21 cm cosmology in the 21st century

Jonathan R Pritchard and Abraham Loeb

Institute for Theory and Computation, Harvard University, 60 Garden St., Cambridge, MA 02138, USA

E-mail: jpritchard@cfa.harvard.edu and aloeb@cfa.harvard.edu

Abstract

Imaging the Universe during the first hundreds of millions of years remains one of the exciting challenges facing modern cosmology. Observations of the redshifted 21 cm line of atomic hydrogen offer the potential of opening a new window into this epoch. This will transform our understanding of the formation of the first stars and galaxies and of the thermal history of the Universe. A new generation of radio telescopes is being constructed for this purpose with the first results starting to trickle in. In this review, we detail the physics that governs the 21 cm signal and describe what might be learnt from upcoming observations. We also generalize our discussion to intensity mapping of other atomic and molecular lines.

(Some figures may appear in colour only in the online journal)

This article was invited by K Kirby.

Contents

1. Introduction

Our understanding of cosmology has matured significantly over the last 20 years. In that time, observations of the Universe from its infancy, 400 000 years after the Big Bang, through to the present day, some 13.7 billion years later, have given us a basic picture of how the Universe came to be the way it is today. Despite this progress much of the first billion years of the Universe, a period when the first stars and galaxies formed, is still an unobserved mystery.

Astronomers have an advantage over archaeologists in that the finite speed of light gives them a way of looking into the past. The further away an object is located the longer the light that it emits takes to reach an observer today. The image recorded at a telescope is therefore a picture of the object long ago when the light was first emitted. The construction

of telescopes both on the Earth, such as Keck, Subaru and Very Large Telescope (VLT), and in space, such as the Hubble Space Telescope, has enabled astronomers to directly observe galaxies out to distances corresponding to a time when the Universe was a billion years old.

Added to this, observations at microwave frequencies reveal the cooling afterglow of the Big Bang. This cosmic microwave background (CMB) decoupled from the cosmic gas 400 000 years after the Big Bang when the Universe cooled sufficiently for protons and electrons to combine to form neutral hydrogen. Radiation from this time is able to reach us directly, providing a snapshot of the primordial Universe.

Despite current progress, connecting these two periods represents a considerable challenge. Our understanding of structure is based upon the observation of small perturbations in the temperature maps of the CMB. These indicate that the early Universe was inhomogeneous at the level of 1 part in 100 000. Over time the action of gravity causes the growth of these small perturbations into larger non-linear structures, which collapse to form sheets, filaments and halos. These non-linear structures provide the framework within which galaxies form via the collapse and cooling of gas until the density required for star formation is reached.

The theoretical picture is well established, but the middle phase is largely untested by observations. To improve on this astronomers are pursuing two main avenues of attack. The first is to extend existing techniques by building larger, more sensitive, telescopes at a variety of wavelengths. On the ground, there are plans for optical telescopes with an aperture diameter of 24–39 m—the Giant Magellan Telescope (GMT), the Thirty Meter Telescope (TMT) and the European Extremely Large Telescope (E-ELT)—that would be able to detect an individual galaxy out to redshifts $z > 10$. In space, the James Webb Space Telescope (JWST) will operate at infrared wavelengths and potentially image some of the first galaxies at $z \sim 10$–15. Other efforts involve the Atacama Large Millimeter/Submillimeter Array (ALMA), which will observe the molecular gas that fuels star formation in galaxies during reionization ($z = 8 - 10$). These efforts target individual galaxies although the objects of interest are far enough away that only the brightest sources may be seen.

This review focuses on an alternative approach based upon making observations of the redshifted 21 cm line of neutral hydrogen. This 21 cm line is produced by the hyperfine splitting caused by the interaction between electron and proton magnetic moments. Hydrogen is ubiquitous in the Universe, amounting to ∼75% of the gas mass present in the intergalactic medium (IGM). As such, it provides a convenient tracer of the properties of that gas and of major milestones in the first billion years of the Universe's history.

The 21 cm line from gas during the first billion years after the Big Bang redshifts to radio frequencies 30–200 MHz making it a prime target for a new generation of radio interferometers currently being built. These instruments, such as the Murchison Widefield Array (MWA), the LOw Frequency ARray (LOFAR), the Precision Array to Probe the Epoch of Reionization (PAPER), the 21 cm Array (21CMA) and the Giant Meter-wave Radio Telescope (GMRT), seek to detect

the radio fluctuations in the redshifted 21 cm background arising from variations in the amount of neutral hydrogen. Next generation instruments (e.g. the Square Kilometre Array (SKA)) will be able to go further and might make detailed maps of the ionized regions during reionization and measure properties of hydrogen out to $z = 30$. These observations constrain the properties of the IGM and by extension the cumulative impact of light from all galaxies, not just the brightest ones. In combination with direct observations of the sources they provide a powerful tool for learning about the first stars and galaxies. They will also provide information about active galactic nuclei (AGNs), such as quasars, by observing the ionized bubbles surrounding individual AGNs.

In addition to learning about galaxies and reionization, 21 cm observations have the potential to inform us about fundamental physics too. Part of the signal traces the density field giving information about neutrino masses and the initial conditions from the early epoch of cosmic inflation in the form of the power spectrum. However spin-temperature fluctuations driven by astrophysics also contribute to the signal. Getting at this cosmology is a challenge, since the astrophysical effects must be understood before cosmology can be disentangled. One possibility is to exploit the effect of redshift space distortions, which also produce 21 cm fluctuations but directly trace the density field. In the long term, 21 cm cosmology may allow precision measurements of cosmological parameters by opening up large volumes of the Universe to observation.

The goal of this review is to summarize the physics that determines the 21 cm signal, along with a comprehensive overview of related astrophysics. Figure 1 provides a summary of the 21 cm signal showing the key features of the signal with the relevant cosmic time, frequency and redshift scales indicated. The earliest period of the signal arises in the period after thermal decoupling of the ordinary matter (baryons) from the CMB, so that the gas is able to cool adiabatically with the expansion of the Universe. In these cosmic 'Dark Ages', before the first stars have formed, the first structures begin to grow from the seed inhomogeneties thought to be produced by quantum fluctuations during inflation. The cold gas can be seen in a 21 cm absorption signal, which has both a mean value (shown in the bottom panel) and fluctuations arising from variation in density (shown in the top panel). Once the first stars and galaxies form, their light radically alters the properties of the gas. Scattering of Lyα photons leads to a strong coupling between the excitation of the 21 cm line spin states and the gas temperature. Initially, this leads to a strong absorption signal that is spatially varying due to the strong clustering of the rare first generation of galaxies. Next, the x-ray emission from these galaxies heats the gas leading to a 21 cm emission signal. Finally, UV photons ionize the gas producing dark holes in the 21 cm signal within regions of ionized bubbles surrounding groups of galaxies. Eventually all of the hydrogen gas, except for that in a few dense pockets, is ionized.

Throughout this review, we will make reference to parameters describing the standard ΛCDM cosmology. These describe the mass densities in non-relativistic matter $\Omega_m = 0.26$, dark energy $\Omega_\Lambda = 0.74$ and baryons $\Omega_b = 0.044$ as a fraction of the critical mass density. We further parametrize
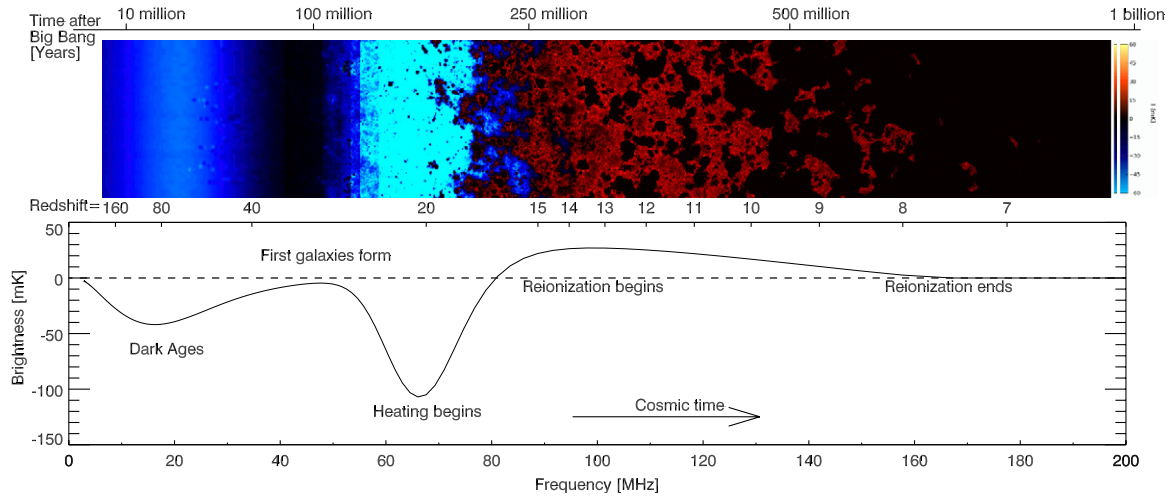
Figure 1. The 21 cm cosmic hydrogen signal. (*a*) Time evolution of fluctuations in the 21 cm brightness from just before the first stars formed through to the end of the reionization epoch. This evolution is pieced together from redshift slices through a simulated cosmic volume [1]. Coloration indicates the strength of the 21 cm brightness as it evolves through two absorption phases (purple and blue), separated by a period (black) where the excitation temperature of the 21 cm hydrogen transition decouples from the temperature of the hydrogen gas, before it transitions to emission (red) and finally disappears (black) owing to the ionization of the hydrogen gas. (*b*) Expected evolution of the sky-averaged 21 cm brightness from the 'Dark Ages' at redshift 200 to the end of reionization, sometime before redshift 6 (solid curve indicates the signal; dashed curve indicates $T_{\mathrm{b}} = 0$). The frequency structure within this redshift range is driven by several physical processes, including the formation of the first galaxies and the heating and ionization of the hydrogen gas. There is considerable uncertainty in the exact form of this signal, arising from the unknown properties of the first galaxies. Reproduced with permission from [2]. Copyright 2010 Nature Publishing Group.

the Hubble parameter $H_0 = 100h \, \mathrm{km \, s^{-1} \, Mpc^{-1}}$ with $h = 0.74$. Finally, the spectrum of fluctuations is described by a logarithmic slope or 'tilt' $n_{\mathrm{S}} = 0.95$, and the variance of matter fluctuations today smoothed on a scale of $8h^{-1} \, \mathrm{Mpc}$ is $\sigma_8 = 0.8$. The values quoted are indicative of those found by the latest measurements [3].

The layout of this review is as follows. We first discuss the basic atomic physics of the 21 cm line in section 2. In section 3, we turn to the evolution of the sky-averaged 21 cm signal and the feasibility of observing it. In section 4 we describe 3D 21 cm fluctuations, including predictions from analytical and numerical calculations. After reionization, most of the 21 cm signal originates from cold gas in galaxies (which is self-shielded from the background of ionizing radiation). In section 5 we describe the prospects for intensity mapping (IM) of this signal as well as using the same technique to map the cumulative emission of other atomic and molecular lines from galaxies without resolving the galaxies individually. The 21 cm forest that is expected against radio-bright sources is described in section 6. Finally, we conclude with an outlook for the future in section 7.

We direct interested readers to a number of other worthy reviews on the subject. Reference [4] provides a comprehensive overview of the entire field, and [5] takes a more observationally orientated approach focusing on the near term observations of reionization.

2. Physics of the 21 cm line of atomic hydrogen

2.1. Basic 21 cm physics

As the most common atomic species present in the Universe, hydrogen is a useful tracer of local properties of the gas.

The simplicity of its structure—a proton and electron—belies the richness of the associated physics. In this review, we will be focusing on the 21 cm line of hydrogen, which arises from the hyperfine splitting of the 1S ground state due to the interaction of the magnetic moments of the proton and the electron. This splitting leads to two distinct energy levels separated by $\Delta E = 5.9 \times 10^{-6} \, \mathrm{eV}$, corresponding to a wavelength of 21.1 cm and a frequency of 1420 MHz. This frequency is one of the most precisely known quantities in astrophysics having been measured to great accuracy from studies of hydrogen masers [6].

The 21 cm line was theoretically predicted by van de Hulst in 1942 [7] and has been used as a probe of astrophysics since it was first detected by Ewen and Purcell in 1951 [8]. Radio telescopes look for emission by warm hydrogen gas within galaxies. Since the line is narrow with a well measured rest frame frequency it can be used in the local Universe as a probe of the velocity distribution of gas within our galaxy and other nearby galaxies. The 21 cm rotation curves are often used to trace galactic dynamics. Traditional techniques for observing 21 cm emission have only detected the line in relatively local galaxies, although the 21 cm line has been seen in absorption against radio-loud background sources from individual systems at redshifts $z \lesssim 3$ [9, 10]. A new generation of radio telescopes offers the exciting prospect of using the 21 cm line as a probe of cosmology.

In passing, we note that other atomic species show hyperfine transitions that may be useful in probing cosmology. Of particular interest are the 8.7 GHz hyperfine transition of $^3\mathrm{He}^+$ [11, 12], which could provide a probe of helium reionization, and the 92 cm deuterium analogue of the 21 cm line [13]. The much lower abundance of deuterium and $^3\mathrm{He}$ compared with neutral hydrogen makes it more difficult to take advantage of these transitions.

In cosmological contexts the 21 cm line has been used as a probe of gas along the line of sight to some background radio source. The detailed signal depends upon the radiative transfer through gas along the line of sight. We recall the basic equation of radiative transfer for the specific intensity I_ν (per unit frequency ν) in the absence of scattering along a path described by coordinate s [14]:

$$\frac{dI_\nu}{ds} = -\alpha_\nu I_\nu + j_\nu, \tag{1}$$

where absorption and emission by gas along the path are described by the coefficients α_ν and j_ν, respectively.

To simplify the discussion, we will work in the Rayleigh–Jeans limit, appropriate here since the relevant photon frequencies ν are much smaller than the peak frequency of the CMB blackbody. This allows us to relate the intensity I_ν to a brightness temperature T by the relation $I_\nu = 2k_B T\nu^2/c^2$, where c is the speed of light and k_B is Boltzmann's constant. We will also make use of the standard definition of the optical depth $\tau = \int ds\, \alpha_\nu(s)$. With this we may rewrite (1) to give the radiative transfer for light from a background radio source of brightness temperature T_R along the line of sight through a cloud of optical depth τ_ν and uniform excitation temperature T_{ex} so that the observed temperature T_b^{obs} at a frequency ν is given by

$$T_b^{obs} = T_{ex}(1 - e^{-\tau_\nu}) + T_R(\nu)e^{-\tau_\nu}. \tag{2}$$

The excitation temperature of the 21 cm line is known as the spin temperature T_S. It is defined through the ratio between the number densities n_i of hydrogen atoms in the two hyperfine levels (which we label with a subscript 0 and 1 for the 1S singlet and 1S triplet levels, respectively)

$$n_1/n_0 = (g_1/g_0)\exp(-T_\star/T_S), \tag{3}$$

where $g_1/g_0 = 3$ is the ratio of the statistical degeneracy factors of the two levels, and $T_\star \equiv hc/k\lambda_{21\,cm} = 0.068$ K.

With this definition, the optical depth of a cloud of hydrogen is then

$$\tau_\nu = \int ds\, [1 - \exp(-E_{10}/k_B T_S)]\sigma_0\phi(\nu)n_0, \tag{4}$$

where $n_0 = n_H/4$ with n_H being the hydrogen density, and we have denoted the 21 cm cross-section as $\sigma(\nu) = \sigma_0\phi(\nu)$, with $\sigma_0 \equiv 3c^2 A_{10}/8\pi\nu^2$, where $A_{10} = 2.85 \times 10^{-15}$ s^{-1} is the spontaneous decay rate of the spin–flip transition, and the line profile is normalized so that $\int \phi(\nu)\, d\nu = 1$. To evaluate this expression we need to find the column length as a function of frequency $s(\nu)$ to determine the range of frequencies $d\nu$ over the path ds that correspond to a fixed observed frequency ν_{obs}. This can be done in one of two ways: by relating the path length to the cosmological expansion $ds = -c\, dz/(1 + z)H(z)$ and the redshifting of light to relate the observed and emitted frequencies $\nu_{obs} = \nu_{em}/(1 + z)$ or assuming a linear velocity profile locally $\nu = (d\nu/ds)s$ (the well-known Sobolev approximation [15]) and using the Doppler law $\nu_{obs} = \nu_{em}(1 - \nu/c)$ self-consistently to $\mathcal{O}(\nu/c)$. Since the latter case describes the well-known Hubble law

in the absence of peculiar velocities these two approaches give identical results for the optical depth. The latter picture brings out the effect of peculiar velocities that modify the local velocity–frequency conversion.

The optical depth of this transition is small at all relevant redshifts, yielding a differential brightness temperature

$$\delta T_b = \frac{T_S - T_R}{1 + z}(1 - e^{-\tau_\nu}) \tag{5}$$

$$\approx \frac{T_S - T_R}{1 + z}\tau \tag{6}$$

$$\approx 27x_{HI}(1 + \delta_b)\left(\frac{\Omega_b h^2}{0.023}\right)\left(\frac{0.15}{\Omega_m h^2}\frac{1 + z}{10}\right)^{1/2}$$
$$\times \left(\frac{T_S - T_R}{T_S}\right)\left[\frac{\partial_r v_r}{(1 + z)H(z)}\right] \text{mK}, \tag{7}$$

Here x_{HI} is the neutral fraction of hydrogen, δ_b is the fractional overdensity in baryons and the final term arises from the velocity gradient along the line of sight $\partial_r v_r$.

The key to the detectability of the 21 cm signal hinges on the spin temperature. Only if this temperature deviates from the background temperature, will a signal be observable. Much of this review will focus on the physics that determines the spin temperature and how spatial variation in the spin temperature conveys information about astrophysical sources.

Three processes determine the spin temperature: (i) absorption/emission of 21 cm photons from/to the radio background, primarily the CMB; (ii) collisions with other hydrogen atoms and with electrons; and (iii) resonant scattering of Lyα photons that cause a spin–flip via an intermediate excited state. The rate of these processes is fast compared with the deexcitation time of the line, so that to a very good approximation the spin temperature is given by the equilibrium balance of these effects. In this limit, the spin temperature is given by [16]

$$T_S^{-1} = \frac{T_\gamma^{-1} + x_\alpha T_\alpha^{-1} + x_c T_K^{-1}}{1 + x_\alpha + x_c}, \tag{8}$$

where T_γ is the temperature of the surrounding bath of radio photons, typically set by the CMB so that $T_\gamma = T_{CMB}$; T_α is the color temperature of the Lyα radiation field at the Lyα frequency and is closely coupled to the gas kinetic temperature T_K by recoil during repeated scattering and x_c, x_α are the coupling coefficients due to atomic collisions and scattering of Lyα photons, respectively. The spin temperature becomes strongly coupled to the gas temperature when $x_{tot} \equiv x_c + x_\alpha \gtrsim 1$ and relaxes to T_γ when $x_{tot} \ll 1$.

Two types of background radio sources are important for the 21 cm line as a probe of astrophysics. Firstly, we may use the CMB as a radio background source. In this case, $T_R = T_{CMB}$ and the 21 cm feature is seen as a spectral distortion to the CMB blackbody at appropriate radio frequencies (since fluctuations in the CMB temperature are small $\delta T_{CMB} \sim 10^{-5}$ the CMB is effectively a source of uniform brightness). The distortion forms a diffuse background that can be studied across the whole sky in a similar way to CMB anisotropies. Observations at different frequencies probe different spherical
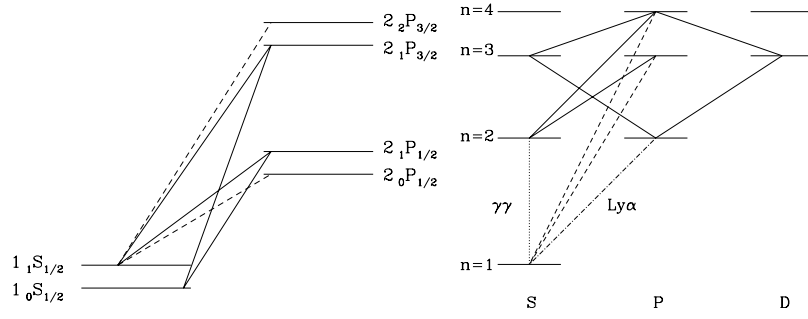
Figure 2. Left panel: hyperfine structure of the hydrogen atom and the transitions relevant for the Wouthuysen–Field effect [25]. Solid line transitions allow spin–flips, while dashed transitions are allowed but do not contribute to spin–flips. Right panel: illustration of how atomic cascades convert Lyn photons into Lyα photons. Reproduced with permission from [25]. Copyright 2006 Wiley.

shells of the observable Universe, so that a 3D map can be constructed. This is the main subject of section 4.

The second situation uses a radio-loud point source, for example a radio-loud quasar, as the background. In this case, the source will always be much brighter than the weak emission from diffuse hydrogen gas, $T_R \gg T_S$, so that the gas is seen in absorption against the source. The appearance of lines from regions of neutral gas at different distances to the source leads to a 'forest' of lines known as the '21 cm forest' in analogy to the Lyα forest. The high brightness of the background source allows the 21 cm forest to be studied with high frequency resolution so probing small-scale structures ($\sim$kpc) in the IGM. For useful statistics, many lines of sight to different radio sources are required, making the discovery of high redshift radio sources a priority. We leave discussion of the 21 cm forest to section 6.

Note that we have a number of different quantities with units of temperature, many of which are not true thermodynamic temperatures. T_R and δT_b are measures of radio intensity. T_S measures the relative occupation numbers of the two hyperfine levels. T_α is a colour temperature describing the photon distribution in the vicinity of the Lyα transition. Only the CMB blackbody temperature T_{CMB} and T_K are genuine thermodynamic temperatures.

2.2. Collisional coupling

Collisions between different particles may induce spin–flips in a hydrogen atom and dominate the coupling in the early Universe where the gas density is high. Three main channels are available: collisions between two hydrogen atoms and collisions between a hydrogen atom and an electron or a proton. The collisional coupling for a species i is [4, 16]

$$x_c^i \equiv \frac{C_{10}}{A_{10}} \frac{T_\star}{T_\gamma} = \frac{n_i \kappa_{10}^i}{A_{10}} \frac{T_\star}{T_\gamma}, \qquad (9)$$

where C_{10} is the collisional excitation rate and κ_{10}^i is the specific rate coefficient for spin deexcitation by collisions with species i (in units of cm^3 s^{-1}).

The total collisional coupling coefficient can be written as

$$x_c = x_c^{HH} + x_c^{eH} + x_c^{pH}$$
$$= \frac{T_\star}{A_{10} T_\gamma} \left[\kappa_{1-0}^{HH}(T_k) n_H + \kappa_{1-0}^{eH}(T_k) n_e + \kappa_{1-0}^{pH}(T_k) n_p \right], \quad (10)$$

where κ_{1-0}^{HH} is the scattering rate between hydrogen atoms, κ_{1-0}^{eH} is the scattering rate between electrons and hydrogen atoms and κ_{1-0}^{pH} is the scattering rate between protons and hydrogen atoms.

The collisional rates require a quantum mechanical calculation. Values for κ_{1-0}^{HH} have been tabulated as a function of T_k [17, 18], the scattering rate between electrons and hydrogen atoms κ_{1-0}^{eH} was considered in [19] and the scattering rate between protons and hydrogen atoms κ_{1-0}^{pH} was considered in [20]. Useful fitting functions exist for these scattering rates: the H–H scattering rate is well fit in the range $10\,K < T_K < 10^3\,K$ by $\kappa_{1-0}^{HH}(T_K) \approx 3.1 \times 10^{-11} T_K^{0.357} \exp(-32/T_K)\,cm^3\,s^{-1}$ [21]; and the e–H scattering rate is well fit by $\log(\kappa_{1-0}^{eH}/cm^3\,s^{-1}) = -9.607 + 0.5 \log T_K \times \exp[-(\log T_K)^{4.5}/1800]$ for $T \leqslant 10^4\,K$ and $\kappa_{1-0}^{eH}(T_K > 10^4\,K) = \kappa_{1-0}^{eH}(10^4\,K)$ [22].

During the cosmic Dark Ages, where the coupling is dominated by collisional coupling the details of the process become important. For example, the above calculations make use of the assumption that the collisional cross-sections are independent of velocity; the actual velocity dependance leads to a non-thermal distribution for the hyperfine occupation [23]. This effect can lead to a suppression of the 21 cm signal at the level of 5%, which although small is still important from the perspective of using the 21 cm signal from the Dark Ages for precision cosmology.

2.3. Wouthuysen–Field effect

For most of the redshifts that are likely to be observationally probed in the near future collisional coupling of the 21 cm line is inefficient. However, once star formation begins, resonant scattering of Lyα photons provides a second channel for coupling. This process is generally known as the Wouthuysen–Field effect [16, 24] and is illustrated in figure 2, which shows the hyperfine structure of the hydrogen 1S and 2P levels. Suppose that hydrogen is initially in the hyperfine singlet state. Absorption of a Lyα photon will excite the atom into either of the central 2P hyperfine states (the dipole selection rules $\Delta F = 0, 1$ and no $F = 0 \rightarrow 0$ transitions make the other two hyperfine levels inaccessible). From here emission of a Lyα photon can relax the atom to either of the two ground state hyperfine levels. If relaxation takes the atom to the ground

level triplet state then a spin–flip has occurred. Hence, resonant scattering of Lyα photons can produce a spin–flip.

The physics of the Wouthuysen–Field effect is considerably more subtle than this simple description would suggest. We may write the coupling as

$$x_\alpha = \frac{4P_\alpha}{27A_{10}} \frac{T_\star}{T_\gamma},$$ (11)

where P_α is the scattering rate of Lyα photons. Here we have related the scattering rate between the two hyperfine levels to P_α using the relation $P_{01} = 4P_\alpha/27$, which results from the atomic physics of the hyperfine lines and assumes that the radiation field is constant across them [26].

The rate at which Lyα photons scatter from a hydrogen atom is given by

$$P_\alpha = 4\pi \chi_\alpha \int d\nu \, J_\nu(\nu)\phi_\alpha(\nu),$$ (12)

where $\sigma_\nu \equiv \chi_\alpha \phi_\alpha(\nu)$ is the local absorption cross-section, $\chi_\alpha \equiv (\pi e^2/m_e c) f_\alpha$ is the oscillation strength of the Lyα transition, $\phi_\alpha(\nu)$ is the Lyα absorption profile and $J_\nu(\nu)$ is the angle-averaged specific intensity of the background radiation field (by number).

