Statistical detection of the EoR signal

Jouke Jensma

Contents

1. Introduction

The events preceding the Epoch of Reionization (EoR) show the necessity of having a reionization phase in the evolution of the Universe. The recombination epoch marks the stage in the evolution of the Universe where matter starts to dominate the evolution of the structure in the Universe. During recombination the free protons and electrons in the plasma of the photon-baryon fluid proceeded to recombine. Neutral hydrogen and helium started to form as soon as it became energetically favorable. Heavier elements are negligible. This recombination marked the start of the photon-baryon fluid decoupling. After decoupling at $z \approx 1100$, the Universe became transparent to photons, that we observe nowadays as the Cosmic Microwave Background (CMB). Without the influence of the photons baryons were free to fall into the potential wells of dark matter (DM) that were seeded by primordial inhomogeneities and continue to evolve with them.

As the Universe expands the background radiation cools and redshifts out of the visible part of the spectrum and the Universe becomes “dark”. This epoch is called the Dark Ages (DA). After recombination the Universe turns neutral and remains so until the first structures collapse to form the first radiation emitting astrophysical sources. This is the stage in which densities grow to high enough values so that they ignite the first stars and QSOs which proceed to reionize the Universe. As the Universe evolves further the number and radiation output of these sources increase until eventually the ionized regions around them percolate to fill the whole Intergalactic Medium (IGM) and heat it. The epoch in which the Universe undergoes the transition from neutral to ionized is called the Epoch of Reionization.

Given that hydrogen is by far the most abundant element in the Universe the level of ionization is determined by the value of its ionized fraction (x_{HII}). Initially x_{HII} is almost zero, but as the first sources turn on and start to ionize the Universe it gradually grows to 1. This reionization starts off with small ionized bubbles around the sources that continue to expand outwards as more of the surroundings get ionized. Eventually ionized bubbles merge with the bubbles of neighbouring sources until they fill the whole Universe.

When the Universe turned from neutral to ionized is constrained by observational evidence using the Lyman-α (Ly-α) forest. The Ly-α forest is the part of the spectrum from a source which shows jagged features. The dips in this spectrum are thought to be caused by absorption of the radiation from distant QSOs in the Ly-α line. Using observations of nearby QSOs the neutral fraction for the Universe at $z < 6$ is estimated to be about 10^{-4}, showing that the Universe at these redshifts is highly ionized. The main constraint on the end of the reionization process is given by the Ly-α-forest seen in the spectra of the high redshift Sloan Digital Sky Survey (SDSS) quasars (Fan et al., 2003, 2006). The key evidence is in the observation of the Gunn-Peterson troughs at $z > 6.3$ (Gunn and Peterson, 1965). This is interpreted as increased absorption by Ly-α absorbers. This is interpreted as evidence for an increase in the neutral fraction of hydrogen at $z > 6.3$, marking the tail end of the reionization process.

Another important constraint on reionization comes from CMB experiments that measure the Thomson scattering optical depth. This constraint is an integral constraint on the global electron density. This optical depth has been determined to be $\tau = 0.084 \pm 0.016$ (Dunkley et al., 2009). It attains its value when CMB photons scatter off residual electrons left over after recombination and off electrons that were freed during the EoR. Instantaneous reionization would have occurred at $z \sim 11$ to get this optical depth. However, instantaneous reionization is not physical and a more gradual process is more likely. This means that reionization could have occurred within an approximate window of redshifts in the range of $6.5 \lesssim z \lesssim 20$.

To ionize hydrogen an energy of 13.6 eV is required, which falls in the ultraviolet part of the radiation spectrum. The sources responsible for reionization must therefore be able to emit sufficient radiation with energies equal to 13.6 eV or above such that the Universe is fully reionized by $z = 6.5$. Possible candidates for reionization are the first or second generation of stars, miniqsos (powered by black holes) or even annihilating or decaying DM.

How the EoR exactly evolved is uncertain. Many different reionization scenarios are allowed within the current constraints. In addition to the constraints already mentioned a number of weaker ones exist, such as the Soft X-ray background (SXRB) measurement, which puts limits on reionization via miniqsos; the infrared background, this constrains the radiation spectrum of the first stars. Most of the weaker constraints are severely model-dependent, rendering them less useful. For the exact evolution of reionization to be known direct observations are required.

The EoR can be detected with the 21-cm line of neutral hydrogen. This hyperfine transition line of atomic hydrogen (in the ground state) arises due to interactions between the electron and the proton spins. The triplet state has electron and proton spin parallel with respect to each other, antiparallel for the singlet state. Since this is a forbidden transition the lifetime of the triplet state is extremely long: about 2.85×10^{15} seconds ($\sim$ 10 Myr). However, because of the abundance of neutral hydrogen in the Universe the 21-cm line is visible. For 21-cm radiation to be emitted the triplet state must constantly be pumped, several mechanisms exist that do this.

Collisions between particles can populate the triplet state. Collisions can occur between several particles, for example H-H (electron exchange causing the required spin flip), H-e (electron exchange or the collision causing a spin flip) or H-p (collisions). Each of these has its own efficiency in pumping the electrons to the upper level. Collisions are not bound by electronic transition rules to which radiative transitions are subject and can therefore excite an electron in the singlet state directly to the triplet state.

The Wouthuysen-Field effect is the radiative mixing of levels. A Ly-α photon can excite an electron from the singlet state to an excited state (see the solid lines in Figure 1 for allowed transitions). The electron can then either decay back to the same level or, more likely, decay to the triplet state. These transitions require an energy of 10.2 eV, i.e. a Ly-α photon. The cross section for this line is large, enhancing the probability of this transition to occur.

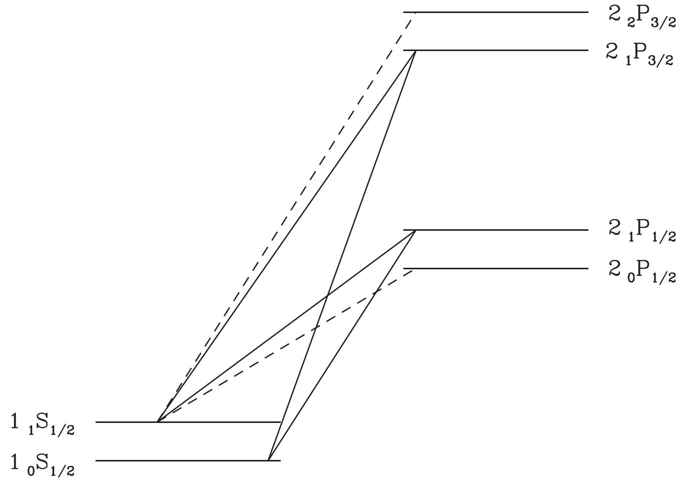


Fig. 1: Image adopted from Field (1958). 21-cm radiation is emitted when an electron decays from the $1\ {}_1S_{1/2}$ level to the $1\ {}_0S_{1/2}$ level. Solid lines denote electronic transitions which help mix the levels which is a prerequisite before 21-cm radiation can be emitted.

X-ray photons have no chance of mixing levels directly via the Wouthuysen-Field effect. Because of their high energies, any interaction they will have is ionization. However, it is possible for them to contribute to level mixing indirectly. Shull and van Steenberg (1985) have shown that secondary processes are very important for heating, exciting and ionizing hydrogen atoms in the IGM. X-ray photons have long mean free paths but when they interact with the IGM they deposit a large amount of energy in it.

Secondary processes occur when an X-ray photon ionizes an atom. A fraction of the energy is lost in ionization but the remainder is transferred to the electron. The interaction probability for this electron is much higher than the photons' and it quickly dissipates its energy by either heating the IGM, ionizing atomic hydrogen or exciting it. The secondary excitations mix levels via exciting electrons to the excited state or directly to the triplet state. Secondary ionizations cascade down and repeat the process, until the point where all the energy has been dissipated such that the remainder is transferred as heat.

The ground state of the 21-cm line is described by its spin temperature T_s. The spin temperature is defined in terms of the ratio of level populations of the hydrogen atom,

$$\frac{n_u}{n_l} = \frac{g_u}{g_l} e^{-T_*/T_s}. \tag{1}$$

Here n_u is the number of electrons in the triplet state and n_l the number of electrons in the singlet state. The g's denote the statistical weights of the levels with a ratio of 3/1 (triplet vs. singlet). $T_* = \frac{E_{10}}{k_B} = 0.0681$ K and denotes the excitation temperature of the 21-cm line. Since the spin temperature $T_s \gg T_*$ in astrophysical applications, there are ~3 times as many electrons in the upper level as in the lower level.

