The Rise of the First Stars: Supersonic Streaming, Radiative Feedback, and 21-cm Cosmology

Rennan Barkana[a,b,c,d]

[a]Raymond and Beverly Sackler School of Physics and Astronomy, Tel Aviv University, Tel Aviv 69978, Israel
[b]Sorbonne Universités, Institut Lagrange de Paris (ILP), Institut d'Astrophysique de Paris,
UPMC Univ Paris 06/CNRS, France
[c]Department of Astrophysics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK
[d]Perimeter Institute for Theoretical Physics, 31 Caroline St N., Waterloo, ON N2L 2Y5, Canada

---

Abstract

Understanding the formation and evolution of the first stars and galaxies represents one of the most exciting frontiers in astronomy. Since the universe was filled with hydrogen atoms at early times, the most promising method for observing the epoch of the first stars is to use the prominent 21-cm spectral line of hydrogen. Current observational efforts are focused on the cosmic reionization era, but observations of the pre-reionization cosmic dawn are also beginning and promise exciting discoveries. While observationally unexplored, theoretical studies predict a rich variety of observational signatures from the astrophysics of the early galaxies that formed during cosmic dawn. As the first stars formed, their radiation (plus that from stellar remnants) produced feedback that radically affected both the intergalactic medium and the character of newly-forming stars. Lyman-α radiation from stars generated a strong 21-cm absorption signal, observation of which is currently the only feasible method of detecting the dominant population of galaxies at redshifts as early as $z \sim 25$. Another major player is cosmic heating; if due to soft X-rays, then it occurred fairly early ($z \sim 15$) and produced the strongest pre-reionization signal, while if it is due to hard X-rays, as now seems more likely, then it occurred later and may have dramatically affected the 21-cm sky even during reionization. In terms of analysis, much focus has gone to studying the angle-averaged power spectrum of 21-cm fluctuations, a rich dataset that can be used to reconstruct the astrophysical information of greatest interest. This does not, however, diminish the importance of finding additional probes that are complementary or amenable to a more model-independent analysis. Examples include the global (sky-averaged) 21-cm spectrum, and the line-of-sight anisotropy of the 21-cm power spectrum. Another striking feature may result from a recently recognized effect of a supersonic relative velocity between the dark matter and gas. This effect enhanced large-scale clustering and, if early 21-cm fluctuations were dominated by small galactic halos, it produced a prominent pattern on 100 Mpc scales. Work in this field, focused on understanding the whole era of reionization and cosmic dawn with analytical models and numerical simulations, is likely to grow in intensity and importance, as the theoretical predictions are finally expected to confront 21-cm observations in the coming years.

Keywords:
first stars, cosmic reionization, 21-cm cosmology, galaxy formation, cosmology

---

Preprint submitted to Physics Reports *May 13, 2016*

Contents

1 Introduction and Overview 3

2 Galaxy Formation: Basic Theory 8
2.1 Cosmological background . . . 8
2.2 Linear perturbation theory . . . 10
2.3 Non-linear collapse . . . 12
2.4 Baryons: linear evolution, pressure, and cooling . . . 15

3 Galaxy Formation: High-Redshift Highlights 19
3.1 Large fluctuations in the galaxy number density . . . 19
3.2 Simulations at high redshift: challenges and approaches . . . 23
3.3 The very first stars . . . 28

4 21-cm Cosmology 29
4.1 Basic physics . . . 29
4.2 Low-temperature corrections . . . 37
4.3 Anisotropy of the 21-cm signal . . . 40
4.4 Observational aspects . . . 43

5 The Supersonic Streaming Velocity 46
5.1 Cosmological origins . . . 46
5.2 Effect on star formation in early halos . . . 48
5.3 Consequences . . . 49

6 Cosmic milestones of early radiative feedback 50
6.1 Reionization . . . 50
6.2 Lyα coupling and Lyman-Werner feedback . . . 55
6.3 Cosmic heating . . . 58

7 21-cm Signatures of the First Stars 62
7.1 21-cm signatures of reionization . . . 62
7.2 21-cm signatures of Lyα coupling and LW feedback . . . 65
7.3 Large 21-cm fluctuations from early cosmic heating . . . 68
7.4 Late heating and reionization . . . 80
7.5 The global 21-cm spectrum . . . 81

8 Summary and Conclusions 84

9 Acknowledgments 88

Email address: `barkana@tau.ac.il` (Rennan Barkana)

1. Introduction and Overview

Galaxies around us have been mapped systematically out to a redshift $z \sim 0.3$ by recent large surveys [1, 2]. The observed galaxy distribution shows a large-scale filament-dominated "cosmic web" pattern that is reproduced by cosmological numerical simulations [3]. This structure is well-understood theoretically [4] as arising from the distribution of the primordial density fluctuations, which drove hierarchical structure formation in the early universe. Recent observations have been pushing a new frontier of early cosmic epochs, with individual bright galaxies detected reliably from as early as $z = 11.1$ [5], which corresponds to $t \sim 400$ Myr after the Big Bang. However, it is thought that the bulk of the early stars formed in a large number of very small galactic units, which will be difficult to observe individually. In particular, high-resolution numerical simulations show that the truly earliest stars formed within $\sim 10^6 M_\odot$ dark matter halos [6, 7]. These simulations can only follow small cosmic volumes, and thus begin to form stars much later than in the real universe, but analytical methods show that the very first such stars within our light cone must have formed at $z \sim 65$ (age $t \sim 35$ Myr) [8, 9].

The best hope of observing the bulk population of early stars is via the cosmic radiation fields that they produced. The mean radiation level traces the cosmic star formation rate, while spatial fluctuations reflect the clustering of the underlying sources, and thus the masses of their host halos. In particular, the hyperfine spin-flip transition of neutral hydrogen (H I) at a wavelength of 21 cm (Figure 1) is potentially the most promising probe of the gas and stars at early times. Observations of this line at a wavelength of $21 \times (1 + z)$ cm can be used to slice the universe as a function of redshift z (or, equivalently, distance along the line of sight), just like any atomic resonance line in combination with the cosmological redshift. Together with the other two dimensions (angular position on the sky), 21-cm cosmology can thus be used to obtain a three-dimensional map of the diffuse cosmic H I distribution [10], in the previously unexplored era of redshifts $\sim 7 - 200$.

Absorption or emission by the gas along a given line of sight changes the 21-cm brightness temperature T_b, measured relative to the temperature of the background source, which here is the cosmic microwave background (CMB) [11]. The observed T_b is determined by the spin temperature T_S, an effective temperature that describes the relative abundance of hydrogen atoms in the excited hyperfine level compared to the ground state. Primordial density inhomogeneities imprinted a three-dimensional power spectrum of 21-cm intensity fluctuations on scales down to ~ 10 kpc (all sizes henceforth are comoving unless indicated otherwise), making it the richest dataset on the sky [12]. The potential yield of 21-cm observations is further increased by the expected anisotropy of the 21-cm power spectrum [13, 14, 15, 16].

The 21-cm signal vanished at redshifts above $z \sim 200$, when the gas kinetic temperature, T_k, was close to the CMB temperature, $T_{\rm CMB}$, making the gas invisible with respect to the CMB background. Subsequently, the gas cooled adiabatically, faster than the CMB, and atomic collisions kept the spin temperature T_S of the hyperfine level population below $T_{\rm CMB}$, so that the gas appeared in 21-cm absorption [17]. As the Hubble expansion continued to rarefy the gas, radiative coupling of T_S to $T_{\rm CMB}$ started to dominate over collisional coupling of T_S to T_k and the 21-cm signal began to diminish.

Once stars began to form, their radiation produced feedback on the intergalactic medium (IGM) and on other newly-forming stars, and substantially affected the 21-cm radiation. The first feedback came from the ultraviolet (UV) photons produced by stars between the Lyα and Lyman limit wavelengths (i.e., energies in the range of $10.2 - 13.6$ eV). These photons propagated freely through the Universe and some of them redshifted or scattered into the Lyα reso-

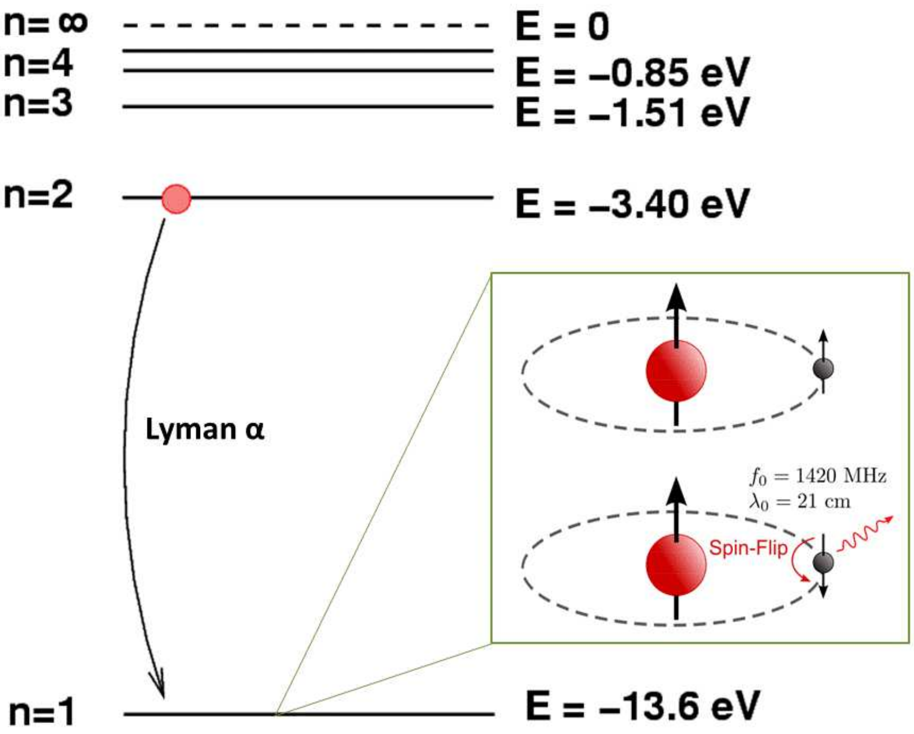

Figure 1: The energy levels of hydrogen. The ionization energy of hydrogen is 13.6 eV, and the Lyman-α (Lyα) line (10.2 eV) corresponds to the $n = 2$ to $n = 1$ transition. The spin-flip transition (inset on the right) is a much lower energy splitting (5.87×10^{-6} eV) within the ground state ($n = 1$) of hydrogen, corresponding to a wavelength of 21 cm and a frequency of 1420 MHz. Credits: Main portion: Michael Richmond (http://spiff.rit.edu/classes/phys301/lectures/spec_lines/spec_lines.html); Inset: Tiltec via Wikimedia Commons.

$1+z$	Observed 21-cm Frequency [MHz]	Cosmic Age [Myr]
1	1.42 GHz	13.8 Gyr
2	710	5.88 Gyr
3	473	3.29 Gyr
4	355	2.15 Gyr
7	203	934
10	142	547
15	94.7	297
20	71.0	192
25	56.8	137
30	47.3	104
40	35.5	67.5
50	28.4	48.2
60	23.7	36.5

Table 1: The observed frequency corresponding to 21-cm radiation from a source at redshift z, and the age of the Universe, listed versus $1+z$. Units are as in the column labels except where indicated otherwise.

nance, and coupled T_S to T_k once again [11] through the Wouthuysen-Field [18, 19] effect by which the two hyperfine states are mixed through the absorption and re-emission of a Lyα photon. Meanwhile, Lyman-Werner (LW) photons in nearly the same energy range (11.2 – 13.6 eV) dissociated molecular hydrogen and eventually ended the era of primordial star formation driven by molecular cooling [20], leading to the dominance of larger halos. X-ray photons also propagated far from the emitting sources and began early on to heat the gas [11]. Once T_S grew larger than $T_{\rm CMB}$, the gas appeared in 21-cm emission over the CMB level. Emission of UV photons above the Lyman limit by the same galaxies initiated the process of cosmic reionization, creating ionized bubbles in the neutral gas around these galaxies. Figure 2 shows a brief summary of early cosmic history, and Table 1 lists the observed frequency corresponding to 21-cm radiation from various redshifts, as well as the age of the Universe.

The subject of cosmic reionization began in earnest after Gunn & Peterson (1965) [22] used a just-discovered quasar to show that the Universe around it was highly ionized. This led to much theoretical work on how the Universe might have been reionized. The subject of 21-cm cosmology is a more recent one. Hogan & Rees (1979) [10] worked out the basic ideas and noted that 21-cm observations could probe the properties of cosmic gas including its density, temperature, and spin temperature (which, they noted, could be different from the kinetic temperature). Scott & Rees (1990) [17] revisited the subject, now in the modern context of galaxy formation in a Universe dominated by cold dark matter; they were the first to note that 21-cm cosmology could probe reionization. Madau et al. (1997) [11] first considered 21-cm radiation during cosmic dawn, before the epoch of reionization (EOR)[1], and highlighted the eras of Lyα coupling and of early cosmic heating. However, 21-cm cosmology was relatively slow in developing. For example, in a major review of the field in 2001 [23], we devoted 3 pages out of 114 to this topic, which at the time was considered only one of many promising avenues in the field.

---

[1] Two notes on common usage: The era/epoch of reionization is often denoted "EOR" in the literature; "Cosmic dawn" usually refers to the period between the formation of the first stars until the beginning of the EOR, although sometimes it is used as a general name for the entire period including the EOR.

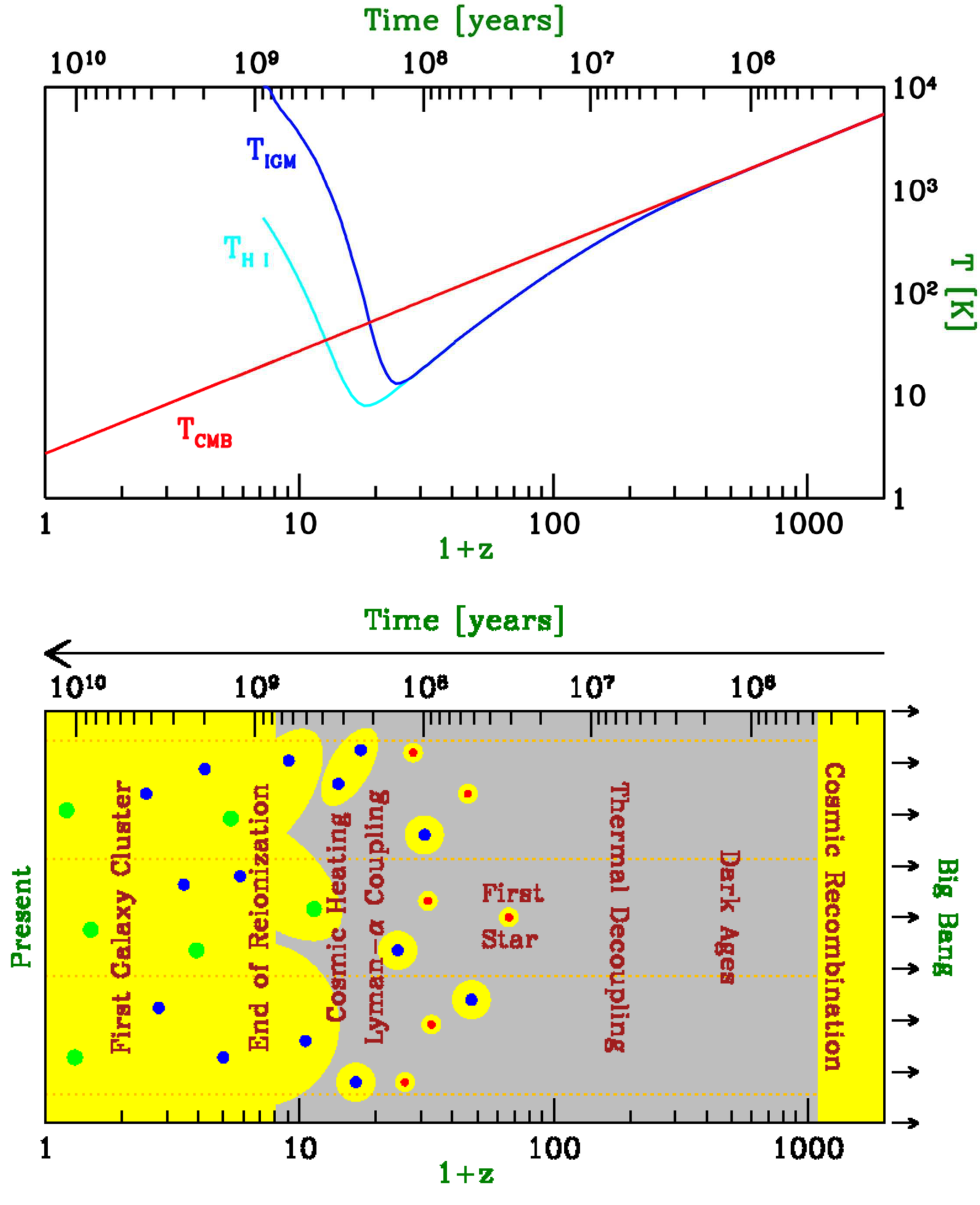

Figure 2: Overview of cosmic history, with the age of the universe shown on the top axis and the corresponding redshift (plus one) on the bottom axis. Bottom panel: Yellow represents ionized hydrogen, and gray is neutral. Observers probe the cosmic gas using the absorption of background light (dotted lines) by atomic hydrogen. Stars formed in halos whose typical size continually grew with time, going from the first generation that formed through molecular-hydrogen cooling (red dots), to the larger galaxies that formed through atomic cooling and likely dominated cosmic reionization (blue dots), all the way to galaxies as massive as the Milky Way (green dots). Top panel: Corresponding sketch of the cosmic mean gas temperature within the IGM, including neutral regions only (cyan) or also ionized regions (blue) assumed to be at 10,000 K; these are compared to the CMB temperature (red curve). The gas was initially thermally coupled to the CMB, until it adiabatically cooled more rapidly, and was then heated first by X-ray heating and subsequently by reionization. Updated and expanded version of a Figure from [21].

Cosmic reionization remained the dominant subject in the field of the first stars for some time longer. After several years of confusion about the basic character of reionization (see § 6.1), the now-standard paradigm was developed. Barkana & Loeb (2004) [24] showed that the surprisingly strong clustering of high-redshift halos leads to large ionized bubbles due to groups of clustered galaxies, and to an inside-out topology (with high-density regions reionizing early, leaving the voids for last). Furlanetto et al. (2004) [25] created a quantitative analytical model that yielded the first predictions of the distribution of H II bubble sizes, showing that 10 Mpc (comoving) is typical for the central stage of reionization. This picture of reionization based on semi-analytic models [24, 25] was then verified by large-scale numerical simulations, starting with Iliev et al. (2006) [26]. The theoretical expectation that the bubbles of reionization were large provides a critical boost for observational efforts to discover the resulting 21-cm fluctuations, since if higher angular resolution were required, this would make it harder to reach the sensitivity needed to detect the faint cosmic signal.

Cosmic reionization was initially thought to be the only source of fluctuations available for 21-cm interferometers (other than primordial density and temperature fluctuations, which create significantly smaller signals than those driven by galaxies and their strongly enhanced clustering). The earlier 21-cm events of cosmic dawn pointed out by Madau et al. (1997) [11] were considered to be highly uniform, since the photons that drove them (Lyα photons in the case of Lyα coupling, and X-ray photons in the case of cosmic heating) can travel $\sim$ 100 Mpc in a neutral Universe before interacting. Unless rare objects such as quasars dominated, this seemed to imply a uniform cosmic transition that could only be seen with global 21-cm measurements that track the sky-averaged spectrum [27]. Cosmic dawn was opened up to interferometric observations when Barkana & Loeb (2005) [28] applied the same idea of unusually large fluctuations in the abundance of early galaxies, which had helped understand reionization, to earlier epochs. Spatial fluctuations in the Lyα intensity were shown to have led, in fact, to rather large 21-cm fluctuations from the Lyα coupling era. The same idea was then applied by Pritchard & Furlanetto (2007) [29] to the X-ray background during the cosmic heating era, showing that a large signal of 21-cm fluctuations should be expected in this case as well.

The entire story of 21-cm cosmology as described thus far is at the moment purely theoretical, but a great international effort is underway to open up the observational field of 21-cm cosmology. In particular, several arrays of low-frequency radio telescopes have been constructed (and are now operating) in order to detect the 21-cm fluctuations from cosmic reionization (and beyond). Current efforts include the Murchison Wide-field Array (MWA [30]), the Low Frequency Array (LOFAR [31]), the Giant Metrewave Radio Telescope (GMRT [32]), and the Precision Array to Probe the Epoch of Reionization (PAPER [33]), and future plans have been made for the Hydrogen Epoch of Reionization Array (HERA; http://reionization.org/) and the Square Kilometer Array (SKA; https://www.skatelescope.org/); a 21-cm cosmology pathfinder of the latter is the New Extension in Nançay Upgrading LOFAR (NenuFAR). Although the expected foregrounds (dominated by Galactic synchrotron) are much brighter than the 21-cm signal, they are not expected to include sharp spectral features. Thus, although ongoing experiments are expected to yield noisy maps, the prospects for extraction of the 21-cm signal (and from it the reionization history) are quite promising, using the key statistic of the 21-cm power spectrum [34, 35, 36] as well as other statistics [37, 38, 39, 40]. A different approach is to measure the total sky spectrum and detect the global reionization signal arising from the overall disappearance of atomic hydrogen [41, 42, 43]; current and future efforts (some also targeting eras before reionization) include the Experiment to Detect the Global EOR Step (EDGES [44]), the Large Aperture Experiment to Detect the Dark Ages (LEDA; http://www.cfa.harvard.edu/LEDA/), and the lunar-orbiting Dark

Ages Radio Explorer (DARE; http://lunar.colorado.edu/dare/).

A novel effect that was only noticed recently is the supersonic velocity difference between the gas and the dark matter [45]. This intriguing effect (often called the streaming velocity) is predicted to have influenced the gas distribution at high redshift as well as early galactic halos.

The plan for this review is to first lay out the theoretical groundwork for galaxy formation, in general (§ 2) and at high redshift in particular (§ 3), followed by the basics of 21-cm cosmology (§ 4). What follows is a detailed discussion of the velocity streaming effect and its consequences (§ 5). Next, we discuss in detail the milestones of radiative feedback during early cosmic history (§ 6), and then outline their 21-cm signatures (§ 7). Finally, we summarize the review and conclude with a general outlook on the field (§ 8).

2. Galaxy Formation: Basic Theory

The fundamental theoretical understanding of galaxy formation as related to the earliest generations of stars and galaxies has been reviewed extensively [23, 46]. Here we provide a brief updated overview and summarize some useful results and formulas.

2.1. Cosmological background

In General Relativity, the metric for a space which is spatially homogeneous and isotropic is the Robertson-Walker metric, which can be written in the form

$$ds^2 = dt^2 - a^2(t)\left[\frac{dR^2}{1-kR^2} + R^2\left(d\theta^2 + \sin^2\theta\, d\phi^2\right)\right] , \qquad (1)$$

where $a(t)$ is the cosmic scale factor which describes expansion in time, and (R, θ, ϕ) are spherical comoving coordinates. The constant k determines the geometry of the metric; it is positive in a closed universe, zero in a flat universe, and negative in an open universe. Observers at rest at fixed (R, θ, ϕ) remain at rest, with their physical separation increasing with time in proportion to $a(t)$. A given observer sees a nearby observer at physical distance D receding at the Hubble velocity $H(t)D$, where the Hubble constant at time t is $H(t) = d\ln a(t)/dt$. Light emitted by a source at time t is observed at the present time t_0 with a redshift $z = 1/a(t) - 1$, where we set $a(t_0) \equiv 1$.

The Einstein field equations of General Relativity yield the Friedmann equation [47, 48]

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

which relates the expansion of the universe to its matter-energy content. For each component of the energy density ρ, with an equation of state $p = p(\rho)$, the density ρ varies with $a(t)$ according to the equation of energy conservation

$$d(\rho R^3) = -p d(R^3) . \qquad (3)$$

With the critical density

$$\rho_C(t) \equiv \frac{3H^2(t)}{8\pi G} \qquad (4)$$

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

$$\Omega \equiv \frac{\rho}{\rho_C} . \qquad (5)$$

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

$$\frac{H(t)}{H_0} = \left[\frac{\Omega_m}{a^3} + \Omega_\Lambda + \frac{\Omega_r}{a^4} + \frac{\Omega_k}{a^2} \right]^{1/2} , \tag{6}$$

where we define H_0 and $\Omega_0 = \Omega_m + \Omega_\Lambda + \Omega_r$ to be the present values of H and Ω, respectively, and we let

$$\Omega_k \equiv -\frac{k}{H_0^2} = 1 - \Omega_0 \ . \tag{7}$$

In the particularly simple Einstein-de Sitter (EdS) model ($\Omega_m = 1$, $\Omega_\Lambda = \Omega_r = \Omega_k = 0$), the scale factor varies as $a(t) \propto t^{2/3}$. Even models with non-zero Ω_Λ or Ω_k approach the Einstein-de Sitter behavior at high redshifts, i.e., when

$$(1+z) \gg \max \left[\Omega_k/\Omega_m, (\Omega_\Lambda/\Omega_m)^{1/3} \right] , \tag{8}$$

as long as we do not reach extremely early times at which Ω_r cannot be neglected. The approach to EdS is particularly rapid in practice given that current observations imply $\Omega_k \approx 0$. In this EdS regime (which we will also refer to as the high-z regime), $H(t) \approx 2/(3t)$.

We assume the best-fit cosmological parameters of the ΛCDM (cosmological constant Λ plus cold dark matter) model, based on the first-year data of the Planck satellite [49]: $h = 0.678$ (where the present Hubble constant is $H_0 = 100\,h$ km s^{-1} Mpc^{-1}), a flat Universe with matter density parameter $\Omega_m = 0.307$ and the rest (adding up to unity) in a cosmological constant, and $\Omega_b = 0.0482$. In convenient units, the comoving cosmic mean density of matter in the Universe is:

$$\bar{\rho}_m = 3.91 \times 10^{10} \left(\frac{\Omega_m h^2}{0.141} \right) \frac{M_\odot}{\text{Mpc}^3} , \tag{9}$$

where the physical density at redshift z is higher by a factor of $(1+z)^3$. Also, in the high-z regime,

$$H(z) \approx H_0 \frac{\sqrt{\Omega_m}}{a^{3/2}} , \tag{10}$$

and the age of the universe is

$$t \approx \frac{2}{3\,H_0\,\sqrt{\Omega_m}} (1+z)^{-3/2} = 5.49 \times 10^8 \left(\frac{\Omega_m h^2}{0.141} \right)^{-1/2} \left(\frac{1+z}{10} \right)^{-3/2} \text{yr} \ . \tag{11}$$

Another useful quantity is the comoving (or particle) horizon, the maximum distance from which light could have traveled to an observer in the age of the universe. It thus represents a causal horizon for physical influences (other than gravity). Its general expression, and its approximate value in the high-redshift regime defined above (not including an early period of inflation), are:

$$\eta = \int_0^t \frac{c\,dt'}{a(t')} \approx 5.05 \left(\frac{\Omega_m h^2}{0.141} \right)^{-1/2} \left(\frac{1+z}{10} \right)^{-1/2} \text{Gpc} \ . \tag{12}$$

We note that cosmologists often explicitly take out the Hubble constant in expressions, e.g., distances in cosmology are expressed in units of h^{-1} Mpc (and wavenumbers in h Mpc^{-1}). This is to some degree a remnant of an earlier time when h was uncertain by nearly a factor of two. Now that h has been determined to equal 0.7 up to a few percent, it may be preferable to simply use units of Mpc in papers, and specify the assumed h in case it is needed for precise comparisons.

In the standard hot Big Bang model, the universe is initially hot and the energy density is dominated by radiation. The transition to matter domination occurs at $z \sim 3400$, but the universe remains hot enough that the gas is ionized, and electron-photon scattering effectively couples the baryonic matter and the radiation. At $z \sim 1100$ the temperature drops below ~ 3000 K and protons and electrons recombine to form neutral hydrogen. The photons then decouple and travel freely until the present, when they are observed as the CMB.

2.2. Linear perturbation theory

Observations of the CMB (e.g., [49]) show that the universe at recombination was extremely uniform, but with large-scale fluctuations in the energy density and gravitational potential of roughly one part in 10^5. Such small fluctuations, generated in the early universe, grew over time due to gravitational instability, and eventually led to the formation of galaxies and the large-scale structure observed in the present universe.

We distinguish here between physical/proper and comoving coordinates. Using vector notation, the physical coordinate $\mathbf{r}$ corresponds to a comoving position $\mathbf{x} = \mathbf{r}/a$. In a homogeneous universe with density ρ, we describe the cosmological expansion in terms of an ideal pressure-less fluid of particles each of which is at fixed $\mathbf{x}$, expanding with the Hubble flow $\mathbf{v} = H(t)\mathbf{r}$ where $\mathbf{v} = d\mathbf{r}/dt$. Onto this uniform expansion we impose small perturbations, given by a relative density perturbation

$$\delta(\mathbf{x}) \equiv \frac{\rho(\mathbf{r})}{\bar{\rho}} - 1 \, , \tag{13}$$

where the mean fluid density is $\bar{\rho}$, with a corresponding peculiar velocity $\mathbf{u} \equiv \mathbf{v} - H\mathbf{r}$. Then the fluid is described by the continuity and Euler equations in comoving coordinates [50, 51]. The gravitational potential ϕ is given in turn by the Poisson equation, in terms of the density perturbation. This fluid description is valid for describing the evolution of collisionless cold dark matter particles until different particle streams cross. After such "shell-crossing", the individual particle trajectories must in general be followed, but this typically occurs only after perturbations have grown to become non-linear. Similarly, baryons can be described as a pressure-less fluid as long as their temperature is negligibly small, but non-linear collapse leads to the formation of shocks in the gas.

For small perturbations $\delta \ll 1$, the fluid equations can be linearized and combined to yield

$$\frac{\partial^2 \delta}{\partial t^2} + 2H \frac{\partial \delta}{\partial t} = 4\pi G \bar{\rho} \delta \, . \tag{14}$$

This linear equation has in general two independent solutions, the so-called growing and decaying modes. Starting with random initial conditions, the growing mode comes to dominate the density evolution. Thus, until it becomes non-linear, the density perturbation maintains its shape in comoving coordinates and grows in proportion to a growth factor $D(t)$. In the Einstein-de Sitter model (or, at high redshift, in other models as well) the growth factor is simply proportional to $a(t)$. Given the standard normalization of $D(a = 1) = 1$, in EdS we would simply have $D(a) = a$, while in ΛCDM with our standard parameters, at high redshift we have $D(a) \approx 1.28a$. In other words, the recent dominance by the cosmological constant in ΛCDM suppresses the linear growth of structure down to the present by a factor of 1.28 compared to a universe that continued to follow the EdS model.

More generally, there is also a decaying mode that declines with time rapidly after matter-radiation equality, as $a^{-3/2}$ in EdS. More importantly, in the presence of baryons, the difference between the distribution of dark matter and baryons persists for much longer (see § 2.4).

The spatial form of the initial density fluctuations can be described in Fourier space, in terms of Fourier components $\delta_{\mathbf{k}}$, where

$$\delta_{\mathbf{k}} = \int d^3x\, \delta(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}} \; ; \quad \delta(\mathbf{x}) = \int \frac{d^3k}{(2\pi)^3}\, \delta_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}} \; . \tag{15}$$

We note that the $(2\pi)^3$ factor is sometimes switched (or split) between these two equations, so care must be taken when comparing results that use different conventions for this factor within the definitions of the Fourier transform and its inverse. Here we have introduced the comoving wavevector $\mathbf{k}$, whose magnitude k is the comoving wavenumber which is equal to 2π divided by the wavelength.

The Fourier description is particularly simple for fluctuations generated by cosmic inflation [48]. Inflation generates perturbations given by a Gaussian random field, in which different $\mathbf{k}$-modes are statistically independent, each with a random phase. The statistical properties of the fluctuations are determined by the variance of the different $\mathbf{k}$-modes, and the variance is described in terms of the power spectrum $P(k)$ as follows:

$$\langle \delta_{\mathbf{k}} \delta^*_{\mathbf{k}'} \rangle = (2\pi)^3\, P(k)\, \delta_D^{(3)}\left(\mathbf{k} - \mathbf{k}'\right) \; , \tag{16}$$

where $\delta_D^{(3)}$ is the three-dimensional Dirac delta function. Note that $P(k)$ has units of volume, or, more generally, the power spectrum of some quantity has units of volume times the square of the units of that quantity.

In standard models, inflation produces a primordial power-law spectrum $P(k) \propto k^n$ with $n \sim 1$. Perturbation growth in the radiation-dominated and then matter-dominated universe results in a modified final power spectrum, characterized by a turnover at a scale of order the horizon cH^{-1} at matter-radiation equality, and a small-scale asymptotic shape of $P(k) \propto k^{n-4}$. On large scales the power spectrum evolves in proportion to the square of the growth factor, and this simple evolution is termed linear evolution. On small scales, the power spectrum changes shape due to the additional non-linear gravitational growth of perturbations, yielding the non-linear (also called “full”) power spectrum. The overall amplitude of the power spectrum is not specified by current models of inflation, and it is usually set observationally using the CMB temperature fluctuations or local measures of large-scale structure.

Since density fluctuations may exist on all scales, in order to determine the formation of objects of a given size or mass it is useful to consider the statistical distribution of the smoothed density field. Using a window function $W(\mathbf{y})$ normalized so that $\int d^3\mathbf{y}\, W(\mathbf{y}) = 1$, the smoothed density perturbation field, $\int d^3\mathbf{y}\delta(\mathbf{x}+\mathbf{y})W(\mathbf{y})$, itself follows a Gaussian distribution with zero mean. For the particular choice of a spherical top-hat, in which W is constant within a sphere of radius R and is zero outside it, the smoothed perturbation field measures the fluctuations in the mass M in spheres of radius R. Indeed, a halo of mass M forms out of an initial (i.e., when $\delta \to 0$) region of comoving radius R, where

$$M = \frac{4}{3}\pi\bar{\rho}_0 R^3 = 1.64 \times 10^8 \left(\frac{\Omega_m h^2}{0.141}\right)\left(\frac{R}{100\text{ kpc}}\right)^3 M_\odot \; . \tag{17}$$

The inverse relation is:

$$R = 84.8 \left(\frac{\Omega_m h^2}{0.141} \right)^{-1/3} \left(\frac{M}{10^8 M_\odot} \right)^{1/3} \text{kpc} \, . \tag{18}$$

The normalization of the present power spectrum is often specified by the value of $\sigma_8 \equiv \sigma(R = 8h^{-1}\text{Mpc})$. For the top-hat, the smoothed perturbation field is denoted δ_R or δ_M (in reference to the equivalent mass M). The variance $\langle \delta_M \rangle^2$ is

$$S(M) = \sigma^2(M) = \sigma^2(R) = \int_0^\infty \frac{dk}{2\pi^2}\, k^2 P(k) \left[\frac{3 j_1(kR)}{kR} \right]^2 \, , \tag{19}$$

where $j_1(x) = (\sin x - x \cos x)/x^2$. The function $\sigma(M)$ plays a crucial role in estimates of the abundance of collapsed objects, as described below. We note that Eq. 19 in the limit of no smoothing (i.e., $R \to 0$) shows that the contribution of power at wavenumber k per $\log k$ to the variance at a point is $k^3 P(k)/(2\pi^2)$. The relative (dimensionless) fluctuation level at k is defined as the root mean square, i.e., the square root of this contribution to the variance, and in 21-cm cosmology, if the cosmic mean 21-cm brightness temperature at some redshift is $\langle T_\mathrm{b} \rangle$, then the 21-cm fluctuation level δT_b at k (usually in units of mK) is defined as:

$$\delta T_\mathrm{b} = \langle T_\mathrm{b} \rangle \sqrt{\frac{k^3 P(k)}{2\pi^2}} \, , \tag{20}$$

in terms of the dimensionless 21-cm power spectrum $P(k)$ (i.e., the power spectrum of relative fluctuations in the 21-cm brightness temperature).

2.3. Non-linear collapse

The small density fluctuations evidenced in the CMB grow over time as described in the previous subsection, until the perturbation δ becomes of order unity, and the full non-linear gravitational problem must be considered. The dynamical collapse of a dark matter halo can be solved analytically only in cases of particular symmetry, with the simplest case being that of spherical symmetry. Although this model is restricted in its direct applicability, the results of spherical collapse have turned out to be surprisingly useful in understanding the properties and distribution of halos in models based on cold dark matter.

In spherical collapse, at the moment when the top-hat collapses to a point, the extrapolated overdensity as predicted by linear theory is [50] $\delta_L = 1.686$ in the Einstein-de Sitter model, with a weak dependence on Ω_m and Ω_Λ in the more general case. Thus, a top-hat collapses at redshift z if its linear overdensity extrapolated to the present day (also termed the critical density of collapse) is

$$\delta_\mathrm{crit}(z) = \frac{1.686}{D(z)} \, , \tag{21}$$

where we again set $D(z = 0) = 1$.

Even a slight violation of the exact symmetry of the initial perturbation can prevent the top-hat from collapsing to a point. Instead, the halo reaches a state of virial equilibrium by violent relaxation (phase mixing). Using the virial theorem $U = -2K$ to relate the potential energy U to the kinetic energy K in the final state, the final overdensity relative to the critical density at the collapse redshift is found to be $\Delta_c = 18\pi^2 \simeq 178$ in the Einstein-de Sitter model. This theoretical value is slightly modified in ΛCDM, but conventionally the EdS value (or even the rougher value of 200) is often used to *define* the virial radius $r_{\rm vir}$ and the virial masses of halos in numerical simulations and in analyses of observations.

Quantitatively, a halo of mass M collapsing at redshift z (assumed high enough for the EdS limit) thus has a physical virial radius

$$r_{\rm vir} = 1.51 \left(\frac{\Omega_m h^2}{0.141}\right)^{-1/3} \left(\frac{M}{10^8\ M_\odot}\right)^{1/3} \left(\frac{\Delta_c}{18\pi^2}\right)^{-1/3} \left(\frac{1+z}{10}\right)^{-1}\ {\rm kpc}\ , \tag{22}$$

a corresponding circular velocity,

$$V_c = \left(\frac{GM}{r_{\rm vir}}\right)^{1/2} = 16.9 \left(\frac{\Omega_m h^2}{0.141}\right)^{1/6} \left(\frac{M}{10^8\ M_\odot}\right)^{1/3} \left(\frac{\Delta_c}{18\pi^2}\right)^{1/6} \left(\frac{1+z}{10}\right)^{1/2}\ {\rm km\ s^{-1}}\ , \tag{23}$$

and a virial temperature

$$T_{\rm vir} = \frac{\mu m_p V_c^2}{2k_{\rm B}} = 1.03\times 10^4 \left(\frac{\Omega_m h^2}{0.141}\right)^{1/3} \left(\frac{\mu}{0.6}\right) \left(\frac{M}{10^8\ M_\odot}\right)^{2/3} \left(\frac{\Delta_c}{18\pi^2}\right)^{1/3} \left(\frac{1+z}{10}\right)\ {\rm K}\ , \tag{24}$$

where μ is the mean molecular weight in units of the proton mass m_p. Note that the value of μ depends on the ionization state of the gas; $\mu = 0.59$ for a fully ionized primordial gas, $\mu = 0.61$ for a gas with ionized hydrogen but only singly-ionized helium, and $\mu = 1.22$ for neutral primordial gas.

Although spherical collapse captures some of the physics governing the formation of halos, structure formation in cold dark matter models proceeds hierarchically. At early times, most of the dark matter is in low-mass halos, and these halos continuously accrete and merge to form high-mass halos. Numerical simulations of hierarchical halo formation indicate a roughly universal spherically-averaged density profile for the resulting halos (Navarro, Frenk, & White 1997 [52], hereafter NFW), though with considerable scatter among different halos. The NFW profile has the form

$$\rho(r) \propto \frac{1}{c_{\rm N} x (1 + c_{\rm N} x)^2}\ , \tag{25}$$

where $x = r/r_{\rm vir}$ and $c_{\rm N}$ is the concentration parameter.

In addition to characterizing the properties of individual halos, a critical prediction of any theory of structure formation is the abundance of halos, i.e., the number density of halos as a function of mass, at any redshift. While the number density of halos can be measured for particular cosmologies in numerical simulations, an analytical model helps us gain physical understanding, can be used to explore the dependence of the halo abundance on the cosmological parameters, and can be extrapolated to regimes that cannot be reached by current simulations. It is also a starting point towards building models for the abundances of galaxies and galaxy clusters.