Making use of this expression, we can express the coupling as

$$x_\alpha = \frac{16\pi^2 T_\star e^2 f_\alpha}{27A_{10}T_\gamma m_e c} S_\alpha J_\alpha,$$ (13)

where J_α is the specific flux evaluated at the Lyα frequency. Here we have introduced $S_\alpha \equiv \int dx \phi_\alpha(x) J_\nu(x)/J_\infty$, with J_∞ being the flux away from the absorption feature, as a correction factor of order unity to describe the detailed structure of the photon distribution in the neighbourhood of the Lyα resonance.

Equation (13) can be used to calculate the critical flux required to produce $x_\alpha = S_\alpha$. We rewrite (13) as $x_\alpha = S_\alpha J_\alpha/J_\alpha^C$, where $J_\alpha^C \equiv 1.165 \times 10^{10}[(1 + z)/20] \, \text{cm}^{-2} \, \text{s}^{-1} \, \text{Hz}^{-1} \, \text{sr}^{-1}$. The critical flux can also be expressed in terms of the number of Lyα photons per hydrogen atom $J_\alpha^C/n_H = 0.0767[(1 + z)/20]^{-2}$. In practice, this condition is easy to satisfy once star formation begins.

The above physics couples the spin temperature to the colour temperature of the radiation field, which is a measure of the shape of the radiation field as a function of frequency in the neighbourhood of the Lyα line defined by [27]

$$\frac{h}{k_B T_c} = -\frac{d \log n_\nu}{d\nu},$$ (14)

where $n_\nu = c^2 J_\nu/2\nu^2$ is the photon occupation number. Some care must to taken with this definition; other definitions that do not obey detailed balance can be found in the literature.

Typically, $T_c \approx T_K$, because in most cases of interest the optical depth to Lyα scattering is very large leading to a large number of scatterings of Lyα photons that bring the radiation field and the gas into local equilibrium for frequencies near the line centre [28]. At the level of microphysics this relation occurs through the process of scattering Lyα photons in the neighbourhood of the Lyα resonance, which leads to a distinct feature in the frequency distribution of photons. Without going into the details, one can understand the formation of this feature in terms of the 'flow' of photons in frequency. Redshifting with the cosmic expansion leads to a flow of photons from high to low frequency at a fixed rate. As photons flow into the Lyα resonance they may scatter to larger or smaller frequencies. Since the cross-section is symmetric, one would expect the net flow rate to be preserved. However, each time a Lyα photon scatters from a hydrogen atom it will lose a fraction of its energy $h\nu/m_p c^2$ due to the recoil of the atom. This loss of energy increases the flow to lower energy and leads to a deficit of photons close to line centre. As this feature develops scattering redistributes photons leading to an asymmetry about the line. This asymmetry is exactly that required to bring the distribution into local thermal equilibrium with $T_c \approx T_K$.

The shape of this feature determines S_α and, since recoils source an absorption feature, ensures $S_\alpha \leqslant 1$. At low temperatures, recoils have more of an effect and the suppression of the Wouthuysen–Field effect is most pronounced. If the IGM is warm then this suppression becomes negligible [29–32]. The above discussion has neglected processes whereby the distribution of photons is changed by spin-exchanges. Including this complicates the determination of T_S and T_c considerably since they must then be iterated to find a self-consistent solution for the level- and photon-populations [30]. However, the effect of spin–flips on the photon distribution is small $\lesssim 10\%$.

A useful approximation for S_α is outlined in [31]: $S_\alpha \approx \exp(-1.79\alpha)$, where $\alpha \equiv \eta(3a/2\pi\gamma)^{1/3}$, $a = \Gamma/(4\pi\Delta\nu_D)$, Γ the inverse lifetime of the upper 21 cm level, $\Delta\nu_D/\nu_0 = (2k_B T_K/mc^2)^{1/2}$ is the Doppler parameter, ν_0 the line centre frequency, $\gamma = \tau_{GP}^{-1}$ and $\eta = (h\nu_0^2)/(mc^2\Delta\nu_D)$ is the mean frequency drift per scattering due to recoil, which is accurate at the 5% level provided that $T_K \gtrsim 1$ K and the Gunn–Peterson optical depth τ_{GP} is large.

In the astrophysical context, we will primarily be interested in photons redshifting into the Lyα resonance from frequencies below the Lyβ resonance. In addition, Lyα photons can be produced by atomic cascades from photons redshifting into higher Lyman series resonances. These atomic cascades are illustrated in figure 2, where the probability of converting a Lyn photon into a Lyα photon is set by the atomic rate coefficients and can be found in tabular form in [25, 30]. For large n, approximately 30% conversion is typical. These photons are injected into the Lyα line rather than being redshifted from outside of the line. This changes their contribution to the Wouthuysen–Field coupling since the photon distribution is now one-sided. Similar processes to those described above apply to the redistribution of these photons, and they can lead to an important amplification of the Lyα flux.

This discussion gives a sense of some of the subtleties that go into determining the strength of the Lyα coupling. These effects can modify the 21 cm signal at the ~10% level, which will be important as observations begin to detect 21 cm fluctuations. At this stage, it appears that the underlying atomic physics is understood, although the details of Lyα radiative transfer still requires some work.
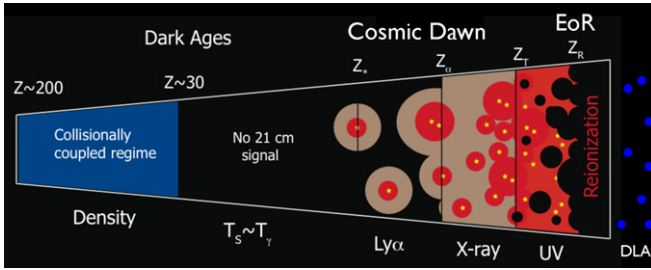
Figure 3. Cartoon of the different phases of the 21 cm signal. The signal transitions from an early phase of collisional coupling to a later phase of Lyα coupling through a short period where there is little signal. Fluctuations after this phase are dominated successively by spatial variation in the Lyα, x-ray and ionizing UV radiation backgrounds. After reionization is complete there is a residual signal from neutral hydrogen in galaxies.

3. Global 21 cm signature

3.1. Outline

Next we examine the cosmological context of the 21 cm signal. We may express the 21 cm brightness temperature as a function of four variables $T_b = T_b(T_K, x_i, J_\alpha, n_H)$, where x_i is the volume-averaged ionized fraction of hydrogen. In calculating the 21 cm signal, we require a model for the global evolution of and fluctuations in these quantities. Before looking at the evolution of the signal quantitatively, we will first outline the basic picture to delineate the most important phases.

An important feature of T_b is that its dependence on each of these quantities saturates at some point, for example once the Lyα flux is high enough the spin and kinetic gas temperatures become tightly coupled and further variation in J_α becomes irrelevant to the details of the signal. This leads to conceptually separate regimes where variation in only one of the variables dominating fluctuations in the signal. These different regimes can be seen in figure 1 and are shown in schematic form in figure 3 for clarity. We now discuss each of these phases in turn.

- $200 \lesssim z \lesssim 1100$. The residual free electron fraction left after recombination allows Compton scattering to maintain thermal coupling of the gas to the CMB, setting $T_K = T_\gamma$. The high gas density leads to effective collisional coupling so that $T_S = T_\gamma$ and we expect $\bar{T}_b = 0$ and no detectable 21 cm signal.
- $40 \lesssim z \lesssim 200$. In this regime, the gas cools adiabatically so that $T_K \propto (1 + z)^2$ leading to $T_K < T_\gamma$ and collisional coupling sets $T_S < T_\gamma$, leading to $\bar{T}_b < 0$ and an early absorption signal. At this time, T_b fluctuations are sourced by density fluctuations, potentially allowing the initial conditions to be probed [23, 33].
- $z_\star \lesssim z \lesssim 40$. As the expansion continues, decreasing the gas density, collisional coupling becomes ineffective and radiative coupling to the CMB sets $T_S = T_\gamma$, and there is no detectable 21 cm signal.
- $z_\alpha \lesssim z \lesssim z_\star$. Once the first sources switch on at $z_\star$, they emit both Lyα photons and x-rays. In general, the emissivity required for Lyα coupling is significantly less than that for heating T_K above T_γ. We therefore expect

a regime where the spin temperature is coupled to cold gas so that $T_S \sim T_K < T_\gamma$ and there is an absorption signal. Fluctuations are dominated by density fluctuations and variation in the Lyα flux [25, 34, 35]. As further star formation occurs the Lyα coupling will eventually saturate ($x_\alpha \gg 1$), so that by a redshift z_α the gas will everywhere be strongly coupled.

- $z_h \lesssim z \lesssim z_\alpha$. After Ly$\alpha$ coupling saturates, fluctuations in the Lyα flux no longer affect the 21 cm signal. By this point, heating becomes significant and gas temperature fluctuations source T_b fluctuations. While T_K remains below T_γ we see a 21 cm signal in absorption, but as T_K approaches T_γ hotter regions may begin to be seen in emission. Eventually by a redshift z_h the gas will be heated everywhere so that $\bar{T}_K = T_\gamma$.
- $z_T \lesssim z \lesssim z_h$. After the heating transition, $T_K > T_\gamma$ and we expect to see a 21 cm signal in emission. The 21 cm brightness temperature is not yet saturated, which occurs at z_T, when $T_S \sim T_K \gg T_\gamma$. By this time, the ionization fraction has likely risen above the per cent level. Brightness temperature fluctuations are sourced by a mixture of fluctuations in ionization, density and gas temperature.
- $z_r \lesssim z \lesssim z_T$. Continued heating drives $T_K \gg T_\gamma$ at z_T and temperature fluctuations become unimportant. $T_S \sim T_K \gg T_\gamma$ and the dependence on T_S may be neglected in equation (7), which greatly simplifies analysis of the 21 cm power spectrum [36]. By this point, the filling fraction of H II regions probably becomes significant and ionization fluctuations begin to dominate the 21 cm signal [37].
- $z \lesssim z_r$. After reionization, any remaining 21 cm signal originates primarily from collapsed islands of neutral hydrogen (damped Lyα systems).

Most of these epochs are not sharply defined, and so there could be considerable overlap between them. In fact, our ignorance of early sources is such that we cannot definitively be sure of the sequence of events. The above sequence of events seems most likely and can be justified on the basis of the relative energetics of the different processes and the probable properties of the sources. We will discuss this in more detail as we quantify the evolution of the sources.

Perhaps the largest uncertainty lies in the ordering of z_α and z_h. Reference [38] explores the possibility that $z_h > z_\alpha$, so that x-ray preheating allows collisional coupling to be important before the Lyα flux becomes significant. Simulations of the very first mini-quasar [21, 39] also probe this regime and show that the first luminous x-ray sources can have a great impact on their surrounding environment. We note that these studies ignored Lyα coupling, and that an x-ray background may generate significant Lyα photons [35], as we discuss in section 3.5. Additionally, while these authors looked at the case where the production of Lyα photons was inefficient, one can consider the case where heating is much more efficient. This can be the case where weak shocks raise the IGM temperature very early on [40] or if exotic particle physics mechanisms such as dark matter annihilation are important. Clearly, there is still considerable uncertainty in the exact evolution of the signal making the potential implications of measuring the 21 cm signal very exciting.

3.2. Evolution of global signal

Having outlined the evolution of the signal qualitatively, we will turn to the details of making quantitative predictions. In calculating the 21 cm signal it will help us to treat the IGM as a two phase medium. Initially, the IGM is composed of a single mostly neutral phase left over after recombination. This phase is characterized by a gas temperature T_K and a small fraction of free electrons x_e. This is the phase that generates the 21 cm signal.

Once galaxy formation begins, energetic UV photons ionize H II regions surrounding, first individual galaxies and then clusters of galaxies. These UV photons have a very short mean free path in a neutral medium leading to the ionized H II regions having a very sharp boundary (although the boundary can be softened if the ionizing photons are particularly hard [41]). We may therefore treat the ionized H II bubbles as a second phase in the IGM characterized by a volume filling fraction x_i (provided that the free electron fraction is small x_i is approximately the mean ionization fraction). We will assume that these bubbles are fully ionized and that the temperature inside the bubbles is fixed at $T_{H II} = 10^4$ K determining the collisional recombination rate inside these bubbles. Since the photons that redshift into the Lyα resonance initially have long mean free paths, we may treat the Lyα flux J_α as being the same in both phases (although in practice, since there is no 21 cm signal from the fully ionized bubbles, it is only the Lyα flux in the mostly neutral phase that matters). To determine the 21 cm signal at a given redshift, we must calculate the four quantities x_i, x_e, T_K and J_α. We begin by describing the evolution of the gas temperature T_K,

$$\frac{dT_K}{dt} = \frac{2T_K}{3n}\frac{dn}{dt} + \frac{2}{3k_B}\sum_j \frac{\epsilon_j}{n}. \tag{15}$$

Here, the first term accounts for adiabatic cooling of the gas due to the cosmic expansion while the second term accounts for other sources of heating/cooling j with ϵ_j the heating rate per unit volume for the process j.

Next, we consider the volume filling fraction x_i and the ionization of the neutral IGM x_e

$$\frac{dx_i}{dt} = (1 - x_e)\Lambda_i - \alpha_A C x_i^2 n_H, \tag{16}$$

$$\frac{dx_e}{dt} = (1 - x_e)\Lambda_e - \alpha_B(T) x_e^2 n_H. \tag{17}$$

In these expressions, we define Λ_i to be the rate of production of ionizing photons per unit time per baryon applied to H II regions, Λ_e is the equivalent quantity in the bulk of the IGM, $\alpha_A = 4.2 \times 10^{-13}$ cm^3 s^{-1} is the case-A recombination coefficient at $T = 10^4$ K, $\alpha_B(T)$ is the case-B recombination rate (whose temperature dependence can be obtained from [42]), and $C \equiv \langle n_e^2 \rangle / \langle n_e \rangle^2$ is the clumping factor.

Superficially these look the same, since in each case the ionization rate is a balance between ionizations and recombinations. The main distinction lies in the manner in which we treat the recombinations. In the fully ionized bubbles, recombinations occur in those dense clumps of material capable of self-shielding against ionizing radiation. These overdense regions will have a locally enhanced recombination rate, making it important to account for the inhomogeneous distribution of matter through the clumping factor C. Since recombinations will occur on the edge of these neutral clumps, secondary photons produced by the recombinations will likely be absorbed inside the clumps rather than in the mean IGM, justifying the use of case-A recombination [43]. In contrast, recombinations in the bulk of the neutral IGM will occur at close to mean density in gas with temperature T_K. Here recombination radiation will be absorbed in the IGM, so we must use case-B recombination. By keeping track of this carefully our evolution matches that of RECFAST [42].

This two phase approximation will eventually break down should x_e become close to unity, indicating that most of the IGM has been ionized and that there is no clear distinction between ionized bubbles and a neutral bulk IGM. In most of our models, x_e remains small until the end of reionization making this a reasonable approximation.

3.3. Growth of H II regions

The growth of ionized H II regions is governed by the interplay between ionization and recombination, both of which contain considerable uncertainties. We may write the ionization rate per hydrogen atom as

$$\Lambda_i = A_{He} f_{esc} N_{ion} \dot{\rho}_\star(z), \tag{18}$$

with N_{ion} being the number of ionizing photons per baryon produced in stars, f_{esc} the fraction of ionizing photons that escape the host halo and A_{He} a correction factor for the presence of helium. Here $\dot{\rho}_\star(z)$ is the star-formation rate (SFR) density as a function of redshift, which is still poorly known observationally. For a present-day initial mass function of stars, $N_{ion} \sim 4 \times 10^3$, whereas for very massive ($>10^2 M_\odot$) stars of primordial composition, $N_{ion} \sim 10^5$ [44, 45].

We model the SFR as tracking the collapse of matter, so that we may write the SFR per (comoving) unit volume

$$\dot{\rho}_\star(z) = \bar{\rho}_b^0 f_\star \frac{d}{dt} f_{coll}(z), \tag{19}$$

where $\bar{\rho}_b^0$ is the cosmic mean baryon density today and $f_\star$ is the fraction of baryons converted into stars. This formalism is appropriate for $z \gtrsim 10$, as at later times star formation as a result of mergers becomes important.

With these assumptions, we may rewrite the ionization rate per hydrogen atom as

$$\Lambda_i = \zeta(z) \frac{df_{coll}}{dt}, \tag{20}$$

where $f_{coll}(z)$ is the fraction of gas inside collapsed objects at z and the ionization efficiency parameter ζ is given by

$$\zeta = A_{He} f_\star f_{esc} N_{ion}. \tag{21}$$

This model for x_i is motivated by a picture of H II regions expanding into neutral hydrogen [46]. In calculating f_{coll},

we use the Sheth–Tormen [47] mass function dn/dm and determine a minimum mass m_{min} for collapse by requiring the virial temperature $T_{vir} \geqslant 10^4$ K, appropriate for cooling by atomic hydrogen. Decreasing this minimum galaxy mass, say to the virial temperature corresponding to molecular hydrogen cooling (~ 300 K), will allow star formation to occur at earlier times, shifting the features that we describe in redshift.

The sources of ionizing photons in the early Universe are believed to have been primarily galaxies. However, the properties of these galaxies are currently only poorly constrained. Recent observations with the Hubble Space Telescope provide some of the best constraints on early galaxy formation. Faint galaxies are identified as being at high redshift using a 'Lyα dropout technique' where a naturally occurring break in the galaxy spectrum at the Lyα wavelength 1216 Å is seen in different colour filters as a galaxy is redshifted. So far, galaxies at redshifts up to $z \sim 10$ have been found providing information on the sources of reionization. There are unfortunately considerable limitations on the existing surveys owing to their small sky coverage, which makes it unclear whether those galaxies seen are properly representative, and the limited frequency coverage. Even more problematic for our purposes is that the optical frequencies at which the galaxies are seen do not correspond to the UV photons that ionize the IGM. Our limited understanding of the mass distribution of the emitting stars introduces an uncertainty in the number of ionizing photons per baryon N_{ion} is emitted by galaxies. There is also considerable uncertainty in the fraction of ionizing photons f_{esc} that escape the host galaxy to ionize the IGM.

The recombination rate is primarily important at late times once a significant fraction of the volume has already been ionized. At this stage, dense clumps within an ionized bubble can act as sinks of ionizing photons slowing or even stalling further expansion of the bubble. The degree to which gas resides in these dense clumps is an important uncertainty in modelling reionization. Important hydrodynamic effects, such as the evaporation of gas from a halo as a result of photoionization heating [48], can significantly modify the clumping factor.

A simple model for the clumping factor [43] assumes that the Universe will be fully ionized up to some critical overdensity Δ_c. If the probability distribution for the gas density $P_V(\Delta)$ is specified, we may then write the clumping factor as

$$C \equiv \int^{\Delta_c} d\Delta\, \Delta^2 P_V(\Delta). \quad (22)$$

The quantity $P_V(\Delta)$ can be modelled analytically starting from a consideration of behaviour of low density voids and accounting for Gaussian initial conditions. The analytic form resembles that of a Gaussian with a power law tail [43] and can be measured from simulations [49]. To accurately capture the clumping one should self-consistently perform a full hydrodynamical simulation of reionization, since thermal feedback can modify the gas density distribution [48].

To set the critical density Δ_c, we account for the patchy nature of reionization, which proceeds via the expansion and overlap of ionized bubbles [50]. The size of bubbles will become limited if the mean free path of ionized photons

becomes shorter than the size of the bubble, for example if the bubble contains many small self-shielded absorbers. The mean free path of ionizing photons can be related to the underlying density field as

$$\lambda_i = \lambda_0 [1 - F_V(\Delta_i)]^{-2/3}. \quad (23)$$

Here λ_0 is an unknown normalization constant that was found by [43] in the context of simulations at $z = 2$–4 to be well fit by $\lambda_0 H(z) = 60$ km s^{-1}. This scaling relationship is likely to be very approximate, but we make use of it for convenience. With this we can fix Δ_i within an ionized bubble by setting the relevant $R_b = \lambda_i(\Delta_c)$. We then average the clumping factor over the distribution of bubble sizes (discussed in more details later) to get the mean clumping factor.

3.4. Heating and ionization

To determine the heating rate, we must integrate equation (15) and therefore we must specify which heating mechanisms are important. At high redshifts, the dominant mechanism is Compton heating of the gas arising from the scattering of CMB photons from the small residual free electron fraction. Since these free electrons scatter readily from the surrounding baryons this transfers energy from the CMB to the gas. Compton heating serves to couple T_K to T_γ at redshifts $z \gtrsim 150$, but becomes ineffective below that redshift. In our context, it serves to set the initial conditions before star formation begins. The heating rate per particle for Compton heating is given by [51]

$$\frac{2}{3}\frac{\epsilon_{Compton}}{k_B n} = \frac{x_e}{1 + f_{He} + x_e}\frac{T_\gamma - T_K}{t_\gamma}\frac{u_\gamma}{\bar{u}_\gamma}(1+z)^4, \quad (24)$$

where f_{He} is the helium fraction (by number), u_γ is the energy density of the CMB, $\sigma_T = 6.65 \times 10^{-25}$ cm^2 is the Thomson cross-section, and we define

$$t_\gamma^{-1} = \frac{8\bar{u}_\gamma \sigma_T}{3m_e c} = 8.55 \times 10^{-13} \text{ yr}^{-1}. \quad (25)$$

At lower redshifts, the growth of non-linear structures leads to other possible sources of heat. Shocks associated with large scale structure occur as gas separates from the Hubble flow and undergoes turnaround before collapsing onto a central overdensity. After turnaround different fluid elements may cross and shock due to the differential accelerations. Such turnaround shocks could provide considerable heating of the gas at late times [40].

Another source of heating is the scattering of Lyα photons off hydrogen atoms, which leads to a slight recoil of the nucleus that saps energy from the photon. It was initially believed that this would provide a strong source of heating sufficient to prevent the possibility of seeing the 21 cm signal in absorption. Early calculations showed that by the time the scattering rate required for Lyα photons to couple the spin and gas temperatures was reached, the gas would have been heated well above the CMB temperature [52]. These early estimates, however, did not account for the way the distribution of Lyα photon energies was changed by scattering.

This spectral distortion is a part of the photons coming into equilibrium with the gas and serves to greatly reduce the heating rate [29, 31, 32, 52]. While Lyα heating can be important it typically requires very large Lyα fluxes and so is most relevant at late times and may be insufficient to heat the gas to the CMB temperature alone.

The most important source of energy injection into the IGM is likely via x-ray heating of the gas [29, 53–55]. While shock heating dominates the thermal balance in the present-day Universe, during the epoch we are considering they heat the gas only slightly before x-ray heating dominates. For sensible source populations, Lyα heating is mostly negligible compared with x-ray heating [32, 56].

Since x-ray photons have a long mean free path, they are able to heat the gas far from the source, and can be produced in large quantities once compact objects are formed. The comoving mean free path of an x-ray with energy E is [4]

$$\lambda_X \approx 4.9 \bar{x}_{H_I}^{-1/3} \left(\frac{1+z}{15} \right)^{-2} \left(\frac{E}{300\,\text{eV}} \right)^3 \text{Mpc}. \quad (26)$$

Thus, the Universe will be optically thick over a Hubble length to all photons with energy below $E \sim 2[(1+z)/15]^{1/2} \bar{x}_{H_I}^{1/3}$ keV. The E^{-3} dependence of the cross-section means that heating is dominated by soft x-rays, which fluctuate on small scales. In addition, though, there will be a uniform component to the heating from harder x-rays.

X-rays heat the gas primarily through photoionization of H I and He I: this generates energetic photo-electrons, which dissipate their energy into heating, secondary ionizations, and atomic excitation. With this in mind, we calculate the total rate of energy deposition per unit volume as

$$\epsilon_X = 4\pi \sum_i n_i \int d\nu\, \sigma_{\nu,i} J_\nu (h\nu - h\nu_{\text{th},i}), \quad (27)$$

where we sum over the species $i = $ H I, He I and He II, n_i is the number density of species i, $h\nu_{\text{th}} = E_{\text{th}}$ is the threshold energy for ionization, $\sigma_{\nu,i}$ is the cross-section for photoionization and J_ν is the number flux of photons of frequency ν.

We may divide this energy into heating, ionization and excitation by inserting the factor $f_i(\nu, x_e)$, defined as the fraction of energy converted into form i at a specific frequency. This allows us to calculate the contribution of x-rays to both the heating and the partial ionization of the bulk IGM. The relevant division of the x-ray energy depends on both the x-ray energy E and the free electron fraction x_e and can be calculated by Monte Carlo methods. This partitioning of x-ray energy in this way was first calculated by [57] and subsequently updated [58, 59]. In the following calculations, we make use of fitting formula for the $f_i(\nu)$ calculated by [57], which are approximately independent of ν for $h\nu \gtrsim 100\,\text{eV}$, so that the ionization rate is related to the heating rate by a factor $f_{\text{ion}}/(f_{\text{heat}} E_{\text{th}})$.

The x-ray number flux is found from

$$J_X(z) = \int_{\nu_{\text{th}}}^{\infty} d\nu\, J_X(\nu, z), \quad (28)$$

$$= \int_{\nu_{\text{th}}}^{\infty} d\nu \int_z^{z_\star} dz' \frac{(1+z)^2}{4\pi} \frac{c}{H(z')} \hat{\epsilon}_X(\nu', z') e^{-\tau},$$

where $\hat{\epsilon}_X(\nu, z)$ is the comoving photon emissivity for x-ray sources and ν' is the emission frequency at z' corresponding to an x-ray frequency ν at z

$$\nu' = \nu \frac{(1+z')}{(1+z)}. \quad (29)$$

The optical depth is given by

$$\tau(\nu, z, z') = \int_z^{z'} \frac{dl}{dz''} dz'' \left[n_{H_I} \sigma_{H_I}(\nu'') + n_{He_I} \sigma_{He_I}(\nu'') + n_{He_{II}} \sigma_{He_{II}}(\nu'') \right], \quad (30)$$

where we calculate the cross-sections using the fits of [60]. Care must be taken here, as the cross-sections have a strong frequency dependence and the x-ray photon frequency can redshift considerably between emission and absorption. In practice, the abundance of He II may be neglected [53].