The type of astrophysical source influences T_s because they change level populations of neutral hydrogen. Field (1958) found that the spin temperature is given by

$$T_s^{-1} = \frac{T_{CMB}^{-1} + y_\alpha T_k^{-1} + y_c T_k^{-1}}{1 + y_\alpha + y_c}, \tag{2}$$

where the spin temperature T_s is a weighted mean of the CMB temperature and the kinetic temperature T_k. The coefficients y_α and y_c are due to Ly-α coupling and collisional coupling respectively. The y_α coupling term describes the Wouthuysen-Field effect and secondary collisional excitations and y_c describes collisions. In the absence of X-ray sources the contribution for collisions can often be ignored because the average density in the IGM is too low to be of importance.

In the absence of any other coupling mechanisms the spin temperature is coupled to the CMB temperature. In this case the 21-cm radiation is not visible (see equation 2).

Within the near future several telescopes are expected to measure the 21-cm line. These include Low Frequency Array (LOFAR), Murchison Wide-field Array (MWA) and Giant Metrewave Radio Telescope (GMRT). These will eventually be followed up by Square Kilometre Array (SKA), promising even higher resolution, sensitivity and frequency range. For the remainder of this report we will assume observations of LOFAR.

Radio telescopes observe brightness temperatures, which can be derived from the radiative transfer equation as

$$T_b = T_b(0)e^{-\tau_\nu} + T_s(1 - e^{-\tau_\nu}). \tag{3}$$

Here $T_b(0)$ is the temperature of the (black body) source. The apparent $T_b(0)$ decreases as the optical depth τ_ν increases. $e^{-\tau_\nu}$ is then the transmission fraction and consequently $(1 - e^{-\tau_\nu})$ is then the absorption fraction. T_s is the spin temperature of the medium through which the radiation propagates. T_b is the brightness temperature received. An interferometer (such as LOFAR) measures fluctuations in the brightness temperature, which is the temperature difference between the brightness temperature and the CMB temperature.

LOFAR will directly detect the neutral hydrogen during the EoR through the 21-cm transition. However, the noise per resolution element in these measurements is expected to be much higher than the 21-cm signal itself. Furthermore, the resolution will be on the order of 3 arcmin which will be insufficient to individually resolve the ionizing sources. Hence, the detection of the 21-cm signal with LOFAR will be of a statistical nature. The statistics should quantify how reionization proceeded, allowing for direct constraints on the evolution.

In order to help investigate the evolution of reionization a radiative transfer code was developed, called Bubble Expansion Around Radiative Sources (BEARS). Its purpose is to simulate cosmological EoR signals in a fast and cheap way. Using BEARS we obtain maps of the differential brightness temperature δT_b and of T_s. These maps are corrupted by incorporating the effects of limited resolution on it (beam smearing). Furthermore, foregrounds from the sky are added. These add noise to the maps, decreasing the signal to noise ratio. A corrupted 21-cm map corresponds to what LOFAR will observe.

The purpose of this project is to statistically characterize the EoR process when all the contributions to the signal are taken into account, including the evolution of the spin temperature. In order to do this we investigated methods to create mock 21-cm Probability Density Function (PDF)s in order to examine the influence of the component distributions of the neutral fraction, the cosmological density field and the spin temperature. While the analytical forms of the neutral fraction and the cosmological density field are known, this is not the case for the spin temperature distribution. Our aim was to parametrize the spin temperature contribution with a multi-parameter model. This parametrization could help us understand the evolution of the Ly-α photon distribution and of the kinetic temperature distribution. Aside from this we want to investigate if it is possible to extract higher order statistics from simulated data sets of the EoR signal that have been corrupted such that they are representative of what LOFAR will observe.

We have three important results to show. The first result shows that it is possible to create mock 21-cm PDFs using either direct integration using the analytical form of the distributions of the neutral fraction, the cosmological density field and the spin temperature contribution $Q = (T_s - T_{CMB})/T_s$, or via multiplication of random samples drawn from these component distributions. This can be used as a tool to investigate the influence of the distributions of the various components on the 21-cm signal.

Our second result is the discovery that the distribution of the parameter Q (defined as $Q = (T_s - T_{CMB})/T_s$) is an excellent tracer of early reionization physics. This distribution is obtained through the T_s simulations and has been parametrized with a multi-parameter model. The evolution of these parameters can be connected to the evolution of the Universes' early reionization phase. We show that Q traces mostly the IGM temperature evolution and the Ly-α photon distribution evolution.

Our third result is that it is possible to reconstruct moments from corrupted 21-cm maps. Accurate estimation of these moments allows us to accurately reconstruct higher order statistics such as the skewness and kurtosis of the 21-cm signal. These statistics can quantify how the EoR proceeds.

This report is structured as follows: in section 2 details of the simulation used to obtain the strength of the cosmological signal are described. We proceed with using these simulations in section 3 where the statistical properties of the signal are computed and analyzed. In section 3.3 the distribution of the Q component in the cosmological signal is discussed. The methods used to create mock 21-cm PDFs are discussed as well. The details of the fitting process of this distribution are described in section 3.7. The influence of LOFAR on the signal is discussed in 4. We conclude in section 5.

2. Simulations

The advantage of running simulations is that one knows exactly at each grid point (x, y, z) the values of the physical environment (such as the neutral fraction x_{HI}, the spin temperature T_s, the kinetic temperature T_k and the cosmological overdensity δ) and the cosmology for the simulation box ($h, \Omega_m, \Omega_b, \Omega_\Lambda, n_s$ and σ_8). This makes analyzing the details of the realization straightforward.

2.1. N-body simulation

The publicly available GADGET-2 code (see Springel et al., 2005) has been used to obtain a cosmological particle distribution. This simulation is the same as the one used in Thomas et al. (2009). The initial conditions have been generated with CMBFAST using glass-like initial conditions (Seljak and Zaldarriaga, 1996) and were linearly evolved down to $z = 127$ using the Zel'dovich approximation. The boxsize for the simulation was 100 h^{-1} comoving Mpc with 512^3 particles. This box has a DM particle mass resolution of 4.9×10^8 h^{-1} $M_\odot$. The simulated Universe is a flat ΛCDM Universe with $\Omega_m = 0.238, \Omega_b = 0.0418, \Omega_\Lambda = 0.762, \sigma_8 = 0.74, n_s = 0.951$ and $h = 0.73$. From this simulation we obtain 35 snapshots between redshifts 12 to 6. Halos were identified using a Friends-of-Friends algorithm as described in Davis et al. (1985), with a linking length of 0.2. The smallest halo mass that can be resolved is on the order of 10^{10} h^{-1} $M_\odot$. The particle distributions were converted to density fields via a Gaussian smoothing kernel with standard deviation $2 \times 100/512$ h^{-1} comoving Mpc.

2.2. Radiative post-processing with BEARS

The density fields and the halo catalogue are put into the BEARS code (Thomas et al., 2009; Thomas and Zaroubi, 2010). BEARS is a spherically symmetric 1D radiative transfer code. The free parameters allowed include the source spectrum, the source lifetime, redshift of turn-on and lastly, the clumping factor of the surrounding matter. Because of its high speed, using BEARS allows one to explore a wide variety of reionization scenarios quickly. From Thomas et al. (2009) we adopt the QSO and hybrid models. The QSO model has miniqsos as reionizing sources whereas the hybrid model is composed of a mixture of stars and QSOs.

BEARS places in each halo a radiation source. The number of ionizing photons emitted depends on the halo mass and is therefore assigned via a semi-analytical model. The radiative transfer equations are solved in 1D and these return profiles of the spin temperature T_s, the kinetic temperature T_k and the neutral fraction x_{HI}. These profiles are then embedded around the halos found in the halo catalogue. The differential brightness temperature δT_b can then be computed. Unlike most other codes, BEARS does not assume that $T_s \gg T_{CMB}$, but computes the spin temperature T_s selfconsistently. Especially in early reionization this is important since fluctuations in δT_b are not only driven by primordial density fluctuations but also by fluctuations in T_s.

2.3. The absorption and emission regime of the 21-cm PDF

Absorption and emission of the 21-cm line are two important concepts that frequently arise when discussing the 21-cm signal. It is also relatively poorly adressed in the literature as other authors have often assumed that $T_s \gg T_{CMB}$ in their simulations. The latter is a good assumption for mid-to-late reionization, but it also means that the 21-cm signal is always seen in emission, which is patently wrong at the early stage of the process.

It was only with the latest generation of radiative transfer codes that the spin temperature was computed self-consistently, allowing for signals visible in absorption with respect to the CMB (Baek et al., 2009; Pritchard and Furlanetto, 2007; Thomas et al., 2009; Thomas and Zaroubi, 2010). A region is defined to be visible in absorption when $\delta T_b < 0$. It is called visible in emission when $\delta T_b > 0$. From equation 4 it is obvious that absorption is possible only when $T_s < T_{CMB}$, and emission when $T_s > T_{CMB}$.