A simple analytical model that has become the basis for work in this field was developed by Press & Schechter (1974) [53]. The model is based on the ideas of a Gaussian random field of density perturbations, linear gravitational growth, and spherical collapse. To determine the abundance of halos at a redshift z, in this model we use δ_M, the density field smoothed on a mass scale M, as defined in the previous subsection. Although the model is based on the initial conditions, it is usually expressed in terms of redshift-zero quantities. Thus, we use the linearly-extrapolated density field, i.e., the initial density field at high redshift extrapolated to the present by simple multiplication by the relative linear growth factor. A useful entity is the 'present power spectrum', which refers to the initial power spectrum, linearly-extrapolated to the present without including non-linear evolution. Since δ_M is distributed as a Gaussian variable with zero mean and standard deviation $\sigma(M)$ [which is determined by Eq. (19) from the present power spectrum], the probability that δ_M is greater than some δ equals $(1/2)\,\mathrm{erfc}\left[\delta/\left(\sqrt{2}\,\sigma\right)\right]$. The fundamental ansatz is to identify this probability with the fraction of dark matter particles that are part of collapsed halos of mass greater than M, at redshift z. There are two additional ingredients: First, the value used for δ is $\delta_{\mathrm{crit}}(z)$ given in Eq. (21), which is the critical density of collapse found for a spherical top-hat (extrapolated to the present since $\sigma(M)$ is calculated using the present power spectrum); and second, the fraction of dark matter in halos above M is multiplied by an additional factor of 2 in order to ensure that every particle ends up as part of some halo with $M > 0$. Thus, the final formula for the mass fraction in halos above M at redshift z is

$$F(>M|z) = \mathrm{erfc}\left[\frac{\delta_{\mathrm{crit}}(z)}{\sqrt{2}\,\sigma(M)}\right] . \tag{26}$$

This ad-hoc factor of 2 is necessary, since otherwise only positive fluctuations of δ_M would be included. Bond et al. (1991) [54] found a more satisfactory derivation of this correction factor, using a different ansatz. In their derivation, the factor of 2 originates from the so-called "cloud-in-cloud" problem: For a given mass M, even if δ_M is smaller than $\delta_{\mathrm{crit}}(z)$, it is possible that the corresponding region lies inside a region of some larger mass $M_L > M$, with $\delta_{M_L} > \delta_{\mathrm{crit}}(z)$. In this case the original region should be counted as belonging to a halo of mass M_L. Bond et al. (1991) [54] showed that, under certain assumptions, the additional contribution results precisely in a factor of 2 correction. We note that this work is the basis of the extended Press-Schechter model, which is mentioned later in this review.

The halo abundance (or halo mass function) is

$$\frac{dn}{dM} = \frac{\bar{\rho}_m}{M}\left|\frac{dS}{dM}\right| f(\delta_{\mathrm{crit}}(z), S)\ , \tag{27}$$

where dn is the comoving number density of halos with masses in the range M to $M + dM$, $S = \sigma^2(M)$ is the variance on scale M, and $f(\delta_{\mathrm{crit}}(z), S)\,dS$ is defined to be the mass fraction contained at z within halos with mass in the range corresponding to S to $S + dS$. In the Press-Schechter model [53],

$$f_{\mathrm{PS}}(\delta_{\mathrm{crit}}(z), S) = \frac{1}{\sqrt{2\pi}}\,\frac{\nu}{S}\,\exp\left[-\frac{\nu^2}{2}\right] , \tag{28}$$

where $\nu = \delta_{\mathrm{crit}}(z)/\sqrt{S}$ is the number of standard deviations that the critical collapse overdensity at z represents on the mass scale M corresponding to the variance S.

The classic Press-Schechter [53] model for the halo mass function fits numerical simulations only roughly, and in particular it substantially underestimates the abundance of the rarest, most massive halos. The halo mass function of Sheth & Tormen (1999) [55], with modified best-fit parameters [56], fits numerical simulations much more accurately [57]. It is given by:

$$f_{\mathrm{ST}}(\delta_{\mathrm{crit}}(z), S) = A'\,\frac{\nu}{S}\,\sqrt{\frac{a'}{2\pi}}\left[1 + \frac{1}{(a'\nu^2)^{q'}}\right]\exp\left[-\frac{a'\nu^2}{2}\right] , \tag{29}$$

with best-fit parameters $a' = 0.75$ and $q' = 0.3$, and where normalization to unity is ensured by taking $A' = 0.322$.

In addition to the overall, mean abundance of halos, another key question in cosmology and galaxy formation is the spatial distribution of the halo number density. In particular, since halos form due to gravity, massive halos should form in larger numbers in regions of high overall density than in low-density voids. Thus, density fluctuations are expected to lead to spatial fluctuations in the halo number density. This leads to the concept of halo (or galaxy) bias, a now-standard concept in galaxy formation [53, 58, 59, 60, 54, 61]. Particularly simple is the case of linear bias, i.e., when the distribution of galaxies is a proportionally amplified ("biased") version of that of the underlying density of matter. Mathematically this means that the relative fluctuations in the number density of galactic halos ($\delta_{\rm g}$) are proportional to the relative fluctuations in the underlying density of matter δ:

$$\delta_{\rm g} = b\,\delta\ , \tag{30}$$

where b is the linear bias factor (or simply "the bias").

This simple result is expected to be reasonably accurate when looking at fluctuations on large scales s (usually tens of comoving Mpc or more). Several conditions must be satisfied for this to be true. The scale s must be much larger than the spatial scales involved in forming the individual galactic halos whose clustering is being considered; this allows a separation of scales that is the basis of a simple approximation called a peak-background split [60]. Also, in order to avoid non-linear effects, the scale s must be large enough that typically $\delta \ll 1$, i.e., the variance is small when the density field is averaged on the scale s. Similarly, $\delta_{\rm g} \ll 1$ on that scale is advisable. Finally, gravity must dominate galaxy formation, or at least, any other effects (such as astrophysical feedbacks) must operate on much smaller scales than s. Of these conditions, the first two tend to be more favorable at high redshifts, since the galaxies are typically small and thus associated with small formation scales, and density fluctuations on large scales are still relatively small. However, the last two conditions become more problematic, as discussed in great detail in the rest of this review. High-redshift galaxies are highly biased, so their fluctuations are much larger than those in the underlying density (section 3.1); and since early galaxies were typically small, they were susceptible to a variety of external feedbacks that operate on scales of order 100 Mpc, including the supersonic streaming velocity (section 5) and various radiative feedbacks (section 6).

2.4. Baryons: linear evolution, pressure, and cooling

Baryons play a major role in cosmology. On the largest scales, their coupling to the photons in the early universe leaves them clustered differently from the dark matter, with the difference decaying away only gradually. On smaller scales, the baryonic pressure suppresses gravitational growth. Most directly, of course, the baryons are important since stars form out of baryons that cool and collapse to high density. To get started, we note a useful number: the cosmic mean number density of hydrogen (including both neutral and ionized forms) is

$$\bar{n}_H(z) = 1.89 \times 10^{-7} \left(\frac{\Omega_b h^2}{0.0221} \right) (1+z)^3 \ {\rm cm}^{-3}\ , \tag{31}$$

assuming that 76% of the baryon mass density is in hydrogen. The number density of helium is smaller by a factor of 12.7.

As noted in § 2.2, in the presence of dark matter only, the linear perturbations are dominated by a growing mode that is $\propto a$ in EdS, as the decaying mode drops rapidly, $\propto a^{-3/2}$ in EdS. On large scales, baryons also respond to gravity only (after cosmic recombination), but their initial conditions are different, as their strong coupling to the CMB photons up to recombination suppresses their sub-horizon fluctuations. Thus, cosmic recombination begins a period of baryonic infall, during which the baryons gradually catch up with the dark matter perturbations [50]. Specifically, if we denote the perturbations of the dark matter and baryon density $\delta_{\rm dm}$ and $\delta_{\rm b}$, respectively, and their mass fractions within the total matter density $f_{\rm dm} = (\Omega_m - \Omega_b)/\Omega_m$ and $f_{\rm b} = \Omega_b/\Omega_m$, then it is useful to work with the perturbation of the total density, $\delta_{\rm tot} = f_{\rm dm}\delta_{\rm dm} + f_{\rm b}\delta_{\rm b}$, and the difference $\delta_{\rm diff} = \delta_{\rm b} - \delta_{\rm tot}$. In linear perturbation theory, $\delta_{\rm tot}$ has the usual growing and decaying modes (i.e., $\propto a$ and $\propto a^{-3/2}$ in EdS), while the two solutions for $\delta_{\rm diff}$ are constant ($\propto 1$) and $\propto a^{-1/2}$ in EdS. Thus, the baryon perturbation $\delta_{\rm b}$ approaches $\delta_{\rm tot}$ gradually from below. This approach can be described through their relative difference. If we approximately include only the dominant modes, this key quantity decays as [62]

$$r_{\rm LSS} \equiv \frac{\delta_{\rm diff}}{\delta_{\rm tot}} \approx -\frac{0.3\%}{a}\ . \tag{32}$$

This decay is slow enough that the gradual baryonic infall is in principle observable in high-redshift 21-cm measurements [63] and perhaps also in the distribution of galaxies at low redshift [64, 65].

During this era of baryonic infall, and before cosmic heating from radiative astrophysical sources, the gas cooled adiabatically with the expansion. Traditional calculations [50, 66, 67] assumed a uniform speed of sound for the gas at each redshift, but a more careful consideration of the combination of adiabatic cooling and Compton heating substantially modifies the temperature perturbations on all scales [68, 69, 63, 70].

On small scales, the density evolution is no longer purely gravitational, as the gas pressure suppresses baryon perturbations. The relative force balance at a given time can be characterized by the Jeans scale, which is the minimum scale at which a small perturbation will grow due to gravity overcoming pressure gradients. If the gas has a uniform sound speed $c_{\rm s}$, then the comoving Jeans wavenumber is

$$k_{\rm J} = \frac{a}{c_{\rm s}} \sqrt{4\pi G \bar{\rho}_m}\ . \tag{33}$$

In the simple limit where the gas cools adiabatically (after thermally decoupling from the CMB at $z \sim 150$), this gives a characteristic Jeans mass (defined in terms of a sphere of diameter equal to the Jeans wavelength) [23]

$$M_{\rm J} \equiv \frac{4\pi}{3} \left(\frac{\pi}{k_{\rm J}}\right)^3 \bar{\rho}_m = 5.89 \times 10^3 \left(\frac{\Omega_m h^2}{0.141}\right)^{-1/2} \left(\frac{\Omega_b h^2}{0.0221}\right)^{-3/5} \left(\frac{1+z}{10}\right)^{3/2}\ M_\odot\ . \tag{34}$$

The Jeans mass, however, is not the whole story, since it is related only to the evolution of perturbations at a given time. When the Jeans mass itself varies with time, the overall suppression of the growth of perturbations depends on a time-averaged Jeans mass, the filtering mass [71]. To define it, we start from the regime of large-scale structure (i.e., scales too large to be affected by pressure, but much smaller than the horizon and the scale of baryon acoustic oscillations), where, as noted above, $r_{\rm LSS}$ does not depend on k, and is simply a function of redshift. On smaller scales, the next-order term describing the difference between the baryons and dark matter is the k^2 term [71], and the filtering wavenumber $k_{\rm F}$ and corresponding mass scale $M_{\rm F}$ are defined through [62]

$$\frac{\delta_{\rm b}}{\delta_{\rm tot}} = 1 + r_{\rm LSS} - \frac{k^2}{k_{\rm F}^2}\ ; \quad M_{\rm F} \equiv \frac{4\pi}{3} \left(\frac{\pi}{k_{\rm F}}\right)^3 \bar{\rho}_m\ . \tag{35}$$

This filtering mass scale captures how the whole history of the evolving Jeans mass affects the final baryon perturbations that result at a given time. Starting at early times, since the baryon fluctuations are very small before cosmic recombination, the gas pressure (which depends on $\delta_{\rm b}$) starts out small, so the filtering mass starts from low values and rises with time up to a value of $\sim 3 \times 10^4 M_\odot$ [62] around redshift 30. It then drops due to the cooling cosmic gas, but the drop is very gradual (reaching $\sim 2 \times 10^4 M_\odot$ at $z = 10$ in the absence of cosmic heating or reionization) due to the remaining after-effects of the suppression of gas infall at higher redshifts. This behavior is significantly different from the Jeans mass, which declines rapidly with time (as in Eq. 34) and drops below $10^4 M_\odot$ at $z = 13$.

What makes the filtering mass even more useful is that it seems to offer in many situations a good estimate of the minimum halo mass that manages to accrete a significant amount of gas (e.g., 50% of the cosmic baryon fraction). It is natural to expect some relation between this characteristic, minimum halo mass and the filtering mass, since the gas fraction in a collapsing halo reflects the total amount of gas that was able to accumulate in the collapsing region during the entire, extended collapse process. For example, a sudden change in gas temperature immediately begins to affect gas motions (through the pressure-gradient force), but has only a gradual, time-integrated effect on the overall amount of gas in a given region. In this way, the minimum accreting mass is analogous to the linear-theory filtering mass. However, the former is defined within the deeply non-linear regime, so the two masses may not necessarily agree quantitatively.

Gnedin [72] first compared the filtering mass to the characteristic mass in numerical simulations, suggesting that they are approximately equal in the post-reionization universe in which the IGM is hot and ionized. However, he used a non-standard definition of the filtering mass that equals 8 times the standard definition given above. Subsequently, higher-resolution simulations did not find a clear relation between the theoretically calculated filtering mass and the characteristic mass measured in post-reionization simulations [73, 74]. However, the heating within simulations of inhomogeneous reionization is complex, and thus the filtering mass (which depends on the thermal history) is difficult to compute directly. The filtering mass has been shown to agree to within a factor of ~ 1.5 with the characteristic mass measured in simulations at higher redshifts, throughout the era prior to significant cosmic heating or reionization, as well as after a controlled, sudden heating [75, 76]. Thus, the issue of the possible usefulness of the filtering mass after reionization has not been settled, but alternative models have been recently proposed to fit results from post-reionization simulations [77, 78].

The conclusion is that prior to cosmic heating and reionization, gas is expected to accumulate significantly in dark matter minihalos down to a mass of $\sim 3 \times 10^4 M_\odot$ [76]. This minimum accretion mass later rises during cosmic heating and even more rapidly within ionized regions during cosmic reionization. In addition, even at the highest redshifts, the minimum mass is boosted in regions of significant streaming velocity (see § 5 below).

We end this section with a brief summary of cooling. Figure 3 shows the cooling curve for primordial gas, prior to metal enrichment. Primordial *atomic* gas can radiate energy only once hydrogen or helium are significantly ionized, so such cooling is limited to gas at temperatures above $\sim 10^4$ K. At high redshifts, most of the gas is in halos with relatively low masses, so that even if the accreted gas is shocked and heated to the virial temperature (Eq. 24), it is unable to cool. However, in the presence of even a small ionized hydrogen fraction, molecular hydrogen can acquire sufficient abundance to provide significant cooling [79], and its rotational and vibrational transitions allow cooling down to below 10^3 K. Further details about primordial gas cooling are reviewed elsewhere [23].

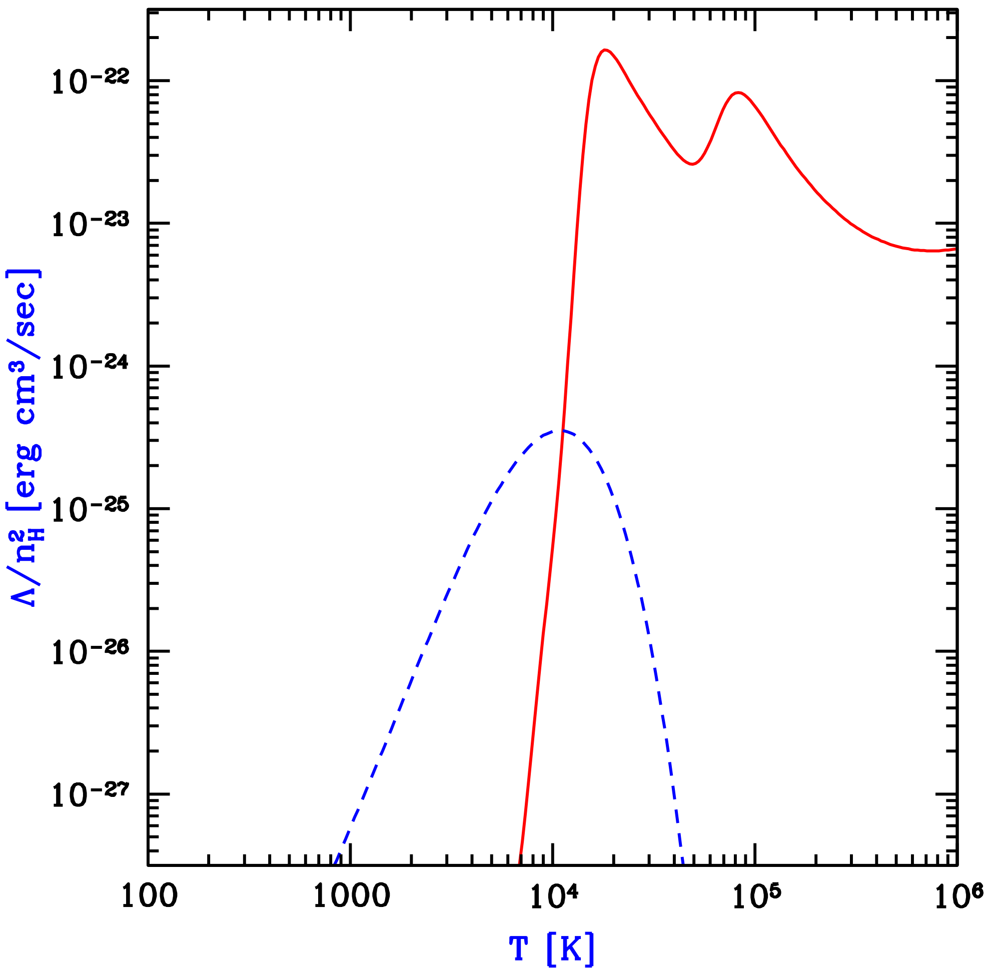

Figure 3: Cooling rates as a function of temperature for a primordial gas composed of atomic hydrogen and helium, as well as molecular hydrogen, in the absence of any metals or external radiation. The plotted quantity Λ/n_H^2 is roughly independent of density (unless $n_H \gg 10\ \mathrm{cm}^{-3}$), where Λ is the volume cooling rate (in erg/sec/cm^3). The solid line shows the cooling curve for an atomic gas, with the characteristic peaks due to collisional excitation of H I and He II. The dashed line shows the additional contribution of molecular cooling, assuming a molecular abundance equal to 0.1% of n_H. From [23].

3. Galaxy Formation: High-Redshift Highlights

In this section we expand on several features of galaxy formation that are particularly important at high redshifts. We first discuss fluctuations in the density of galaxies, which are generally important in cosmology, but at high redshift the fluctuations become unusually large and this has significant consequences that reverberate throughout this review. We then discuss various challenges of numerical simulations and approaches to deal with them. While simulations have become an indispensable tool in cosmology, it is important to bear in mind that they have fundamental limitations, some of them specific to, or worsening at, high redshifts. For example, while simulations at low redshift can be continually tested by and improved based on astronomical observations, this is not currently possible (or is at least far more limited) at high redshift. Finally, we discuss the formation of the very first stars, obviously a subject of great theoretical and numerical interest, hopefully with significant observational traces as well.

3.1. Large fluctuations in the galaxy number density

A broad, common thread runs through much of the recent theoretical development of cosmic reionization and 21-cm cosmology: The density of galaxies (or stars) varies spatially, with the fluctuations becoming surprisingly large at high redshift, even on quite large cosmological scales [24]. This can be understood from the standard theory of galaxy formation as due to the fact that the first galaxies represented rare peaks in the cosmic density field.

As an analogy, imagine searching on Earth for mountain peaks above 5000 meters. The 200 such peaks are not at all distributed uniformly but instead are found in a few distinct clusters on top of large mountain ranges. Similarly, in order to find the early galaxies, one must first locate a region with a large-scale density enhancement, and then galaxies will be found there in abundance. For a more detailed argument, note that galactic halos form roughly in regions where the (linearly extrapolated) density perturbation reaches above a fixed threshold value $\delta_{\rm crit}(z)$ (see section 2.3). Now, the total density at a point is the sum of contributions from density fluctuations on various scales (Figure 4). For initial perturbations from inflation (which follow the statistics of a Gaussian random field), the fluctuations on different scales are statistically independent. Thus, the same small-scale density fluctuations are added, in different regions, to various long-wavelength density fluctuations. In an over-dense region on large scales, the small-scale fluctuations only need to supply the missing amount needed to reach $\delta_{\rm crit}(z)$, while in a large-scale void, the same small-scale fluctuations must supply a total density of $\delta_{\rm crit}(z)$ plus the extra density missing within the void. This means that a larger fraction of the volume within the over-dense region will reach above $\delta_{\rm crit}(z)$ in total density, and thus more halos will form there. Now, at high redshift, when density fluctuations had not yet had time for much gravitational growth, the effective threshold value $\delta_{\rm crit}(z)$ is many times larger than the typical density fluctuation on the scales that form galactic halos. In other words, each halo represents a many-σ fluctuation. Under Gaussian statistics, the fraction of points above $t\sigma$ changes rapidly with t, once t is $2-3$ or higher. Thus, the abundance of halos in a given region changes rapidly with small changes of the mean density in the region (and this mean density is set by large-scale density modes). The density of star formation is thus expected to show strongly biased (i.e., amplified) fluctuations on large scales [24]. These large-scale fluctuations at high redshift, and their great observational importance, had for a long time been underestimated, in part because the limited range of scales available to numerical simulations put these fluctuations mostly out of their reach.

Figure 5 illustrates a further effect, which is that the limited box size of simulations leads to a delay of halo formation, or equivalently, an underestimate of the abundance of halos (and stars) at

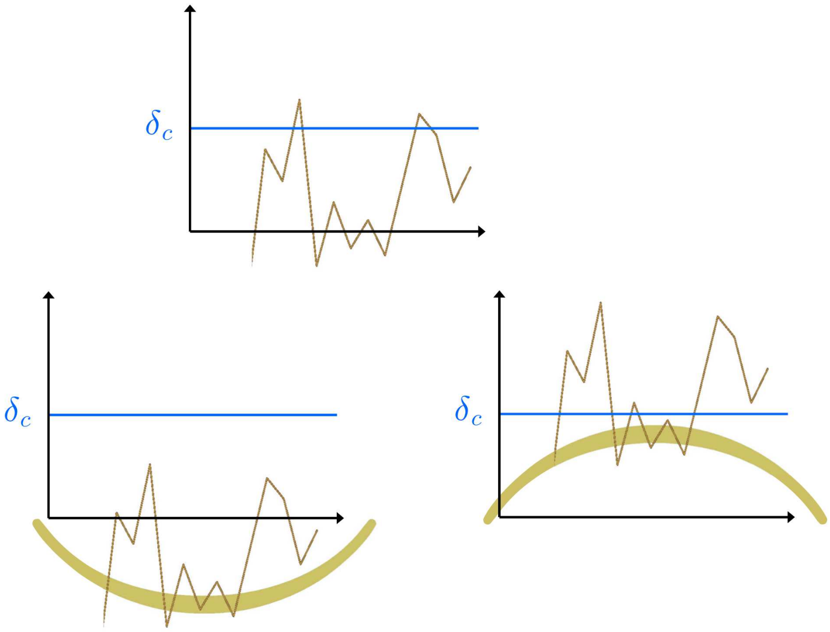

Figure 4: Simple illustration of the large bias of high-redshift galaxies, which is the main idea driving the character of reionization [24] and the 21-cm fluctuations during cosmic dawn [28]. To form a halo, the total (linearly-extrapolated) density fluctuation must reach a value δ_c [denoted $\delta_{\rm crit}(z)$ in Eq. (21)], from the sum of large-scale and small-scale density fluctuations. Thus, a large-scale void (bottom left) might have no halos, a typical region (top) a couple small halos, while a region with a large-scale overdensity (bottom right) will have many halos, both small and large. See text for additional explanation.

any given time. The reason is that the periodic boundary conditions within the finite simulation box artificially set the amplitude of large-scale modes (above the box size) to zero. There are many such volumes in the real Universe, with various mean densities (that follow a Gaussian distribution, within linear perturbation theory). Since galaxies (especially at high redshift) are highly biased, most of them form in those volumes that have an unusually high mean density. Thus, a simulated volume at the cosmic mean density is not representative of the locations of stars.

This limitation of simulations is most acute for the very first star in the Universe, a challenge of special interest for simulators because it represents in principle a perfectly clean problem, before the first entrance of the complexities of astrophysical feedback from prior star formation. The very first star formed in a very rare high-density region. Indeed, the large size of the real Universe allowed such a rare fluctuation to be found somewhere by chance, but it is unlikely to be found within a small simulation box, even if the simulation has the right abundance of galaxies (while real simulations, in addition, artificially lower this abundance when setting the mean density in the box to the cosmic mean density). For example, one of the first high-resolution "first star" simulations formed its first star only at redshift 18.2 [81], while analytical methods show that the first stars must have formed at $z \sim 65$ [8, 9] within our past light cone (i.e., so that we can in principle see them as they formed), or a further $\Delta z \sim 6$ earlier [82] within the entire volume of the observable Universe (so that we can see them or their remnants after they formed). On this point, we note that there were some early, rough analytical estimates of the formation redshift of the very first stars [83, 84].

More generally, Barkana & Loeb (2004) [24] developed a hybrid model that allows one to predict the modified halo mass function in regions of various sizes and various average densities. As noted in section 2.3, for the cosmic mean halo abundance, the classic Press-Schechter [53] model works only roughly, while the halo mass function of Sheth & Tormen (1999) [55] (with modified best-fit parameters [56]) fits numerical simulations much more accurately [57]. Now, a generalization of the Press-Schechter model known as the extended Press-Schechter model [54] allows the prediction of the halo mass function in a given volume (of given size and mean density) compared to the cosmic mean mass function. No simple generalization of this type is known for the Sheth-Tormen mass function, but Barkana & Loeb [24] pointed out that this problem can be overcome since the prediction of the extended Press-Schechter model for the change *relative* to the cosmic mean mass function has been shown to provide a good fit to numerical simulations over a wide range of parameters [61, 85, 56]. Thus, the Barkana & Loeb [24] hybrid model starts with the Sheth-Tormen mass function and applies a correction based on the extended Press-Schechter model. The model gives a good match to simulations even in volumes that strongly deviate from the cosmic mean halo function (Figure 5).

The idea of unusually large fluctuations in the abundance of early galaxies first made a major impact on studies of cosmic reionization, leading to the conclusion that reionization occurs inside-out, with typical H II bubbles that are larger and thus easier to observe than previously thought [24] (see § 6.1). The same idea soon found another important application in a different regime, leading to the prediction of 21-cm fluctuations from earlier times during cosmic dawn. The study of fluctuations in the intensity of early cosmic radiation fields began with Lyα radiation [28] (see § 6.2) and continued to other fields including the X-rays responsible for early cosmic heating [29] (see § 6.3). These are all sources of 21-cm fluctuations, and are thus the main targets for 21-cm radio interferometers. Clearly, the idea of substantial large-scale fluctuations in galaxy numbers is a driver of much of the current theoretical and observational interest in 21-cm cosmology as a way to probe the era of early galaxy formation. The recent discovery

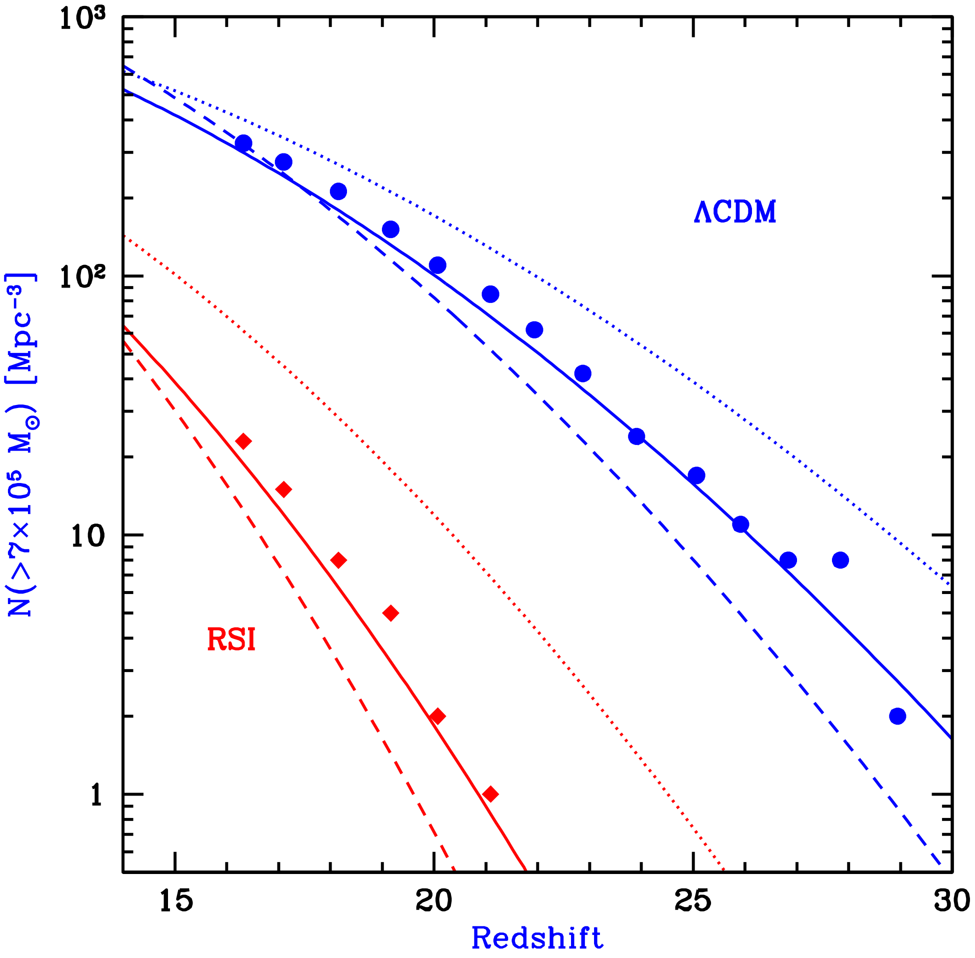

Figure 5: Halo mass function at high redshift in a 1 Mpc box at the cosmic mean density. Data points show the number of halos above mass $7 \times 10^5 M_\odot$ as measured in simulations (from Figure 5 of [80]) with two different sets of cosmological parameters, the scale-invariant ΛCDM model of [80] (upper curves), and their running scalar index (RSI) model (lower curves). Each data set is compared with three theoretically predicted curves. The simulated values are well below the cosmic mean of the halo mass function (dotted lines). However, the prediction of the Barkana & Loeb (2004) [24] hybrid model (solid lines) takes into account the periodic boundary conditions of the small simulation box and matches the simulation results fairly well. The pure extended Press-Schechter model (dashed lines) is too low. From [24].

of the streaming velocity (see § 5) has added a new flavor to this general theme.

3.2. Simulations at high redshift: challenges and approaches

In this subsection we discuss several aspects of simulations of the high-redshift Universe. First, we discuss some challenges and limitations of current numerical simulations, particularly when applied to early galaxy formation at high redshifts. Some of the issues we discuss can be addressed with additional study (e.g., setting the initial conditions accurately), while other difficulties are likely to remain for the foreseeable future (such as uncertainties related to star formation and stellar feedback). We then briefly discuss other approaches: analytical models and semi-numerical simulations.

We begin with a number of challenges that are important to recognize when evaluating the results of numerical simulations. As explained in the previous subsection, the large size of the real Universe implies that stars began to form very early. More generally, halos of various masses (or circular velocities) are predicted to have begun to form much earlier than the typical redshifts we are accustomed to, both from current numerical simulations and current observations. Figure 6 shows that while the very first star formed (in our past light cone) via molecular cooling at $z \sim 65$, the first generation of more massive atomic-cooling halos formed at $z \sim 47$ [8]. While the Milky Way halo mass is fairly typical in today's Universe, the very first such halo formed at $z \sim 11$, and the first Coma cluster halo at $z \sim 1.2$.

A direct simulation of the entire observable universe out to the spherical shell at redshift 70 would require a simulated box of length 25,000 Mpc on a side. Actual simulations, which often form a "first star" at redshift 20 or 30, effectively explore a very different environment from $z \sim 65$, in terms of the CMB temperature, the cosmic and virial halo densities (of both the dark matter and gas), the halo merger history, and high-redshift effects such as the difference between the power spectra of baryons and dark matter (discussed further below). Even if simulations do not attempt to approach the very first star, critical physical effects at high redshift push simulations towards the requirement of large boxes. The fact that the typical bubble scale of cosmic reionization is tens of Mpc (see § 6.1) already implies a minimum box size of ~ 100 Mpc for this era. However, the streaming velocity (§ 5), which is important early on, has a typical coherence scale of ~ 100 Mpc, and the radiation fields responsible for early feedback (§ 6) – Lyα coupling, Lyman-Werner feedback, and cosmic heating – fluctuate significantly on a similar scale. In particular, hard X-rays heat from afar and can extend the heating era into cosmic reionization (§ 6.3 and § 7.4).

A significant presence of any one of these effects is enough to force any reasonable simulation during these epochs to a minimum box size of ~ 400 Mpc. Another consideration that pulls in the same direction is that observations of the 21-cm signal are easier (and currently only possible) on large scales. The sensitivity of a radio interferometer is degraded as the angular resolution is increased [Eq. (67)]. Thus, numerical simulations are squeezed between the need to cover a huge volume, on the one hand, and the need to adequately resolve each halo, on the other hand. This becomes especially demanding at early times, when most of the star formation occurs in very low-mass halos. Consider, for example, an N-body simulation of a 400 Mpc box in which $10^6 M_\odot$ halos are resolved into 500 dark matter particles each. Extensive tests [89] show that this resolution is necessary in order to determine the overall properties of an individual halo (such as halo mass) just crudely, to within a factor of two; for better accuracy or to determine properties such as star formation, more particles are required. Even with just 500, this would require a simulation with a total of 10^{15} particles, much higher than numbers that are currently

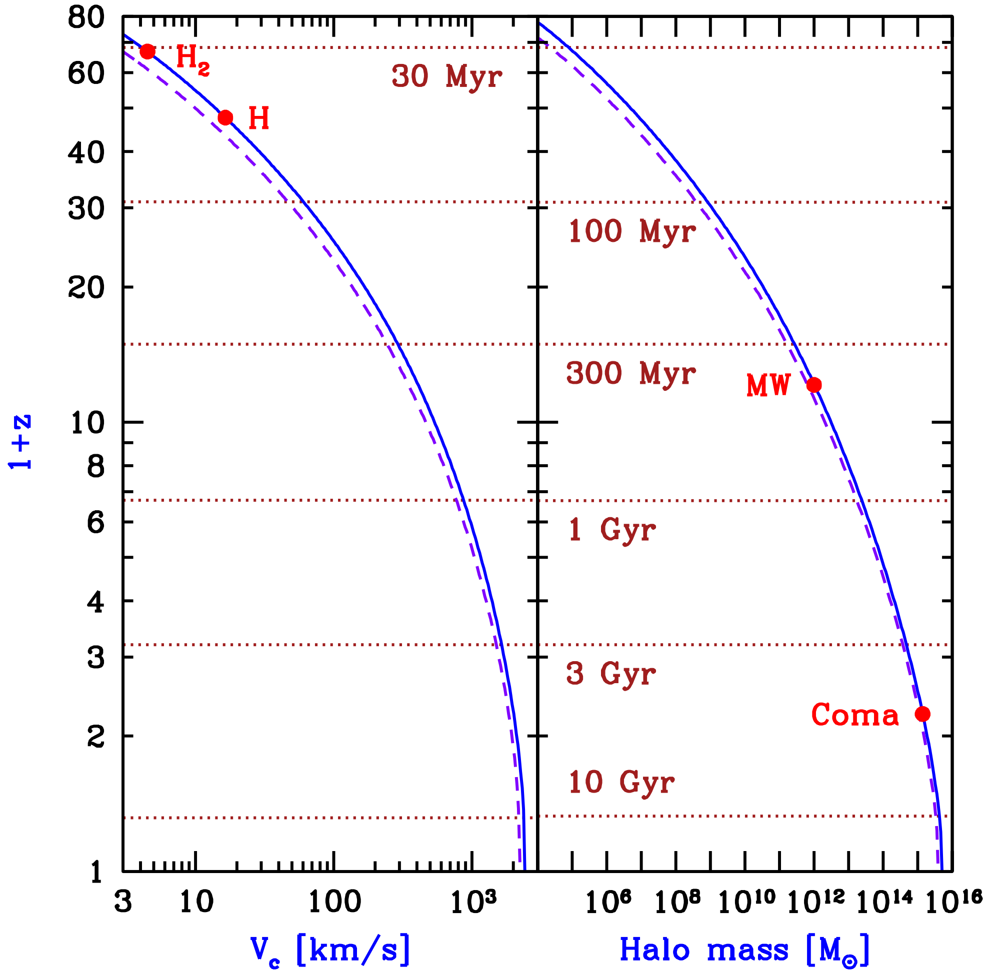

Figure 6: The median redshift for the first appearance (in our past light cone) of various populations of halos: either halos above a minimum circular velocity (left panel) or a minimum mass (right panel). Dots indicate in particular the first star-forming halo in which H_2 allows the gas to cool, the first galaxy that forms via atomic cooling (H), as well as the first galaxy as massive as our own Milky Way and the first cluster as massive as Coma. The horizontal lines indicate the elapsed time since the Big Bang. The results from two sets of cosmological parameters (solid curves [86] and dashed curves [87]) illustrate the systematic error due to current uncertainties in the values of the cosmological parameters. From [8].

feasible. Truly resolving star formation within these halos also requires hydrodynamics and radiative transfer at very high resolution.

Naoz et al. (2006) [8] pointed out another limitation of current simulations, namely that they do not determine their initial conditions accurately enough for achieving precise results for high-redshift halos, especially those hosting the very first stars. Simulations assume Gaussian random initial fluctuations as might be generated by a period of cosmic inflation in the early Universe. The evolution of these fluctuations can be calculated exactly as long as they are small, with the linearized Einstein-Boltzmann equations. The need to begin the simulation when fluctuations are still linear forces numerical simulations of the first star-forming halos to start at very high redshifts (much higher than starting redshifts in common use that are often around $z = 200$). According to spherical collapse, a halo forming at redshift $z_{\rm form}$ has an extrapolated linear overdensity of $\delta = \delta_c \sim 1.7$. Since it grows roughly with the EdS growing mode, the corresponding perturbation (in the dark matter) is $\delta \sim 13\% \, [(1 + z_{\rm form})/66]$ at cosmic recombination, and $\delta \sim 6\% \, [(1 + z_{\rm form})/66]$ at matter-radiation equality (see Figure 7). The perturbation reaches $\delta \sim 1\% \, [(1 + z_{\rm form})/66]$ extremely early, at $z \sim 10^6$. It re-enters the horizon (after having left during inflation) when $\delta \sim 0.2\% \, [(1 + z_{\rm form})/66]$ at $z \sim 3 \times 10^7$; precision at this level would require setting initial conditions with a non-linear General Relativistic calculation.

In addition to the problem of non-linearity, there is also the influence of early cosmic history on the linear and (more challengingly) non-linear initial conditions. Effects that must be taken into account include the contribution of the radiation to the cosmic expansion, suppression of sub-horizon perturbations in the photon density by the radiation pressure, and the coupling of the baryons to the photons which suppresses baryon perturbations until cosmic recombination. Within a spherical collapse calculation, Naoz et al. [8, 62] calculated halo formation including all these effects (Figure 7), and found that they result in an earlier formation redshift for the first star by 3.3% in $1 + z$ (compared to using the standard results from spherical collapse). The extended period at high redshift when the baryon perturbations remain suppressed compared to the dark matter is the main cause of this shift in the formation time, but the contribution of the photons to the expansion of the universe also makes a significant contribution. A 3.3% change in $1 + z$ at $z \sim 65$ corresponds to a 4.8% change in the age of the universe, and to an order of magnitude change in the abundance of halos at a given redshift at $z \sim 65$. The shift in $1 + z$ for the formation of a given halo goes down with time but is still 1% at $z \sim 20$. In addition, early cosmic history has a major impact (by factors of two or more) on the amount (and distribution) of gas that accumulates in the halos that hosted the first stars [76] (see the discussion of the filtering mass in § 2.4); this effect is increased further by the presence of the streaming velocity (see § 5). Therefore, even mild precision in numerical simulations of the formation of the first stars requires a calculation of these effects on halo formation, in combination with the above-mentioned issue of non-linearity going back to extremely early times.

Thus, while some processes are calculated with very high precision in numerical simulations, there are much larger effects that must be confronted before the results can be considered to be accurate. Even in the limit of the very first stars, ostensibly a very clean problem for numerical simulations, the effects just discussed make the problem difficult, even if all relevant physical processes can someday be included and numerical convergence fully demonstrated. The current status of numerical simulation results on the formation of the first stars is summarized below (§ 3.3).