X-rays may be produced by a variety of different sources with three main candidates at high redshifts being identified as starburst galaxies, supernova remnants (SNRs), and miniquasars [61–63]. Galaxies with high rates of star formation produce copious numbers of x-ray binaries, whose total x-ray luminosity can be considerable. Two populations of x-ray binaries may be identified in the local Universe distinguished by the mass of the donor star which feeds its black hole companion—low-mass x-ray binaries (LMXBs) and high-mass x-ray binaries (HMXBs). The short life time of HMXBs ($t_{\text{HMXB}} \sim 10^7$ yr) leads the x-ray luminosity L_X^{HMXB} to track the SFR. At the same time, the longer lived LMXBs ($t_{\text{LMXB}} \sim 10^{10}$ yr) tracks the total mass of stars formed. Since we will focus on the early Universe and on the first billion years of evolution, when few LMXBs are expected to have formed, the dominant contribution to L_X in galaxies is likely to be from HMXBs [64]. This has conventionally been defined in terms of a parameter f_X such that the emissivity per unit (comoving) volume per unit frequency

$$\hat{\epsilon}_X(z, \nu) = \hat{\epsilon}_X(\nu) \left(\frac{\dot{\rho}_\star(z)}{M_\odot\, \text{yr}^{-1}\, \text{Mpc}^{-3}} \right), \quad (31)$$

where $\dot{\rho}_\star$ is the SFR density and the spectral distribution function is a power law with index α_S

$$\hat{\epsilon}_X(\nu) = \frac{L_0}{h\nu_0} \left(\frac{\nu}{\nu_0} \right)^{-\alpha_S - 1}, \quad (32)$$

and the pivot energy $h\nu_0 = 1$ keV. We assume emission within the band 0.2–30 keV and set $L_0 = 3.4 \times 10^{40} f_X$ erg s^{-1} Mpc^{-3}, where f_X is a highly uncertain constant factor [63]. For a spectral index $\alpha_S = 1.5$, roughly corresponding to that for starburst galaxies, $f_X = 1$ corresponds to the emission of approximately 560 eV for every baryon converted into stars.

This normalization was chosen so that, with $f_X = 1$, the total x-ray luminosity per unit SFR is consistent with that observed in starburst galaxies at the present epoch [64, 65]. Since then improved observations have revised this figure owing to better separation of the contribution from LMXBs and HMXBs. While the data are still as of yet fairly patchy and show considerable scatter $f_X \approx 0.2$ seems a better fit to the

most recent data in the local Universe [66, 67]. Extrapolating observations from the present day to high redshift is fraught with uncertainty, and we note that this normalization is very uncertain and probably evolves with redshift [68]. In particular, the metallicity evolution of galaxies with redshift is likely to impact the ratio of black holes to neutron stars that form the compact object in the HMXBs and with it the efficiency of x-ray production. Additionally, the fraction of stars in binaries may evolve with redshift and is only poorly constrained at high redshifts [69].

Other sources of x-rays are inverse Compton scattering of CMB photons from the energetic electrons in supernova remnants. Estimates of the luminosity of such sources is again highly uncertain, but of a similar order of magnitude as from HMXBs [61]. Like HMXBs, the x-ray luminosity from supernovae remnants is expected to track the SFR. Finally, miniquasars—accretion onto black holes with intermediate masses in the range $10^{1-5} M_\odot$—can produce significant levels of x-rays. Since the early formation of black holes depends sensitively on the source of seed black holes and their subsequent merger history there is again considerable uncertainty. For simplicity, we will assume that miniquasars similarly track the SFR. In reality, of course, their evolution could be considerably more complex [70, 71].

The total x-ray luminosity at high redshift is constrained by observations of the present-day unresolved soft x-ray background (SXRB). An early population of x-ray sources would produce hard x-rays that would redshift to lower energies contributing to this background. Since there will be faint x-ray sources at lower redshift that also contribute to this background, the SXRB can be used to place a conservative upper limit on the amount of x-ray production at early times. This rules out complete reionization by x-rays but allows considerable latitude for heating [72].

Since heating requires considerably less energy than ionization, f_X is still relatively unconstrained with values as high as $f_X \lesssim 10^3$ possible without violating constraints from the CMB polarization anisotropies on the optical depth for electron scattering. Constraining this parameter will mark a step forward in our understanding of the thermal history of the IGM and the population of x-ray sources at high redshifts.

3.5. Coupling

Finally, we need to specify the evolution of the Lyα flux. This is produced by stellar emission ($J_{\alpha,\star}$) and by x-ray excitation of H I ($J_{\alpha,X}$). Photons emitted by stars, between Lyα and the Lyman limit, will redshift until they enter a Lyman series resonance. Subsequently, they may generate Lyα photons via atomic cascades [25, 30]. The Lyα flux from stars $J_{\alpha,\star}$ arises from a sum over the Lyn levels, with the maximum n that contributes $n_{max} \approx 23$ determined by the size of the H II region of a typical (isolated) galaxy (see [34] for details). The average Lyα background is then

$$J_{\alpha,\star}(z) = \sum_{n=2}^{n_{max}} J_\alpha^{(n)}(z), \tag{33}$$

$$= \sum_{n=2}^{n_{max}} f_{recycle}(n) \int_z^{z_{max}(n)} dz' \frac{(1+z)^2}{4\pi} \frac{c}{H(z')} \hat{\epsilon}_\star(\nu'_n, z'),$$

where $z_{max}(n)$ is the maximum redshift from which emitted photons will redshift into the level n Lyman resonance, ν'_n is the emission frequency at z' corresponding to absorption by the level n at z, $f_{recycle}(n)$ is the probability of producing a Lyα photon by cascade from level n and $\hat{\epsilon}_\star(\nu, z)$ is the comoving photon number emissivity for stellar sources. We connect $\hat{\epsilon}_\star(\nu, z)$ to the SFR in the same way as for x-rays in equation (32), and define $\hat{\epsilon}_\star(\nu)$ to be the spectral distribution function of the stellar sources.

Stellar sources typically have a spectrum that falls rapidly above the Lyβ transition. We consider models with present-day (Population I and II) and very massive (Population III) stars. In each case, we take $\hat{\epsilon}_\star(\nu)$ to be a broken power law with one index describing emission between Lyα and Lyβ, and a second describing emission between Lyβ and the Lyman limit (see [25] for details). The details of the signal depend primarily on the total Lyα emissivity and not on the shape of the spectrum (but see [54, 73] for details of how precision measurements of the 21 cm fluctuations might say something about the source spectrum).

For convenience, we define a parameter controlling the normalization of the Lyα emissivity f_α by setting the total number of Lyα photons emitted per baryon converted into stars as $N_\alpha = f_\alpha N_{\alpha,ref}$ where we take the reference values appropriate for normal (so-called, Population I and II) stars $N_{\alpha,ref} = 6590$ [34, 74]. For comparison, in this notation, the very massive (III) stars have [44], $N_\alpha = 3030$ ($f_\alpha = 0.46$), when the contribution from higher Lyman series photons is included. We expect the value of f_α to be close to unity, since stellar properties are relatively well understood.

Photoionization of H I or He I by x-rays may also lead to the production of Lyα photons. In this case, some of the primary photo-electron's energy ends up in excitations of H I [57], which on relaxation may generate Lyα photons [35, 52, 73]. The rate at which Lyα photons are produced $\epsilon_{X,\alpha}$ can be calculated in the same way as the contribution to heating by x-rays, but with the appropriate change in the fraction of the total x-ray energy that goes into excitations rather than heating.

Calculating the fraction of energy that goes into producing Lyα photons is a problem that has been considered by a number of authors. Early work [57, 75] focused on the amount of energy that went into atomic excitations as a whole, but we require only the fraction that leads to Lyα production. Although excitations to the 2P level will always generate Lyα photons, only some fraction of excitations to other levels will lead to Lyα generating cascades. The rest will end with two photon decay from the 2S level. This has been addressed by Monte Carlo simulation of the x-ray scattering process using up to date cross-sections in [58, 59]. These simulations find that around $p_\alpha \approx 0.7$ of the total energy that goes into excitation ends up as Lyα photons, consistent with simple estimates based on the atomic cross-sections [54].

This Lyα flux $J_{\alpha,X}$ produced by x-ray excitation may be found by balancing the rate at which Lyα photons are produced via cascades with the rate at which photons redshift out of the Lyα resonance [35], giving

$$J_{\alpha,X} = \frac{c}{4\pi} \frac{\epsilon_{X,\alpha}}{h\nu_\alpha} \frac{1}{H\nu_\alpha}. \tag{34}$$

The relative importance of Lyα photons from x-rays or directly produced by stars is highly dependent upon the nature of the sources that existed at high redshifts. Furthermore, it can vary significantly from place to place. In general, x-rays with their long mean free path seem likely to dominate the Lyα flux far from sources while the contribution from stellar sources dominates closer in [35].

3.6. Astrophysical sources and histories

In the above sections, we have outlined the mathematical formalism for describing the 21 cm signal and have omitted a detailed discussion of the sources. This was deliberate; although we have a reasonable understanding of the physical processes involved, our knowledge of the properties of early sources of radiation is highly uncertain.

Many models of galaxy formation assume that the first stars to form from the collapse of primordial gas are very massive ($\sim$10–100 $M_\odot$) Population III stars [44]. This is predicated on the inference that the absence of coolants more efficient than molecular hydrogen leads to monolithic collapse into a single massive star rather than fragmentation into many lower mass stars. This assumption has recently begun to be challenged by new numerical simulations that use 'sink particles' to better follow the collapsing gas for many dynamical times. Such simulations show that fragmentation into many $\sim$0.1–1 $M_\odot$ stars may be the preferred channel of star formation [76]. This would naturally explain tentative observations of low-mass metal-free stars [77] and could lead to a much higher fraction of early x-ray binaries [69]. Once earlier generations of star formation have enriched the IGM with metals low-mass Population II stars will begin to form due to more efficient gas cooling [78]. Different predictions for the mode of star formation will lead to quite different IGM histories.

We have three radiation backgrounds to account for— ionizing UV, x-ray and Lyman series photons (identified as those photons with energy 10.2 eV $\leqslant E <$ 13.6 eV). For each of these radiation fields we must specify a single parameter: the ionization efficiency ζ, the x-ray emissivity f_X and the Lyα emissivity f_α. These parameters enter our model as a factor multiplying the SFR and are therefore individually degenerate with the star-formation efficiency $f_\star$. This split provides a natural separation between the physics of the sources and the SFR and, in practice, one might imagine using observations of the SFR by other means as a way of breaking the degeneracy between them. In addition to these parameters, we must specify the minimum mass halo in which galaxies form M_{min} and make use of the Sheth–Tormen mass function of dark matter halos.

We now show results for the 21 cm global signal that explore this parameter space to give a sense of how the signal depends on these astrophysical parameters. Model A uses $(N_{ion,IGM}, f_\alpha, f_X, f_\star) = (200, 1, 1, 0.1)$ giving $z_{reion} = 6.47$ and $\tau = 0.063$. Model B uses $(N_{ion,IGM}, f_\alpha, f_X, f_\star) = (600, 1, 0.1, 0.2)$ giving $z_{reion} = 9.76$ and $\tau = 0.094$. Model C uses $(N_{ion,IGM}, f_\alpha, f_X, f_\star) = (3000, 0.46, 1, 0.15)$ giving $z_{reion} = 11.76$ and $\tau = 0.115$.
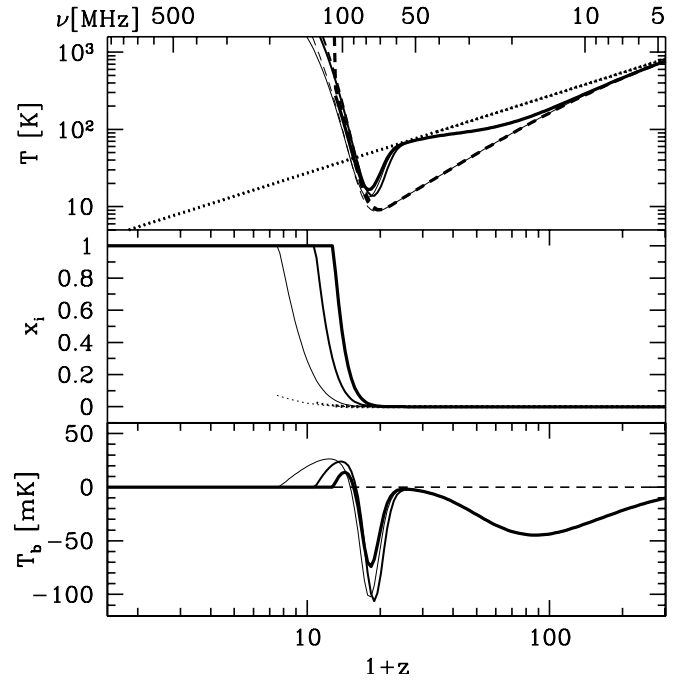

Figure 4. Top panel: evolution of the CMB temperature T_{CMB} (dotted curve), the gas kinetic temperature T_K (dashed curve) and the spin temperature T_S (solid curve). Middle panel: evolution of the gas fraction in ionized regions x_i (solid curve) and the ionized fraction outside these regions (due to diffuse x-rays) x_e (dotted curve). Bottom panel: evolution of mean 21 cm brightness temperature T_b. In each panel we plot curves for model A (thin curves), model B (medium curves) and model C (thick curves). Reproduced with permission from [79]. Copyright 2008 American Physical Society.

Figure 4 shows several examples of the global 21 cm signal and the associated evolution in the neutral fraction and gas temperatures. While the details of the models may vary considerably, all show similar basic properties. At high redshift, $10 \gg z \lesssim 200$, the gas temperature cools adiabatically faster than the CMB (since the residual fraction of free electrons is insufficient to couple the two temperatures). At the same time, collisional coupling is effective at coupling spin and gas temperatures leading to the absorption trough seen at the right of the lower panel. The details of this trough are fixed by cosmology and therefore may be predicted relatively robustly. The minimum of this trough corresponds to the point at which collisional coupling starts to become relatively ineffective.

Once star formation begins, the spin and gas temperatures again become tightly coupled leading to a second, potentially deeper, absorption trough. The minimum of this trough corresponds to the point when x-ray heating switches on heating the gas above the CMB temperature leading to an emission signal. The signal then reaches the curve for a saturated signal ($T_S \gg T_{CMB}$) briefly before the ionization of neutral hydrogen diminishes it.

The ordering of these events is determined primarily by the energetics of the processes involved and by the basic properties of the reasonable source spectra. For example, ionization requires at least one ionizing photon with energy $E \geqslant 13.6$ eV per baryon while depositing only $\sim$10% of that energy per

baryon would heat the gas to $T_K \gtrsim 10^4$ K. However, the details of the shape of the curve after star formation begins are highly uncertain—in our model we have neglected any possible redshift evolution in the various photon emissivity parameters—but the basic structure of one emission feature and two absorption troughs are likely to be robust. By determining the positions of the various turning points in the signal one could hope to constrain the underlying astrophysics and learn about the first stars and galaxies.

3.7. Exotic heating

One of the key points to take away from this discussion is that the 21 cm global signal plays the role of a very sensitive calorimeter of the IGM gas temperature. Provided that the coupling is saturated and that the IGM is close to neutral there is a direct connection between the 21 cm brightness temperature and the IGM temperature. Many models of physics beyond the Standard Model make concrete predictions for exotic heating of the IGM. For example, dark matter annihilation in the early Universe can act as a source of x-rays leading to heating. In this subsection, we consider some of the possibilities that have been advanced for exotic heating mechanisms and discuss the possibility of constraining them.

Perhaps the most commonly considered source of heating in the Dark Ages is that of dark matter annihilation [80–84]. Dark matter is widely assumed to explain the observed galaxy rotation curves as well as the detailed features of the CMB acoustic peaks. Simple models of dark matter production and freeze out in the early Universe lead to a prediction for the annihilation cross-section required to leave a freeze out abundance corresponding to the measured value of the dark matter density parameter, Ω_{DM}.

Depending on the dark matter mass, which for fixed Ω_{DM} determines the dark matter number density n_{DM}, annihilation of dark matter in the later Universe may be an important source of heating. It is important to note that there are two regimes in which dark matter annihilation may be important. Since the annihilation rate scales as n_{DM}^2 the rate may be large at early times where n_{DM} has yet to be diluted by the cosmic expansion. Alternatively, dark matter annihilation can become important once significant numbers of collapsed dark matter halos form, leading to a local enhancement in the dark matter density [85].

Alternatives to dark matter annihilation include dark matter decaying into Standard Model particles or photons leading to the deposition of energy in the IGM [81, 82]. The different density dependence of dark matter decay and annihilation might lead to distinguishable redshift evolution of the heating. Further, models where dark matter contains an internal excited state that relaxes to the ground state releasing energy have been proposed [86].

Many other scenarios for exotic heating of the IGM have been put forward, emphasizing the interest in a new technique for distinguishing models of new physics. Primordial black holes produced in the early Universe may evaporate after recombination if their masses lie in the range 10^{14}–10^{17} g [87] and the Hawking radiation given off could provide a strong heating source [88]. Moving cosmic strings produce wakes
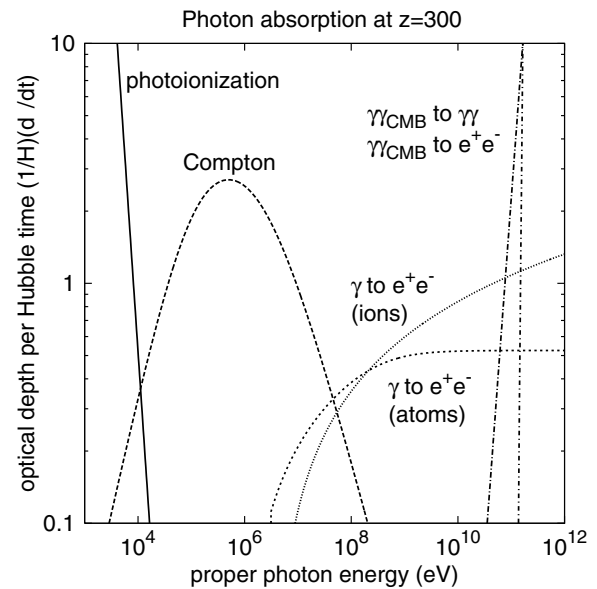

Figure 5. Optical depths per time for various photon-IGM processes, in units of the Hubble time, at $z = 300$, assuming a neutral IGM. These include processes which deposit energy directly into the IGM (pair production and photoionization), processes which redistribute photons ($2\gamma \rightarrow 2\gamma$) and ones that do both (Compton). At very low energies, photoionization is the dominant process; at very high energies, $e^\pm$ pair production dominates. Reproduced with permission from [87].

that stir the IGM imparting heat into the gas. These were originally put forward as a source of density fluctuations for seeding the growth of structure. While ruled out for this purpose, cosmic strings might be further constrained via their heating effect on the IGM [89].

Incorporating the heating effect arising from exotic sources requires a knowledge of the energy spectrum of photons produced by the source, which must then be carefully processed to determine how much of the radiative energy is ultimately deposited into the IGM. The cross-section for photon absorption has a number of minima, which reflect windows at which the IGM is transparent to photons so that rather than being absorbed they may propagate to the present as a diffuse background. For most scenarios, this consideration greatly constrains the amount of energy deposited as heat in the IGM. Figure 5 illustrates the various processes that dominate the loss of energy from an energetic photon at $z = 300$.

3.8. Detectability of the global signal with small numbers of dipoles

The global 21 cm signal could potentially be measured by absolute temperature measurements as a function of frequency, averaged over the sky. Since the global signal is constant over different large patches of the sky, experimental efforts to measure it do not need high angular resolution and can be carried out with just a single dipole. The attempted measurement is complicated however by the need to remove galactic foregrounds, which are much larger than the desired signal. Foreground removal is predicated on the assumption of spectral smoothness of the foregrounds in contrast to

the frequency structure of the signal. This should allow removal of the foregrounds by, for example, fitting a low order polynomial to the foregrounds leaving the 21 cm signal in the residuals. This methodology requires very precise calibration of the instrumental frequency response, which could otherwise become confused with the foregrounds.

The first experimental efforts to detect the 21 cm global signal have been carried out by the COsmological Reionization Experiment (CORE) [90] and the Experiment to Detect the Reionization Step (EDGES) [91]. These have been analysed using a tanh model of reionization that depends upon the redshift of reionization z_r and its duration Δz. EDGES is presently able to rule out the most rapid models of reionization that occur over a redshift interval as short as $\Delta z < 0.06$ [92]. These first experimental efforts should be seen as the first steps along a road that may lead to considerably better constraints. Other experiments using different experimental approaches are underway. Some of these use individual dipoles, such as the Shaped Antenna measurement of the RAdio Spectrum (SARAS) [93] and the Broadband Instrument for the Global HydrOgen ReionizatioN Signal (BIGHORNS) [94], while others are exploring ways of using many dipoles as with the Large-aperture Experiment to Detect the Dark Ages (LEDA) [95].

Theoretical estimates for the ability of a single dipole experiment to constrain models of the 21 cm signal can be made via the Fisher matrix formalism [96]. For a single dipole experiment, the Fisher matrix may be written as [97]

$$F_{ij} = \sum_{n=1}^{N_{\text{channel}}} (2 + B t_{\text{int}}) \frac{\mathrm{d} \log T_{\text{sky}}(\nu_n)}{\mathrm{d} p_i} \frac{\mathrm{d} \log T_{\text{sky}}(\nu_n)}{\mathrm{d} p_j}, \quad (35)$$

where t_{int} is the total integration time (before systematics limit the performance), and we divide the total bandwidth B into N_{channel} frequency bins $\{\nu_n\}$ running between $[\nu_{\text{min}}, \nu_{\text{max}}]$. For the 21 cm global signature, our observable is the antennae temperature $T_{\text{sky}}(\nu) = T_{\text{fg}}(\nu) + T_{\text{b}}(\nu)$, where we assume the dipole sees the full sky so that spatial variations can be ignored. Best case errors on the parameters $\{p_i\}$, which include both foreground and signal model parameters, are then given by $\sigma_i \leqslant \sqrt{F_{ii}^{-1}}$.

Such estimates show that global 21 cm experiments should be able to constrain realistic reionization models with $\Delta z \lesssim 2$ [97, 98]. The results of integrating for 500 h between 100 and 200 MHz with a single dipole are shown in figure 6, where the reionization history has been parametrized with a tanh function.

In addition to constraining reionization, global 21 cm experiments might be used to probe the thermal evolution of the IGM at redshifts $z > 12$. Such high redshifts are very difficult to probe via the 21 cm fluctuations (discussed later) since they require very large collecting areas. Global experiments bypass this requirement, but still suffer from the larger foregrounds at lower frequencies. By going to high redshifts such experiments could place constraints on x-ray heating and Lyα coupling giving information about when the first black holes and galaxies form, respectively. The absorption feature resulting from this physics can potentially be larger ($\sim$100 mK) making
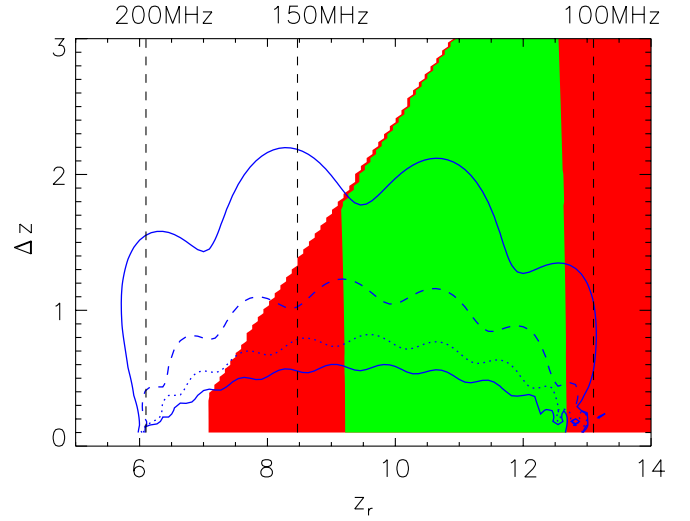

Figure 6. 95% constraint region on a tanh reionization model $T_{\text{b}}(z) = T_{21} \tanh[(z - z_{\text{r}})/\Delta z]$ of the end of reionization for an EDGES-like experiment assuming $N_{\text{poly}} = 3$ (solid curve), 6 (dashed curve), 9 (dotted curve) and 12 (dotted–dashed curve). Also plotted are the 68 and 95% contours for the WMAP5 electron scattering optical depth constraint combined with a prior that $x_i(z = 6.5) > 0.95$ (green and red coloured regions). Reproduced with permission from [97]. Copyright 2010 American Physical Society.
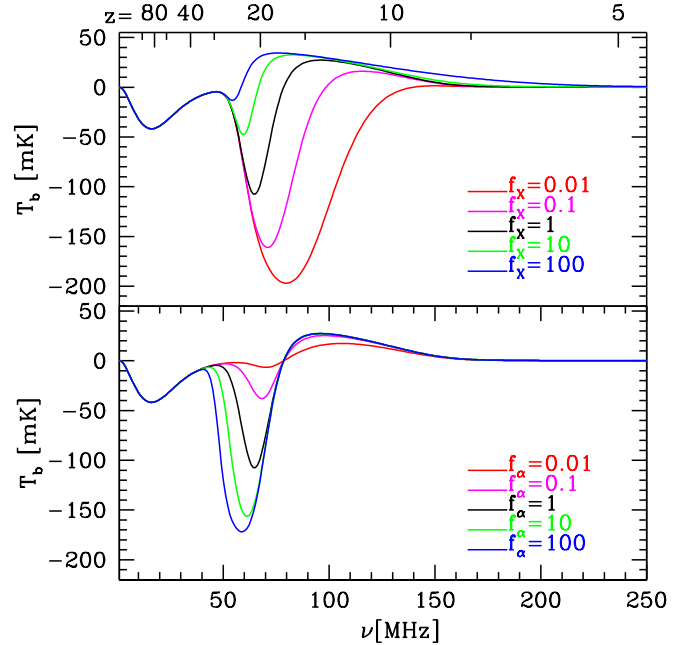

Figure 7. Dependence of 21 cm signal on the x-ray (top panel) and Lyα (bottom panel) emissivity. In each case, we consider examples with the emissivity reduced or increased by a factor of up to 100. Note that in our model f_X and f_α are really the product of the emissivity and the star-formation efficiency. Reproduced with permission from [97]. Copyright 2010 American Physical Society.

it a good target for observations. Figure 7 shows how the global 21 cm signal can vary with different values of f_X and f_α. Measuring the global signal would offer a useful avenue for distinguishing these models although there is some degeneracy between the two parameters.

EDGES-type experiments at frequencies $\nu < 100$ MHz are underway from the ground and a lunar orbiting dipole

experiment—the Dark Ages Radio Experiment (DARE) [99]—has also been proposed. Lunar orbit offers a number of advantages including being shielded from terrestrial radio frequency interference (RFI) while on the far side of the Moon and the ability to use the Moon to chop the beam aiding instrumental calibration [100]. DARE would be targeted at the 40–120 MHz range ideal for measuring the deep absorption feature and determining f_X and f_α.

