Title
Physical Cosmology From the 21-cm Line

Permalink
https://escholarship.org/uc/item/73f1b12f

Author
Furlanetto, Steven R

Publication Date
2019-09-26

Peer reviewed
Chapter 3

Physical Cosmology From the 21-cm Line

Steven R. Furlanetto, University of California, Los Angeles

Excerpted from The Cosmic 21-cm Revolution: Charting the first billion years of our Universe, Ed. Andrei Mesinger (Bristol: IOP Publishing Ltd) AAS-IOP ebooks, http://www.iopscience.org/books/aas

We describe how the high-$z$ 21-cm background can be used to improve both our understanding of the fundamental cosmological parameters of our Universe and exotic processes originating in the dark sector. The 21-cm background emerging during the cosmological Dark Ages, the era between hydrogen recombination and the formation of the first luminous sources (likely at $z \sim 30$), is difficult to measure but provides several powerful advantages for these purposes: in addition to the lack of astrophysical contamination, it will allow probes of very small scale structure over a very large volume. Additionally, the 21-cm background is sensitive to the thermal state of the intergalactic hydrogen and therefore probes any exotic processes (including, e.g., dark matter scattering or decay and primordial black holes) during that era. After astrophysical sources have formed, cosmological information can be separated from astrophysical effects on the 21-cm background through methods such as redshift space distortions, joint modeling, and by searching for indirect effects on the astrophysical sources themselves.

3.1 Introduction

The previous chapter has shown how the first galaxies and black holes have enormous implications for the 21-cm background. However, all these astrophysical processes occur within the framework of cosmological structure formation – a process we would like to probe to understand the fundamental properties of our Universe. This chapter will examine ways in which the 21-cm background can be used to probe the cosmology. Just as fluctuations in the cosmic microwave background and galaxy distribution can be used as probes of cosmology, so can the H I distribution at $z > 10$. We shall see that the 21-cm background offers an unparalleled probe of the matter distribution in our Universe [65]: interferometric measurements
can, at least in principle, map the distribution of gas over a wide range in both redshift and physical scale. Moreover, we have seen that the amplitude of the 21-cm signal depends sensitively on the thermal state of the IGM. Although the combination of adiabatic expansion and X-ray heating determines that state in the standard scenario, any “exotic” process – like dark matter decay, scattering between dark matter and baryons, X-rays from primordial black holes, etc. – that changes this energy balance will leave a signature in the 21-cm background. The low temperature of the hydrogen gas before any astrophysical X-ray background forms means that the 21-cm line is an exceptionally sensitive calorimeter for these processes.

In this chapter, we will review some of these potential cosmological probes. We begin in section 3.2 with a discussion of cosmology in the “Dark Ages” before structure forms – an era that should be uncontaminated by astrophysics, although it is also extraordinarily hard to observe. Then, in section 3.3, we consider how cosmological information can be extracted from the signal even in the presence of astrophysical processes. Finally, in section 3.4, we briefly point out that 21-cm measurements can offer strong synergies with other cosmological probes.

3.2 Cosmology in the Dark Ages

3.2.1 Setting the Stage: the Standard Cosmological Paradigm

Before astrophysical sources turn on, the 21-cm background depends on the thermal evolution of the intergalactic medium (IGM) and the earliest stages of structure formation in the Universe. We will therefore first describe these processes in the context of the standard cosmological paradigm.

Thermal Evolution

Let us begin with the thermal evolution. If it were thermally isolated, the IGM gas would simply cool adiabatically as the Universe expands. For an ideal gas this cooling rate can be written as $(\gamma - 1)(\dot{\rho}_b/\rho_b)\bar{T}_e$, where $\rho_b$ is the baryon density, $\bar{T}_e$ is the electron temperature (equal to the hydrogen temperature in this regime), and $\gamma = 5/3$ is the adiabatic index of a mono-atomic gas. For gas at the mean density, the factor $(\dot{\rho}_b/\rho_b) = -3H$ due to the Hubble expansion.

