The physics and early history of the intergalactic medium

2007 Rep. Prog. Phys. 70 627

(http://iopscience.iop.org/0034-4885/70/4/R02)

IOP PUBLISHING REPORTS ON PROGRESS IN PHYSICS

Rep. Prog. Phys. **70** (2007) 627–657 doi:10.1088/0034-4885/70/4/R02

The physics and early history of the intergalactic medium

Rennan Barkana[1] **and Abraham Loeb**[2]

[1] School of Physics and Astronomy, The Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel
[2] Astronomy Department, Harvard University, 60 Garden Street, Cambridge, MA 02138, USA

E-mail: barkana@wise.tau.ac.il and aloeb@cfa.harvard.edu

Received 20 November 2006, in final form 7 February 2007
Published 16 March 2007
Online at stacks.iop.org/RoPP/70/627

Abstract

The intergalactic medium—the cosmic gas that fills the great spaces between the galaxies—is affected by processes ranging from quantum fluctuations in the very early Universe to radiative emission from newly formed stars. This gives the intergalactic medium a dual role as a powerful probe both of fundamental physics and of astrophysics. The heading of fundamental physics includes conditions in the very early Universe and cosmological parameters that determine the age of the Universe and its matter content. The astrophysics refers to chapters of the long cosmic history of stars and galaxies that are being revealed through the effects of stellar feedback on the cosmic gas. This review describes the physics of the intergalactic medium, focusing on recent theoretical and observational developments in understanding early cosmic history. In particular, the earliest generation of stars is thought to have transformed the Universe from darkness to light and to have had an enormous impact on the intergalactic medium. Half a million years after the Big Bang the Universe was filled with atomic hydrogen. As gravity pulled gas clouds together, the first stars ignited and their radiation turned the surrounding atoms back into free electrons and ions. From the observed spectral absorption signatures of the gas between us and distant sources, we know that the process of reionization pervaded most of space a billion years after the Big Bang, so that only a small fraction of the primordial hydrogen atoms remained between galaxies. Knowing exactly when and how the reionization process happened is a primary goal of cosmologists, because this would tell us when the early stars and black holes formed and in what kinds of galaxies. The distribution and clustering of these galaxies is particularly interesting since it is driven by primordial density fluctuations in the dark matter.

This article was invited by Professor J Silk.

Cosmic reionization is beginning to be understood with the help of theoretical models and computer simulations. Numerical simulations of reionization are computationally challenging, as they require radiative transfer across large cosmological volumes as well as sufficiently high resolution to identify the sources of the ionizing radiation in the infant Universe. Rapid progress in our understanding is expected with additional observational input. A wide variety of instruments currently under design—including large-aperture infrared telescopes on the ground or in space (JWST), and low-frequency radio telescope arrays for the detection of redshifted 21 cm radiation—will probe the first sources of light during an epoch in cosmic history that has been largely unexplored so far. The new observations and the challenges for theoretical models and numerical simulations will motivate intense work in this field over the coming decade.

(Some figures in this article are in colour only in the electronic version)

Contents

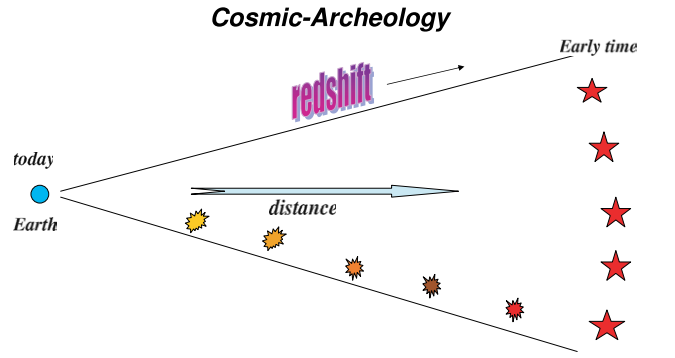

Figure 1. Cosmology is like archaeology. The deeper one looks, the older is the layer that is revealed, owing to the finite propagation speed of light (illustration from Loeb 2006).

1. Introduction

1.1. The cosmic history of the intergalactic medium

When we look at our image reflected off a mirror at a distance of 1 m, we see the way we looked 6.7 ns ago, the light travel time to the mirror and back. If the mirror is spaced 10^{19} cm $\simeq$ 3 pc away, we will see the way we looked twenty one years ago. Light propagates at a finite speed, and so by observing distant regions, we are able to see what the Universe looked like in the past, a light travel time ago (figure 1). The statistical homogeneity of the Universe on large scales guarantees that what we see far away is a fair statistical representation of the conditions that were present in our region of the Universe a long time ago.

This fortunate situation makes cosmology an empirical science. We do not need to guess how the Universe evolved. Using telescopes we can simply see how it appeared at earlier cosmic times. In principle, this allows the entire 13.7 billion year cosmic history of our Universe to be reconstructed by surveying the galaxies and other sources of light to large distances (figure 2). Since a greater distance means a fainter flux from a source of a fixed luminosity, the observation of the earliest sources of light requires the development of sensitive instruments and poses challenges to observers.

To measure distance, astronomers use the characteristic emission patterns of hydrogen and other chemical elements in the spectrum of each galaxy to measure its cosmological redshift z. As the Universe expands, photon wavelengths get stretched as well, so that the spectrum we observe today is shifted from the emitted one by a factor of $(1 + z)$ in wavelength. This then implies that the Universe has expanded by a factor of $(1 + z)$ in linear dimension since that time, and cosmologists can calculate the corresponding distance and cosmic age for the source galaxy. Large telescopes have allowed astronomers to observe faint galaxies that are so far away that we see them more than ten billion years in the past. Thus, we know directly that galaxies were in existence as early as 850 million years after the Big Bang, at a redshift of $z \sim 6.5$ (Hu *et al* 2002, White *et al* 2003, Iye *et al* 2006).

We can in principle image the Universe only if it is transparent. Less than 400 000 years after the Big Bang, the cosmic gas was ionized and the Universe was opaque to Thomson scattering by the free electrons in the dense plasma. Thus, telescopes cannot be used to image the infant Universe at earlier times (or redshifts $\gtrsim 10^3$). The earliest possible image of

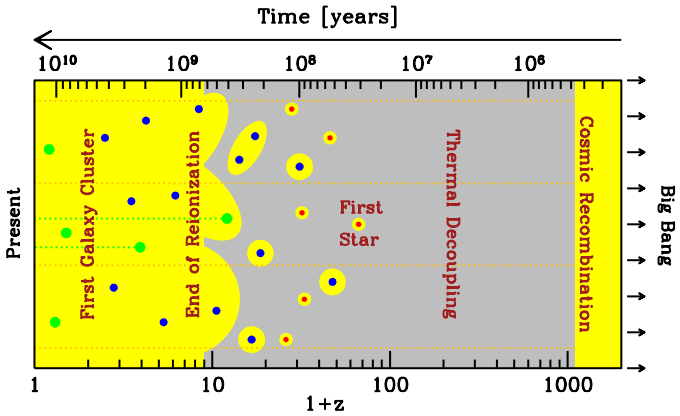

Figure 2. Overview of cosmic history, with the age of the Universe shown on the top axis and the corresponding redshift on the bottom axis. Yellow represents regions where the hydrogen is ionized, and grey, neutral regions. Stars form in galaxies located within dark matter concentrations whose typical mass grows with time, starting with $\sim 10^5 M_\odot$ (red [smallest] circles) for the host of the first star, rising to 10^7–$10^9 M_\odot$ (blue [medium-size] circles) for the sources of reionization and reaching $\sim 10^{12} M_\odot$ (green [large] circles) for present-day galaxies like our own Milky Way. Astronomers probe the evolution of the cosmic gas using the absorption of background light (dotted lines) by atomic hydrogen along the line-of-sight. The classical technique uses absorption by the Lyman-α resonance of hydrogen of the light from bright quasars located within massive galaxies, while a new type of astronomical observation will use the 21 cm line of hydrogen with the cosmic microwave background as the background source (illustration from Barkana (2006)).

the Universe was recorded by the COBE and WMAP satellites (Bennett *et al* 1996, Spergel *et al* 2006), which recorded the temperature distribution of the cosmic microwave background (CMB) on the sky (figure 3).

The CMB, the relic radiation from the fiery beginning of the Universe, is indeed another major probe of observational cosmology. The Universe cools as it expands, so it was initially far denser and hotter than it is today. For hundreds of thousands of years the cosmic gas consisted of a plasma of protons, electrons and a slight mix of light nuclei, sustained by the intense thermal motion of these particles. Just like the plasma in our own Sun, the ancient cosmic plasma emitted and scattered a strong field of visible and ultraviolet photons. As mentioned above, about 400 000 years after the Big Bang the temperature of the Universe dipped for the first time below a few thousand degrees Kelvin. The protons and electrons were now moving slowly enough that they could attract each other and form hydrogen atoms, in a process known as cosmic recombination. With the scattering of the energetic photons now much reduced, the photons continued travelling in straight lines, mostly undisturbed except that cosmic expansion has redshifted them into the microwave regime. The emission temperature of the observed spectrum of these CMB photons is the same in all directions to one part in 100 000 (figure 3), which reveals that conditions were extremely uniform in the early Universe.

It was just before the moment of cosmic recombination (when matter started to dominate in energy density over radiation) that gravity entered the scene. Since that time, gravity has amplified the tiny fluctuations in temperature and density observed in the CMB data (Spergel *et al* 2006). Regions that started out slightly denser than average began to contract because the gravitational forces were also slightly stronger than average in these regions. Eventually, after hundreds of millions of years of contraction, galaxies and the stars within them were able to form. This process, however, would have taken too long to explain the abundance of galaxies today, if it involved only the observed cosmic gas. Instead, gravity is strongly enhanced by the

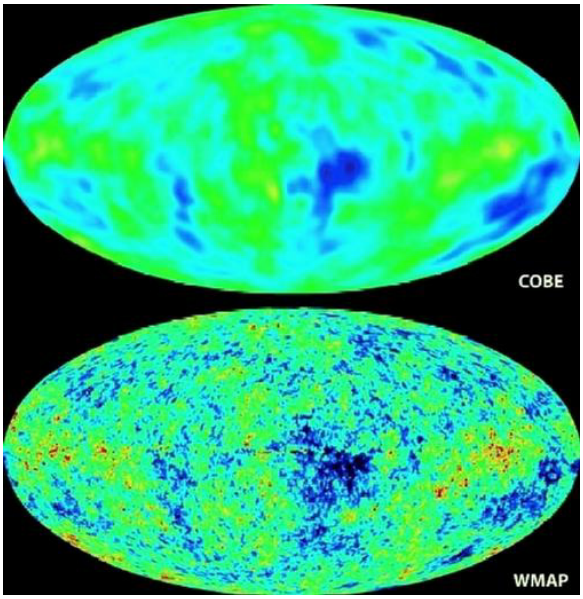

Figure 3. Images of the Universe shortly after it became transparent, taken by the COBE and WMAP satellites (see http://map.gsfc.nasa.gov/ for details). The slight density inhomogeneities in the otherwise uniform Universe imprinted a map of hot and cold spots (shown here as different colours [shades]) in the CMB that is observed today. The existence of these anisotropies was predicted three decades before the technology for taking these images became available, in a number of theoretical papers including Sachs and Wolfe (1967), Rees and Sciama (1968), Silk (1968), Sunyaev and Zeldovich (1970), and Peebles and Yu (1970).

presence of dark matter—an unknown substance that makes up the vast majority (84%) of the cosmic density of matter. The motion of stars and gas around the centres of nearby galaxies indicates that each is surrounded by an extended mass of dark matter, and so dynamically relaxed dark matter concentrations are generally referred to as 'halos'.

According to the standard cosmological model, the dark matter is cold (abbreviated as CDM), i.e. it behaves as a collection of collisionless particles that started out at matter domination with negligible thermal velocities and have evolved exclusively under gravitational forces. The model explains how both individual galaxies and the large-scale patterns in their distribution originated from the small initial density fluctuations. On the largest scales, observations of the present galaxy distribution have indeed found the same statistical patterns as seen in the CMB, enhanced as expected by billions of years of gravitational evolution (Eisenstein *et al* 2005, Cole *et al* 2005). On smaller scales, the model describes how regions that were denser than average collapsed due to their enhanced gravity and eventually formed gravitationally bound halos, first on small spatial scales and later on larger ones. In this hierarchical model of galaxy formation, the small galaxies formed first and then merged or accreted gas to form larger galaxies. At each snapshot of this cosmic evolution, the abundance of collapsed halos, whose masses are dominated by dark matter, can be computed from the initial conditions using numerical simulations. The common understanding of galaxy formation is based on the notion that stars formed out of the gas that cooled and subsequently condensed to high densities in the cores of some of these halos.

Gravity thus explains how some gas is pulled into the deep potential wells within dark matter halos and forms the galaxies. One might naively expect that the gas outside halos

would remain mostly undisturbed. However, observations show that it has not remained neutral (i.e. in atomic form) until the present. To learn about diffuse gas pervading the space outside and between galaxies (referred to as the intergalactic medium or IGM (Field 1972)), astronomers study its absorption signatures in the spectra of distant quasars, the brightest long-lived astronomical objects. Quasars' great luminosities are believed to be powered by gas accretion onto black holes weighing up to a few billion times the mass of the Sun that are situated in the dense centres of massive galaxies. As the surrounding gas spirals in towards the black hole sink, its excess rotation yields viscous dissipation of heat that makes the gas glow brightly into space, creating a luminous source visible from afar.