4. 21 cm tomography

The 21 cm global signal can be viewed as a zeroth order approximation to the full 21 cm signal, as it is averaged over large angular scales. The full 3D signal will be highly inhomogeneous as a result of the spatial variation in the different radiation fields and properties of the IGM. In this section, we consider the physics underlying 21 cm brightness fluctuations in 3D and detail the existing techniques for calculating the statistical properties of the signal.

Ultimately, one might wish to make use of fully numerical simulations of the relevant physics and so produce detailed maps of the 21 cm signal along the light cone. At present, the large dynamic range required and the computational cost make this a dream for the future. For the moment, it is important to make use of a variety of analytic, semi-numerical and numerical techniques to calculate the expected 21 cm signal. These different methods complement one another in speed, accuracy and detail, as we describe below.

Fluctuations in the 21 cm signal may be expanded to linear order [4]

$$\delta_{T_b} = \beta_b \delta_b + \beta_x \delta_x + \beta_\alpha \delta_\alpha + \beta_T \delta_T - \delta_{\partial v}, \tag{36}$$

where each δ_i describes the fractional variation in the quantity i and we include fluctuations in the baryon density (b), neutral fraction (x), Lyα coupling coefficient (α), gas temperature (T) and line-of-sight peculiar velocity gradient (∂v). The expansion coefficients are given by

$$\beta_b = 1 + \frac{x_c}{x_{tot}(1 + x_{tot})}, \tag{37}$$

$$\beta_x = 1 + \frac{x_c^{HH} - x_c^{eH}}{x_{tot}(1 + x_{tot})},$$

$$\beta_\alpha = \frac{x_\alpha}{x_{tot}(1 + x_{tot})},$$

$$\beta_T = \frac{T_\gamma}{T_K - T_\gamma} + \frac{1}{x_{tot}(1 + x_{tot})}$$

$$\times \left(x_c^{eH} \frac{d \log \kappa_{10}^{eH}}{d \log T_K} + x_c^{HH} \frac{d \log \kappa_{10}^{HH}}{d \log T_K} \right).$$

In this expression, we have treated all the terms as being of a similar size, but it is important to realize that fluctuations in x_H can be of order unity. This means that terms in higher order of δ_x, which one might naively think to be small, can still contribute at a significant level to the power spectrum.

In general, homogeneity and isotropy of the Universe suggest that the power spectrum of brightness temperature fluctuations should be spherically symmetric in Fourier space, i.e. it should only depend on $k = |\boldsymbol{k}|$ for a wavevector $\boldsymbol{k}$ of

a given Fourier mode. However, redshift space distortions induced by peculiar velocities break this symmetry since the direction to the observer becomes important and so only a cylindrical symmetry is preserved. This symmetry may be useful in separating signal from foregrounds, which typically do not share this symmetry (see e.g. [101]). In Fourier space and at linear order, we may write the peculiar velocity term $\delta_{\partial v} = -\mu^2 \delta$ [102], where μ is cosine of the angle between the line of sight and the wavevector $\boldsymbol{k}$ of the Fourier mode. With this, we may use (36) to form the power spectrum

$$\begin{aligned}
P_{T_b}(k, \mu) = {} & P_{bb} + P_{xx} + P_{\alpha\alpha} + P_{TT} + 2P_{bx} \\
& + 2P_{b\alpha} + 2P_{bT} + 2P_{x\alpha} + 2P_{xT} + 2P_{\alpha T} \\
& + P_{x\delta x\delta} + \text{other quartic terms} \\
& + 2\mu^2 \left(P_{b\delta} + P_{x\delta} + P_{\alpha\delta} + P_{T\delta} \right) \\
& + \mu^4 P_{\delta\delta} \\
& + 2P_{x\delta\delta_{\partial v}} + P_{x\delta_{\partial v}\delta_{\partial v}x} \\
& + \text{other quartic terms with } \delta_{\partial v}. \tag{38}
\end{aligned}$$

Here we note that all quartic terms must be quadratic in x_H and separate them depending upon whether they contain powers of $\delta_{\partial v}$ or not. Those that contain powers of $\delta_{\partial v}$ will not be isotropic and will lead to the angular dependence of P_{T_b} (see [103] for further discussion).

We may rewrite equation (38) in a more compact form

$$P_{T_b}(k, \mu) = P_{\mu^0}(k) + \mu^2 P_{\mu^2}(k) + \mu^4 P_{\mu^4}(k) + P_{f(k,\mu)}(k, \mu), \tag{39}$$

where we have grouped those quartic terms with anomalous μ dependence into the term $P_{f(k,\mu)}(k, \mu)$. In principle, high precision measurements of the 3D power spectrum will allow the separation of $P_{T_b}(k, \mu)$ into these four terms by their angular dependence on powers of μ^2 [104]. The contribution of the $P_{f(k,\mu)}(k, \mu)$ term, with its more complicated angular dependence, threatens this decomposition [103]. Since this term is only important during the final stages of reionization, we will not discuss it in detail in this paper noting only that the angular decomposition by powers of μ^2 may not be possible when ionization fluctuations are important.

It is unclear whether the first generation of 21 cm experiments will be able to achieve the high signal to noise required for this separation [103]. Instead, they might measure the angle averaged quantity

$$\bar{P}_{T_b}(k) = P_{\mu^0}(k) + P_{\mu^2}(k)/3 + P_{\mu^4}(k)/5 \tag{40}$$

(where we neglect the $P_{f(k,\mu)}(k, \mu)$ term). One typically plots the power per logarithmic interval $\Delta = [k^3 P(k)/2\pi^2]^{1/2}$.

4.1. Redshift space distortions

Peculiar velocity effects can have a significant effect on the 21 cm signal. At linear order, the effects of peculiar velocities are well understood [102, 104, 105] since the $\delta_{\partial v}$ term is simply related to the total density field. However, as the density field evolves and non-linear corrections to the velocity field become important the picture can change in ways that are not yet well understood.

15

Redshift space effects become important because our observations are made in frequency space, while theory makes predictions most directly in coordinate space. The conversion between the two is affected by the local bulk velocity of the gas. We may write the comoving distance to an object as

$$\chi_z = \int_0^z \frac{c\,\mathrm{d}z'}{(1+z')\mathcal{H}(z')}, \tag{41}$$

where we introduce $\mathcal{H}$ as the comoving Hubble parameter. Using $\mathcal{H}$ makes the notion compact and comes from introducing the conformal time coordinate η related to the proper time t via $\mathrm{d}\eta = \mathrm{d}t/a(t)$, where $a(t)$ is the usual scale factor, so that the comoving Hubble parameter $\mathcal{H} = (1/a)\mathrm{d}a(\eta)/\mathrm{d}\eta$.

Including the effects of peculiar velocity, the true coordinate distance to an object with measured redshift z is

$$\chi = \chi_z - \boldsymbol{v}(\boldsymbol{x}) \cdot \hat{\boldsymbol{n}}/\mathcal{H}|_z. \tag{42}$$

Writing our coordinates as $\boldsymbol{x} = \chi\hat{\boldsymbol{n}}$ in real space and $\boldsymbol{s} = \chi_z\hat{\boldsymbol{n}}$ in redshift space gives the mapping between the two as

$$\boldsymbol{s} = \boldsymbol{x} + [\boldsymbol{v}(\boldsymbol{x}) \cdot \hat{\boldsymbol{n}}/\mathcal{H}]\hat{\boldsymbol{n}}. \tag{43}$$

Accounting for this difference in the coordinate systems leads to the redshift space distortions. In linear theory, we have

$$\nabla \cdot \boldsymbol{v}(\boldsymbol{x}) = -\mathcal{H}\delta(\boldsymbol{x}), \tag{44}$$

where we are assuming the Universe to be matter dominated so that the growth factor $f \equiv \mathrm{d}\log D_+/\mathrm{d}\log a = 1$. The Fourier transform of this result introduces the factor μ^2. More generally, large peculiar velocities can lead to the so-called 'finger of God' effects from virialized structures and greatly complicate efforts to separate components via the angular structure of the power spectrum [106–108]. Mistakenly attempting to use the form (40) would lead to significantly biased results [107], and so new estimators calibrated by simulations are needed.

4.2. Ionization fluctuations

Reionization is a complicated process involving the balancing of ionizing photons originating in highly clustered collections of galaxies and recombinations in dense clumps of matter. It is perhaps surprising then that one can produce remarkably robust models of the topology of reionization by simply counting the number of ionizing photons. This basic insight lies at the centre of the analytic calculation of ionization fluctuations [37, 109].

Imagine a spherical region of gas containing a total mass of ionized gas m_{ion}. As an ansatz for determining whether this region of gas will be ionized, we can ask whether the region contains a quantity of galaxies sufficient to ionize it. Connecting to the language we set up earlier when discussing the 21 cm global signal, we ask whether the following condition is satisfied:

$$m_{\mathrm{ion}} \geqslant \zeta m_{\mathrm{gal}}, \tag{45}$$

where ζ is the ionizing efficiency and m_{gal} is the total mass in galaxies. Note this is essentially the condition that the galaxies

produce enough ionizing photons to have ionized all the gas within the region.

This condition equates to asking if the collapse fraction exceeds a critical value $f_{\mathrm{coll}} \geqslant f_{\mathrm{x}} \equiv \zeta^{-1}$. From the definition of the collapse fraction (making use of the Press–Schechter [110] mass fraction for analytic simplicity), these can be translated into a condition on the mass overdensity if a region is to self-ionize

$$\delta_{\mathrm{m}} \geqslant \delta_{\mathrm{x}}(m, z) \equiv \delta_{\mathrm{c}}(z) - \sqrt{2}K(\zeta)\left[\sigma_{\mathrm{min}}^2 - \sigma^2(m)\right]^{1/2}, \tag{46}$$

where $K(\zeta) = \mathrm{erf}^{-1}(1 - \zeta^{-1})$. This condition for self-ionization can be used to calculate the probability distribution of ionized regions or bubble sizes $n_{\mathrm{bub}}(m)$ by reference to the excursion set formalism [111]. Smoothing a Gaussian density field on decreasing mass scales corresponds to a random walk in overdensity. Once the overdensity for a region crosses the ionization threshold, the mass enclosed will be ionized. This is similar to the mass function calculations of Press–Schechter or Sheth–Tormen, except with a mass dependent rather than constant barrier. The distribution of bubbles sizes found from this analytic calculation has been shown to be a good match to numerical reionization simulations at the same neutral fraction [112, 113].

To connect the bubble distribution (a one-point statistic) to the power spectrum (a two-point statistic) requires extra thought. It is possible to generalise the basic excursion set formalism to keep track of two correlated random walks corresponding to two spatially separated locations. This very directly gives the two-point correlation function for the ionization (or density) field [114, 115]. Unfortunately the resulting expressions are somewhat complicated to deal with and simpler more approximate calculations can be more useful. We follow [37], which incorporates some simple ansatzes for the form of P_{xx} and $P_{\mathrm{x}\delta}$ based upon the expected clustering properties of the bubbles.

4.3. Fluctuations in the coupling

Next we consider fluctuations in the Lyα coupling [25, 34]. Provided that we neglect the mild temperature dependence of S_α [32], the fluctuations in the coupling are simply sourced by fluctuations in the flux and we may write $\delta_\alpha = \delta_{J_\alpha}$.

Density perturbations at redshift z' source fluctuations in J_α seen by a gas element at redshift z via three effects. First, the number of galaxies traces, but is biased with respect to, the underlying density field. As a result an overdense region will contain a factor $[1 + b(z')\delta]$ more sources, where $b(z')$ is the (mass-averaged) bias, and will emit more strongly. Next, photon trajectories near an overdense region are modified by gravitational lensing, increasing the effective area by a factor $(1 + 2\delta/3)$. Finally, peculiar velocities associated with gas flowing into overdense regions establish an anisotropic redshift distortion, which modifies the width of the region contributing to a given observed frequency. Given these three effects, we can write $\delta_\alpha = \delta_{J_\alpha} = W_\alpha(k)\delta$, where we compute the window function $W_{\alpha,\star}(k)$ for a gas element at z by adding the coupling

due to Lyα flux from each of the Lyn resonances and integrating over radial shells [34],

$$
\begin{aligned}
W_{\alpha,\star}(k) = {} & \frac{1}{J_{\alpha,\star}} \sum_{n=2}^{n_{\max}} \int_z^{z_{\max}(n)} \mathrm{d}z' \frac{\mathrm{d}J_\alpha^{(n)}}{\mathrm{d}z'} \\
& \times \frac{D(z')}{D(z)} \left\{ [1 + b(z')] j_0(kr) - \frac{2}{3} j_2(kr) \right\},
\end{aligned} \tag{47}
$$

where $D(z)$ is the linear growth function, $r = r(z, z')$ is the distance to the source and the $j_l(x)$ are spherical Bessel functions of order l. The first term in brackets accounts for galaxy bias while the second describes velocity effects. The ratio $D(z')/D(z)$ accounts for the growth of perturbations between z' and z. Each resonance contributes a differential comoving Lyα flux $\mathrm{d}J_\alpha^{(n)}/\mathrm{d}z'$, calculated from equation (33). This analytic prescription has been bound to match later simulations fairly well [116]. At present, simulations generally do not account for the full Lyα radiative transfer so this agreement is not unexpected and comparisons with future, more detailed simulations will be needed.

On large scales, $W_{\alpha,\star}(k)$ approaches the average bias of sources, while on small scales it dies away rapidly encoding the property that the Lyα flux becomes more uniform. In addition to the fluctuations in $J_{\alpha,\star}$, there will be fluctuations in $J_{\alpha,X}$. We discuss these below, but note in passing that the effective value of W_α is the weighted average $W_\alpha = \sum_i W_{\alpha,i}(J_{\alpha,i}/J_\alpha)$ of the contribution from stars and x-rays.

This description of the coupling fluctuations neglects one important piece of physics. Namely, it assumes that UV photons redshift until they reach the line centre of a Lyman series resonance and only then they scatter. At redshifts before reionization the Universe is filled with neutral hydrogen leading to an optical depth for scattering of Lyα photons that is very large $\tau_\alpha \gtrsim 10^5$. While this is reduced for higher series transitions, the Universe is optically thick for all but the highest n transitions. This has the consequence that photons will preferentially scatter in the wings of the line and only a few will make it to the line centre.

The related implications have been investigated in the literature [117–119]. Since UV photons scatter in the wings of the line they will travel a significantly reduced distance from the source before scattering. This reduces the size of the coupled region around a source and acts to steepen the flux profile surrounding a source. These effects increase the variance in the Lyα flux on small scales. This can be incorporated into our analytic formalism as a modification to the $\mathrm{d}J_\alpha^{(n)}/\mathrm{d}z'$ term in (47) to account for the modified flux profile around individual sources. For our purposes, we will neglect this since it adds considerable numerical complication without modifying the qualitative features of the power spectrum.

Once the effect of the optical depth is taken into account, another limitation of this approach becomes apparent. We have assumed that the photons propagate through a uniform IGM with mean properties. Density inhomogeneities and especially velocity flows may modify the propagation and scattering of UV photons leading to extra fluctuations that have not been accounted for here.

This becomes especially apparent when one recalls that the sources are placed in ionized bubbles. This means that on small scales there are regions where there is no coupling because there is no neutral hydrogen. A simple way to account for this is by changing the lower limit of (47) to $z_{\mathrm{H\,II}}$, the typical size of an ionized region [119]. Clearly, if the patch of gas being considered is closer to a source than the H II region that surrounds that source then the patch will be ionized and there will be no signal. This leads to a reduction in power on small scales. For the sharp cutoff of [119] one numerically sees oscillation in the power spectrum, but in reality averaging over a distribution of ionized bubble sizes would remove these oscillations yielding a smooth reduction in power.

4.4. Formalism for temperature and ionization fluctuations from x-rays

We next turn to fluctuations in the gas temperature and the free electron fraction in the IGM. Following on from our discussion of the global history, we will assume that the evolution of T_K and x_e is driven by the effect of x-rays and will consider the fluctuations δ_T, and δ_e that arise from clustering of the x-ray sources. In doing so, we will follow the approach of [54].

Before plunging into the calculation, it is worth detailing some of the issues that face radiative transfer of x-ray photons. The cross-section for x-ray photoionization depends sensitively on photon energy E, $\sigma \sim E^{-3}$, so that we must keep track of separate energies. Most importantly, this energy dependence means that the IGM is optically thick for soft x-rays ($E \sim 20\,\mathrm{eV}$), but optically thin for hard x-rays ($E \gtrsim 1\,\mathrm{keV}$). Moreover, heating is a continuous process and the temperature of a gas element depends on its past history. This is different from UV coupling where only the Lyα flux at a given redshift is important. We must therefore be careful to track the change in fluctuations in the heating and integrate these to get the temperature fluctuations at a later time. Insight can be gained by looking at the results of 1D numerical simulations of x-ray radiative transfer [53, 55].

The prescription we adopt here describes x-ray fluctuations produced by the clustering of x-ray sources. We neglect the possibility of Poisson fluctuations in the distribution of x-ray sources, since it is not clear how to calculate this. Poisson fluctuations in the Lyα flux were considered in [34], but the formalism is not readily applicable to heating. As a result, this formalism is most appropriate in scenarios where x-ray sources are relatively common and would not be appropriate for describing heating by extremely rare quasars [120].

We begin by forming equations for the evolution of δ_T and δ_e (the fractional fluctuation in x_e) by perturbing equations (15) and (17) (see also [51, 121]). This gives

$$
\frac{\mathrm{d}\delta_T}{\mathrm{d}t} - \frac{2}{3}\frac{\mathrm{d}\delta}{\mathrm{d}t} = \sum_i \frac{2\bar{\Lambda}_{\mathrm{heat},i}}{3k_B \bar{T}_K}[\delta_{\Lambda_{\mathrm{heat},i}} - \delta_T], \tag{48}
$$

$$
\frac{\mathrm{d}\delta_e}{\mathrm{d}t} = \frac{(1 - \bar{x}_e)}{\bar{x}_e}\bar{\Lambda}_e[\delta_{\Lambda_e} - \delta_e] - \alpha_A C \bar{x}_e \bar{n}_H[\delta_e + \delta], \tag{49}
$$

where an overbar denotes the mean value of that quantity and $\Lambda = \epsilon/n$ is the ionization or heating rate per baryon.

We note that the fluctuation in the neutral fraction δ_x is simply related to the fluctuation in the free electron fraction by $\delta_x = -x_e/(1 - x_e)\delta_e$.

Our challenge then is to fill in the right-hand side of these equations by calculating the fluctuations in the heating and ionizing rates. We will focus here on Compton and x-ray heating processes. Perturbing equation (24) we find that the contribution of Compton scattering to the right-hand side of equation (48) becomes [51]

$$\frac{2\bar{\Lambda}_{\text{heat,C}}}{3k_B\bar{T}_K}[\delta_{\Lambda_{\text{heat,C}}} - \delta_T] = \frac{\bar{x}_e}{1 + f_{\text{He}} + \bar{x}_e}\frac{a^{-4}}{t_\gamma}$$
$$\times \left[4\left(\frac{\bar{T}_\gamma}{\bar{T}_K} - 1\right)\delta_{T_\gamma} + \frac{\bar{T}_\gamma}{\bar{T}_K}(\delta_{T_\gamma} - \delta_T)\right], \quad (50)$$

where δ_{T_γ} is the fractional fluctuation in the CMB temperature, and we have ignored the effect of ionization variations in the neutral fraction outside of the ionized bubbles, which are small. Before recombination, tight coupling sets $T_K = T_\gamma$ and $\delta_T = \delta_{T_\gamma}$. This coupling leaves a scale dependent imprint in the temperature fluctuations, which slowly decreases in time. We will ignore this effect, as it is small ($\sim$10%) below $z = 20$ and once x-ray heating becomes effective any memory of these early temperature fluctuations is erased. At low z, the amplitude of fluctuations in the CMB δ_{T_γ} becomes negligible, and equation (50) simplifies. Our main challenge then is to calculate the fluctuations in the x-ray heating. We shall achieve this by paralleling the approach we took to calculating fluctuations in the Lyα flux from a population of stellar sources.

First, note that for x-rays $\delta_{\Lambda_{\text{ion}}} = \delta_{\Lambda_{\text{heat}}} = \delta_{\Lambda_\alpha} = \delta_{\Lambda_X}$, as the rate of heating, ionization and production of Lyα photons differ only by constant multiplicative factors (provided that we may neglect fluctuations in x_e, which are small, and focus on x-ray energies $E \gtrsim 100$ eV). In each case, fluctuations arise from variation in the x-ray flux. We then write $\delta_{\Lambda_X} = W_X(k)\delta$ and obtain

$$W_X(k) = \frac{1}{\bar{\Lambda}_X}\int_{E_{\text{th}}}^\infty dE \int_z^{z_\star} dz' \frac{d\Lambda_X(E)}{dz'}$$
$$\times \frac{D(z')}{D(z)}\left\{[1 + b(z')]j_0(kr) - \frac{2}{3}j_2(kr)\right\}, \quad (51)$$

where the contribution to the energy deposition rate by x-rays of energy E emitted with energy E' from between redshifts z' and $z' + dz'$ is given by

$$\frac{d\Lambda_X(E)}{dz'} = \frac{4\pi}{h}\sigma_\nu(E)\frac{dJ_X(E, z)}{dz'}(E - E_{\text{th}}), \quad (52)$$

where $\sigma_\nu(E)$ is the cross-section for photoionization, E_{th} is the minimum energy threshold required for photoionization and $\bar{\Lambda}_X$ is obtained by performing the energy and redshift integrals. Note that rather than having a sum over discrete levels, as in the Lyα case, we must integrate over the x-ray energies. The differential x-ray number flux is found from equation (28).

The window function $W_X(k)$ gives us a 'mask' to relate fluctuations to the density field; its scale dependence means that it is more than a simple bias. The typical sphere of influence of the sources extends out to several Mpc. On scales smaller than this, the shape of $W_X(k)$ will be determined by the details of the x-ray source spectrum and the heating cross-section. On larger scales, the details of the heated regions remain unresolved so that $W_X(k)$ will trace the density fluctuations.

An x-ray is emitted with energy E' at a redshift z' and redshifts to an energy E at redshift z, where it is absorbed. To calculate W_X we perform two integrals in order to capture the contribution of all x-rays produced by sources at redshifts $z' > z$. The integral over z' counts x-rays emitted at all redshifts $z' > z$ which redshift to an energy E at z; the integral over E then accounts for all the x-rays of different energies arriving at the gas element. Together, these integrals account for the full x-ray emission history and source distribution. Many of these x-rays have travelled considerable distances before being absorbed. The effect of the intervening gas is accounted for by the optical depth term in J_X. Soft x-rays have a short mean free path and are absorbed close to the source; hard x-rays will travel further, redshifting as they go, before being absorbed. Correctly accounting for this redshifting when calculating the optical depth is crucial as the absorption cross-section shows strong frequency dependence. In our model, heating is dominated by soft x-rays, from nearby sources, although the contribution of harder x-rays from more distant sources cannot be neglected.

Returning now to the calculation of temperature fluctuations, to obtain solutions for equations (48) and (49), we let $\delta_T = g_T(k, z)\delta$, $\delta_e = g_e(k, z)\delta$, $\delta_\alpha = W_\alpha(k, z)\delta$ and $\delta_{\Lambda_X} = W_X(k, z)\delta$ [102]. Since we do not assume these quantities to be independent of scale we must solve the resulting equations for each value of k. Note that we do not include the scale dependence induced by coupling to the CMB [51]. In the matter dominated limit, we have $\delta \propto (1 + z)^{-1}$ and so obtain

$$\frac{dg_T}{dz} = \left(\frac{g_T - 2/3}{1 + z}\right) - Q_X(z)[W_X(k) - g_T] - Q_C(z)g_T, \quad (53)$$

$$\frac{dg_e}{dz} = \left(\frac{g_e}{1 + z}\right) - Q_I(z)[W_X(k) - g_e] + Q_R(z)[1 + g_e], \quad (54)$$

where we define

$$Q_I(z) \equiv \frac{(1 - \bar{x}_e)}{\bar{x}_e}\frac{\bar{\Lambda}_{\text{ion},X}}{(1 + z)H(z)}, \quad (55)$$

$$Q_R(z) \equiv \frac{\alpha_A C \bar{x}_e \bar{n}_H}{(1 + z)H(z)}, \quad (56)$$

$$Q_C(z) \equiv \frac{\bar{x}_e}{1 + f_{\text{He}} + \bar{x}_e}\frac{(1 + z)^3}{t_\gamma H(z)}\frac{T_\gamma}{\bar{T}_K}, \quad (57)$$

and

$$Q_X(z) \equiv \frac{2\bar{\Lambda}_{\text{heat,X}}}{3k_B\bar{T}_K(1 + z)H(z)}. \quad (58)$$

These are defined so that Q_R and Q_I give the fractional change in x_e per Hubble time as a result of recombination and ionization, respectively. Similarly, Q_C and Q_X give the fractional change in $\bar{T}_K$ per Hubble time as a result of Compton and x-ray heating. Immediately after recombination Q_C is large, but it becomes negligible once Compton heating becomes ineffective at $z \sim 150$. The Q_R term

becomes important only towards the end of reionization, when recombinations in clumpy regions slows the expansion of H II regions. Only the Q_X and Q_I terms are relevant immediately after sources switch on. We must integrate these equations to calculate the temperature and ionization fluctuations at a given redshift and for a given value of k.

These equations illuminate the effect of heating. First, consider g_T, which is related to the adiabatic index of the gas γ_a by $g_T = \gamma_a - 1$, giving it a simple physical interpretation. Adiabatic expansion and cooling tends to drive $g_T \rightarrow 2/3$ (corresponding to $\gamma_a = 5/3$, appropriate for a monoatomic ideal gas), but when Compton heating is effective at high z, it deposits an equal amount of heat per particle, driving the gas towards isothermality ($g_T \rightarrow 0$). At low z, when x-ray heating of the gas becomes significant, the temperature fluctuations are dominated by spatial variation in the heating rate ($g_T \rightarrow W_X$). This embodies the higher temperatures closer to clustered sources of x-ray emission. If the heating rate is uniform $W_X(k) \approx 0$, then the spatially constant input of energy drives the gas towards isothermality $g_T \rightarrow 0$.