Numerical simulation of galaxy formation *beyond* the very first star (in a given cosmological region) faces even bigger problems, which can be summarized with one word: feedback. Long-distance feedback directly from stellar radiation is generated by Lyα photons (reaching out to

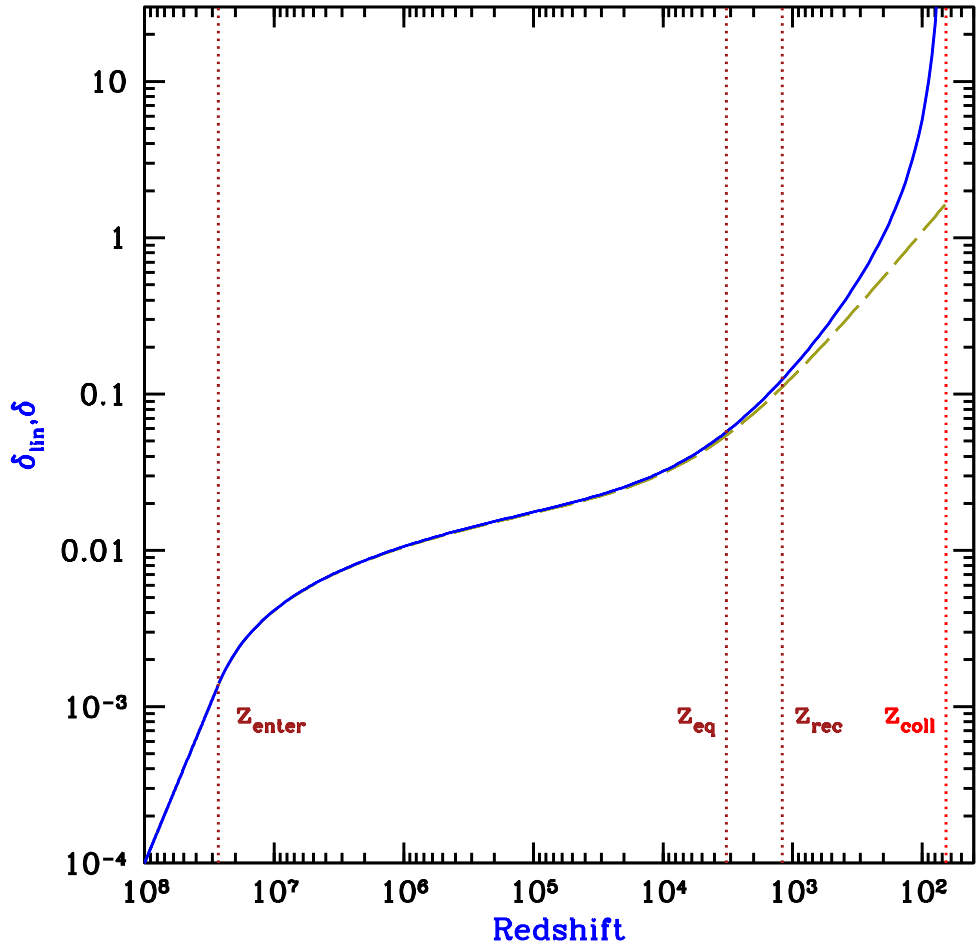

Figure 7: Evolution of the fractional overdensity δ for a spherical region containing $10^5 M_\odot$ that collapses at $z = 66$ (approximately corresponding to the host halo of the very first star in our past light cone). We show the fully non-linear δ (solid curve) and the linearly-extrapolated δ (dashed curve). We indicate the redshifts of halo collapse ($z_{\rm coll}$), cosmic recombination ($z_{\rm rec}$), matter-radiation equality ($z_{\rm eq}$), and entry into the horizon ($z_{\rm enter}$). Note that the overdensity shown here corresponds to synchronous gauge. From [8].

$\sim$ 300 Mpc), Lyman-Werner photons (out to $\sim$ 100 Mpc), and UV ionizing photons (initially absorbed in the immediate surroundings, but reaching up to $\sim$ 70 Mpc by the end of reionization [90]). Some stars have strong stellar winds, and some explode in supernovae, which deposit thermal and kinetic energy as well as metals. Stellar remnants such as X-ray binaries produce X-rays which include hard photons that reach cosmological distances. Central black holes may also produce thermal and kinetic feedback, as well as UV and X-ray radiation. Most types of radiation that are responsible for feedback can be partially absorbed or scattered within the emitting galaxy or its immediate surroundings, another important process that depends on the detailed, small-scale distribution of gas and metals. Given the basic uncertainties about the detailed physics even of well-observed present-day astrophysical phenomena such as magnetic fields, dust, supernovae, the stellar initial mass function, and central black holes, ab-initio numerical simulations that are truly self-contained do not seem feasible. Once these various feedback effects begin to operate, they strongly affect the properties of subsequent generations of stars and galaxies, so that many observable predictions become strongly dependent on the generation and results of feedback. Numerical simulations can offer increasingly precise gravity, hydrodynamics, and radiative transfer, but are often limited by simplistic models of star formation and feedback that are inserted by hand. A major issue with astrophysical sources is that truly simulating their formation process, detailed structure, and feedback would require resolving length scales that are around 20 orders of magnitude smaller than the cosmological distances reached by some of the photons responsible for radiative feedback. The resulting vast gulf between the resolution of cosmological simulations and that of reality means that increasing resolution does not necessarily imply convergence towards the correct final answer; there could be multiple regimes of apparent convergence as additional levels of resolution uncover new physical processes.

On the opposite end from simulations are analytical (or semi-analytical) models. These models are very flexible, can be easily used to explore a wide variety of astrophysical possibilities and to incorporate a range of astrophysical uncertainties, and can be directly fit to data in order to determine the parameters of well-fit models. Such models can also be made more quantitatively accurate by basing them on fits to the results of numerical simulations of early galaxy formation. However, analytical models are also significantly limited. In 21-cm cosmology, perhaps their biggest limitation is that they usually must assume linear perturbations. While large-scale density fluctuations are indeed fairly small at early times, the large bias of high-redshift galaxies (§ 3.1) leads to quite non-linear fluctuations in the radiative, astrophysical sources of 21-cm fluctuations. In addition, the highly non-linear fluctuations on small scales do not completely average out when smoothing on large scales (as in real observations). This is due to additional non-linear relationships in 21-cm cosmology such as the dependence of 21-cm temperature on gas temperature (Eq. 50 or 56). Thus, analytical calculations based on assuming linear perturbations and linear bias are quite limited in their accuracy (An important example is the discussion in § 4.3 of non-linear limits on the accuracy of the linear result for the anisotropy of the 21-cm power spectrum).

The limitations of both numerical simulations and analytical models have led to the rise of an intermediate approach that combines some of the advantages of both. This method is termed hybrid, or semi-numerical simulation. While there are several specific approaches, the basic idea is to calculate physical processes directly on large scales, where everything is relatively simple, and indirectly on small, highly non-linear scales. On the small scales, halos and their properties are often adopted from semi-analytical models that have been fitted to numerical simulation results, or sometimes directly from the outputs of N-body (i.e., gravity-only) simulations plus some assumptions about star formation and other astrophysics. On the large scales, radiation such as X-rays, LW, and Lyα photons can be directly summed from all sources, albeit with a few approximations (e.g., the optical depth calculated assuming the cosmic mean density, and multiple scattering of Lyα photons treated approximately). Also, for reionization, such codes usually employ an approximation based on an analytical model for the distribution of H II bubble sizes [25] (§ 6.1); fortunately, the resulting ionized bubble distribution is quite similar to the results of radiative transfer, except in the fine (small-scale) details (see Figure 23 in § 6.1). A successful, publicly-available semi-numerical code in 21-cm cosmology is 21CMFAST [91]; results from this code and from the code developed by the author's group [92] are shown in § 7.

To summarize this subsection, numerical simulations of early galaxies offer the potential advantages of fully realistic source halo distributions and accurate gravity, hydrodynamics, and radiative transfer. However, much of the vitality of the field comes from the major uncertainties associated with the formation of, and feedback from, astrophysical sources. For example, it is possible that most early stars were much more massive and thus brighter than modern stars, or that a relatively large amount of gas collected within massive mini-quasars in the centers of galaxies. These astrophysical uncertainties will very likely be resolved only based on direct observational evidence. As we contemplate the range of possible observational predictions, it is much easier to explore a wide variety of astrophysical possibilities with simple analytical models or semi-numerical hybrid methods that combine processes on a large-scale grid with a sub-grid model based on numerical simulation results. Once the observations come in, there will be a need to fit astrophysical parameters to the data, and this requires a flexible framework and cannot be done directly with numerical simulations; once a well-fit model has been found, though, simulations may offer the best way to compare it in detail with the observations. It is important to note that discoveries in the field of the first stars and 21-cm cosmology (as summarized throughout this review) are often driven by large-scale processes, so due to the limited reach of simulations, many have come first from analytical or semi-numerical methods.

3.3. The very first stars

In the previous two subsections we discussed the limitations of numerical simulations in general, and those of the first stars in particular. Still, simulations remain our best tool for trying to understand and predict the detailed properties of the first stars. This subject has been extensively reviewed elsewhere [93, 94], but we briefly summarize it in this section. In principle, the formation of primordial stars is a clean numerical problem, as the initial conditions (including the distribution of the gas and dark matter and the chemical and thermal history of the gas) are cosmological and not yet affected by astrophysical feedback. One possible (though still speculative) complication is the generation and amplification of magnetic fields in the early universe in time for them to affect the formation of the first stars [95, 96, 97, 93].

As mentioned at the end of § 2.4, under cosmological conditions, gas cooling in small early halos is possible only via molecular hydrogen cooling. Studies of the non-equilibrium chemistry of H_2 formation and destruction [98, 99, 100, 101, 102] concluded that H_2 formation in a collapsing small halo is dominated by the H^- channel, in which the residual free electrons from cosmic recombination act as catalysts:

$$H + e^- \to H^- + \gamma \ ; \quad H^- + H \to H_2 + e^- \ . \tag{36}$$

Numerical simulation of the formation of a first (so-called Population III, or Pop III) star via H_2 cooling in a primordial minihalo of $10^5 - 10^6 M_\odot$ has proven to be a difficult problem, as initial results that established a prediction of single very massive stars have recently been replaced by a new paradigm of multiple stellar systems with a range of masses. Indeed, the first generation of simulations indicated the formation of massive Pop III stars of $\sim 100 M_\odot$. Such stars would be short-lived, generate extremely strong ionizing radiation and stellar winds, and end up producing massive black hole seeds or pair-instability supernovae. The expectation of massive stars was consistent between early simulations evolving an artificial overdensity with a smooth particle hydrodynamics (SPH) code (Figure 8) and simulations that directly employed cosmological initial conditions along with the impressive resolution of an adaptive mesh refinement (AMR) code (Figure 9).

Even for a given set of initial conditions for star formation, the final properties of the resulting stars depend on a complex process of proto-stellar evolution. It was initially thought that the rapid accretion rates characteristic of primordial star-forming regions at high-redshift would naturally lead to isolated Pop III stars of $100 M_\odot$ or more. However, some simulations [103] then showed the possible formation of binaries (Figure 10), and further semi-analytical and numerical simulation studies [104, 105, 106, 107, 108, 109] have found that the clumps have sufficient angular momentum to form a disk, and that the rapid accretion onto the disk causes it to fragment due to gravitational instability. While it is too early to draw final conclusions, the best bet currently is that Pop III stars formed with a wide range of different masses, but on average were significantly heavier than later generations of stars (Figure 11).

4. 21-cm Cosmology

An overview of the basic features and early development of 21-cm cosmology was given in § 1. In this section we present the basic physics in greater detail, then focus on some important low-temperature corrections, and discuss the important subject of anisotropy in the 21-cm signal. Finally, we give a brief overview of the observational aspects of 21-cm cosmology, focusing on the power spectrum. More details of 21-cm physics and observational techniques are available in specific reviews of 21-cm cosmology [88, 110, 111].

4.1. *Basic physics*

The basic physics of the hydrogen spin transition is determined as follows. At the low densities typical in cosmological applications, the gas is far from full thermal equilibrium, and a single temperature cannot accurately describe the occupancy of various atomic levels. In particular, the relative occupancy of the spin levels is usually described in terms of the hydrogen spin temperature T_S, which is an effective temperature that determines the emission or absorption properties of the 21-cm line. Specifically, T_S is defined by

$$\frac{n_1}{n_0} = 3 \, \exp\left\{ -\frac{T_*}{T_S} \right\} \,, \tag{37}$$

where n_0 and n_1 are the number densities of the singlet and triplet hyperfine levels in the atomic ground state ($n = 1$), respectively, and $T_* = 0.0682$ K is defined by $k_B T_* = E_{21}$, where the energy of the 21-cm transition is $E_{21} = 5.87 \times 10^{-6}$ eV, corresponding to a frequency of 1420 MHz (and a precise wavelength of $\lambda_{21} = 21.1$ cm). The factor of 3 in Eq. (37) is the ratio of statistical weights, i.e., it arises from the degeneracy factor of the spin 1 excited state. In particular, $T_S \rightarrow \infty$ would correspond to having the singlet and triplet levels populated in their statistical 1:3 ratio, $T_S \rightarrow 0$ would mean an empty excited state, while a population inversion (not expected in the cosmological 21-cm field) would correspond to negative T_S. Since T_* is such

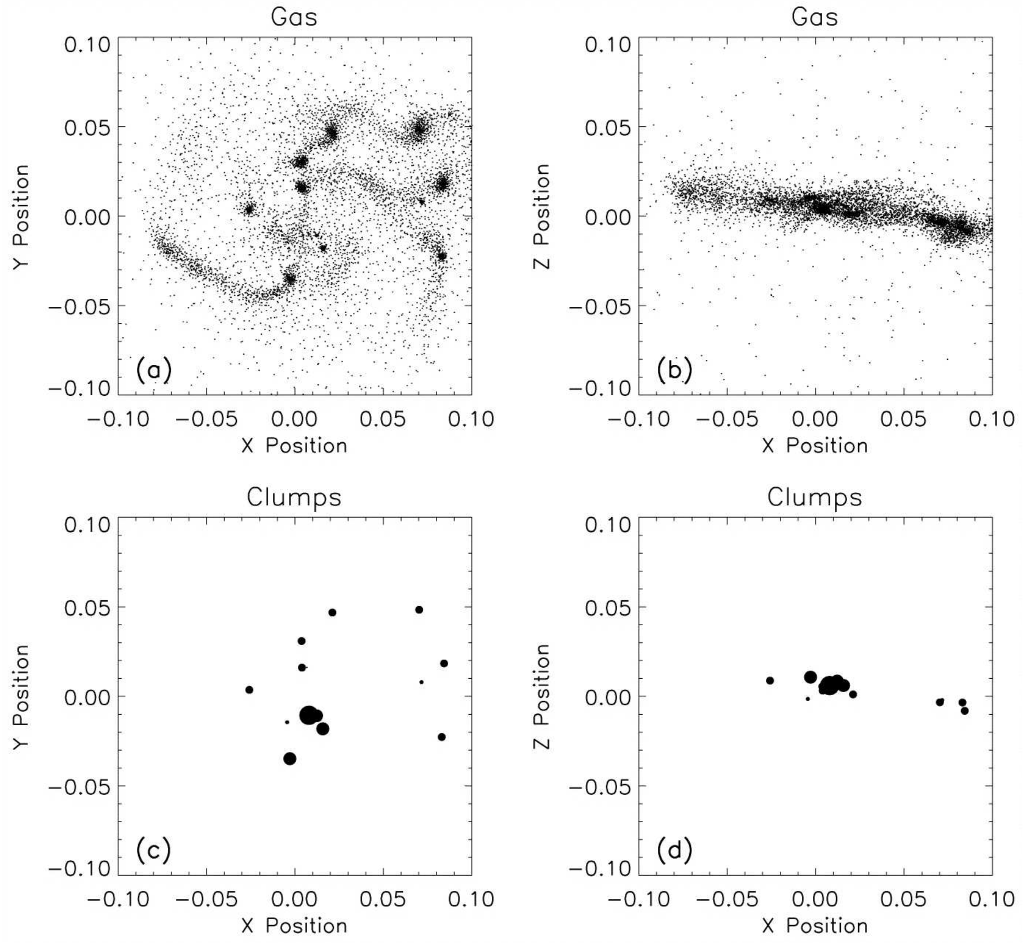

Figure 8: Gas and clump morphology at $z = 28$ in the first-star simulation of Bromm et al. (1999) [6]. Top row: The remaining gas in the diffuse phase. Bottom row: The distribution of clumps, where the four increasing dot sizes denote increasing clump masses ($> 10^2 M_\odot$, $> 10^3 M_\odot$, $> 5\times10^3 M_\odot$, $> 10^4 M_\odot$). Left panels: Face-on view. Right panels: Edge-on view. The length of the box is 30 pc. The gas has settled into a flattened configuration with a number of dominant, massive clumps. From [6].

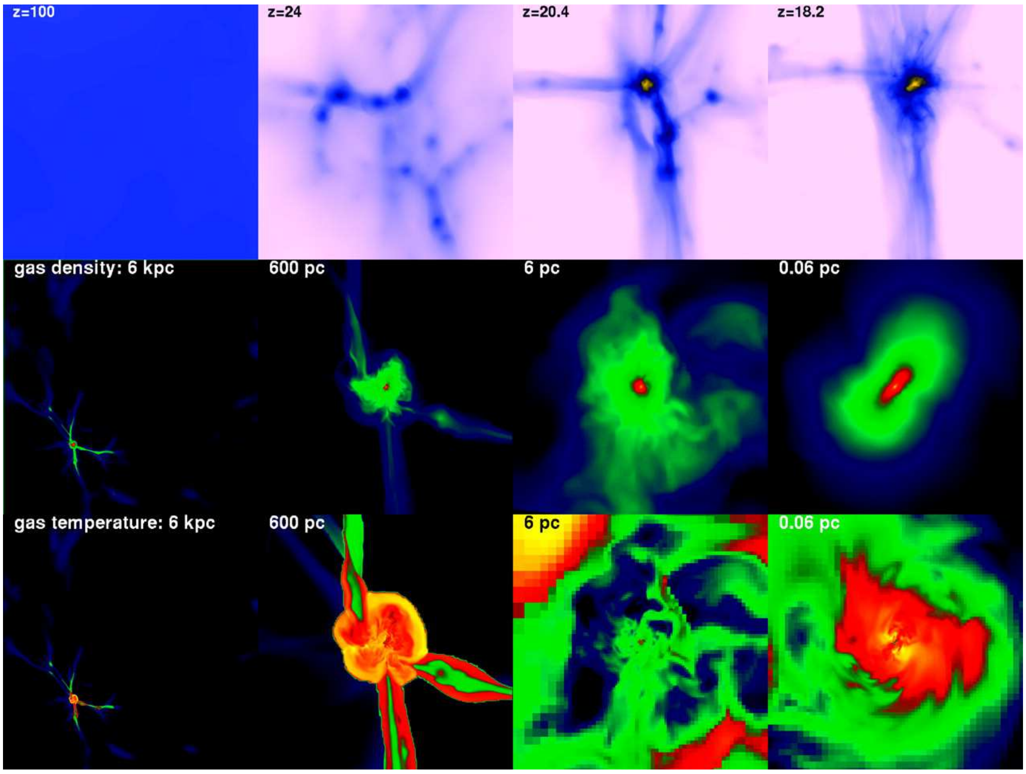

Figure 9: The first star in a simulation by Abel et al. (2002) [7]. Top row: Projection of gas density on a 600 pc scale (all distances are physical in this Figure), at several redshifts. Other two rows: Slices of gas density or temperature on several different scales, all at the final redshift of the simulation ($z = 18.2$). From left to right, the two bottom rows show: large-scale filaments; the virial accretion shock; the H_2 cooled, high-redshift molecular cloud analog; and a warm core containing $\sim 100\,M_\odot$ of gas. From [7].

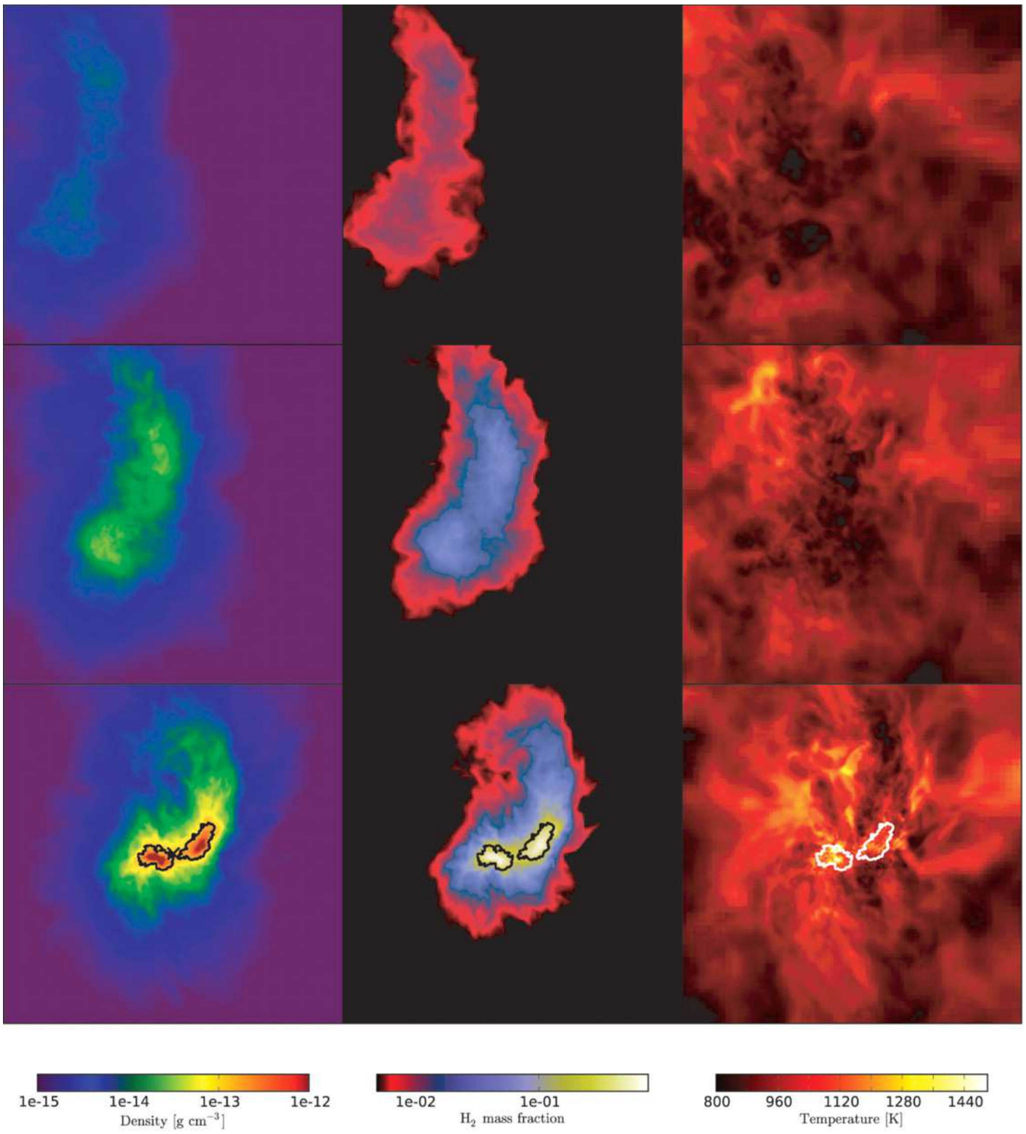

Figure 10: The first stars may have been binaries, according to a simulation by Turk et al. (2009) [103]. Shown is the average density (left column), H_2 mass fraction (middle), and temperature (right), projected through a cube 3500 AU on a side. The bottom row (in which the two separate gravitationally-bound cores are outlined with thick lines) is at the end of the simulation, with the other rows showing earlier times by 555 years (middle) or 1146 years (top). From [103].

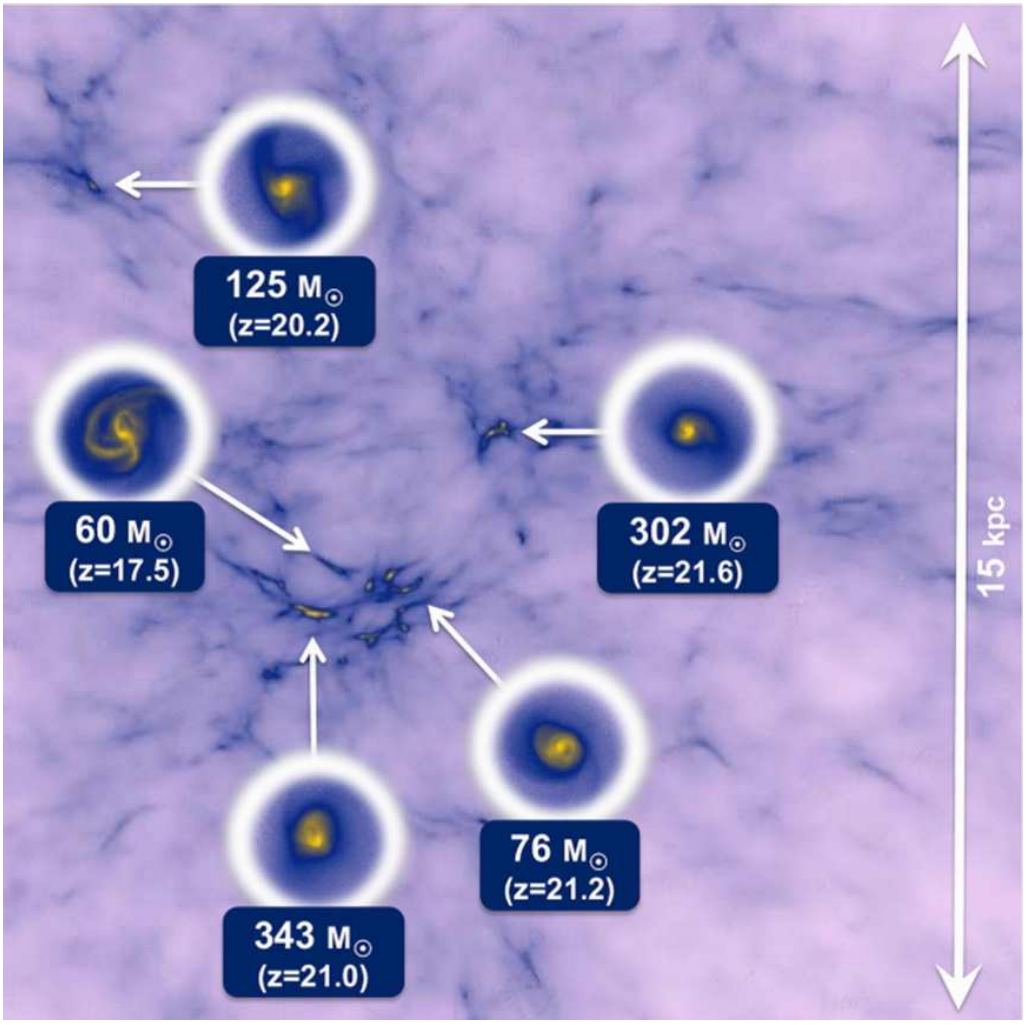

Figure 11: The first stars may have had a range of masses, based on a simulation by Hirano et al. (2014) [108]. The projected gas density is shown at $z = 25$. Five primordial star-forming clouds are highlighted, with each circle showing a zoom-in to the central parsec at the formation time of the star; its formation redshift and stellar mass are listed. From [108].

a low temperature, in what follows we make the standard assumption that all other temperatures (including T_S) are much higher.

A patch of neutral hydrogen at the mean density and with a uniform T_S produces an optical depth at 21 cm [observed at $21(1+z)$ cm] of

$$\tau(z) = \frac{3c\lambda_{21}^2 h_P A_{10} n_{H\ I}}{32\pi k_B T_S (1+z)(dv_r/dr)} , \tag{38}$$

where h_P is Planck's constant, $A_{10} = 2.85 \times 10^{-15}$ s^{-1} is the spontaneous decay rate of the hyperfine transition, $n_{H\ I}$ is the number density of hydrogen atoms, and dv_r/dr is the gradient of the radial velocity along the line of sight, with v_r being the physical radial velocity and r the comoving distance. In a fully-neutral, homogeneous universe, $n_{H\ I} = \bar{n}_H(z)$ and $dv_r/dr = H(z)/(1+z)$ in terms of the Hubble parameter H. Assuming the high-redshift (EdS) form for $H(z)$ (see § 2.1), this yields

$$\tau(z) = 9.85 \times 10^{-3} \left(\frac{T_{\rm CMB}}{T_S}\right)\left(\frac{\Omega_b h}{0.0327}\right)\left(\frac{\Omega_m}{0.307}\right)^{-1/2}\left(\frac{1+z}{10}\right)^{1/2} , \tag{39}$$

where T_S and $T_{\rm CMB}$ are measured at z. Since the brightness temperature through the IGM is $T_{\rm b}^z = T_{\rm CMB} e^{-\tau} + T_S(1-e^{-\tau})$, the observed mean differential antenna temperature relative to the CMB is [11]

$$T_{\rm b} = (1+z)^{-1}(T_S - T_{\rm CMB})(1-e^{-\tau}) \simeq 26.8\,{\rm mK} \left(\frac{\Omega_b h}{0.0327}\right)\left(\frac{\Omega_m}{0.307}\right)^{-1/2}\left(\frac{1+z}{10}\right)^{1/2}\left(\frac{T_S - T_{\rm CMB}}{T_S}\right) , \tag{40}$$

where $\tau \ll 1$ is assumed (the relative correction to the linear term that we kept is $\tau/2$) and $T_{\rm b}$ has been redshifted to redshift zero. We use here the now standard notation of $T_{\rm b}$ for this final quantity. Note that the brightness temperature is simply a measure of intensity in equivalent temperature units, defined in terms of the Rayleigh-Jeans limit of the Planck spectrum:

$$I_\nu = 2k_B T_{\rm b} \frac{\nu^2}{c^2} . \tag{41}$$

Note that in 21-cm cosmology, the CMB is certainly deep in the Rayleigh-Jeans limit, as its Planck spectrum peaks at a wavelength of ~ 2 mm, while the observed (redshift 0) wavelengths of relevance to us here are three orders of magnitude larger.

The IGM is observable when T_S differs from $T_{\rm CMB}$, which is reasonable since $T_S = T_{\rm CMB}$ implies a kind of thermal equilibrium between the ground-state hyperfine levels of hydrogen and the CMB background, meaning that the net effect of the gas is neither absorption nor added emission above the background. The key question for 21-cm observations is thus the value of the spin temperature. For intergalactic hydrogen it is determined by three processes. First, by direct absorption and emission (both spontaneous and stimulated) of 21-cm photons from/into the radio background (which at high redshifts is simply the CMB), the hyperfine levels of hydrogen tend to thermalize with the CMB, making the IGM unobservable. If other processes shift the hyperfine level populations away from such a thermal equilibrium, then the gas becomes observable against the CMB in emission or in absorption. In the presence of the CMB alone, the spin states would reach thermal equilibrium with $T_S = T_{\rm CMB} = 2.725(1+z)$ K on a time-scale of $T_*/(T_{\rm CMB}A_{10}) \simeq 3 \times 10^5 (1+z)^{-1}$ yr. This time-scale is much shorter than the age of the universe at all redshifts after cosmological recombination.

On the other hand, at high densities the spin temperature comes into equilibrium with the regular, kinetic temperature T_K that describes the random velocities of the hydrogen atoms. This equilibrium is enforced by collisions, which involve energies of order $k_B T_K$, and drive T_S towards T_K [112]. Collisionally-induced transitions are effective at high redshift, but become less effective compared to the CMB at low redshift. This may seem surprising given that as the universe expands, the mean energy density of radiation decreases faster than that of matter, and the comparison here is between two-body interactions of the hydrogen atom with either a photon or a second atom. Part of the explanation is that while the total radiation energy density goes as $T_{\rm CMB}^4$ (and thus decreases rapidly with time), the relevant energy density for the 21-cm coupling is that at a fixed physical wavelength of 21 cm; this is only proportional to $T_{\rm CMB}$ in the Rayleigh-Jeans limit of the Planck spectrum of the CMB (Eq. 41). In addition, the collisional rate coefficient (see below) depends strongly on temperature in the relevant range, and it decreases very rapidly as the gas cools with time. Thus, if collisions were the only coupling mechanism of the spin temperature with the kinetic temperature, the cosmic gas would disappear at 21 cm below $z \sim 30$.

Instead, 21-cm cosmology down to $z \sim 7$ is made possible by a subtle atomic effect worked out nearly 50 years before its cosmological significance became widely recognized. This effect is 21-cm coupling as an indirect consequence of the scattering of much higher-energy Lyα photons [18, 19]. Continuum UV photons produced by early radiation sources redshift by the Hubble expansion into the local Lyα line at a lower redshift, or are injected at Lyα after redshifting and cascading down from higher Lyman lines. These photons mix the spin states via the Wouthuysen-Field (hereafter WF) effect whereby an atom initially in the $n = 1$ state absorbs a Lyα photon (of wavelength $\lambda_\alpha = 1216$ Å), and the spontaneous decay that returns it from $n = 2$ to $n = 1$ can result in a final spin state that is different from the initial one (These various energy levels are illustrated in Figure 1). The WF effect drives T_S to the so-called "color temperature" T_C, defined so that the spin-flip transition rates due to Lyα photons upwards (P_{01}^α) and downwards (P_{10}^α) are related by [113]:

$$\frac{P_{01}^\alpha}{P_{10}^\alpha} = 3\left(1 - \frac{T_*}{T_C}\right) . \tag{42}$$

The color temperature enters since the $0 \rightarrow 1$ and $1 \rightarrow 0$ scattering events are caused by photons with slightly different frequencies. It is the equivalent temperature of a blackbody spectrum that would yield this transition rate ratio. In general (i.e., including the case of a non-blackbody radiation background), the color temperature is determined by the shape of the radiation spectrum near Lyα, and is related to the photon intensity J through [114]

$$\frac{h}{k_B T_C} = \frac{2}{\nu} - \frac{d \ln J}{d\nu} . \tag{43}$$

Given CMB scattering (which pulls $T_S \rightarrow T_{\rm CMB}$), atomic collisions ($T_S \rightarrow T_K$), and Ly$\alpha$ scattering ($T_S \rightarrow T_C$), the spin temperature becomes a weighted mean [113]:

$$T_S^{-1} = \frac{T_{\rm CMB}^{-1} + x_c T_K^{-1} + x_\alpha T_C^{-1}}{1 + x_{\rm tot}} , \tag{44}$$

where $x_{\rm tot} = x_c + x_\alpha$ and the combination that appears in $T_{\rm b}$ (Eq. 40) is then:

$$\frac{T_S - T_{\rm CMB}}{T_S} = \frac{x_{\rm tot} - T_{\rm CMB}\left(x_c T_K^{-1} + x_\alpha T_C^{-1}\right)}{1 + x_{\rm tot}} . \tag{45}$$

Here we have used the notation from Barkana & Loeb (2005) [28] in terms of the coupling coefficients x_c and x_α for collisions and Lyα scattering, respectively. They are given by [11]

$$x_c = \frac{4\kappa_{1-0}(T_k)\, n_H T_\star}{3A_{10} T_{\rm CMB}} \ , \tag{46}$$

where the collisional rate coefficient $\kappa_{1-0}(T_k)$ is tabulated as a function of T_k [115, 116], and

$$x_\alpha = \frac{4P_\alpha T_*}{27A_{10} T_{\rm CMB}} \ , \tag{47}$$

in terms of the Lyα scattering rate P_α. Expressed in terms of the proper Lyα photon intensity J_α (defined as the spherical average of the number of photons hitting a gas element per unit area per unit time per unit frequency per steradian),

$$x_\alpha = \frac{16\pi^2 T_\star e^2 f_\alpha}{27A_{10}\, m_e c\, T_{\rm CMB}} J_\alpha \ , \tag{48}$$

except for a low-temperature correction (see the next subsection), where $f_\alpha = 0.4162$ is the oscillator strength of the Lyα transition.

The neutral IGM is highly opaque to resonant scattering, which involves energy transfers between the atomic motion and the photons, and tends to drive a kind of thermal equilibrium between the photon energy distribution near Lyα and the kinetic motion of the atoms. This makes T_C very close to T_K [117], except for another low-temperature correction (see the next subsection). In the high-temperature approximation, equations (44) and (45) simplify to:

$$T_S^{-1} = \frac{T_{\rm CMB}^{-1} + x_{\rm tot} T_K^{-1}}{1 + x_{\rm tot}} \ , \tag{49}$$

and

$$\frac{T_S - T_{\rm CMB}}{T_S} = \frac{x_{\rm tot}}{1 + x_{\rm tot}} \left(1 - \frac{T_{\rm CMB}}{T_K}\right) \ . \tag{50}$$

Below $z \sim 200$, the gas is mostly thermally decoupled from the CMB and $T_K < T_{\rm CMB}$ (until significant X-ray heating), so that 21-cm observations are possible since collisions or Lyα scattering provide an effective mechanism coupling T_S to T_K. While Eq. (40) gives the 21-cm brightness temperature in a fully-neutral, homogeneous universe, in the real Universe $T_{\rm b}$ fluctuates. It is proportional in general to the gas density, and in partially ionized regions $T_{\rm b}$ is proportional to the neutral hydrogen fraction. Fluctuations in the velocity gradient term in Eq. (38) leads to a line-of-sight anisotropy in the 21-cm signal (§ 4.3). Also, if $T_S > T_{\rm CMB}$ then the IGM is observed in emission, and when $T_S \gg T_{\rm CMB}$ the emission level saturates at a level that is independent of T_S. On the other hand, if $T_S < T_{\rm CMB}$ then the IGM is observed in absorption, and if $T_S \ll T_{\rm CMB}$ the absorption strength is a factor $\sim T_{\rm CMB}/T_S$ larger (in absolute value) than the saturated emission level. In addition, once the Universe fills up with Lyα radiation and the WF effect turns on (this is the Lyα coupling transition, with its peak usually defined as the point when $x_{\rm tot} = 1$ due mostly to x_α), the rapid rise expected during the early stages of cosmic star formation implies that soon afterwards $x_\alpha \gg 1$ and $T_{\rm b}$ saturates to a value that no longer depends on x_α. As a result of these various considerations, a number of cosmic events (§ 6) are expected to leave observable signatures in the redshifted 21-cm line (§ 7).

4.2. Low-temperature corrections

There are two corrections to the 21-cm coupling due to Lyα scattering, which can be important in low-temperature gas. Both arise from a careful consideration of the multiple scatterings of the photons near the Lyα resonance with the hydrogen atoms, and how these scatterings affect the energy distribution of the photons near the resonance, resulting in a change in the 21-cm coupling. One correction is due to a difference between the color temperature and the kinetic temperature of the gas, and the other due to a modified Lyα scattering rate. We attempt here to clear up confusion in some of the literature on this subject.

An accurate determination of the Lyα color temperature requires a careful consideration of radiative scattering including atomic recoil and energy transfer due to spin exchange. In the limit of a high optical depth to Lyα scattering (an excellent approximation in the cosmological context),

$$T_C = T_K \left(\frac{1 + T_{\rm se}/T_K}{1 + T_{\rm se}/T_S} \right) , \tag{51}$$

which differs significantly from T_K once temperatures approach $T_{\rm se}$, which is given by

$$T_{\rm se} = \frac{m_H c^2}{9 k_B} \left(\frac{\lambda_\alpha}{\lambda_{21}} \right)^2 = 0.402 \text{ K} , \tag{52}$$

where m_H is the mass of a hydrogen atom. Eq. (51) is easily solved simultaneously with Eq. (44), yielding results that have precisely the same form as equations (49) and (50) if we replace $x_{\rm tot}$ by an $x_{\rm tot,eff}$ in which we adopt an effective value $x_{\alpha,\rm eff} = x_\alpha/(1 + T_{\rm se}/T_K)$.

The second effect modifies the relation between J_α (defined as the naive Lyα photon intensity, i.e., not including the modification due to multiple scattering) and the actual Lyα scattering rate P_α. The final result is to multiply Eq. (48) by an extra factor S_α, which depends on T_K as well as the Gunn-Peterson [22] optical depth to Lyα absorption, which for neutral gas at the cosmic mean density is

$$\tau_{\rm GP} = \frac{\pi e^2 f_\alpha \lambda_\alpha n_{\rm H\ I}}{m_e c H} = 6.62 \times 10^5 \left(\frac{\Omega_b h}{0.0327} \right) \left(\frac{\Omega_m}{0.307} \right)^{-1/2} \left(\frac{1+z}{10} \right)^{3/2} , \tag{53}$$

where in the second equality we used the high-redshift form of the Hubble parameter $H(z)$.

The scattering-rate correction factor S_α is due to the fact that the H atoms recoil in each scattering, and near the center of the Lyα line, frequent scatterings with atoms make the photons lose energy faster. Thus, the number of photons per unit energy at any instant is smaller than would have been expected without recoil, leading to a suppression in the scattering rate (i.e., $S_\alpha < 1$). The actual value of S_α is derived from solving the radiative transfer equation for the photons including scattering and energy losses. The result is

$$S_\alpha = e^{-0.0128 \left(\tau_{\rm GP}/T_K^2 \right)^{1/3}} , \tag{54}$$

with T_K in Kelvin (in this equation only). Thus, the final results including both low-temperature corrections are

$$T_S^{-1} = \frac{T_{\rm CMB}^{-1} + x_{\rm tot,eff} T_K^{-1}}{1 + x_{\rm tot,eff}} , \tag{55}$$

and

$$\frac{T_S - T_{\rm CMB}}{T_S} = \frac{x_{\rm tot,eff}}{1 + x_{\rm tot,eff}} \left(1 - \frac{T_{\rm CMB}}{T_K} \right) , \tag{56}$$

where $x_{\rm tot,eff} = x_c + x_{\alpha,\rm eff}$, and

$$x_{\alpha,\rm eff} = x_\alpha \left(1 + \frac{T_{\rm se}}{T_K}\right)^{-1} \exp\left[-2.06 \left(\frac{\Omega_b h}{0.0327}\right)^{1/3} \left(\frac{\Omega_m}{0.307}\right)^{-1/6} \left(\frac{1+z}{10}\right)^{1/2} \left(\frac{T_K}{T_{\rm se}}\right)^{-2/3}\right] . \tag{57}$$

Eq. (56) shows that even with the low-temperature correction, whether we get 21-cm emission or absorption is determined solely by whether T_K is larger or smaller than $T_{\rm CMB}$ (which seems reasonable based on thermodynamics), while at a given T_K, the absolute value of $T_{\rm b}$ increases monotonically with $x_{\alpha,\rm eff}$. The low-temperature corrections simply reduce the effective value of x_α and thus reduce the absolute value of $T_{\rm b}$ and delay the onset of Lyα coupling and Lyα saturation (the latter is when $T_S \to T_K$). Note that we wrote the scattering-rate correction in Eq. (57) in terms of $T_{\rm se}$ for ease of comparison with the color-temperature correction.