The Lyman-alpha (Lyα) resonance line of hydrogen at a wavelength of 1216 Å has been widely used to trace hydrogen gas through its absorption of quasar light (Gunn and Peterson 1965). The expansion of the Universe gives this tool an important advantage common to all spectral absorption probes. Since the wavelength of every photon is stretched as the Universe expands, the rest-frame absorption at 1216 Å by a gas element at redshift z is observed today at a wavelength of $1216(1 + z)$ Å. The absorptions of the different gas elements along the line-of-sight are therefore distributed over a broad range of wavelengths, making it possible to measure the distribution of intergalactic hydrogen.

Lyα absorption shows that the IGM has been a hot plasma at least from a cosmic age of 850 million years ($z \sim 6.5$) until today (White *et al* 2003). Thus, the hydrogen must have been ionized for a second time after it became neutral at cosmic recombination. Radiation from the first generations of stars is a plausible source for the ionizing photons that transformed the IGM.

Absorption at the Lyα resonance is so strong that it becomes difficult to use as observations approach the reionization epoch where the density of neutral hydrogen becomes high (White *et al* 2003). As described below, cosmologists believe that a different method, termed '21 cm cosmology', will allow us to measure how the reionization process developed over time and to test theoretical predictions of the properties of the earliest galaxies (Barkana and Loeb 2001, Loeb 2006).

1.2. The expanding Universe

The modern physical description of the Universe as a whole can be traced back to Einstein, who argued theoretically for the so-called 'cosmological principle': that the distribution of matter and energy must be homogeneous and isotropic on the largest scales. Today isotropy is well established (see the review by Wu *et al* (1999)) for the distribution of faint radio sources, optically selected galaxies, the x-ray background, and most importantly the cosmic microwave background (hereafter, CMB; see, e.g. Bennett *et al* (1996)). The constraints on homogeneity are less strict, but a cosmological model in which the Universe is isotropic but significantly inhomogeneous in spherical shells around our special location is also excluded (Goodman *et al* 1995).

Cosmological solutions of General Relativity predict the evolution of the cosmic scale factor $a(t)$, defined so that the physical separation between observers at rest (with respect to the general expansion) increases with time in proportion to $a(t)$. A given observer sees a nearby observer at physical distance D receding at the Hubble velocity $H(t)D$, where the Hubble constant at time t is $H(t) = \mathrm{d}\,a(t)/\mathrm{d}t$. Light emitted by a source at time t with an emitted wavelength $\lambda_{\rm emit}$ is observed at $t = 0$ with a redshifted wavelength $\lambda_{\rm obs}$; the source redshift z, defined from $\lambda_{\rm obs} = \lambda_{\rm emit} \times (1 + z)$, is related to the cosmic scale factor at the time of emission by $z = 1/a(t) - 1$, where we set $a(t = 0) \equiv 1$ for convenience.

The Einstein field equations of General Relativity yield the Friedmann equation (e.g. Weinberg (1972) and Kolb and Turner (1990))

$$H^2(t) = \frac{8\pi G}{3}\rho - \frac{k}{a^2}, \tag{1}$$

which relates the expansion of the Universe to its matter-energy content. The constant k determines the geometry of the Universe; it is positive in a closed Universe, zero in a flat Universe and negative in an open Universe. For each component of the energy density ρ, with an equation of state $p = p(\rho)$, the density ρ varies with $a(t)$ according to the equation of energy conservation

$$\mathrm{d}(\rho R^3) = -p\mathrm{d}(R^3). \tag{2}$$

With the critical density

$$\rho_C(t) \equiv \frac{3H^2(t)}{8\pi G} \tag{3}$$

defined as the density needed for $k = 0$, we define the ratio of the total density to the critical density as

$$\Omega \equiv \frac{\rho}{\rho_C}. \tag{4}$$

With Ω_m, Ω_Λ and Ω_r denoting the present contributions to Ω from matter (including cold dark matter as well as a contribution Ω_b from baryons), vacuum density (cosmological constant) and radiation, respectively, the Friedmann equation becomes

$$\frac{H(t)}{H_0} = \left[\frac{\Omega_\mathrm{m}}{a^3} + \Omega_\Lambda + \frac{\Omega_\mathrm{r}}{a^4} + \frac{\Omega_k}{a^2}\right], \tag{5}$$

where we define H_0 and $\Omega_0 = \Omega_\mathrm{m} + \Omega_\Lambda + \Omega_\mathrm{r}$ to be the present values of H and Ω, respectively, and we let

$$\Omega_k \equiv -\frac{k}{H_0^2} = 1 - \Omega_\mathrm{m}. \tag{6}$$

In the particularly simple Einstein–de Sitter model ($\Omega_\mathrm{m} = 1$, $\Omega_\Lambda = \Omega_\mathrm{r} = \Omega_k = 0$), the scale factor varies as $a(t) \propto t^{2/3}$. Even models with non-zero Ω_Λ or Ω_k approach the Einstein–de Sitter behaviour at high redshift, i.e. when $(1+z) \gg |\Omega_\mathrm{m}^{-1} - 1|$ (as long as Ω_r can be neglected). In this high-z regime the age of the Universe is

$$t \approx \frac{2}{3H_0\sqrt{\Omega_\mathrm{m}}}\,(1+z)^{-3/2}\,. \tag{7}$$

Recent observations confine the standard set of cosmological parameters to a relatively narrow range. In particular, we seem to live in a ΛCDM cosmology (with Ω_k so small that it is usually assumed to equal zero) with an approximately scale-invariant primordial power spectrum of density fluctuations, i.e. $n \approx 1$ where the initial power spectrum is $P(k) \propto k^n$ in terms of the wavenumber k. Also, the Hubble constant today is written as $H_0 = 100h$ km s^{-1} Mpc^{-1} in terms of h, and the overall normalization of the power spectrum is specified in terms of σ_8, the root-mean-square amplitude of mass fluctuations in spheres of radius 8 h^{-1} Mpc. For example, the best-fit cosmological parameters matching the three-year WMAP data together with large-scale gravitational lensing observations (Spergel *et al* 2006) are $\sigma_8 = 0.826$, $n = 0.953$, $h = 0.687$, $\Omega_\mathrm{m} = 0.299$, $\Omega_\Lambda = 0.701$ and $\Omega_\mathrm{b} = 0.0478$. A different cosmological parameter set, also based on the CMB data together with other large-scale structure measurements is (Viel *et al* 2006) $\sigma_8 = 0.785$, $n = 0.957$, $h = 0.723$ $\Omega_\mathrm{m} = 0.253$, $\Omega_\Lambda = 0.747$ and $\Omega_\mathrm{b} = 0.0425$. The difference between these two parameter sets roughly represents current $1-\sigma$ parameter uncertainties.

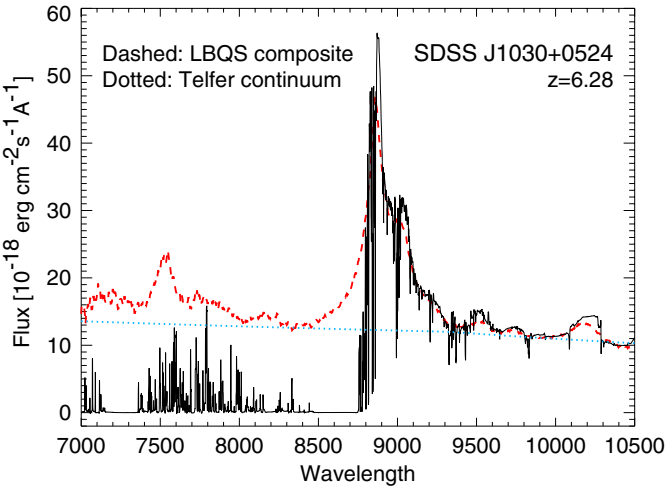

Figure 4. Using Lyα absorption in quasar spectra to probe the ionization state of the IGM. This figure from White *et al* (2003) shows the observed spectrum of a $z = 6.28$ quasar (solid curve), and the expected unabsorbed emission (dashed curve), based on an average over many quasars seen at lower redshifts. The unabsorbed emission is a sum of smooth emission (the 'continuum', dotted curve) plus emission features from atomic resonances ('emission lines').

2. Atomic physics of the intergalactic medium

2.1. Radiative absorption and scattering

Resonant Lyα absorption has thus far proved to be the best probe of the state of the IGM. The optical depth to absorption by a uniform intergalactic medium is (Gunn and Peterson 1965)

$$\tau_s = \frac{\pi e^2 f_\alpha \lambda_\alpha n_{\rm HI}(z)}{m_e c H(z)} \approx 6.45 \times 10^5 x_{\rm HI} \left(\frac{\Omega_{\rm b} h}{0.0315}\right) \left(\frac{\Omega_{\rm m}}{0.3}\right)^{-1/2} \left(\frac{1+z}{10}\right)^{3/2}, \quad (8)$$

where $H \approx 100h\,{\rm km\,s^{-1}\,Mpc^{-1}}\Omega_{\rm m}^{1/2}(1+z)^{3/2}$ is the Hubble parameter at redshift z, $f_\alpha = 0.4162$ and $\lambda_\alpha = 1216$ Å are the oscillator strength and the wavelength of the Lyα transition; $n_{\rm HI}(z)$ is the neutral hydrogen density at z (assuming primordial abundances) $\Omega_{\rm m}$ and $\Omega_{\rm b}$ are the present-day density parameters of all matter and of baryons, respectively and $x_{\rm HI}$ is the average fraction of neutral hydrogen. In the second equality we have implicitly considered high redshifts.

Lyα absorption is thus highly sensitive to the presence of even trace amounts of neutral hydrogen. The lack of full absorption in quasar spectra then implies that the IGM has been very highly ionized during much of the history of the Universe, from the present out to high redshift. At redshifts approaching six, however, the optical depth increases and the observed absorption becomes very strong. An example of this is shown in figure 4, taken from White *et al* (2003), where an observed quasar spectrum is compared with the unabsorbed expectation for the same quasar. The prominent Lyα emission line, which is produced by radiating hot gas near the quasar itself, is centred at a wavelength of 8850 Å, which for the redshift (6.28) of this quasar corresponds to a rest-frame 1216 Å. Above this wavelength, the original emitted quasar spectrum is seen, since photons emitted with wavelengths higher than 1216 Å redshift to higher wavelengths during their journey towards us and never encounter resonance lines of hydrogen atoms. Shorter-wavelength photons, however, redshift until they hit the local 1216 Å and are then absorbed by any existing hydrogen atoms. The difference between the

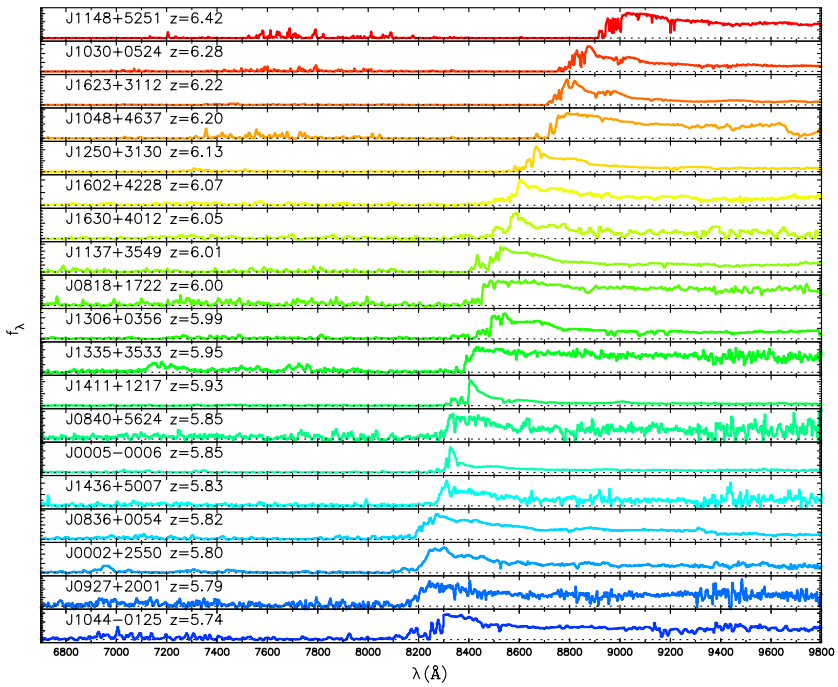

Figure 5. Spectra of 19 quasars with redshifts $5.74 < z < 6.42$ from the *Sloan Digital Sky Survey*, taken from Fan *et al* (2005). For some of the highest-redshift quasars, the spectrum shows no transmitted flux shortwards of the Lyα wavelength at the quasar redshift (the so-called 'Gunn–Peterson trough'), indicating a non-negligible neutral fraction in the IGM.

unabsorbed expectation and the actual observed spectrum can be used to measure the amount of absorption and thus to infer the atomic hydrogen density. In this particular quasar, this difference is very large (i.e. the observed flux is near zero) just to the blue of the Lyα emission line.

Several quasars beyond $z \sim 6.1$ show in their spectra such a Gunn–Peterson trough, a blank spectral region at wavelengths shorter than Lyα at the quasar redshift (figure 5). The detection of Gunn–Peterson troughs indicates a rapid change (Fan *et al* 2002, 2006b, White *et al* 2003) in the neutral content of the IGM at $z \sim 6$ and hence a rapid change in the intensity of the background ionizing flux. However, even a small atomic hydrogen fraction of $\sim 10^{-3}$ would still produce nearly complete Lyα absorption. The lower absorption efficiencies of higher Lyman transitions allow much stronger constraints to be derived from observations of Lyβ and Lyγ absorption (Fan *et al* 2006a).