However, the gas is not actually thermally isolated: free electrons may exchange energy with CMB photons through Compton scattering. Although cosmological recombination at $z \sim 1100$ results in a nearly neutral universe, a small fraction $\bar{x}_e \sim 10^{-4}$ of electrons are “frozen out” following the recombination process. These free electrons scatter off CMB photons and, for a long period, maintain thermal equilibrium with that radiation field. The timescale for Compton cooling is

$$t_C = \left(\frac{8\sigma_T a_{rad} T^4_\gamma}{3m_e c}\right)^{-1} = 1.2 \times 10^8 \left(\frac{1+z}{10}\right)^{-4} \text{yr}, \quad (3.1)$$

where $T_\gamma \propto (1+z)$ is the background radiation temperature (in this case, the CMB), $\sigma_T$ is the Thomson cross section, $a_{rad}$ is the radiation constant, and $m_e$ is the electron mass.
3.2. Cosmology in the Dark Ages

Figure 3.1: Thermal and ionization history during the Dark Ages (panels a and b, respectively). The free electron fraction decreases rapidly after recombination at $z \sim 1100$ and then “freezes-out” at later times. Meanwhile, Compton scattering keeps $\bar{T}_e \approx T_\gamma$ until $z \sim 200$, after which the declining CMB energy density and small residual ionized fraction are no longer sufficient to maintain thermal equilibrium between the gas and CMB. Past that point, $\bar{T}_e \propto (1 + z)^2$ as appropriate for an adiabatically expanding non-relativistic gas. These results were produced with the publicly available code RECFAST (http://www.astro.ubc.ca/people/scott/recfast.html).

Including both adiabatic cooling and Compton heating, the temperature evolution of gas at the mean cosmic density, is therefore described by

$$\frac{d\bar{T}_e}{dt} = \frac{\bar{x}_e}{(1 + \bar{x}_e)} \left[ \frac{T_\gamma - \bar{T}_e}{t_c(z)} \right] - 2H\bar{T}_e. \tag{3.2}$$

The first term describes Compton heating. For an electron-proton gas, $\bar{x}_e = n_e/(n_e + n_H)$ where $n_e$ and $n_H$ are the electron and hydrogen densities; the relation is more complicated when helium is included. This prefactor appears because the electrons must share the energy they gain from Compton scattering with the other particles. The last term on the right-hand-side of equation (3.2) yields the adiabatic scaling $\bar{T}_e \propto (1 + z)^2$ in the absence of Compton scattering.

The temperature evolution therefore depends on the residual fraction of free electrons after cosmological recombination [59, 1, 13]. It is easy to estimate this residual fraction, but in detail depends on the complex physics of hydrogen recombination. In a simple picture,
the hydrogen recombination rate is
\[ \frac{d\bar{x}_e}{dt} = -\alpha_B(T_e)^2 \bar{x}_e \bar{n}_H \] (3.3)

where \( \alpha_B \propto T_e^{-0.7} \) is the case-B recombination coefficient. (The case-B coefficient ignores recombinations to the ground state, which generate a new ionizing photon and so do not change the net ionized fraction.) In the standard cosmology, the fractional change in \( \bar{x}_e \) per Hubble time is therefore
\[ \frac{\dot{\bar{n}}_e}{H^{-1} \bar{n}_e} \approx 7x(1+z)^{0.8}, \] (3.4)

where \( n_e \) is the electron fraction. Electrons “freeze-out” and cease to recombine effectively when this factor becomes of order unity; after that point, the Hubble expansion time is shorter than the recombination time. More precise numerical calculations account for the large photon density during cosmological recombination, line emission, and recombinations to higher energy levels, amongst other factors [59, 1, 13]. Figure 3.1 shows the result of one such calculation, which yields \( x \approx 3 \times 10^{-4} \) by \( z \approx 200 \).

Inserting this electron density into equation (3.2), we find that the small fraction of residual electrons maintains thermal equilibrium between the gas and CMB down to \( z \approx 200 \), when Compton heating finally becomes inefficient. Figure 3.1 shows a more exact calculation: note how the gas and CMB temperatures begin to depart at \( z \sim 200 \), after which the gas follows the expected adiabatic cooling track.

### Density Fluctuations

Measurements of fluctuations in the 21-cm background depend on variations in the density, spin temperature, and ionization fraction. We therefore briefly consider here how fluctuations in those quantities can emerge during the Dark Ages. A complete treatment of the power spectrum and structure formation is well outside the scope of this work, but we will summarize the key points for understanding the 21-cm signal.

Density fluctuations grow through gravity and, in the linear regime, their statistical properties can be calculated precisely. Combining the mass and momentum conservation equations for a perfect fluid with the Poisson equation for the gravitational potential yields an evolution equation for the Fourier transform \( \delta_k \) of the fractional density perturbation \( \delta \),
\[ \frac{\partial^2 \delta_k}{\partial t^2} + 2H \frac{\partial \delta_k}{\partial t} = 4\pi G \bar{\rho} \delta - \frac{c_s^2 k^2}{a^2} \delta_k, \] (3.5)

where the last term is the pressure force (which vanishes for cold dark matter) and \( c_s^2 \) is the sound speed. This linear equation has two independent solutions, one of which grows in time and eventually comes to dominate the density evolution. \(^1\)