These results are based on Chuzhoy & Shapiro (2006) [118], who found simple and accurate final expressions based on an approximate analytical solution (that was also found earlier in a different context [119]). The calculation of S_α was first carried out by Chen & Miralda-Escudé (2004) [120] (based on a numerical solution to an approximate form of the radiative transfer equation developed earlier [121, 122]), but they made a numerical error and were off by about a factor of 2. Hirata (2006) [123] gave complicated fitting formulas to numerical solutions for both corrections, but the results given above agree with those formulas to within a relative error of a few percent or better within the relevant parameter range. Furlanetto & Pritchard (2006) [124] developed higher-order analytical solutions and also compared them to full numerical solutions. Contrary to statements in the literature [88], no iteration is necessary in order to include the low-temperature corrections; the results summarized in this section are accurate at all $T \gtrsim 1$ K, except at very high temperatures (> 1000 K) which in the real Universe are reached only after the Lyα coupling has saturated (and so these corrections no longer matter). Note also that the scattering correction factor S_α, while calculated slightly differently for the continuum (redshifting) Lyα photons and the injected (from higher-level cascades) Lyα photons, has the same value in the two cases, to high accuracy.

The quantitative results are illustrated in Figure 12. The scattering correction dominates over the correction from the color temperature. In practice, the observable effects of the low-temperature corrections could be important in the real Universe during the Lyα coupling era. These corrections affect 21-cm fluctuations only when the Lyα coupling is significant but has not yet saturated (since in the saturated limit, 21-cm observations are independent of x_α and its corrections). As long as the cosmic gas cools, the strengthening reduction in the effective x_α slows the rise of Lyα coupling; once the gas reaches its minimum temperature and begins to warm up, the declining low-temperature effect then accelerates Lyα saturation. In realistic models (see § 6 and § 7), $x_\alpha \sim 1$ is expected at $z \sim 25$, when the gas has cooled to ~ 15 K, while temperatures as low as $\sim 7-8$ K may be reached at $z \sim 17$ (e.g., in the plausible case of late heating), although x_α is then expected to already be fairly large. Thus, the low-temperature corrections may affect $T_{\rm b}$ by up to $\sim 20\%$ within this redshift range.

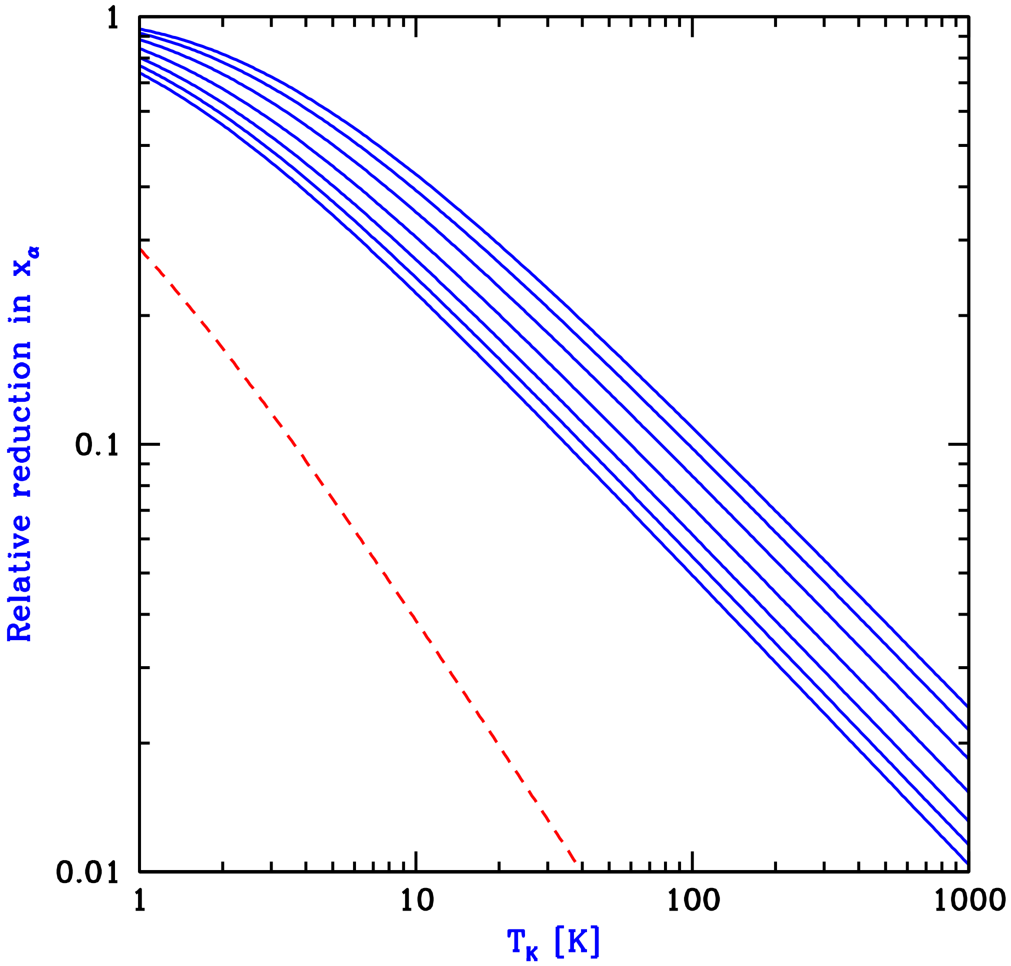

Figure 12: Relative reduction in x_α, i.e., $1 - x_{\alpha,\mathrm{eff}}/x_\alpha$, versus the kinetic gas temperature T_K. We show the total reduction (solid curves) including both the scattering and color-temperature corrections, at redshifts 7, 9, 12, 17, 25, 35, and 45 (from bottom to top), and the reduction from the modified color temperature only (dashed curve).

4.3. Anisotropy of the 21-cm signal

As explained previously, the 21-cm signal on the sky is potentially an extremely rich dataset. This signal is intrinsically three dimensional, covering the full sky over a wide range of redshifts. Even if 21-cm fluctuations are only measured statistically in terms of the isotropically-averaged power spectrum of fluctuations, this power spectrum versus redshift should yield a powerful dataset that can probe a wide range of the physics and astrophysics of the first stars and galaxies (as explored in detail in § 7).

The fluctuations in 21-cm cosmology are potentially even richer, as a result of a particular form of anisotropy that is expected due to gas motions along the line of sight [125, 126, 13]. This anisotropy, expected in any measurement of density that is based on a spectral resonance or on redshift measurements, results from velocity compression. The point is that spectral absorption is determined directly by the velocity (along the line of sight) of gas rather than its position. As an extreme example, a slab of neutral hydrogen with no internal motions will all appear to be at the same redshift from an observer, producing enormous absorption at one particular frequency and thus appearing like a huge density enhancement at the corresponding redshift, even though the real, physical density need not be high (if the slab extends over a long distance along the line of sight).

More generally, consider a photon traveling along the line of sight that resonates with absorbing atoms at a particular point. In a uniform, expanding universe, the absorption optical depth encountered by this photon probes a particular narrow strip of atoms, since the expansion of the universe makes all other atoms move with a relative velocity that takes them outside the narrow frequency width of the resonance line. If there is a density peak, however, near the resonating position, the increased gravity will reduce the expansion velocities around this point and bring more gas into the resonating velocity width. Thus, near a density peak, the velocity gradient tends to increase the 21-cm optical depth above and beyond the direct increase due to the gas density itself. This effect is sensitive only to the line-of-sight component of the gradient of the line-of-sight component of the velocity of the gas, and thus causes an observed anisotropy in the 21-cm power spectrum even when all physical causes of the fluctuations are statistically isotropic. Barkana & Loeb (2005) [13] showed that this anisotropy is particularly important in the case of 21-cm fluctuations. When all fluctuations are linear, the 21-cm power spectrum takes the form [13]

$$P_{21-\mathrm{cm}}(\mathbf{k}) = P_{\mathrm{iso}}(k) + 2\mu^2 P_{\rho-\mathrm{iso}}(k) + \mu^4 P_\rho(k)\ , \tag{58}$$

where $\mu = \cos\theta$ in terms of the angle θ between the wavevector $\mathbf{k}$ of a given Fourier mode and the line of sight, $P_{\mathrm{iso}}(k)$ is the isotropic power spectrum that would result from all sources of 21-cm fluctuations without velocity compression, $P_\rho(k)$ is the power spectrum of gas density fluctuations, and $P_{\rho-\mathrm{iso}}(k)$ is the Fourier transform of the cross-correlation between the density and all (isotropic) sources of 21-cm fluctuations. Here the velocity gradient has led to the appearance of the density power spectrum due to their simple relationship via the continuity equation. The three power spectra can more generally be denoted according to the power of μ that multiplies each term:

$$P_{21-\mathrm{cm}}(\mathbf{k}, z) = P_{\mu^0}(k, z) + 3\mu^2 P_{\mu^2}(k, z) + 5\mu^4 P_{\mu^4}(k, z)\ , \tag{59}$$

where we have defined the coefficients according to their angle-averaged size (e.g., P_{μ^4} is defined accounting for $\langle\mu^4\rangle = 1/5$), and have written the redshift dependence explicitly.

Given this anisotropic form, measuring the power spectrum as a function of μ should yield three separate power spectra at each redshift [13]. These probe, in turn, the 21-cm fluctuations without the velocity gradient term (through the μ-independent term); basic cosmology (through the intrinsic density power spectrum, measurable from the μ^4 term even when complex astrophysical processes contribute to the other terms); and additional information about the nature and properties of the various sources of 21-cm fluctuations (through the μ^2 term, which measures the cross-correlation between density fluctuations and the total isotropic 21-cm fluctuations).

In practice, 21-cm fluctuations on small scales are quite non-linear, and this non-linearity cannot be completely decoupled from large scales. In other words, even if the fluctuations are linear on a particular large scale, the way the fluctuations on that scale are measured is via a Fourier decomposition of the overall 21-cm fluctuations, which include non-linear, small-scale fluctuations. This small-scale averaging may to some degree cancel out, or largely result in an overall, simple bias factor, but the fact that the averaging involves non-linearity makes the interpretation of even large-scale measurements somewhat model-dependent. This is the double-edged sword of small-scale 21-cm fluctuations: on the one hand, they make 21-cm cosmology potentially a much larger dataset than CMB anisotropies [12], but on the other hand, they make 21-cm fluctuations more susceptible to non-linear effects (see the related discussion in § 3.2 of non-linear limits on the accuracy of analytical models).

Numerical investigations during cosmic reionization [35, 127, 128, 129] suggest that indeed, the decomposition of the line-of-sight anisotropy is more complex than the simple linear limit. It remains an incontrovertible fact, though, that the line-of-sight anisotropy makes 21-cm cosmology richer. The anisotropy allows three separate power spectra to be measured at each redshift, or more generally, a two-dimensional function of k and μ. At worst, the interpretation of this large dataset will be somewhat complicated and will need to be studied numerically, but in any case the anisotropy makes the 21-cm technique more powerful. There, are, moreover, two important caveats to these numerical studies. First, they focused on reionization (dominated by UV photons), which is a particularly difficult case as it makes the 21-cm fluctuations intrinsically non-linear on small scales, since the ionization fraction basically jumps from zero to unity in going from a neutral region to an H II bubble. And second, they focused on the μ^4 term and its promised yield of the primordial power spectrum; this term, though, is usually the smallest of the three anisotropic terms (as it does not benefit from the large bias of galaxies which enhances terms dominated by astrophysical radiation), so it is most susceptible to non-linear contamination.

Recently, Fialkov et al. (2015) [130] reconsidered the anisotropic 21-cm power spectrum using a semi-numerical simulation that covered a wide period of early cosmic history. Focusing on the dominant anisotropic term (P_{μ^2}), they showed that the anisotropy is large and thus potentially measurable at most redshifts, and it acts as a model-independent cosmic clock that tracks the evolution of 21-cm fluctuations over various eras (see Figure 13). Also, they predicted a redshift window during cosmic heating (at $z \sim 15$) when the anisotropy is small, during which the shape of the 21-cm power spectrum on large scales is determined directly by the average radial distribution of the flux received from X-ray sources at a typical point. This makes possible a direct and, again, model-independent, reconstruction of the X-ray spectrum of the earliest sources of cosmic heating.

The velocity gradient anisotropy that we have just discussed is well known in the context of galaxy redshift surveys [125], where it is often referred to as “redshift-space distortions”. In that case, it is used not as an additional probe of galaxies but of fundamental cosmology, since it allows a measurement of the amplitude of the velocity field (a recent example is [131]), which is related to the rate of change of the growth factor (section 2.2). A similar velocity gradient anisotropy also arises in the context of the Lyα forest. In that case, measurements are mostly one-dimensional (i.e., along the line of sight), so redshift distortions are more difficult to extract,

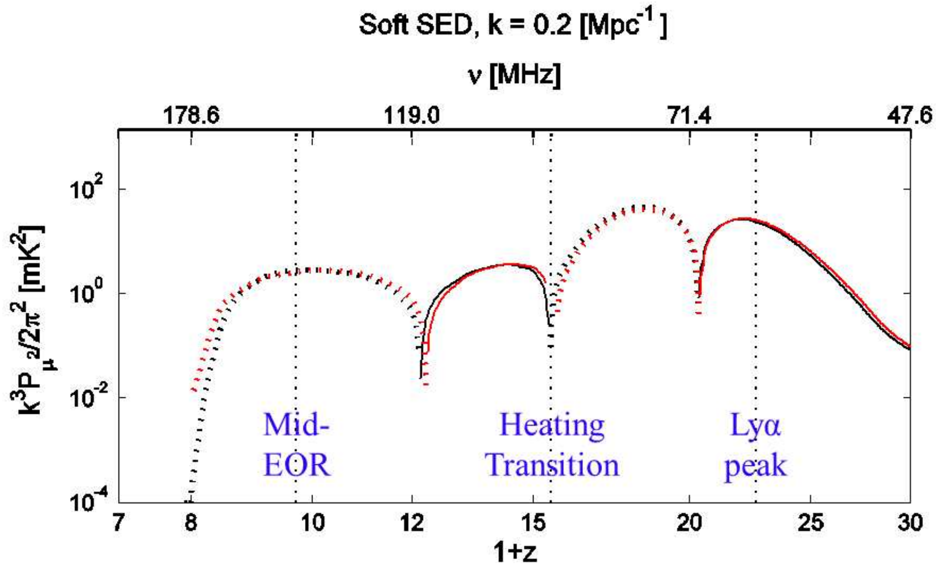

Figure 13: The main anisotropic term of the 21-cm power spectrum, $P_{\mu^2}(k, z)$, shown in terms of the corresponding variance of the 21-cm fluctuations, at wavenumber $k = 0.2$ Mpc^{-1}. The comparison of the actual value [reconstructed by fitting the form of Eq. (59) to mock observations] (red) with that from assuming perfect linear separation [as in Eq. (58)] (black) shows that this quantity withstands non-linearities and can be reconstructed accurately. This quantity, which measures the cross-correlation between density fluctuations and 21-cm fluctuations, is sometimes positive (solid lines) and sometimes negative (dashed lines), as it tracks early history like a cosmic clock. It is negative during the EOR as a direct reflection of inside-out reionization (§ 6.1): higher density implies more galaxies which implies less neutral hydrogen, hence an inverse cross-correlation of density and the 21-cm signal. It is positive during the Lyα coupling era, as more galaxies imply stronger Lyα radiation and a stronger 21-cm (absorption) signal. During the cosmic heating era, it changes sign at the heating transition (when the cosmic H I gas is first heated above the CMB temperature), the point at which heating a gas element switches from reducing the size of the 21-cm (absorption) signal to enhancing the size of the (emission) signal. The particular model shown here assumes cosmic heating by a soft power-law X-ray spectrum (see § 6.3). From [130].

though they do affect observations [132].

An additional source of 21-cm anisotropy is the light-cone anisotropy [133]. While redshift can be converted to position in order to create three-dimensional cubes for calculating the 21-cm power spectrum, the line-of-sight direction is intrinsically different from directions on the sky. The reason is that the look-back time changes with the radial distance, and the character of the 21-cm fluctuation sources evolves with time, which results in a line-of-sight effect that introduces anisotropy. A significant anisotropy can be generated on large scales near the end of reionization [133], as has been further studied in numerical simulations [134, 135, 136, 137]. It is important to clarify a possible confusing issue here (see, e.g., a clear explanation in [134]). The light-cone anisotropy refers to 21-cm fluctuations, which will be observed by radio interferometer experiments. Interferometers measure the relative fluctuations at each redshift, and are not sensitive to the mean of the 21-cm intensity at each redshift. Mathematically, this is equivalent (for a flat sky) to not being able to measure **k** modes that point directly along the line of sight ($\mu = 1$). Once the mean at each redshift is properly removed, the light-cone effect on the power spectrum is then mainly that the measured power spectrum is a redshift average of the real power spectrum, since any frequency slice corresponds to a range of redshifts within our past light cone. Looking towards the future, the light-cone anisotropy can in principle be reduced as data become available with improved sensitivity and larger fields of views, allowing the power spectrum to be measured from thin redshift slices that minimize the light-cone effect (though the slice should not be thinner than one wavelength, which implies some remaining averaging when measuring power on large scales).

Finally, if 21-cm data are analyzed using assumed cosmological parameters that differ from the true ones, this causes an additional Alcock-Paczyński [138] anisotropy that can be used to constrain cosmological parameters [14, 15]; in particular, the technique of Eq. (58) can be extended, in principle permitting (in the limit of linear fluctuations) a separate probe of this anisotropy using the μ^6 term that it induces in the 21-cm power spectrum [16].

4.4. Observational aspects

Attempts to measure the cosmological 21-cm signal must deal with the much stronger foreground emission, dominated by synchrotron radiation from electrons in the Milky Way, with other radio sources added on. Indeed, the brightness temperature of the sky for typical high-latitude, relatively quite portions of the sky, is [88]

$$T_{\rm sky} \sim 180 \left(\frac{\nu}{180\,{\rm MHz}} \right)^{-2.6} {\rm K}\ . \tag{60}$$

This steep increase of foreground emission with decreasing frequency is the reason that 21-cm observations become more difficult with increasing redshift; distortion of the radio signal due to refraction within the Earth's ionosphere also increases with redshift, down to the critical plasma frequency of $\nu \sim 20$ MHz below which the ionosphere becomes opaque. The sky emission in Eq. (60) must be compared to the expected signal of typically a few tens of mK (sky averaged), with fluctuations of order several mK. The reason that this tiny signal may be observable, even on top of a foreground that is brighter by at least a factor of 10^4, is that the foreground is produced by synchrotron emission which inherently produces a very smooth frequency spectrum.

There are a number of approaches to observing the 21-cm signal from high redshifts. The simplest, in principle, is measuring the global 21-cm signal, i.e., the sky-averaged, cosmic mean emission as a function of frequency (i.e., redshift). This can be done with a single dish (or dipole), but requires a very accurately calibrated instrument to enable foreground subtraction. Indeed, the sensitivity of a single dish [139] is

$$\Delta T \sim \frac{T_{\rm sys}}{\sqrt{\Delta\nu\, t_{\rm int}}}\ . \tag{61}$$

Assuming that the system temperature is approximately equal to that of the foreground [Eq. (60)], and taking a bandwidth of $\Delta\nu = 5$ MHz centered at $z = 10$ ($\nu = 129$ MHz), a sensitivity of $\Delta T = 10$ mK only requires an integration time $t_{\rm int}$ of 6 minutes. Thus, the real issue with global 21-cm experiments is not raw sensitivity, but the ability to clean out the smooth foreground emission to a spectral accuracy of one part in 10^4 or 10^5. In practice, the need to subtract out the smoothly-varying foreground implies a simultaneous removal of the smoothly-varying part of the desired 21-cm signal. Thus, the absolute level of the global signal likely cannot be measured, but its variation with frequency may be measurable, particularly when the frequency gradient of the 21-cm signal is large during the rises or declines that accompany various milestones of early cosmic evolution (see section 7.5).

The other main approach is to make an interferometric map of the 21-cm signal. In this case, much more information is available at each redshift than just a single mean temperature. With a sufficiently high signal-to-noise ratio, direct tomography/imaging can reveal the full spatial distribution of the 21-cm signal. However, even if the maps themselves are noisy, statistical measures such as the 21-cm power spectrum can be computed with high accuracy, and used to extract many of the most interesting aspects of cosmic dawn such as the properties of the galaxies that existed at various times. In the case of an interferometer, one basic consideration is the achievable angular resolution θ_D (and corresponding comoving spatial resolution r_D), determined by the diffraction limit corresponding to the longest array baseline $D_{\rm max}$ [88]:

$$\theta_D \sim \frac{\lambda}{D_{\rm max}} \sim 7.'3 \left(\frac{1+z}{10}\right)\left(\frac{D_{\rm max}}{1\,{\rm km}}\right)^{-1}\ ;\ \ r_D \sim 20 \left(\frac{h}{0.68}\right)^{-1}\left(\frac{1+z}{10}\right)^{1.2}\left(\frac{D_{\rm max}}{1\,{\rm km}}\right)^{-1}\ {\rm Mpc}\ . \tag{62}$$

For an array of N radio antennae (or stations), each with an effective collecting area $A_{\rm eff}$, the resulting field of view $\Omega_{\rm FoV} = \lambda^2/A_{\rm eff}$ corresponds to an angular diameter $\theta_{\rm FoV}$ (and comoving distance $r_{\rm FoV}$):

$$\theta_{\rm FoV} \equiv \sqrt{\frac{4\Omega_{\rm FoV}}{\pi}} = 5.^\circ 1 \left(\frac{1+z}{10}\right)\left(\frac{A_{\rm eff}}{700\,{\rm m}^2}\right)^{-1/2}\ ;\ \ r_{\rm FoV} \sim 0.86 \left(\frac{h}{0.68}\right)^{-1}\left(\frac{1+z}{10}\right)^{1.2}\left(\frac{A_{\rm eff}}{700\,{\rm m}^2}\right)^{-1/2}\ {\rm Gpc}\ . \tag{63}$$

In the line-of-sight direction, the comoving length corresponding to a bandwidth $\Delta\nu$ is

$$r_{\Delta\nu} \sim 18 \left(\frac{\Delta\nu}{1\,{\rm MHz}}\right)\left(\frac{1+z}{10}\right)^{1/2}\left(\frac{\Omega_m h^2}{0.141}\right)^{-1/2}\ {\rm Mpc}\ . \tag{64}$$

Another commonly noted quantity is the total collecting area of the array

$$A_{\rm coll} = N A_{\rm eff} = 1.8\times 10^5 \left(\frac{N}{250}\right)\left(\frac{A_{\rm eff}}{700\,{\rm m}^2}\right)\ {\rm m}^2\ , \tag{65}$$

where we have used illustrative values based roughly on the planned first phase of the Square Kilometer Array [140] (though note that $A_{\rm eff}$ is actually expected to vary with frequency).

A key quantity for interferometric arrays is the sensitivity to power spectrum measurements. We assume the simple approximation of antennae distributed over a core area $A_{\rm core}$ in such a way that the *uv*-density (i.e., the density in visibility space which is equivalent to a Fourier transform of the sky) is uniform, and a single beam (i.e., we do not include here the technique of multi-beaming which can speed up surveys). In this case, the power-spectrum error due to thermal noise is [141, 140]

$$\Delta T_{\rm PS}^{\rm thermal} = \sqrt{\frac{2}{\pi}}\, k^{3/4} \left[D_c^2\, \Delta D_c\, \Omega_{\rm FoV}\right]^{1/4} \frac{T_{\rm sys}}{\sqrt{\Delta\nu\, t_{\rm int}}} \frac{1}{N} \sqrt{\frac{A_{\rm core}}{A_{\rm eff}}}\ , \tag{66}$$

which yields an approximate value of

$$\begin{aligned} \Delta T_{\rm PS}^{\rm thermal} \quad \sim \quad & 0.13 \left(\frac{k}{0.1\,{\rm Mpc}^{-1}}\right)^{3/4} \left(\frac{1+z}{10}\right)^{3.3} \left(\frac{t_{\rm int}}{1000\,{\rm hr}}\right)^{-1/2} \left(\frac{\Delta\nu}{1\,{\rm MHz}}\right)^{-1/4} \\ & \times \left(\frac{A_{\rm eff}}{700\,{\rm m}^2}\right)^{-3/4} \left(\frac{A_{\rm core}}{3.8\times 10^5\,{\rm m}^2}\right)^{1/2} \left(\frac{N}{250}\right)^{-1} \left(\frac{\Omega_m h^2}{0.141}\right)^{-1/8} \ {\rm mK}\ , \end{aligned} \tag{67}$$

where D_c is the comoving distance to redshift z, and ΔD_c equals $r_{\Delta\nu}$ from above. The thermal noise thus increases with k, and typically dominates the expected power spectrum errors on small scales. Attempting to improve the angular resolution by increasing $D_{\rm max}$ would typically imply an increase in $A_{\rm core} \propto D_{\rm max}^2$ as well, and thus a worsening power-spectrum sensitivity at all k.

The uncertainty in comparing data to models is usually dominated on large scales by sample variance (sometimes termed "cosmic variance"), which gives a relative error that is roughly proportional to the inverse square root of the number of modes of wavenumber k that fit into the survey volume. Assuming a cylindrical volume and a bin width of $\Delta k \sim k$ [assumptions also made in Eq. (66)], this yields [141]

$$\Delta T_{\rm PS}^{\rm sample} \approx T_{\rm PS} \sqrt{\frac{8\pi}{k^3 r_{\rm FoV}^2\, r_{\Delta\nu}}}\ , \tag{68}$$

where $T_{\rm PS}$ is the root-mean-square 21-cm brightness temperature fluctuation at wavenumber k. The resulting approximate value is

$$\begin{aligned} \Delta T_{\rm PS}^{\rm sample} \quad \sim \quad & 0.087 \left(\frac{T_{\rm PS}}{2\,{\rm mK}}\right) \left(\frac{k}{0.1\,{\rm Mpc}^{-1}}\right)^{-3/2} \left(\frac{1+z}{10}\right)^{-1.5} \left(\frac{\Delta\nu}{1\,{\rm MHz}}\right)^{-1/2} \\ & \times \left(\frac{A_{\rm eff}}{700\,{\rm m}^2}\right)^{1/2} \left(\frac{h}{0.68}\right) \left(\frac{\Omega_m h^2}{0.141}\right)^{1/4} \ {\rm mK}\ . \end{aligned} \tag{69}$$

We note, though, that these noise estimates (both thermal noise and sample variance) may in a sense be overestimated, since they are calculated for a narrow bandwidth at a single redshift (e.g., 1 MHz around $1 + z = 20$ corresponds to $\Delta z \sim 0.3$). If a theoretical model is fit to data covering a wide range of redshifts, then the model in a sense smoothes the data over the various redshifts, yielding effectively lower noise overall. Of course, this conclusion is not model-independent as it relies on the smooth variation with redshift typically assumed in any model, a smoothness that ties together, within such a combined fit, the data measured at various redshifts. A model-independent way to try to reduce the errors would be to simply average the data over wide redshift bins, but that would erase some information about the redshift evolution, as well as features of the power spectrum that may only appear prominently at particular redshifts. A more direct observational approach is to use the flexibility available in balancing the amount of integration time spent per field (with more time leading to lower thermal noise), on the one hand, and the total number of separate fields of view observed (with more fields reducing the sample variance), on the other hand.

5. The Supersonic Streaming Velocity

Current observational efforts in 21-cm cosmology (and high-redshift astronomy more generally) are focused on the reionization era (redshift $z \sim 10$), with earlier times considered more difficult to observe. However, recent work suggests that at least in the case of 21-cm cosmology, the pre-reionization, $z \sim 20$ era of even earlier galaxies may produce very interesting signals that make observational exploration quite promising. One argument for this is based on a recently noticed effect on early galaxy formation that had been previously neglected. We discuss here this supersonic streaming velocity, which has also been reviewed recently in detail [142].

5.1. Cosmological origins

Up until recently, studies of early structure formation were based on initial conditions from linear perturbation theory. However, Tseliakhovich & Hirata (2010) [45] pointed out an important effect that had been missing in this treatment. At early times, the electrons in the ionized gas scattered strongly with the then-energetic CMB photons, so that the baryons moved together with the photons in a strongly-coupled fluid. On the other hand, the motion of the dark matter was determined by gravity, as it did not otherwise interact with the photons. Thus, the initial inhomogeneities in the universe led to the gas and dark matter having different velocities. When the gas recombined at $z \sim 1100$, it was moving relative to the dark matter, with a relative velocity that varied spatially. The root-mean-square value at recombination was ~ 30 km/s, which was supersonic (Mach number ~ 5). The streaming velocity then gradually decayed as $\propto 1/a$, like any peculiar velocity, but remained supersonic (getting down to around Mach 2) until the onset of cosmic heating. This is true for the root-mean-square value, but the streaming velocity was lower in some regions, and up to a few times higher in others.

Figure 14 shows the contribution of fluctuations on various scales to the variance of the velocity difference. This highlights two important properties of this relative motion. First, there is no contribution from small scales, so that the relative velocity is uniform in patches up to a few Mpc in size; the velocity is generated by larger-scale modes, up to ~ 200 Mpc in wavelength. The uniformity on small scales is critical as it allows a separation of scales between the spatial variation of the velocity (on large scales) and galaxy formation (on small scales). Each individual high-redshift mini-galaxy forms out of a small region (~ 20 kpc for a $10^6 M_\odot$ halo) that can be accurately approximated as having a uniform, local baryonic wind, or a uniform stream of baryons; the relative velocity is thus also referred to as the “streaming velocity”. The second important feature of Figure 14 is the strong baryon acoustic oscillation (BAO) signature. Arising from the acoustic oscillations of the photon-baryon fluid before recombination, this strong BAO signature is a potentially observable fingerprint of the effect of this relative motion, as is further detailed below.

The relative motion between the dark matter and baryons was not in itself a surprise (it had been known for decades), but before 2010 it had not been noticed that this effect was both important and dropped within the standard approach. The standard initial conditions for both analytical calculations and numerical simulations had been generated based on linear perturbation theory, in which each **k** mode evolves independently. Indeed, the relative velocity is negligible if any

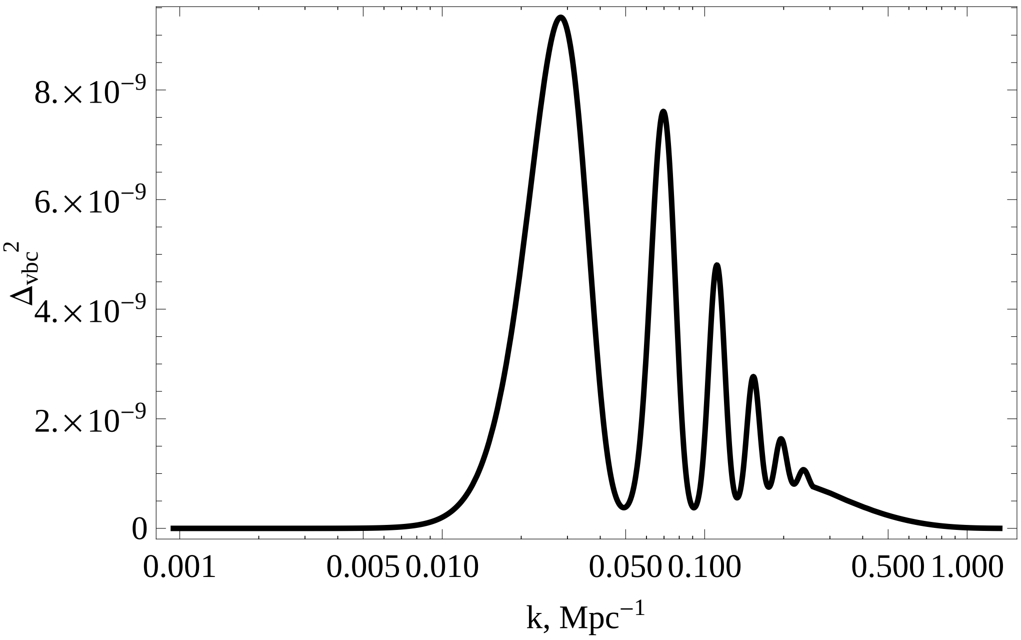

Figure 14: The contribution of various scales to the mean squared velocity difference between the baryons and dark matter (at the same position) at recombination. The contribution per $\log k$ of fluctuations at wavenumber k is shown vs. k. From [45].

single scale is considered. However, it *is* important as an effect of large scales (which contribute to the velocity difference) on small scales (which dominate early galaxy formation). Specifically, the relative motion makes it harder for small-scale overdensities in the dark matter to gravitationally accrete the streaming gas. Now, observing such small scales directly would require far higher resolution than is currently feasible in radio astronomy at high redshift. Nonetheless, the relative motion is immensely important because of its effect on star formation. Since stellar radiation strongly affects 21-cm emission from the surrounding IGM, 21-cm cosmology offers an indirect probe of the relative velocity effect.

5.2. Effect on star formation in early halos

The effect of the streaming velocity on early star formation can be usefully separated into three effects, both for physical understanding and for the purposes of analytical modeling. This also tracks the development of the subject. The first effect of the streaming velocity on halos to be analyzed was the suppression of the abundance of halos [45]. Since the baryons do not follow the dark matter perturbations as closely as they would without the velocity effect, linear fluctuations in the total density are suppressed on small scales (where the gravitationally-induced velocities are comparable to or smaller than the relative velocity). According to the standard theoretical models for understanding the abundance of halos as a function of mass [53, 54] (§ 2.3), this should result in a reduction of the number density of high-redshift halos of mass up to $\sim 10^6 M_{\odot}$ [45], a mass range that is expected to include most of the star-forming halos at early times.

The next effect to be noted [143] was that separately from the effect on the number of halos that form, the relative velocity also suppresses the gas content of each halo that does form. It was initially claimed [143] that this second effect results in 2 mK, large-scale 21-cm fluctuations during Lyα coupling, with a power spectrum showing a strong BAO signature due to the streaming velocity effect. These conclusions were qualitatively on the mark but were later seriously revised quantitatively. In particular, it turned out [144, 9] that the gas-content effect is a minor one on star-forming halos, and is mainly important for the lower-mass gas minihalos that do not form stars.

Meanwhile, many groups began to run small-scale numerical simulations that followed individual collapsing halos subject to the streaming velocity [145, 146, 147, 148, 149, 150, 151]. In particular, two simulations [146, 147] first indicated the presence of a third effect, i.e., that the relative velocity substantially increases the minimum halo mass for which stars can form from gas that cools via molecular hydrogen cooling. The intuitive explanation is that even if a halo does manage to form (albeit with a reduced gas content), it does not contain the same dense gas core that it would in the absence of the streaming velocity. The reason is that the densest part of the halo (which is where stars first manage to form) comes together well before the rest of the halo, and is thus strongly disrupted by the streaming velocity (which is high at early times); thus, after a halo forms in the presence of the streaming velocity, it is necessary to wait longer for a dense core to develop and bring about star formation. Given these simulation results on the increase in the minimum halo mass for star formation, a physically-motivated fit [9] allowed the development of a general analytical model of early star formation that includes the effect of density as well as all three effects of the streaming velocity on star formation.

Figure 15 illustrates some of the results of the numerical simulation studies of the effect of the streaming velocity on galaxy formation. As expected, a larger velocity suppresses gas accretion more strongly, in particular reducing the amount of dense gas at the centers of halos. But beyond just this general trend, the relative velocity effect gives rise to very interesting dynamics on small scales. It disrupts gas accretion in an asymmetric way, so that filaments of accreting gas are

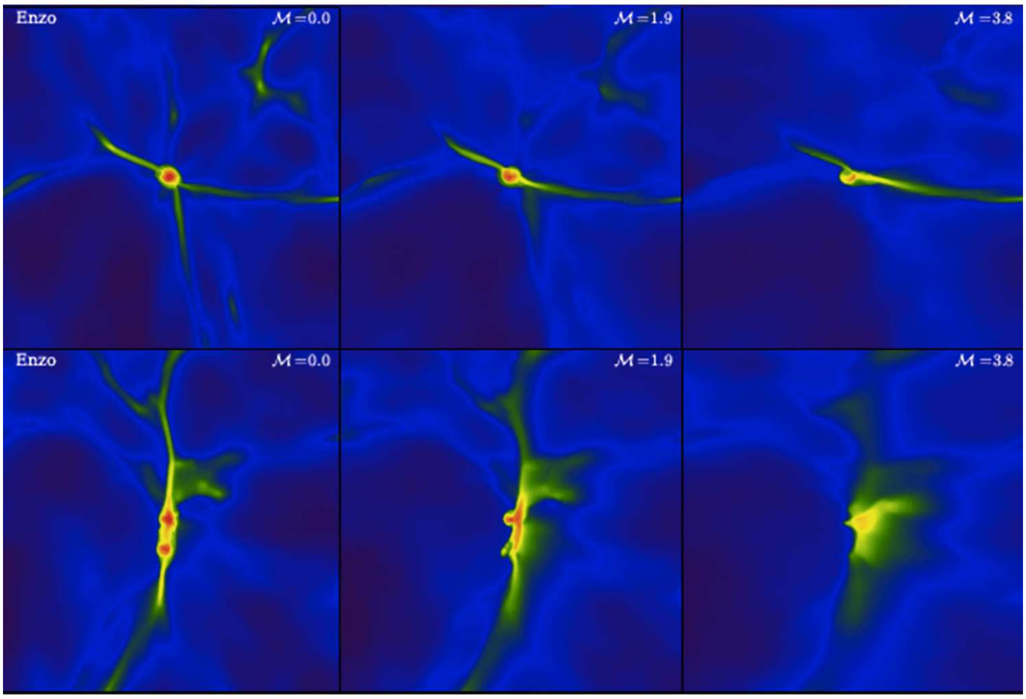

Figure 15: The effect of relative velocity on individual halos, from numerical simulations (including gravity and hydrodynamics). The colors indicate the gas density, which ranges from $10^{-26}\mathrm{g/cm}^3$ (blue) to $10^{-23}\mathrm{g/cm}^3$ (red). Two halos are shown at $z = 20$, with a total halo mass of $2 \times 10^6 M_\odot$ (top panels) or $8 \times 10^5 M_\odot$ (bottom panels). Panels show the result for gas initially moving to the right with a relative velocity of 0 (left panels), 1 (middle panels), or 2 (right panels) in units of the root-mean-square value of the relative velocity. $\mathcal{M}$ indicates the corresponding Mach number at $z = 20$. From [148].

disrupted more easily if they are perpendicular to the local wind direction. In addition, halos that form in regions of relatively high velocity develop supersonic wakes as they move through the wind.

5.3. Consequences

The immediate major consequence of the streaming velocity effect is the change in the large-scale distribution of the first stars in the Universe, and the resulting pattern embedded in the 21-cm sky at very high redshift. All of this is discussed below, particularly in § 7.3, where the distribution of the streaming velocity field is also shown (Figure 33). Here we note some other interesting consequences of the streaming velocity that have been suggested.

Although the relative velocity only affected low-mass halos at high redshifts, those halos were the progenitors of later, more massive galaxies. Thus, the streaming velocity may have indirectly left a mark on later galaxies through its influence on their star-formation histories and, thus, on their current luminosity (through their old stellar content and perhaps its feedback on the formation of younger stars). This signature may be observable in galaxy surveys, and could affect probes of dark energy through measurements of BAO positions in the galaxy power spectrum [152]; indeed, current data imply an upper limit of 3.3% on the fraction of the stars in luminous red galaxies that are sensitive to the relative velocity effect [153]. More directly, the early streaming velocity effect on star formation in dwarf galaxies could leave remnants in their properties as measured today, e.g., in the low-mass satellites of the Milky Way [154].

We note, though, that when considering these effects on later galaxies, it is important to keep in mind the modulation of star formation by other effects, in particular LW radiation (§ 6.2) that suppresses molecular hydrogen cooling, and reionization, which suppresses gas accretion through photoheating feedback (§ 6.1). These effects suppressed star formation in larger halos than the streaming velocity itself, which means that they affected later progenitors of current galaxies (containing a larger fraction of the final, present-day stellar content). The distribution of LW feedback may have reflected in part the initial relative velocity pattern [155], since the LW radiation itself was produced by stars in small halos, but reionization occurred later, likely due to more massive halos (§ 6.1) that were not affected much by the streaming velocity. Thus, photoheating likely did not carry a significant signature of the streaming velocity field.

Moving towards higher redshifts, as mentioned, the streaming velocity likely did not significantly affect the main stages of cosmic reionization. However, it suppressed the formation of earlier cosmic populations, perhaps including supermassive black holes at $z > 15$ [156]. More intriguing (and speculative) are ideas on opposite effects, whereby a large streaming velocity may have produced a unique environment that allowed some objects to form. A large relative velocity may have delayed star formation enough to allow a direct collapse to a massive black hole [157], or it may have produced a baryonic density peak that was sufficiently displaced from dark matter to allow the formation of an early globular cluster [158]. Moving on to the dark ages ($z > 30$), the supersonic streaming velocity had a number of significant effects on the 21-cm power spectrum at both large and small scales [159].

6. Cosmic milestones of early radiative feedback

6.1. Reionization

The reionization of the Universe is an old subject. The observation of transmitted flux shortward of the Lyα wavelength of quasars indicated in 1965 that the modern Universe is highly ionized [22]. While this led to a gradual growth of literature on the theoretical development of cosmic reionization (as summarized, e.g., in [23]), calculations in the context of modern cosmological models of hierarchical galaxy formation were first made in the 1990's. These included the first numerical simulations of cosmic reionization [160, 161], and analytical calculations [162, 163, 164, 165, 166, 167] that mostly followed the overall, global progress of reionization, based on counting the ionizing photons from the rapidly rising star formation while accounting for recombinations. Exploration of the 21-cm signatures of reionization began in one of these numerical simulations [161] and in theoretical papers by Shaver et al. (1999) and Tozzi et al. (2000) [139, 27].