While only resonant Lyα absorption is important at moderate redshifts, the damping wing of the Lyα line plays a significant role when neutral fractions of order unity are considered at $z \gtrsim 6$. The scattering cross-section of the Lyα resonance line by neutral hydrogen is given by (section 23 of Peebles 1993)

$$\sigma_\alpha(\nu) = \frac{3\lambda_\alpha^2 \Lambda_\alpha^2}{8\pi} \frac{(\nu/\nu_\alpha)^4}{4\pi^2(\nu-\nu_\alpha)^2 + (\Lambda_\alpha^2/4)(\nu/\nu_\alpha)^6}, \tag{9}$$

where $\Lambda_\alpha = (8\pi^2 e^2 f_\alpha / 3 m_e c \lambda_\alpha^2) = 6.25 \times 10^8\,\mathrm{s}^{-1}$ is the Lyα ($2\mathrm{p} \rightarrow 1\mathrm{s}$) decay rate, $f_\alpha = 0.4162$ is the oscillator strength, and $\lambda_\alpha = 1216\,\text{Å}$ and $\nu_\alpha = (c/\lambda_\alpha) = 2.47 \times 10^{15}\,\mathrm{Hz}$ are the wavelength and frequency of the Lyα line. The term in the numerator is responsible for the classical Rayleigh scattering.

While reionization is a quite inhomogeneous process (as we discuss below), we consider here a simple illustrative case of instantaneous reionization. Consider a source at a redshift z_s beyond the redshift of reionization, $z_{\rm reion}$, and the corresponding scattering optical depth of a uniform, neutral IGM of hydrogen density $n_{\rm H,0}(1+z)^3$ between the source and the reionization redshift. The optical depth is a function of the observed wavelength $\lambda_{\rm obs}$,

$$\tau(\lambda_{\rm obs}) = \int_{z_{\rm reion}}^{z_s} {\rm d}z \, \frac{c{\rm d}t}{{\rm d}z} \, n_{\rm H,0}(1+z)^3 \sigma_\alpha \left[\nu_{\rm obs}(1+z)\right], \tag{10}$$

where $\nu_{\rm obs} = c/\lambda_{\rm obs}$ and

$$\frac{{\rm d}t}{{\rm d}z} = [(1+z)H(z)]^{-1} = H_0^{-1} \times \left[\Omega_{\rm m}(1+z)^5 + \Omega_\Lambda (1+z)^2 + (1-\Omega_{\rm m}-\Omega_\Lambda)(1+z)^4\right]^{-1/2}. \tag{11}$$

At wavelengths longer than Lyα at the source, the optical depth obtains a small value; these photons redshift away from the line centre along its red wing and never resonate with the line core on their way to the observer. Considering only the regime in which $|\nu - \nu_\alpha| \gg \Lambda_\alpha$, we may ignore the second term in the denominator of equation (9). This leads to an analytical result for the red damping wing of the Gunn–Peterson trough (Miralda-Escudé 1998).

At wavelengths shorter than 912 Å, the photons are absorbed when they photoionize atoms of hydrogen or helium. The detailed absorption cross-sections are summarized in Barkana and Loeb (2001). For rough estimates, the average photoionization cross-section for a mixture of hydrogen and helium with cosmic abundances can be approximated in the range of $54 < h\nu \lesssim 10^3$ eV as $\sigma_{bf} \approx \sigma_0(\nu/\nu_{\rm H,0})^{-3}$, where $\sigma_0 \approx 6 \times 10^{-17}$ cm^2 (Miralda-Escudé 2000). Before reionization, in a Universe filled with atoms, this yields a very high optical depth for the absorption of ionizing photons with tens of eVs in energy. The bound-free optical depth only becomes of order unity in the extreme ultraviolet (UV) to soft x-rays, around $h\nu \sim 0.1$ keV, a regime which is unfortunately difficult to observe due to galactic absorption (Miralda-Escudé 2000).

2.2. The spin temperature of the 21 cm transition of hydrogen

The ground state of hydrogen exhibits hyperfine splitting involving the spins of the proton and the electron. The state with parallel spins (the triplet state) has a slightly higher energy than the state with anti-parallel spins (the singlet state). The 21 cm line associated with the spin-flip transition from the triplet to the singlet state is often used to detect neutral hydrogen in the local Universe. At high redshift, the occurrence of a neutral pre-reionization IGM offers the prospect of detecting the first sources of radiation and probing the reionization era by mapping the 21 cm emission from neutral regions. While its energy density is estimated to be only a 1% correction to that of the CMB, the redshifted 21 cm emission should display angular structure as well as frequency structure due to inhomogeneities in the gas density field (Hogan and Rees 1979, Scott and Rees 1990), hydrogen ionized fraction and spin temperature (Madau *et al* 1997). Indeed, a full mapping of the distribution of H I as a function of redshift is possible in principle.

The basic physics of the hydrogen spin transition is determined as follows (for a more detailed treatment, see Madau *et al* (1997) and Furlanetto *et al* (2006)). The ground-state hyperfine levels of hydrogen tend to thermalize with the CMB background, making the IGM unobservable. If other processes shift the hyperfine level populations away from thermal equilibrium, then the gas becomes observable against the CMB in emission or in absorption. The relative occupancy of the spin levels is usually described in terms of the hydrogen spin

temperature T_S, defined by

$$\frac{n_1}{n_0} = 3 \exp\left\{-\frac{T_*}{T_S}\right\} , \tag{12}$$

where n_0 and n_1 refer, respectively, to the singlet and triplet hyperfine levels in the atomic ground state ($n = 1$), and $T_* = 0.068\,\mathrm{K}$ is defined by $k_B T_* = E_{21}$, where the energy of the 21 cm transition is $E_{21} = 5.9 \times 10^{-6}\,\mathrm{eV}$, corresponding to a frequency of 1420 MHz. In the presence of the CMB alone, the spin states reach thermal equilibrium with $T_S = T_{CMB} = 2.725(1+z)\,\mathrm{K}$ on a time-scale of $T_*/(T_{CMB} A_{10}) \simeq 3 \times 10^5 (1+z)^{-1}\,\mathrm{yr}$, where $A_{10} = 2.87 \times 10^{-15}\,\mathrm{s}^{-1}$ is the spontaneous decay rate of the hyperfine transition. This time-scale is much shorter than the age of the Universe at all redshifts after cosmological recombination.

The IGM is observable when the kinetic temperature T_k of the gas differs from T_{CMB} and an effective mechanism couples T_S to T_k. Collisional de-excitation of the triplet level (Purcell and Field 1956) dominates at very high redshift, when the gas density (and thus the collision rate) is still high, but once a significant galaxy population forms in the Universe, the spin temperature is affected also by an indirect mechanism that acts through the scattering of Lyα photons (Wouthuysen 1952, Field 1958). Continuum UV photons produced by early radiation sources redshift by the Hubble expansion into the local Lyα line at a lower redshift. These photons mix the spin states via the Wouthuysen–Field process whereby an atom initially in the $n = 1$ state absorbs a Lyα photon, and the spontaneous decay which returns it from $n = 2$ to $n = 1$ can result in a final spin state which is different from the initial one. Since the neutral IGM is highly opaque to resonant scattering, and the Lyα photons receive Doppler kicks in each scattering, the shape of the radiation spectrum near Lyα is determined by T_k (Field (1959b), but see also Hirata (2006) and Chuzhoy and Shapiro (2006)), and the resulting spin temperature (assuming $T_S \gg T_*$) is then a weighted average of T_k and T_{CMB} (Field 1959a):

$$T_S = \frac{T_{CMB} T_k (1 + x_{tot})}{T_k + T_{CMB} x_{tot}}, \tag{13}$$

where $x_{tot} = x_\alpha + x_c$ is the sum of the radiative and collisional threshold parameters. These parameters are

$$x_\alpha = \frac{P_{10} T_\star}{A_{10} T_{CMB}} \tag{14}$$

and

$$x_c = \frac{4\kappa_{1-0}(T_k)\, n_H T_\star}{3 A_{10} T_{CMB}} , \tag{15}$$

where P_{10} is the indirect de-excitation rate of the triplet $n = 1$ state via the Wouthuysen–Field process, related to the total scattering rate P_α of Lyα photons by $P_{10} = 4P_\alpha/27$ (Field 1958). Also, the atomic coefficient $\kappa_{1-0}(T_k)$ is tabulated as a function of T_k (Allison and Dalgarno 1969; Zygelman 2005). Collisions can be particularly important in collapsed halos (Iliev *et al* 2003). Note that we have adopted the modified notation (i.e. in terms of x_α and x_c) of Barkana and Loeb (2005b). The coupling of the spin temperature to the gas temperature becomes substantial when $x_{tot} \gtrsim 1$; in particular, $x_\alpha = 1$ defines the thermalization rate (Madau *et al* 1997) of P_α:

$$P_{th} \equiv \frac{27 A_{10} T_{CMB}}{4 T_*} \simeq 7.6 \times 10^{-12} \left(\frac{1+z}{10}\right) \mathrm{s}^{-1}. \tag{16}$$

A patch of neutral hydrogen at the mean density and with a uniform T_S produces (after correcting for stimulated emission) an optical depth at a present-day (observed) wavelength of $21(1+z)$ cm,

$$\tau(z) = 9.0 \times 10^{-3} \left(\frac{T_{CMB}}{T_S}\right) \left(\frac{\Omega_b h}{0.03}\right) \left(\frac{\Omega_m}{0.3}\right)^{-1/2} \left(\frac{1+z}{10}\right)^{1/2}, \tag{17}$$

assuming a high redshift $z \gg 1$. The observed spectral intensity I_ν relative to the CMB at a frequency ν is measured by radio astronomers as an effective brightness temperature T_b of blackbody emission at this frequency, defined using the Rayleigh–Jeans limit of the Planck radiation formula: $I_\nu \equiv 2k_B T_b \nu^2/c^2$.

The brightness temperature through the IGM is $T_b = T_{CMB}e^{-\tau} + T_S(1 - e^{-\tau})$, so the observed differential antenna temperature of this region relative to the CMB is (Madau *et al* (1997), with the Ω_m dependence added)

$$\begin{aligned} T_b &= (1+z)^{-1}(T_S - T_{CMB})(1 - e^{-\tau}) \\ &\simeq 28\,\text{mK} \left(\frac{\Omega_b h}{0.033}\right) \left(\frac{\Omega_m}{0.27}\right)^{-1/2} \left(\frac{1+z}{10}\right)^{1/2} \left(\frac{T_S - T_{CMB}}{T_S}\right), \end{aligned} \tag{18}$$

where $\tau \ll 1$ is assumed and T_b has been redshifted to redshift zero. Note that the combination that appears in T_b is

$$\frac{T_S - T_{CMB}}{T_S} = \frac{x_{tot}}{1 + x_{tot}} \left(1 - \frac{T_{CMB}}{T_k}\right). \tag{19}$$

In overdense regions, the observed T_b is proportional to the overdensity, and in partially ionized regions T_b is proportional to the neutral fraction. Also, if $T_S \gg T_{CMB}$ then the IGM is observed in emission at a level that is independent of T_S. On the other hand, if $T_S \ll T_{CMB}$ then the IGM is observed in absorption at a level that is enhanced by a factor of T_{CMB}/T_S. As a result, a number of cosmic events are expected to leave observable signatures in the redshifted 21 cm line, as discussed below in further detail.

3. Galaxy formation

3.1. Growth of linear perturbations

As noted in section 1, observations of the CMB show that the Universe at cosmic recombination (redshift $z \sim 10^3$) was remarkably uniform apart from spatial fluctuations in the energy density and in the gravitational potential of roughly one part in 10^5. The primordial inhomogeneities in the density distribution grew over time and eventually led to the formation of galaxies as well as galaxy clusters and large-scale structure. In the early stages of this growth, as long as the density fluctuations on the relevant scales were much smaller than unity, their evolution can be understood with a linear perturbation analysis.

Different physical processes contributed to the perturbation growth (e.g. Peebles (1980) and Ma and Bertschinger (1995)). In the absence of other influences, gravitational forces due to density perturbations imprinted by inflation would have driven parallel perturbation growth in the dark matter, baryons and photons. However, since the photon sound speed is of the order of the speed of light, the radiation pressure produced sound waves on a scale of the order of the horizon and suppressed sub-horizon perturbations in the photon density. The baryonic pressure similarly suppressed perturbations in the gas below the (much smaller) so-called baryonic *Jeans* scale. Since the formation of hydrogen at recombination had decoupled the cosmic gas from its mechanical drag on the CMB, the baryons subsequently began to fall into the pre-existing gravitational potential wells of the dark matter.

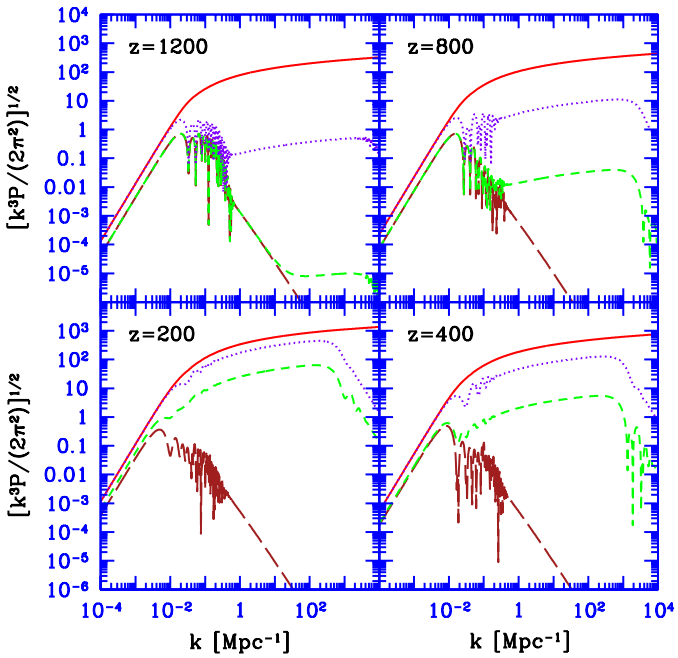

Figure 6. Power spectra of density and temperature fluctuations versus comoving wavenumber, at redshifts 1200, 800, 400 and 200, adopted from Naoz and Barkana (2005). We consider fluctuations in the CDM density (solid curves), baryon density (dotted curves), baryon temperature (short-dashed curves) and photon temperature (long-dashed curves).