While the fluctuations are small – and thus very nearly Gaussian – the density field can be accurately characterized by the power spectrum,
\[ P(k) = \langle \delta_k \delta_k^* \rangle = (2\pi)^3 \delta^D(k - k')P(k). \] (3.6)

\(^1\)Note that this solution uses Eulerian perturbation theory, which breaks down when \( \delta \) is still relatively small. It suffices during the Dark Ages, but greater accuracy is necessary during later eras.
Figure 3.2: Power spectra for density and temperature fluctuations as a function of comoving wavenumber at four different redshifts. The curves show dimensionless power spectra for the CDM density (solid), baryon density (dotted), baryon temperature (short-dashed), and photon temperature (long-dashed). These curves do not include the relative streaming of the baryons and cold dark matter. Reproduced from Naoz, S. & Barkana, R., “Growth of linear perturbations before the era of the first galaxies, Monthly Notices of the Royal Astronomical Society, vol. 362, pp. 1047–1053. Copyright OUP 2005.

The power spectrum is the expectation value of the Fourier amplitude at a given wavenumber; for a homogeneous, isotropic universe it depends only on the magnitude of the wavenumber.\(^2\) The power spectrum therefore represents the variance in the density field as a function of smoothing scale.

The solid curve in Figure 3.2 shows the cold dark matter power spectrum at several redshifts during the Dark Ages (taken from [56]). The most obvious feature is the flattening at \(k \sim 0.1 \text{ Mpc}^{-1}\), which results from stagnation in the growth of small-scale structure during the radiation era. The power spectrum is otherwise quite simple. However, the 21-cm line will probe the fluctuations in the IGM gas, which we will require more physics to understand.

For example, equation (3.2) describes the evolution of the mean IGM temperature. However, that field actually fluctuates as well, for two reasons [56]. First, the CMB temperature itself has fluctuations, so each electron will scatter off a different local \(T_\gamma\): the power spectrum of the CMB temperature fluctuations is shown by the long-dashed curve in Figure 3.2. Additionally, the adiabatic expansion term depends on the local density, because

\(^2\)Henceforth we will suppress the \(k\) subscript for notational simplicity.
gravity slows the expansion of overdense regions and hence decreases the cooling rate (and of course in underdense regions, the cooling accelerates). Thus, the IGM will be seeded by small temperature fluctuations reflecting its density structure.

To describe these fluctuations, we write $\delta T$ as the fractional gas temperature fluctuation and $\delta \gamma$ as the photon density fluctuation (so that $\delta \gamma = 4 \delta T$, the fractional CMB temperature fluctuation). Then the perturbed version of equation (3.2) is

$$\frac{d \delta T}{dt} = \frac{2}{3} \frac{d \delta b}{dt} + \frac{x_e(t)}{t_C(z)} \left[ \delta \gamma \left( \frac{\bar{T}_\gamma}{\bar{T}_e} - 1 \right) + \frac{\bar{T}_\gamma}{\bar{T}_e} (\delta \bar{T}_\gamma - \delta T) \right],$$  \hspace{1cm} (3.7)

where the first term describes adiabatic cooling due to expansion (allowing for variations in the expansion rate) and the second accounts for variations in the rate of energy exchange through Compton scattering (which can result from variations in either the gas or photon temperatures).

Meanwhile, the fluctuations in the baryon temperature affect the baryon density evolution as well. Equation (3.5) implicitly assumed that temperature fluctuations were driven (only) by density fluctuations; allowing a more general relation, we obtain [56]

$$\frac{\partial^2 \delta}{\partial t^2} + 2H \frac{\partial \delta}{\partial t} = \frac{3}{2} H^2 (\Omega_c \delta_c + \Omega_b \delta_b) - \frac{k^2}{a^2} \frac{k_B \bar{T}_e}{\mu m_H} (\delta_b + \delta_T).$$ \hspace{1cm} (3.8)

This, together with equations (3.7), (3.2), and a more precise version of (3.3) for the temperature and ionized fraction evolution, provide a complete set of equations to trace the density and temperature evolution (modulo one more effect that we will discuss next).

Figure 3.2 shows the resulting power spectra for the dark matter density, baryon density, baryon temperature, and photon temperature perturbations at four different redshifts. The photon fluctuations are not directly observable, but the others can in principle be probed through the 21-cm line. The photon perturbations are strongly suppressed on scales below the sound horizon thanks to their large pressure. Near recombination, the baryonic perturbations are also suppressed on these scales, especially in the temperature, because they interact so strongly with the CMB. However, after recombination, the baryons fall into the dark matter potential wells, with their perturbations rapidly growing, and temperature fluctuations also grow thanks largely to the variations in the adiabatic cooling rate. The turnover at very small scales in the baryonic power spectrum is due to the finite pressure of the gas.