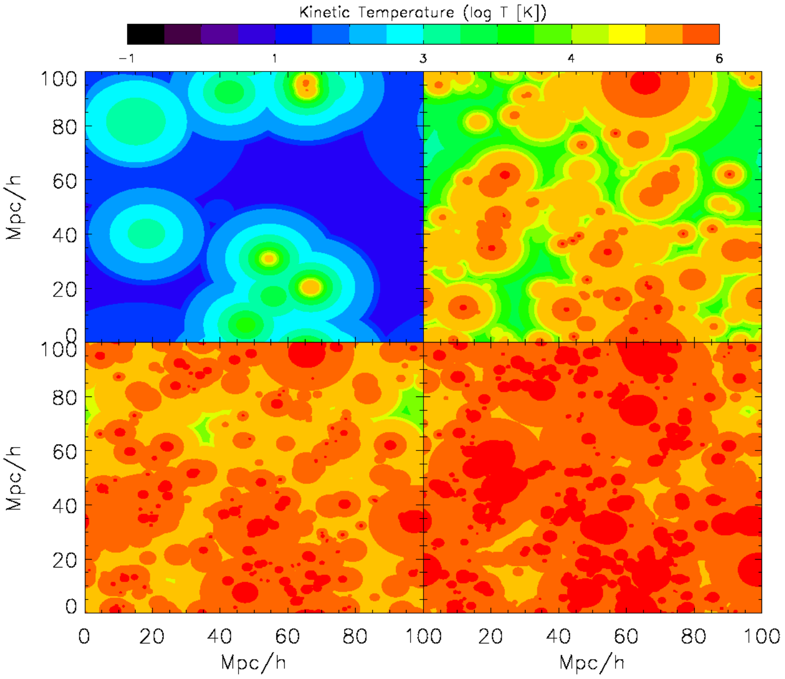


Fig. 2: Kinetic temperatures for the simulation of the QSO model. Slices are shown from redshifts 10, 8, 7, 6 from top left to bottom right. The Universe starts out cold but as photons propagate throughout it (and eventually get absorbed), it gradually heats up. Image courtesy of Thomas and Zaroubi (2010).

2.4. Simulation results from BEARS

Here we will simulate different reionization scenarios in order to explore their influence on the various statistics that we consider in this project. The reionizing sources chosen for the QSO model are miniqsos. The hybrid model is composed of a mixture of miniqsos and stars. These miniqsos have a power law energy spectrum ($\propto E^{-1}$), which has relatively more high energy photons than e.g. stars. For this reason they are thought to be extra efficient in reheating the Universe. The limits on the spectrum are from 200 eV up to 100 keV. The normalization is done such that they emit at 10% of the Eddington luminosity. Star populations are assigned luminosities using the population synthesis code of Bruzual and Charlot (2003). Furthermore, the stellar masses and ages were determined by the simulation. The mass distribution follows a Chabrier IMF.

The spin temperature can be coupled to the kinetic temperature T_k or the CMB temperature (see equation 2). These temperatures, combined with the coupling strength coefficients, determine the value of the spin temperature. The dominant coupling strength coefficient is the Ly-α coefficient describing secondary collisional excitations. This coefficient is $\propto J_0/T_k$ where J_0 is the local Ly-α flux density. The distribution of the local Ly-α flux density therefore influences the distribution of the spin temperature. Fluctuations in this flux density will give rise to fluctuations in the spin temperature and therefore in the 21-cm signal.

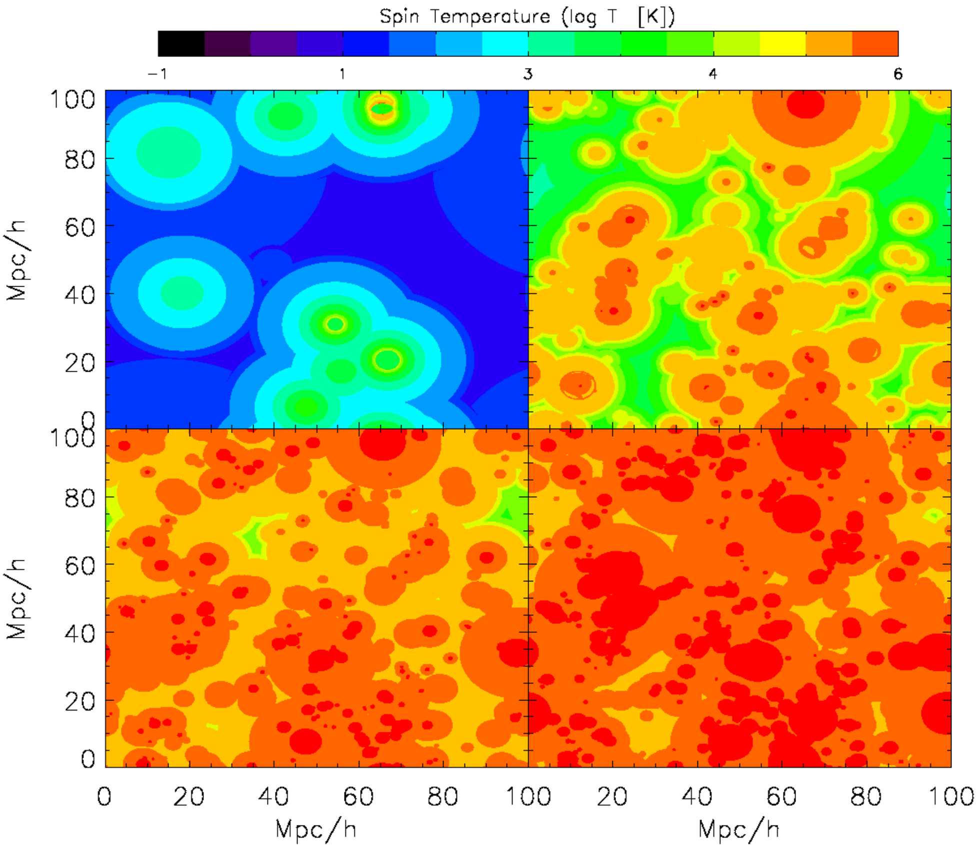
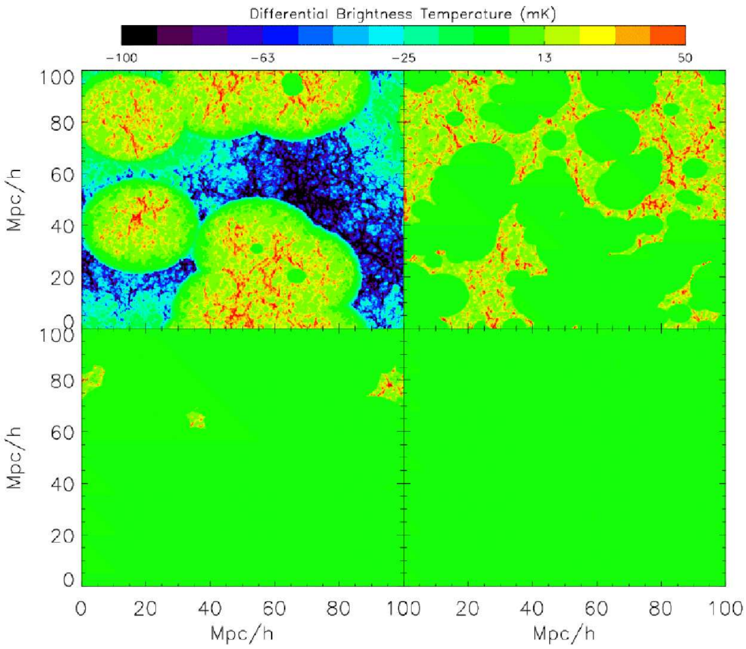


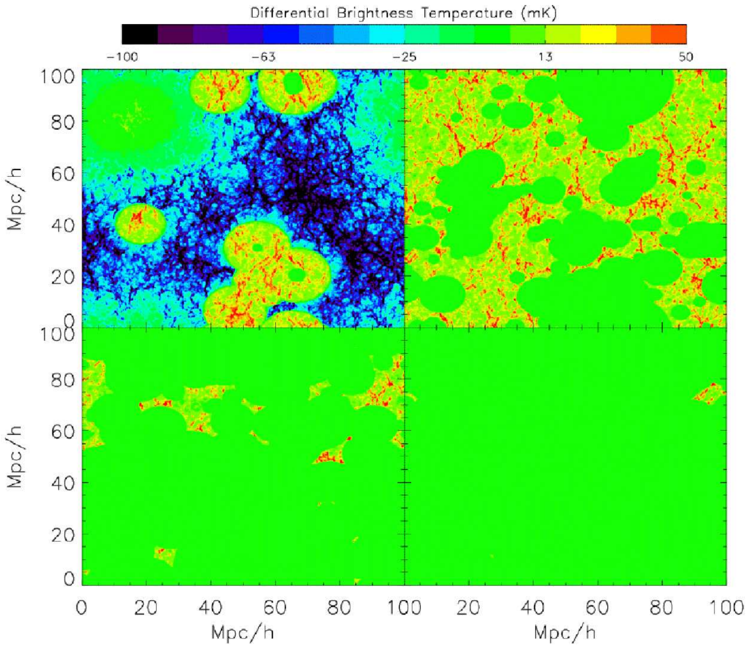
Fig. 3: Spin temperatures for the simulation of the QSO model. Slices are shown from redshifts 10, 8, 7, 6 from top left to bottom right. log T_{CMB} is 1.48, 1.39, 1.34 and 1.28 respectively. For spin temperatures lower than these values the Universe is visible in absorption, emission otherwise. z = 10 shows a Universe mostly visible in absorption, but has already turned to emission by z = 8. Image courtesy of Thomas and Zaroubi (2010).