Spatial fluctuations developed in the gas temperature as well as in the gas density. Both the baryons and the dark matter were affected on small scales by the temperature fluctuations through the gas pressure. Compton heating due to scattering of the residual free electrons (constituting a fraction $\sim 10^{-4}$) with the CMB photons remained effective, keeping the gas temperature fluctuations tied to the photon-temperature fluctuations, even for a time after recombination. The growth of linear perturbations can be calculated with the standard CMBFAST code (Seljak and Zaldarriaga (1996), see http://www.cmbfast.org), after a modification to account for the fact that the speed of sound of the gas also fluctuates spatially (Yamamoto *et al* 1997, 1998, Naoz and Barkana 2005).

The magnitude of the fluctuations in the CDM and baryon densities, and in the baryon and photon temperatures, is shown in figure 6, in terms of the dimensionless combination $[k^3 P(k)/(2\pi^2)]^{1/2}$, where $P(k)$ is the corresponding power spectrum of fluctuations in terms of the comoving wavenumber k of each Fourier mode. After recombination, two main drivers affect the baryon density and temperature fluctuations, namely, the thermalization with the CMB and the gravitational force that attracts the baryons to the dark matter potential wells. As shown in the figure, the density perturbations in all species grow together on scales where gravity is unopposed, outside the horizon (i.e. at $k \lesssim 0.01\,\mathrm{Mpc}^{-1}$ at $z \sim 1000$). At $z = 1200$ the perturbations in the baryon-photon fluid oscillate as acoustic waves on scales of the order of the sound horizon ($k \sim 0.01\,\mathrm{Mpc}^{-1}$), while smaller-scale perturbations in both the photons and the baryons are damped by photon diffusion (Silk damping) and the drag of the diffusing photons on the baryons. On sufficiently small scales the power spectra of baryon density and temperature roughly assume the shape of the dark matter fluctuations

(except for the gas-pressure cutoff at the very smallest scales), due to the effect of gravitational attraction on the baryon density and of the resulting adiabatic expansion on the gas temperature. After the mechanical coupling of the baryons to the photons ends at $z \sim 1000$, the baryon density perturbations gradually grow towards the dark matter perturbations because of gravity. Similarly, after the thermal coupling ends at $z \sim 200$, the baryon temperature fluctuations are driven by adiabatic expansion towards a value of 2/3 of the baryon density fluctuations. As the figure shows, by $z = 200$ the baryon infall into the dark matter potentials is well advanced and adiabatic expansion is becoming increasingly important in setting the baryon temperature.

3.2. Formation of the first stars

Theoretical expectations for the properties of the first galaxies (Barkana and Loeb 2001) are based on the standard cosmological model outlined in section 1. The formation of the first bound objects marked the central milestone in the transition from the initial simplicity (discussed in the previous subsection) to the present-day complexity. Stars and quasars output copious radiation and also produced explosions and outflows that brought into the IGM chemical products from stellar nucleosynthesis and enhanced magnetic fields. However, the formation of the very first stars, in a Universe that had not yet suffered such feedback, remains a well-specified problem for theorists.

Stars form when huge amounts of matter collapse to enormous densities. However, the process can be stopped if the pressure exerted by the hot intergalactic gas prevents outlying gas from falling into dark matter concentrations. As the gas falls into a dark matter halo, it forms shocks due to converging supersonic flows and in the process heats up and can only collapse further by first radiating its energy away. This restricts this process of collapse to very large clumps of dark matter that are around 100 000 times the mass of the Sun. Inside these clumps, the shocked gas loses energy by emitting radiation from excited molecular hydrogen that formed naturally within the primordial gas mixture of hydrogen and helium (Peebles 1984, Haiman *et al* 1996, Tegmark *et al* 1997).

The first stars are expected to have been quite different from the stars that form today in the Milky Way. The higher pressure within the primordial gas due to the presence of fewer cooling agents suggests that fragmentation only occurred into relatively large units, in which gravity could overcome the pressure. Due to the lack of carbon, nitrogen and oxygen—elements that would normally dominate the nuclear energy production in modern massive stars—the first stars must have condensed to extremely high densities and temperatures before nuclear reactions were able to heat the gas and balance gravity. These unusually massive stars produced high luminosities of UV photons, but their nuclear fuel was exhausted after 2–3 million years, resulting in a huge supernova or in collapse to a massive black hole. For additional details about the properties of the first stars, see the review by Bromm and Larson (2004).

Advances in computing power have made possible detailed numerical simulations of how the first stars formed (Bromm *et al* 2002, Abel *et al* 2002, Bromm and Loeb 2004, Yoshida *et al* 2006). These simulations begin in the early Universe, in which dark matter and gas are distributed uniformly, apart from tiny variations in density and temperature that are statistically distributed according to the patterns observed in the CMB. In order to span the vast range of scales needed to simulate an individual star within a cosmological context, the latest code (Yoshida *et al* 2006) follows a box 0.3 Mpc in length and zooms in repeatedly on the densest part of the first collapsing cloud that is found within the simulated volume. The simulation follows gravity, hydrodynamics, and chemical processes in the primordial gas, and resolves a scale 10 orders of magnitudes smaller than that of the simulated box. While the resolved scale is still three orders of magnitude larger than the size of the Sun, these simulations have

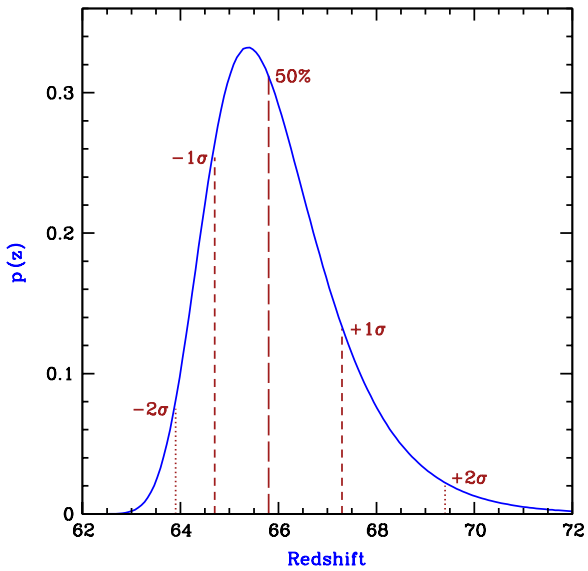

Figure 7. The probability of finding the first observable star in the Universe, as a function of redshift (from Naoz *et al* 2006). Vertical lines show the median redshift as well as the central $1-\sigma$ (68%) range and $2-\sigma$ (95%) range. This result is based on an analytical model that accounts for the size of the Universe and for three essential ingredients: the light travel time from distant galaxies, Poisson and density fluctuations on all scales and the effect of very early cosmic history on galaxy formation.

established that the first stars formed within halos containing $\sim 10^5 M_\odot$ in total mass, and indicate that the first stars most likely weighed $\sim 100 M_\odot$ each.

To estimate *when* the first stars formed we must remember that the first 100 000 solar mass halos collapsed in regions that happened to have a particularly high density enhancement very early on. There was initially only a small abundance of such regions in the entire Universe, so a simulation that is limited to a small volume is unlikely to find such halos until much later. Simulating the entire Universe is well beyond the capabilities of current simulations, but analytical models predict that the first observable star in the Universe (figure 7) probably formed 30 million years after the Big Bang (Naoz *et al* 2006), less than a quarter of one percent of the Universe's total age of 13.7 billion years.

Although stars were extremely rare at first, gravitational collapse increased the abundance of galactic halos and star formation sites with time (figure 2). Radiation from the first stars is expected to have eventually dissociated all the molecular hydrogen in the intergalactic medium, leading to the domination of a second generation of larger galaxies where the gas cooled via radiative transitions in atomic hydrogen and helium (Haiman *et al* 1997). Atomic cooling occurred in halos of mass above $\sim 10^8 M_\odot$, in which the infalling gas was heated above 10 000 K and became ionized. The first galaxies to form through atomic cooling are expected to have formed around redshift 45 (Naoz *et al* 2006), and such galaxies were likely the main sites of star formation by the time reionization began in earnest. As the IGM was heated above 10 000 K by reionization, its pressure jumped and prevented the gas from accreting into newly forming halos below $\sim 10^9 M_\odot$ (Rees 1986). The first Milky-Way-sized halo $M = 10^{12} M_\odot$ is predicted to have formed 400 million years after the Big Bang (Naoz *et al* 2006), but such halos have become typical galactic hosts only in the last five billion years.

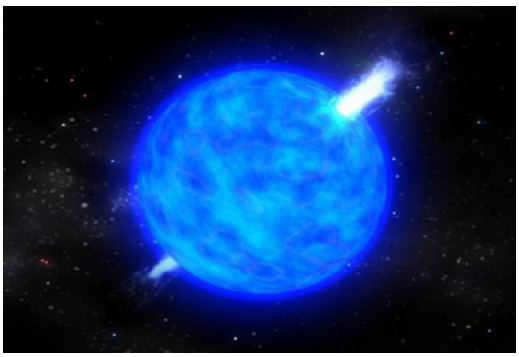

Figure 8. Illustration of a long-duration gamma-ray burst in the popular 'collapsar' model (Zhang *et al* 2003). The collapse of the core of a massive star (which lost its hydrogen envelope) to a black hole generates two opposite jets moving out at a speed close to the speed of light. The jets drill a hole in the star and shine brightly towards an observer who happens to be located within with the collimation cones of the jets. The jets emanating from a single massive star are so bright that they can be seen across the Universe out to the epoch when the first stars formed. Upcoming observations by the *Swift* satellite will have the sensitivity to reveal whether Pop III stars served as progenitors of gamma-ray bursts (for more information see swift.gsfc.nasa.gov/).

3.3. Gamma-ray bursts: probing the first stars one star at a time

Gamma-ray bursts (GRBs) are believed to originate in compact remnants (neutron stars or black holes) of massive stars. Their high luminosities make them detectable out to the edge of the visible Universe (Lamb and Reichart 2000, Ciardi and Loeb 2000). GRBs offer the opportunity to detect the most distant (and hence earliest) population of massive stars, the so-called Population III (or Pop III), one star at a time (figure 8). In the hierarchical assembly process of halos that are dominated by cold dark matter (CDM), the first galaxies should have had lower masses (and lower stellar luminosities) than their more recent counterparts. Consequently, the characteristic luminosity of galaxies or quasars is expected to decline with increasing redshift (but note that star formation may favour more massive halos due to feedback effects, similar to the 'downsizing' observed at low redshift; e.g. Bundy *et al* (2006)). GRB afterglows, which already produce a peak flux comparable to that of quasars or starburst galaxies at $z \sim 1$–2, are therefore expected to outshine any competing source at the highest redshifts, when the first dwarf galaxies formed in the Universe.

GRBs, the electromagnetically brightest explosions in the Universe, should be detectable out to redshifts $z > 10$. High-redshift GRBs can be identified through infrared photometry, based on the Lyα break induced by absorption of their spectrum at wavelengths below $1.216\,\mu\mathrm{m}\,[(1+z)/10]$. Follow-up spectroscopy of high-redshift candidates can then be performed on a 10 m class telescope. GRB afterglows offer the opportunity to detect stars as well as to probe the metal enrichment level (Furlanetto and Loeb 2003) of the intervening IGM. Recently, the ongoing *Swift* mission (Gehrels *et al* 2004) has detected a GRB originating at $z \simeq 6.3$ (e.g. Haislip *et al* (2006)), thus demonstrating the viability of GRBs as probes of the early Universe.

Another advantage of GRBs is that the GRB afterglow flux at a given observed time lag after the γ-ray trigger is not expected to fade significantly with increasing redshift, since higher redshifts translate to earlier times in the source frame, during which the afterglow is intrinsically brighter (Ciardi and Loeb 2000). For standard afterglow lightcurves and spectra, the increase in the luminosity distance with redshift is compensated by this *cosmological time-stretching* effect (Barkana and Loeb 2004a) as shown in figure 9.

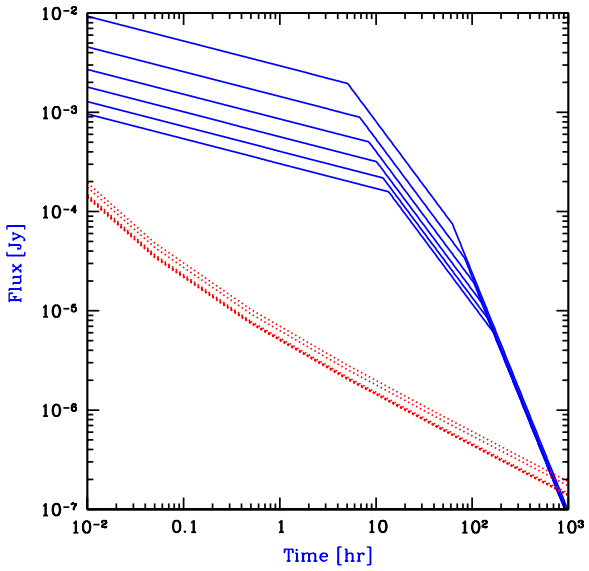

Figure 9. GRB afterglow flux as a function of time since the γ-ray trigger in the observer frame (taken from Barkana and Loeb (2004a)). The flux (solid curves) is calculated at the redshifted Lyα wavelength. The dotted curves show the planned detection threshold for the James Webb Space Telescope (JWST), assuming a spectral resolution $R = 5000$ with the near infrared spectrometer, a signal to noise ratio of 5 per spectral resolution element, and an exposure time equal to 20% of the time since the GRB explosion (see http://www.ngst.stsci.edu/nms/main/). Each set of curves shows a sequence of redshifts, namely $z = 5$, 7, 9, 11, 13 and 15, respectively, from top to bottom.