There soon began a more detailed discussion of the structure and character of reionization, important issues for a variety of observational probes of the era of reionization, especially 21-cm cosmology. A commonly-assumed simple model was that of instantaneous reionization, often adopted in calculations of the effect of reionization on the CMB. This was supported by simulations [161, 168] that showed a rapid "overlap" stage whereby the transition from individual H II regions around each galaxy to nearly full reionization was rapid ($\Delta z \sim 0.1$). Fast reionization would have made it easier to detect reionization through a sudden jump in the number of faint Lyα sources [169, 170] (given the strong Lyα absorption due to a neutral IGM).

These same simulations also found that the H II regions during reionization were typically quite small, below a tenth of a Mpc for most of reionization until a sudden sharp rise (to larger than the simulation box) once only 30% of the hydrogen mass (occupying 15% of the volume) remained neutral. Predictions made on this basis [171] were bad news for 21-cm observations, which will find it difficult to reach the angular resolution required to see such small features within the cosmological 21-cm signal. Modeling of the effect of reionization on secondary CMB anisotropies through the kinetic Sunyaev-Zel'dovich effect (whereby the velocities of free electrons created by reionization changed the energies of the fraction of CMB photons that rescattered) also assumed that the ionized bubble scale would be very small unless quasars were dominant [172, 173, 174].

Another basic issue about reionization is its structure/topology. At this time, both analytical models and numerical simulations [175, 168] suggested that reionization would be outside-in (with most ionizing photons leaking to the voids and reionizing them first, leaving the dense regions for later) rather than inside-out (which is when the high-density regions around the sources reionize before the low-density voids). All of the just-noted conclusions were based on numerical simulations with box sizes below 10 Mpc. A simulation of a 15 Mpc box found some ionized regions as large as 3 Mpc [176]. An even larger, 30 Mpc simulation [177] considered a field (average) region and a proto-cluster (i.e., an overdense region), and found substantial differences between their reionization histories (thus suggesting fluctuations on quite large scales), but still supported an outside-in reionization (since the proto-cluster reionized later than the field region). In hindsight, most of the results summarized in this and the previous two paragraphs were incorrect or confusing.

The now-accepted paradigm of reionization began to emerge when Barkana & Loeb (2004) [24] realized that the surprisingly strong clustering of high-redshift halos (see section § 3.1) leads to H II bubbles driven by multiple clustered galaxies rather than individual galaxies[2] (see Figures 16 and 17). This clustering is significant even on scales of tens of Mpc, leading to typical bubble sizes during reionization that are larger than the total box size of most numerical simulations of reionization at the time. The strong bias of high-redshift galaxies also settled the issue of the topology of reionization [24], showing that it is inside-out; while the recombination rate was higher in overdense regions because of their higher gas density, these regions still reionized first, despite the need to overcome the higher recombination rate, since the number of ionizing sources in these regions was increased even more strongly as a result of the strong bias of galaxies[3]. The outside-in picture, though, is still useful, as it seems likely to apply to the internal structure of individual H II bubbles and to the post-reionization universe. Another important revision was in the common view of the effect of reionization on the abundance of dwarf galaxies in various environments [24].

The next big step was taken by Furlanetto et al. (2004) [25], who created an analytical model for the distribution of H II bubble sizes (Figure 18), based on an ingenious application of the

---

[2] This paper [24] was first submitted in August 2003 but was only published 11 months later due to initial resistance to its novel conclusions.

[3] Quantitatively, the number of hydrogen atoms that must be initially ionized in each region is proportional to its density, i.e., the effective linear bias (Eq. 30) for this quantity is unity. The number of recombinations that must be overcome goes as density squared, so its effective bias is 2. The high-redshift galaxies that are thought to have sourced reionization likely had a bias above 2 throughout reionization, with a more typical value of 5 or 10.

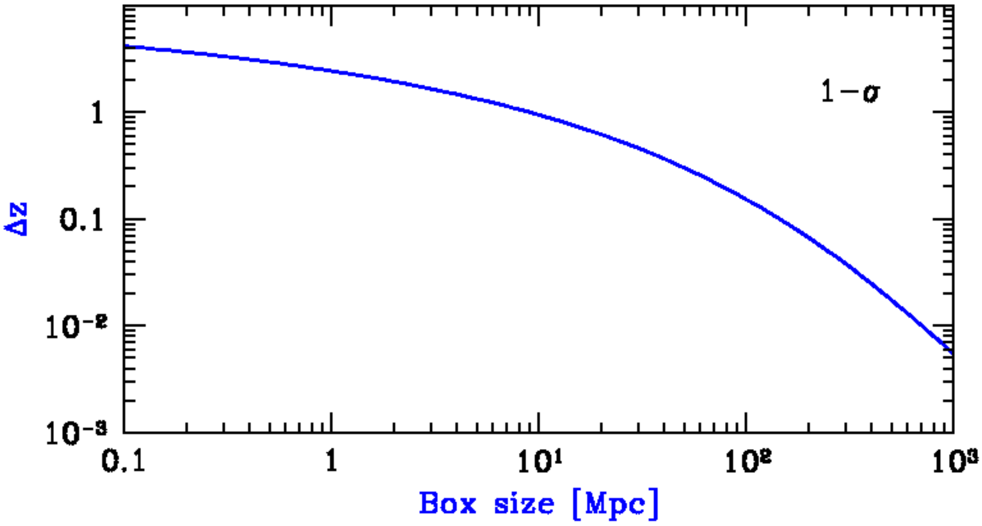

Figure 16: Cosmic $1 - \sigma$ scatter in the redshift of reionization, or any other phenomenon that depends on the fraction of gas in galaxies, versus the size of a rectangular region (in the Universe or in a simulation). When expressed as a shift in redshift, the scatter is predicted to be approximately independent of the typical mass of galactic halos. Regions of size 10 Mpc are not representative and do not yield an overall picture of reionization, since different regions of that size reionize at redshifts that differ by a $1 - \sigma$ scatter of $\Delta z \sim 1$. One hundred Mpc boxes are required in order to decrease Δz to well below unity (~ 0.15). From [24].

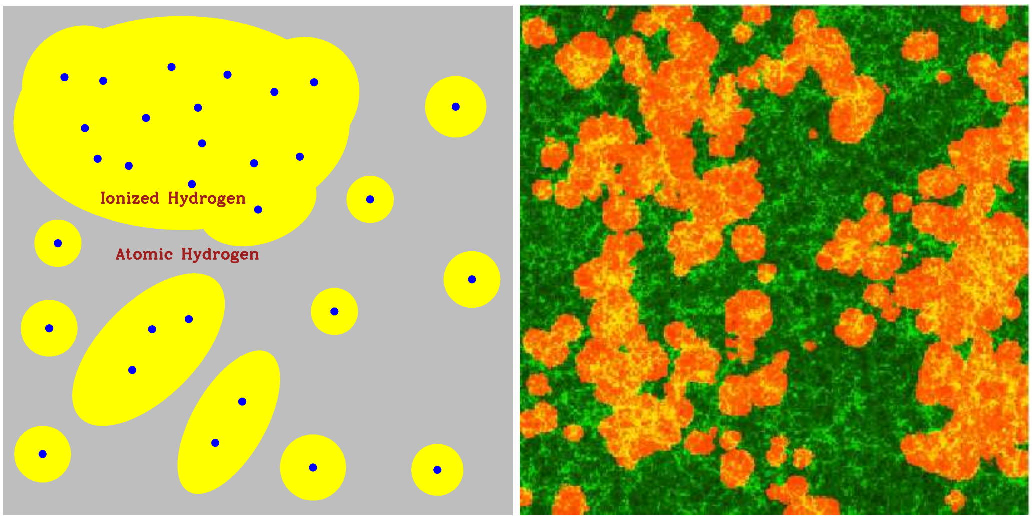

Figure 17: During reionization, ionized bubbles were created by clustered groups of galaxies [24]. The illustration (left panel, from [21]) shows how regions with large-scale overdensities formed large concentrations of galaxies (dots) whose ionizing photons produced large ionized bubbles. At the same time, other large regions had a low density of galaxies and were still mostly neutral. A similar pattern has been confirmed in large-scale numerical simulations of reionization (e.g., the right panel shows a two-dimensional slice from a 150 Mpc simulation box [178]). Multiple-source bubbles likely dominated the ionized volume from as early as $z \sim 20$ [179].

extended Press-Schechter model [54]. This showed how the typical size rises gradually during reionization, from a few Mpc to tens of Mpc during the main stages, and allowed an estimate of the resulting 21-cm power spectrum during reionization. This picture of reionization based on semi-analytic models [24, 25] was then confirmed by several numerical simulations that reached sufficiently large scales with boxes of $\sim$ 100 Mpc in size (e.g., [26, 180, 181]). The simulations indeed showed the dominance of large bubbles due to large groups of strongly-clustered galaxies, though it should be noted that the price of such large boxes was (and remains) a limited ability to resolve the small galaxies that were likely the dominant sources of reionization.

This realization, that reionization was characterized by strong fluctuations on large scales even if the individual galaxies that caused it were small, has been very important and influential. It has helped motivate the large number of observational efforts currently underway in 21-cm cosmology (§ 1), since large-scale fluctuations are easier to detect (as they do not require high angular resolution; see Eq. (67)).

Today there remain some major uncertainties about reionization that will likely only be resolved by 21-cm measurements (§ 7.1). In terms of the overall timing, the best current constraint comes from large-angle polarization measurements of the CMB which capture the effect of the rescattering of CMB photons by the reionized IGM. The latest measured optical depth of $5.5 \pm 0.9\%$ [182] implies a reionization midpoint at a redshift of $7.5 - 9$ in realistic models (with reionization completing somewhere in the range $z = 6 - 8$). However, the best-fit optical depth has changed substantially with every new measurement (declining over time), and in general it is more difficult to constrain small values of the optical depth since the corresponding reionization signature on the CMB is then smaller compared to systematic errors. The CMB results do strongly limit the high-redshift onset of reionization, with a limit of less than $\sim$ 10% completion by $z = 10$ [183]. There have long been hints of a late end to reionization at $z \sim 6$ [184, 185, 186, 187, 188], but they have been controversial due to the expected large fluctuations in the cosmic ionizing background even after full reionization of the low-density IGM [189, 190, 191, 192, 193].

As far as the typical halo masses that hosted the dominant sources of reionization, it is expected that Lyman-Werner radiation dissociated molecular hydrogen early on [20], so that by the central stages of reionization star formation required atomic cooling, with a minimum halo mass for star-formation of $\sim 10^8 M_\odot$. As reionization proceeded, the hot gas within ionized regions raised the gas pressure and prevented it from falling into small gravitational potential wells; this photoheating feedback gradually eliminated star formation in halos up to a mass of $\sim 3 \times 10^9 M_\odot$, as has been studied in many calculations and numerical simulations [194, 195, 98, 196, 197, 198, 161, 199, 200, 72, 201, 77, 78]. In particular, this means that an era of active star formation in dwarf galaxies prior to reionization may be observable directly with next-generation telescopes [202, 203], or in the star formation histories of massive high-redshift galaxies [204], although this depends also on the effectiveness of supernova feedback in small galaxies [205, 206].

Another interesting issue related to reionization is that of minihalos, i.e., low-mass halos that collect gas but do not form stars due to the lack of sufficient cooling. These minihalos formed in large numbers, clustered strongly around ionizing sources, and contained enough gas to effectively block most ionizing photons [207, 208]. However, the minihalos naturally photoevaporated once engulfed by H II regions [209, 210], making their effect on reionization (which they delay) and on 21-cm emission only modest [211, 212]. We note that due to their low masses, minihalos were also strongly affected by the baryon - dark matter streaming velocity (§ 5).

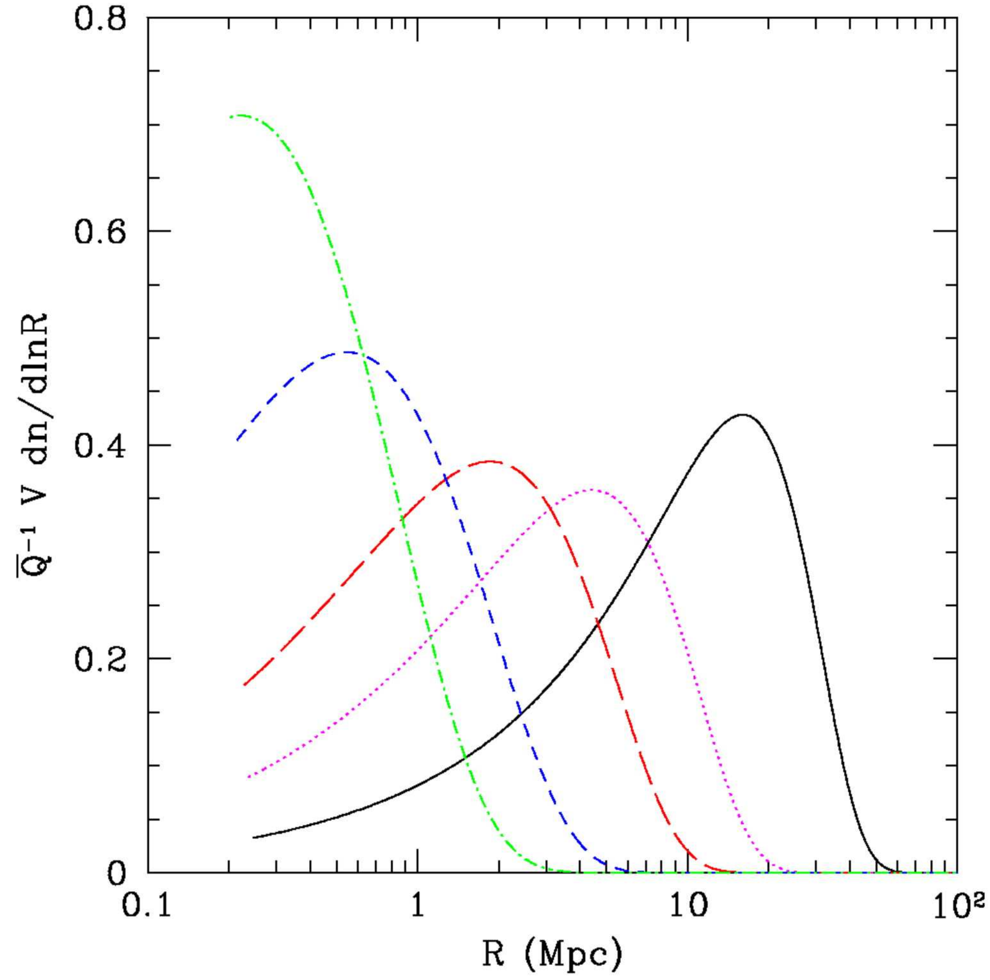

Figure 18: Distribution of H II bubble sizes during reionization. The fraction of the ionized volume in bubbles of radius R is shown per $\log R$ interval. Dot-dashed, short-dashed, long-dashed, dotted, and solid lines are for z = 18, 16, 14, 13, and 12, respectively, in a model in which the cosmic ionized fraction at these times is 0.037, 0.11, 0.3, 0.5, and 0.74, respectively. From [25].

6.2. Lyα coupling and Lyman-Werner feedback

The general course of cosmic history as relevant to 21-cm cosmology was outlined in § 1, and the physics of the 21-cm transition (including Lyα coupling) was described in detail in § 4. Here we briefly summarize Lyα coupling and LW feedback, as they are among the most important observable events in early cosmic history.

The IGM can be observed in 21-cm emission or absorption, relative to the CMB background, only if the hyperfine levels of the hydrogen atom are not in equilibrium with the CMB. This means that the spin temperature must differ from the CMB temperature. At the highest redshifts, atomic collisions overcome the scattering of CMB photons and drive the spin temperature to the kinetic temperature of the gas. However, this becomes ineffective at $z \sim 30$, and the spin temperature then approaches the CMB temperature. Luckily for 21-cm cosmologists, stellar Lyα photons come to the rescue [11], moving the spin temperature back towards the kinetic temperature through the indirect Wouthuysen-Field effect [18, 19]. The Lyα coupling era refers to the time during which the Lyα flux reaches and passes the level needed for effective 21-cm coupling.

Unlike reionization and heating, Lyα coupling and Lyman-Werner feedback are not cosmic events that change the overall state of the IGM. Lyα coupling is basically a 21-cm event, and it is important because of the prospect of detecting 21-cm emission from the early era ($z \sim 20-30$ [213, 214]) of Lyα coupling. A 21-cm observation of Lyα coupling (see § 7.2 for more details) is the only currently feasible method of detecting the dominant population of galaxies from such high redshifts and measuring their properties, either through a global 21-cm detection of the strong mean absorption signal or by interferometric measurement of the substantial 21-cm fluctuations expected from this era [28]. While still far from the very first stars at $z \sim 65$ [8, 9], this is the highest redshift range currently envisioned for observing the dominant galaxy population, a feat which would be very exciting.

LW feedback is a major feedback effect on the first stars. It indirectly affects the IGM and the 21-cm sky through its effect on the radiative output from stars (including Lyα, X-ray, and ionizing radiation). LW feedback dissociates molecular hydrogen and thus it ended star formation driven by molecular cooling [20] in halos of $\sim 10^6 M_\odot$ [7, 6]. If the overall (time-averaged) star-formation efficiency in such small, early halos was significant, then their LW radiation is expected to have produced significant feedback early on ($z \sim 20-25$) [20, 215, 213, 155], at a time when these halos still dominated the global star formation. This feedback strengthened gradually as the LW intensity increased, as has been found in numerical simulations that imposed a LW background on forming early galaxies (either constant with time [216, 217, 218] or increasing more realistically [219]). Because of its gradual rise, LW feedback did not actually halt or reduce the global star formation, but it did slow down the otherwise rapid rise of star formation at high redshifts. Like other inhomogeneous negative feedbacks, LW feedback increased cosmic equality by first suppressing the sites of earliest star formation [215, 213, 155] (Figure 19). While some LW photons reached out to a distance of ~ 100 Mpc from each source, the feedback was more local than that; emission from distant sources was absorbed more weakly, so that half the effective LW flux seen at a given point came from sources within ~ 15 Mpc away (Figure 20).

A discussion of the 21-cm signatures of the Lyα coupling and LW feedback eras is deferred to § 7.2. We note that in this topic it is essential to include the baryon - dark matter streaming velocity (§ 5) as well, since it affects the same halos as the LW feedback, and these same halos may have dominated star formation during the Lyα coupling era.

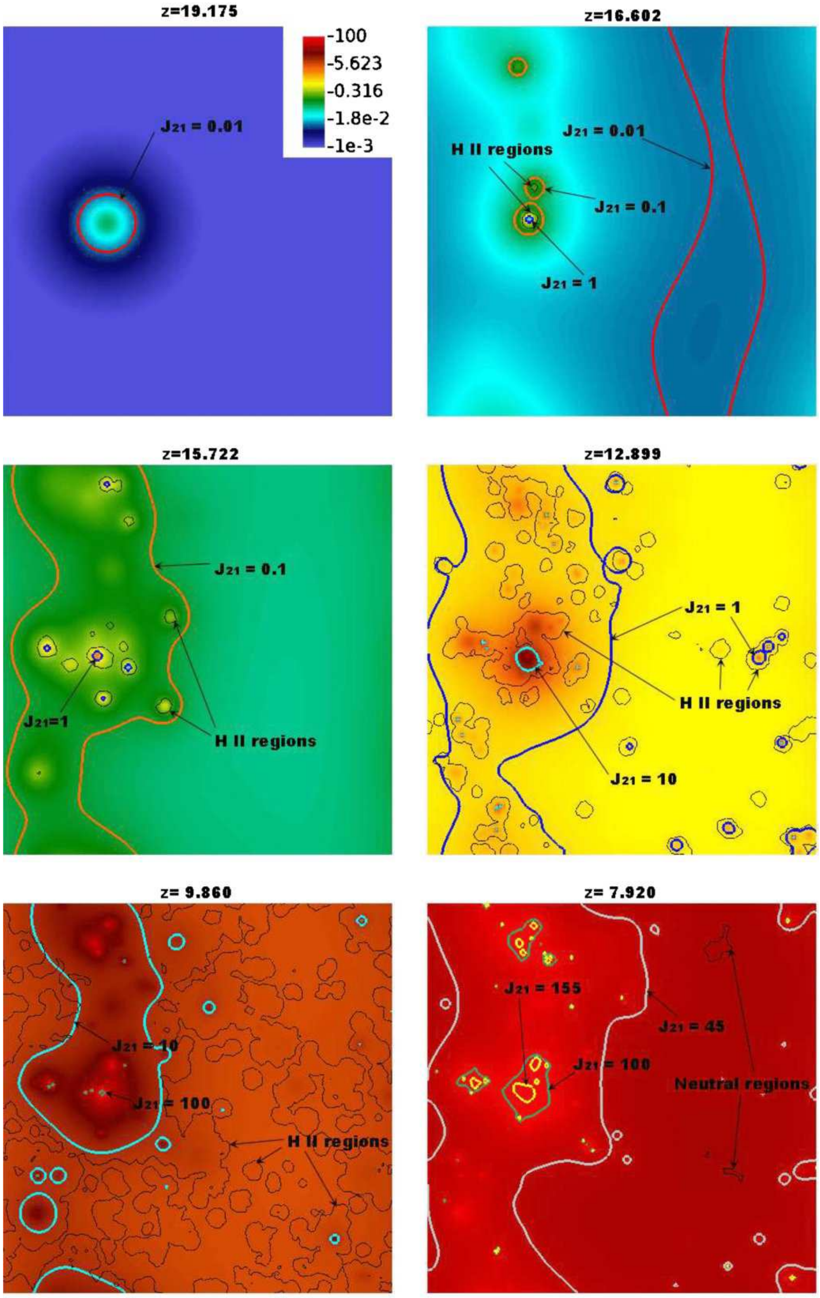

Figure 19: Spatial images from a simulation showing the isocontours of patchy reionization and the patchy H_2-dissociating background on a planar slice through a box of volume $(35/h\ \mathrm{Mpc})^3$ at various epochs. The level of $J_{\mathrm{LW},21}$ (the LW photon intensity in units of 10^{-21} erg cm^{-1} s^{-1} Hz^{-1} sr^{-1}) on the grid is depicted by various colors, with the range $[10^{-3} - 10^{2}]$ shown on the inset of the top-left panel. On top of each $J_{\mathrm{LW},21}$ color map, contours of thick colored lines represent different $J_{\mathrm{LW},21}$ levels (red, orange, blue, cyan, and green corresponding to $J_{\mathrm{LW},21} = 0.01$, 0.1, 1, 10, and 100, respectively). The black lines represent ionization fronts. From [215].

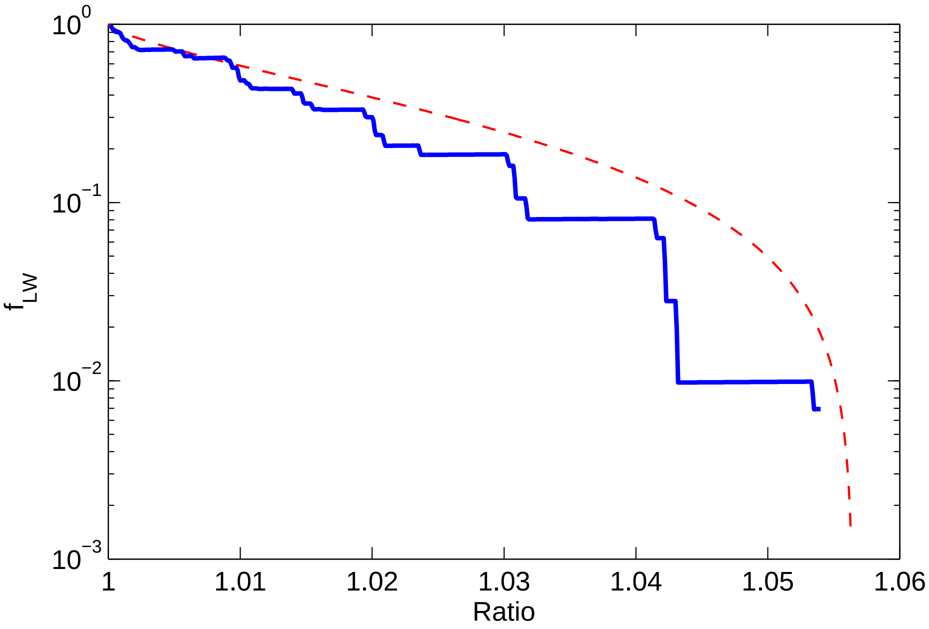

Figure 20: The relative effectiveness of causing H_2 dissociation in an absorber at z_a due to stellar radiation from a source at z_s, shown versus the ratio $R \equiv (1+z_s)/(1+z_a)$ since in this form this function is independent of redshift. The complex result (solid curve) incorporates the expected stellar spectrum of Population III stars [220, 28], along with the full list of 76 relevant LW lines [20]. Beyond the max shown $R = 1.054$ (which corresponds to 104 comoving Mpc at $z = 20$), $f_{\rm LW}$ immediately drops by five orders of magnitude. Also shown is a commonly used approximation [215] (dashed curve) which is based on a flat, averaged LW spectrum. Both functions are normalized to unity at $R = 1$. From [155].

6.3. Cosmic heating

Before discussing heating in the context of 21-cm cosmology, we begin with a brief summary of the basic physics of X-ray heating. The comoving mean free path of an X-ray photon, to photoelectric absorption in a universe of neutral fraction $x_{\rm H\,I}$, is[4] [88]

$$\lambda_X \approx 51\, x_{\rm H\,I}^{-1} \left(\frac{1+z}{10}\right)^{-2} \left(\frac{E}{0.5\,{\rm keV}}\right)^{3} \ {\rm Mpc}\ . \tag{70}$$

For photons of energy $E \gtrsim 1$ keV, λ_X becomes a significant fraction of the horizon (Eq. 12), and in that case cosmological redshift effects lead to a substantial loss of energy between emission and absorption (plus there is a significant time delay between these two events). Once the X-rays are absorbed, the resulting (primary) fast electrons then interact with the surrounding gas through the processes of collisional excitation, ionization, and electron-electron scattering. These secondary processes quickly distribute the original X-ray energy into ionization (of hydrogen and helium in the IGM), heating (i.e., thermalized energy), and excitation (which results in low energy photons that then escape, so that the energy is effectively lost). The fraction of energy that goes into heating varies with the ionization fraction of the background medium, from around a third of the energy in a neutral medium up to nearly all of the energy in a highly ionized one [88, 221, 222, 223].

It has long been known that the Universe was reionized at an early time (§ 6.1) and thus heated to at least $\sim 10,000$ K by the ionizing photons. While reionization was a major phase transition in the IGM, the question of whether the gas had been radiatively pre-heated prior to reionization is also important. Significant pre-heating of the IGM directly affects 21-cm observations, and also produces some photoheating feedback (though much weaker than that due to reionization).

The dependence of the 21-cm brightness temperature on the kinetic temperature T_K of the gas takes the form $T_{\rm b} \propto [1 - T_{\rm CMB}/T_K]$ (Eq. 50 or 56). Thus, the midpoint of the heating era, or the central moment of the "heating transition", refers to the moment when the mean gas temperature is equal to that of the CMB, so that the cosmic mean $T_{\rm b}$ is zero; actually, the latter would be true in a universe with purely linear fluctuations, but non-linearities delay the time when $\langle T_{\rm b} \rangle = 0$ by an extra $\Delta z \sim 0.5$ [155]. Also, clearly $T_{\rm b}$ is more sensitive to cold gas than to hot gas (relative to the CMB temperature). Indeed, at early times the 21-cm absorption can be very strong (depending on how much the gas cools), but at late times, once $T_K \gg T_{\rm CMB}$, $T_{\rm b}$ becomes independent of T_K and the 21-cm emission is said to be in the "saturated heating" regime.

For a long time it was confidently predicted that the universe was well into the saturated heating regime once cosmic reionization got significantly underway. The stage for this widespread belief was set by the landmark paper in 21-cm cosmology by Madau et al. (1997) [11]. They considered several possible heating sources, mainly X-rays from quasars (later observed to disappear rapidly at $z > 3$, e.g., [224]) and heating from Lyα photons (later shown to be negligibly small [120, 225, 226]). However, stellar remnants – particularly X-ray binaries (Figure 21) – have become the most plausible source of cosmic heating. This is the result of a combination of basic facts: 1) X-rays travel large distances even through a neutral IGM; 2) Large populations of X-ray binaries should have formed among the stellar remnants associated with the significant

---

[4] In Eq. (70), the power-law dependence of λ_X on $x_{\rm H\,I}$ is -1; it has sometimes been incorrectly listed as +1/3 [88] or -1/3 [29].

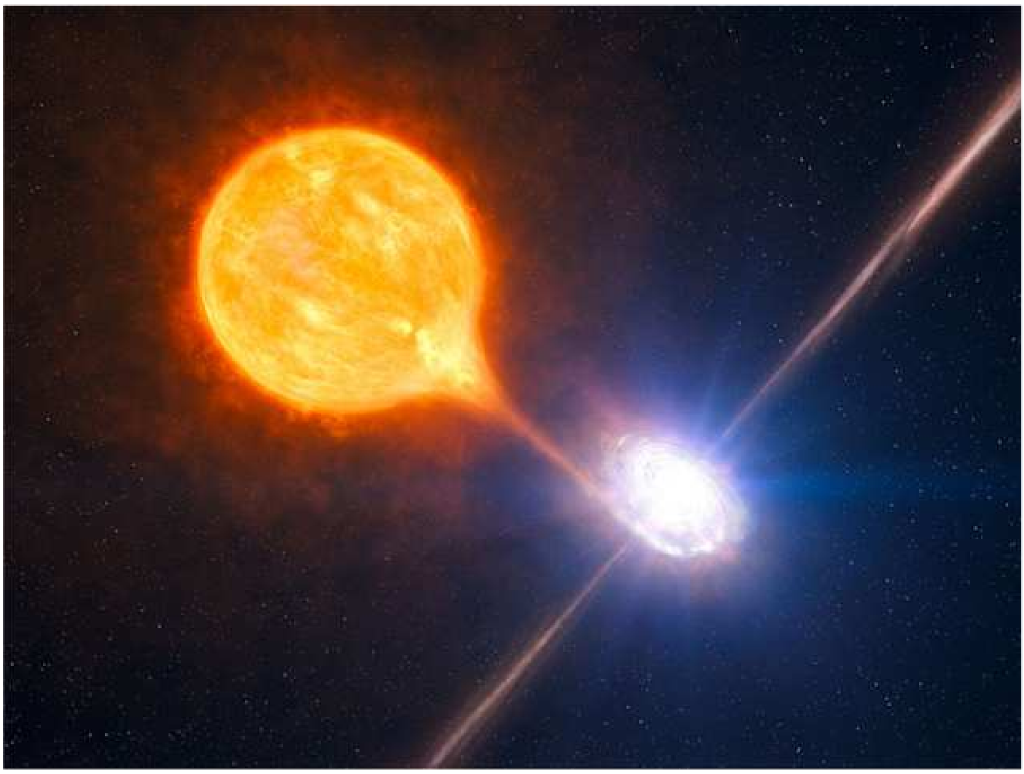

Figure 21: The most plausible sources for cosmic heating before reionization are early X-ray binaries, dominated by black-hole binaries [228] like the one illustrated here, where material from a companion spills onto a black hole, resulting in X-ray emission from its accretion disk. Credit: ESO drawing from http://en.wikipedia.org/wiki/File:A_stellar_black_hole.jpg .

cosmic star formation that we know must have occurred in order to reionize the universe; 3) Observations of the local Universe suggest not only that X-ray binaries form wherever star formation is found, but that their relative populations increase by an order of magnitude at the low metallicity expected for high-redshift galaxies [227, 228, 229, 230, 231].

Even with X-ray binaries as the plausible source, the common expectation of saturated heating before reionization had remained, and had been assumed in many mock analyses made in preparation for upcoming data ([232] is a recent example). A key reason for this is that until recently, calculations of cosmic X-ray heating [41, 29, 233, 91, 234] had assumed power-law spectra that place most of the X-ray energy at the low-energy end, where the mean free path of the soft X-rays is relatively short. This means that most of the emitted X-rays are absorbed soon after they are emitted, before much energy is lost due to cosmological effects. The absorbed energy is then enough to heat the gas by the time of reionization to $\sim$ 10 times the temperature of the CMB [92].

However, Fialkov et al. (2014) [92] recognized that the assumed X-ray spectrum is a critical parameter for both the timing of cosmic heating and the resulting 21-cm signatures. The average radiation from X-ray binaries is actually expected to have a much harder spectrum (Figure 22) whose energy content (per logarithmic frequency interval) peaks at $\sim$ 3 keV. Photons above a (roughly redshift-independent) critical energy of ~ 1 keV have such a long mean free path that by the start of reionization, most of these photons have not yet been absorbed, and the absorbed ones came from distant sources that were effectively dimmed due to cosmological redshift effects. This reduces the fraction of the X-ray energy absorbed as IGM heat by about a factor of 5, enough to push the moment of the heating transition into the expected redshift range of cosmic reionization (and thus, we will refer to this case as late heating). For this and other reasons, the spectrum of the X-ray heating sources is a key parameter for 21-cm cosmology, as further discussed in § 7.3.

Based on low-redshift observations, other potential X-ray sources appear sub-dominant compared to X-ray binaries. One such source is thermal emission from hot gas in galaxies, which has a relatively soft X-ray spectrum. Its X-ray luminosity in local galaxies [238] is (for a given star-formation rate) about a third of that of X-ray binaries. Given the above-mentioned order-of-magnitude increase expected in the emission from X-ray binaries at high redshift, the thermal gas would have to be highly efficient at high redshift in order to contribute significantly. Also, some theoretical arguments suggest that X-rays produced via Compton emission from relativistic electrons in galaxies could be significant at high redshift [239], though again the increase would have to be very large compared to such emission in low-redshift sources; the expected spectrum in this case (flat from ~ 100 eV to ~ 100 GeV) would deliver most of the energy above 1 keV and thus count as a hard spectrum in terms of 21-cm signatures. Another possible heating source, large-scale structure shocks, is likely ineffective [240, 241, 149].

A natural X-ray source to consider is the population of bright quasars. As noted above, while quasars are believed to dominate the X-ray background at low redshift [242], their rapid decline beyond $z \sim 3$ [224] suggests that their total X-ray luminosity (including an extrapolation of their observed luminosity function) is sub-dominant compared to X-ray binaries during and prior to reionization [228]. The rarity of quasars at early times is natural since they seem to be hosted mainly by halos comparable in mass to our own Milky Way; the Lyα absorption signature of gas infall provides direct evidence for this [243]. More promising for early heating, perhaps, is the possibility of a population of mini-quasars, i.e., central black holes in early star-forming halos. This must be considered speculative, since the early halos were so small compared to galactic halos in the present universe that the corresponding black-hole masses are expected to fall in a very different range from observed quasars, specifically within the intermediate black-hole range ($10^2 - 10^4 M_\odot$) that local observations have probed only to a limited extent [244]. Thus, the properties of these mini-quasars are highly uncertain, and various assumptions can allow them to produce either early or late heating [245, 226]. Local observations can be used to try to estimate the possible importance of mini-quasars. An internal feedback model that is consistent with observations of local black-hole masses as well as high-redshift quasar luminosity functions [246] indicates a mini-quasar contribution that is somewhat lower than X-ray binaries [92], though the uncertainties are large. Regarding the spectrum, standard models of accretion disks [247] around black holes predict that the X-ray spectrum of mini-quasars [245] should peak at $1 - 5$ keV, making it a hard spectrum for cosmic heating that is quite similar to that of X-ray binaries.

Regardless of the source of X-rays, an important parameter is the degree of absorption in high-redshift halos compared to locally observed galaxies. If we assume that the gas density in high-redshift halos increases proportionally with the cosmic mean density, then the column density through gas (within a galaxy or a halo) is proportional to $(1 + z)^2 M_{\rm halo}^{1/3}$. This simple relation suggests that absorption of X-rays should increase at high redshift, since the redshift dependence should have a stronger effect than the decrease of the typical halo mass. However,

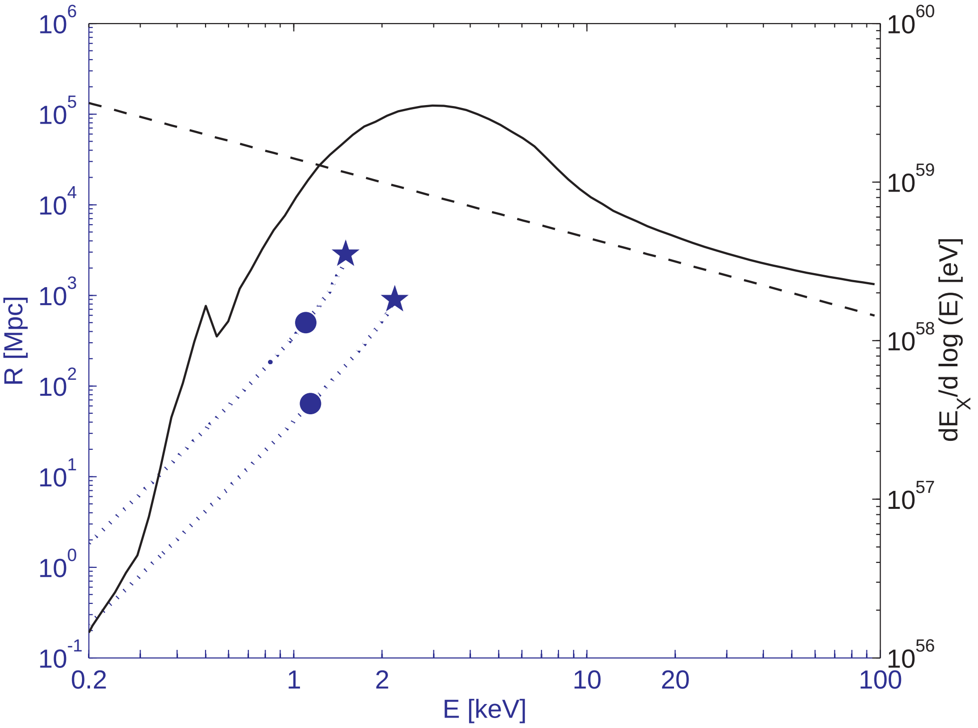

Figure 22: X-ray spectra, mean free paths, and horizons. The expected spectrum of X-ray binaries at high redshift (solid curve) from population synthesis models [235, 228] is compared with the soft power-law spectrum (dashed curve) adopted until recently [41, 29, 233, 91, 234]. Both indicate the distribution into X-ray photons with energy E of the total X-ray energy E_X produced per solar mass of newly-formed stars. The X-ray emission of X-ray binaries should be dominated by the most massive systems in their high (that is, bright) state [228], which is dominated by thermal disk emission, with little emission expected or seen [235, 236, 237] below 1 keV. Also shown are the mean free paths (dotted curves) of X-ray photons arriving at $z = 10$ (top) or $z = 30$ (bottom). For each of these redshifts, also indicated are the effective horizon for X-rays (defined as a $1/e$ drop-off, like a mean free path) from the combined effect of cosmological redshift and time retardation of sources (•), and the distance to $z = 65$ (⋆), the formation redshift of the first star [8, 9] (at which the mean free path curves are cut off). Note the separate y axes that indicate energy content for the spectra (right) or comoving distance for the other quantities (left). From [92].

complex astrophysics could substantially affect this conclusion, since the lower binding energy of the gas in low-mass halos could make it easier to clear out more of the blockading gas. Given the large uncertainty in internal absorption (on top of the other uncertainties in source properties), it is likely that only 21-cm observations will determine the precise characteristics of the high-redshift sources of cosmic heating.

7. 21-cm Signatures of the First Stars

Ongoing and planned interferometric observations in 21-cosmology hope to reach a sub-mK sensitivity level [35, 88] (see also § 1). The best current observational upper limit is from PAPER [33]: 22.4 mK at a wavenumber range of $k = 0.1 - 0.35$ Mpc^{-1} at $z = 8.4$, around an order of magnitude away from plausible predictions (or two orders of magnitude in terms of the power spectrum). Global 21-cm experiments (measuring the total sky spectrum) are also being pursued, with the best result thus far (from the EDGES experiment) [44] being a lower limit of $\Delta z > 0.06$ for the duration of the reionization epoch. In the next few subsections we focus on 21-cm fluctuations, and consider global experiments separately in § 7.5.

7.1. 21-cm signatures of reionization

In § 6.1 we discussed the important realizations that reionization was driven by groups of galaxies, the early galaxies were strongly clustered on large scales, and reionization had an inside-out topology. These features of reionization should all be observable with 21-cm cosmology. Figure 23 shows an example of 21-cm maps during reionization, as predicted by numerical simulations; a semi-numerical model gives a quite similar reionization field though it differs in the fine details. Another example is shown in Figure 24, which is from a simulation that computes the ionization, Lyα, and X-ray fields.

The typical evolution of the 21-cm power spectrum during cosmic reionization is illustrated in Figure 25, using an analytical model [36] that was shown to be in reasonable agreement with numerical simulations. Early on, when the cosmic ionized fraction is $\sim$ 10%, the 21-cm power spectrum simply traces the baryon density power spectrum (assuming here the limit of saturated Lyα coupling and saturated heating). As reionization advances, H II bubbles form around individual sources and begin to overlap between nearby sources, giving the 21-cm power spectrum an extra hump on large scales, with the corresponding k gradually decreasing as the typical size of the bubbles increases. At the final stages of reionization, the 21-cm intensity probes the distribution of remaining neutral gas in large-scale underdensities, and at the very end, atomic hydrogen remains only within galaxies. Figure 25 also illustrates how the 21-cm power spectrum can be used to probe the properties of the galaxies that are the sources of reionization. By artificially setting various values for the minimum circular velocity (or mass) of halos that dominate star formation, it is possible to simulate cases where small galaxies dominate or where large galaxies do (the latter case illustrating a situation where internal feedback is highly effective within small galaxies). Placing a fixed total amount of ionizing intensity within a smaller number of more massive halos has a number of effects on the 21-cm power spectrum; large halos are rarer and more strongly biased/clustered, leading to a higher power spectrum (in amplitude), a more prominent H II bubble bump that extends to somewhat larger scales, and a more rapid reionization process (in terms of the corresponding redshift range).