**Relative Streaming of Baryons and Cold Dark Matter**

There is one additional effect on the baryonic power spectrum that may provide insight into cosmology during this era: “streaming” of baryonic matter relative to dark matter [66]. As a relativistic fluid, CMB photons have a very high pressure that drives acoustic waves throughout that component. While these photons are coupled to the baryons through Compton scattering, they can drag the baryonic component along with them – it is these acoustic waves in the photon-baryon fluid that we see in CMB fluctuations. Once recombination occurs, the radiation drag force decreases, and the baryons begin to fall into the potential wells of dark matter fluctuations (which have not participated in the acoustic waves). This transition can be seen in the dotted and short-dashed curves in Figure 3.2. Because the radiation sound
speed is $\sim c/\sqrt{3}$; it is near the causal horizon at the time of recombination, corresponding to $\sim 150$ comoving Mpc today, where they can be observed as “baryon acoustic oscillations” in the matter power spectrum (and, because their physical scale is well-known, used as a standard ruler to measure cosmological parameters).

There is an additional effect relevant to high redshifts, however. Even after the radiation driving becomes ineffective, the baryons are left with a relic velocity imprinted by the pressure force with a root-mean-square (rms) speed of $v_{bc} \approx 10^{-4} c = 30$ km s$^{-1}$, which decays with redshift as $1/a$ [66]. This streaming velocity between the baryonic and dark matter components is coherent over very large scales – comparable to the acoustic scale – and fades only slowly over time. Figure 3.3 shows the variance of the relative velocity perturbations as a function of the mode wavenumber $k$ at $z = 15$. The power is substantial on scales up to the sound horizon at decoupling ($\sim 140$ comoving Mpc), but it declines rapidly at $k > 0.5$ Mpc$^{-1}$, indicating that the streaming velocity is coherent over a scale of several comoving Mpc. Therefore, in the rest-frame of small-scale fluctuations (such as those that will eventually collapse into galaxies), the baryons appeared to be moving coherently as a “wind,” which can suppress the formation of the first galaxies [17, 23].
3.2.2 The Global 21-cm Signal During the Dark Ages

With the thermal history from Figure 3.1, we can easily calculate the global 21-cm signal throughout the Dark Ages – without any way of generating a Lyman-α background, the competition between the CMB and particle collisions sets the spin temperature. The left panel of Figure 3.4 shows the result: once the gas temperature begins to fall below the CMB temperature at \( z \sim 150 \), the 21-cm signal can be seen in absorption. At these early times, the gas is relatively dense and has a high enough temperature for collisional coupling to be substantial, so \( T_S \rightarrow T_K \). However, as the gas expands and cools, collisional coupling becomes inefficient, and eventually the spin temperature begins to return to \( T_\gamma \). By the time we expect star formation to begin in earnest \( (z < 30) \), \( T_S \approx T_\gamma \), so the 21-cm line is nearly invisible.

The global 21-cm signal is therefore sensitive to the IGM thermal history, so any process that affects the temperature evolution can also be probed by the 21-cm line. This method provides a particularly powerful probe of non-standard physics because the low gas temperature during this period makes the 21-cm line a sensitive calorimeter of additional heating (or cooling) and/or of an excess radio background over and above the CMB [22].

While many such processes tend to heat the IGM and therefore decrease the amplitude of the 21-cm signal (e.g. [9, 24, 61, 14]), the recent claim of a detection of a 21-cm absorption feature at 78 MHz \( (z \sim 17) \) by the EDGES collaboration [7] has triggered interest in non-standard models that amplify the 21-cm signal (though note that the EDGES signal has not yet been independently confirmed, and with such a challenging analysis systematics and foreground contamination remain a concern [26, 19, 64, 8]). The EDGES feature is more than twice as deep as deep as expected if gas follows the standard thermal history, even if
3.2. COSMOLOGY IN THE DARK AGES

astrophysical sources were able to drive $T_S \rightarrow T_K$ and no X-ray heating had yet occurred. It therefore requires new physics that either cools the IGM or increases the radio background against which the gas absorbs. A multitude of processes has been proposed to explain the anomalous detection, including dark matter-baryon scattering [54, 3, 63, 27], millicharged dark matter [55, 5, 33], axions [50], neutrino decay [12], charge sequestration [21], quark nuggets [34], dark photons [29], and interacting dark energy [16], and it has been used to constrain additional exotic processes like dark matter annihilation [11].