Compare Figures 2 and 3 (the kinetic temperature map and the spin temperature map). There is no visible difference between the two images for most of the map. The spin temperature and the kinetic temperature are fully coupled, i.e. the radiation field intensity is sufficient to fully couple the temperatures. Conversely, if the radiation field intensity is too low then T_s will not be coupled to T_k. This will give rise to fluctuations in the spin temperature. These fluctuations are caused by the distribution of Ly-α flux density in the Universe, as well as the distribution of the kinetic temperature.

Figures 4a and 4b show that the Universe is partially visible in absorption at z = 10. This happens in the voids of the map. Hot bubbles have formed around the radiative sources where the Universe is seen in emission. By z = 8 and lower however the Universe is entirely visible in emission. This is mostly due to reheating of the Universe (compare with the map of the kinetic temperature in Figure 2). The differences between the QSO and hybrid model are minimal: the most pronounced features are the average size of the emission bubbles in the early Universe and the speed of the emission bubbles disappearing due to ionization. These are more spread out as there is relatively more energy in X-ray photons.



(a) Differential brightness temperatures for the simulation of the QSO model. Slices are shown from redshifts 10, 8, 7, 6 from top left to bottom right. The absorption phase at $z = 10$ is shortlived: the Universe is already mostly visible in emission by $z = 8$. Ongoing ionization of the Universe drives the differential brightness temperature to 0 as ionizing bubbles start to overlap (see $z = 8$ to 6). Image courtesy of Thomas and Zaroubi (2010).



(b) Differential brightness temperatures for the simulation of the hybrid model. Slices are shown from redshifts 10, 8, 7, 6 from top left to bottom right. Like the QSO model, the absorption phase at $z = 10$ is shortlived. The size of the emission bubbles is comparatively smaller. Ongoing ionization of the Universe drives the differential brightness temperature to 0 as ionizing bubbles start to overlap (see $z = 8$ to 6). Image courtesy of Thomas and Zaroubi (2010).

3. Statistics

The differential brightness temperature for an optically thin medium in an expanding Universe is written as

$$\delta T_b = 20\,\mathrm{mK}\,(1+\delta)\left(\frac{\mathrm{x_{HI}}}{\mathrm{h}}\right)\left(1-\frac{\mathrm{T_{CMB}}}{\mathrm{T_s}}\right)\left(\frac{\Omega_\mathrm{b}\mathrm{h}^2}{0.0223}\right)\left[\left(\frac{(1+\mathrm{z})}{10}\right)\left(\frac{0.24}{\Omega_\mathrm{m}}\right)\right]^{1/2}\left[\frac{\mathrm{H(z)}/(1+\mathrm{z})}{\mathrm{d}\mathbf{v}_{||}/\mathrm{d}\mathbf{r}_{||}}\right]. \tag{4}$$

The strength of the 21-cm signal depends partially on its environment. The neutral fraction x_{HI}, the cosmological matter density contrast δ, the spin temperature T_s and the peculiar velocity contribution $\mathrm{d}\mathbf{v}_{||}/\mathrm{d}\mathbf{r}_{||}$ are all quantities defined in 3D. Furthermore, the strength of the signal depends on the cosmology. Parameters such as Ω_b (baryonic content of the Universe) and Ω_m (matter content in the Universe) enter the equation as well. The distribution of samples of δT_b in a simulation box composes the 21-cm PDF. The shape of the PDF will be determined by the physical environment and the cosmology. For some examples of the 21-cm PDF, see Figures 4a and 4b.

3.1. The 21-cm PDF, its skewness and kurtosis

The 21-cm PDF is important because it characterizes the shape of the 21-cm signal. The evolution of the PDF may help constrain the evolution of reionization. Statistics (such as moments of samples of the 21-cm signal) characterize this evolution quantitatively. From these moments the skewness and kurtosis can be computed.

The second, third and fourth moments have been computed for the full data sets (containing 512^3 data points) and likewise for the limited data sets. The word limited refers to the fact that LOFAR observations do not yield $\sim 10^3 \times 512^3$ independent data samples. A 100 Mpc box was simulated, however, a $5° \times 5°$ field of view corresponds to a box of size $\sim$ 1 Gpc. Rather, at a given redshift the number of independent samples is limited by the resolution, field of view and frequency resolution of a telescope. If the 21-cm PDF does not evolve considerably with redshift one can bin in the frequency direction to increase the number of samples. This can be done safely as long as the “coherence” of the 21-cm signal is preserved along the frequency direction.

Since the full data set is much larger than the number of independent samples we compute the estimated moments 50 times via a Monte Carlo approach. The error contours in Figures 8, 9 and 10 represent the 2σ limits of the relevant quantities. This is done to estimate what the result of taking a limited number of samples is on the statistic.

The moments as estimated from the full data set are in reasonable agreement with the moments as estimated from the limited data set using the Monte Carlo approach. They do not agree perfectly (especially for the higher order moments) because of the influence of outliers. The difference between the fourth order moment for the limited and full data set is about 0.9%. The estimate of the moment from the Monte Carlo approach is taken to be the mean of the distribution of moments. Because we will only have access to a limited number of data samples after observations we choose to plot only those results of the estimation from the limited number of samples in Figures 8, 9 and 10. Likewise for the statistics shown in Figures 11 and 12.

Observations with LOFAR will yield a noisy data set. We assume the noise data to be known. These data sets will allow us to reconstruct the EoR signal. The noisy data set is computed by first taking the uncorrupted dataset containing the pure cosmological signal. This uncorrupted dataset is smoothed with a kernel that corresponds to the LOFAR resolution and then renormalized to preserve the original variance. The noise map (or noise data) is subject to the same process. The smoothed and renormalized noise is then added to this dataset (for details on adding the noise, see section 4.1). This yields the dataset which is called the noisy data. In principle interferometric effects will have to be taken into account as well.

Disentangling the components so the EoR data can be extracted is made easy by the additivity property of the cumulants κ of a distribution. This additivity property is that $\kappa_n(x + y) = \kappa_n(x) + \kappa_n(y)$ for $n \geq 2$. In all figures the estimate of the statistic from noisy data is plotted and compared with the statistic as determined from the "pure" signal case as a check on the results.

Computing the statistics

The skewness and kurtosis are used to quantify the evolution of the reionization process. These are composed of combinations of moments of samples. The estimation of the moments of samples from data will therefore be discussed first. These estimates of the moments are then used to compute the skewness and kurtosis. Accurate estimation of the moments is required before the skewness and kurtosis can be computed. The second moment is commonly called the variance. The variance is the simplest statistic for the problem of detecting the EoR signal.

Mathematically the sample variance is defined as

$$\begin{aligned} \sigma^2 &= E((X - \mu)^2) \\ &= \frac{1}{n} \sum_{i=1}^{n} (y_i - \langle y \rangle)^2. \end{aligned} \tag{5}$$

This is the biased sample variance but the number of samples n is always large enough such that the difference with the unbiased estimator is negligible. The y_i correspond to samples of a data set.

Since LOFAR is an interferometer (which measures only fluctuations), $\mu = \langle y \rangle = 0$ and the sample variance becomes the second raw moment μ_2. Furthermore, it has the property of being positive everywhere. It is also the easiest statistic to estimate from noisy data, because it is the lowest order meaningful statistic.

As explained in Jelić et al. (2008), the EoR signal is expected to be detected statistically as an excess variance over the noisy data. Making use of the fact that the distributions have zero mean, it is possible to derive how to compute the second moment of the EoR signal when the second moments of the noisy data and noise data are given. The excess variance is the second moment of the EoR signal:

$$\mu_2(\mathrm{EoR}) = \mu_2(\mathrm{noise} + \mathrm{EoR}) - \mu_2(\mathrm{noise}). \tag{6}$$

Here we have used that the second cumulant $\kappa_2 = \mu_2$. The third moment is important in estimating the skewness of a distribution. The third central moment is defined as

$$\mu_3 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \langle y \rangle)^3. \tag{7}$$

In a similar manner as above we can derive that

$$\mu_3(\mathrm{EoR}) = \mu_3(\mathrm{noise} + \mathrm{EoR}) - \mu_3(\mathrm{noise}), \tag{8}$$

where we have used that the third cumulant κ_3 equals μ_3.