The behaviour of g_e is similarly straightforward to interpret. At high redshift, when the IGM is dense and largely neutral, the ionization fraction is dominated by the recombination rate, driving $g_x \rightarrow -1$, because denser regions recombine more quickly. As the density decreases and recombination becomes ineffective, the first term of equation (54) gradually drives $g_x \rightarrow 0$. Again, once ionization becomes important, the ionization fraction is driven towards tracking spatial variation in the ionization rate ($g_x \rightarrow W_X$). Note that, because the ionization fraction in the bulk remains less than a few per cent, fluctuations in the neutral fraction remain negligibly small at all times.

Fluctuations in the temperature g_T attempt to track the heating fluctuations $W_X(k)$, but two factors prevent this. First, until heating is significant, the effect of adiabatic expansion tends to smooth out variations in g_T. Second, g_T responds to the integrated history of the heating fluctuations, so that it tends to lag W_X somewhat. When the bulk of star formation has occurred recently, as when the SFR is increasing with time, then there is little lag between g_T and W_X. In contrast, when the SFR has reached a plateau or is decreasing the bulk of the x-ray flux originates from noticeably higher z and so g_T tends to track the value of W_X at this higher redshift. On small scales, the heating fluctuations are negligible and g_T returns to the value of the (scale independent) uniform heating case.

4.5. Evolution of the full power spectrum

Having described the different components of the 21 cm power spectrum, we now need to put them together. The 21 cm power spectrum is a 3D quantity observed as a function of scale and redshift, much like a movie evolving with time on a 2D screen. Displaying this information on static 2D paper is therefore challenging.

Figure 8 shows the evolution of the power spectrum as a function of redshift for several fixed k-values. Four key epochs can be picked out. At early times $z \gtrsim 30$ before star formation, the power spectrum rises towards a peak at $z \approx 50$
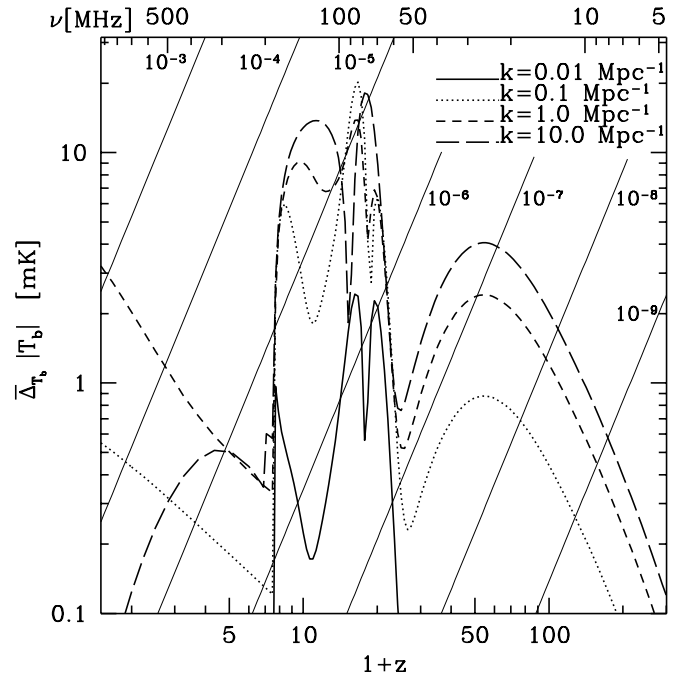

Figure 8. Evolution of power spectrum fluctuations. The different curves show $P(k, z)$ as a function of z at fixed k for $k = 0.01, 0.1, 1, 10 \, \text{Mpc}^{-1}$. Diagonal lines show $\epsilon T_{\text{fg}}(\nu)$, the foreground temperature reduced by a factor ϵ ranging from 10^{-3} to 10^{-9} to indicate the level of foreground removal required to detect the signal. Reproduced with permission from [79]. Copyright 2008 American Physical Society.

and falls thereafter as the 21 cm power spectrum tracks the density field modulated by the mean brightness temperature. Once stars switch on, there is a period of complicated evolution as coupling and temperature fluctuations become important. Next, ionization fluctuations become important culminating in the loss of signal at the end of reionization. Thereafter, a weaker signal arises from the remaining neutral hydrogen in dense clumps that grows as structures continue to collapse.

Diagonal lines in figure 8 show the amplitude of the mean foreground reduced by a factor ranging from 10^{-3} to 10^{-9} (note that it is really fluctuations in the foregrounds that set the upper limit, but this is less well known than the mean level). This gives a measure of the difficulty of removing the foregrounds and the sensitivity of the instrument required for a detection. This diagonal lines serve to identify the most readily accessible epochs. The 10^{-5} line identifies the level required to (i) access density fluctuations at $z \lesssim 4$, (ii) access ionization fluctuations at $z \approx 6$–12 and (iii) detect fluctuations from heating and coupling at $z \approx 20$. It is perhaps surprising that a similar sensitivity is needed for all three of these regimes, since the foregrounds grow monotonically with decreasing frequency. This is explained by the strong evolution in the amplitude of the signal. The fourth epoch at $z \approx 30$–50 of the Dark Ages is considerably more difficult to detect, requiring the 10^{-7} level of foreground removal.

4.6. Other sources of fluctuations

We have focused on fluctuations in the 21 cm signal that arise from spatial variation in three main radiation fields: UV, Lyα

and x-rays. Other sources of fluctuations from the non-linear growth of structure are possible.

Since the 21 cm signal arises from neutral hydrogen it is interesting to examine the densest regions of neutral hydrogen. These occur in those dense clumps that have achieved the critical density for collapse, but that lack the mass for efficient cooling of the gas that would lead to star formation. This requirement is satisfied for halos with virial temperature below 10^4 K provided that only atomic hydrogen is available for cooling or for halos with $T_{vir} \lesssim 300$ K if molecular hydrogen is present. These minihalos should be abundant in the early Universe, although at later times external ionizing radiation may cause them to evaporate. Furthermore, their formation may be prevented with moderate gas heating that raises the Jean's mass suppressing the collapse of these low-mass objects. The high density in these regions implies that collisions can provide the main coupling mechanism and due to the high temperature of the virialized gas these minihalos are bright in 21 cm emission [122–125].

When baryons and photons decouple at $z \approx 1, 100$ there is a sudden drop in the pressure supporting the baryons against gravitational collapse and they begin to fall into dark matter overdensities. It was recently realized that the relative velocity between the baryons and dark matter exceeds the local sound speed (which drops dramatically from the relativistic value of $c/\sqrt{3}$ to $\sim \sqrt{k_B T/m_p}$) leading to supersonic flows [126]. These flows have the potential to suppress the formation of the first gas clouds by preventing the baryons from collapsing into dark matter halos with low escape velocities [127]. This suppression of galaxy formation in minihalos may have important consequences for the 21 cm signal during its earliest phases—for example, delaying the onset of Lyα coupling. It has even been suggested that the increased modulation of the earliest galaxies might boost the Lyα coupling fluctuations significantly [128]. Ultimately, it appears that as the higher mass halos required for atomic cooling begin to collapse, the memory of this effect is greatly reduced [129, 130]. It is possible that by suppressing the building blocks of larger galaxies mild echos of this effect might be detectable in late galaxies in a similar way to the effects of inhomogeneous reionization [131–133].

In the local Universe, shocks are known to be an important mechanism for IGM heating. Shocks heat the gas directly converting bulk motion into thermal energy. If magnetic fields are present in the shock, strong shocks can also accelerate charged particles. This can lead to radiative emission of photons at energies from radio to x-ray energies due to inverse Compton scattering of CMB photons from the charged particle. The resulting x-rays can further heat the IGM. At high redshift, shocks can be broadly divided into two categories: strong shocks around large scale structure and weak shocks in the diffuse IGM due to the low sound speed of the gas.

Large scale structure shocks occur as gas surrounding an overdense region decouples from the Hubble flow and undergoes turnaround. As the gas begins to collapse inwards derivations from spherical symmetry will lead to shell crossing and shocking of the gas. These shocks have been discussed both analytically [40] and observed in simulations [21, 125].

The temperature distribution of shocked gas can be estimated using a Press–Schecter formalism under the assumption that all gas that has undergone turn around has shocked and estimating the shock temperature from the characteristic peculiar velocity of the collapsing overdense region. These models show that, in the absence of x-ray heating, the thermal energy of the IGM is dominated by large scale shocks. Since these shocks trace the collapse of structure, they only become significant at redshifts $z \lesssim 20$. These shocks cannot play an important role in ionizing the IGM, since only a small fraction of the baryons participate in them and they generate $\lesssim 1$ eV per participating baryon at $z \gtrsim 10$ [134].

Finally, we note that there is one final radiation background that could in principle lead to fluctuations in the 21 cm signal: the diffuse radio background. In our calculations, we have assumed that the ambient 21 cm radiation field is dominated by the CMB, so that $T_\gamma = T_{CMB}$. This is likely to be a good assumption, but one can imagine a situation in which the 21 cm flux might be influenced by nearby radio-bright sources. Since 21 cm photons have a large mean free path a diffuse radio background may build up in the same way as the diffuse x-ray background grows. Fluctuations in this radio background would arise from clustering of the radio sources. We note that the only period where these fluctuations would be important are in the regime where Lyα and collisional coupling are unimportant, so that the spin temperature relaxes to the ambient radiation temperature T_γ.

4.7. Simulation techniques

In the previous sections we have focused on analytical descriptions of calculating the 21 cm signal. These provide a framework for understanding the underlying physics that governs the signal. They also provide a method for rapidly exploring the dependence of the 21 cm power spectrum on a wide range of astrophysical parameters and determining their relative importance. Ultimately, the interpretation of observations requires comparison of data to theoretical predictions at a more detailed level. This necessitates the production of maps from numerical techniques. In this section we describe a hierarchy of different levels of numerical approximations that allow more quantitative comparisons. A more detailed review can be found in [135].

Closest to the analytic techniques described in this review are semi-numerical techniques. These are primarily techniques for simulating the reionization field based upon an extension of the excursion set formalism [40, 111, 112]. It has been realized that the spectrum of ionized fluctuations depends primarily on a single parameter—the ionized fraction x_i [136]. Once x_i is fixed the ionization pattern can be calculated by filtering the density field on progressively smaller scales and asking if a region is capable of self-ionization. Only those regions capable of self-ionization are taken to be ionized; this amounts to photon counting. This technique forms the basis of a number of codes, 21 cmFAST [137], SIMFAST21 [138], which can rapidly calculate the 21 cm signal with reasonable accuracy. To add in fluctuations in the Lyα and temperature these codes also evaluate the equivalent

of the analytic calculations from section 4.4. The relevant expressions are convolutions of the density field with a kernel describing the astrophysics and so may be rapidly evaluated on a numerical grid via fast Fourier transform (FFT) techniques. Other semi-numerical techniques exist such as BEARS [139], which is based upon painting spherically symmetric ionization, heating or coupling profiles from a library of 1D radiative transfer calculations appropriately scaled for a halo's SFR and formation time onto the density field.

Fully numerical simulations are the ultimate destination for these calculations, but as of yet it is difficult to achieve the required simulation volume and dynamic range required for the full calculation of the 21 cm signal. Numerical simulations can be broadly split into two classes: those that fully account for gas hydrodynamics and those that do not. The latter are essentially dark matter N-body calculations with a prescription for painting baryons onto the simulation in a post-processing step. Radiative transfer prescriptions are then applied in a further post-processing step to calculate the evolution of the ionized regions. Such calculations [106, 136, 140, 141] are of great utility in describing the large scale properties of the 21 cm signal. Full hydrodynamic calculations [142] are typically restricted to smaller volumes making them unrepresentative of the cosmic volume. They have the advantage of self-consistently evolving the dark matter, baryons and radiation field. This allows the evolution of photon sinks, in the form of minihalos and dense clumps, to be studied in detail. As computers improve these simulations will grow in size.

Finally, most of the simulation work to date has focused on reionization and made the assumption that $T_S \gg T_{CMB}$ ignoring the effect of spin-temperature fluctuations. For predicting the 21 cm signal it is important that analytic calculations of these T_S variation be verified numerically. Work on incorporating Lyα and x-ray propagation into reionization simulations exists, but is at an early stage of development [118, 143, 144]. These calculations require keeping track of the radiative transfer of photons in many frequency bins and so becomes numerically expensive.

4.8. Detectability of the 21 cm signal

Our focus in this review is on the physics underlying the 21 cm signal, but it is appropriate to pause for a moment and consider the instrumental requirements for detecting the signal. We direct the interested reader to the review in [5], which covers the near term prospects for detecting the 21 cm signal from reionization in some detail.

The 21 cm line from the epoch of reionization (FOR) is redshifted to meter wavelengths requiring radio frequency observations. While a typical radio telescope is made of a single large dish, an interferometer composed of many dipole anntenae is the preferred design for 21 cm observations. Cross-correlating the signals from individual dipoles allows a beam to be synthesized on the sky. Using dipoles allows for arrays with very large collecting areas and a large field of view suitable for surveys. This comes at the cost of large computational demands and, driven by the long wavelengths, relatively poor angular resolution ($\sim$10 arcmin, which corresponds to the predicted bubble size at the middle of reionization).

The sensitivity of these arrays is determined in part by the distribution of the elements, with a concentrated core configuration giving the highest sensitivity to the power spectrum [145]. The desire for longer baselines to boost the angular resolution in order to remove radio point sources places constraints on the compactness achievable in practice.

The variance of a 21 cm power spectrum measurement with a radio array for a single $\boldsymbol{k}$-mode with line of sight component $k_{||} = \mu k$ is given by [145]:

$$\sigma_P^2(k, \mu) = \frac{1}{N_{field}}$$
$$\times \left[\bar{T}_b^2 P_{21}(k, \mu) + T_{sys}^2 \frac{1}{B t_{int}} \frac{D^2 \Delta D}{n(k_\perp)} \left(\frac{\lambda^2}{A_e} \right)^2 \right]^2. \quad (59)$$

We restrict our attention to modes in the upper-half plane of the wavevector $\boldsymbol{k}$, and include both sample variance and thermal detector noise assuming Gaussian statistics. The first term on the right-hand side of the above expression provides the contribution from sample variance, while the second describes the thermal noise of the radio telescope. The thermal noise depends upon the system temperature T_{sys}, the survey bandwidth B, the total observing time t_{int}, the comoving distance $D(z)$ to the centre of the survey at redshift z, the depth of the survey ΔD, the observed wavelength λ and the effective collecting area of each antennae tile A_e. The effect of the configuration of the antennae is encoded in the number density of baselines $n_\perp(k)$ that observe a mode with transverse wavenumber $k_\perp$ [146]. Observing a number of fields N_{field} further reduces the variance.

4.9. Statistics beyond the power spectrum

In our discussion of the 21 cm fluctuations, we have focused on the power spectrum as a statistical quantity that can be measured from maps of the sky. When experiments lack the sensitivity to image the 21 cm fluctuations at high signal-to-noise ratio it is important to compress the information into statistical quantities that can be accurately measured. If 21 cm fluctuations were Gaussian then the power spectrum would contain all information about the signal. However, the 21 cm signal is highly non-Gaussian since the presence of ionized bubbles or spheres that have been heated imprints definite features in space. It is therefore important to think about non-Gaussian statistics that can be used to extract information that is not contained in the power spectrum from 21 cm observations.

The simplest form of non-Gaussianity that arises in the 21 cm signal is the primordial non-Gaussianity induced in the density field during inflation. This is often characterized by the f_{NL} parameter defined by assuming a quadratic correction to the inflaton potential so that schematically $\phi = \phi_G + f_{NL}\phi_G^2$, with ϕ the full inflaton potential and ϕ_G the Gaussian potential. This form of non-Gaussianity shows up clearly in the bispectrum, the Fourier transform of the three-point correlation function [147]. Measuring f_{NL} is an important goal of cosmology since it can effectively distinguish between different classes of inflationary potential. Unfortunately, f_{NL} is expected to be small ($\sim (1 - n_s) \sim 0.05$ for slow roll inflation models) and constraints from the CMB and galaxy

surveys are relatively weak [148]. One long term hope is that observations of the pristine 21 cm signal at $z > 30$ could lead to very stringent non-Gaussianity constraints, since such surveys can probe very large volumes of space [149, 150].

In the near term, studies of 21 cm non-Gaussianity will be most useful for learning about astrophysics in more detail. Ionized bubbles during reionization will imprint a particular form of non-Gaussianity on the 21 cm signal that should contain useful information about the sizes and topology of ionized regions. The challenge is to develop statistics matched to that form of non-Gaussianity, which is currently an unsolved problem. Some examples of statistics that have been explored are the one-point function (or probability distribution function) of the 21 cm brightness field, which develops a skewness as reionization leads to many pixels with zero signal [40, 151]; the Euler characteristic [152], which determines the number of holes in a connected surface; and there are many other possibilities including the bispectrum, wavelets and threshold statistics [153] that have yet to be properly explored. In addition to new statistics, it is possible for signatures of non-Gaussianity to show up as a modification to the shape of the power spectrum [154].

4.10. Prospects for cosmology

In this review, we have mostly focused on the ability of the 21 cm signal to constrain different aspects of astrophysics and the physics that needs to be understood in order to do so. In the first instance the effects of the astrophysics of galaxy formation needs to be understood before fundamental physics can be addressed. Once this is understood there is considerable potential for addressing cosmology. 21 cm observations represent one of the only ways of accessing the large volumes and linear modes present at high redshifts. Studies show that 21 cm experiments have considerable potential to improve on measurements of cosmological parameters [103, 155], on models of inflation [156, 157], and on measurements of the neutrino mass [158].

We have touched upon the possibility of constraining exotic heating and structure formation in the early Universe via the 21 cm global signal. These experiments are primarily sensitive to key transitions such as the onset of Lyα coupling, the onset of heating and the reionization of the Universe. These key moments can be affected by different heating mechanisms in similar ways and so most likely will provide upper limits on the contribution of exotic physics. Such elements as heating by dark matter decay or annihilation [81–84], evaporating primordial black holes [87], or other such physics might all be constrained.

Ideally one would target the cosmic Dark Ages, where the 21 cm signal depends on well-understood linear physics. In this regime, one would have the possibility of making precision measurements similar to the CMB, but at many different redshifts. This represents the long term goal of 21 cm observations as a tool of cosmology. Unfortunately, many technical challenges need to be overcome before this is realistic. Foregrounds become much larger relative to the signal at the low frequencies relevant for the cosmic Dark

Ages. This requires large experiments to achieve the required sensitivity. Since the sensitivity of an interferometer to the 21 cm power spectrum typically decreases on smaller angular scales as a result of fewer long baselines, large angular scales are easier to access. It was shown in [159] that an experiment with a $10 \, \text{km}^2$ collecting area could detect a constrained isocurvature mode via its signature on large angular scales in the 21 cm power spectrum at $z \approx 30$. In contrast, detecting the running of the primordial power spectrum via the 21 cm power spectrum on scales $k \gtrsim 1 \, \text{Mpc}^{-1}$ requires experiments with $\gtrsim 10 \, \text{km}^2$ to achieve the necessary sensitivity [157].

These low frequencies will probably require escaping the Earth's ionosphere to space or the lunar surface [160]. The payoff would be unprecedented precision on cosmological parameters and a precise picture of the earliest stages of structure formation. This could lend insight into fundamental physics of phenomena like variations in the fine-structure constant [161] and the presence of cosmic superstrings [162].

More accessible in the near future, is the epoch of reionization (EoR) where the astrophysics of star formation must be disentangled to get at cosmology. This will likely involve the use of redshift space distortions and astrophysical modelling as a way of removing the astrophysics leaving the cosmological signal behind. In the future, with sensitivity sufficient to image the 21 cm fluctuations one can imagine developing sophisticated algorithms to exploit the full 3D information in the maps to separate astrophysics, much as one currently removes foregrounds in the CMB. This has not been studied in great detail, but preliminary work on modelling out the contribution of reionization appears promising.

It has been shown in simulations that just one parameter—the ionized fraction x_i—does a good job of describing the ionization power spectrum at the few per cent level. This arises from well-understood physics and amounts to a simple photon counting argument. Thus, at first order we might hope to be able to predict the ionization contribution to the 21 cm signal during reionization and marginalize over it in the same manner as the bias of galaxies in galaxy surveys. At higher levels of precision, accounting for the modifications from the clustering of the sources or sinks of ionizing photons will become important. Modelling of the ionization power spectrum need only degrade cosmological constraints by factors of a few [155].

5. Intensity mapping in atomic and molecular lines

5.1. 21 cm intensity mapping and dark energy

Reionization eliminated most of the neutral hydrogen in the Universe, but not all. In overdense regions the column depth can be sufficient for self-shielding to preserve neutral hydrogen. Examples of such pockets of neutral hydrogen are seen as damped Lyα systems in quasar spectra. Whereas the pre-reionization 21 cm signal provides information about the topology of ionized region, the post-reionization signal describes the clustering of collapsed halos based on the underlying density field of matter.

The notion of IM of 21 cm bright galaxies was introduced in several papers [163–165], and similar ideas were discussed
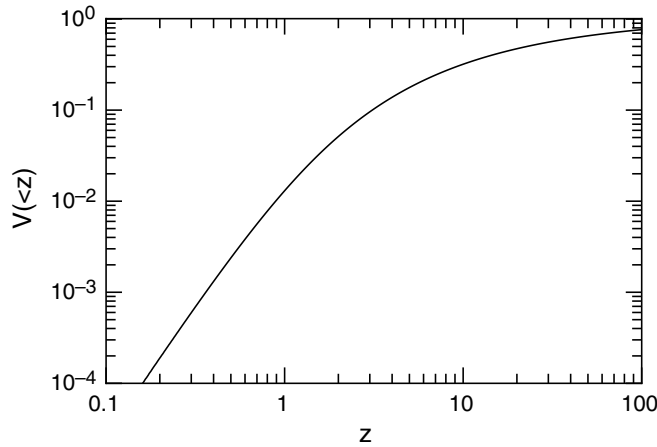
Figure 9. The fraction of the total comoving volume of the observable Universe that is available up to a redshift z. Reproduced with permission from [163]. Copyright 2008 American Physical Society.

in [166]. The starting point was the many experimental efforts to detect the 21 cm signal from the EoR when the IGM is close to fully neutral. After reionization approximately 1% of the baryons are contained in neutral hydrogen ($\Omega_{H_I} \approx 0.01$). Although this reduces the signal significantly relative to a fully neutral IGM, the amplitude of galactic synchrotron emission (which provides the dominant foreground) is several orders of magnitude less at the frequencies corresponding to 21 cm emission at redshifts $z = 1$–3. Consequently, the signal-to-noise ratio for radio experiments targeted at IM might be expected to be comparable to experiments targeted at the EoR. In both cases, the new technologies driven by developments in computing power makes possible previously unattempted observations.

Such observations would be extremely important for our understanding of the Universe. It is perhaps humbling to realize that existing observations from the CMB and galaxy surveys fill only a small fraction of the potentially observable Universe. Figure 9(a) shows the available comoving volume out to a given redshift. At present, the deepest sky survey over a considerable fraction of the sky is the first Sloan Digital Sky Survey (SDSS), whose luminous red galaxies (LRG) sample has a mean redshift of $z \approx 0.3$. By measuring the 21 cm intensity fluctuations at $z = 1$–3 the comoving volume probed by experiment would be increased by two orders of magnitude. This is particularly important since our ability to constrain cosmological parameters depends critically on the survey volume. In general, constraints scale with volume V as $\sigma \propto V_{\text{eff}}^{-1/2}$ since a larger volume means that more independent Fourier modes can be constrained. For example, one might improve constraints on quantities such as the neutrino mass and the running of the primordial power spectrum [167].

Perhaps the greatest utility for IM surveys is in constraining dark energy via measurements of baryon acoustic oscillations (BAOs) in the galaxy power spectrum. BAOs arise from the same physics that produces the spectacular peak structure in the CMB. These oscillations have a characteristic wavelength set by the sound horizon at recombination and so provide a 'standard ruler'. Measurements of this distance scale can be used to probe the geometry of the Universe and to constrain its matter content. Measurements of the oscillations in the CMB at $z \approx 1100$ and more locally in the BAOs at $z \lesssim 3$ would provide exquisite tests of the flatness of the Universe and of the equation of state of the dark energy. Redshifts in the range $z \approx 1$–3 are of great interest since they cover the regime in which dark energy begins to dominate the energy budget of the Universe. While galaxy surveys will begin to probe this range in the next decade, covering sufficiently large areas to the required depth is very challenging for galaxy surveys. This has been considered in depth by the Dark Energy Task Force (DETF) [168].

To get a sense of the observational requirements of BAO observations with 21 cm IM, we consider some numbers. Beyond the third peak of the BAO non-linear evolution begins to wash out the signal. The peak of the third BAO peak has a comoving wavelength of $35h^{-1}$ Mpc, which for adequate samples requires pixels perhaps half that size. At $z = 1.5$ the angular scale corresponding to $18h^{-1}$ Mpc is 20 arcmin, which sets the required angular resolution of the instrument. Estimating the H I mass enclosed in such a volume is uncertain, but based on the observed value of $\Omega_{H_I} \sim 10^{-3}$ is $\sim 2 \times 10^{12} M_\odot$.

The mean brightness temperature T_b of the 21 cm line can be estimated using

$$T_b = 0.3(1+\delta)\left(\frac{\Omega_{H_I}}{10^{-3}}\right)\left(\frac{\Omega_m + a^3 \Omega_\Lambda}{0.29}\right)^{-1/2}$$
$$\times \left(\frac{1+z}{2.5}\right)^{1/2} \text{mK}, \tag{60}$$

where $1 + \delta = \rho_g / \bar{\rho}_g$ is the normalized neutral gas density and $a = (1+z)^{-1}$ is the scale factor. The amplitude of the signal is therefore considerably smaller than that of the 21 cm signal during reionization. However, the foregrounds are reduced by a factor of $[(1+8)/(1+1.5)]^{-2.6} \sim 30$, redeeming the situation. Diffuse foregrounds may be removed in a similar fashion to the 21 cm signal at high redshift. Figure 10 shows the relevant region of the BAO that could be detected. The top excluded region accounts for the non-linear growth of structures, which erases the baryon wiggles in the power spectrum. The left exclusion region comes from the finite volume of the Universe, which limits the largest wavelength mode that fits within the survey. The right region is a rough estimate of the limit from foregrounds based upon a simple differencing of neighbouring frequency channels to remove the correlated foreground component. The white accessible region demonstrates that the first three BAO peaks could be accessible with IM experiments.