An important question is how to fit the 21-cm data that are expected soon from the cosmic reionization era. In general, the 21-cm power spectrum during reionization is a complex superpo-

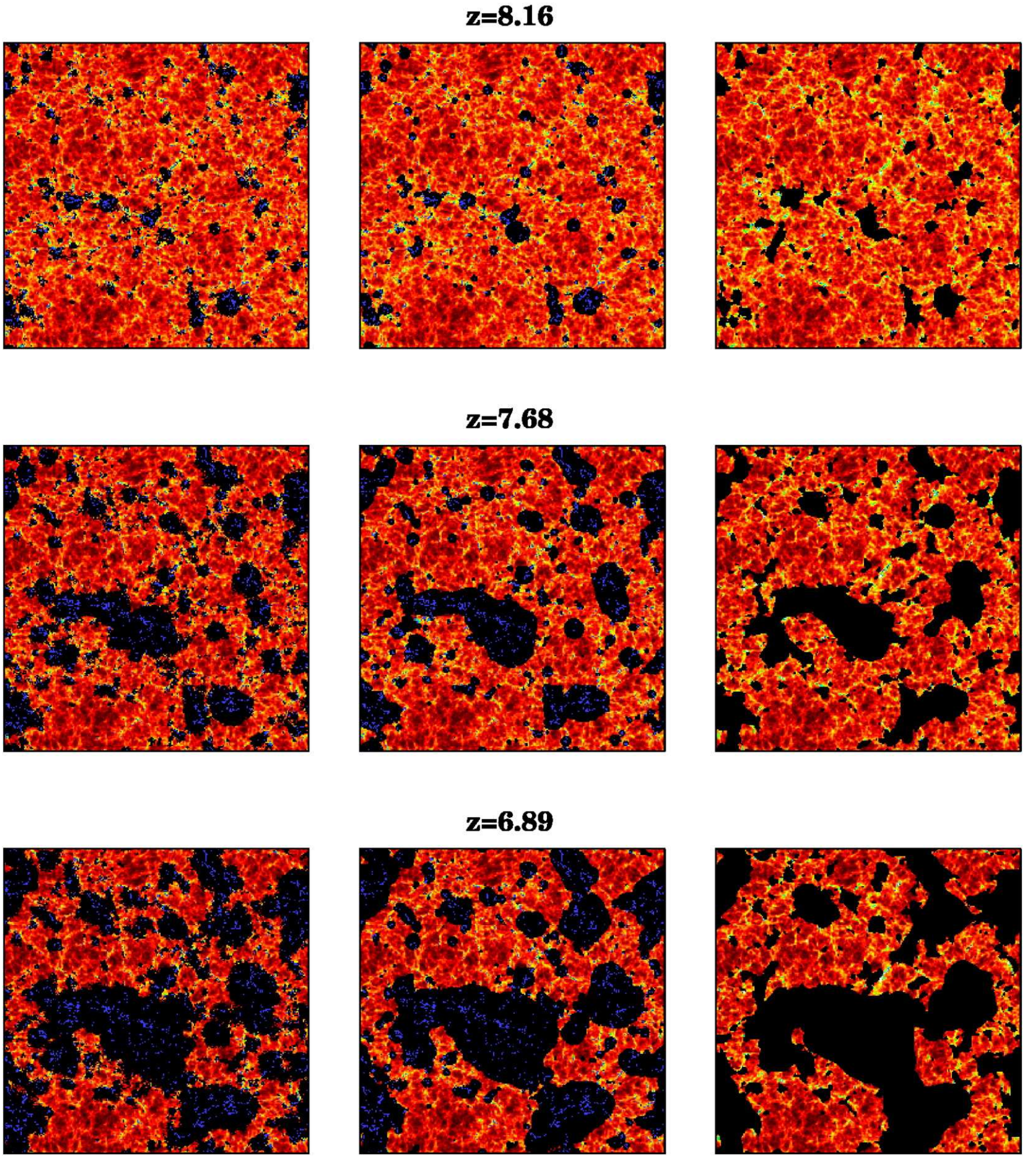

Figure 23: 21-cm maps during reionization, in radiative transfer simulations versus a semi-numeric approach. Each map is 94 Mpc on a side and 0.36 Mpc deep. The ionized fractions are 0.11, 0.33, and 0.52 for $z = 8.16$, 7.26, and 6.89, respectively. Left column: Radiative transfer calculation with ionizing sources (blue dots). Middle column: Halo smoothing procedure with sources from the N-body simulation. Right column: Matching semi-numerical model based on [25] and using the initial, linear dark matter overdensity. From [180].

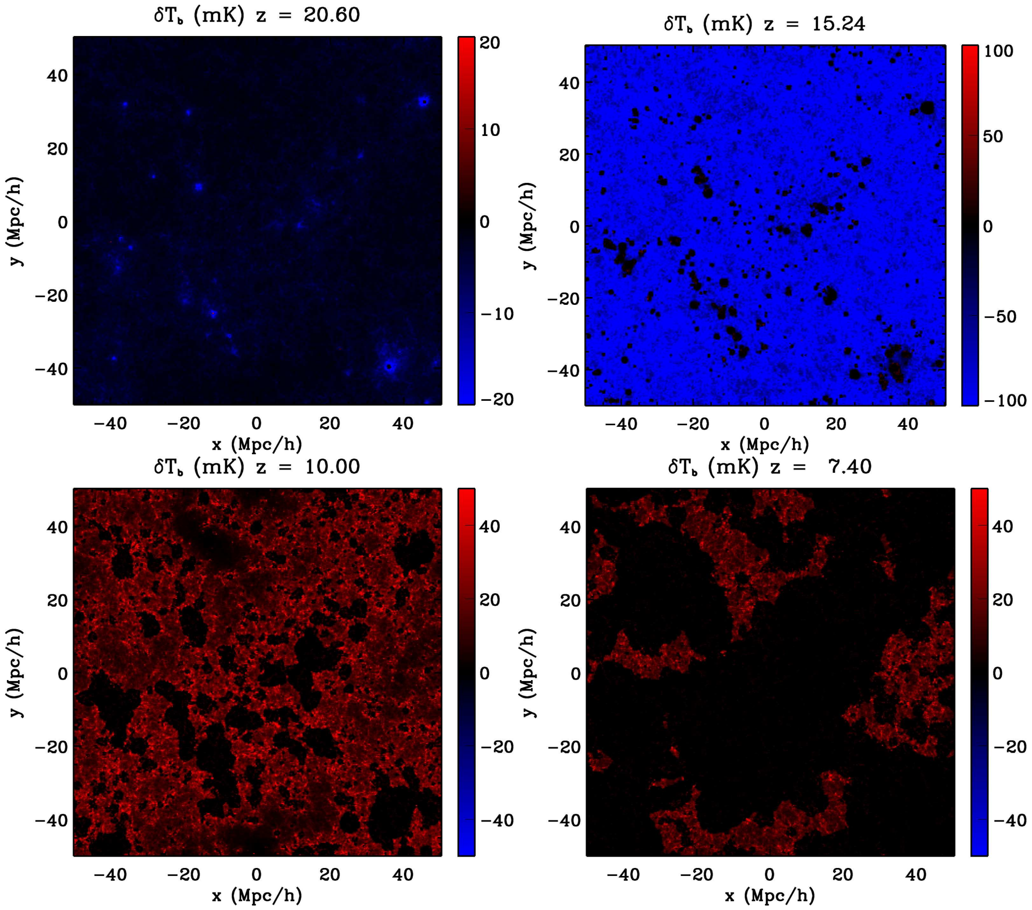

Figure 24: 21-cm maps from a 100/h Mpc simulation box that includes inhomogeneous Lyα and X-ray radiation fields, in addition to reionization. The ionized fractions are 0.0002, 0.03, 0.35, and 0.84 for z = 20.60, 15.24, 10.00, and 7.40, respectively. From [181].

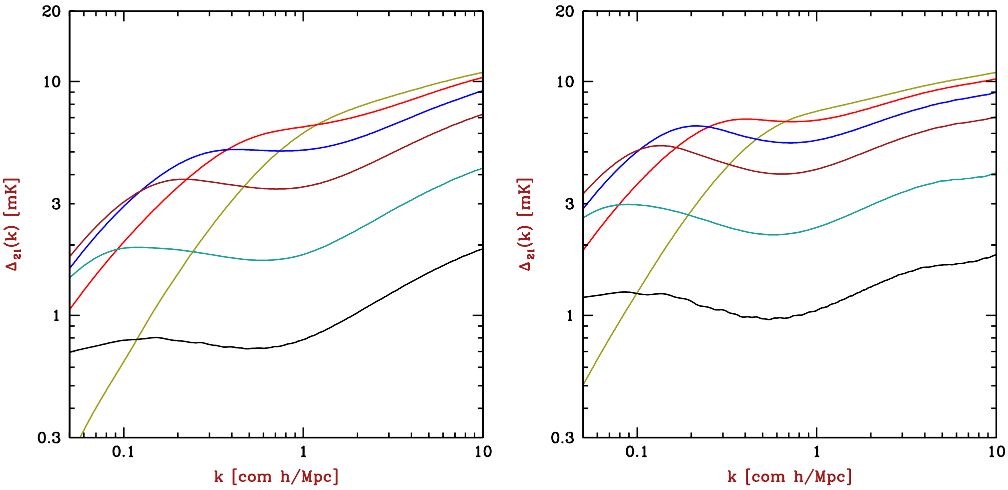

Figure 25: Evolution of the 21-cm power spectrum throughout reionization, for a model that sets the cosmic mean ionized fraction $\bar{x}_i = 98\%$ at $z = 6.5$. Shown are times when $\bar{x}_i = 10\%$, 30%, 50%, 70%, 90%, and 98% (from top to bottom at large k). At the very end of reionization, atomic hydrogen remains only within galaxies (this gas is not included in these plots). The panels show two different possibilities for the masses of galactic halos, assuming a minimum circular velocity for star formation of $V_c = 35$ km/s (left panel) or 100 km/s (right panel). From [36].

sition of the fluctuations in density and ionization (and possibly heating: see § 7.4); in order to interpret it quantitatively and reconstruct the history of reionization and of early galaxy formation, a flexible model is needed. Fitting to data cannot be done directly with numerical simulations, and is difficult even with a faster-running semi-numerical code. Thus, the first maximum likelihood fitting of mock data [36] was done with the analytical model noted above. The computational efficiency of this approach made it possible to employ a flexible six-parameter model that parameterized the uncertainties in the properties of high-redshift galaxies; specifically, the parameters were the coefficients of quadratic polynomial approximations to the redshift evolution of two parameters: the minimum circular velocity of galactic halos, and the overall efficiency of ionizing photon production within galaxies. The conclusion (see Figure 26) was that observations with a first-generation experiment should measure the cosmic ionized fraction to ~ 1% accuracy at the very end of reionization, and a few percent accuracy around the mid-point of reionization. The mean halo mass hosting the ionizing sources should be measurable to better than 10% accuracy when reionization is 2/3 of the way through, and to 20% accuracy throughout the central stage of reionization [36]. Recently the semi-numerical code 21CMFAST [91], in a sped-up version that employs some approximations, has been incorporated directly within 21CMMC, a Monte Carlo Markov Chain statistical analysis code. One result derived with this code (see Figure 27) is that combining three observations (at $z = 8$, 9 and 10) of the 21-cm power spectrum will allow upcoming 21-cm arrays to accurately constrain the basic parameters of reionization [250].

7.2. *21-cm signatures of Lyα coupling and LW feedback*

As previously discussed, the idea of unusually large fluctuations in the abundance of early galaxies (§ 3.1) first made a major impact on studies of cosmic reionization (§ 6.1). The same idea was also key in opening up cosmic dawn, prior to reionization, to interferometric 21-cm

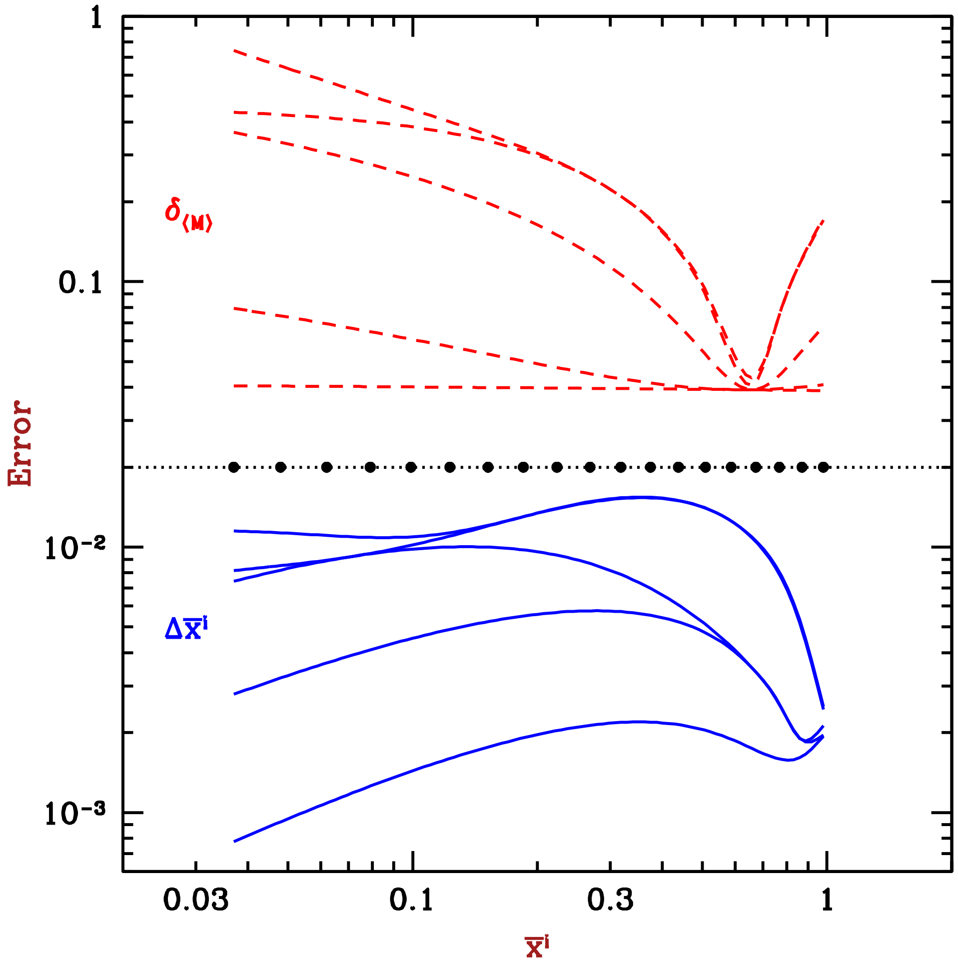

Figure 26: Expected reconstruction errors throughout cosmic reionization, from fitting models to mock data of the 21-cm power spectrum (with the expected errors of a first-generation experiment). The models are based on an analytical model [248, 249] that is in reasonable agreement [36] with numerical simulations of reionization. The x-axis shows the stage of reionization, i.e., the fraction of the IGM that has been reionized ($\overline{x_i}$). Models of varying degrees of flexibility are considered, with 2–6 free parameters (bottom to top in each set of curves). The input model of the mock universe sets the end of reionization (defined as 98% of the IGM being ionized) at $z = 6.5$, with galactic halos assumed to have a minimum circular velocity (Eq. 23) $V_c = 35$ km/s. A horizontal dashed line separates the two areas of the plot that show the expected relative error in the intensity-weighted mean mass of galactic halos (top) and the *absolute* error in the ionized fraction (bottom). Dots on the horizontal line show the values of $\bar{x}_i$ corresponding to the 19 assumed observed redshifts (in the range $z = 6.5 - 12$). From [36].

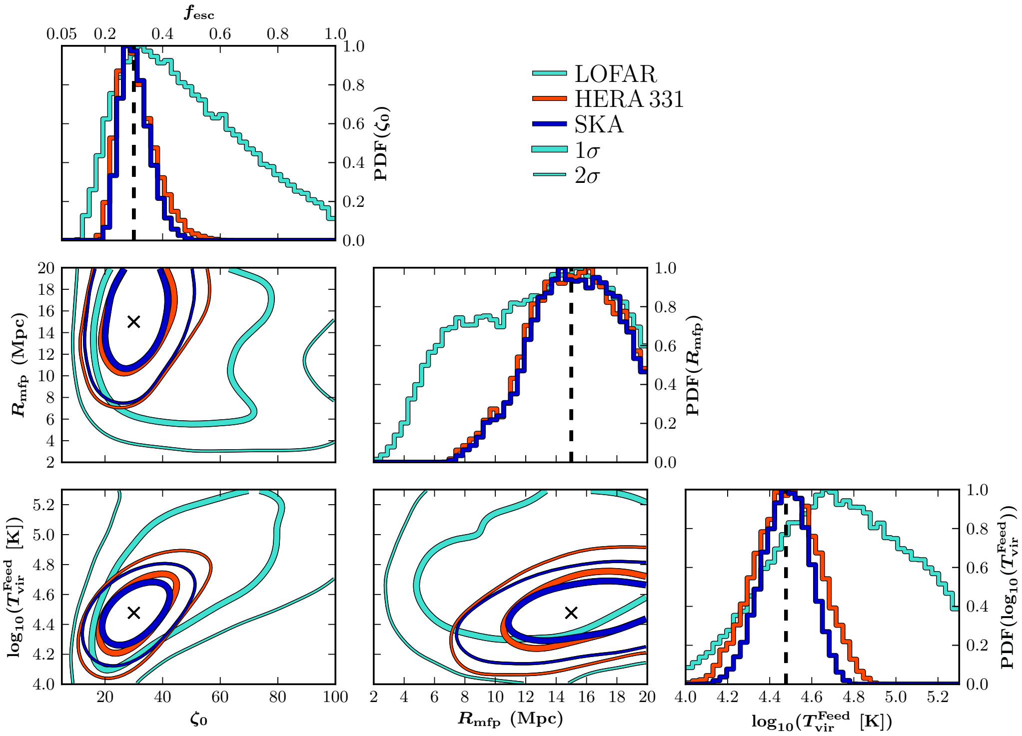

Figure 27: The recovered constraints from 21CMMC on various reionization model parameters from combining three independent (z = 8, 9 and 10) 1000 hr observations of the 21 cm power spectrum. Three different telescope arrays are compared: the current LOFAR (turquoise), and the future HERA (red) and SKA phase 1 (blue). Across the diagonal panels, the 1D marginalized PDFs are shown for the recovered reionization parameters [the ionizing efficiency ζ_0, showing the corresponding escape fraction of ionizing photons $f_{\rm esc}$ on the top; $R_{\rm mfp}$, the mean free path of ionizing photons within ionized regions; and $\log_{10}(T_{\rm vir}^{\rm Feed})$, where $T_{\rm vir}^{\rm Feed}$ is the minimum virial temperature of star-forming halos], with the input model parameter value indicated by a vertical dashed line. In the three panels below the diagonal, 2D joint marginalized likelihood contours are shown for various pairs out of the three reionization parameters. The 1σ (thick) and 2σ (thin) contours are shown, with crosses marking the input parameter values. From [250].

observations, by launching the study of fluctuations in the intensity of early cosmic radiation fields. The fact that fluctuations in the galaxy number density cause fluctuations even in the intensity of long-range radiation was first shown, specifically for the Lyα radiation background, by Barkana & Loeb (2005) [28]. The spin temperature of hydrogen atoms in the IGM is coupled to the gas temperature indirectly through the Wouthuysen-Field effect [18, 19], which involves the absorption of Lyα photons (§ 4). While it had been previously known [11, 120] that this Lyα coupling likely occurred in the IGM due to Lyα photons emitted by early stars at $z \sim 20-30$, this radiation background had been assumed to be uniform. This intuition was based on the fact that each atom sees Lyα radiation from sources as far away as $\sim$ 300 Mpc. However, it turns out that relatively large, potentially observable, 21-cm fluctuations are generated during the era of initial Lyα coupling, for two reasons: fluctuations in the number density of the (highly biased) early galaxies are significant even on scales of order 100 Mpc, and also a significant fraction of the Lyα flux received by each atom comes from sources at smaller distances. Since relatively few galaxies contribute most of the flux seen at any given point, Poisson fluctuations can be significant as well, producing correlated 21-cm fluctuations (since a single galaxy contributes Lyα flux to many surrounding points in the IGM). If observed, the Lyα fluctuation signal would not only constitute the first detection of these early galaxies, but the shape and amplitude of the resulting 21-cm power spectrum would also probe their average properties [28] (Figure 28).

This discovery of Lyα fluctuations has led to a variety of follow-up work, including more precise analyses of the atomic cascades of Lyman series photons [123, 251]. Also, a significant boost is predicted in the 21-cm power spectrum from Lyα fluctuations due to the repeated scattering of the photons from stars on their way to the hydrogen atoms, out in the wing of the Lyα line [252, 253, 254] (Figure 29). The repeated scatterings mean that the Lyα photons do not reach as far (in the fixed time until they redshift into – and then out of – the line), which decreases the overall large-scale smoothing and thus increases the predicted level of 21-cm fluctuations. Moreover, the increased sensitivity to Lyα photons from short distances makes the overall 21-cm power spectrum sensitive to the sizes of H II regions at this very early stage in reionization (Figure 29). Note that in addition to direct stellar emission, Lyα photons are also produced in the IGM from X-ray ionization; however, despite early overestimates [29], the contribution of these Lyα photons in typical models is $\sim 1\%$ compared to stellar Lyα photons [214].

As discussed in § 6.2, LW feedback is an important feedback effect on early galaxies, as it dissociates molecular hydrogen and eventually ends star formation driven by molecular cooling [20]. Thus, it affects 21-cm fluctuations indirectly by changing the amount and distribution of star formation [213]. The effect becomes particularly striking once the baryon - dark matter streaming velocity (§ 5) is included. Assuming that star formation is dominated by $10^6 M_\odot$ halos at very high redshift, the streaming velocity strongly affects them and produces a distinctive BAO signature in the 21-cm fluctuations (§ 7.3). Since LW feedback affects star formation in precisely the same halos that are affected by the streaming velocity, the effectiveness of the feedback has a major effect on 21-cm observations [155] (Figure 30). This is particularly important since there is a substantial uncertainty in the strength of LW feedback on early star formation (although this subject has been explored somewhat with numerical simulations: § 6.2); thus, the prospect that 21-cm observations over a range of redshifts will detect the time evolution of the LW feedback is quite interesting.

7.3. Large 21-cm fluctuations from early cosmic heating

As discussed in detail in § 6.3, until recently it was expected that the universe had been preheated well before cosmic reionization. This early heating was thought to be likely due to the

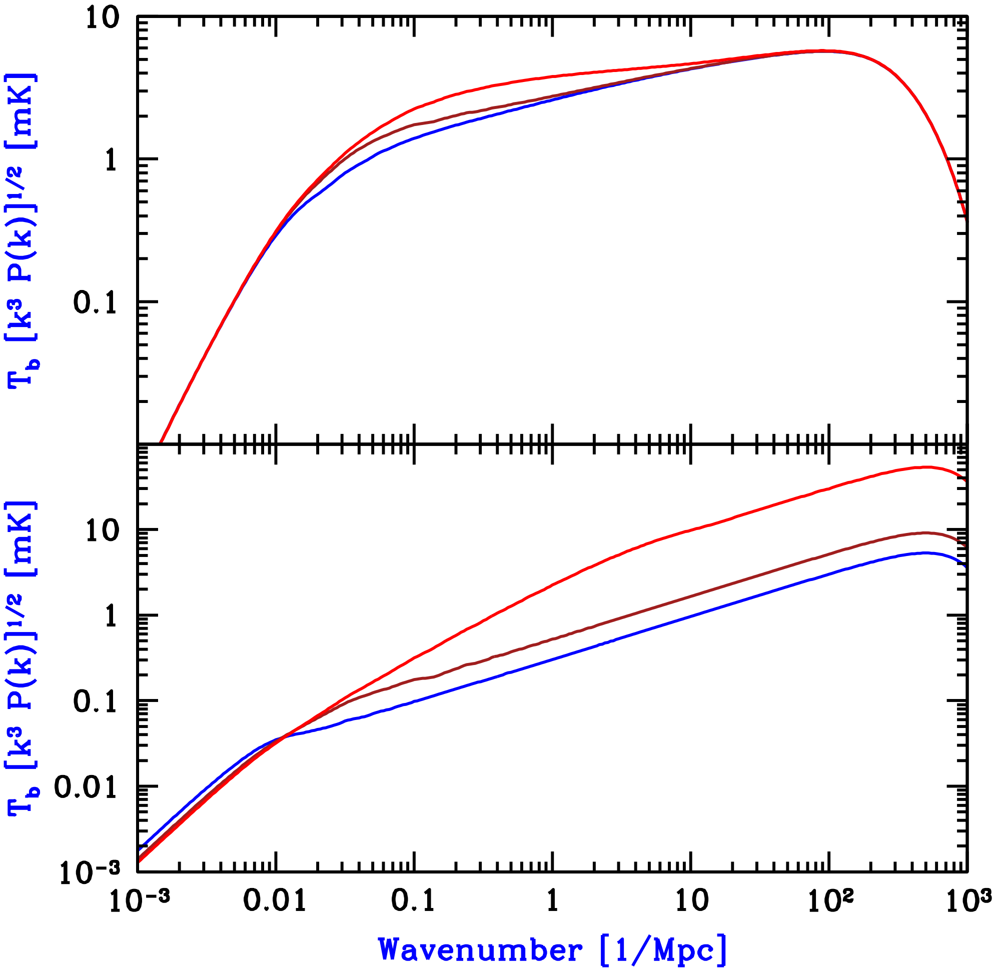

Figure 28: Power spectra of 21-cm brightness fluctuations (in mK units) from Lyα fluctuations, plotted versus (comoving) wavenumber. Shown are two components of the power spectrum that in principle can be separated (in the limit of linear perturbations) based on the line-of-sight anisotropy of the 21-cm fluctuations [13] (§ 4.3): P_{μ^2} (top panel) contains contributions directly from density fluctuations and from the density-induced fluctuations in galaxy density and therefore in Lyα flux, and $P_{\mathrm{un}-\delta}$ (bottom panel) is due to Lyα fluctuations from Poisson fluctuations in galaxy numbers. These results are for galaxies formed via atomic cooling in halos at $z = 20$, with a star formation efficiency set to produce the Lyα coupling transition at this redshift. They also assume linear fluctuations, and that the IGM gas cooled adiabatically down to this redshift. Each set of solid curves includes, from bottom to top at $k = 0.1$ Mpc^{-1}, stellar radiation emitted up to Lyβ, Lyδ, or full Lyman-band emission, all assuming Pop III stars. Note that the results shown here from the first such prediction [28] were later updated (Figure 29).

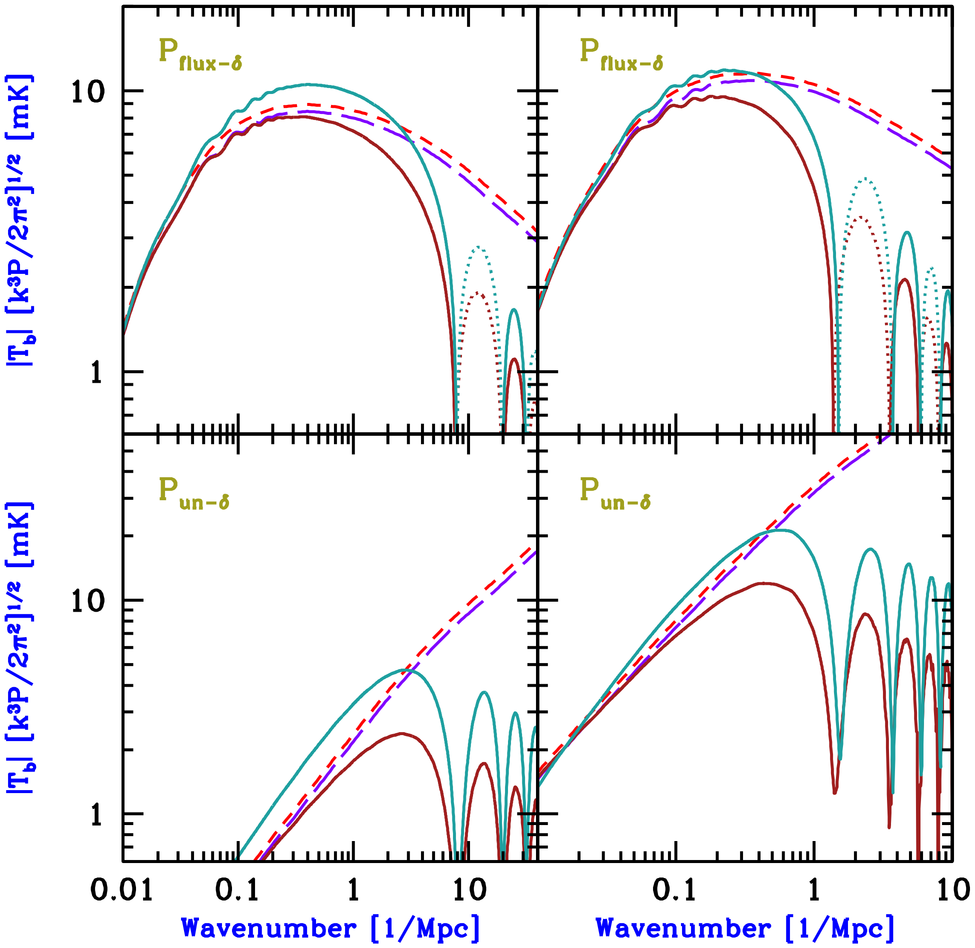

Figure 29: 21-cm power spectrum P (in terms of the brightness temperature fluctuation at wavenumber k) as a function of k. Shown are two components of the power spectrum that in principle can be separated (in the limit of linear perturbations) based on the line-of-sight anisotropy of the 21-cm fluctuations [13] (§ 4.3): $P_{\rm flux-\delta}$ (top panels) contains the contribution of density-induced Lyα fluctuations and $P_{\rm un-\delta}$ (bottom panels) is due to Lyα fluctuations from Poisson fluctuations in galaxy numbers. Compared here are the earlier result from [28] (including the correction from [123, 251]) (short-dashed curves), the result corrected to use the precise density and temperature power spectra from [70] (long-dashed curves), and from [254] the same calculation with a cutoff due to individual H II regions around galaxies (solid curves, the lower of each pair), and the full calculation (higher solid curve of each pair) which also includes the redistribution of photons due to scattering in the wing of the Lyα line. Two possible examples are shown for galactic halos, where their minimum circular velocity is assumed to be V_c = 16.5 km/s (left panels, corresponding to atomic cooling) or V_c = 35.5 km/s (right panels, an example of a case where internal feedback makes lower-mass halos inefficient at star formation). Negative portions are shown dotted in absolute value. Note that these results assume the simple case of a fixed H II region size around all galaxies; more realistically, the small-scale ringing seen in this Figure may be smoothed out by a scatter in H II region sizes, but the overall shape and the peak of each curve are more robust predictions. From [254].

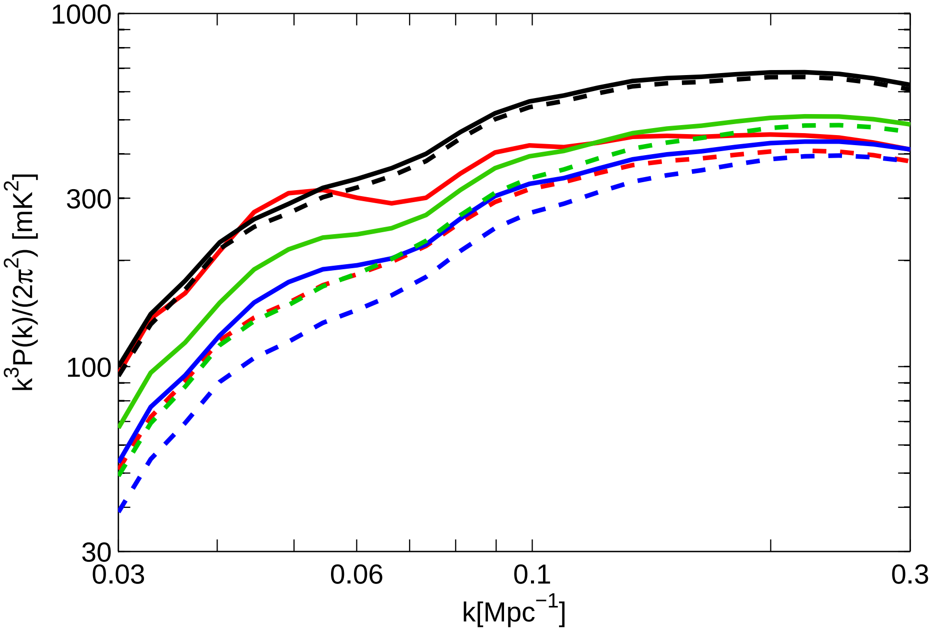

Figure 30: Power spectra of the 21-cm brightness temperature for various strengths of LW feedback: no feedback (red), relatively weak feedback (blue), relatively strong feedback (green) and saturated feedback (i.e., no H_2 molecules; black); each case is shown either with (solid) or without (dashed) the streaming velocity. The weak and strong feedback cases roughly bracket current uncertainties, although recent simulations [219] give some support to the strong case. The results in each case are shown at a time of particularly strong heating fluctuations, a $\Delta z = 3$ earlier (i.e., higher z) than the moment when the cosmic mean 21-cm temperature is zero. The streaming velocity increases and flattens the large-scale power spectrum, and strengthens the BAOs (e.g., at the large-scale peak at $k = 0.05$ Mpc^{-1}); this effect (which is wiped out in the limiting case of saturated feedback) is partially suppressed by the LW feedback. This Figure from [155] assumed the case of early cosmic heating by a soft X-ray spectrum (§ 7.3); in the more likely case of late heating by a hard X-ray spectrum (§ 7.4), the combined effect of LW feedback and the streaming velocity would be more difficult to observe with heating fluctuations, but would still be observable during the somewhat earlier era of Lyα fluctuations.

high heating efficiency of the soft X-ray spectrum that had been assumed in calculations of cosmic heating. Soft X-rays are absorbed in the neutral IGM over relatively short distances, making heating a local phenomenon that can potentially give rise to large temperature fluctuations in the early IGM. Indeed, when combined with the idea of unusually large fluctuations in the abundance of early galaxies (§ 3.1), the expectation of large-scale fluctuations in ionization (§ 6.1) and in the Lyα radiation background (§ 7.2) can be extended to the X-ray background. The first calculation of heating due to an inhomogeneous X-ray background, by Pritchard & Furlanetto (2007) [29], applied to X-rays a similar method as in the Lyα case [28]; integrating the heating over time to find the distribution of gas temperatures, the result was the prediction of another era of detectably large 21-cm fluctuations (Figure 31).

As discussed in § 3.2, while numerical simulations are the best, most accurate method for studying early galaxy formation on small scales, they are unable to simultaneously cover large volumes. Simulations that successfully resolve the tiny mini-galaxies that dominated star formation at early times are limited to $\sim$ 1 Mpc volumes, and cannot explore the large cosmological scales that might be accessible to 21-cm observations (which are currently limited to low resolution). On the other hand, analytical calculations are limited to linear (plus sometimes weakly non-linear) scales, and thus cannot directly probe the non-linear astrophysics of halo and star formation. Even if the results of simulations are incorporated within them, analytical approaches assume small fluctuations and linear bias (see the end of section 2.3), assumptions that break down in the current context, where the stellar density varies by orders of magnitude on scales of a few Mpc. Even on 100 Mpc scales, fluctuations in the gas temperature are as large as order unity (see below). Thus, linear, analytical calculations can only yield rough estimates, even for large-scale fluctuations.

As a result of these considerations, perhaps the best current method to generate observable 21-cm predictions from the era of early galaxies is with a hybrid, semi-numerical code that combines linear theory and full calculations on large scales with analytical models and the results of numerical simulations on small scales. Such methods have been compared with numerical simulations of reionization [180, 181], and have also been used to predict the effect of the streaming velocity on high-redshift galaxy formation [45, 143]. Figure 32 shows a prediction of the 21-cm signatures of X-ray heating made with the semi-numerical code 21CMFAST [234]. The light-cone slices show the progression through cosmic 21-cm history: collisional decoupling during the dark ages (black, far-right region), Lyα coupling (black to yellow transition), X-ray heating (yellow to blue), and reionization (blue to black).

In the case of soft X-ray heating sources, heating fluctuations are the largest, most promising source of pre-reionization 21-cm fluctuations, but even in this case there remains a large uncertainty in predicting the signal. The redshift at which this signal peaks depends on the overall efficiency of X-ray production, with higher efficiency leading to an earlier cosmic heating era. This uncertainty is not too problematic since planned observations will cover a wide redshift range and find the signal if it is there. Given the correct redshift, the strength of the signal still depends on the typical mass of the galactic halos that hosted these sources. The more massive the halos, the more highly biased (clustered) they are expected to have been, thus producing a larger 21-cm fluctuation signal. However, the baryon - dark matter streaming velocity (§ 5) greatly cuts down this uncertainty, as it boosts the expected signal from low-mass halos nearly to the same level as that from high-mass halos. Observational predictions that include the streaming velocity were achieved with a semi-numerical method [255].

This approach built upon previous semi-numerical methods used for high-redshift galaxy formation [45, 143, 91]. It used the known statistical properties of the initial density and ve-

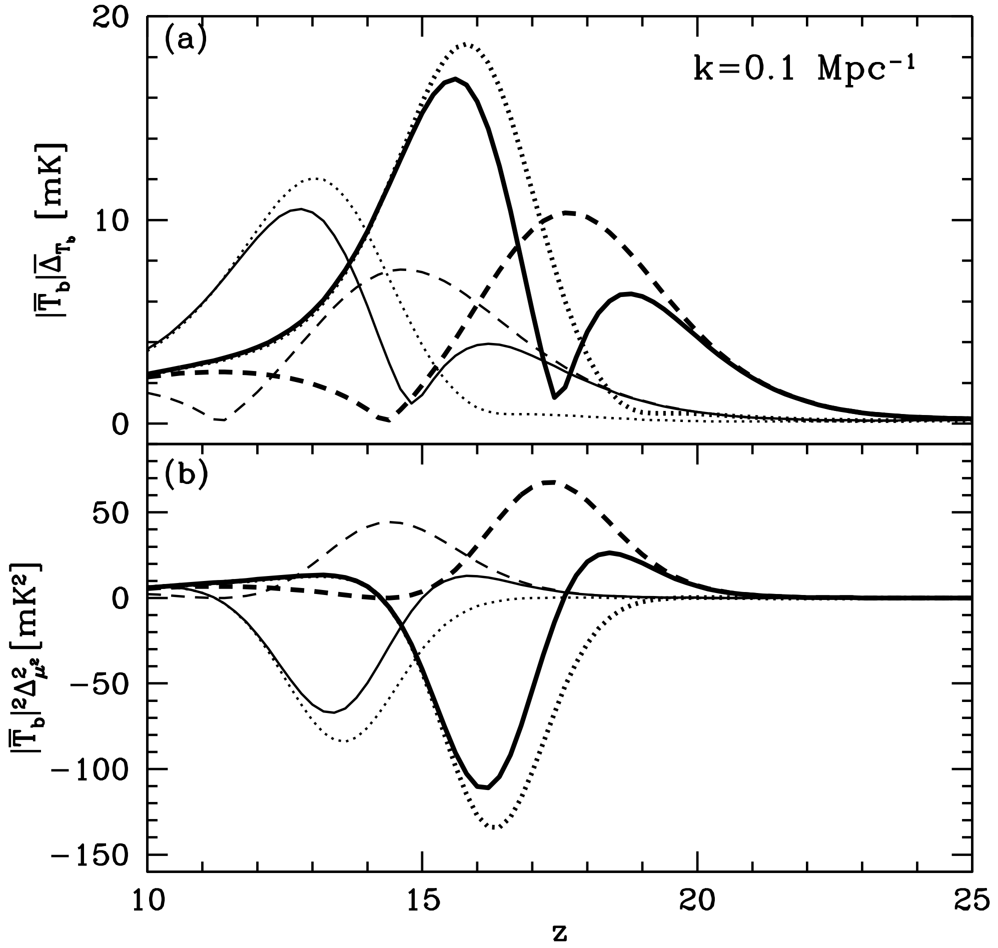

Figure 31: Power spectra of 21-cm brightness fluctuations from temperature fluctuations during cosmic X-ray heating [29]. Shown are the total isotropically-averaged fluctuation (top panel) and the main anisotropic term P_{μ^2} (bottom panel) from the line-of-sight anisotropy of the 21-cm fluctuations [13] (§ 4.3). These quantities are shown at a wavenumber $k = 0.1$ Mpc^{-1}, including the effects of heating fluctuations only (dotted curves), Lyα fluctuations only (dashed curves), or both (solid curves). Two models are shown, one corresponding to Pop II stars (thick curves) and the other to Pop III stars (thin curves). Note that this Figure from [29] assumed linear fluctuations, early heating by a soft spectrum of X-ray sources, and did not include the boost in the Lyα fluctuations by a factor of $\sim$ 1.5 (Figure 29) from multiple scattering.