The fourth moment is important for estimating the kurtosis of a distribution. Using cumulants and rewriting to central moments (using that the fourth cumulant $\kappa_4 = \mu_4 - 3\mu_2^2$) we can derive that

$$\mu_4(\mathrm{EoR}) = \mu_4(\mathrm{noise} + \mathrm{EoR}) - 3\mu_2(\mathrm{noise} + \mathrm{EoR})^2 - \mu_4(\mathrm{noise}) + 3\mu_2(\mathrm{noise})^2 + 3\mu_2(\mathrm{EoR})^2. \tag{9}$$

Note that this statistic requires the estimate of the variance of the EoR signal. Any uncertainties in the variance will propagate into the uncertainty of the fourth moment.

The main problem with computing the statistics however is that it is difficult to accurately extract higher order moments from noisy data. From the definition of moments ($\mu_n = \int x^n f(x)dx$) it can be seen that if at some point $x\,f(x)$ as obtained from observations (due to noise or other contaminating effects) deviates from the "pure" $f_{pure}(x)$ the estimate of the nth order statistic is off by $x^n(|f(x) - f_{pure}(x)|)$. This deviation gets larger for each order of the statistic. Later on this effect is seen in the size of the error contours on the estimates of the EoR signal from noisy data.

Now that the estimates of the moments of distributions are known the skewness and kurtosis can be computed. In estimating these statistics the mean of the estimate of the moments is used as the proper value. Error bars on the figures on skewness and kurtosis are computed with error analysis using the uncertainties on the moments (unlike the Monte Carlo approach used in estimating the moments).

Skewness is defined mathematically as

$$\gamma_1 = \frac{\mu_3}{\sigma^3}, \tag{10}$$

where μ_3 is the third central moment of a distribution and σ its standard deviation (the root of the second central moment).

It has the following interpretation: for a normal distribution, the skewness is zero. If the distribution is not normal and has outliers then it is said to have positive skewness if the outliers lie right of the mean, and negative otherwise. Skewness can also be thought of as a measure of symmetry of a distribution in units of σ^3.

The excess kurtosis is defined mathematically as

$$\gamma_1 = \frac{\mu_4}{\sigma^4} - 3, \tag{11}$$

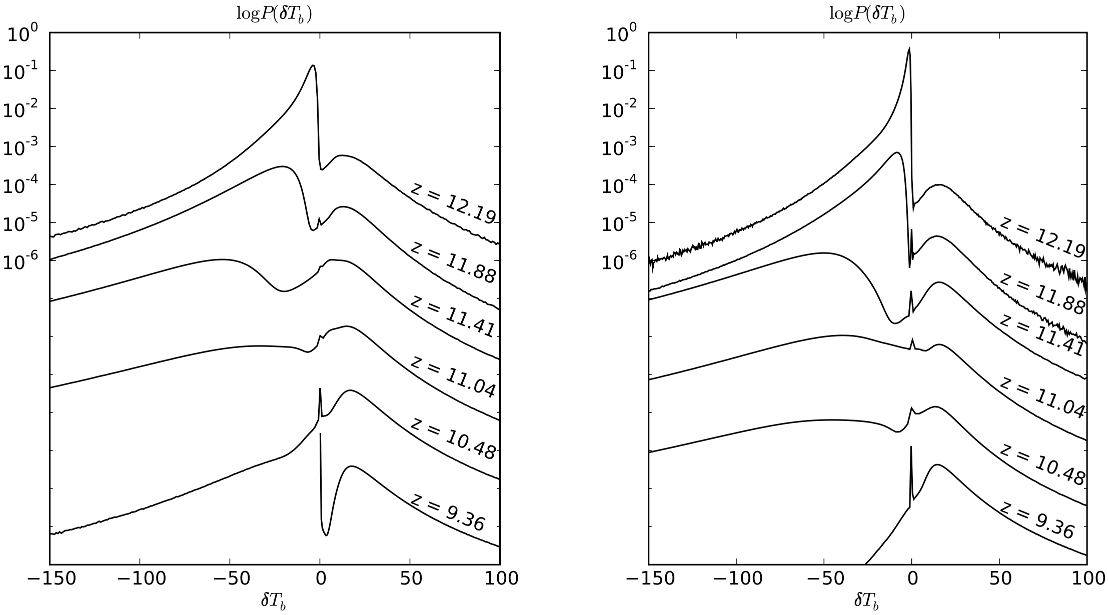
where μ_4 is the fourth central moment of a distribution and σ its standard deviation. For the normal distribution, $\sigma^2 = 1$ and the excess kurtosis has a value of 0. Unlike the standard definition of kurtosis, this definition can turn negative.

It has the following interpretation: for a normal distribution, the excess kurtosis has a value of zero. This is the standard "peakedness" of a distribution. If a distribution is more "peaked", i.e. higher kurtosis, then the peak is higher than that of the normal distribution but the distribution also has wider and thicker tails. Negative kurtosis has a more rounded peak and thinner tails. A high value for the kurtosis means that most of the variance of the PDF is due to extreme deviations (the PDF gets thicker and longer tails). Another way of thinking about the kurtosis (a fourth order statistic) is that it is a measure of fluctuations in the skewness (a third order statistic).

3.2. The 21-cm PDF from the simulations

Evolution of the 21-cm PDF for different reionization scenarios is always subject to the same physical principles. In global lines the evolution of the 21-cm PDF of the QSO and hybrid model will therefore be similar as well. For this reason we will use Figure 4a as the main example in discussing the evolution. The same physical principles can be applied to Figure 4b just as easily.

Most authors argue that in early reionization (and ignoring spin temperature fluctuations) the shape of the 21-cm PDF is expected to be Gaussian. This is because the Universe is largely neutral everywhere



(a) 21-cm PDF of the QSO model.

(b) 21-cm PDF of the Hybrid model.

Fig. 4: Each subsequent distribution in both panels has its value decreased with an additional factor of 100 for clarity of presentation. Lines from top to bottom display distributions for the redshifts 12.19, 11.88, 11.41, 11.04, 10.48 and 9.36.

and fluctuations in the PDF are then driven by primordial fluctuations in the density field δ. These latter fluctuations in δ are Gaussian in the linear regime. However, when fluctuations in the spin temperature are taken into account the resulting 21-cm PDF is no longer Gaussian. The 21-cm PDF shown in Figure 4a is clearly non-Gaussian (where the top line corresponds to the highest redshift, decreasing as you go down the lines). Note that the ordinates are in logarithmic scale.

In the top line of Figure 4a a roughly bimodal PDF is seen, one peak centered in absorption and the other in emission. If Ly-α photons exist to decouple T_s from T_{CMB} to T_k (which is lower than T_{CMB} at this time), an absorption signal arises. The amplitude of the absorption signal is mostly determined by the temperature of the IGM and the strength of the coupling of T_s to T_k. The absorption peak is much larger in amplitude than the emission peak by about two orders of magnitude in this early reionization phase. The Universe is mostly visible in absorption at $z = 12.19$.

The second line (at $z = 11.88$) shows the Universe with a relatively stronger absorption signal than emission signal. The absorption signals that exist have higher amplitude. This increase in amplitude is due to better coupling of T_s to T_k. T_k is still low because heating has not started yet because photons still need to percolate throughout the Universe. As more Ly-α photons come into existence the coupling strength increases and this has the effect of better coupling T_s to T_k, which will continue to increase the amplitude of the absorption signal for as long as the IGM is not significantly heated.

For the cold IGM (regions devoid of photons with the ability to heat the IGM) at these redshifts the matter temperature is on the order of a factor ~ 10 lower than the CMB temperature (see Couchman, 1985). This can enhance the amplitude of the δT_b signal 10-fold. The trend to more negative δT_b is a consequence of a steadily dropping spin temperature towards T_k. This is due to increasing Ly-α flux density.

At the same redshift we see a peak is start to appear around $\delta T_b = 0$. This is a consequence of the first ionized regions appearing in the Universe. The timescale for recombination in the IGM is large enough that when something gets ionized, it stays ionized. Compare with the lowest line in Figure 4a where a significant fraction of the points is ionized and stays at $\delta T_b = 0$. Once the Universe is ionized no further evolution of the PDF is expected.

For the late Universe (the lower lines in Figure 4a), the emission peak is mostly dominated by structures that are not that overdense that they have been ionized (because of the expected correlation between ionization fraction and high-density regions, ignoring recombination). These are the intermediate δ structures. The IGM has been heated to high temperatures and in this case $T_s \gg T_{CMB}$ and the 21-cm signal is independent of T_s. Summarizing: $\delta \gg 1, T_s \gg T_{CMB}$ and $x_{HI} = 1$. An intermediate δ structure will continue to collapse, thereby increasing the amplitude of its 21-cm signal, until the point where it becomes ionized and shifts back towards $\delta T_b = 0$.