GRB afterglows have smooth (broken power-law) continuum spectra unlike quasars which show strong spectral features (such as broad emission lines or the so-called 'blue bump') that complicate the extraction of IGM absorption features. In particular, the continuum extrapolation into the Lyα damping wing (see section 2.1) during the epoch of reionization is much more straightforward for the smooth UV spectra of GRB afterglows than for quasars with an underlying broad Lyα emission line (Barkana and Loeb 2004a). However, the interpretation may be complicated by the presence of damped Lyα absorption by dense neutral hydrogen in the immediate environment of the GRB within its host galaxy. Since GRBs originate from the dense environment of active star formation, such damped absorption is expected and indeed has been seen, including in the most distant GRB at $z = 6.3$ (Totani *et al* 2006).

3.4. The epoch of reionization

Given the understanding described above of how many galaxies formed at various times, the course of reionization can be determined Universe-wide by counting photons from all sources of light (Arons and Wingert 1972, Shapiro and Giroux 1987, Meiksin and Madau 1993, Miralda-Escudé and Ostriker 1990, Tegmark *et al* 1994, Kamionkowski *et al* 1994, Fukugita and Kawasaki 1994, Shapiro *et al* 1994, Haiman and Loeb 1997). Both stars and black holes contribute ionizing photons, but the early Universe is dominated by small galaxies which in the local Universe have central black holes that are disproportionately small, and indeed quasars are rare above redshift 6 (Fan *et al* 2003). Thus, stars most likely dominated the production of ionizing UV photons during the reionization epoch (although high-redshift galaxies should have also emitted x-rays from accreting black holes and accelerated particles

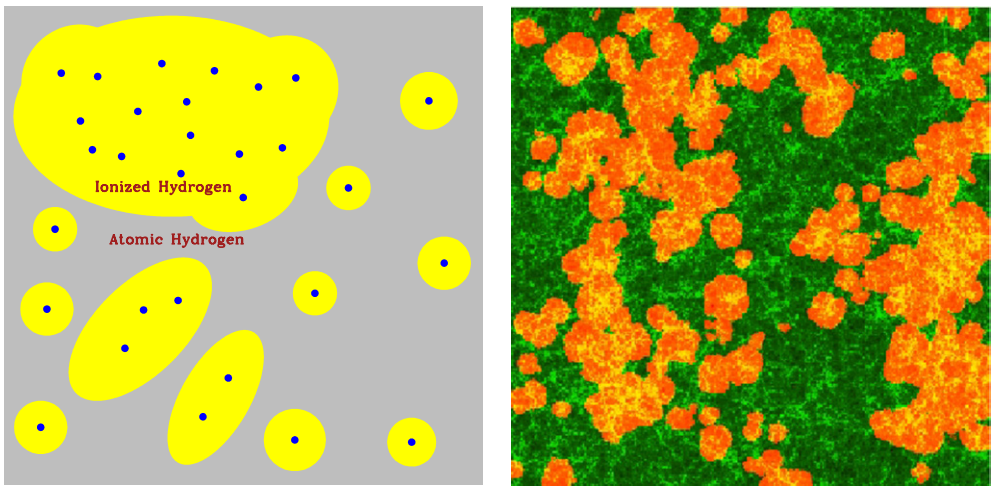

Figure 10. The spatial structure of cosmic reionization. The illustration (left panel, based on Barkana and Loeb (2004b)) shows how regions with large-scale overdensities form large concentrations of galaxies (dots) whose ionizing photons produce enormous joint ionized bubbles (upper left). At the same time, galaxies are rare within large-scale voids, in which the IGM is still mostly neutral (lower right). A numerical simulation of reionization (right panel, from Mellema *et al* (2006)) indeed displays such variation in the sizes of ionized bubbles (orange [light shade]), shown overlayed on the density distribution (green [dark shade]).

in collisionless shocks (Oh 2001)). Since most stellar ionizing photons are only slightly more energetic than the 13.6 eV ionization threshold of hydrogen, they are absorbed efficiently once they reach a region with substantial neutral hydrogen (e.g. see section (2.1)). This makes the IGM during reionization a two-phase medium characterized by highly ionized regions separated from neutral regions by sharp ionization fronts (see figure 10).

We can obtain a first estimate of the requirements of reionization by demanding one stellar ionizing photon for each hydrogen atom in the IGM. If we conservatively assume that stars within the early galaxies were similar to those observed locally, then each star produced $\sim$4000 ionizing photons per baryon. Star formation is observed today to be an inefficient process, but even if stars in galaxies formed out of only $\sim$10% of the available gas, it was still sufficient to accumulate a small fraction (of order 0.1%) of the total baryonic mass in the Universe into galaxies in order to reionize the entire IGM. More accurate estimates of the actual required fraction account for the formation of some primordial stars (which were massive, efficient ionizers, as discussed above), and for recombinations of hydrogen atoms at high redshifts and in dense regions.

From studies of quasar absorption lines at $z \sim 6$ we know that the IGM is highly ionized a billion years after the Big Bang (see the review by Fan *et al* 2006b). There are hints, however, that some large neutral hydrogen regions persist at these early times (Wyithe and Loeb 2004a, Mesinger and Haiman 2004, Lidz *et al* 2006) and so this suggests that we may not need to go to much higher redshifts to begin to see the epoch of reionization. (Interestingly, these inferences were deduced long before the latest WMAP results were announced, when it was widely believed that reionization occurred much earlier.) We now know that the Universe could not have fully reionized earlier than an age of 300 million years, since WMAP observed the effect of the freshly created plasma at reionization on the large-scale polarization anisotropies of the CMB and this limits the reionization redshift (Spergel *et al* 2006); an earlier reionization,

when the Universe was denser, would have created a stronger scattering signature that would be inconsistent with the WMAP observations. In any case, the redshift at which reionization ended only constrains the overall cosmic efficiency of ionizing photon production. In comparison, a detailed picture of reionization as it happens will teach us a great deal about the population of young galaxies that produced this cosmic phase transition.

A key point is that the spatial distribution of ionized bubbles is determined by clustered groups of galaxies and not by individual galaxies. At such early times galaxies were strongly clustered even on very large scales (up to tens of Mpc), and these scales therefore dominate the structure of reionization (Barkana and Loeb 2004b). The basic idea is simple (Kaiser 1984). At high redshift, galactic halos are rare and correspond to rare high density peaks. As an analogy, imagine searching on Earth for mountain peaks above 5000 m. The 200 such peaks are not at all distributed uniformly but instead are found in a few distinct clusters on top of large mountain ranges. Given the large-scale boost provided by a mountain range, a small-scale crest need only provide a small additional rise in order to become a 5000 m peak. The same crest, if it formed within a valley, would not come anywhere near 5000 m in total height. Similarly, in order to find the early galaxies, one must first locate a region with a large-scale density enhancement, and then galaxies will be found there in abundance.

The ionizing radiation emitted from the stars in each galaxy initially produces an isolated ionized bubble. However, in a region dense with galaxies the bubbles quickly overlap into one large bubble, completing reionization in this region while the rest of the Universe is still mostly neutral (figure 10). Most importantly, since the abundance of rare density peaks is very sensitive to small changes in the density threshold, even a large-scale region with a small enhanced density (say, 10% above the mean density of the Universe) can have a much larger concentration of galaxies than in other regions (e.g. a 50% enhancement). On the other hand, reionization is harder to achieve in dense regions, since the protons and electrons collide and recombine more often in such regions, and newly-formed hydrogen atoms need to be reionized again by additional ionizing photons. However, the overdense regions end up reionizing first since the number of ionizing sources in these regions is increased so strongly (Barkana and Loeb 2004b). The large-scale topology of reionization is therefore inside out, with underdense voids reionizing only at the very end of reionization, with the help of extra ionizing photons coming in from their surroundings (which have a higher density of galaxies than the voids themselves). This is a key prediction awaiting observational testing.

Detailed analytical models that account for large-scale variations in the abundance of galaxies (Furlanetto *et al* 2004) confirm that the typical bubble size starts well below a Mpc early in reionization, as expected for an individual galaxy, rises to 5–10 Mpc during the central phase (i.e. when the Universe is half ionized) and then by another factor of $\sim$5 towards the end of reionization. These scales are given in comoving units that scale with the expansion of the Universe, so that the actual sizes at a redshift z were smaller than these numbers by a factor of $1 + z$. Numerical simulations have only recently begun to reach the enormous scales needed to capture this evolution (Ciardi *et al* 2003, Mellema *et al* 2006, Zahn *et al* 2006). Accounting precisely for gravitational evolution on a wide range of scales but still crudely for gas dynamics, star formation and the radiative transfer of ionizing photons, the simulations confirm that the large-scale topology of reionization is inside out and that this topology can be used to study the abundance and clustering of the ionizing sources (figures 10 and 11).

Wyithe and Loeb (2004b) showed that the characteristic size of the ionized bubbles at the end of reionization can be calculated based on simple considerations that only depend on the power spectrum of density fluctuations and the redshift. As the size of an ionized bubble increases, the time it takes a 21 cm photon to traverse it gets longer. At the same time, the variation in the time at which different regions reionize becomes smaller as the regions grow

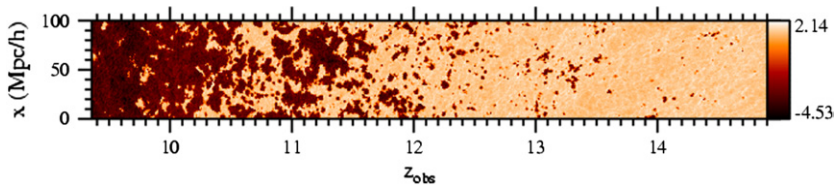

Figure 11. Close-up of cosmic evolution during the epoch of reionization, as revealed in a predicted 21 cm map of the IGM based on a numerical simulation (from Mellema *et al* (2006)). This map is constructed from slices of the simulated cubic box of side 150 Mpc (in comoving units), taken at various times during reionization, which for the parameters of this particular simulation spans a period of 250 million years from redshift 15 down to 9.3. The vertical axis shows position χ in units of Mpc h^{-1} (where $h = 0.7$). This two-dimensional slice of the sky (one linear direction on the sky versus the line-of-sight or redshift direction) shows $\log_{10}(T_b)$, where T_b (in mK) is the 21 cm brightness temperature relative to the CMB. Since neutral regions correspond to strong emission (i.e. a high T_b), this slice illustrates the global progress of reionization and the substantial large-scale spatial fluctuations in reionization history. Observationally it corresponds to a narrow strip half a degree in length on the sky observed with radio telescopes over a wavelength range 2.2–3.4 m (with each wavelength corresponding to 21 cm emission at a specific line-of-sight distance and redshift).

larger. Thus, there is a maximum size above which the photon crossing time is longer than the cosmic variance in ionization time. Regions bigger than this size will be ionized at their near side by the time a 21 cm photon will cross them towards the observer from their far side. They would appear to the observer as one-sided and hence signal the end of reionization. These 'light cone' considerations imply a characteristic size for the ionized bubbles of ∼10 physical Mpc at $z \sim 6$ (equivalent to 70 Mpc today). This result implies that future radio experiments should be tuned to a characteristic angular scale of tens of arcminutes and have a minimum frequency band-width of 5–10 MHz for an optimal detection of 21 cm brightness fluctuations near the end of reionization.

3.5. Post-reionization suppression of low-mass galaxies

After the ionized bubbles overlapped in each region, the ionizing background increased sharply, and the IGM was heated by the ionizing radiation to a temperature $T_{IGM} \gtrsim 10^4$ K. Due to the substantial increase in the IGM pressure, the smallest mass scale into which the cosmic gas could fragment, the so-called Jeans mass, increased dramatically, changing the minimum mass of forming galaxies (Rees 1986, Efstathiou 1992, Gnedin and Ostriker 1997, Miralda-Escudé and Rees 1998; Benson *et al* 2003).

Gas infall depends sensitively on the Jeans mass. When a halo more massive than the Jeans mass begins to form, the gravity of its dark matter overcomes the gas pressure. Even in halos below the Jeans mass, although the gas is initially held up by pressure, once the dark matter collapses its increased gravity pulls in some gas (Haiman *et al* 1996). Thus, the Jeans mass is generally higher than the actual limiting mass for accretion. Before reionization, the IGM is cold and neutral, and the Jeans mass plays a secondary role in limiting galaxy formation compared with cooling. After reionization, the Jeans mass is increased by several orders of magnitude due to the photoionization heating of the IGM, and hence begins to play a dominant role in limiting the formation of stars. Gas infall in a reionized and heated Universe has been investigated in a number of numerical simulations. Thoul and Weinberg (1996) inferred, based on a spherically-symmetric collapse simulation, a reduction of ∼50% in the collapsed gas mass due to heating, for a halo of circular velocity $V_c \sim 50$ km s^{-1} at $z = 2$, and

a complete suppression of infall below $V_c \sim 30\,\mathrm{km\,s^{-1}}$. Kitayama and Ikeuchi (2000) also performed spherically symmetric simulations but included self-shielding of the gas and found that it lowers the circular velocity thresholds by $\sim 5\,\mathrm{km\,s^{-1}}$. Three-dimensional numerical simulations (Quinn *et al* 1996, Weinberg *et al* 1997, Navarro and Steinmetz 1997) found a significant suppression of gas infall in even larger halos ($V_c \sim 75\,\mathrm{km\,s^{-1}}$), but this was mostly due to a suppression of late infall at $z \lesssim 2$.