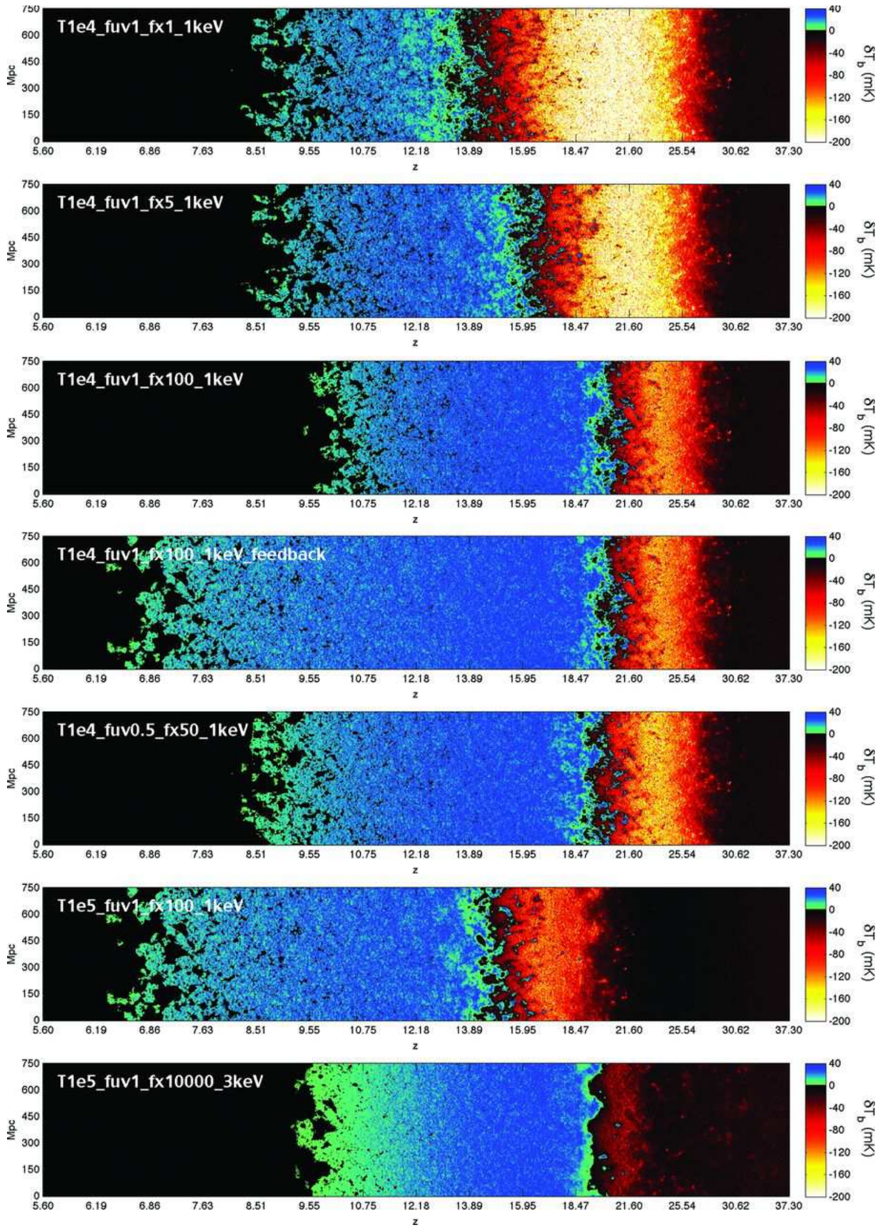

Figure 32: Slices through the predicted 21 cm signal for models calculated with the semi-numerical code 21CMFAST. The slices show the evolution of the 21-cm brightness temperature with distance along the light cone, with the redshift indicated on the *x*-axis, and the *y*-axis showing spatial structure at each redshift. In the model name, T indicates the minimum assumed virial temperature of galactic halos (10^4 K corresponds to atomic cooling), fuv and fx parameterize the ionizing and X-ray efficiency, respectively, and the final number indicates the mean X-ray photon energy ("1 keV" indicates a soft power-law starting at 0.3 keV, with a mean photon energy of 0.9 keV; these quantities are 3 times larger for the "3 keV" case). These models assume heating via soft X-rays, except for the extreme (bottom-most) model in which very intense X-rays dominate reionization (not just heating). From [234].

locity perturbations to generate a realistic sample universe on large, linear scales. This was followed by a calculation of the stellar content of each pixel on the grid using a model [9] previously developed to describe the streaming velocity effect on galaxy formation; this includes analytical models as well as fits to the results of small-scale numerical simulations. Like other semi-numerical codes, it assumed standard initial perturbations (e.g., from a period of inflation), where the density and velocity components are Gaussian random fields.

Velocities are coherent on larger scales than the density, due to the extra factor of $1/k$ in the velocity from the continuity equation that relates the two fields. This is clearly apparent in the example shown in Figure 33 of a thin slice of a simulated volume. The density field fluctuates on relatively small scales, while the velocity field shows a larger-scale cosmic web, with coherent structure on scales of order 100 Mpc. This means that the largest scales will be dominated by the pattern due to the velocity effect, as long as the streaming velocity significantly affects star formation.

The resulting distribution of stellar density at $z = 20$ is also shown in Figure 33. Note the large biasing (i.e., amplification of fluctuations) of the stars: density fluctuations ranging up to $\pm 50\%$ yield (without including the streaming velocity) a field of stellar density that varies by over a factor of 20 (when both fields are smoothed on a 3 Mpc scale). The velocity effect produces a more prominent cosmic web on large scales, marked by large coherent regions that have a low density of stars, separated by ribbons or filaments of high star formation. The effect is much more striking at higher redshifts (Figure 34), and it thus substantially alters the feedback environment of the very first generations of stars. The various types of radiation that produce feedback spread out to a considerable distance from each source, but this distance is typically not as large as the span of the velocity-induced features. This means that regions of low velocity (and thus high star formation) experience radiative feedback substantially earlier than regions of high velocity (low star formation). Thus, the substantial effect of the velocities on early star formation makes early feedback much more inhomogeneous than previously thought.

Observationally, these degree-scale fluctuations affect various cosmic radiation backgrounds, and in particular the history of 21-cm emission and absorption. As noted above, in the presence of soft X-ray heating sources, the heating fluctuations produce the largest pre-reionization 21-cm fluctuations, typically from sometime after the Lyα coupling has mostly saturated. As for the LW flux, here we consider the case of negligible LW feedback (as was assumed in Figures 33 and 34), but below we bracket the effect of the LW flux by also considering the opposite limiting case where the LW transition has already saturated (i.e., completely destroyed hydrogen molecules); the effect of various strengths of LW feedback was discussed in more detail in § 7.2.

Figure 35 shows the gas temperature distribution at $z = 20$, assumed to be at the heating transition, i.e., when the mean H I gas temperature was equal to that of the CMB. Regions where the gas moved rapidly with respect to the dark matter (dark red regions, top right panel of Figure 33) produced fewer stars (dark blue regions, bottom right panel of Figure 33) and thus a lower X-ray intensity, leaving large regions with gas that is still colder than the CMB by a factor of several (dark blue regions, top right panel of Figure 35). The spatial reach of X-rays results in a gas temperature distribution that is smoother than the distribution of stars, and this brings out the effect of large-scale fluctuations and thus highlights the contrast between the effect of density and velocity fluctuations.

During the heating transition (§ 6.3), the 21-cm brightness temperature (shown in the bottom panels of Figure 35) mainly measures the gas (kinetic) temperature T_K, although it is also proportional to the gas density (and to the square root of $1 + z$). The form of the dependence, $T_{\rm b} \propto 1 - T_{\rm CMB}/T_K$, makes the 21-cm intensity more sensitive to cold gas than to hot gas (relative

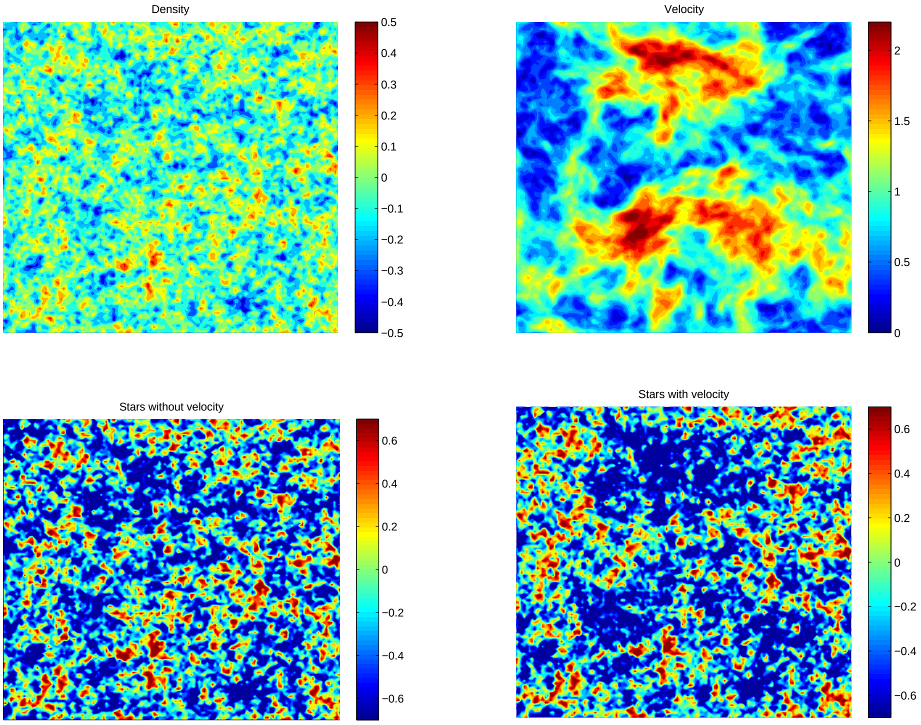

Figure 33: Effect of the streaming velocity on the number density of stars at redshift 20. The large-scale density (top left panel) and velocity (top right panel) fields are shown on the top. For the density field, the fractional perturbation relative to the mean is shown, at $z = 20$; for the velocity field, the magnitude of the relative motion in units of the root-mean-square value is shown (the map is independent of redshift in these relative units). For the same slice, the traditional calculation (lower left panel), which includes the effect of density only, is compared to the new prediction (lower right panel), which includes the effect of the same density field plus that of the streaming velocity. The colors in the bottom panels correspond to the logarithm (base 10) of the gas fraction in units of its cosmic mean value in each case. The panels all show a 3 Mpc thick slice (the pixel size of the grid in the semi-numerical code) from a simulated volume 384 Mpc on a side (based on [255], but taken from a different box from the one shown in the Figures in [255], i.e., for a different set of random initial conditions).

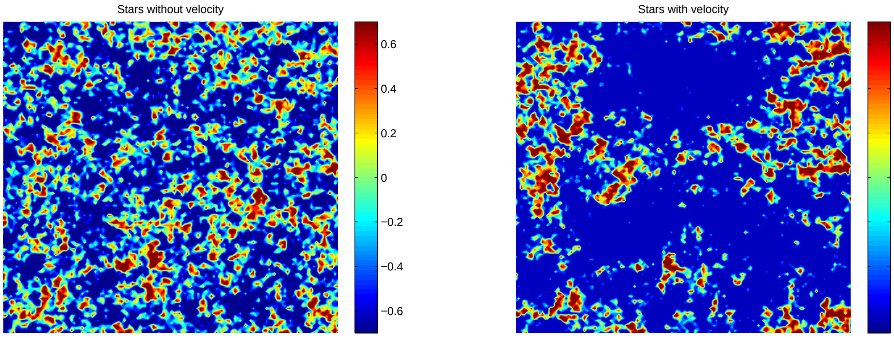

Figure 34: Effect of the streaming velocity on the number density of stars at redshift 40. For the same slice as in Figure 33, we compare the traditional calculation (left panel), which includes the effect of density only, to the new prediction (right panel), which includes the effect of the same density field plus that of the streaming velocity. The colors correspond to the logarithm of the gas fraction in units of its cosmic mean value in each case. The color scale spans the same range as in Figure 33 for easy comparison.

to the CMB temperature). Thus, the large voids in star formation produced by a high streaming velocity lead to prominent 21-cm absorption (dark blue regions, bottom right panel of Figure 35) seen on top of the pattern from the effect of density fluctuations. These deep 21-cm cold spots are a major observable signature of the effect of the streaming velocity on early galaxies.

While Figure 35 illustrates the detailed pattern that the streaming velocity imprints on the 21-cm intensity distribution, upcoming experiments are expected to yield noisy maps that likely must be analyzed statistically. Figure 36 shows the predicted effect on the power spectrum of the fluctuations in 21-cm intensity [255]. The velocities enhance large-scale fluctuations (blue solid curve compared with red dotted), leading to a flatter power spectrum with prominent baryon acoustic oscillations (reflecting the BAO signature in Figure 14). The signal is potentially observable with a redshift 20 version of current instruments (green dashed curve). If there is complete LW feedback (solid purple curve), then the small galaxies that rely on molecular-hydrogen cooling are unable to form; the larger galaxies that dominate in that case are almost unaffected by the streaming velocity, so the 21-cm power spectrum reverts to the density-dominated shape (compare the solid purple and red dotted curves), but it becomes even higher since more massive galactic halos are even more strongly biased.

Thus, regardless of the strength of the LW feedback (or other negative feedback effects on small galaxies), the 21-cm power spectrum at the peak of the heating transition should feature large fluctuations on observable scales. Beyond just detection of the signal, only a mild additional accuracy is necessary in order to determine whether feedback has suppressed star formation in the smallest halos. If it has not, then the velocity effect produces strong BAOs on top of a flattened power spectrum, in particular raising it by a factor of 4 on large scales ($k = 0.05$ Mpc^{-1}, wavelength 130 Mpc, observed angle 2/3 of a degree) where the experimental sensitivity is optimal. If this characteristic shape is observed it would confirm that million mass halos dominated galaxy formation at this early epoch.

While Figure 36 considers a single redshift, similar observations over the full $\Delta z \sim 6$ redshift range of significant heating fluctuations could actually detect the slow advance of the LW feedback process, during which the power spectrum is predicted to continuously change shape,

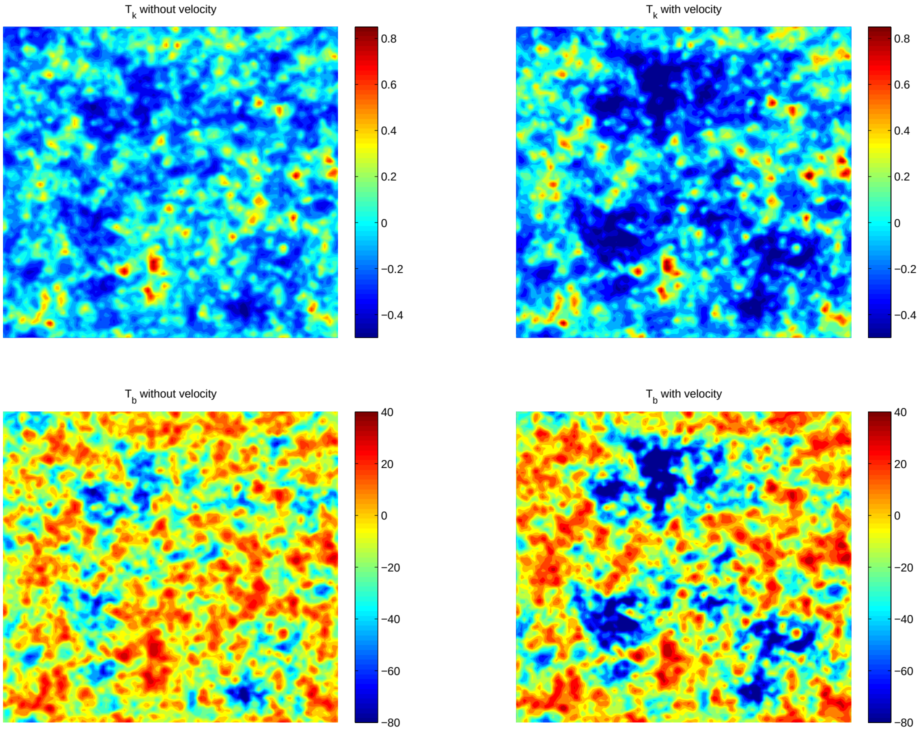

Figure 35: Effect of the streaming velocity on the gas temperature T_k and on the 21-cm brightness temperature T_b at redshift 20. For the same slice as in Figure 33, we compare the traditional calculations (left-hand panels), which include the effect of density only, to the new predictions (right-hand panels), which include the effect of density and streaming velocity. The colors in the top panels correspond to the logarithm of the gas (kinetic) temperature in units of the CMB temperature at $z = 20$. The colors in the bottom panels correspond to the 21-cm brightness temperature in millikelvin units. Note that the observed wavelength of this 21-cm radiation is redshifted by the expansion of the universe to 4.4 meters (corresponding to a frequency of 68 MHz).

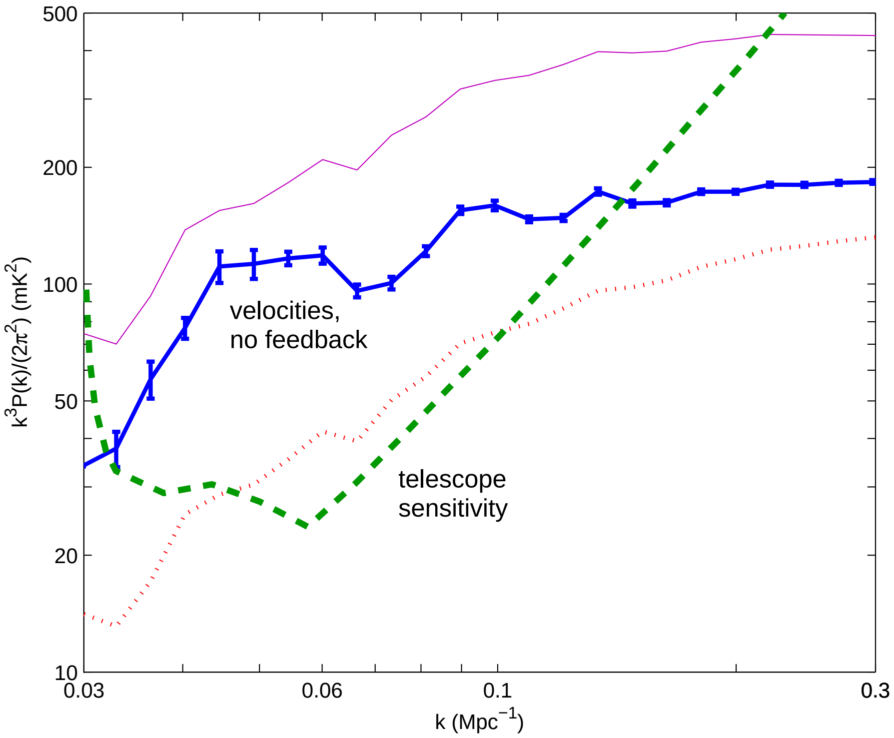

Figure 36: Signature of the streaming velocity in the 21-cm power spectrum, at the peak of the X-ray heating transition. The prediction is shown including the streaming velocity effect (blue solid curve) or with the effect of density only (red dotted curve), both for the case of a late LW transition for which the LW feedback is still negligible at the heating transition. These predictions are compared to the projected 1-σ telescope sensitivity (green dashed curve) based on a 1000-hour observation with an instrument like the Murchison Wide-field Array or the Low Frequency Array but designed to operate at 50–100 MHz [35], including an estimated degradation factor due to foreground removal [256]; this sensitivity is defined as the signal that would yield a measurement with a signal-to-noise ratio of unity in each k bin of size $\Delta k = 0.5k$ averaged over an 8 MHz frequency band (where only thermal noise is included). Future experiments like the Square Kilometer Array should reach a better sensitivity by more than an order of magnitude [35]. To allow for the possibility of feedback, the prediction is also shown in the opposite limit of maximum feedback, i.e., an early LW transition that has already saturated (purple solid curve). In this plot, the heating transition has been fixed at $z = 20$ for easy comparison among the various cases. Error bars on the main prediction curve indicate the $1-\sigma$ sample variance in one simulation box. From [255].

gradually steepening as the BAO signature weakens towards low redshift (see Figure 30 in § 7.2). This is all the case if the Universe was heated by soft X-rays. If it was heated by hard X-rays (see the next subsection), then the heating peak is largely erased, but similar effects of the streaming velocity are expected on the 21-cm signal during the $z \sim 25$ fluctuation peak from the Lyman-α coupling transition (§ 7.2).

7.4. Late heating and reionization

As discussed in § 6.3, it was recently realized that the hard X-ray spectrum characteristic of X-ray binaries, the most plausible source of early cosmic heating, is predicted to have produced a relatively late heating, possibly encroaching on the reionization era. The effect of this on the global 21-cm signal is discussed in § 7.5. Here we discuss the key consequences for 21-cm fluctuations.

A major effect of X-ray heating by a hard spectrum is the suppression of 21-cm fluctuations due to heating. Under the previously assumed soft spectra, the short typical distance traveled by the X-ray photons was found to produce large fluctuations in the gas temperature and thus in the 21-cm intensity around the time of the heating transition, regardless of when this transition occurred [29, 233, 255] (§ 7.3). However, the larger source distances associated with a hard spectrum lead to a much more uniform heating, with correspondingly low temperature fluctuations even around the time of the heating transition, when the 21-cm intensity is quite sensitive to the gas temperature. This trend is strengthened by late heating, as it occurs at a time when the heating sources are no longer as rare and strongly biased as they would be in the case of an earlier heating era. Thus, heating with a hard X-ray spectrum is predicted to produce a new signature in the 21-cm fluctuation signal: a deep *minimum* during reionization [92]. This results from the low level of gas temperature fluctuations in combination with a suppression of the 21-cm impact of other types of fluctuations (i.e., in density and ionization); in particular, right at the heating transition, the cosmic mean 21-cm intensity is (very nearly) zero, and thus all fluctuations other than those in the gas temperature disappear (to linear order) from the 21-cm sky.

This effect is visually apparent in simulated maps (Figure 37). In upcoming observations, it is likely to be apparent in the measured 21-cm power spectrum (Figure 38). Depending on the parameters, the deep minimum (reaching below 1 mK) may occur at any time during reionization, but is likely to occur before its mid-point. Previously, the fluctuation signal was expected to lie within a narrow, well-defined range, allowing for a relatively straight-forward interpretation of the data in terms of the progress of reionization; the possibility of a hard X-ray spectrum, however, introduces a variety of possibilities, making it likely that modeling of the 21-cm data will involve an analysis of the interplay of heating and reionization.

If a sufficient sensitivity level can be achieved, a low minimum in the 21-cm power spectrum during reionization would be a clear signature of late heating due to a hard X-ray spectrum. Indeed, a clear observational indication that this feature corresponds to a cosmic milestone is that the minima at all $k > 0.5$ Mpc^{-1} should occur at essentially the same redshift (namely the true redshift of the heating transition); lower wavenumbers correspond to larger scales than the typical X-ray mean free path, leading to a more complicated evolution and to minima delayed to lower redshifts (see also Figure 41). More generally, observations of the 21-cm power spectrum over a broad range of wavenumbers will clearly probe the X-ray spectrum of the sources of cosmic heating [92, 257, 258, 130].

Beyond reionization, heating by high-energy X-rays removes the previously expected signal from an early heating transition (§ 7.3) at $z \sim 15 - 20$, but leaves in place the similar $z \sim 20 - 25$ signal from the Lyman-α coupling transition that is likely detectable with the Square

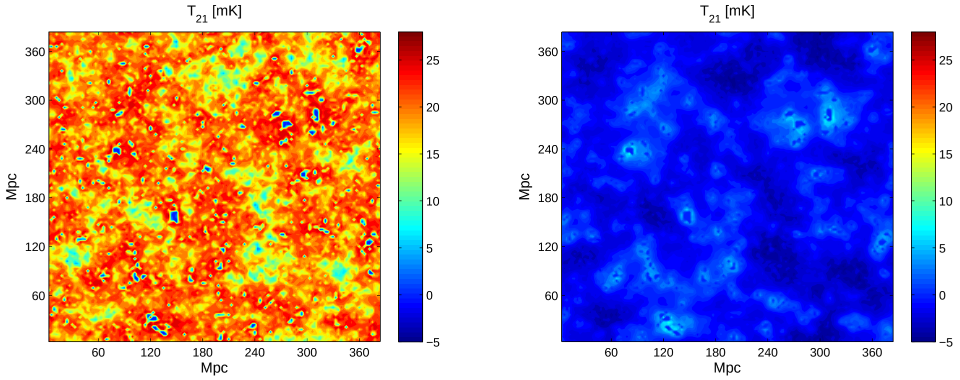

Figure 37: Example of a predicted 21-cm map from a semi-numerical code, at $z = 12.1$, comparing the case of heating sources with a hard X-ray spectrum (right panel) and the previously-assumed soft spectrum (left panel), shown on a common scale. For the hard spectrum, this redshift corresponds to the cosmic heating transition. In this comparison, both cases have the same underlying distribution of star formation at a given redshift, so they have the same ionized patches (at an early stage of reionization, when 14% of the IGM has been reionized) and a similar distribution pattern of gas temperature and of 21-cm temperature. However, the difference is visually striking, in that the map for the hard spectrum is strongly suppressed in terms of both the typical value of T_b and the typical size of its fluctuations. From [258].

Kilometre Array (§ 7.2); actually, in this case the Lyα peak is stronger and more extended in redshift, since it is not cut off by early heating as in the case of soft X-rays [258]. It could also affect other observations of high-redshift galaxies. For example, since late heating implies weak photoheating feedback during the cosmic heating era, low-mass halos may continue to produce copious stars in each region right up to its local reionization; note though that internal feedback (arising from supernovae or mini-quasars) could still limit star formation in small halos.

7.5. The global 21-cm spectrum

This section thus far has focused on 21-cm fluctuations, and in particular the 21-cm power spectrum. The power spectrum encodes a lot of information about the various sources of 21-cm fluctuations, and it is a rich dataset consisting of an entire function of wavenumber at each redshift, or potentially even much more than that due to the line-of-sight anisotropy (§ 4.3). This information can hopefully be extracted from data obtained with radio interferometers, after dealing with the expected thermal noise and sample variance, foreground residuals, and artifacts of the imperfectly-known responses of the radio antennae and receivers.

A very different approach is to measure the total sky spectrum and detect the redshift evolution of the global, cosmic mean 21-cm intensity. A global experiment requires a simple, relatively cheap setup (an all-sky antenna) compared to the fluctuation experiments, and the total sky naturally yields a higher signal-to-noise ratio and a spectrally smoother foreground than found in small patches (which are the basic units of the fluctuation experiments). In order to make success more likely, observations can focus on constraining sharp frequency features, without attempting to measure the absolute cosmological 21-cm emission level (which is much harder). During reionization, there should be a decrease in the global 21-cm emission due to the overall disappearance of atomic hydrogen (§ 6.1). This global step, while not sudden, is still expected to be fairly sharp in frequency. At higher redshifts, a sharp decrease towards negative brightness temperature should occur due to the rise of the first stars as a result of Lyα coupling of the cold

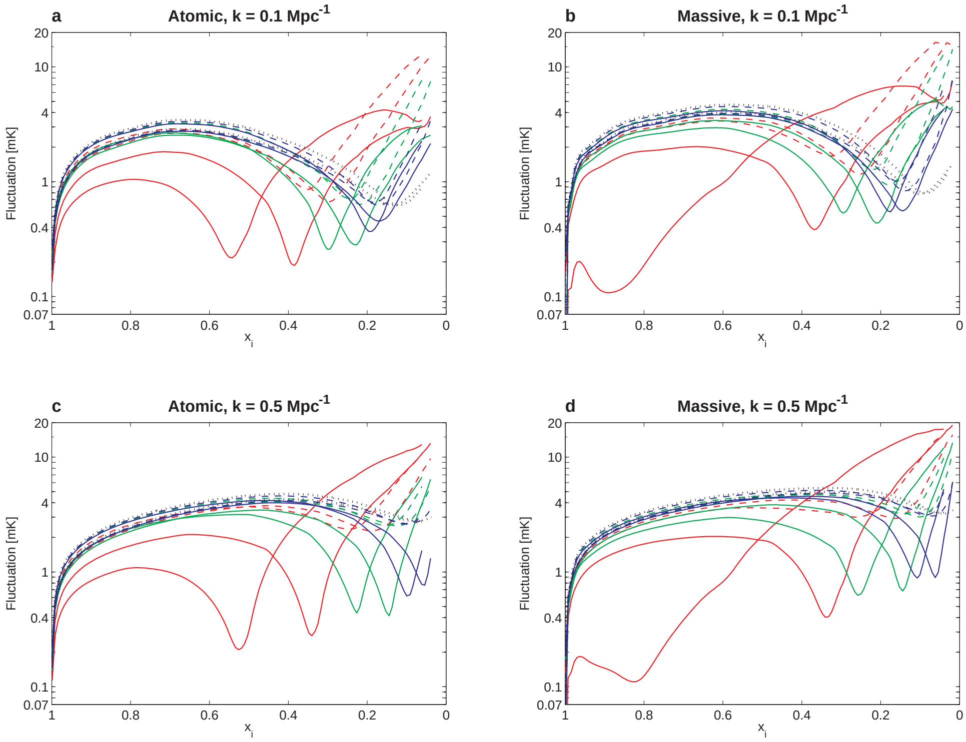

Figure 38: The fluctuation level of the 21-cm brightness temperature is shown versus the ionized (mass) fraction of the universe x_i (starting on the right from $z = 15$). We compare results obtained with a hard X-ray binary spectrum [228] (solid curves) to those with the previously-adopted soft spectrum (dashed curves), and show the commonly-assumed saturated heating case for reference (black dotted curves). The curves show various cases, in order to give a reasonable idea of the range of parameter space given current uncertainties. Thus, the best guess for the X-ray efficiency (green curves) is shown along with an efficiency lower (red curves) or higher (blue curves) by a factor of $\sqrt{10}$, each with either early or late reionization (given current uncertainties about its timing: see § 6.1). The fluctuation is shown at a wavenumber $k = 0.1$ Mpc^{-1} (top panels) or $k = 0.5$ Mpc^{-1} (bottom panels), for two possible cases of galactic halos, either a minimum halo mass set by atomic cooling (left panels) or halos that are ten times more massive (right panels). The lower k value roughly tracks large-scale fluctuations (heating early on, and ionized bubbles later), while the higher k value corresponds to a smaller scale (though one that can still be measured accurately with current experiments) and thus tracks more closely the evolution of density fluctuations. To illustrate the effect of the X-ray spectrum on the results, consider the fluctuation level at $k = 0.5$ Mpc^{-1} at the mid-point of reionization (i.e., $x_i = 0.5$); the parameter space explored here gives a possible range of 3.6–4.9 mK for the soft spectrum, while the hard spectrum gives a much broader range of 0.3–4.4 mK. Note also that the latter values are typically much lower than the often-assumed limit of saturated heating (which gives a corresponding range of 4.1–5.1 mK). From [92].

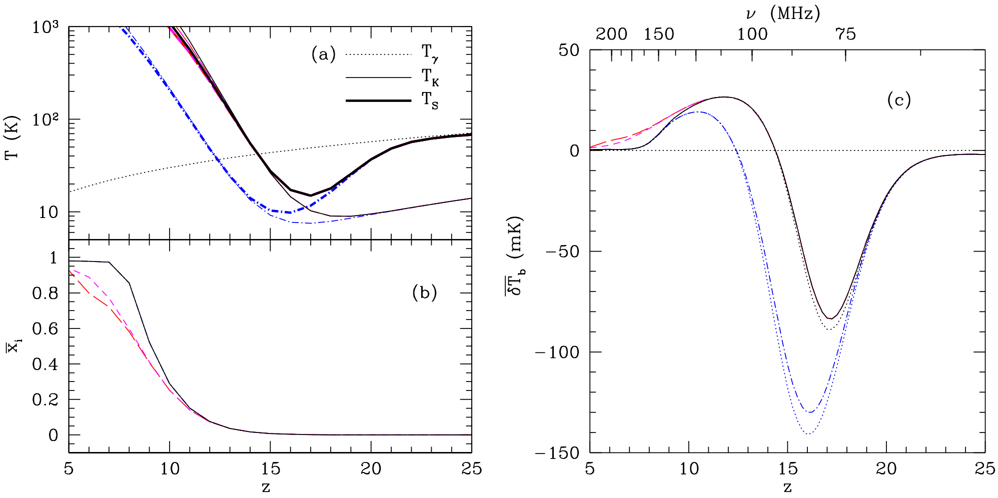

Figure 39: Examples of predicted global 21-cm histories, showing how they reflect the cosmic history of star formation and its various effects on the IGM. A range of parameters are shown in order to reflect a reasonable range of uncertainty: the best-guess X-ray efficiency (solid curves), a lower efficiency by a factor of 5 (dot-dashed curves), and the addition of two possible models for photoheating feedback (short- and long-dashed curves). Panel (a) shows the CMB (T_γ), gas kinetic (T_K) and spin (T_S) temperatures (dotted, thin, and thick solid curves, respectively). Panel (b) shows the progress of reionization, in terms of the cosmic mean ionized fraction $\bar{x}_i$. Panel (c) shows the resulting global mean 21-cm brightness temperature measured with respect to the CMB; in this panel, the two dotted lines show $T_{\rm b}$ if shock heating is ignored. Note that this panel shows the observed frequency on top in addition to the redshift on the bottom. All models here assume Pop II stars and a soft X-ray spectrum of heating sources. From [41].

IGM (§ 6.2), followed by a sharp rise up to positive values due to cosmic heating (§ 6.3). Thus, a detection of the global signal would trace the overall cosmic history of the first stars through their effect on 21-cm emission (Figure 39). Maximum-likelihood analyses of data fitting show that global 21-cm measurements during cosmic reionization should be able to detect a wide range of realistic models and measure the main features of the reionization history while constraining the key properties of the ionizing sources; this is true in analyses (that assumed the saturated heating limit) using a flexible toy model [42] or a ΛCDM-based model [43], though the results are rather sensitive to assumptions on just how difficult it will be to remove the effect of the foregrounds.

The late heating (§ 6.3 and § 7.4) expected due to the hard spectrum of X-ray binaries has a particularly important effect on the global 21-cm signal. The effect of late heating is to give the cosmic gas more time to cool adiabatically to well below the CMB temperature, thus producing mean 21-cm absorption that reaches a maximum depth in the range −110 to −180 mK at $z \sim 15-19$ (Figure 40). This may make it easier for experiments to detect the global 21-cm spectrum from before reionization and thus probe the corresponding early galaxies. Global experiments are most sensitive to the frequency derivative of the 21-cm brightness temperature; late heating extends the steep portion of the spectrum to higher frequencies, moving the maximum positive derivative to a ∼ 10% higher frequency (where the foregrounds are significantly weaker) while also changing the value of this maximum derivative by ±10%. On the other hand, at lower redshift, late heating significantly suppresses the global step from reionization, which suggests that global 21-cm experiments should focus instead on the earlier eras of Lyα coupling and

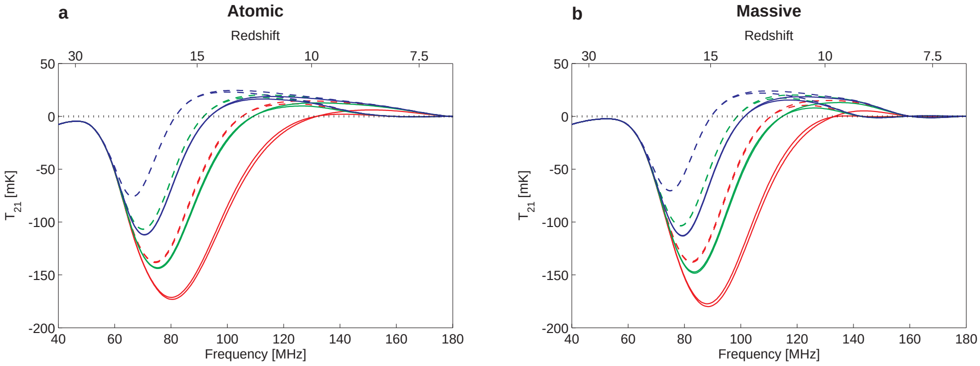

Figure 40: The effect of late cosmic heating on the global 21-cm spectrum. The cosmic mean 21-cm brightness temperature (T_{21}) relative to the CMB is shown versus observed frequency (with the redshift indicated on top), for the hard X-ray binary spectrum [228] (solid curves) and for the previously-adopted soft spectrum (dashed curves); note also the fiducial dotted line at $T_{21} = 0$. Various cases are shown, in order to give a reasonable idea of the range of parameter space given current uncertainties; the notation matches that in Figure 38. From [92].

cosmic heating.

8. Summary and Conclusions

The study of the first stars, galaxies, and black holes, and their effect on the 21-cm sky, is entering a critical stage. While this subject has been developing theoretically for several decades, including a great acceleration in recent years, observationally this field is in its infancy. Thus, we are about to experience that pinnacle of excitement that comes with the first clash of a scientific theory with experimental data. In such a collision of theoretical expectations with reality, there are several possible outcomes. The predictions can be perfectly verified, an outcome that makes the successful theorists gleeful and proud, but at the same time is extremely boring. At the other extreme, the predictions can fail completely, making the theorists a laughing-stock, but revealing previously unexpected cosmic events, which makes this possibility the most exciting one. Neither of these extreme possibilities is expected in the case of 21-cm cosmology. The shear magnitude of the uncertainty about high-redshift astrophysics makes the first possibility unlikely, even in the absence of exotic cosmic events such as dark matter decay. On the other hand, complete failure is made unlikely by the fact that the theory is grounded in solid atomic physics as well as models of galaxy formation that are significantly constrained by observations of the current Universe, at one end, and the CMB at the other (initial condition) end. Thus, the most likely outcome is an intermediate one, where the overall framework of theoretical expectations will be confirmed, but with some, hopefully interesting and significant, surprises, such as an unexpected, new class of astrophysical sources (which will be noticed if it dominated one of the types of radiation that drove the 21-cm emission). Regardless of the precise outcome, it is likely that once a clear detection of the 21-cm signal from early cosmic history is achieved, the field will get a big boost, analogous to the development of CMB observations and theory after the first detection of CMB temperature fluctuations by the COBE satellite. This breakthrough moment for 21-cm cosmology will hopefully occur within the next few years, and will be followed up with confirmations and more detailed measurements soon afterwards.

A great wealth of data is potentially available in 21-cm cosmology (§ 6 and § 7). Even just the isotropically-averaged 21-cm power spectrum, measured as a function of wavenumber and redshift, is a rich data set that probes many details of the various cosmological and astrophysical sources of 21-cm fluctuations (see Figure 41). A number of cosmic events leave clear signatures in the power spectrum, but the redshifts of the associated features (such as the peaks) vary with scale, since several different sources of 21-cm fluctuations contribute at any given time, and these sources vary differently with scale. In the model shown in Figure 41, for which reionization ends at $z \sim 7$, the reionization peak of fluctuations occurs in the range $z = 7.5 - 9$ depending on wavenumber. While the uncertainties are still large, it now seems that the IGM was most likely heated by X-ray sources with a hard spectrum (§ 6.3 and § 7.4), a possibility not considered until recently; in this case, the cosmic heating transition produces a clear minimum on small scales, but a weak heating peak remains on the largest scales that are larger than the typical distance traveled even by hard X-rays. Continuing with Figure 41, the Lyα peak occurs in this example at $z = 18 - 20$, and (generally in the case of late heating) it is both the strongest and highest-redshift signal from the first stars (In the case of a soft X-ray spectrum, the heating peak is somewhat higher than the Lyα peak [258]). We note that additional theoretical uncertainties result from the complexity of the astrophysics during early times, including substantial transitions in the basic character of star formation expected due to various types of stellar feedback such as supernova outflows, LW radiation, and metal enrichment. The dark ages, during which 21-cm emission is not significantly affected by astrophysical sources and becomes a purely cosmological probe, begin at $z > 30$; at this point the predicted fluctuation signal is quite low, and since the galactic foreground increases rapidly with redshift (with the brightness temperature of the sky $\propto (1+z)^{2.6}$ [88]), observations of this era lie in the somewhat distant future.

Actually measuring a data set like that shown in Figure 41 would obviously constitute an amazing advance in our understanding of cosmic history. However, it is important to also look for robust, model-independent signatures that can convincingly confirm and complement the results obtained from fitting parameterized models to the (angle-averaged) 21-cm power spectrum. This is particularly needed in a field looking to probe a new, unexplored regime of cosmic history. Luckily, the field of 21-cm cosmology has indeed emerged as a very rich one. For example, the line-of-sight anisotropy of the power spectrum (§ 4.3) is potentially an immensely important source of additional information, and it has only begun to be explored. It can provide a number of model-independent probes of early galaxies that would complement inferences made based on the angle-averaged power spectrum. In particular, the dominant μ^2 term of the anisotropy acts as a cosmic clock, its sign changing as it tracks various cosmic milestones; for instance, measuring it to be negative during reionization would directly confirm the inside-out topology of this transition (i.e., where overdense regions reionize first). Another example of a possible model-independent signature is the streaming velocity with its associated strong BAO features. In addition, the global 21-cm spectrum (§ 7.5) is a wonderfully complementary probe of the same cosmic history. One way to express this is that the 21-cm fluctuations can be written as a product of the mean intensity and its relative fluctuations, and information on the global spectrum helps to separate these two quantities and thus break a degeneracy.