Although the EDGES signal at $z \sim 17$ is very likely past the end of the Dark Ages – after the first astrophysical sources formed – the same new physics would have affected the Dark Ages signal. The solid curves in Figure 3.4 illustrate how models that invoke excess cooling to explain the EDGES result (shown schematically by the gray contours) could also greatly amplify the Dark Ages signal. (Here the curves use a phenomenological parameterized cooling model as in [49]; physically-motivated models will differ in the details.) (Note that the signal still vanishes at $z \sim 30$ because collisional coupling is still inefficient.) In these cases, even though the exotic physics have implications at relatively low redshifts, Dark Ages observations serve to break degeneracies between astrophysics and that new physics.

3.2.3 The Power Spectrum During the Dark Ages

As we have discussed, maps of 21-cm emission provide a sensitive probe of the power spectrum of density and temperature fluctuations [32, 43]. Figure 3.5, from [58], shows examples of how modes of the 21-cm fluctuation power spectrum evolve in the standard cosmology. Although the strongest fluctuations occur at $z \sim 10$, when astrophysical sources dominate, the Dark Ages signal are not too far behind: this is because, although the fractional density fluctuations are small at those times, the mean temperature can be relatively large during the absorption era – even in the standard cosmology.

The 21-cm power spectrum during the Dark Ages offers several advantages over other probes of the density field. Because they use the cosmological redshift to establish the distance to each observed patch, 21-cm measurements probe three-dimensional volumes — unlike the CMB, which probes only a narrow spherical shell around recombination. Additionally, the 21-cm line does not suffer from Silk damping (photon diffusion), which suppresses the CMB fluctuations on relatively large scales. The number of independent modes accessible through this probe is therefore [39]

$$N_{21\text{cm}} \sim 8 \times 10^{11} \left( \frac{k_{\text{max}}}{3 \text{ Mpc}^{-1}} \right)^3 \left( \frac{\Delta \nu}{\nu} \right) \left( \frac{1+z}{100} \right)^{-1/2},$$

(3.9)

where $\Delta \nu$ is the bandwidth of the observation. The choice of $k_{\text{max}}$ — the smallest physical scale to be probed — is not obvious. The Jeans length during the Dark Ages corresponds to $k_{\text{max}} \sim 1000 \text{ Mpc}^{-1}$. Accessing these small-scale modes in three dimensions would require an enormous instrument, but our relatively conservative choice in equation (3.9) shows that even a more modest effort provides a massive improvement over the information contained in all the measurable modes of the CMB, $N_{\text{CMB}} \sim 10^7$. Moreover, because the density fluctuations are still small at such early epochs, these modes remain in the linear or mildly non-linear regime, allowing a straightforward interpretation of them in terms of the fundamental parameters of our Universe [35].
Figure 3.5: 21-cm fluctuations are substantial during the Dark Ages. The curves show the amplitude of the 21-cm brightness temperature fluctuations at several different wavenumbers from the Dark Ages to low redshifts, in the standard model of cosmology. Red diagonal lines compare these fluctuations to the foreground brightness temperature (from Galactic synchrotron): each scales the foreground by the number shown. Note that exotic cooling scenarios described in §3.2.2 could significantly amplify these fluctuations. From [58]. Copyright IOP Publishing. Reproduced with permission. All rights reserved.
Figure 3.6: The 21-cm power spectrum is a sensitive probe of cosmological parameters. The angular power spectrum of 21-cm fluctuations at $z = 55$ in “standard” cosmologies, varying some of the parameters. The solid and dotted curves use $\Lambda$CDM with the power law index of density fluctuations $n_s = 1$ and 0.98; respectively; the short-dashed curve adds a “running” to the spectral index. The long-dashed curve assumes 10% of the matter density is in the form of massive neutrinos (with masses 0.4 keV), while the dot-dashed curve assumes warm dark matter with particle masses of 1 keV. Reprinted figure from Loeb, A. & Zaldarriaga, M., Phys. Rev. Lett. 92, 21, 211301. Copyright 2004 by the American Physical Society.
In principle, measurements of the 21-cm power spectrum during the Dark Ages will therefore enable a number of precision cosmological measurements, even within the “standard” cosmology. For example, because 21-cm fluctuations extend to such small scales, they expand the dynamic range of power spectrum measurements over several orders of magnitude. This is useful for a number of specific cosmological parameters, as illustrated by the curves in Figure 3.6, but also including the running of the spectral index of the matter power spectrum [43], a key parameter that can test the underlying assumptions of the inflationary paradigm, as well as the total spatial curvature and neutrino masses [43].