When a volume of the IGM is ionized by stars, the gas is heated to a temperature $T_{\mathrm{IGM}} \sim 10^4\,\mathrm{K}$. If quasars dominate the UV background at reionization, their harder photon spectrum leads to $T_{\mathrm{IGM}} > 2 \times 10^4\,\mathrm{K}$. Including the effects of dark matter, a given temperature results in a linear Jeans mass corresponding to a halo circular velocity of

$$V_{\mathrm{J}} \approx 80 \left(\frac{T_{\mathrm{IGM}}}{1.5 \times 10^4 \mathrm{K}} \right)^{1/2} \mathrm{km\,s^{-1}}. \tag{20}$$

In halos with a circular velocity well above V_{J}, the gas fraction in infalling gas equals the universal mean of $\Omega_{\mathrm{b}}/\Omega_{\mathrm{m}}$, but gas infall is suppressed in smaller halos. A simple estimate of the limiting circular velocity, below which halos have essentially no gas infall, is obtained by substituting the virial overdensity for the mean density in the definition of the Jeans mass. The resulting estimate is

$$V_{\mathrm{lim}} = 34 \left(\frac{T_{\mathrm{IGM}}}{1.5 \times 10^4 \mathrm{K}} \right)^{1/2} \mathrm{km\,s^{-1}}. \tag{21}$$

This value is in rough agreement with the numerical simulations mentioned before. A more recent study by Dijkstra *et al* (2004) indicated that at the high redshifts of $z > 10$ gas could nevertheless assemble into halos with circular velocities as low as $v_c \sim 10\,\mathrm{km\,s^{-1}}$, even sometime after the establishment of a UV background.

Although the Jeans mass is closely related to the rate of gas infall at a given time, it does not directly yield the total gas residing in halos at a given time. The latter quantity depends on the entire history of gas accretion onto halos, as well as on the merger histories of halos, and an accurate description must involve a time-averaged Jeans mass. The gas content of halos in simulations is well fit by an expression which depends on the filtering mass, a particular time-averaged Jeans mass (Gnedin and Hui 1998, Gnedin 2000).

The reionization process was not perfectly synchronized throughout the Universe. Large-scale regions with a higher density than the mean tended to form galaxies first and reionized earlier than underdense regions. The suppression of low-mass galaxies by reionization is therefore modulated by the fluctuations in the timing of reionization (Babich and Loeb 2006). Inhomogeneous reionization imprints a signature on the power spectrum of low-mass galaxies. Future high-redshift galaxy surveys hoping to constrain inflationary parameters must properly model the effects of reionization; conversely, they will also place new constraints on the thermal history of the IGM during reionization.

4. 21 cm cosmology

4.1. A handy tool for studying cosmic reionization

The prospect of studying reionization by mapping the distribution of atomic hydrogen across the Universe using its prominent 21 cm spectral line has motivated several teams to design and construct arrays of low-frequency radio telescopes; the Low Frequency Array (http://www.lofar.org/), the Mileura Wide-Field Array (http://www.haystack.mit.edu/ast/arrays/mwa/site/index.html), the Primeval Structure Telescope (http://arxiv.org/abs/astro-ph/0502029) and ultimately the Square Kilometer Array

(http://www.skatelescope.org) will search over the next decade for 21 cm emission or absorption from $z \sim 6.5$–15, redshifted and observed today at relatively low frequencies which correspond to wavelengths of 1.5–4 m.

The idea is to use the resonance associated with the hyperfine splitting in the ground state of hydrogen (see section 2.2). While the CMB spectrum peaks at a wavelength of 2 mm, it provides a still-measurable intensity at metre wavelengths that can be used as the bright background source against which we can see the expected 1% absorption by neutral hydrogen along the line-of-sight (Hogan and Rees 1979, Scott and Rees 1990). The hydrogen gas produces 21 cm absorption if its spin temperature is colder than the CMB and excess emission if it is hotter. Since the CMB covers the entire sky, a complete three-dimensional map of neutral hydrogen can in principle be made from the sky position of each absorbing gas cloud together with its redshift z.

Because the smallest angular size resolvable by a telescope is proportional to the observed wavelength, radio astronomy at wavelengths as large as a metre has remained relatively undeveloped. Producing resolved images even of large sources such as cosmological ionized bubbles requires telescopes which have a kilometre scale. It is much more cost-effective to use a large array of thousands of simple antennas distributed over several kilometres, and to use computers to cross-correlate the measurements of the individual antennas and combine them effectively into a single large telescope. The new experiments are being placed mostly in remote sites, because the cosmic wavelength region overlaps with more mundane terrestrial telecommunications.

In approaching redshifted 21 cm observations, although the first inkling might be to consider the mean emission signal, the signal is orders of magnitude fainter than foreground synchrotron emission from relativistic electrons in the magnetic field of our own Milky Way (Furlanetto *et al* 2006) as well as other galaxies (Di Matteo *et al* 2002). Thus, cosmologists have focused on the expected characteristic variations in $T_{\rm b}$, both with position on the sky and especially with frequency, which signifies redshift for the cosmic signal. The synchrotron foreground is expected to have a smooth frequency spectrum, and so it is possible to isolate the cosmological signal by taking the difference in the sky brightness fluctuations at slightly different frequencies (as long as the frequency separation corresponds to the characteristic size of ionized bubbles). The 21 cm brightness temperature depends on the density of neutral hydrogen. As explained in the previous subsection, large-scale patterns in the reionization are driven by spatial variations in the abundance of galaxies; the 21 cm fluctuations reach ∼5 mK (root-mean-square) in brightness temperature (figure 11) on a scale of 10 Mpc (comoving). While detailed maps will be difficult to extract due to the foreground emission, a statistical detection of these fluctuations (through the power spectrum) is expected to be well within the capabilities of the first-generation experiments now being built (Bowman *et al* 2006, McQuinn *et al* 2006). Current work suggests that the key information on the topology and timing of reionization can be extracted statistically.

While numerical simulations of reionization are now reaching the cosmological box sizes needed to predict the large-scale topology of the ionized bubbles, they do this at the price of limited small-scale resolution. These simulations cannot yet follow in any detail the formation of individual stars within galaxies, or the feedback that stars produce on the surrounding gas, such as photoheating or the hydrodynamic and chemical impact of supernovae, which blow hot bubbles of gas enriched with the chemical products of stellar nucleosynthesis. Thus, the simulations cannot directly predict whether the stars that form during reionization are similar to the stars in the Milky Way and nearby galaxies or to the primordial $100 M_{\odot}$ behemoths. They also cannot determine whether feedback prevents low-mass dark matter halos from forming stars. Thus, models are needed that make it possible to vary all these astrophysical parameters

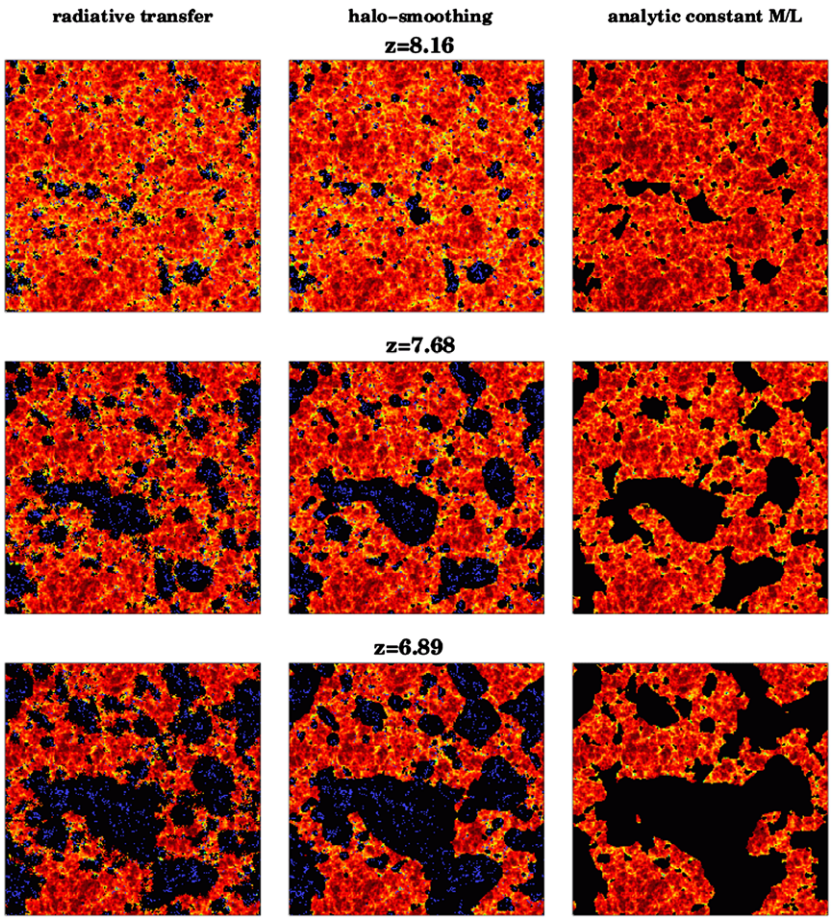

Figure 12. Maps of the 21 cm brightness temperature comparing results of a numerical simulation and of two simpler numerical schemes, at three different redshifts (from Zahn *et al* (2006)). Each map is 65.6 Mpc h^{-1} on a side, with a depth (0.25 Mpc h^{-1}) that is comparable to the frequency resolution of planned experiments. The ionized fractions are $x_i = 0.13$, 0.35 and 0.55 for $z = 8.16$, 7.26 and 6.89, respectively. All three maps show a very similar large-scale ionization topology. *Left column:* Numerical simulation, showing the ionized bubbles (black) produced by the ionizing sources (dots) that form in the simulation. *Middle column:* Numerical scheme that applies the Furlanetto *et al* (2004) analytical model to the final distribution of ionizing sources that form in the simulation. *Right column:* Numerical scheme that applies the Furlanetto *et al* (2004) analytical model to the linear density fluctuations that are the initial conditions of the simulation.

of the ionizing sources and to study the effect on the 21 cm observations. For example, Furlanetto *et al* (2004) developed an analytical model that allows the calculation of the one-point bubble distribution, i.e. the probability distribution (at a given redshift) of the size of the ionizing bubble surrounding a random point in space. Zahn *et al* (2006) have considered numerical schemes that apply the Furlanetto *et al* (2004) model to either the initial conditions of their simulation or to part of its results (figure 12). Barkana (2007) has generalized the Furlanetto *et al* (2004) model and presented an analytical model that calculates the 21 cm two-point correlation (or equivalently, the power spectrum), a quantity of interest for the upcoming 21 cm observations (figure 13).

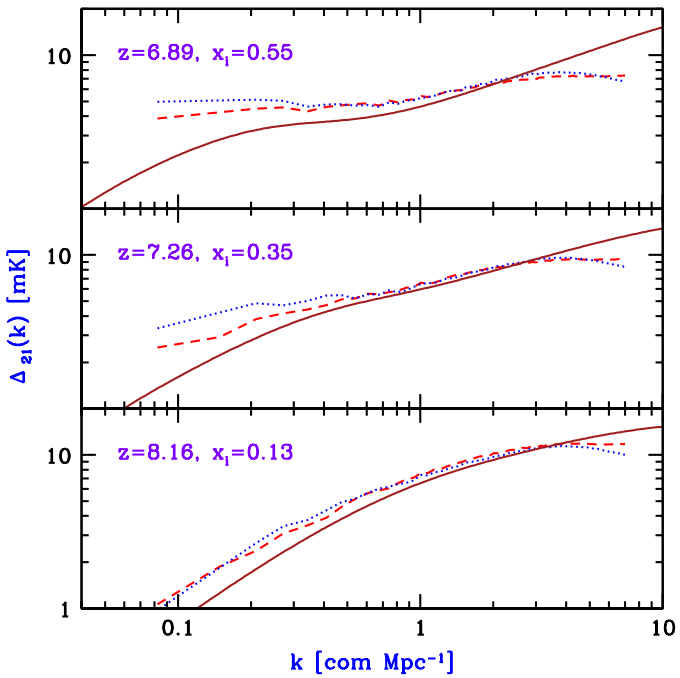

Figure 13. 21 cm power spectrum, comparing the analytical model of Barkana (2007) (solid curves) to those from the simulation of Zahn *et al* (2006) (dashed curves) and from their numerical scheme based on the final ionizing sources (dotted curves). The results are shown at several different redshifts, as indicated in each panel. At each redshift z, the value of the efficiency in the analytical model is adjusted in order to match the mean global ionized fraction $x_{\rm i}$ from the simulation. While the model and the numerical scheme include some approximations and the simulation has resolution and size limitations, they all show the same overall trend: as reionization advances, the power spectrum flattens, with the increased power on large scales ($k \sim 0.1\,{\rm Mpc}^{-1}$) reflecting the increasing size of bubbles that are due to clustered groups of galaxies.

The theoretical expectations presented here for reionization and for the 21 cm signal are based on rather large extrapolations from observed galaxies to deduce the properties of much smaller galaxies that formed at an earlier cosmic epoch. Considerable surprises are thus possible, such as an early population of quasars or even unstable exotic particles that emitted ionizing radiation as they decayed. In any case, the forthcoming observational data in 21 cm cosmology should make the next few years a very exciting time.