In this review we have focused on the 21-cm power spectrum (including its angular anisotropy). There are good reasons for this, even though it is not an open and shut case as in CMB studies, where the power spectrum carries the most important cosmological information in the signal (which is thought to reflect the underlying Gaussian random field of primordial perturbations). In general, there are two different modes for studying galaxies: The collective (galaxy clustering) and the individual (studying individual galaxies). 21-cm cosmology during cosmic dawn

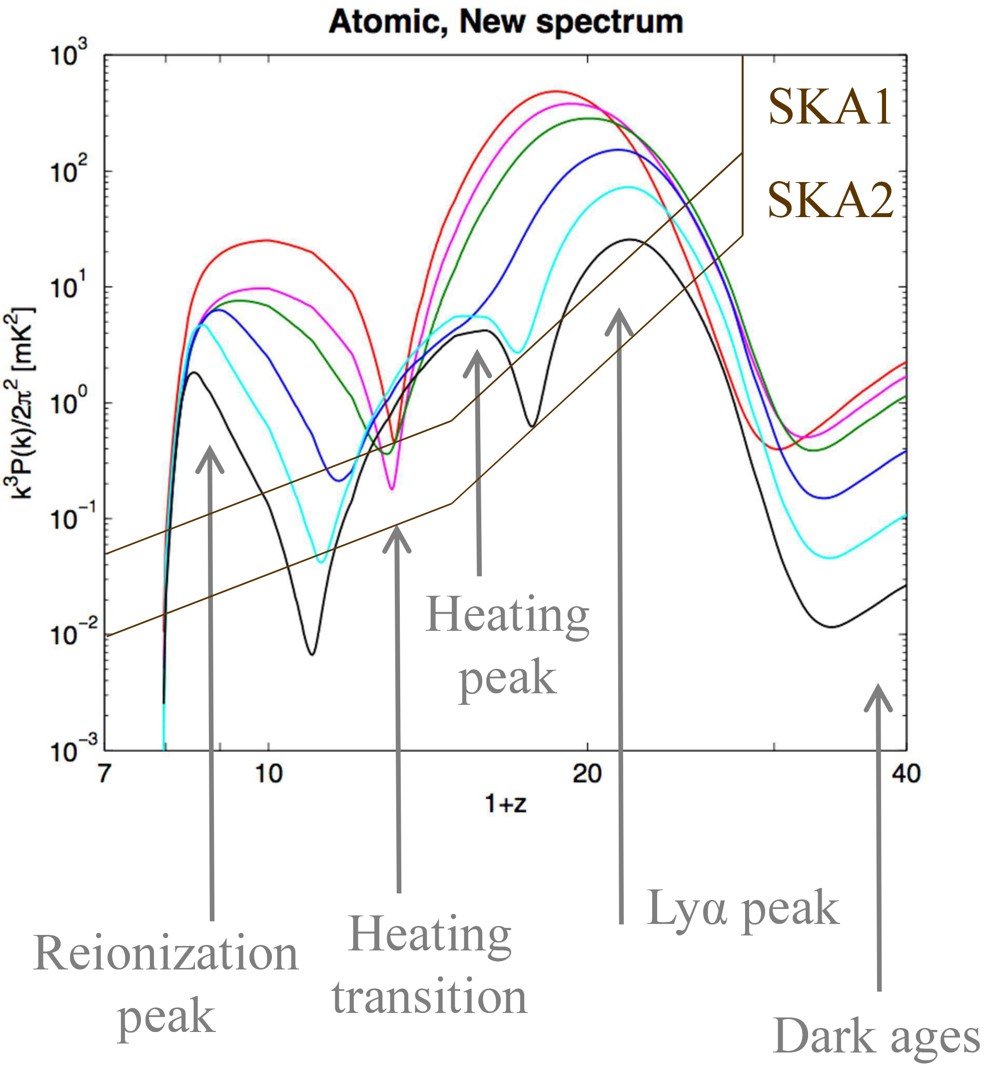

Figure 41: A summary view of the rich complexity of even just the isotropically-averaged 21-cm power spectrum, shown via the evolution with redshift of the squared 21-cm fluctuation at various wavenumbers. Going from small to large scales, shown are $k = 1$ Mpc^{-1} (red), $k = 0.5$ Mpc^{-1} (magenta), $k = 0.3$ Mpc^{-1} (green), $k = 0.1$ Mpc^{-1} (blue), $k = 0.05$ Mpc^{-1} (cyan), and $k = 0.03$ Mpc^{-1} (black). The particular model shown here assumes cosmic heating by a hard X-ray spectrum (§ 6.3 and § 7.4), and that stars form in all halos in which the gas can cool via atomic cooling. Also shown is the approximate observational thermal noise power spectrum expected for the SKA phase 1 and phase 2 (at $k = 0.1$ Mpc^{-1}) [140]. From [258], with added labels and observational sensitivities.

and the EOR will be dominated by the collective regime. The 21-cm fluctuations will be dominated by various radiation fields, and the intensity of those fields at any point will be made up of the contributions of many individual sources, except perhaps in a few rare regions. Thus, the structures that will be seen will be a collective effect, and thus mainly dependent on the clustering of sources. The power spectrum naturally measures this clustering. More specifically, the distribution of sources throughout this era is driven by the underlying density distribution of matter (except for the additional effect of the streaming velocity). This density distribution is determined by the power spectrum, and for linear fluctuations, the 21-cm map is also determined by its power spectrum (which is the underlying power spectrum times a window function, corresponding to a convolution in real space that accounts for the spatial redistribution of photons of the various relevant frequency regimes). It is true that there are some non-linear distortions along the way, but still, on the (relatively large) scales resolvable by upcoming radio arrays, the power spectrum should capture most of the information available in a full image. Indeed, as described throughout this review, the 21-cm power spectrum can be used to reconstruct the most interesting astrophysical information that we desire: at what redshifts Lyα coupling, cosmic heating, and reionization occurred, how fast they progressed, which galactic halos dominated each era, and what the spectrum was of the sources (e.g., the X-ray spectrum in the case of X-ray heating). The most non-linear process is reionization (with its sharp edges in the expected scenario that is dominated by UV photons), but the non-Gaussianity of the ionization field only reflects the rapid absorption of ionizing photons, and may not probe much interesting physics beyond that. Also, in the near future the power spectrum is likely to be the main available observable from the least explored, and thus most exciting, high-redshift regime of the pre-EOR cosmic dawn; imaging from such an early time will be difficult even for the SKA.

That said, the non-Gaussianity of 21-cm fluctuations [259] does make other statistics beyond the power spectrum interesting, including the bispectrum [37, 260], the 21-cm PDF (probability distribution function, i.e., histogram of values of the 21-cm brightness temperature) [38, 39, 178, 40, 261, 262], and the difference PDF (i.e., histogram of T_b differences between pixel pairs) [263, 264]; some of the additional information available in the PDF can be captured by its skewness [265, 266]. Also, while in this review we have focused heavily on the emerging field of 21-cm cosmology, other cosmological probes are making rapid advances and should explore complementary aspects of high-redshift galaxies. The James Webb Space Telescope (JWST; http://www.jwst.nasa.gov/) should discover at least the largest galaxies at early times, as well as rare bright objects (such as supernovae or gamma-ray bursts) in more typical galaxies. The planned generation of larger ground-based optical/IR telescopes, including the Thirty Meter Telescope (TMT; http://www.tmt.org/), the Giant Magellan Telescope (GMT; http://www.gmto.org/), and the European Extremely Large Telescope (E-ELT; http://www.eso.org/public/teles-instr/e-elt/) should give us detailed, spectroscopic information on some of these objects and their surrounding IGM. Imaging the 21-cm sky, as planned for the SKA, will be very interesting around particular bright objects. In another area, the CMB, in addition to its further development as a cosmological probe, may allow the detection of the small-scale signature of CMB scattering by the ionized bubbles during cosmic reionization [172, 173, 174, 267].

We have also discussed in this review the complementary interaction in this field between numerical simulations, analytical (or semi-analytical) models, and semi-numerical methods. Each method has its advantages and disadvantages, and in particular it is important not to overlook the limitations of numerical simulations (§ 3.2).

Another highlight of this work is in pointing out how the idea of unusually large fluctuations in the number density of high-redshift galaxies (§ 3.1) is a common thread that has driven the whole topic of 21-cm fluctuations, from the understanding of the character of reionization (§ 6.1) to the first predictions of large-scale 21-cm fluctuations from the inhomogeneous Lyα (§ 7.2) and X-ray (§ 7.3) backgrounds. It has recently been joined by an exciting new source of large-scale fluctuations, the supersonic streaming velocity (§ 5). This new source comes with a strong signature of baryon acoustic oscillations, making it a potential tool for identifying the presence of tiny, million solar mass halos at very early times. The streaming velocity certainly had a major effect on the first generation of stars, and it may also have had a great significance at redshifts that are observable with 21-cm experiments (§ 7.3), though this depends on just how efficiently such small halos were able to form stars. We have also highlighted a different issue, one of basic 21-cm physics: we have helped clarify the literature regarding the low-temperature corrections to the basic expressions of 21-cm cosmology (§ 4.2).

In this review we have focused on the astrophysical era of 21-cm cosmology that is accessible to upcoming experiments. However, it is also important to keep in mind the great long-term promise of the development of 21-cm cosmology. When 21-cm measurements reach small spatial scales, this will open up a variety of new probes and applications, especially in the dark ages during which 21-cm cosmology will be a clean cosmological probe. For example, 21-cm fluctuations should be present down to much smaller scales than CMB fluctuations (which are cut off by the combination of Silk damping and the width of the surface of last scattering). This implies a far greater potential sensitivity of 21-cm measurements to a small primordial non-Gaussianity [12, 268]. Measuring the primordial power spectrum on small scales will also probe the tilt of the power spectrum and could potentially uncover a cutoff due to dark matter properties (such as in the warm dark matter or fuzzy dark matter [269] models). Also, the gas temperature can in principle be mapped through its effect on the small-scale power spectrum (i.e., the filtering mass discussed in § 2.4) as well as more directly through the anisotropic effect of the thermal smoothing of the 21-cm power spectrum [70]; e.g., if the cosmic gas is radiatively heated to 10^3 K, then the smoothing is expected on a scale of $\sim$ 20 kpc. On small scales, the supersonic streaming velocity also has a significant effect on the 21-cm power spectrum [159]. Further back in time, a 21-cm signal is expected from the cosmological epoch of recombination [270].

We would like to end in the same way that the author concluded a review written a decade ago [21], with the sincere hope of not having to write this again a decade from now: Astronomers are eager to start tuning into the cosmic radio channels of 21-cm cosmology.

9. Acknowledgments

I would like to thank my collaborators over the years on the reviewed work, especially Avi Loeb, my former students Smadar Naoz and Anastasia Fialkov, my current student Aviad Cohen, and Eli Visbal. This work was supported by Israel Science Foundation grant 823/09 and the Ministry of Science and Technology, Israel. This work was partly done within the Labex Institut Lagrange de Paris (ILP, reference ANR-10-LABX-63) part of the Idex SUPER, and received financial state aid managed by the Agence Nationale de la Recherche, as part of the programme Investissements d'avenir under the reference ANR-11-IDEX-0004-02. I also acknowledge a Leverhulme Trust Visiting Professorship at the University of Oxford. This research was supported in part by the Perimeter Institute for Theoretical Physics. Research at Perimeter Institute is supported by the Government of Canada through Industry Canada and by the Province of Ontario through the Ministry of Economic Development & Innovation.

References

[1] H. Aihara, et al., Astroph. J. Supp. 193 (2011) 29; erratum – 195 26.
[2] M. Colless, et al., Mon. Not. R. Astron. Soc. 328 (2001) 1039.
[3] V. Springel, C. S. Frenk, S. D. M. White, Nature 440 (2006) 1137.
[4] J. R. Bond, L. Kofman, D. Pogosyan, Nature 380 (1996) 603.
[5] P. A. Oesch, G. Brammer, P. G. van Dokkum, et al., ApJ 819 (2016) 129
[6] V. Bromm, P. S. Coppi, R. B. Larson, Astroph. J. 527 (1999) L5.
[7] T. Abel, G. L. Bryan, M. L. Norman, Science 295 (2002) 93.
[8] S. Naoz, S. Noter, R. Barkana, Mon. Not. R. Astron. Soc. 373 (2006) L98.
[9] A. Fialkov, R. Barkana, D. Tseliakhovich, C. Hirata, Mon. Not. R. Astron. Soc. 424 (2012) 1335.
[10] C. J. Hogan, M. J. Rees, Mon. Not. R. Astron. Soc. 188 (1979) 791.
[11] P. Madau, A. Meiksin, M. J. Rees, Astroph. J. 475 (1997) 429.
[12] A. Loeb, M. Zaldarriaga, Phys. Rev. Lett. 92 (2004) 211301.
[13] R. Barkana, A. Loeb, Astroph. J. 624 (2005) 65.
[14] A. Nusser, Mon. Not. R. Astron. Soc. 364 (2005) 743.
[15] S. S. Ali, S. Bharadwaj, B. Pandey, Mon. Not. R. Astron. Soc. 363 (2005) 251.
[16] R. Barkana, Mon. Not. R. Astron. Soc. 372 (2006) 259.
[17] D. Scott, M. J. Rees, Mon. Not. R. Astron. Soc. 247 (1990) 510.
[18] S. A. Wouthuysen, Astron. J. 57 (1952) 31.
[19] G. B. Field, Proc. IRE 46 (1958) 240.
[20] Z. Haiman, M. J. Rees, A. Loeb, Astroph. J. 476 (1997) 458; erratum – 484 985.
[21] R. Barkana, Science 313 (2006) 931.
[22] J. E. Gunn, B. A. Peterson, Astroph. J. 142 (1965) 1633.
[23] R. Barkana, A. Loeb, Phys. Rep. 349 (2001) 125.
[24] R. Barkana, A. Loeb, Astroph. J. 609 (2004) 474.
[25] S. R. Furlanetto, M. Zaldarriaga, L. Hernquist, Astroph. J. 613 (2004) 1.
[26] I. T. Iliev, G. Mellema, U.-L. Pen, et al., Mon. Not. R. Astron. Soc. 369 (2006) 1625.
[27] P. Tozzi, P. Madau, A. Meiksin, M. J. Rees, Astrophys. J. 528 (2000) 597.
[28] R. Barkana, A. Loeb, Astroph. J. 626 (2005) 1.
[29] J. R. Pritchard, S. R. Furlanetto, Mon. Not. R. Astron. Soc. 376 (2007) 1680.
[30] J. D. Bowman, M. F. Morales, J. N. Hewitt, Astroph. J. 695 (2009) 183.
[31] G. Harker, Mon. Not. R. Astron. Soc. 405 (2010) 2492.
[32] G. Paciga, T.-C. Chang, Y. Gupta, et al., Mon. Not. R. Astron. Soc. 413 (2011) 1174.
[33] Z. S. Ali, A. R. Parsons, H. Zheng, et al., Astroph. J. 809 (2015) 61.
[34] J. D. Bowman, M. F. Morales, J. N. Hewitt, Astroph. J. 661 (2007) 1.
[35] M. McQuinn, O. Zahn, M. Zaldarriaga, L. Hernquist, S. R. Furlanetto, Astroph. J. 653 (2006) 815.
[36] R. Barkana, Mon. Not. R. Astron. Soc. 397 (2009) 1454.
[37] S. Bharadwaj, S. K. Pandey, Mon. Not. R. Astron. Soc. 358 (2005) 968.
[38] B. Ciardi, P. Madau, Astroph. J. 596 (2003) 1.
[39] S. R. Furlanetto, M. Zaldarriaga, L. Hernquist, Astroph. J. 613 (2004) 16.
[40] K. Ichikawa, R. Barkana, I. T. Iliev, G. Mellema, P. R. Shapiro, Mon. Not. R. Astron. Soc. 406 (2010) 2521.
[41] S. R. Furlanetto, Mon. Not. R. Astron. Soc. 371 (2006) 867.
[42] J. R. Pritchard, A. Loeb, Phys. Rev. D 82 (2010) 023006.
[43] A. Morandi, R. Barkana, Mon. Not. R. Astron. Soc. 424 (2012) 2551.
[44] J. D. Bowman, A. E. E. Rogers, Nature 468 (2010) 796.
[45] D. Tseliakhovich, C. M. Hirata, Phys. Rev. D 82 (2010) 083520.
[46] A. Loeb, S. R. Furlanetto, The First Galaxies in the Universe, Princeton University Press, Princeton, 2013.
[47] S. Weinberg, Gravitation and Cosmology, Wiley, New York, 1972.
[48] E. W. Kolb, M. S. Turner, The Early Universe, Addison-Wesley, Redwood City, CA, 1990.
[49] Planck Collaboration, P. A. R. Ade, N. Aghanim, et al. Astron. & Astroph. 566 (2014) A54.
[50] P. J. E. Peebles, The Large-Scale Structure of the Universe, Princeton University Press, Princeton, 1980.
[51] P. J. E. Peebles, Principles of Physical Cosmology, Princeton University Press, Princeton, 1993.
[52] J. F. Navarro, C. S. Frenk, S. D. M. White, Astrophys. J. 490 (1997) 493.
[53] W. H. Press, P. Schechter, Astroph. J. 187 (1974) 425.
[54] J. R. Bond, S. Cole, G. Efstathiou, N. Kaiser, Astroph. J. 379 (1991) 440.
[55] R. K. Sheth, G. Tormen, Mon. Not. R. Astron. Soc. 308 (1999) 119.
[56] R. K. Sheth, G. Tormen, Mon. Not. R. Astron. Soc. 329 (2002) 61.
[57] A. Jenkins, et al., Mon. Not. R. Astron. Soc. 321 (2001) 372.

[58] N. Kaiser, Astroph. J. 284 (1984) L9.
[59] J. M. Bardeen, J. R. Bond, N. Kaiser, A. S. Szalay, Astroph. J., 304 (1986) 15.
[60] S. Cole, N. Kaiser, Mon. Not. R. Astron. Soc. 237 (1989) 1127.
[61] H. J. Mo, S. D. M. White, Mon. Not. R. Astron. Soc. 282 (1996) 347.
[62] S. Naoz, R. Barkana, Mon. Not. R. Astron. Soc. 377 (2007) 667.
[63] R. Barkana, A. Loeb, Mon. Not. R. Astron. Soc. 363 (2005) L36.
[64] R. Barkana, A. Loeb, Mon. Not. R. Astron. Soc. 415 (2011) 3113.
[65] M. T. Soumagnac, R. Barkana, C. G. Sabiu, et al., 2016, arXiv:1602.01839
[66] C. Ma, E. Bertschinger, Astroph. J. 455 (1995) 7.
[67] U. Seljak, M. Zaldarriaga, Astroph. J. 469 (1996) 437.
[68] K. Yamamoto, N. Sugiyama, H. Sato, Phys. Rev. D, 56 (1997) 7566.
[69] K. Yamamoto, N. Sugiyama, H. Sato, Astroph. J. 501 (1998) 442.
[70] S. Naoz, R. Barkana, Mon. Not. R. Astron. Soc. 362 (2005) 1047.
[71] N. Y. Gnedin, L. Hui, Mon. Not. R. Astron. Soc. 296 (1998) 44.
[72] N. Y. Gnedin, Astroph. J. 542 (2000) 535.
[73] M. Hoeft, G. Yepes, S. Gottlöber, V. Springel, Mon. Not. R. Astron. Soc. 371 (2006) 401.
[74] T. Okamoto, L. Gao, T. Theuns, Mon. Not. R. Astron. Soc. 390 (2008) 920.
[75] S. Naoz, R. Barkana, A. Mesinger, Mon. Not. R. Astron. Soc. 399 (2009) 369.
[76] S. Naoz, N. Yoshida, R. Barkana, Mon. Not. R. Astron. Soc. 416 (2011) 232.
[77] E. Sobacchi, A. Mesinger, Mon. Not. R. Astron. Soc. 432 (2013) L51.
[78] Y. Noh, M. McQuinn, Mon. Not. R. Astron. Soc. 444 (2014) 503.
[79] W. C. Saslaw, D. Zipoy, Nature 216 (1967) 976.
[80] N. Yoshida, A. Sokasian, L. Hernquist, V. Springel, Astroph. J. 598 (2003) 73.
[81] T. Abel, G. L. Bryan, M. L. Norman, Science 295 (2002) 93.
[82] A. Loeb, International Journal of Astrobiology 13 (2014) 337.
[83] J. Miralda-Escudé, Science 300 (2003) 1904.
[84] J. H. Wise, T. Abel, Astroph. J. 629 (2005) 615.
[85] R. Casas-Miranda,H. J. Mo, R. K. Sheth, G. Börner, Mon. Not. R. Astron. Soc. 333 (2002) 730.
[86] D. N. Spergel, R. Bean, O. Doré, et al., Astroph. J. 170 (2007) 377.
[87] M. Viel, M. G. Haehnelt, A. Lewis, Mon. Not. R. Astron. Soc. 370 (2006) L51.
[88] S. R. Furlanetto, S. P. Oh, F. H. Briggs, Phys. Rep. 433 (2006) 181.
[89] V. Springel, L. Hernquist, Mon. Not. R. Astron. Soc. 339 (2003) 312.
[90] J. S. B. Wyithe, A. Loeb, Nature 432 (2004) 194.
[91] A. Mesinger, S. Furlanetto, R. Cen, Mon. Not. R. Astron. Soc. 411 (2011) 955.
[92] A. Fialkov, R. Barkana, E. Visbal, Nature 506 (2014) 197.
[93] V. Bromm, Rep. Prog. Phys. 76 (2013) 112901.
[94] T. Karlsson, V. Bromm, J. Bland-Hawthorn, Rev. Mod. Phys. 85 (2013) 809.
[95] R. Durrer, A. Neronov, Astron. Astroph. Rev. 21 (2013) 62.
[96] L. M. Widrow, D. Ryu, D. R. G. Schleicher, et al., Space Science Rev. 166 (2012) 37.
[97] S. Naoz, R. Narayan, Phys. Rev. Lett. 111 (2013) 051303.
[98] Z. Haiman, A. A. Thoul, A. Loeb, Astrophys. J. 464 (1996) 523.
[99] T. Abel, P. Anninos, Y. Zhang, M. L. Norman, New Astron. 2 (1997) 181.
[100] D. Galli, F. Palla, Astron. Astrophys. 335 (1998) 403.
[101] S. C. O. Glover, T. Abel, Mon. Not. R. Astron. Soc. 388 (2008) 1627.
[102] M. Tegmark, J. Silk, M. J. Rees, A. Blanchard, T. Abel, F. Palla, Astrophys. J. 474 (1997) 1.
[103] M. J. Turk, T. Abel, B. O'Shea, Science 325 (2009) 601.
[104] J. C. Tan, C. F. McKee, Astrophys. J. 603 (2004) 383
[105] A. Stacy, T. H. Greif, V. Bromm, Mon. Not. R. Astron. Soc. 403 (2010) 45.
[106] P. C. Clark, S. C. O. Glover, R. J. Smith, T. H. Greif, R. S. Klessen, V. Bromm, Science 331 (2011) 1040.
[107] T. H. Greif, V. Bromm, P. C. Clark, S. C. O. Glover, R. J. Smith, R. S. Klessen, N. Yoshida, V. Springel, Mon. Not. R. Astron. Soc. 424 (2012) 399.
[108] S. Hirano, T. Hosokawa, N. Yoshida, et al. Astroph. J. 781 (2014) 60.
[109] S. Hirano, T. Hosokawa, N. Yoshida, K. Omukai, H. W. Yorke, Mon. Not. R. Astron. Soc. 448 (2015) 568.
[110] M. F. Morales, J. S. B. Wyithe, Ann. Rev. Astron. Astroph. 48 (2010) 127.
[111] J. R. Pritchard A. Loeb, Rep. Prog. Phys. 75 (2012) 086901.
[112] E. M. Purcell, G. B. Field, Astroph. J. 124 (1956) 542.
[113] G. B. Field, Proc. IRE, 46 (1958) 240.
[114] G. B. Rybicki, Astroph. J. 647 (2006) 709.
[115] A. C. Allison, A. Dalgarno, Astroph. J. 158 (1969) 423.

[116] B. Zygelman, Astroph. J. 622 (2005) 1356.
[117] G. B. Field, Astroph. J. 129 (1959) 551.
[118] L. Chuzhoy, P. R. Shapiro, Astroph. J. 651 (2006) 1.
[119] S. I. Grachev, Astrofizika, 30 (1989) 347.
[120] X. Chen, J. Miralda-Escudé, Astroph. J. 602 (2004) 1.
[121] M. M. Basko, Astrophysics 17 (1981) 69.
[122] G. B. Rybicki, I. P. dell'Antonio, Astroph. J. 427 (1994) 603.
[123] C. M. Hirata, Mon. Not. R. Astron. Soc. 367 (2006) 259.
[124] J. R. Pritchard, S. R. Furlanetto, Mon. Not. R. Astron. Soc. 372 (2006) 1093.
[125] N. Kaiser, Mon. Not. R. Astron. Soc. 227 (1987) 1.
[126] S. Bharadwaj, S. S. Ali, Mon. Not. R. Astron. Soc. 352 (2004) 142.
[127] Y. Mao, P. R. Shapiro, G. Mellema, et al., Mon. Not. R. Astron. Soc. 422 (2012) 926.
[128] H. Jensen, K. K. Datta, G. Mellema, et al., Mon. Not. R. Astron. Soc. 435 (2013) 460.
[129] P. R. Shapiro, Y. Mao, I. T. Iliev, et al., Phys. Rev. Lett. 110 (2013) 151301.
[130] A. Fialkov, R. Barkana, A. Cohen, Phys. Rev. Lett. 114 (2015) 101303.
[131] F. Beutler, S. Saito, H.-J. Seo, et al., Mon. Not. R. Astron. Soc. 443 (2014) 1065.
[132] U. Seljak, J. Cosmo. Astropart. Phys. 3 (2012) 004.
[133] R. Barkana, A. Loeb, Mon. Not. R. Astron. Soc. 372 (2006) 43.
[134] K. K. Datta, G. Mellema, Y. Mao, et al., Mon. Not. R. Astron. Soc. 424 (2012) 1877.
[135] K. Zawada, B. Semelin, P. Vonlanthen, S. Baek, Y. Revaz, Mon. Not. R. Astron. Soc. 439 (2014) 1615.
[136] P. La Plante, N. Battaglia, A. Natarajan, et al., Astroph. J. 789 (2014) 31.
[137] K. K. Datta, H. Jensen, S. Majumdar, et al., Mon. Not. R. Astron. Soc. 442 (2014) 1491.
[138] C. Alcock, B. Paczynski, Nature 281 (1979) 358.
[139] P. A. Shaver, R. A. Windhorst, P. Madau, A. G. de Bruyn, Astron. & Astroph. 345 (1999) 380.
[140] L. Koopmans, J. Pritchard, G. Mellema, J. Aguirre, K. Ahn, R. Barkana, et al., Advancing Astrophysics with the Square Kilometre Array (AASKA14), 2015, p. 1.
[141] G. Mellema, L. V. E. Koopmans, F. A. Abdalla, et al., Experimental Astronomy 36 (2013) 235.
[142] A. Fialkov, Intern. J. Mod. Phys. D 23 (2014) 30017.
[143] N. Dalal, U.-L. Pen, U. Seljak, J. Cosmo. Astropart. Phys. 11 (2010) 7.
[144] D. Tseliakhovich, R. Barkana, C. Hirata, Mon. Not. R. Astron. Soc. 418 (2011) 906.
[145] U. Maio, L. V. E. Koopmans, B. Ciardi, Mon. Not. R. Astron. Soc. 412 (2011) L40.
[146] A. Stacy, V. Bromm, A. Loeb, Astroph. J. 730 (2011) 1.
[147] T. Greif, S. White, R. Klessen, V. Springel, Astroph. J. 736 (2011) 147.
[148] R. M. O'Leary, M. McQuinn, Astroph. J. 760 (2012) 4.
[149] M. McQuinn, R. M. O'Leary, Astroph. J. 760 (2012) 3.
[150] S. Naoz, N. Yoshida, N. Y. Gnedin, Astroph. J. 747 (2012) 128.
[151] S. Naoz, N. Yoshida, N. Y. Gnedin, Astroph. J. 763 (2013) 27.
[152] J. Yoo, N. Dalal, U. Seljak, J. Cosmo. Astropart. Phys. 7 (2011) 18.
[153] J. Yoo, U. Seljak, Phys. Rev. D 88 (2013) 103520.
[154] J. Bovy, C. Dvorkin, Astroph. J. 768 (2013) 70.
[155] A. Fialkov, R. Barkana, E. Visbal, D. Tseliakhovich, C. M. Hirata, Mon. Not. R. Astron. Soc. 432 (2013) 2909.
[156] T. L. Tanaka, M. Li, Z. Haiman, Mon. Not. R. Astron. Soc. 435 (2013) 3559.
[157] T. L. Tanaka, M. Li, Mon. Not. R. Astron. Soc. 439 (2014) 1092.
[158] S. Naoz, R. Narayan, Astroph. J. 791 (2014) L8.
[159] Y. Ali-Haïmoud, P. D. Meerburg, S. Yuan, Phys. Rev. D 89 (2014) 083506.
[160] R. Cen, J. P. Ostriker, Astroph. J. 417 (1993) 404.
[161] N. Y. Gnedin, J. P. Ostriker, Astrophys. J. 486 (1997) 581.
[162] M. Fukugita, M. Kawasaki, Mon. Not. R. Astron. Soc. 269 (1994) 563.
[163] P. R. Shapiro, M. L. Giroux, A. Babul, Astroph. J. 427 (1994) 25.
[164] M. Kamionkowski, D. N. Spergel, N. Sugiyama, Astroph. J. 426 (1994) 57.
[165] M. Tegmark, J. Silk, A. Blanchard, Astroph. J. 420 (1994) 484.
[166] Z. Haiman, A. Loeb, Astroph. J. 483 (1997) 21.
[167] P. Valageas, J. Silk, Astron. & Astroph. 347 (1999) 1.
[168] N. Y. Gnedin, Astroph. J. 535 (2000) 530.
[169] Z. Haiman, M. Spaans, Astroph. J. 518 (1999) 138.
[170] Z. Haiman, Astroph. J. 576 (2002) L1.
[171] C. L. Carilli, N. Y. Gnedin, F. Owen, Astroph. J. 577 (2002) 22.
[172] N. Aghanim, F. X. Desert, J. L. Puget, R. Gispert, Astron. & Astroph. 311 (1996) 1.
[173] A. Gruzinov, W. Hu, Astroph. J. 508 (1998) 435.

[174] M. G. Santos, A. Cooray, Z. Haiman, L. Knox, C.-P. Ma, Astroph. J. 598 (2003) 756.
[175] J. Miralda-Escudé, M. Haehnelt, M. J. Rees, Astroph. J. 530 (2000) 1.
[176] S. R. Furlanetto, A. Sokasian, L. Hernquist, Mon. Not. R. Astron. Soc. 347 (2004) 187.
[177] B. Ciardi, F. Stoehr, S. D. M. White, Mon. Not. R. Astron. Soc. 343 (2003) 1101.
[178] G. Mellema, I. T. Iliev, U.-L. Pen, P. R. Shapiro, Mon. Not. R. Astron. Soc. 372 (2006) 679.
[179] R. Barkana, Mon. Not. R. Astron. Soc. 391 (2008) 727.
[180] O. Zahn, A. Lidz, M. McQuinn, S. Dutta, L. Hernquist, M. Zaldarriaga, S. R. Furlanetto, Astroph. J. 654 (2007) 12.
[181] M. G. Santos, A. Amblard, J. Pritchard, H. Trac, R. Cen, A. Cooray, Astroph. J. 689 (2008) 1.
[182] Planck Collaboration, Aghanim, N., Ashdown, M., et al. 2016, arXiv:1605.02985
[183] Planck Collaboration, Adam, R., Aghanim, N., et al. 2016, arXiv:1605.03507
[184] S. G. Djorgovski, S. Castro, D. Stern, A. A. Mahabal, Astroph. J. 560 (2001) L5.
[185] R. H. Becker, X. Fan, R. L. White, et al., Astron. J. 122 (2001) 2850.
[186] X. Fan, V. K. Narayanan, M. A. Strauss, et al., Astron. J. 123 (2002) 1247.
[187] X. Fan, M. A. Strauss, R. H. Becker, et al., Astron. J. 132 (2006) 117.
[188] J. Schroeder, A. Mesinger, Z. Haiman, Mon. Not. R. Astron. Soc. 428 (2013) 3058.
[189] R. Barkana, New Astron. 7 (2002) 85.
[190] A. Nusser, A. J. Benson, N. Sugiyama, C. Lacey, Astroph. J. 580 (2002) L93.
[191] A. Lidz, L. Hui, M. Zaldarriaga, & R. Scoccimarro, Astroph. J. 579 (2002) 491.
[192] A. Mesinger, S. R. Furlanetto, Mon. Not. R. Astron. Soc. 385 (2008) 1348.
[193] A. Mesinger, Mon. Not. R. Astron. Soc. 407 (2010) 1328.
[194] M. J. Rees, Mon. Not. R. Astron. Soc. 222 (1986) 27.
[195] G. Efstathiou, Mon. Not. R. Astron. Soc. 256 (1992) 43.
[196] A. A. Thoul, D. H. Weinberg, Astroph. J. 465 (1996) 608.
[197] T. Quinn, N. Katz, G. Efstathiou, Mon. Not. R. Astron. Soc. 278 (1996) L49.
[198] D. H. Weinberg, L. Hernquist, N. Katz, Astroph. J. 477 (1997) 8.
[199] J. Miralda-Escudé, M. J. Rees, Astroph. J. 497 (1998) 21.
[200] T. Kitayama, S. Ikeuchi, Astroph. J. 529 (2000) 615.
[201] J. F. Navarro, M. Steinmetz, Astroph. J. 538 (2000) 477.
[202] R. Barkana, A. Loeb, Astroph. J. 539 (2000) 20.
[203] J. S. B. Wyithe, A. Loeb, Nature, 441 (2006) 322.
[204] R. Barkana, A. Loeb, Mon. Not. R. Astron. Soc. 371 (2006) 395.
[205] A. Dekel, J. Silk, Astroph. J. 303 (1986) 39.
[206] J. S. B. Wyithe, A. Loeb, Mon. Not. R. Astron. Soc. 428 (2013) 2741.
[207] Z. Haiman, T. Abel, P. Madau, Astroph. J. 551 (2001) 599.
[208] R. Barkana, A. Loeb, Astroph. J. 578 (2002) 1.
[209] R. Barkana, A. Loeb, Astroph. J. 523 (1999) 54.
[210] P. R. Shapiro, I. T. Iliev, A. C. Raga, Mon. Not. R. Astron. Soc. 348 (2004) 753.
[211] B. Ciardi, E. Scannapieco, F. Stoehr, et al., Mon. Not. R. Astron. Soc. 366 (2006) 689.
[212] B. Yue, B. Ciardi, E. Scannapieco, X. Chen, Mon. Not. R. Astron. Soc. 398 (2009) 2122.
[213] L. N. Holzbauer, S. R. Furlanetto, Mon. Not. R. Astron. Soc. 419 (2012) 718.
[214] A. Fialkov, R. Barkana, A. Pinhas, & E. Visbal, Mon. Not. R. Astron. Soc. 437 (2014) L36.
[215] K. Ahn, P. R. Shapiro, I. T. Iliev, G. Mellema, U. Pen, Astroph. J. 695 (2009) 1430.
[216] M. E. Machacek, G. L. Bryan, T. Abel, Astroph. J. 548 (2001) 509.
[217] J. H. Wise, T. Abel, Astroph. J. 671 (2007) 1559.
[218] B. W. O'Shea, M. L. Norman, Astroph. J. 673, (2008) 14.
[219] E. Visbal, Z. Haiman, B. Terrazas, G. L. Bryan, R. Barkana Mon. Not. R. Astron. Soc. 445 (2014) 107.
[220] V. Bromm, R. P. Kudritzki, A. Loeb, Astroph. J. 552 (2001) 464.
[221] J. M. Shull, M. E. van Steenberg, Astroph. J., 298 (1985) 268.
[222] X. Chen, M. Kamionkowski, Phys. Rev. D 70 (2004) 043502.
[223] S. R. Furlanetto, S. J. Stoever, Mon. Not. R. Astron. Soc. 404 (2010) 1869.
[224] J. D. Silverman, P. J. Green, W. A. Barkhouse, et al., Astroph. J. 679 (2008) 118.
[225] L. Chuzhoy, P. R. Shapiro, Astroph. J. 655 (2007) 843.
[226] B. Ciardi, R. Salvaterra, T. Di Matteo, Mon. Not. R. Astron. Soc. 401 (2010) 2635.
[227] I. F. Mirabel, Proc. IAU Symp. 275, Jets at all Scales, Cambridge University Press, pages 2-8. ed. G. E. Romero, R. A. Sunyaev, T. Belloni [arXiv:1012.4944v1], 2010.
[228] T. Fragos, B. D. Lehmer, S. Naoz, A. Zezas, A. Basu-Zych, Astroph. J. 776 (2013) 31.
[229] I. F. Mirabel, M. Dijkstra, P. Laurent, A. Loeb, J. R. Pritchard, Astron. & Astroph. 528 (2011) A149.
[230] A. R. Basu-Zych, et al., Astroph. J. 762 (2013) 45.

[231] A. R. Basu-Zych, et al., Astroph. J. 774 (2013) 152.
[232] A. H. Patil, S. Zaroubi, E. Chapman, et al., Mon. Not. R. Astron. Soc. 443 (2014) 1113.
[233] P. Christian, A. Loeb, J. Cosmo. Astropart. Phys. 09 (2013) 014.
[234] A. Mesinger, A. Ferrara, D. S. Spiegel, Mon. Not. R. Astron. Soc. 431 (2013) 621.
[235] T. Fragos, et al., Astroph. J. 764 (2013) 41.
[236] J. E. McClintock, R. A. Remillard, In: Compact stellar X-ray sources. Edited by Walter Lewin & Michiel van der Klis, Cambridge Astrophysics Series, 157, 2006.
[237] M. Tamura, et al., Astroph. J. 753 (2012) 65.
[238] S. Mineo, M. Gilfanov, R. Sunyaev, Mon. Not. R. Astron. Soc. 426 (2012) 1870.
[239] S. P. Oh, Astroph. J. 553 (2001) 499.
[240] S. R. Furlanetto, A. Loeb, Astroph. J. 611 (2004) 642.
[241] P. R. Shapiro, K. Ahn, M. A. Alvarez, et al., Astroph. J. 646 (2006) 681.
[242] R. V. Vasudevan, R. F. Mushotzky, P. Gandhi, Astroph. J. 770 (2013) L37.
[243] R. Barkana, A. Loeb, Nature 421 (2003) 341.
[244] N. Lützgendorf, M. Kissler-Patig, N. Neumayer, et al., Astron. & Astroph. 555 (2013) A26.
[245] T. Tanaka, R. Perna, Z. Haiman, Mon. Not. R. Astron. Soc. 425 (2012) 2974.
[246] J. S. B. Wyithe, A. Loeb, Astroph. J. 595 (2003) 614.
[247] N. I. Shakura, R. A. Sunyaev, Astron. & Astroph. 24 (1973) 337.
[248] E. Scannapieco, R. Barkana, Astroph. J. 571 (2002) 585.
[249] R. Barkana, Mon. Not. R. Astron. Soc. 376 (2007) 1784.
[250] B. Greig, A. Mesinger, Mon. Not. R. Astron. Soc. 449 (2015) 4246.
[251] J. R. Pritchard, S. R. Furlanetto, Mon. Not. R. Astron. Soc. 367 (2006) 1057.
[252] L. Chuzhoy, Z. Zheng, Astroph. J. 670 (2007) 912.
[253] B. Semelin, F. Combes, S. Baek, Astron. & Astroph. 474 (2007) 365.
[254] S. Naoz, R. Barkana, Mon. Not. R. Astron. Soc. 385 (2008) 63.
[255] E. Visbal, R. Barkana, A. Fialkov, D. Tseliakhovich, C. M. Hirata, Nature 487 (2012) 70.
[256] A. Liu, M. Tegmark, Mon. Not. R. Astron. Soc. 419 (2012) 3491.
[257] F. Pacucci, A. Mesinger, S. Mineo, A. Ferrara, Mon. Not. R. Astron. Soc. 443 (2014) 678.
[258] A. Fialkov, R. Barkana, Mon. Not. R. Astron. Soc. 445 (2014) 213.
[259] S. Bharadwaj, S. S. Ali, Mon. Not. R. Astron. Soc. 356 (2005) 1519.
[260] S. Saiyad Ali, S. Bharadwaj, S. K. Pandey, Mon. Not. R. Astron. Soc. 366 (2006) 213.
[261] I. T. Iliev, G. Mellema, P. R. Shapiro, et al., Mon. Not. R. Astron. Soc. 423 (2012) 2222.
[262] C. A. Watkinson, J. R. Pritchard, Mon. Not. R. Astron. Soc. 443 (2014) 3090.
[263] R. Barkana, A. Loeb, Mon. Not. R. Astron. Soc. 384 (2008) 1069.
[264] V. Gluscevic, R. Barkana, Mon. Not. R. Astron. Soc. 408 (2010) 2373.
[265] S. Wyithe, M. Morales, Mon. Not. R. Astron. Soc. 379 (2007) 1647.
[266] G. J. A. Harker, et al., Mon. Not. R. Astron. Soc. 393 (2009) 1449.
[267] E. Calabrese, R. Hložek, N. Battaglia, et al., J. Cosmo. Astropart. Phys. 8 (2014) 10.
[268] A. Cooray, Phys. Rev. Lett. 97 (2006) 261301.
[269] W. Hu, R. Barkana, A. Gruzinov, Phys. Rev. Lett. 85 (2000) 1158.
[270] A. Fialkov, A. Loeb, J. Cosmo. Astropart. Phys. 11 (2013) 66.