3.3. Fitting the 21-cm PDF and the Q distribution

From equation 4 it can be seen that the key components of the 21-cm PDF (aside from the cosmology) are the cosmological overdensity field $(1 + \delta)$, the neutral fraction x_{HI} and the spin temperature contribution $Q = (T_s - T_{CMB})/T_s$. The differential brightness temperature is an observable (after removal of foregrounds and instrumental noise), the contribution of the cosmological overdensity field is in principle known and the ionized fraction can be extracted directly from the 21-cm maps, for example with the difference PDF (see Gluscevic and Barkana, 2010). The only unknown is the underlying distribution of the quantity Q. Q is an excellent tracer of the early physics of the reionization process. To illustrate that it is possible to reconstruct the 21-cm distribution from $(1 + \delta)$, x_{HI} and Q maps data from simulations is taken. These provide all required quantities.

Two methods of creating mock 21-cm PDFs have been investigated, the first one being direct integration of the component distributions of the neutral fraction, the cosmological density field and Q. The second method is straightforward multiplication of random sample values.

From direct integration

If the analytical form of all functions in equation 4 are known then the resulting 21-cm distribution can be computed via direct integration. Formally, the distribution $P(t)$ of the product of n distributions is defined as

$$P(t) = \iint P_1(x)P_2(y)\dots P_n(z)\delta(xy\dots z - t)dxdy\dots dz \tag{12}$$

The delta-function makes this a difficult integral to compute numerically. Therefore, for the case where the Universe is completely neutral the distribution of the neutral fraction is taken to be single-valued (1). This allows us to rewrite equation 12 as

$$P(\delta T_b) = \int \frac{1}{\delta} P_1(\delta)P_2(\delta T_b/\delta)d\delta \tag{13}$$

Here $P_1(\delta)$ is the PDF of the cosmological density field and $P_2(\delta T_b/\delta)$ is the contribution of the quantity Q. For $P_2(x)$ our multi-parameter model was used to compute the integral (see section 3.5). This function has limits set on x: $x_{\text{upp}} = 1$ is the upper limit for the case where $T_s \gg T_{CMB}$. The lower limit is set by the coldest temperature of the IGM, given by Couchman (1985). The expected temperature of the IGM without heating is thought to be $T_k(z) = 2.6 \times 10^{-2}(1+z)^2$. The theoretical lower limit is then $x_{\text{low}} = (2.6 \times 10^{-2}(1+z)^2 - T_{CMB,0}(1+z))/(2.6 \times 10^{-2}(1+z)^2)$. To see the feasibility of this approach see Figure 5, which shows the result of computing equation 13.

Note that the horizontal axis does not show δT_b but rather Q. This is a result of the variable of $P_2(x)$. In going from equation 12 to equation 13 the coordinate transformation $\delta Q = u$ was made, and via integrating over u we obtain $\delta T_b = u$. Since δ is mostly of order unity the algorithm uses most of the points where $u \simeq Q$.

From randomly chosen samples

Another method to obtain an estimate of the 21-cm distribution is via multiplication of randomly chosen samples. Suppose we have N samples of the quantities in equation 4, say, $x_{HI,1}, \dots, x_{HI,N}$ and likewise for the other quantities, we can construct a $\delta T_{b,1}$ by multiplying $(1+\delta)_1 x_{HI,1} Q_1$ with the remaining factors in equation 4. Doing this for all N samples yields N samples for δT_b. In taking these samples we take care that the combination of samples are representative of their distributions. A comparison of this method with the data taken from the simulation is shown in Figure 6. Some deviations with the data are expected since the variables are not uncorrelated. For example, for the highest density regions one expects to find relatively high Q-factors because of relatively higher T_k. Regardless, the fact that the sample multiplication works well means that the correlation is rather weak.

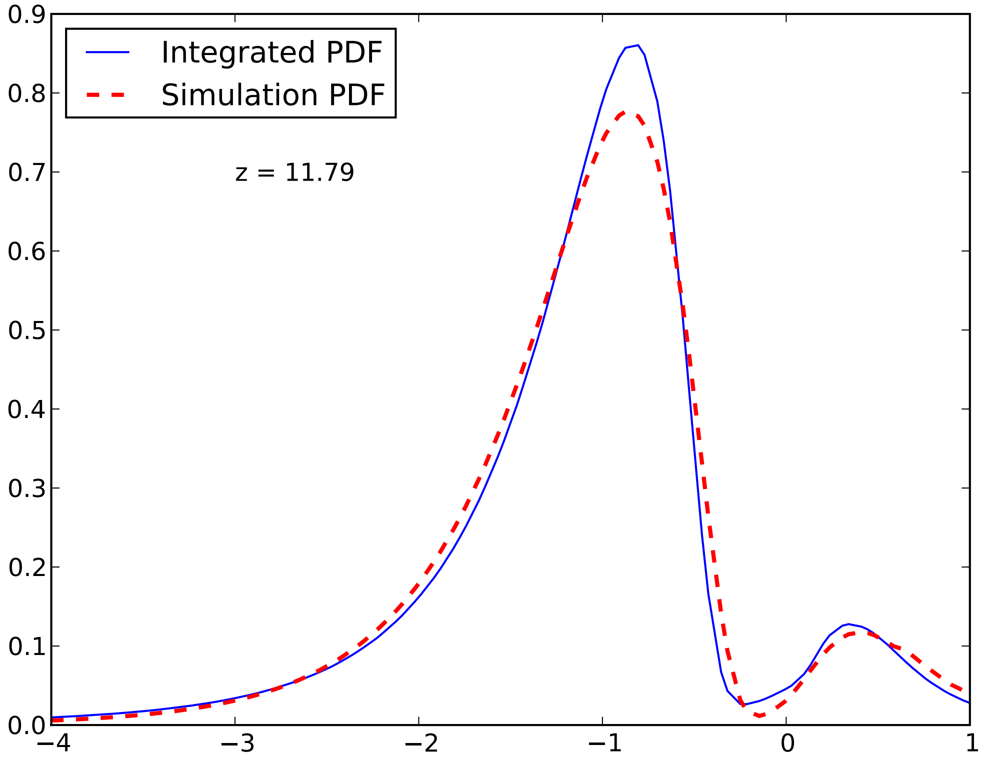


Fig. 5: Direct integration of distributions to compute the 21-cm distribution, using equation 13.

From simulation snapshots

Using the simulation snapshots of δ, x_{HI} and T_s we can compute snapshots of δT_b using equation 4. These computed snapshots contain samples of δT_b values and this allows us to compute the PDF of δT_b directly by creating a histogram. In this section we will use the results from the QSO simulation. This PDF is uncorrupted by the problems that the real data will suffer from. Instrumental noise, Radio Frequency Interference (RFI), beam smearing, incomplete calibration will all degrade the signal-to-noise ratio of the cosmological signal.

The 21-cm PDF is essentially a product of 3 separate distributions, the Gaussian / lognormal distribution of the density field (depending on the amount of nonlinearity in structure formation), the bimodal distribution of the neutral fraction x_{HI} (in the BEARS code) and lastly, the distribution of $Q = (T_s - T_{CMB})/T_s$.

The quantity $Q = (T_s - T_{CMB})/T_s$ is an excellent tracer of decoupling mechanisms in the early stages of reionization, as well as tracing the first heating of the IGM. What types of reionizing sources there exist in the Universe might be constrained via estimating the timescales required to heat the cold IGM to above the T_{CMB} temperature.