4.2. Multiple uses in the era before reionization

A detection of the cosmological 21 cm signal will open a new window on the Universe and likely motivate a second generation of more powerful telescopes. These will be used to obtain three-dimensional maps of atomic hydrogen during reionization as well as statistical power-spectrum measurements at even higher redshifts (using wavelengths at which the foregrounds are brighter and thus more difficult to remove). Since the 21 cm measurements are sensitive to any difference between the hydrogen temperature and the CMB temperature, the potential reach of 21 cm cosmology extends down to a cosmic age of ∼6 million years ($z \sim 200$), when the IGM first cooled below the CMB temperature (an event referred to as "thermal decoupling") due to the cosmic expansion. At those very high redshifts, the cosmic gas was

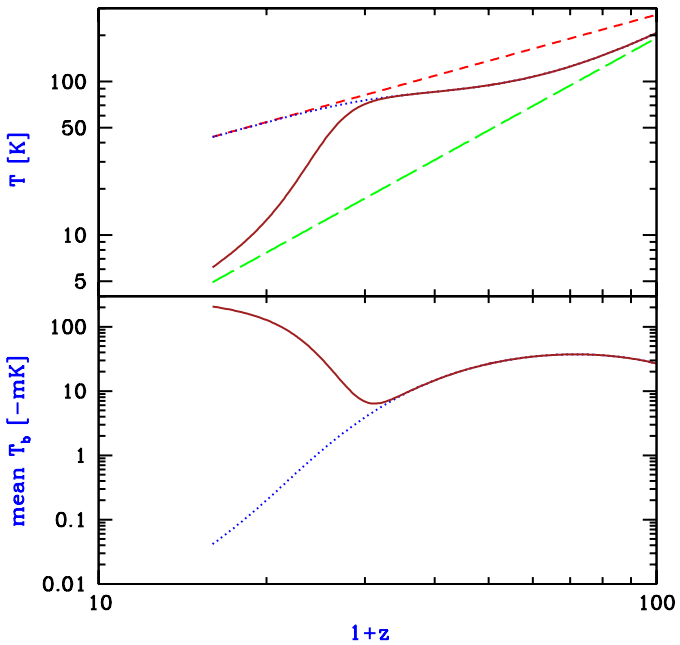

Figure 14. Redshift evolution of various mean temperatures (from Barkana and Loeb (2005b)). The mean 21 cm spin temperature is shown in the upper panel for adiabatic IGM gas with scattering of stellar Lyα photons (solid curve) and without it (dotted curve). Also shown for comparison are the gas temperature (long-dashed curve) and the CMB temperature (short-dashed curve). In adiabatic expansion, the CMB temperature cools approximately as $\propto (1+z)$ and the gas cools faster, $\propto (1+z)^2$. The mean 21 cm brightness temperature offset relative to the CMB is shown in the lower panel, with Lyα scattering (solid curve) and without it (dotted curve). The Lyα flux is calculated assuming that galaxies with Pop III stars form in dark matter halos where the gas cools efficiently via atomic cooling. The star formation efficiency is normalized so that Lyα scattering becomes significant (i.e. brings the value of $x_{\rm tot}$ from section 2.2 to unity) at redshift 20.

so dense that atomic collisions kept the 21 cm spin temperature near that of the gas. As the Universe expanded, if there were no stars, CMB scattering would overcome the decreasing collision rate, drive the spin temperature to the CMB temperature and eliminate the 21 cm signal at $z \lesssim 30$ (figure 14). Instead, the formation of a significant galaxy population by redshift 30 is expected to drive the spin temperature back towards the gas temperature through the indirect mechanism of Lyα scattering (Wouthuysen 1952, Field 1958, Madau *et al* 1997, Chen and Miralda-Escudé 2004). Thus, it should be possible to detect a 21 cm signal throughout this redshift range in absorption (as long as the gas is cooler than the CMB).

At high redshifts prior to reionization, spatial perturbations in the thermodynamic gas properties are linear and can be predicted precisely (see section 3.1). While early collapsed mini-halos are expected to produce a 21 cm signal (Iliev *et al* 2003), in most scenarios this signal is swamped by that from the IGM (Oh and Mack 2003; Furlanetto and Oh 2006). Thus, if the gas is probed with the 21 cm technique then it becomes a promising tool of fundamental, precision cosmology, able to probe the primordial power spectrum of density fluctuations imprinted in the very early Universe, perhaps in an era of cosmic inflation. The 21 cm fluctuations can be measured down to the smallest scales where the baryon pressure suppresses gas fluctuations, while the CMB anisotropies are damped on small scales (through the so-called Silk damping). This difference in damping scales can be seen by comparing the

baryon density and photon-temperature power spectra in figure 6. Since the 21 cm technique is also three-dimensional (while the CMB yields a single sky map), there is a much larger potential number of independent modes probed by the 21 cm signal: $N_{21\,\mathrm{cm}} \sim 3 \times 10^{16}$ compared with $N_{\mathrm{cmb}} \sim 2 \times 10^7$ (Loeb and Zaldarriaga 2004). This larger number should provide a measure of non-Gaussian deviations to a level of $\sim N_{21\,\mathrm{cm}}^{-1/2}$, constituting a test of the inflationary origin of the primordial inhomogeneities which are expected to possess non-Gaussian deviations $\gtrsim 10^{-6}$.

An important cross-check on these measurements is possible by measuring the particular form of anisotropy, expected in the 21 cm fluctuations, that is caused by gas motions along the line-of-sight (Kaiser 1987, Bharadwaj and Ali 2004, Barkana and Loeb 2005a). This anisotropy, expected in any measurement of density that is based on a spectral resonance or on redshift measurements, results from velocity compression. Consider a photon travelling along the line-of-sight that resonates with absorbing atoms at a particular point. In a uniform, expanding Universe, the absorption optical depth encountered by this photon probes only a narrow strip of atoms, since the expansion of the Universe makes all other atoms move with a relative velocity that takes them outside the narrow frequency width of the resonance line. If there is a density peak, however, near the resonating position, the increased gravity will reduce the expansion velocities around this point and bring more gas into the resonating velocity width. This effect is sensitive only to the line-of-sight component of the velocity gradient of the gas and thus causes an observed anisotropy in the power spectrum even when all physical causes of the fluctuations are statistically isotropic. Barkana and Loeb (2005a) showed that this anisotropy is particularly important in the case of 21 cm fluctuations. When all fluctuations are linear, the 21 cm power spectrum takes the form (Barkana and Loeb 2005a)

$$P_{21\,\mathrm{cm}}(\boldsymbol{k}) = \mu^4 P_\rho(k) + 2\mu^2 P_{\rho-\mathrm{iso}}(k) + P_{\mathrm{iso}}, \tag{22}$$

where $\mu = \cos\theta$ in terms of the angle θ between the wavevector $\boldsymbol{k}$ of a given Fourier mode and the line-of-sight, P_{iso} is the isotropic power spectrum that would result from all sources of 21 cm fluctuations without velocity compression, $P_\rho(k)$ is the 21 cm power spectrum from gas density fluctuations alone and $P_{\rho-\mathrm{iso}}(k)$ is the Fourier transform of the cross-correlation between the density and all sources of 21 cm fluctuations. The three power spectra can also be denoted $P_{\mu^4}(k)$, $P_{\mu^2}(k)$ and $P_{\mu^0}(k)$, according to the power of μ that multiplies each term. The prediction for these power spectra at high redshift ($z > 20$), neglecting the effects of any stellar radiation, are shown in figure 15. At these redshifts, the 21 cm fluctuations probe the infall of the baryons into the dark matter potential wells (Barkana and Loeb 2005c). The power spectrum shows remnants of the photon–baryon acoustic oscillations on large scales and of the baryon pressure suppression on small scales (Naoz and Barkana 2005).

Once stellar radiation becomes significant, many processes can contribute to the 21 cm fluctuations. The contributions include fluctuations in gas density, temperature, ionized fraction and Lyα flux. These processes can be divided into two broad categories: the first, related to *'physics'*, consists of probes of fundamental, precision cosmology, and the second, related to *'astrophysics'*, consists of probes of stars. Both categories are interesting—the first for precision measures of cosmological parameters and studies of processes in the early Universe, and the second for studies of the properties of the first galaxies. However, the astrophysics depends on complex non-linear processes (collapse of dark matter halos, star formation, supernova feedback), and must be cleanly separated from the physics contribution, in order to allow precision measurements of the latter. As long as all the fluctuations are linear, the anisotropy noted above allows precisely this separation of the physics from the astrophysics of the 21 cm fluctuations (Barkana and Loeb 2005a). In particular, the $P_{\mu^4}(k)$ is independent of the effects of stellar radiation and is a clean probe of the gas density fluctuations. Once non-linear terms become important, there arises a significant mixing of the different terms;

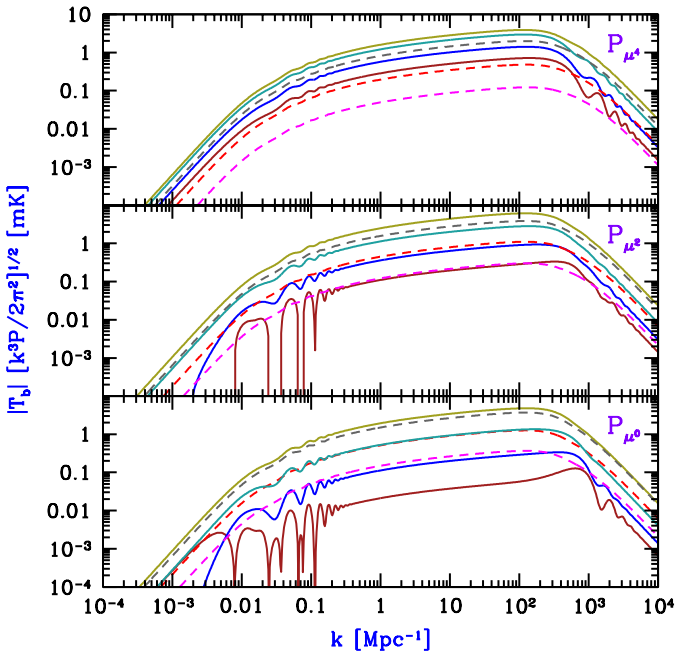

Figure 15. Power spectra of 21 cm brightness fluctuations versus comoving wavenumber (Barkana and Loeb 2005c, Naoz and Barkana 2005). We show the three power spectra that are separately observable, P_{μ^4} (upper panel), P_{μ^2} (middle panel), and P_{μ^0} (lower panel). In each case we show redshifts 200, 150, 100, 50 (solid curves, from bottom to top), 35, 25 and 20 (dashed curves, from top to bottom).

in particular, this occurs on the scale of the ionizing bubbles during reionization (McQuinn *et al* 2006).

The 21 cm fluctuations are affected by fluctuations in the Lyα flux from stars, a result that yields an indirect method to detect and study the early population of galaxies at $z \sim 20$ (Barkana and Loeb 2005b). The fluctuations are caused by biased inhomogeneities in the density of galaxies, along with Poisson fluctuations in the number of galaxies. Observing the power spectra of these two sources would probe the number density of the earliest galaxies and the typical mass of their host dark matter halos. Furthermore, the enhanced amplitude of the 21 cm fluctuations from the era of Lyα coupling considerably improves the practical prospects for their detection. Precise predictions account for the detailed properties of all possible cascades of a hydrogen atom after it absorbs a photon (Hirata 2006, Pritchard and Furlanetto 2006a). Around the same time, x-rays may also start to heat the cosmic gas and to indirectly generate Lyα photons, producing strong 21 cm fluctuations due to fluctuations in the x-ray flux (Chuzhoy *et al* 2006, Pritchard and Furlanetto 2006b).

5. Future prospects

Understandably, astronomers are eager to start tuning into the cosmic radio channels of 21 cm cosmology. The main challenge involves the galactic synchrotron foreground which is several orders of magnitude brighter than the cosmic signal (for details, see Furlanetto *et al* (2006)). Removal of the bright foreground is made feasible by the fact that it does not change for slight shifts in observed wavelength while the cosmological signal does (because different

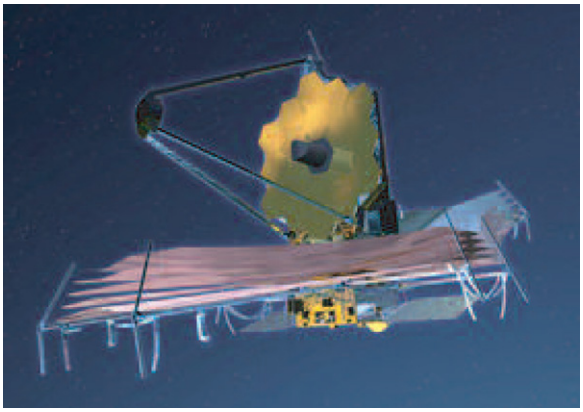

Figure 16. A sketch of the current design for the James Webb Space Telescope, the successor to the Hubble Space Telescope to be launched in 2013 (see http://www.jwst.nasa.gov/). The current design includes a primary mirror made of beryllium which is 6.5 m in diameter as well as an instrument sensitivity that spans the full range of infrared wavelengths of 0.6–28μm and will allow detection of some of the first galaxies in the infant Universe. The telescope will orbit 1.5 million km from Earth at the Lagrange L2 point. Note that the sun shield (the large flat screen in the image) is 22 m$\times$10 m in size.

wavelengths correspond to different slices through the 'swiss cheese' structure of the cosmic hydrogen). Primordial hydrogen produced 21 cm absorption against the CMB sky within the redshift interval $30 \lesssim z \lesssim 200$. Density fluctuations in the gas produced variable absorption and corresponding 21 cm brightness fluctuations during that epoch. Probing these fluctuations in three dimensions down to small scales would offer the biggest data set about the initial conditions of the Universe, with orders of magnitude more bits of information than the CMB anisotropies or galaxy surveys could ever provide. Such information has the potential to bring new clues about small deviations from Gaussianity in the initial density field or about new species of cosmic matter. At redshifts $z \lesssim 20$, the cosmic gas is expected to be heated above the CMB temperature by radiation from the first galaxies, and if so its 21 cm brightness would exceed that of the CMB. At this stage, the 21 cm fluctuations are sourced by the first galaxies which produce ionization fraction fluctuations Lyα intensity fluctuations, temperature fluctuations or direct 21 cm emission from mini-halos.