Observations of the BAO can be used to constrain the dark energy equation of state w, which may be parametrized as $w = w_p + (a_p - a)w'$, where w_p and a_p are the value of the equation of state at a pivot redshift where the errors in w and w' are uncorrelated. Figure 11 shows the ability of an IM telescope covering a square aperture of size 200 m × 200 m subdivided into 16 cylindrical sections each 12.5 m wide and 200 m long. After integrating for 100 days with bandwidth 3 MHz, this experiment is roughly comparable to the stage IV milestone set by the DETF committee [168], and provides very good dark energy constraints. The power of such an experiment makes it a potential alternative to more traditional galaxy surveys [167, 169, 170].
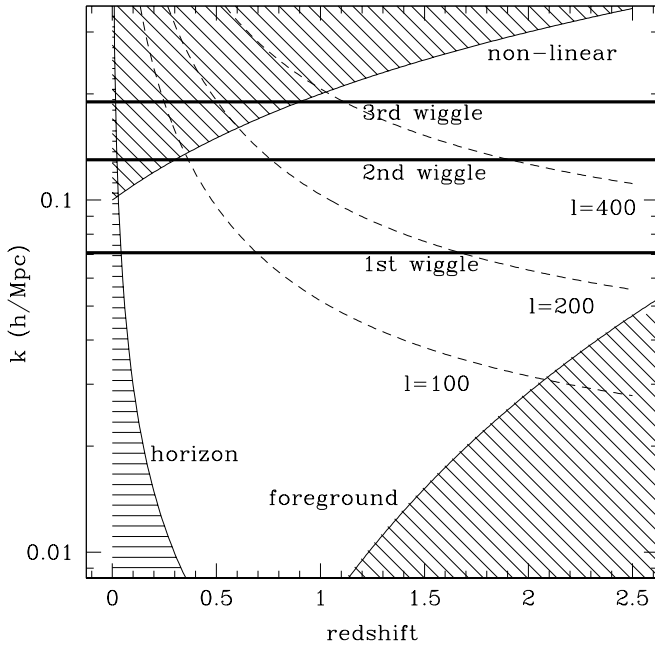
23

Figure 10. The observable parameter space in redshift and in scale (k) for BAO studies. The shaded regions are observationally inaccessible. The horizontal lines indicate the scale of the first three BAO wiggles, and the dashed lines show contours of constant spherical harmonic order ℓ. Reproduced with permission from [165]. Copyright 2008 American Physical Society.
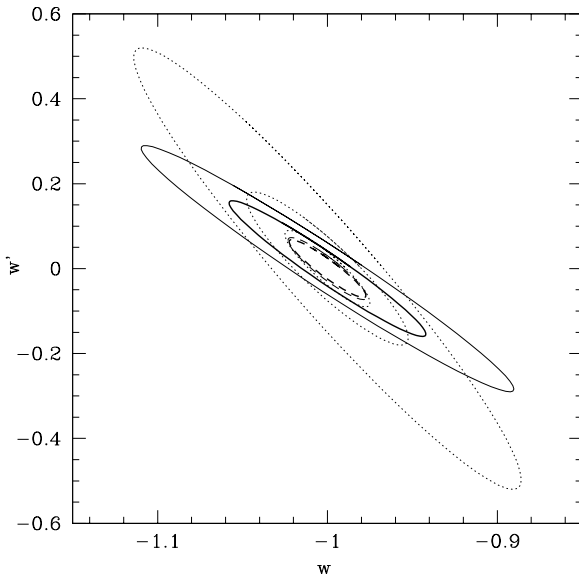

Figure 11. Constraints on the equation of state parameter $w = p/\rho c^2$ (horizontal axis) and its redshift variation for an equation of state $w = w_p + (a_p - a)w'$ (vertical axis) at $z = 0$. The 1-σ contour for IM combined with Planck (inner thick solid line for baseline model, outer thin solid line for worst case), the DETF stage I projects with Planck (outer dotted line), the stage I and III projects with Planck (intermediate dotted line), the stage I, III and IV projects with Planck (inner dotted line), and all above experiments combined (dashed lines, again thick for baseline, thin for worst case; the two contours are nearly indistinguishable). Reproduced with permission from [165]. Copyright 2008 American Physical Society.

Reaching the thermal noise required by a signal of $\sim 300\,\mu$K at frequencies of 600 MHz is within the range of existing telescopes. As a first step towards demonstrating that such observations are possible, the authors of [171] made use of the Green Bank Telescope (GBT) to observe radio spectra corresponding to 21 cm emission at $0.53 < z < 1.12$ over 2 square degrees on the sky. These raw radio data were processed by removing a smooth component, dominated by the foregrounds, to leave a map of intensity fluctuations. Cross-correlating these intensity fluctuations with a catalogue of DEEP2 galaxies in the same field showed a distinct correlation in agreement with theory. These experiments are continuing with the next step to understand instrumental systematics at the level required to cleanly measure the 21 cm intensity auto-power spectrum. They constitute an important step towards demonstrating the feasibility of IM with the 21 cm line.

21 cm IM probes the neutral hydrogen contained within galaxies and one might reasonably worry about how robust a tracer this is of the density field. It is believed that damped Lyα absorbers (DLAs), which contain the bulk of neutral hydrogen, are relatively low-mass systems and so should have relatively low bias (when compared with the highly biased bright galaxies seen in high redshift galaxy surveys) [164]. On top of this, however, one might worry that fluctuations in the ionizing background could lead to spatial variation in the neutral hydrogen content. But after reionization the mean free path for ionizing photons increases and the ionizing background should become fairly uniform with modulation only at the per cent level [170].

5.2. Intensity mapping in other lines

So far, we have described the possibilities for 21 cm IM. The 21 cm line has a number of nice qualities—it is typically optically thin, it is a line well separated in frequency from other atomic lines and hydrogen is ubiquitous in the Universe. It is however only one line out of many that have been observed in local galaxies, which begs the question: Is 21 cm the best line to use and what additional information might be gained by looking at IM in other lines? The analysis of the impact of IM using lines other than the 21 cm line has not yet been fully explored. Because the physical conditions leading to emission by these species are quite varied the cross-species joint analysis of intensity maps is a complex topic that the community has only begun to examine. Because of this variation, the cross-analysis is potentially a rich source of information on conditions at high redshift.

One area where IM in lines other than 21 cm would be particularly interesting is during the EoR. One of the challenges for understanding the first galaxies is the difficulty of placing the galaxies seen in the Hubble Ultradeep Field (HUDF) into a proper context. By focusing on a small patch of sky, the HUDF sees very faint galaxies, but it is unclear how representative this patch is of the whole Universe at that time. For comparison, the full HUDF is approximately 3×3 arcmin in size comparable with the size of an individual ionized bubble, expected to be ~ 10 arcmin in diameter during the middle stages of reionization. Moreover, it is apparent that the
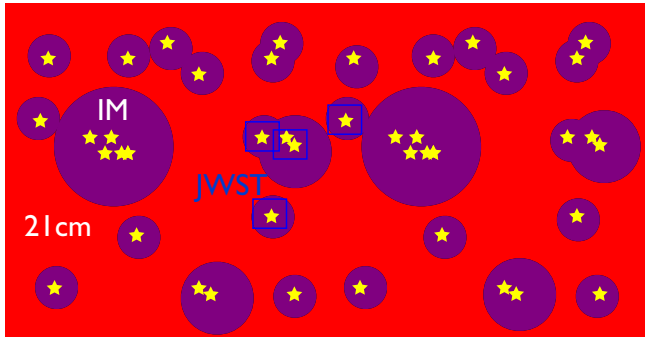
Figure 12. Cartoon of the role IM would play in understanding galaxy formation. Deep galaxy surveys with HST and JWST image the properties of individual galaxies in small fields (blue boxes). 21 cm tomography (red filled region) provides a 'negative space' view of the Universe by determining the properties of the neutral gas surrounding groups of galaxies. IM (purple filled regions) fills in the gaps providing information about the collective properties of groups of galaxies. Together the three would give a complete view of the early generation of galaxies in the infant Universe.

Figure 13. Ratio between line luminosity, L, and SFR, $\dot{M}_*$, for various lines observed in galaxies and taken from table 1 of [174]. For the first seven lines this ratio is measured from a sample of low redshift galaxies. The other lines have been calibrated based on the galaxy M82. Some weaker lines, for example for HCN, have been omitted for clarity.

galaxies seen in the HUDF are the brightest galaxies and that fainter, as yet unseen, galaxies contribute significantly to star formation and reionization. The JWST will see even fainter galaxies and transform our view of the galaxy population at $z \sim 10$, but there will still be a substantial level of star formation that it will not be able to resolve [172].

Combining these galaxy surveys with 21 cm observations and IM would allow a powerful synergy between three independent types of observations directed at understanding the first galaxies and the EoR (illustrated in figure 12). Deep galaxy surveys with HST and JWST would inform us of the detailed properties of small numbers of galaxies during the EoR. 21 cm tomography provides information about the neutral hydrogen gas surrounding groups of galaxies. IM fills in the gaps providing information about the total emission and clustering of the full population of galaxies, even those below the sensitivity threshold of the JWST. Together these three techniques would provide a complete view of galaxies at high redshift and transform our understanding of the origins of galaxy formation.

The first steps towards understanding IM in molecular lines were made by Righi *et al* [173], who considered the possibility that redshifted emission from CO rotational lines might contribute a foreground to CMB experiments. They showed that cross-correlating CMB maps of differing spectral resolution at frequencies of $\nu \approx 30\,\text{GHz}$ could be used to constrain the CO emission of galaxies at $z \approx 3$–8. This might be done, for example, by correlating CMB maps from CBI with those of the Planck LFI (Low Frequency Instrument). These ideas were focused on the notion of 2D maps, but contain the seed of later work.

Visbal and Loeb [175] explored for the first time the notion of looking at a variety of lines in 3D intensity maps. Figure 13 shows the major lines that appear in the emission spectrum of M82. By making a map at the appropriate frequency any of these lines could be studied by IM. Two major challenges arise. The first is that continuum foregrounds are typically larger than the signal from these lines by 2–3 orders of magnitude. This
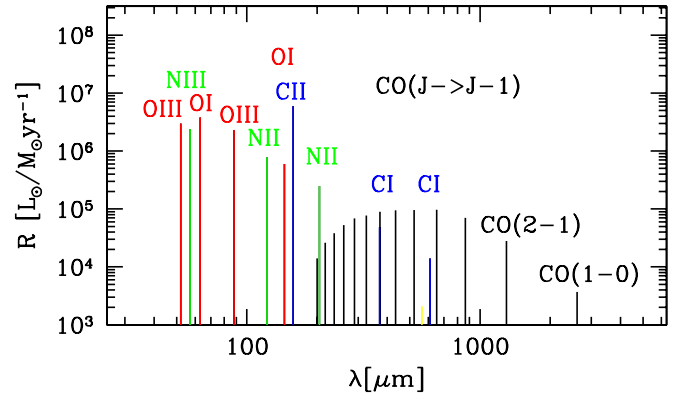
is an identical, although more tractable problem, as occurs in the case of 21 cm studies of the EoR and has been studied in considerable detail. Studies [146, 176, 177] show that, provided the continuum foreground is spectrally smooth in individual sky pixels, it can be removed leaving very little residuals in the cleaned signal.

Potentially more challenging is the issue of line confusion. If we look for the CO(1-0) line (rest frame frequency of 115 GHz) in a map made at 23 GHz (corresponding to emission by CO at $z = 4$) then our map will additionally consider emission from other lines in galaxies at other redshifts (e.g. CO(2-1) from galaxies at $z = 9$). However, the contaminating emission arises from different galaxies which opens the possibility of combining maps at different frequencies corresponding to different lines from the same galaxies as a way of isolating a particular redshift. The emission from lines in the same galaxies will correlate, while emission from lines in galaxies at different redshifts will not[1].

Fortunately, it is possible to statistically isolate the fluctuations from a particular redshift by cross-correlating the emission in two different lines [174, 175]. If one compares the fluctuations at two different wavelengths, which correspond to the same redshift for two different emission lines, the fluctuations will be strongly correlated. However, the signal from any other lines arises from galaxies at different redshifts which are very far apart and thus will have much weaker correlation (see figure 14). In this way, one can measure either the two-point correlation function or power spectrum of galaxies at some target redshift weighted by the total emission in the spectral lines being cross-correlated.

The cross power spectrum at a wavenumber k can be written as

$$P_{1,2}(\vec{k}) = \bar{S}_1 \bar{S}_2 \bar{b}^2 P(\vec{k}) + P_{\text{shot}}, \qquad (61)$$

where $\bar{S}_1$ and $\bar{S}_2$ are the average fluxes in lines 1 and 2, respectively, $\bar{b}$ is the average bias factor of the galactic sources, $P(\vec{k})$ is the matter power spectrum and P_{shot} is the shot-noise

---

[1] This is similar to the suggestion of using the 21 cm map as a template to detect the deuterium hyperfine line [13].
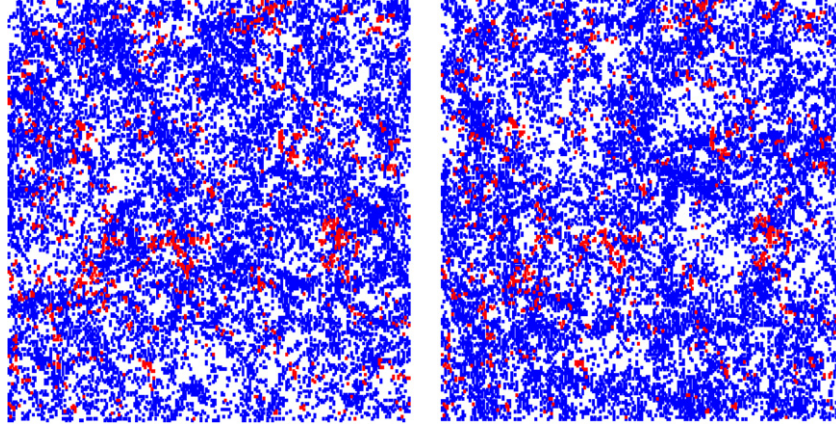
Figure 14. A slice from a simulated realization of line emission from galaxies at an observed wavelength of 441 μm (left) and 364 μm (right) [174]. The slice is in the plane of the sky and spans 250×250 comoving Mpc^2 with a depth of $\Delta\nu/\nu = 0.001$. The colored squares indicate pixels which have line emission greater than 200 Jy Sr^{-1} for the left panel and 250 Jy Sr^{-1} for the right panel. The emission from O I(63 μm) and O III(52 μm) is shown in red on the left and right panels, respectively, originating from the same galaxies at $z = 6$. All of the lines illustrated in figure 13 are included and plotted in blue. Cross-correlating data at these two observed wavelengths would reveal the emission in O I and O III from $z = 6$ with the other emission lines being essentially uncorrelated. Reproduced with permission from [174]. Copyright 2011 SISSA.

power spectrum due to the discrete nature of galaxies. The root-mean-square error in measuring the cross power spectrum at a particular k-mode is given by [175]

$$\delta P_{1,2}^2 = \tfrac{1}{2}(P_{1,2}^2 + P_{1\text{total}}P_{2\text{total}}), \qquad (62)$$

where $P_{1\text{total}}$ and $P_{2\text{total}}$ are the total power spectra corresponding to the first line and second line being cross-correlated. Each of these includes a term for the power spectrum of contaminating lines, the target line and detector noise. Figure 15 shows the expected errors in the determination of the cross power spectrum using the O I(63 μm) and O III(52 μm) lines at a redshift $z = 6$ for an optimized spectrometer on a 3.5 m space-borne infrared telescope similar to SPICA, providing background limited sensitivity for 100 diffraction limited beams covering a square on the sky which is $1.7°$ across (corresponding to 250 comoving Mpc) and a redshift range of $\Delta z = 0.6$ (280 Mpc) with a spectral resolution of $(\Delta\nu/\nu) = 10^{-3}$ and a total integration time of 2×10^6 s.

We emphasize that one can measure the line cross power spectrum from galaxies which are too faint to be seen individually over detector noise. Hence, a measurement of the line cross power spectrum can provide information about the total line emission from all of the galaxies which are too faint to be directly detected. One possible application of this technique would be to measure the evolution of line emission over cosmic time to better understand galaxy evolution and the sources that reionized the Universe. Changes in the minimum mass of galaxies due to photoionization heating of the IGM during reionization could also potentially be measured [175].

As a definite example of the sort of information that might be gathered in molecular line IM, we will discuss CO, the only line other than the 21 cm line to have been somewhat investigated [178–180]. CO emission is widely used as a tracer for the mass of cold molecular gas in a galaxy. This cold molecular gas provides the fuel for star formation and so determining its abundance provides a key constraint on models of star formation and galaxy evolution. Indeed one of the key
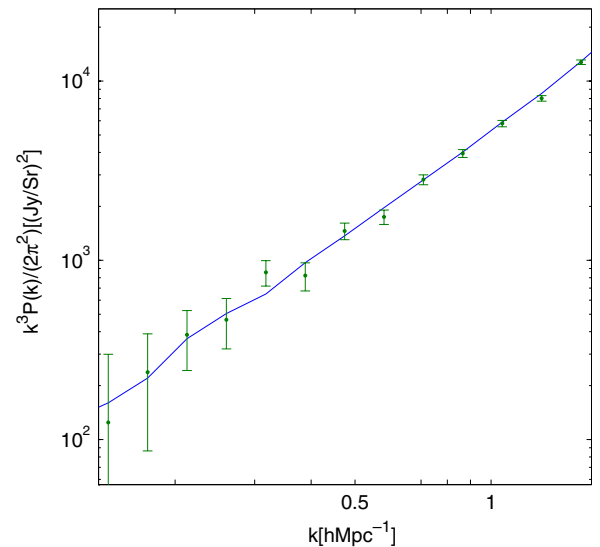

Figure 15. The cross power spectrum of O I(63 μm) and O III(52 μm) at $z = 6$ measured from mock simulation data for a hypothetical infrared space telescope similar to SPICA [174]. The solid line is the cross power spectrum measured when only line emission from galaxies in the target lines is included. The points are the recovered power spectrum when detector noise, contaminating line emission, galaxy continuum emission and dust in our galaxy and the CMB are included. The error bars follow equation (62) with $P_{1\text{total}}$ and $P_{2\text{total}}$ calculated from the simulated data, including detector noise, contaminating line emission and sample variance. Reproduced with permission from [174]. Copyright 2011 SISSA.

science goals for ALMA is to directly image CO emission in individual galaxies at $z \approx 7$. Figure 16 shows the redshift evolution of the mean brightness temperature (a measure of the radio intensity) for the CO(2-1) line as a function of the minimum mass required for a galaxy to be CO bright. These calculations identify a signal strength of 0.01–1 μK at $z = 8$ as the observational target. This represents an order of magnitude increase in sensitivity over existing CMB experiments like CBI and might be enabled by improved techniques with dishes exploiting focal plane arrays.
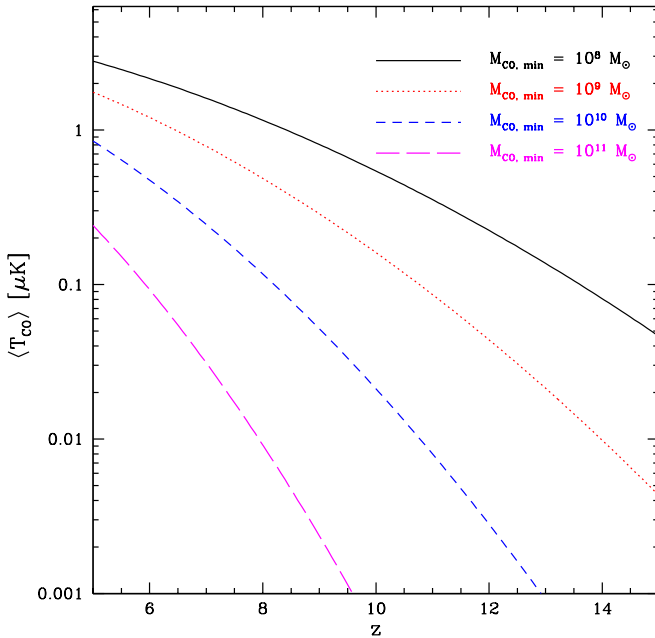
Figure 16. The global mean CO brightness temperature as a function of redshift [179]. The curves show the volume-averaged CO brightness temperature for several different values of $M_{CO,min}$, the minimum halo mass hosting a CO luminous galaxy. In all cases, the minimum host halo mass of star forming galaxies is the atomic cooling mass, $M_{sf,min} \approx 10^8 M_\odot$, and so the curves with larger $M_{CO,min}$ describe models in which galaxies with low SFRs are not CO luminous. If we had instead varied $M_{sf,min}$, while fixing both $M_{CO,min} = M_{sf,min}$ and the SFR density at a given redshift, the model variations would be considerably smaller. Reproduced with permission from [179]. Copyright 2011 American Astronomical Society.

Figure 17. The auto-power spectrum of CO brightness temperature fluctuations [179]. The black solid, red dotted, red dashed, blue dashed and cyan dotted–dashed curves show simulated CO power spectra at different redshifts for various values of $M_{CO,min}$ and fixed duty cycle. The green solid line is the halo model prediction for the signal for $z = 7.3$, $M_{CO,min} = 10^{10} M_\odot$ and $f_{duty} = 0.1$. The green dashed line shows the Poisson term, while the green dotted–dashed curve represents the clustering term. Reproduced with permission from [179]. Copyright 2011 American Astronomical Society.

Figure 17 shows the evolution of the auto-power spectrum of the CO intensity fluctuations. This signal evolves strongly with redshift due to its dependence on $\langle T_{CO} \rangle$ and has a shape that depends on the relative contribution of the clustering of CO bright galaxies and a Poisson shot noise term, which is determined by the distribution of CO galaxy masses. Measurement of the shape and evolution of this power spectrum would therefore give considerable information about the properties of early CO galaxies.

5.3. Cross-correlation of molecular and 21 cm intensity maps

Earlier, we described 21 cm tomography observations with instruments like MWA, LOFAR and PAPER that hope to map the 21 cm emission from neutral hydrogen during the EoR. These maps trace out the ionization field, which is determined by the distribution of galaxies. Regions with many galaxies will tend to have ionized the surrounding IGM leading to a 'hole' in the 21 cm signal. In contrast, since those same galaxies will have undergone considerable star formation and so produced metals, regions with galaxies will appear CO bright. This is clearly seen in the bottom two panels of figure 18—blue CO bright regions correspond to black patches with no signal in the 21 cm map. This anti-correlation can be detected statistically and provides a very direct constraint on the size of the ionized regions. Maps made in lines from
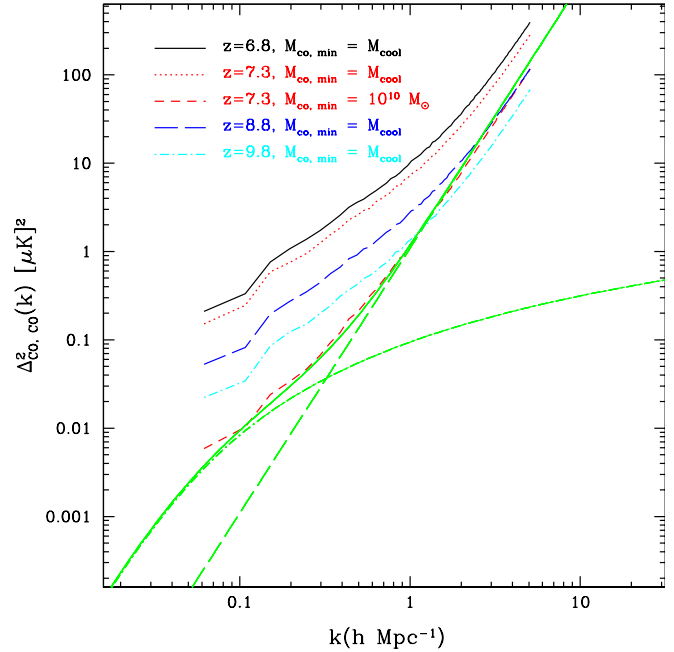
different atoms or molecules might be used in a related way. Each atomic species will reflect different properties of the host galaxies and combining maps may enable a detailed understanding of the host metallicity and internal properties to be achieved.

The model used for these calculations was based upon an empirical calibration of the CO luminosity function to observations of galaxies at $z \lesssim 3$ [179]. This was then extrapolated to higher redshifts as a best guess at how CO galaxies might evolve. More satisfying would be to connect the CO luminosity to the intrinsic properties of the galaxy interstellar medium (ISM). Such modelling of the detailed properties of CO emission from galaxies has been attempted [181]. CO emission in local galaxies is known to originate in giant molecular clouds (GMCs) whose typical size is ~10 pc. These clouds are observed to be optically thick to CO emission and so the empirically determined scaline of the CO luminosity with molecular hydrogen mass, $L_{CO} \propto M_{H_2}$, is best explained by the non-overlap of these clouds in angle and velocity. The relationship then results from the CO luminosity being proportional to the number of clouds. In addition, the detailed physics of the heating of these clouds by stars and AGNs, which determines the excitation state of the different CO rotational levels, must be accounted for. There is increased interest in numerical simulation of the chemistry and properties of GMCs [182], which can inform semi-analytic modelling. Connecting the global properties of CO emission to such ISM details is challenging, but would be greatly rewarding in providing a
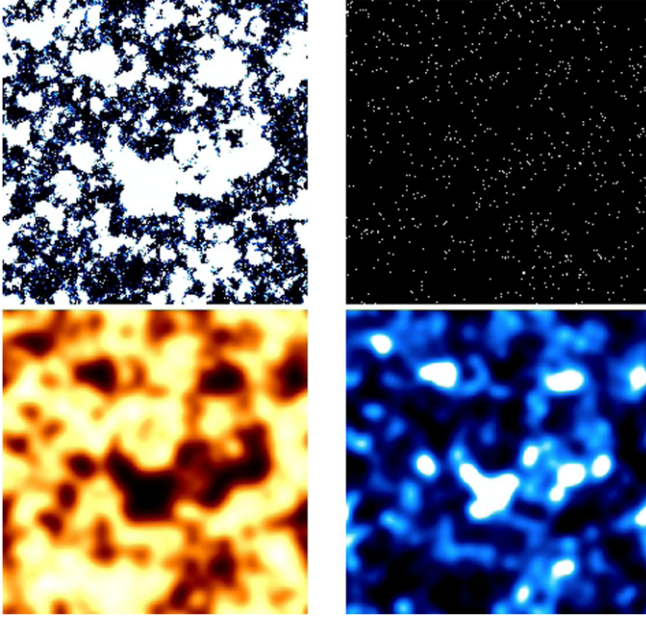
Figure 18. Redshift slices at $z \approx 7$ taken from the simulation of [179] and illustrating the galaxy distribution (top right), ionization field (top left), 21 cm fluctuations (bottom left) and CO intensity fluctuations (bottom right). These last two fields have been smoothed to an angular resolution of 6 arcmin. The CO bright galaxies seen in the CO intensity map are also responsible for ionizing the neutral gas. This can be seen by the correlation between bright CO regions and dark (ionized) patches in the 21 cm map. The associated signal provides a quantitative version of the cartoon in figure 12 [179].

new way of learning about the ISM of galaxies at high redshift. Importantly the chemistry that determines the CO abundance may change dramatically in the denser metal poor ISM of early galaxies [183].