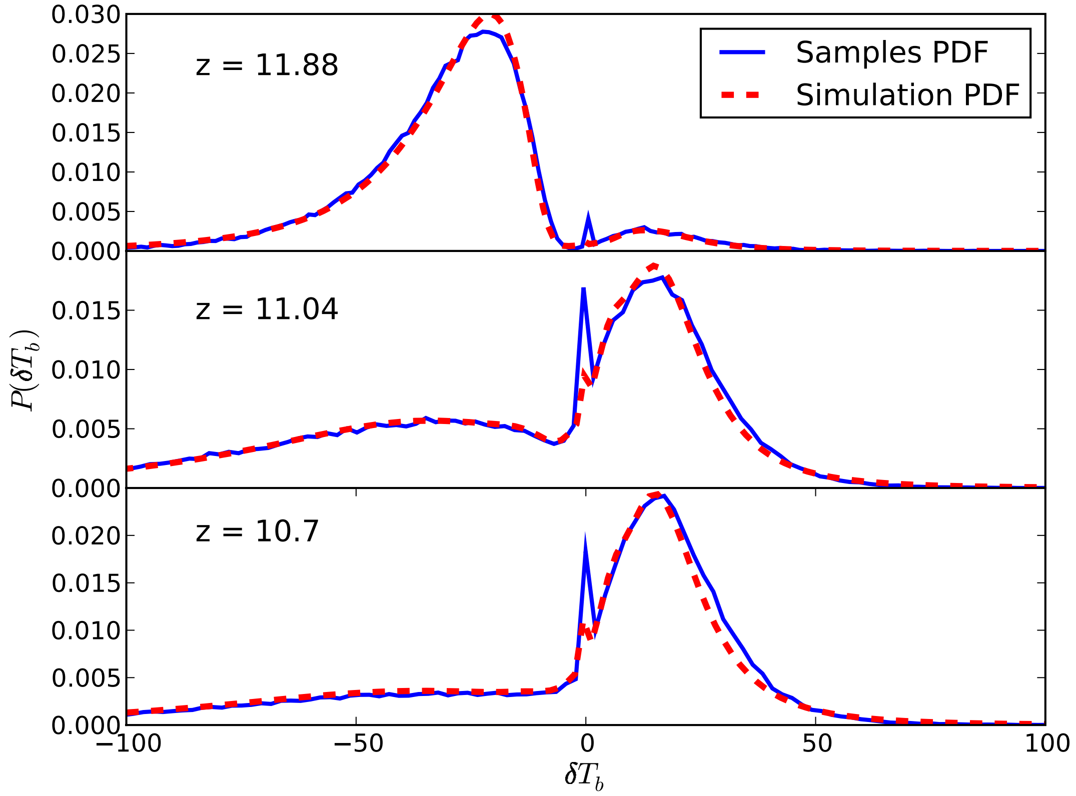


Fig. 6: The solid blue line is the result of the multiplication of samples. The dashed red line are samples obtained from the simulation.

The distribution of the spin temperature itself was examined at first, but it has a shape unsuitable for investigating the physics. One of the reasons is that the domain of the spin temperature is essentially $(0, \infty)$, whereas for Q this is (strictly speaking) $(-\infty, 1]$. However, in practice the lower limit of Q is bounded by the lower limits set on T_k by Couchman (1985) (the theoretical minimum value of Q is on the order of -10).

It turns out that the shape of Q has very distinct properties in the onset of reionization (for some examples, see Figure 7). Therefore we set out to model it. For instance, if T_s is fully coupled to T_k (which it almost always is, except perhaps for very early reionization) then Q is a measure of the temperature of the IGM. If T_s is not fully coupled to T_k then Q is a measure of the first Ly-α photons propagating throughout the Universe.

For several redshifts the distribution of Q points is shown in Figure 7. A rich structure is seen in the distribution. Points where $Q < 0$ correspond to regions where $\delta T_b < 0$. Conversely, points where $Q > 0$ correspond to regions where $\delta T_b > 0$.

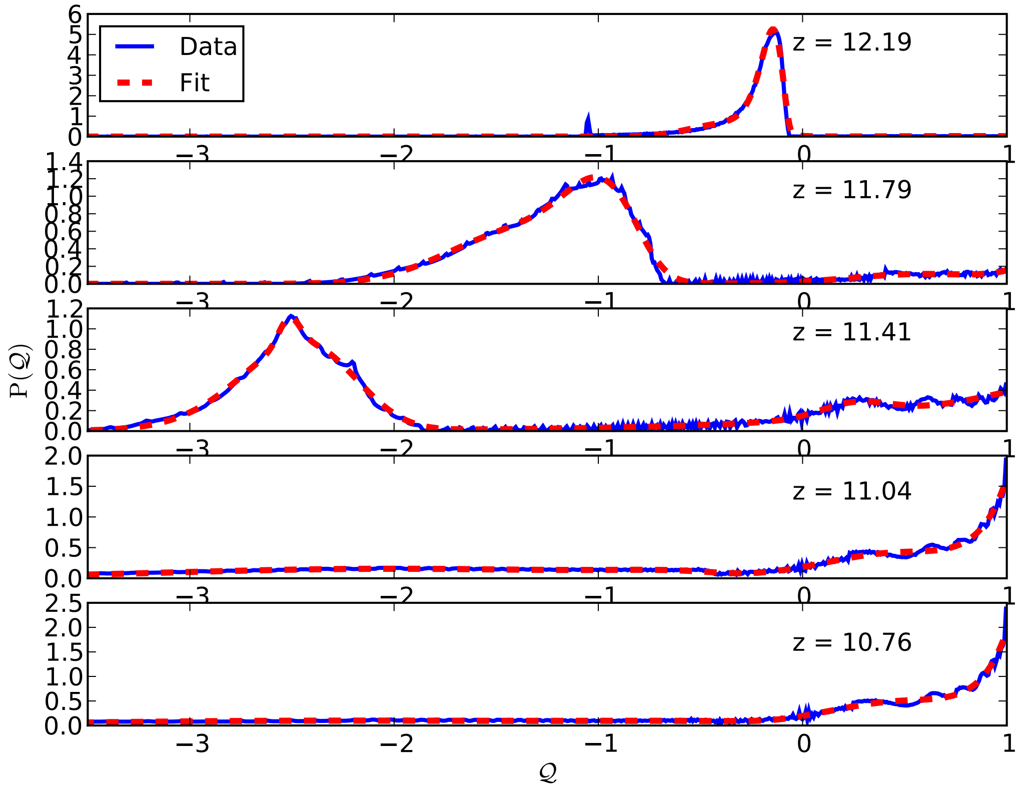


Fig. 7: From top to bottom: distribution of Q for the QSO model for a range of redshifts. The dashed red line is the fit from equation (16). The solid blue line is the simulated data.

Globally, from high redshift to low redshift, points shift towards lower Q, corresponding to a cooling down of the Universe due to its expansion. After a short time this trend is reversed and points shift towards higher Q (the onset of heating), to finally end with all points being at $Q = 1$ (all points sufficiently hot such that $T_s \gg T_{CMB}$).

The physical significance of the Q-distribution is as follows: at the highest redshifts, in the DA, no sources exist and in the absence of coupling mechanisms the spin temperature is coupled to T_{CMB}. As the first sources form, photons start to come into existence and these decouple T_s from T_{CMB} to T_k. Even though the energy requirements for this process are very low (because of the small energy level spacing in the ground state of neutral hydrogen) it does not proceed instantanously. Depending on the amount of sources, their distribution and the source properties it can take several hundred Myr to fully couple T_s to T_k in the entire Universe.

The IGM in the DA has T_k much lower than T_{CMB} and continues to grow colder because of the ongoing expansion of the Universe. The shift of Q to lower values reflects this process. Because the energy requirements for heating are much higher than those for Ly-α level mixing the timescale on which the

IGM gets reheated is much larger as well. This means there will be a phase in reionization in which the spin temperature will be fully coupled to the cold gas. This gives rise to a Universe mostly visible in absorption with $Q < 0$ and thus with $\delta T_b < 0$.

In a slightly later stage points start to move from their lowest value of Q towards larger values of Q. This is a result of heating processes that heat up the IGM. As heating is driving T_k towards T_{CMB} and beyond the value of Q will increase (and eventually change sign). In the later stages of reionization the IGM will be much hotter than T_{CMB} and with an abundance of Ly-α coupling photons T_s will generally exceed T_{CMB} everywhere, causing Q to be 1 everywhere ($Q = 1$ when $T_s \gg T_{CMB}$).

Note that changes at low Q can happen much faster than those at high Q: the temperature difference between two points at Q_2 and Q_1 is: $\Delta T_s = T_{CMB}\left(\frac{1}{1-Q_2} - \frac{1}{1-Q_1}\right)$. Here $Q_2 > Q_1$. As Q_2 approaches 1, the temperature difference approaches infinity. A small energy input in the absorption regime is sufficient to push points into the emission regime. This is seen in the third and fourth panel of Figure 7: the large peak in absorption in the third panel has almost vanished in the fourth panel, even though the redshift difference between the two is small. The progression of points in the emission regime from $Q = 0$ towards $Q = 1$ takes much more time.

3.4. Skewness and kurtosis results

The definitions of the skewness and kurtosis require that the moments of the distributions are known. The results from the estimation of the moments are therefore presented first, followed by the results of the estimation of the skewness and kurtosis.

Estimating the second moment

The results of the estimation can be seen in Figure 8 for both models. This Figure has been obtained by using equation 6. The top panel corresponds to the QSO model. The bottom panel corresponds to the hybrid model. The estimate of the EoR signal agrees well with the true EoR signal. The error contour is very thin, reflecting the fact that the method is precise. However, sometimes the true value lies outside the contour. This underlines the importance that a small set of strong outliers can have on estimation of the statistic: when they are not included the estimate of the EoR signal does not agree with the real value. Repeated estimation of the EoR signal with a Monte Carlo method does not yield accurate standard deviations on the estimate. In this case the estimate is wrong for $z < 6.5$ for the QSO model. This is because the Universe is strongly ionized by now, but with a few outliers (which represent points that have not ionized yet).