The large number of modes is also essential to constraining primordial non-Gaussianity in the cosmological density field, which is another tool for testing inflationary models. CMB measurements already offer constraints, which can best be improved by larger-volume surveys. However, low-z galaxy surveys suffer from contamination, because nonlinear structure formation also generates non-Gaussianity. The “clean,” small-scale Dark Ages 21-cm field is an excellent opportunity to further constrain the non-Gaussianity [10] – in principle, 21-cm Dark Ages measurements can test the generic inflationary picture itself [53].

The potential insights we have discussed so far are useful within the standard cosmology, but the exotic physics mechanisms discussed in the previous section will inevitably affect the 21-cm power spectrum as well, though the implications have only recently begun to be explored. Most obviously, the power spectrum amplitude is proportional to the square of the mean brightness temperature. But more subtle signatures that depend on the particulars of the physics can also appear. For example, if the cooling is triggered by scattering between the baryons and a fraction of the dark matter that has a modest charge, the scattering rate will be modulated by the relative streaming velocity of dark matter and baryons. Such a model therefore leaves distinct features in the 21-cm power spectrum that trace the velocity structure [52, 51]. The detailed implications of most of these exotic models are mostly unexplored, but any model in which (1) the energy exchange depends on local density, velocity, or temperature; or (2) in which background radio energy deposition is inhomogeneous, should generically leave signatures in the power spectrum.

3.3 Cosmology During the Era of Astrophysics

Although the Dark Ages offer a clean and powerful testbed to probe cosmology, in practice they will be difficult to explore. Not only do the astrophysical foregrounds become significantly stronger at low frequencies (as shown by the diagonal lines in Figure 3.5, which increase toward higher redshifts), but Earth’s ionosphere also becomes opaque at very low frequencies [18], which may necessitate observations from space or the Moon.

Because the 21-cm background will likely be observed at lower redshifts – after the first luminous sources have appeared – we will next consider how cosmological information might be extracted from observations during the Cosmic Dawn. The effects we have already described will also affect the global 21-cm signal and the power spectrum during the later era. However, many astrophysical mechanisms (those described in Chapter 2) also affect the 21-cm background during this era, so the challenge is to separate the cosmological information from the astrophysical. In this section, we shall consider some strategies to do that.
3.3. COSMOLOGY DURING THE ERA OF ASTROPHYSICS

3.3.1 Isolating the Matter Power Spectrum

Conceptually, the simplest method is simply to measure the 21-cm power spectrum and fit simultaneously for both astrophysical and cosmological parameters. We will discuss the practicalities of such fits in Chapter 4, but it will suffice for now to note that it is not a trivial exercise, and adding additional cosmological parameters will be a challenge [15, 31, 57]. Moreover, the astrophysical effects can be very strong and hence mask any cosmological effects.

Short of modeling the astrophysics precisely enough to extract cosmology – a strategy whose potential won’t really be known until the astrophysics is better understood – the chief hope for isolating the power spectrum is that there exists an era in which astrophysics processes can largely be ignored. This is not an entirely unreasonable expectation. Recall from Chapter 1 that the 21-cm brightness temperature is

$$T_b(\nu) \approx 9 x_{\text{HI}} (1 + \delta) (1 + z)^{1/2} \left[ 1 - \frac{T_\gamma(z)}{T_S} \right] \left[ \frac{H(z)/(1 + z)}{dv/|dr|} \right] \text{mK.}$$  (3.10)

Astrophysical effects determine the neutral fraction $x_{\text{HI}}$ and the spin temperature $T_S$, but the other factors – density and velocity – are driven by cosmological processes. Thus we can imagine that cosmological information will show up clearly in the power spectrum if, for example, there exists a period in which $T_S \gg T_\gamma$ (so that the temperature effects can be ignored) but in which ionization fluctuations are not yet significant.

Simple estimates show that such a period is far from impossible – but also not guaranteed. Reionization requires at least one ionizing photon per baryon, or of order $\sim 10$ eV of ionizing energy released by stars per baryon. Heating the IGM above the CMB temperature – so that $(1 - T_\gamma/T_S) \approx 1$ requires only $\sim 10^{-2}$ eV (corresponding to a temperature $\sim 100$ K, though only a fraction of the X-ray energy would actually be used to heat the IGM; see §2.2). Thus if early sources produce at least $\sim 10^{-3}$ as much energy in X-rays as they do in ionizing photons, heating would occur before reionization – leaving open the possibility that a period exists in which astrophysics can be ignored. More complex astrophysical processes can also enable such a period as well – for example, strong photoheating feedback can delay reionization relative to X-ray heating [47]. Whether this is more than speculation remains uncertain: calibrating the X-ray luminosity of star-forming galaxies to local measurements suggests that the reionization epoch may overlap with the X-ray heating epoch [48, 57], but if the EDGES measurement is confirmed, heating must actually occur very early in the Cosmic Dawn, well before reionization is complete.