In parallel to the search for redshifted 21 cm fluctuations, future infrared telescopes will search directly for the early galaxies that sculpted the 'swiss cheese' topology in the primordial hydrogen. Within the next decade, NASA plans to launch an infrared space telescope (JWST; figure 16) that will image some of the earliest sources of light (stars and black holes) in the Universe. In parallel, there are several initiatives to construct large-aperture infrared telescopes on the ground with the same goal in mind (see http://www.eso.org/projects/owl/; http://celt.ucolick.org/; http://www.gmto.org/).

The next generation of ground-based telescopes will have an effective diameter of 20–30 m (figure 17), and together with the *JWST* (which will not be affected by the atmosphere) they will be able to image and make spectral studies of the early galaxies at redshifts 10–30. Given that these galaxies also ionized their environments, their locations should correlate with cavities in the 21 cm emission (Wyithe *et al* 2005, Wyithe and Loeb 2006). Within a decade it should be possible to explore the environmental influence of the first galaxies by using both radio and infrared instruments in concert. The challenge is for the theorists to predict this interaction reliably through models and numerical simulations before it is actually observed.

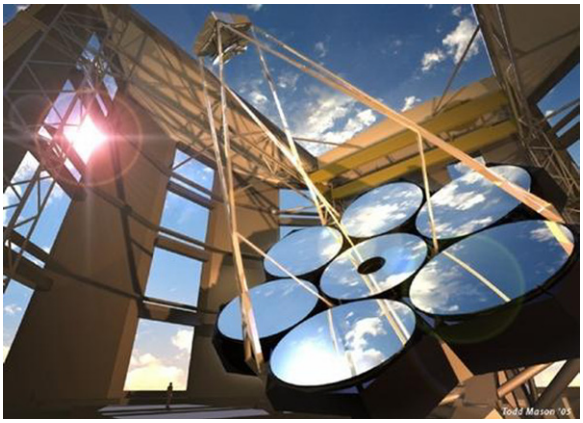

Figure 17. Artist's conception of the design for one of the future giant telescopes that could probe the first generation of galaxies from the ground. The Giant Magellan Telescope (*GMT*) will contain seven mirrors (each 8.4 m in diameter) and will have a resolving power equivalent to a 24.5 m (80 ft) primary mirror. For more details see http://www.gmto.org/.

Acknowledgments

The authors acknowledge support by the Israel—US Binational Science Foundation grant 2004386, the Israel Science Foundation grant 629/05 (for RB) and Harvard University funds (for AL).

References

Abel T L, Bryan G L and Norman M L 2002 *Science* **295** 93
Allison A C and Dalgarno A 1969 *Astrophys. J.* **158** 423
Arons J and Wingert D W 1972 *Astrophys. J.* **177** 1
Babich D and Loeb A 2006 *Astrophys. J.* **640** 1
Barkana R 2006 *Science* **313** 931
Barkana R 2007 *Mon. Not. R. Astron. Soc.* at press
Barkana R and Loeb A 2001 *Phys. Rep.* **349** 125
Barkana R and Loeb A 2004a *Astrophys. J.* **601** 64
Barkana R and Loeb A 2004b *Astrophys. J.* **609** 474
Barkana R and Loeb A 2005a *Astrophys. J. Lett.* **624** L65
Barkana R and Loeb A 2005b *Astrophys. J.* **626** 1
Barkana R and Loeb A 2005c *Mon. Not. R. Astron. Soc. Lett.* **363** L36
Bennett C L *et al* 1996 *Astrophys. J. Lett.* **464** L1
Benson A J, Bower, R G, Frenk C S, Lacey C G, Baugh C M and Cole S 2003 *Astrophys. J.* **599** 38
Bharadwaj S and Ali S S 2004 *Mon. Not. R. Astron. Soc.* **352** 142
Bowman J D, Morales M F and Hewitt J N 2006 *Astrophys. J.* **638** 20
Bromm V, Coppi P S and Larson R B 2002 *Astrophys. J.* **564** 23
Bromm V and Larson R B 2004 *Annu. Rev. Astron. Astrophys.* **42** 79
Bromm V and Loeb A 2004 *New Astron.* **9** 353
Bundy K *et al* 2006 *Astrophys. J.* **651** 120
Chen X and Miralda-Escudé J 2004 *Astrophys. J.* **602** 1
Chuzhoy L, Alvarez M A and Shapiro P R 2006 *Astrophys. J. Lett.* **648** L1
Chuzhoy L and Shapiro P R 2006 *Astrophys. J.* **651** 1
Ciardi B, Ferrara A and White S D M 2003 *Mon. Not. R. Astron. Soc.* **344** L7
Ciardi B and Loeb A 2000 *Astrophys. J.* **540** 687
Cole S *et al* 2005 *Mon. Not. R. Astron. Soc.* **362** 505

Dijkstra M, Haiman Z, Rees M J and Weinberg D H 2004 *Astrophys. J.* **601** 666
Di Matteo T, Perna R, Abel T and Rees M J 2002 *Astrophys. J.* **564** 576
Efstathiou G 1992 *Mon. Not. R. Astron. Soc.* **256** 43
Eisenstein D J *et al* 2005 *Astrophys. J.* **633** 560
Fan X *et al* 2002 *Astron. J.* **123** 1247
Fan X *et al* 2003 *Astron. J.* **125** 1649
Fan X *et al* 2005 *Preprint* astro-ph/0512082
Fan X *et al* 2006a *Astron. J.* **132** 117
Fan X, Carilli C L and Keating B 2006b *Annu. Rev. Astron. Astrophys.* **44** 415
Field G B 1958 *Proc. IRE* **46** 240
Field G B 1959a *Astrophys. J.* **129** 536
Field G B 1959b *Astrophys. J.* **129** 551
Field G B 1972 *Annu. Rev. Astron. Astrophys.* **10** 227
Fukugita M and Kawasaki M 1994 *Mon. Not. R. Astron. Soc.* **269** 563
Furlanetto S R and Loeb A 2003 *Astrophys. J.* **588** 18
Furlanetto S R and Oh S P 2006 *Astrophys. J.* **652** 849
Furlanetto S R, Oh S P and Briggs F 2006 *Phys. Rep.* at press (*Preprint* astro-ph/0608032)
Furlanetto S R, Zaldarriaga M and Hernquist L 2004 *Astrophys. J.* **613** 1
Gehrels N *et al* 2004 *Astrophys. J.* **611** 1005
Gnedin N Y 2000 *Astrophys. J.* **542** 535
Gnedin N Y and Hui L 1998 *Mon. Not. R. Astron. Soc.* **296** 44
Gnedin N Y and Ostriker J P 1997 *Astrophys. J.* **486** 581
Goodman J 1995 *Phys. Rev.* D **52** 1821
Gunn J E and Peterson B A 1965 *Astrophys. J.* **142** 1633
Haiman Z and Loeb A 1997 *Astrophys. J.* **483** 21
Haiman Z, Rees M J and Loeb A 1997 *Astrophys. J.* **476** 458
Haiman Z, Rees M J and Loeb A 1997 *Astrophys. J.* **484** 985 (erratum)
Haiman Z, Thoul A A and Loeb A 1996 *Astrophys. J.* **464** 52
Haislip J *et al* 2006 *Nature* **440** 181
Hirata C M 2006 *Mon. Not. R. Astron. Soc.* **367** 259
Hogan C J and Rees M J 1979 *Mon. Not. R. Astron. Soc.* **188** 791
Hu E M, Cowie L L, McMahon R G, Capak P, Iwamuro F, Kneib J-P, Maihara T and Motohara K 2002 *Astrophys. J. Lett.* **568** L75
Iliev I T, Scannapieco E, Martel H and Shapiro P R 2003 *Mon. Not. R. Astron. Soc.* **341** 81
Iye M *et al* 2006 *Nature* **443** 186
Kaiser N 1984 *Astrophys. J. Lett.* **284** L9
Kaiser N 1987 *Mon. Not. R. Astron. Soc.* **227** 1
Kamionkowski M, Spergel D N and Sugiyama N 1994 *Astrophys. J. Lett.* **426** L57
Kitayama T and Ikeuchi S 2000 *Astrophys. J.* **529** 615
Kolb E W and Turner M S 1990 *The Early Universe* (Redwood City, CA: Addison-Wesley)
Lamb D Q and Reichart D E 2000 *Astrophys. J.* **536** 1
Lidz A, Oh S P and Furlanetto S R 2006 *Astrophys. J. Lett.* **639** L47
Loeb A 2006 *First Light (SAAS-Fee Lecture Notes)* (Berlin: Springer) (*Preprint* astro-ph/0603360)
Loeb A and Zaldarriaga M 2004 *Phys. Rev. Lett.* **92** 211301
Ma C and Bertschinger E 1995 *Astrophys. J.* **455** 7
Madau P, Meiksin A and Rees M J 1997 *Astrophys. J.* **475** 429
McQuinn M, Zahn O, Zaldarriaga M, Hernquist L and Furlanetto S R 2006 *Astrophys. J.* at press (*Preprint* astro-ph/0512263)
Meiksin A and Madau P 1993 *Astrophys. J.* **412** 24
Mellema G, Iliev I T, Pen U-L and Shapiro P R 2006 *Mon. Not. R. Astron. Soc.* at press (*Preprint* astro-ph/0603518)
Mesinger A and Haiman Z 2004 *Astrophys. J. Lett.* **611** 69
Miralda-Escudé J 1998 *Astrophys. J.* **501** 15
Miralda-Escudé J 2000 *Astrophys. J. Lett.* **528** L1
Miralda-Escudé J and Ostriker J P 1990 *Astrophys. J.* **350** 1
Miralda-Escudé J and Rees M J 1998 *Astrophys. J.* **497** 21
Naoz S and Barkana R 2005 *Mon. Not. R. Astron. Soc.* **362** 1047
Naoz S, Noter S and Barkana R 2006 *Mon. Not. R. Astron. Soc. Lett.* at press (*Preprint* astro-ph/0604050)
Navarro J F and Steinmetz M 1997 *Astrophys. J.* **478** 13

Oh S P 2001 *Astrophys. J.* **553** 499
Oh S P and Mack K J 2003 *Mon. Not. R. Astron. Soc.* **346** 871
Peebles P J E 1980 *The Large-Scale Structure of the Universe* (Princeton: Princeton University Press)
Peebles P J E 1984 *Astrophys. J.* **277** 470
Peebles P J E 1993 *Principles of Physical Cosmology* (Princeton: Princeton University Press)
Peebles P J E and Yu J T 1970 *Astrophys. J.* **162** 815
Pritchard J R and Furlanetto S R 2006a *Mon. Not. R. Astron. Soc.* **367** 1057
Pritchard J R and Furlanetto S R 2006b *Mon. Not. R. Astron. Soc.* submitted *Preprint* astro-ph/0607234
Purcell E M and Field G B 1956 *Astrophys. J.* **124** 542
Quinn T, Katz N and Efstathiou G 1996 *Mon. Not. R. Astron. Soc. Lett.* **278** 49
Rees M J 1986 *Mon. Not. R. Astron. Soc.* **222** 27
Rees M J and Sciama D W 1968 *Nature* **217** 511
Sachs R K and Wolfe A M 1967 *Astrophys. J.* **147** 73
Scott D and Rees M J 1990 *Mon. Not. R. Astron. Soc.* **247** 510
Seljak U and Zaldarriaga M 1996 *Astrophys. J.* **469** 437
Shapiro P R and Giroux M L 1987 *Astrophys. J. Lett.* **321** L107
Shapiro P R, Giroux M L and Babul A 1994 *Astrophys. J.* **427** 25
Silk J 1968 *Astrophys. J.* **151** 459
Spergel D N *et al* 2006 *Astrophys. J.* at press (*Preprint* astro-ph/0603449)
Sunyaev R A and Zeldovich Y B 1970 *Astrophys. Space Sci.* **7** 3
Tegmark M, Silk J and Blanchard A 1994 *Astrophys. J.* **420** 484
Tegmark M *et al* 1997 *Astrophys. J.* **474** 1
Thoul A A and Weinberg D H 1996 *Astrophys. J.* **465** 608
Totani T, Kawai N, Kosugi G, Aoki K, Yamada T, Iye M, Ohta K and Hattori T 2006 *Publ. Astron. Soc. Japan* **58** 485
Viel M, Haehnelt M G and Lewis A 2006 *Mon. Not. R. Astron. Soc. Lett.* **370** L51
Weinberg D H, Hernquist L and Katz N 1997 *Astrophys. J.* **477** 8
Weinberg S 1972 *Gravitation and Cosmology* (New York: Wiley)
White R L, Becker R H, Fan X, Strauss M A 2003 *Astron. J.* **126** 1
Wouthuysen S A 1952 *Astron. J.* **57** 31
Wu K K S, Lahav O and Rees M J 1999 *Nature* **397** 225
Wyithe J S B and Loeb A 2004a *Nature* **427** 815
Wyithe J S B and Loeb A 2004b *Nature* **432** 194
Wyithe J S B and Loeb A 2006 *Mon. Not. R. Astron. Soc* submitted (*Preprint* astro-ph/0609734)
Wyithe J S B, Loeb A and Barnes D G 2005 *Astrophys. J.* **634** 715
Yamamoto K, Sugiyama N and Sato H 1997 *Phys. Rev.* D **56** 7566
Yamamoto K, Sugiyama N and Sato H 1998 *Astrophys. J.* **501** 442
Yoshida N, Omukai K, Hernquist L, and Abel T 2006 *Astrophys. J.* at press (*Preprint* astro-ph/0606106)
Zahn O, Lidz A, McQuinn M, Dutta S, Hernquist L, Zaldarriaga M and Furlanetto S R 2006 *Astrophys. J.* at press (*Preprint* astro-ph/0604177)
Zhang W, Woosley S and MacFadyen A I 2003 *Astrophys. J.* **586** 356
Zygelman B 2005 *Astrophys. J.* **622** 1356