Clearly, there is interesting information to be harvested from measurements of intensity fluctuations in line emission from galaxies at high redshifts. In the above discussion we have focused on preliminary results from the study of CO, but similar results apply to the array of other atomic and molecular lines. An important question is how the signal might best be observed and which lines are most important. Some lines, like CO, can be targeted from the ground but others, such as O I(63 μm), lie in the infrared and must be observed from space. The detailed design of the optimal instrument and the requirements for sensitivity and angular resolution are not yet well understood. Furthermore, the statistical tools for removing continuum and line contamination efficiently need to be developed.

A future space-borne infrared telescope such as SPICA could be ideal for IM. As illustrated in figure 15 [174], an optimized instrument on SPICA could measure the O I(63 μm) and O III(52 μm) line signals very accurately out to high redshifts. Such an instrument would require the capability to take many medium resolution spectra at adjacent locations on the sky. However, it may be possible to sacrifice angular resolution since the IM technique does not involve resolving individual sources.

6. 21 cm forest

While most interest in the redshifted 21 cm line has focused on the case where the CMB forms a backlight, an important alternative is the case where the 21 cm absorption is imprinted on the light from radio-loud quasars. Historically, the first attempts to detect the extra-galactic 21 cm line were of this nature [184]. The resulting 1D absorption spectra are expected to show a forest of lines originating from absorption in clumps of neutral hydrogen along the line of sight and have been called the 21 cm forest by analogy to the Lyα forest.

Whereas the Lyα forest is primarily visible at redshifts $z \lesssim 6$, when the Universe is mostly ionized, the 21 cm forest is strongest at higher redshifts where there is a considerable neutral fraction. Additionally, since the 21 cm line is optically thin and does not saturate, the 21 cm forest can contain detailed information about the properties of the IGM as well as collapsed gaseous halos. There is a clear complementarity to the two forests for probing the evolution of the IGM over a wide range of redshifts.

The 21 cm forest differs from the emission signal of the diffuse medium in a number of important ways. First, because the brightness temperature of radio-loud quasars is typically $\sim 10^{11-12}$ K the 21 cm line is always seen in absorption against the source. The strength of the absorption feature depends upon the 21 cm optical depth

$$\tau_{\nu_0}(z) = \frac{3}{32\pi} \frac{h_p c^3 A_{10}}{k_B \nu_0^2} \frac{x_{H\,I} n_H(z)}{T_S(1+z)(\mathrm{d}v_\parallel/\mathrm{d}r_\parallel)} \qquad (63)$$
$$\approx 0.009(1+\delta)(1+z)^{3/2}\frac{x_{H\,I}}{T_S}.$$

We see from this expression that the decrement depends primarily upon the neutral fraction and spin temperature. In contrast to the case where the CMB is the backlight, there is no saturation regime at large values of T_S. The decrement is maximized for a fully neutral and cold IGM—heating or ionizing the gas will reduce the observable signature.

This signature may show several distinct features: (i) a mean intensity decrement blueward of the 21 cm rest-frame frequency whose depth depends on the mean IGM optical depth at that redshift; (ii) small-scale variations in the intensity due to fluctuations in the density, neutral fraction and temperature of the IGM at different points along the line of sight; (iii) transmission windows due to photoionized bubbles along the line of sight; and (iv) deep absorption features arising from the dense neutral hydrogen clouds in dwarf galaxies and minihalos [185, 186]. Figure 19 shows an example of a 21 cm forest spectrum in which a number of these features can be seen.

Since the evolution of the optical depth depends on the mean neutral fraction and the spin temperature, we can understand the evolution of the 21 cm forest based on figure 4. At early times, the IGM is fully neutral and the evolution is dictated by the spin temperature. T_S tracks the gas kinetic temperature and rises from tens of Kelvin to thousands of Kelvin by the end of reionization, causing τ_{ν_0} to fall by several orders of magnitude due to heating alone. Then, as reionization takes place and the neutral fraction drops from $x_H = 1$ to the $x_H \sim 10^{-4}$ seen in the Lyα forest, the optical depth drops even
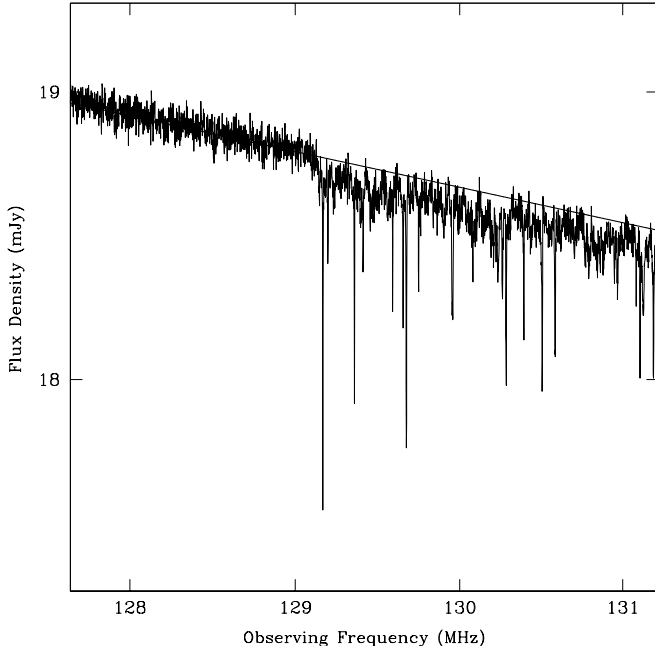
Figure 19. Simulated spectrum from 128 to 131 MHz of a source with brightness $S(120\,\mathrm{MHz}) = 20\,\mathrm{mJy}$ at $z = 10$ using a model spectrum based on that of Cygnus A and assuming H I 21 cm absorption by the IGM. Thermal noise has been added using the specifications of the SKA and assuming 10 days integration with 1 kHz wide spectral channels. The solid line is the model spectrum without noise or absorption. Reproduced with permission from [185]. Copyright 2002 American Astronomical Society.

further. Tracing the evolution of the mean optical depth would provide a useful constraint on the thermal evolution of the IGM and give a clear indication of the end of reionization. Figure 20 shows the evolution of τ_{ν_0} in a model where $T_{\mathrm{S}} = T_{\mathrm{K}}$ showing the very significant evolution in the optical depth with redshift.

Detecting the mean decrement in the 21 cm forest may be challenging since it is relatively weak and requires detailed fitting of the unabsorbed continuum level. A potentially more robust method is to exploit the statistics of individual features. As reionization occurs, the appearance of ionized bubbles will show up in the forest as an increasing number of windows of near total transparency. If these lines could be resolved, the distribution of equivalent widths would give a measure of the process of reionization [185, 186, 188].

It is possible that the forest does not have to be fully resolved and could be analysed statistically instead. For example, the appearance of ionized regions will lead to a change in the variance of the signal in different frequency bins. This has been shown to be an effective discriminant of the end of reionization [187] and has the advantage of not requiring narrow features to be resolved in frequency.

The signal from the diffuse IGM leads to a relatively low optical depth, as seen in (63). The signal from minihalos can be considerably stronger [122, 186, 189]. These are structures that have collapsed and virialized, but because of their low mass do not reach the temperature of 10^4 K required for atomic hydrogen cooling. Their high density contrast and relatively low temperature can lead to optical depths as high as $\tau \sim 0.1$ within a frequency width of $\Delta\nu \sim 2\,\mathrm{kHz}$ [190]. The mean
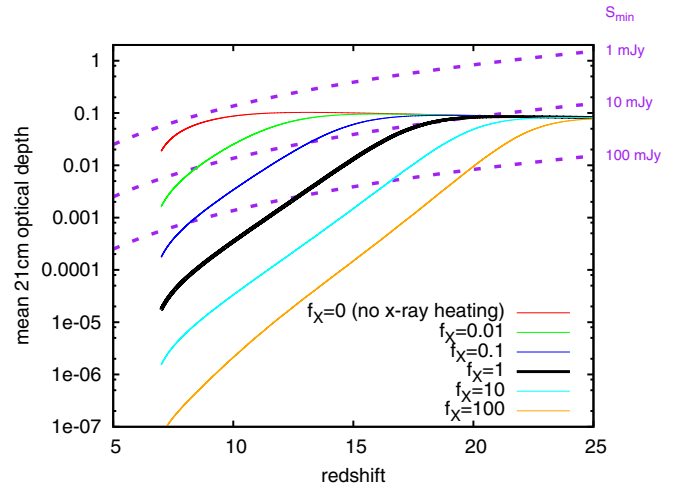

Figure 20. Mean 21 cm optical depth as a function of redshift, for varying f_{X} values: from top to bottom, $f_{\mathrm{X}} = 0$ (no x-ray heating), 0.01, 0.1, 1 (thick black line), 10 and 100. Also included are lines of S_{min} (dashed purple lines) as defined in (65). These lines indicate, for each value of τ on the left-hand axis, the minimum observed flux density of a source that would allow a detection at a signal-to-noise ratio of five of the flux decrement due to absorption, assuming an array with effective area $A_{\mathrm{eff}} = 10^6\,\mathrm{m}^2$, frequency resolution $\Delta\nu_{\mathrm{ch}} = 1\,\mathrm{kHz}$, system temperature $T_{\mathrm{sys}} = 100\,\mathrm{K}\,(\nu/200\,\mathrm{MHz})^{-2.8}$ and integration time $t_{\mathrm{int}} = 1$ week. When the mean optical depth lines (solid) cross above the S_{min} lines (dashed), the flux decrement is detectable at the corresponding redshift. Reproduced with permission from [187].

overdensity required to give an optical depth τ is given by

$$(1 + \delta) \sim 40 \left(\frac{\tau}{0.01}\right)^2 \left(\frac{T_{\mathrm{S}}}{100\,\mathrm{K}}\right)^2 \left(\frac{10}{1 + z}\right)^3. \quad (64)$$

The number density of minihalos is sensitive to the thermal properties of the IGM, which if heated may raise the Jeans mass sufficiently to prevent the collapse of the least massive minihalos that dominate the signal in the forest. Detailed descriptions of line profiles and their statistics have been considered based upon analytic models of the density distribution around halos [186, 189, 191, 192] and in high resolution simulations [125]. Even a single line of sight showing such structures could be used to provide information on early halo formation.

Unfortunately, the sensitivity required to detect the 21 cm forest is challenging. We can consider the thermal noise required to directly detect the flux decrement arising from the mean absorption. Using the relationship $S_\nu = 2k_{\mathrm{B}}T_{\mathrm{a}}/A_{\mathrm{e}}$ to relate the antennae temperature to the flux sensitivity for a single polarization and the radiometer equation gives the minimum source brightness for the decrement to be visible:

$$S_{\mathrm{min}} = 16\,\mathrm{mJy} \left[\frac{S/N}{5} \frac{0.01}{\tau} \frac{10^6\,\mathrm{m}^2}{A_{\mathrm{eff}}} \frac{T_{\mathrm{sys}}}{400\,\mathrm{K}}\right] \sqrt{\frac{1\,\mathrm{kHz}}{\Delta\nu} \frac{1\,\mathrm{week}}{t_{\mathrm{int}}}}, \quad (65)$$

where we have assumed that both polarization states are measured. The value of $A_{\mathrm{eff}}/T_{\mathrm{sys}}$ used is appropriate for SKA giving a sense of the challenge in detecting the forest. The

system temperature is typically set by the sky temperature and is determined by galactic foregrounds.

The major uncertainty in the utility of the 21 cm forest is the existence of radio-loud sources at high redshifts. The most promising set of sources are radio-loud quasars. We are currently ignorant of the number density of radio-loud quasars at high redshifts. Quasars have currently been observed to redshifts as high as $z = 7.1$ [193], but their number density drops rapidly at redshifts $z \gtrsim 3$ [194]. Furthermore, the radio-loud fraction of quasars is poorly understood and may evolve at higher redshifts [195]. Models calibrated to the number of counts at low redshifts lead to predictions of ~ 2000 sources across the sky with $S > 6$ Jy at $8 < z < 12$ [196]. Many of these sources would show up in existing radio surveys, but it is currently impossible to distinguish these high-z radio-bright quasars from low-z radio galaxies. Future large area surveys with near-infrared photometry, such as EUCLID or WFIRST could enable the identification of these high-z quasars providing targets for the SKA [197]. Gamma-ray bursts (GRBs) and hypernovae may provide alternative high-z sources [198–200], but are believed to have a lower maximum brightness at the low frequencies of interest because of synchrotron self-absorption. GRBs do however have the virtue of already having been observed in the desired redshift range.

This rapid decline in radio-bright quasars is unfortunate since the signal is most strong early on before ionization and heating have begun significantly. In any case, it seems likely that only a few sight lines to radio-bright objects will become available for 21 cm forest observations. Since these observations are affected by totally different systematics than 21 cm tomography, they could be very important for identifying the properties of the IGM in the early Universe.

7. Conclusions and outlook

Cosmology in the 21st century is engaged in the process of pushing the boundaries of sensitivity and completeness of our understanding of the Universe. Observations of the redshifted 21 cm line offer a new window into the properties of the Universe at redshifts $z = 1$–150 filling in a crucial gap in observations of the period where the first structures and stars had formed. The 21 cm signal depends upon physics on different scales from the atomic physics of hydrogen, through the astrophysics of star formation, to the physics of cosmological structures. Through this dependence the 21 cm signal has enormous potential to improve our understanding of the Universe. Over the next decade, radio experiments will begin the process of exploiting 21 cm observations and will reveal how much of this potential can be realized.

In this review, we have explored four key ways in which 21 cm observations might be exploited to learn about cosmology and astrophysics. Low angular resolution observations with small numbers of radio dipoles can measure the sky-averaged 'global' 21 cm signal providing information about key milestones in the history of the Universe. Understanding these moments when the first radiation backgrounds illuminated intergalactic space will

place constraints on the properties of the first galaxies and limit the contribution of more exotic physics. Experiments involving just a handful of people, such as EDGES and CoRE, have taken the first steps along this path. There is currently a growing effort to using different radio dipole designs and attempting to combine the benefits of interferometers with individual absolutely calibrated dipoles. There is a significant challenge in separating a relatively smooth 21 cm signal from smooth galactic foregrounds and much hard work ahead before these experiments yield robust results.

More ambitiously, collections of thousands of dipoles, digitally combined into radio interferometers will have the raw sensitivity to measure fluctuations in the 21 cm signal, providing detailed information about how the Universe was heated and ionized. The first generation of such instruments—LOFAR, MWA, PAPER and 21CMA—have or will soon be completed and data are starting to become available. These first instruments will peer through a hazy window of low sensitivity, but should determine when the EoR took place. Looking further ahead, the SKA will see this period clearly mapping the ionized structures and providing information about heating and coupling of the IGM during the formation of the first galaxies. The combination of 21 cm observations with observations of the CMB and with surveys of high redshift galaxies and information from the Lyman alpha forest should enable the reionization history and details of reionization to be inferred [201, 202].

21 cm observations also have the potential to constrain fundamental cosmology and represent a path to a new level of precision cosmology. Exploiting 21 cm observations for fundamental cosmology will be a challenging task and it is unclear how much can realistically be hoped for. During the EoR astrophysics mixes with cosmology in the 21 cm fluctuations and until the performance of instruments is better understood it is not clear whether these two can be separated. In the further future, 21 cm observations of the cosmic Dark Ages provide a long term hope that most of the volume of the Universe could one day be mapped and used for cosmology.

Applying similar techniques at lower redshifts allows for IM of the neutral hydrogen in galaxies, providing an alternative to traditional redshift surveys of galaxies in the optical–infrared bands. In contrast to the traditional approach, only the cumulative line emission from many galaxies over large volumes is being mapped, and galaxies are not being resolved individually. This IM technique could help shed light on the properties of dark energy through the use of BAOs as a standard ruler to measure the expansion of the Universe. Further, IM in other molecular and atomic lines can complement 21 cm tomography allowing a complete picture of the formation of stars and metals to be developed.

Finally, observations of individual radio-bright sources, such as quasars, might allow observations of the 21 cm forest. This provides another window into the properties of the IGM at the end of reionization that depends on very different systematics than 21 cm tomography. The 21 cm forest would allow the small-scale properties of the IGM to be studied in great detail and so constrain the properties of dark matter. This is a powerful technique and the main uncertainty is the

abundance of target sources. Large area near-infrared surveys should begin to answer this question soon, as more high redshift quasars are found and followed up to determine the radio-loud fraction. Just a handful of sufficiently bright radio sources at $z \gtrsim 7$ would make this a very valuable probe of reionization and the IGM.

21 cm cosmology is a new field with much yet to be discovered, but it is one with great potential for the future. Much of this potential stems from breakthroughs in computing power which make 'digital radio astronomy' feasible and allow for extremely large radio interferometers. It is to be hoped that in the decades to come 21 cm observations will transform our understanding of the cosmic dawn and the EoR pushing farther our detailed understanding of the cosmos.

Acknowledgments

References

[1] Santos M G, Amblard A, Pritchard J, Trac H, Cen R and Cooray A 2008 Cosmic reionization and the 21 cm signal: comparison between an analytical model and a simulation *Astrophys. J.* **689** 1–16

[2] Pritchard J and Loeb A 2010 Cosmology: hydrogen was not ionized abruptly *Nature* **468** 772–3

[3] Komatsu E *et al* 2011 Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: cosmological interpretation *Astrophys. J. Suppl.* **192** 18

[4] Furlanetto S R, Oh S P and Briggs F H 2006 Cosmology at low frequencies: the 21 cm transition and the high-redshift Universe *Phys. Rep.* **433** 181–301

[5] Morales M F and Wyithe J S B 2010 Reionization and cosmology with 21-cm fluctuations *Annu. Rev. Astron. Astrophys.* **48** 127–71

[6] Goldenberg H M, Kleppner D and Ramsey N F 1960 Atomic hydrogen maser *Phys. Rev. Lett.* **5** 361–2

[7] van de Hulst H C 1945 Radiogolven uit het Wereldruim (Radio waves from space) *Ned. Tijdschr. Natuurkd.* **11** 210–21

[8] Ewen H I and Purcell E M 1951 Observation of a line in the galactic radio spectrum: radiation from galactic hydrogen at 1,420 Mc./sec *Nature* **168** 356

[9] Kanekar N, Chengalur J N and Lane W M 2007 H I 21-cm absorption at $z \sim 3.39$ towards PKS 0201+113 *Mon. Not. R. Astron. Soc.* **375** 1528–36

[10] Srianand R, Gupta N, Petitjean P, Noterdaeme P and Ledoux C 2010 Detection of 21-cm, H_2 and deuterium absorption at $z > 3$ along the line of sight to J1337+3152 *Mon. Not. R. Astron. Soc.* **405** 1888–900

[11] Bagla J S and Loeb A 2009 The hyperfine transition of ^{3}He$^+$ as a probe of the intergalactic medium arXiv:0905.1698

[12] McQuinn M and Switzer E R 2009 Redshifted intergalactic ^{3}He$^+$ 8.7 GHz hyperfine absorption *Phys. Rev.* D **80** 063010

[13] Sigurdson K and Furlanetto S R 2006 Measuring the primordial deuterium abundance during the cosmic dark ages *Phys. Rev. Lett.* **97** 091301

[14] Rybicki G B and Lightman A P 1986 *Radiative Processes in Astrophysics* (New York: Wiley)

[15] Sobolev V V 1957 The diffusion of Lα radiation in nebulae and stellar envelopes *Sov. Astron.* **1** 678

[16] Field G B 1958 Excitation of the hydrogen 21 cm line *Proc. IRE* **46** 240

[17] Allison A C and Dalgarno A 1969 Spin change in collisions of hydrogen atoms *Astrophys. J.* **158** 423

[18] Zygelman B 2005 Hyperfine level-changing collisions of hydrogen atoms and tomography of the dark age Universe *Astrophys. J.* **622** 1356–62

[19] Furlanetto S R and Furlanetto M R 2007 Spin-exchange rates in electron–hydrogen collisions *Mon. Not. R. Astron. Soc.* **374** 547–55

[20] Furlanetto S R and Furlanetto M R 2007 Spin exchange rates in proton–hydrogen collisions *Mon. Not. R. Astron. Soc.* **379** 130–4

[21] Kuhlen M, Madau P and Montgomery R 2006 The spin temperature and 21 cm brightness of the intergalactic medium in the pre-reionization era *Astrophys. J.* **637** L1–L4

[22] Liszt H 2001 The spin temperature of warm interstellar H I *Astron. Astrophys.* **371** 698–707

[23] Hirata C M and Sigurdson K 2007 The spin-resolved atomic velocity distribution and 21-cm line profile of dark-age gas *Mon. Not. R. Astron. Soc.* **375** 1241–64

[24] Wouthuysen S A 1952 On the excitation mechanism of the 21-cm (radio-frequency) interstellar hydrogen emission line *Astrophys. J.* **57** 31

[25] Pritchard J R and Furlanetto S R 2006 Descending from on high: Lyman-series cascades and spin-kinetic temperature coupling in the 21-cm line *Mon. Not. R. Astron. Soc.* **367** 1057–66

[26] Meiksin A 2000 Detecting the epoch of first light in 21-CM radiation *Perspectives on Radio Astronomy: Science with Large Antennae Arrays* ed M P van Harlem (Dwingeloo: ASTRON) p 37

[27] Rybicki G B 2006 Improved Fokker–Planck equation for resonance-line scattering *Astrophys. J.* **647** 709–18

[28] Field G B 1959 The time relaxation of a resonance-line profile *Astrophys. J.* **129** 551

[29] Chen X and Miralda-Escudé J 2004 The spin-kinetic temperature coupling and the heating rate due to Lyα scattering before reionization: predictions for 21 centimeter emission and absorption *Astrophys. J.* **602** 1–11

[30] Hirata C M 2006 Wouthuysen-field coupling strength and application to high-redshift 21-cm radiation *Mon. Not. R. Astron. Soc.* **367** 259–74

[31] Chuzhoy L and Shapiro P R 2007 Heating and cooling of the early intergalactic medium by resonance photons *Astrophys. J.* **655** 843–6

[32] Furlanetto S R and Pritchard J R 2006 The scattering of Lyman-series photons in the intergalactic medium *Mon. Not. R. Astron. Soc.* **372** 1093–1103

[33] Loeb A and Zaldarriaga M 2004 Measuring the small-scale power spectrum of cosmic density fluctuations through 21 cm tomography prior to the epoch of structure formation *Phys. Rev. Lett.* **92** 211301

[34] Barkana R and Loeb A 2005 Detecting the earliest galaxies through two new sources of 21 centimeter fluctuations *Astrophys. J.* **626** 1–11

[35] Chen X and Miralda-Escude J 2008 The 21 cm signature of the first stars *Astrophys. J.* **684** 18–33

[36] Santos M G and Cooray A 2006 Cosmological and astrophysical parameter measurements with 21-cm anisotropies during the era of reionization *Phys. Rev.* D **74** 083517

[37] Furlanetto S R, Zaldarriaga M and Hernquist L 2004 The growth of H II regions during reionization *Astrophys. J.* **613** 1–15

[38] Nusser A 2005 The spin temperature of neutral hydrogen
     during cosmic pre-reionization *Mon. Not. R. Astron. Soc.*
     359 183–90

[39] Kuhlen M and Madau P 2005 The first miniquasar *Mon. Not.
     R. Astron. Soc.* **363** 1069–82

[40] Furlanetto S R and Loeb A 2004 Large-scale structure shocks
     at low and high redshifts *Astrophys. J.* **611** 642–54

[41] Geil P M and Wyithe J S B 2009 Modification of the 21-cm
     power spectrum by quasars during the epoch of
     reionization *Mon. Not. R. Astron. Soc.* **399** 1877–87

[42] Seager S, Sasselov D D and Scott D 1999 A new calculation
     of the recombination epoch *Astrophys. J.* **523** L1–5

[43] Miralda-Escudé J, Haehnelt M and Rees M J 2000
     Reionization of the inhomogeneous Universe *Astrophys. J.*
     530 1–16

[44] Bromm V, Kudritzki R P and Loeb A 2001 Generic spectrum
     and ionization efficiency of a heavy initial mass function
     for the first stars *Astrophys. J.* **552** 464–72

[45] Zackrisson E, Rydberg C-E, Schaerer D, Östlin G and Tuli M
     2011 The spectral evolution of the first galaxies: I. James
     Webb Space Telescope detection limits and color criteria
     for Population III galaxies *Astrophys. J.* **740** 13

[46] Barkana R and Loeb A 2001 In the beginning: the first
     sources of light and the reionization of the Universe *Phys.
     Rep.* **349** 125–238

[47] Sheth R K and Tormen G 1999 Large-scale bias and the peak
     background split *Mon. Not. R. Astron. Soc.* **308** 119–26

[48] Pawlik A H, Schaye J and van Scherpenzeel E 2009 Keeping
     the Universe ionized: photoheating and the clumping
     factor of the high-redshift intergalactic medium *Mon. Not.
     R. Astron. Soc.* **394** 1812–24

[49] Bolton J S and Becker G D 2009 Resolving the high redshift
     Lyα forest in smoothed particle hydrodynamics
     simulations *Mon. Not. R. Astron. Soc.* **398** L26–L30

[50] Furlanetto S R and Oh S P 2005 Taxing the rich:
     recombinations and bubble growth during reionization
     Mon. Not. R. Astron. Soc. **363** 1031–48

[51] Naoz S and Barkana R 2005 Growth of linear perturbations
     before the era of the first galaxies *Mon. Not. R. Astron.
     Soc.* **362** 1047–53

[52] Madau P, Meiksin A and Rees M J 1997 21 centimeter
     tomography of the intergalactic medium at high redshift
     Astrophys. J. **475** 429

[53] Venkatesan A, Giroux M L and Shull J M 2001 Heating and
     ionization of the intergalactic medium by an early x-ray
     background *Astrophys. J.* **563** 1–8

[54] Pritchard J R and Furlanetto S R 2007 21-cm fluctuations
     from inhomogeneous x-ray heating before reionization
     Mon. Not. R. Astron. Soc. **376** 1680–94

[55] Zaroubi S, Thomas R M, Sugiyama N and Silk J
     2007 Heating of the intergalactic medium by
     primordial miniquasars *Mon. Not. R. Astron. Soc.*
     375 1269–79