3.3.2 Redshift Space Distortions

To this point, we have largely ignored the last factor in equation (3.10): the velocity gradient. However, it offers another route to extracting cosmological information. Usually, we expect the fluctuations from the other terms – density, ionization fraction, Ly\(\alpha\) flux, and temperature

However, note that recent work has shown that the ionization field may be constructed as a perturbative expansion around the density field in certain regimes, which suggests precision modeling may indeed be possible [45, 28].
– to be isotropic, because the processes responsible for them have no preferred direction [e.g., \( \delta(k) = \delta(k) \)]. However, peculiar velocity gradients introduce anisotropic distortions through the “Kaiser effect” [30], which emerge because 21-cm observations use the line’s observed frequency as a proxy for the distance of the cloud that produced it.

Consider a spherical overdense region. Because of its enhanced gravitational potential, the region expands less quickly than an average region of the same mass. Therefore, to an observer using redshift as a distance indicator, the apparent radial size of the overdense region is smaller than that of the average region. Of course, its transverse size can be measured by its angular extent on the sky so is unaffected by the velocity structure. Thus, to the observer, the spherical region is distorted, appearing larger along the plane of the sky than in the radial direction. An underdense region is distorted in the opposite way: because it expands faster than average, it appears larger in the radial direction than along the plane of the sky. Thus redshift space distortions will exaggerate intrinsic density fluctuations, but they do so in an anisotropic way – making this source of fluctuations separable from others [2, 44].

To see these effects, we start by labeling the coordinates in redshift space with \( s \). Assuming that the radial extent of the volume is small, so that the Hubble parameter \( H \) is constant throughout the volume, these coordinates are related to the real space \( r \) by

\[
\mathbf{s}(\mathbf{r}) = \mathbf{r} + \frac{\mathbf{U}(\mathbf{r})}{H},
\]

where \( \mathbf{U}(\mathbf{r}) = \mathbf{v} \cdot \hat{\mathbf{x}} \) is the radial component of the peculiar velocity.

Now consider a set of particles with number density \( n(\mathbf{r}) \) that are biased with respect to the dark matter by a factor \( b \). Number conservation demands that the fractional overdensity in redshift space is related to that in real space via \([1 + \delta_s(s)]d^3s = [1 + \delta(r)]d^3r\). The Jacobian of the transformation is

\[
d^3s = d^3r \left[ 1 + \frac{U(r)}{r} \right]^2 \left[ 1 + \frac{dU(r)}{dr} \right],
\]

because only the radial component of the volume element, \( r^2dr \), changes from real to redshift space. Thus the density observed in redshift space increases if the peculiar velocity gradient is smaller than the Hubble flow, while the redshift space density will be smaller if the peculiar velocity gradient is larger. Thus, assuming \(|U(r)| \ll Hr\),

\[
\delta_s(r) = \delta(r) - \left( \frac{d}{dr} + \frac{2}{r} \right) \frac{U(r)}{H}.
\]

Importantly, the peculiar velocity field itself is a function of the dark matter density field, as described qualitatively above. More rigorously, in Fourier space the components of the peculiar velocity \( u_k \) are directly related to those of the density field, because the latter sources the gravitational fluctuations that drive the velocity gradients (e.g., [30]):

\[
u_k = -i a H f(\Omega) \frac{k}{\hat{k}} \delta \hat{k},
\]

where \( f(\Omega) \) is a function of the growth rate of cosmological perturbations.
Equation (3.13) shows that there are two corrections from the redshift space conversion. To see which of these dominates, consider a plane wave perturbation, \( U \propto e^{i k \cdot r} \). Then the derivative term is \( \sim k U / H_0 \) while the last term is \( \sim U / H_0 r \). But \( r \) is the median distance to the survey volume, and \( k \) corresponds to a mode entirely contained inside it, so \( kr \gg 1 \), and we may neglect the last term. If we further make the small-angle approximation and make a Fourier transform, so that \( \hat{x} \) is also approximately a constant over the relevant volume, the Fourier transform of equation (3.13) is

\[
\delta_s(k) = \delta(k) \left[ 1 + \beta \mu_k^2 \right],
\]

(3.15)

where \( \mu_k = \hat{k} \cdot \hat{x} \) is the cosine of the angle between the wave vector and the line of sight. Here \( \beta = f(\Omega_m) / b \) corrects for a possible bias between the tracers we are studying and the growth rate of dark matter perturbations. For the case of 21-cm fluctuations in the IGM gas, the bias factor is very close to unity except below the Jeans filtering scale.

The redshift-space distortions therefore provide an anisotropic amplification to the background signal, because only modes along the line of sight are affected. Averaged over all modes, these distortions amplify the signal by a factor \( \approx \langle (1 + \mu^2)^2 \rangle \approx 1.87 \) [6].

For the purposes of extracting cosmological information, the anisotropies are helpful in that they imprint angular structure on the signal, which may allow us to separate the many contributions to the total power spectrum [2]. Brightness temperature fluctuations in Fourier space have the form

\[
\delta_{21} = \mu^2 \beta \delta + \delta_{iso}
\]

(3.16)

where we have collected all the statistically isotropic terms – including those due to astrophysics – into \( \delta_{iso} \). Neglecting “second-order” terms (see below) and setting \( \beta = 1 \), the total power spectrum can be written

\[
P_{21}(k) = \mu^4 P_{\delta\delta} + 2 \mu^2 P_{\delta_{iso}\delta} + P_{\delta_{iso}\delta_{iso}}.
\]

(3.17)

(Here we have written the normal density power spectrum as \( P_{\delta\delta} \) for clarity.) By separately measuring these three angular components (which requires, in principle, estimates at just a few values of \( \mu \)), we can isolate the contribution from density fluctuations \( P_{\delta\delta} \). Measuring this component, without any astrophysical contributions, will provide the desired cosmological constraints.

However, in practice the angular dependence of the power spectrum will not be so simple. To write equation (3.17), we must neglect “second-order” terms in the perturbation expansion of the 21-cm field, such as the density and the ionization field perturbations. But the latter is not actually a small term (because, at least in the standard reionization scenarios, \( x_{HI} = 0 \) or 1), so its contributions do not decrease rapidly in higher-order terms [36]. During reionization, these additional terms complicate the angular dependence and will significantly complicate attempts to separate the \( \mu^n \) powers during reionization more difficult [44, 60]. Moreover, if the first H II regions are highly biased – thus overlapping the regions with the largest peculiar velocities – the redshift space distortions can be suppressed [46, 42]. The redshift space distortions are also more complicated if the heating and reionization eras overlap [25].
CHAPTER 3. PHYSICAL COSMOLOGY FROM THE 21-CM LINE

3.3.3 Indirect Effects of Cosmology on the 21-cm Background

Finally, it is also worth noting that cosmological processes can have direct effects on astrophysical sources and hence indirect effects on the 21-cm background. For example, consider warm dark matter. If dark matter has a non-zero velocity dispersion, then it can easily escape from shallow potential wells, suppressing the formation of small dark matter haloes. Because the first phases of galaxy formation occur in small haloes, warm dark matter delays galaxy formation, which in turn delays the formation of Lyman-α, X-ray, and ionizing backgrounds and changes the timing (and potentially spatial fluctuations) of the 21-cm background (e.g., [4, 68, 41]). Of course, this dark matter effect is degenerate with astrophysical processes (for example, strong feedback in small galaxies may also suppress their star formation rates) so requires careful analysis, and other cosmological changes can have qualitatively similar effects (e.g., [67]), although such effects may be distinct from large swathes of astrophysical parameter space [62].

Another example is the interaction between astrophysical processes and the relative streaming between baryons and dark matter. The streaming velocity also suppresses the formation of small baryonic haloes, because a shallow potential well cannot accrete gas traveling by it at sufficiently large velocity. However, unlike in the case of warm dark matter, the streaming effect is spatially variable, so, at least in some circumstances, the streaming effect will imprint spatial structure on the radiation backgrounds and hence on the 21-cm power spectrum [17, 23, 52]. Figure 3.7 shows an example of this effect during the era in which the first stars appear.

Finally, dark matter annihilation offers an other interesting example of the interaction
between cosmology and astrophysics. The heating from dark matter annihilation occurs very uniformly. If astrophysically-driven X-ray heating begins within such a pre-heated medium, the associated large-scale peak in the power spectrum occurs in emission rather than absorption, providing a distinct signature for (some) dark matter annihilation scenarios [20, 40].

3.4 21-cm Cosmology in a Larger Context

Although the focus of this book is on the 21-cm line itself, it is worth emphasizing that observations of the high-$z$ spin-flip background will ultimately be combined with many other observations. For cosmological measurements, it is therefore useful to understand synergies between the 21-cm background and other probes [38, 37]. Some of the key advantages of the 21-cm line have already been described: it can probe small physical scales and high-redshifts. Additionally, it can break degeneracies within other probes. Figure 3.8 shows an example. If the 21-cm line can measure the reionization history, it provides an independent estimate of the CMB optical depth to electron scattering, which is otherwise nearly degenerate with the amplitude of the initial power spectrum – and thus allow a precision measurement of that parameter.
Bibliography