[56] Ciardi B, Salvaterra R and Di Matteo T 2010 Lyα versus
     x-ray heating in the high-*z* intergalactic medium *Mon. Not.
     R. Astron. Soc.* **401** 2635–40

[57] Shull J M and van Steenberg M E 1985 X-ray secondary
     heating and ionization in quasar emission-line clouds
     Astrophys. J. **298** 268–74

[58] Valdés M and Ferrara A 2008 The energy cascade from
     warm dark matter decays *Mon. Not. R. Astron. Soc.*
     387 L8–L12

[59] Furlanetto S R and Stoever S J 2010 Secondary ionization
     and heating by fast electrons *Mon. Not. R. Astron. Soc.*
     404 1869–78

[60] Verner D A, Ferland G J, Korista K T and Yakovlev D G
     1996 Atomic data for astrophysics: II. New analytic FITS
     for photoionization cross sections of atoms and ions
     Astrophys. J. **465** 487

[61] Oh S P 2001 Reionization by hard photons: I. X-rays from
     the first star clusters *Astrophys. J.* **553** 499–512

[62] Glover S C O and Brand P W J L 2003 Radiative feedback
     from an early x-ray background *Mon. Not. R. Astron. Soc.*
     340 210–26

[63] Furlanetto S R 2006 The global 21-centimeter background
     from high redshifts *Mon. Not. R. Astron. Soc.* **371** 867–78

[64] Grimm H-J, Gilfanov M and Sunyaev R 2003 High-mass
     x-ray binaries as a star formation rate indicator
     in distant galaxies *Mon. Not. R. Astron. Soc.* **339** 793–809

[65] Gilfanov M, Grimm H-J and Sunyaev R 2004 L_X-SFR
     relation in star-forming galaxies *Mon. Not. R. Astron. Soc.*
     347 L57–L60

[66] Lehmer B D, Alexander D M, Bauer F E, Brandt W N,
     Goulding A D, Jenkins L P, Ptak A and Roberts T P 2010
     A Chandra perspective on galaxy-wide x-ray binary
     emission and its correlation with star formation rate and
     stellar mass: new results from luminous infrared galaxies
     Astrophys. J. **724** 559–71

[67] Mineo S, Gilfanov M and Sunyaev R 2011 The collective
     x-ray luminosity of HMXB as a SFR indicator *Astron.
     Nachr.* **332** 349

[68] Dijkstra M, Gilfanov M, Loeb A and Sunyaev R 2012
     Constraints on the redshift evolution of the L X-SFR
     relation from the cosmic x-ray backgrounds *Mon. Not. R.
     Astron. Soc.* **421** 213–23

[69] Mirabel I F, Dijkstra M, Laurent P, Loeb A and Pritchard J R
     2011 Stellar black holes at the dawn of the Universe
     Astron. Astrophys. **528** A149

[70] Madau P, Rees M J, Volonteri M, Haardt F and Oh S P 2004
     Early reionization by miniquasars *Astrophys. J.*
     604 484–94

[71] Ricotti M and Ostriker J P 2004 X-ray pre-ionization
     powered by accretion on the first black holes: I. A model
     for the WMAP polarization measurement *Mon. Not. R.
     Astron. Soc.* **352** 547–62

[72] Dijkstra M, Haiman Z and Loeb A 2004 A limit from the
     x-ray background on the contribution of quasars to
     reionization *Astrophys. J.* **613** 646–54

[73] Chuzhoy L, Alvarez M A and Shapiro P R 2006 Recognizing
     the first radiation sources through their 21 cm signature
     Astrophys. J. **648** L1–L4

[74] Leitherer C, Schaerer D, Goldader J D, González Delgado
     R M, Robert C, Kune D F, de Mello D F, Devost D and
     Heckman T M 1999 Starburst99: synthesis models for
     galaxies with active star formation *Astrophys. J. Suppl.*
     123 3–40

[75] Shull J M 1979 Heating and ionization by x-ray
     photoelectrons *Astrophys. J.* **234** 761–4

[76] Clark P C, Glover S C O, Klessen R S and Bromm V 2011
     Gravitational fragmentation in turbulent primordial gas
     and the initial mass function of Population III stars
     Astrophys. J. **727** 110

[77] Caffau E *et al* 2011 An extremely primitive star in the
     galactic halo *Nature* **477** 67–9

[78] Bromm V and Loeb A 2003 The formation of the first
     low-mass stars from gas with low carbon and oxygen
     abundances *Nature* **425** 812–4

[79] Pritchard J R and Loeb A 2008 Evolution of the 21 cm
     signal throughout cosmic history *Phys. Rev. D*
     78 103511

[80] Chen X and Kamionkowski M 2004 Particle decays during
     the cosmic dark ages *Phys. Rev. D* **70** 043502

[81] Furlanetto S R, Oh S P and Pierpaoli E 2006 Effects of dark
     matter decay and annihilation on the high-redshift 21 cm
     background *Phys. Rev. D* **74** 103502

[82] Valdés M, Ferrara A, Mapelli M and Ripamonti E 2007
     Constraining dark matter through 21-cm observations
     Mon. Not. R. Astron. Soc. **377** 245–52

[83] Belikov A V and Hooper D 2009 How dark matter reionized the Universe *Phys. Rev.* D **80** 035007

[84] Slatyer T R, Padmanabhan N and Finkbeiner D P 2009 CMB constraints on WIMP annihilation: energy absorption during the recombination epoch *Phys. Rev.* D **80** 043526

[85] Cumberbatch D T, Lattanzi M and Silk J 2010 Signatures of clumpy dark matter in the global 21 cm background signal *Phys. Rev.* D **82** 103508

[86] Finkbeiner D P, Padmanabhan N and Weiner N 2008 CMB and 21-cm signals for dark matter with a long-lived excited state *Phys. Rev.* D **78** 063530

[87] Mack K J and Wesley D H 2008 Primordial black holes in the dark ages: observational prospects for future 21 cm surveys arXiv:0805.1513

[88] Ricotti M, Ostriker J P and Mack K J 2008 Effect of primordial black holes on the cosmic microwave background and cosmological parameter estimates *Astrophys. J.* **680** 829–45

[89] Brandenberger R H, Danos R J, Hernández O F and Holder G P 2010 The 21 cm signature of cosmic string wakes *J. Cosmol. Astropart. Phys.* JCAP12(2010)028

[90] Chippendale A P, Subrahmanyan E and Ekers R 2005 The cosmological reionization experiment *New Techniques and Results in Low Frequency Radio Astronomy* (*Hobart, Australia, 6–10 Dec. 2005*)

[91] Bowman J D, Rogers A E E and Hewitt J N 2008 Toward empirical constraints on the global redshifted 21 cm brightness temperature during the epoch of reionization *Astrophys. J.* **676** 1–9

[92] Bowman J D and Rogers A E E 2010 A lower limit of $\Delta z \gtrsim 0.06$ for the duration of the reionization epoch *Nature* **468** 796–8

[93] Subrahmanyan R 2011 private communication

[94] Tingay S 2011 private communication

[95] Greenhill L 2011 private communication

[96] Eisenstein D J, Hu W and Tegmark M 1999 Cosmic complementarity: joint parameter estimation from cosmic microwave background experiments and redshift surveys *Astrophys. J.* **518** 2–23

[97] Pritchard J R and Loeb A 2010 Constraining the unexplored period between the dark ages and reionization with observations of the global 21-cm signal *Phys. Rev.* D **82** 023006

[98] Morandi A and Barkana R 2011 Studying cosmic reionization with observations of the global 21 cm signal arXiv:1102.2378

[99] Burns J O, Lazio T J W, Bale S D, Bowman J D, Bradley R F, Carilli C L, Furlanetto S R, Harker G J A, Loeb A and Pritchard J R 2012 Probing the first stars and black holes in the early Universe with the Dark Ages Radio Explorer (DARE) *Adv. Space Res.* **49** 433–50

[100] Harker G J A, Pritchard J R, Burns J O and Bowman J D 2012 An MCMC approach to extracting the global 21-cm signal during the cosmic dawn from sky-averaged radio observations *Mon. Not. R. Astron. Soc.* **419** 1070–84

[101] Morales M F and Hewitt J 2004 Toward epoch of reionization measurements with wide-field radio observations *Astrophys. J.* **615** 7–18

[102] Bharadwaj S and Ali S S 2004 The cosmic microwave background radiation fluctuations from H I perturbations prior to reionization *Mon. Not. R. Astron. Soc.* **352** 142–6

[103] McQuinn M, Zahn O, Zaldarriaga M, Hernquist L and Furlanetto S R 2006 Cosmological parameter estimation using 21 cm radiation from the epoch of reionization *Astrophys. J.* **653** 815–34

[104] Barkana R and Loeb A 2005 A method for separating the physics from the astrophysics of high-redshift 21 centimeter fluctuations *Astrophys. J.* **624** L65–8

[105] Wang X and Hu W 2006 Redshift space 21 cm power spectra from reionization *Astrophys. J.* **643** 585–97

[106] Mellema G, Iliev I T, Pen U-L and Shapiro P R 2006 Simulating cosmic reionization at large scales: II. The 21-cm emission features and statistical signals *Mon. Not. R. Astron. Soc.* **372** 679–92

[107] Shaw J R and Lewis A 2008 Nonlinear redshift-space power spectra *Phys. Rev.* D **78** 103512

[108] Mao Y, Shapiro P R, Mellema G, Iliev I T, Koda J and Ahn K 2011 Redshift space distortion of the 21 cm background from the epoch of reionization: I. Methodology re-examined arXiv:1104.2094

[109] Barkana R and Loeb A 2004 Unusually large fluctuations in the statistics of galaxy formation at high redshift *Astrophys. J.* **609** 474–81

[110] Press W H and Schechter P 1974 Formation of galaxies and clusters of galaxies by self-similar gravitational condensation *Astrophys. J.* **187** 425–38

[111] Bond J R, Cole S, Efstathiou G and Kaiser N 1991 Excursion set mass functions for hierarchical Gaussian fluctuations *Astrophys. J.* **379** 440–60

[112] Zahn O, Lidz A, McQuinn M, Dutta S, Hernquist L, Zaldarriaga M and Furlanetto S R 2007 Simulations and analytic calculations of bubble growth during hydrogen reionization *Astrophys. J.* **654** 12–26

[113] Zahn O, Mesinger A, McQuinn M, Trac H, Cen R and Hernquist L E 2011 Comparison of reionization models: radiative transfer simulations and approximate, seminumeric models *Mon. Not. R. Astron. Soc.* **414** 727–38

[114] Scannapieco E and Barkana R 2002 An analytical approach to inhomogeneous structure formation *Astrophys. J.* **571** 585–603

[115] Barkana R 2007 On correlated random walks and 21-cm fluctuations during cosmic reionization *Mon. Not. R. Astron. Soc.* **376** 1784–92

[116] Santos M G, Silva M B, Pritchard J R, Cen R and Cooray A 2011 Probing the first galaxies with the Square Kilometer array *Astron. Astrophys.* **527** A93

[117] Chuzhoy L and Zheng Z 2007 Radiative transfer effect on ultraviolet pumping of the 21 cm line in the high-redshift Universe *Astrophys. J.* **670** 912–18

[118] Semelin B, Combes F and Baek S 2007 Lyman-alpha radiative transfer during the epoch of reionization: contribution to 21-cm signal fluctuations *Astron. Astrophys.* **474** 365–74

[119] Naoz S and Barkana R 2008 Detecting early galaxies through their 21-cm signature *Mon. Not. R. Astron. Soc.* **385** L63–L67

[120] Alvarez M A, Pen U-L and Chang T-C 2010 Enhanced detectability of pre-reionization 21 cm structure *Astrophys. J.* **723** L17–L21

[121] Barkana R and Loeb A 2005 Probing the epoch of early baryonic infall through 21-cm fluctuations *Mon. Not. R. Astron. Soc.* **363** L36–L40

[122] Iliev I T, Shapiro P R, Ferrara A and Martel H 2002 On the direct detectability of the cosmic dark ages: 21 centimeter emission from minihalos *Astrophys. J.* **572** L123–6

[123] Iliev I T, Scannapieco E, Martel H and Shapiro P R 2003 Non-linear clustering during the cosmic dark ages and its effect on the 21-cm background from minihaloes *Mon. Not. R. Astron. Soc.* **341** 81–90

[124] Furlanetto S R and Oh S P 2006 Redshifted 21 cm emission from minihalos before reionization *Astrophys. J.* **652** 849–56

[125] Shapiro P R, Ahn K, Alvarez M A, Iliev I T, Martel H and Ryu D 2006 The 21 cm background from the cosmic dark ages: minihalos and the intergalactic medium before reionization *Astrophys. J.* **646** 681–90

[126] Tseliakhovich D and Hirata C 2010 Relative velocity of dark matter and baryonic fluids and the formation of the first structures *Phys. Rev.* D **82** 083520

[127] Tseliakhovich D, Barkana R and Hirata C M 2011 Suppression and spatial variation of early galaxies and minihaloes *Mon. Not. R. Astron. Soc.* **418** 906–15

[128] Dalal N, Pen U-L and Seljak U 2010 Large-scale BAO signatures of the smallest galaxies *J. Cosmol. Astropart. Phys.* JCAP11(2010)007

[129] Stacy A, Bromm V and Loeb A 2011 Effect of streaming motion of baryons relative to dark matter on the formation of the first stars *Astrophys. J.* **730** L1

[130] Maio U, Koopmans L V E and Ciardi B 2011 The impact of primordial supersonic flows on early structure formation, reionization and the lowest-mass dwarf galaxies *Mon. Not. R. Astron. Soc.* **412** L40–L44

[131] Babich D and Loeb A 2006 Imprint of inhomogeneous reionization on the power spectrum of galaxy surveys at high redshifts *Astrophys. J.* **640** 1–7

[132] Pritchard J R, Furlanetto S R and Kamionkowski M 2007 Galaxy surveys, inhomogeneous re-ionization and dark energy *Mon. Not. R. Astron. Soc.* **374** 159–67

[133] Wyithe J S B and Loeb A 2007 The imprint of cosmic reionization on galaxy clustering *Mon. Not. R. Astron. Soc.* **382** 921–36

[134] Dopita M A, Krauss L M, Sutherland R S, Kobayashi C and Lineweaver C H 2011 Re-ionizing the Universe without stars *Astrophys. Space. Sci.* **335** 345–52

[135] Trac H and Gnedin N Y 2011 Computer simulations of cosmic reionization *Adv. Sci. Lett.* **4** 228–43

[136] McQuinn M, Lidz A, Zahn O, Dutta S, Hernquist L and Zaldarriaga M 2007 The morphology of H II regions during reionization *Mon. Not. R. Astron. Soc.* **377** 1043–63

[137] Mesinger A, Furlanetto S and Cen R 2011 21CMFAST: a fast, seminumerical simulation of the high-redshift 21-cm signal *Mon. Not. R. Astron. Soc.* **411** 955–72

[138] Santos M G, Ferramacho L, Silva M B, Amblard A and Cooray A 2011 Fast large volume simulations of the 21-cm signal from the reionization and pre-reionization epochs *Mon. Not. R. Astron. Soc.* **406** 2421–32

[139] Thomas R M *et al* 2009 Fast large-scale reionization simulations *Mon. Not. R. Astron. Soc.* **393** 32–48

[140] Trac H and Cen R 2007 Radiative transfer simulations of cosmic reionization: I. Methodology and initial results *Astrophys. J.* **671** 1–13

[141] Partl A M, Maselli A, Ciardi B, Ferrara A and Müller V 2011 Enabling parallel computing in CRASH *Mon. Not. R. Astron. Soc.* **414** 428–44

[142] Trac H, Cen R and Loeb A 2008 Imprint of inhomogeneous hydrogen reionization on the temperature distribution of the intergalactic medium *Astrophys. J.* **689** L81–L84

[143] Baek S, Di Matteo P, Semelin B, Combes F and Revaz Y 2009 The simulated 21 cm signal during the epoch of reionization: full modeling of the Ly-α pumping *Astron. Astrophys.* **495** 389–405

[144] Baek S, Semelin B, Di Matteo P, Revaz Y and Combes F 2010 Reionization by UV or x-ray sources *Astron. Astrophys.* **523** A4

[145] Lidz A, Zahn O, McQuinn M, Zaldarriaga M and Hernquist L 2008 Detecting the rise and fall of 21 cm fluctuations with the Murchison Widefield Array *Astrophys. J.* **680** 962–74

[146] McQuinn M, Zahn O, Zaldarriaga M, Hernquist L and Furlanetto S R 2006 Cosmological parameter estimation using 21 cm radiation from the epoch of reionization *Astrophys. J.* **653** 815–34

[147] Maldacena J 2003 Non-Gaussian features of primordial fluctuations in single field inflationary models *J. High Energy Phys.* JHEP05(2003)013

[148] Komatsu E *et al* 2009 Five-year Wilkinson Microwave Anisotropy Probe observations: cosmological interpretation *Astrophys. J. Suppl.* **180** 330–76

[149] Cooray A 2006 21-cm background anisotropies can discern primordial non-Gaussianity *Phys. Rev. Lett.* **97** 261301

[150] Pillepich A, Porciani C and Matarrese S 2007 The bispectrum of redshifted 21 centimeter fluctuations from the dark ages *Astrophys. J.* **662** 1–14

[151] Harker G J A *et al* 2009 Detection and extraction of signals from the epoch of reionization using higher-order one-point statistics *Mon. Not. R. Astron. Soc.* **393** 1449–58

[152] Friedrich M M, Mellema G, Alvarez M A, Shapiro P R and Iliev I T 2011 Topology and sizes of H II regions during cosmic reionization *Mon. Not. R. Astron. Soc.* **413** 1353–72

[153] Lee K-G and Spergel D N 2011 Threshold probability functions and thermal inhomogeneities in the Lyα forest *Astrophys. J.* **734** 21

[154] Joudaki S, Doré O, Ferramacho L, Kaplinghat M and Santos M G 2011 Primordial non-Gaussianity from the 21 cm power spectrum during the epoch of reionization *Phys. Rev. Lett.* **107** 131304

[155] Mao Y, Tegmark M, McQuinn M, Zaldarriaga M and Zahn O 2008 How accurately can 21 cm tomography constrain cosmology? *Phys. Rev.* D **78** 023529

[156] Barger V, Gao Y, Mao Y and Marfatia D 2009 Inflationary potential from 21 cm tomography and Planck *Phys. Lett.* B **673** 173–8

[157] Adshead P, Easther R, Pritchard J and Loeb A 2011 Inflation and the scale dependent spectral index: prospects and strategies *J. Cosmol. Astropart. Phys.* JCAP02(2011)021

[158] Pritchard J R and Pierpaoli E 2008 Constraining massive neutrinos using cosmological 21 cm observations *Phys. Rev.* D **78** 065009

[159] Gordon C and Pritchard J R 2009 Forecasted 21 cm constraints on compensated isocurvature perturbations *Phys. Rev.* D **80** 063535

[160] Jester S and Falcke H 2009 Science with a lunar low-frequency array: from the dark ages of the Universe to nearby exoplanets *New Astron. Rev.* **53** 1–26

[161] Khatri R and Wandelt B D 2007 21-cm radiation: a new probe of variation in the fine-structure constant *Phys. Rev. Lett.* **98** 111301

[162] Khatri R and Wandelt B D 2008 Cosmic (Super)string constraints from 21 cm radiation *Phys. Rev. Lett.* **100** 091302

[163] Loeb A and Wyithe J S B 2008 Possibility of precise measurement of the cosmological power spectrum with a dedicated survey of 21 cm emission after reionization *Phys. Rev. Lett.* **100** 161301

[164] Wyithe J S B and Loeb A 2008 Fluctuations in 21-cm emission after reionization *Mon. Not. R. Astron. Soc.* **383** 606–14

[165] Chang T-C, Pen U-L, Peterson J B and McDonald P 2008 Baryon acoustic oscillation intensity mapping of dark energy *Phys. Rev. Lett.* **100** 091303

[166] Battye R A, Davies R D and Weller J 2004 Neutral hydrogen surveys for high-redshift galaxy clusters and protoclusters *Mon. Not. R. Astron. Soc.* **355** 1339–47

[167] Visbal E, Loeb A and Wyithe S 2009 Cosmological constraints from 21 cm surveys after reionization *J. Cosmol. Astropart. Phys.* JCAP10(2009)030

[168] Albrecht A *et al* 2006 Report of the Dark Energy Task Force arXiv:astro-ph/0609591

[169] Wyithe J S B, Loeb A and Geil P M 2008 Baryonic acoustic oscillations in 21-cm emission: a probe of dark energy out to high redshifts *Mon. Not. R. Astron. Soc.* **383** 1195–209

[170] Wyithe J S B and Loeb A 2009 The 21-cm power spectrum after reionization *Mon. Not. R. Astron. Soc.* **397** 1926–34

[171] Chang T-C, Pen U-L, Bandura K and Peterson J B 2010 An intensity map of hydrogen 21-cm emission at redshift z ≈ 0.8 *Nature* **466** 463–5

[172] Barkana R and Loeb A 2000 Identifying the reionization redshift from the cosmic star formation rate *Astrophys. J.* **539** 20–5

[173] Righi M, Hernández-Monteagudo C and Sunyaev R A 2008 Carbon monoxide line emission as a CMB foreground: tomography of the star-forming Universe with different spectral resolutions *Astron. Astrophys.* **489** 489–504

[174] Visbal E, Trac H and Loeb A 2011 Demonstrating the feasibility of line intensity mapping using mock data of galaxy clustering from simulations *J. Cosmol. Astropart. Phys.* JCAP08(2011)010

[175] Visbal E and Loeb A 2010 Measuring the 3D clustering of undetected galaxies through cross correlation of their cumulative flux fluctuations from multiple spectral lines *J. Cosmol. Astropart. Phys.* JCAP11(2010)016

[176] Wang X, Tegmark M, Santos M G and Knox L 2006 21 cm tomography with foregrounds *Astrophys. J.* **650** 529–37

[177] Liu A and Tegmark M 2011 A method for 21 cm power spectrum estimation in the presence of foregrounds *Phys. Rev.* D **83** 103006

[178] Carilli C L 2011 Intensity mapping of molecular gas during cosmic reionization *Astrophys. J.* **730** L30

[179] Lidz A, Furlanetto S R, Oh S P, Aguirre J, Chang T-C, Doré O and Pritchard J R 2011 Intensity mapping with carbon monoxide emission lines and the redshifted 21 cm line *Astrophys. J.* **741** 70

[180] Gong Y, Cooray A, Silva M B, Santos M G and Lubin P 2011 Probing reionization with intensity mapping of molecular and fine-structure lines *Astrophys. J.* **728** L46

[181] Obreschkow D, Heywood I, Klöckner H-R and Rawlings S 2009 A heuristic prediction of the cosmic evolution of the Co-luminosity functions *Astrophys. J.* **702** 1321–35

[182] Shetty R, Glover S C, Dullemond C P and Klessen R S 2011 Modelling CO emission: I. CO as a column density tracer and the X factor in molecular clouds *Mon. Not. R. Astron. Soc.* **412** 1686–700

[183] Sternberg A, Dalgarno A, Herbst E and Pei Y 2011 Molecular clouds at the reionization epoch *EAS Publ. Ser.* **52** 43–6

[184] Field G B 1959 An attempt to observe neutral hydrogen between the galaxies *Astrophys. J.* **129** 525

[185] Carilli C L, Gnedin N Y and Owen F 2002 H I 21 centimeter absorption beyond the epoch of reionization *Astrophys. J.* **577** 22–30

[186] Furlanetto S R and Loeb A 2002 The 21 centimeter forest: radio absorption spectra as probes of minihalos before reionization *Astrophys. J.* **579** 1–9

[187] Mack K J and Wyithe J S B 2011 Detecting the redshifted 21 cm forest during reionization arXiv:1101.5431

[188] Xu Y, Chen X, Fan Z, Trac H and Cen R 2009 The 21 cm forest as a probe of the reionization and the temperature of the intergalactic medium *Astrophys. J.* **704** 1396–404

[189] Xu Y, Ferrara A and Chen X 2011 The earliest galaxies seen in 21 cm line absorption *Mon. Not. R. Astron. Soc.* **410** 2025–42

[190] Furlanetto S R 2006 The 21-cm forest *Mon. Not. R. Astron. Soc.* **370** 1867–75

[191] Xu Y, Ferrara A, Kitaura F S and Chen X 2010 Searching for the earliest galaxies in the 21 cm forest *Sci. China Phys. Mech. Astron.* **53** 1124–9

[192] Meiksin A 2011 The micro-structure of the intergalactic medium: I. The 21 cm signature from dynamical minihaloes *Mon. Not. R. Astron. Soc.* **417** 1480–509

[193] Mortlock D J *et al* 2011 A luminous quasar at a redshift of z = 7.085 *Nature* **474** 616–9

[194] Hopkins P F, Richards G T and Hernquist L 2007 An observational determination of the bolometric quasar luminosity function *Astrophys. J.* **654** 731–53

[195] Jiang L, Fan X, Ivezić Ž, Richards G T, Schneider D P, Strauss M A and Kelly B C 2007 The radio-loud fraction of quasars is a strong function of redshift and optical luminosity *Astrophys. J.* **656** 680–90

[196] Haiman Z, Quataert E and Bower G C 2004 Modeling the counts of faint radio-loud quasars: constraints on the supermassive black hole population and predictions for high redshift *Astrophys. J.* **612** 698–705

[197] Willott C J *et al* 2010 The Canada–France High-z Quasar Survey: nine new quasars and the luminosity function at redshift 6 *Astron. J.* **139** 906–18

[198] Ioka K and Mészáros P 2005 Radio afterglows of gamma-ray bursts and hypernovae at high redshift and their potential for 21 centimeter absorption studies *Astrophys. J.* **619** 684–96

[199] de Souza R S, Yoshida N and Ioka K 2011 Populations III.1 and III.2 gamma-ray bursts: constraints on the event rate for future radio and x-ray surveys *Astron. Astrophys.* **533** A32

[200] Campisi M A, Maio U, Salvaterra R and Ciardi B 2011 Population III stars and the long gamma-ray burst rate *Mon. Not. R. Astron. Soc.* **416** 2760–7

[201] Pritchard J R, Loeb A and Wyithe J S B 2010 Constraining reionization using 21-cm observations in combination with CMB and Lyα forest data *Mon. Not. R. Astron. Soc.* **408** 57–70

[202] Mitra S, Choudhury T R and Ferrara A 2011 Reionization constraints using principal component analysis *Mon. Not. R. Astron. Soc.* **413** 1569–80