Estimating the third moment

The results of the estimation can be seen in Figure 9 for both models. The third moment has been estimated using equation 8. The method succeeds rather well in estimating the global features of the third moment. However, the third moment suffers from the same problem as the second moment does: some outliers have such strong influence that the estimation does not always coincide with the true value. A nice example of how difficult it is to estimate moments accurately is the third moment of the noise. From the theory we expect this to be zero (the noise is modelled by a Gaussian which is symmetric and therefore the third moment is expected to be zero). The third moment is centered around zero but the estimate is not as accurate as it could be.

Estimating the fourth moment

The results of the estimation can be seen in Figure 10 for both models. The fourth moment has been estimated using equation 9. The estimate of the fourth moment is the most accurate in the redshift range $11.8 \gtrsim z \gtrsim 10.2$ for the QSO model. For the hybrid model this range is $11.3 \gtrsim z \gtrsim 9.5$. This is because the fourth moment of the EoR and noise is stronger than that of the noise. Another problem is that the estimate of the EoR signal is dependent on the estimate of the fourth moment of distributions of the EoR signal + noise and noise, which are difficult to estimate. For the remaining redshift ranges the estimate is again precise but inaccurate. However, it is able to capture global features of the signal.

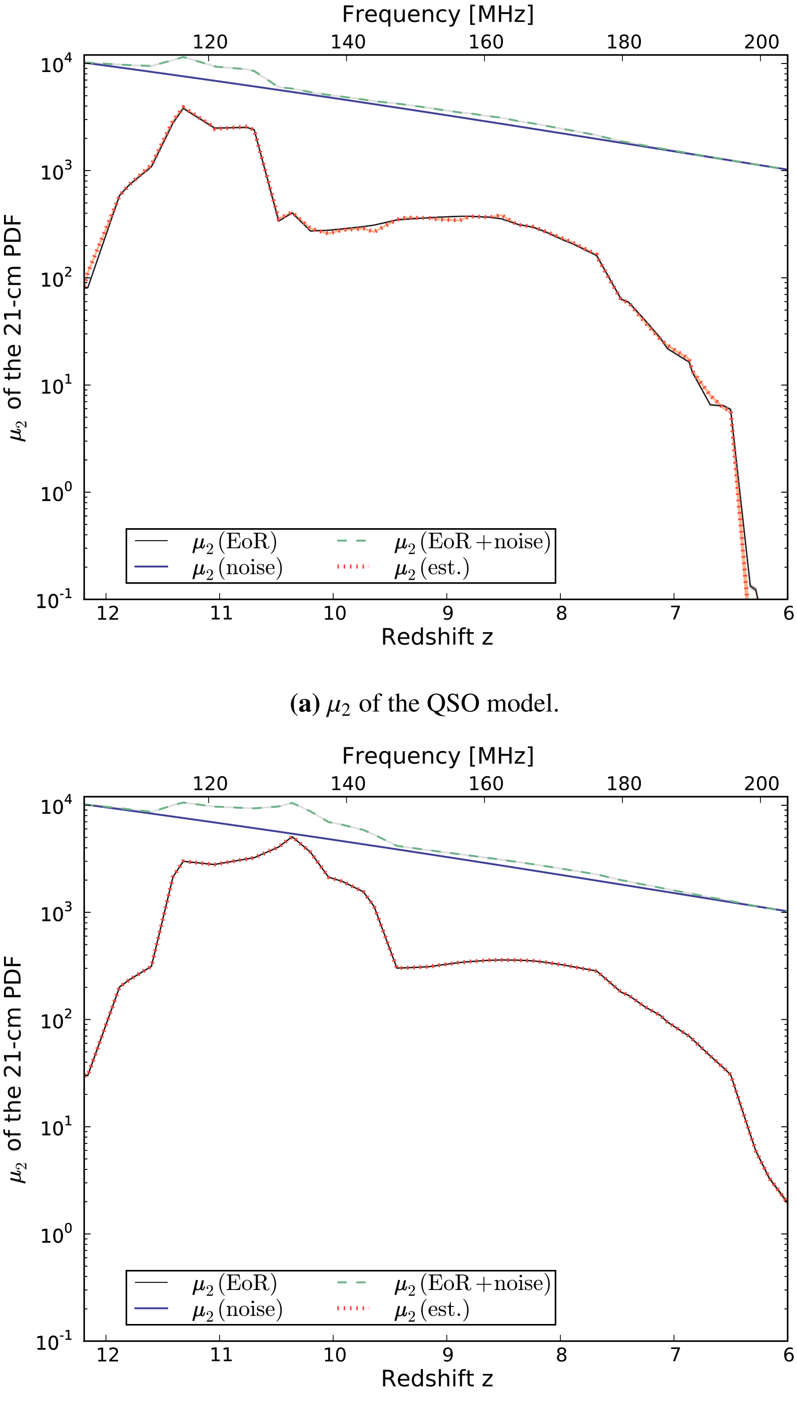
The skewness for the 21-cm PDF is shown in Figure 11 for both models. However, as the evolution of the skewness is similar for both models (as they are subject to the same physical principles) we choose to discuss only the QSO model. Six stages in the evolution of the skewness can be distinguished in this image. The transition between these six can be connected to several stages of the reionization process (see Table 1).

Stage A (z = 12.2 – 11.8)	The skewness is negative in this regime and grows to positive values. Outliers lie mostly to the left of the mean. The increase is *not* interpreted as outliers shifting more towards the mean or to the right of the mean, instead the mean is shifting farther left (compare with Figure 4a). The mean shifting to more negative values of δT_b is because at $z = 12.2$ T_s is not fully coupled to T_k, instead it is somewhere inbetween T_k and T_{CMB}. As the coupling grows stronger with redshift the spin temperature approaches T_k and eventually couples, enhancing the strength of the absorption signal. This causes the mean to shift towards more negative values.
Stage B (z = 11.8 – 11.4)	The skewness is positive in this regime and stays approximately constant over this redshift range. The PDF in Figure 4a however does change considerably. In this redshift regime the emission peak at positive δT_b values is gaining in strength while the center of the absorption peak is shifting towards increasingly negative values. The skewness being negative indicates that the absorption signal is still significant in the Universe. The first traces of ionized regions start to appear, judging from the rising peak at $\delta T_b = 0$. That the skewness stays constant is a coincidence of the mean of the distribution shifting left (which would increase the skewness) in such a way that it compensates for the outliers (which stay predominantly left of the mean). The underlying cause of this is points still in the process of coupling T_s to T_k.
Stage C (z = 11.4 – 10.4)	Stage C sees a decrease in the skewness in this redshift range. From Figure 4a we see that the Universe is now mostly visible in emission. The mean of the PDF is rapidly shifting towards positive values but does not become positive yet. This shift to positive values is because the gas is continually being heated and since T_s is coupled to T_k the average differential brightness temperature rises. The shift to the right is the cause of the skewness becoming more negative, there are now essentially more points left of the mean.

Stage D (z = 10.4 − 10.2)	This redshift regime is interesting because the skewness undergoes a transition from negative values to positive values. After $z = 10.7$ the mean of the PDF makes the transition from negative to positive values. Most of the gas in the Universe now has temperatures above the CMB temperature. In this redshift regime the absorption peak is sufficiently weak that emission is now the dominant component in the Universe. So clearly this regime is the transition between an absorption- and an emission-dominated Universe. The time between $z = 12.2$ and $z = 10.7$ is approximately 115 Myr. This is the timescale required for the first photons to permeate the Universe and heat it to above CMB temperatures. The increase of the emission peak is the cause of the skewness becoming positive.
Stage E (z = 10.2 − 8.5)	In this regime the skewness stays roughly constant. The mean of the PDF is slowly shifting to the left. This effect alone would increase the skewness, however, as ionization of the Universe proceeds more points in the PDF shift towards $\delta T_b = 0$, balancing this increase. The net effect is that the skewness stays approximately constant.
Stage F (z = 8.5 − 6.0)	The mean is rapidly converging to $\delta T_b = 0$. As reionization proceeds points shift towards $\delta T_b = 0$. The effect of the mean converging and points shifting left would lead to an expected decrease in skewness. However, there are a few points in the Universe with high δT_b that function as "extreme outliers", causing a sharp rise in the skewness.

However, the global trend is a rise from negative values to positive values, with a flat plateau inbetween $8.4 \lesssim z \lesssim 10.2$, and a rise afterwards. The first rise is due to the ongoing heating of the Universe, moving from an absorption-dominated phase to an emission-dominated phase. The plateau is the result of the following effect: δT_b points are shifting towards $\delta T_b = 0$ because of ongoing ionization in the Universe , which would increase the skewness. However, the overall distribution of δT_b points is shifting such that the mean moves to the left as well so that the end result is a constant skewness. For $z < 8.4$ a sharp rise in the skewness is seen. The mean of the distribution is very near $\delta T_b = 0$ with the remaining non-ionized points in the Universe causing positive skew.

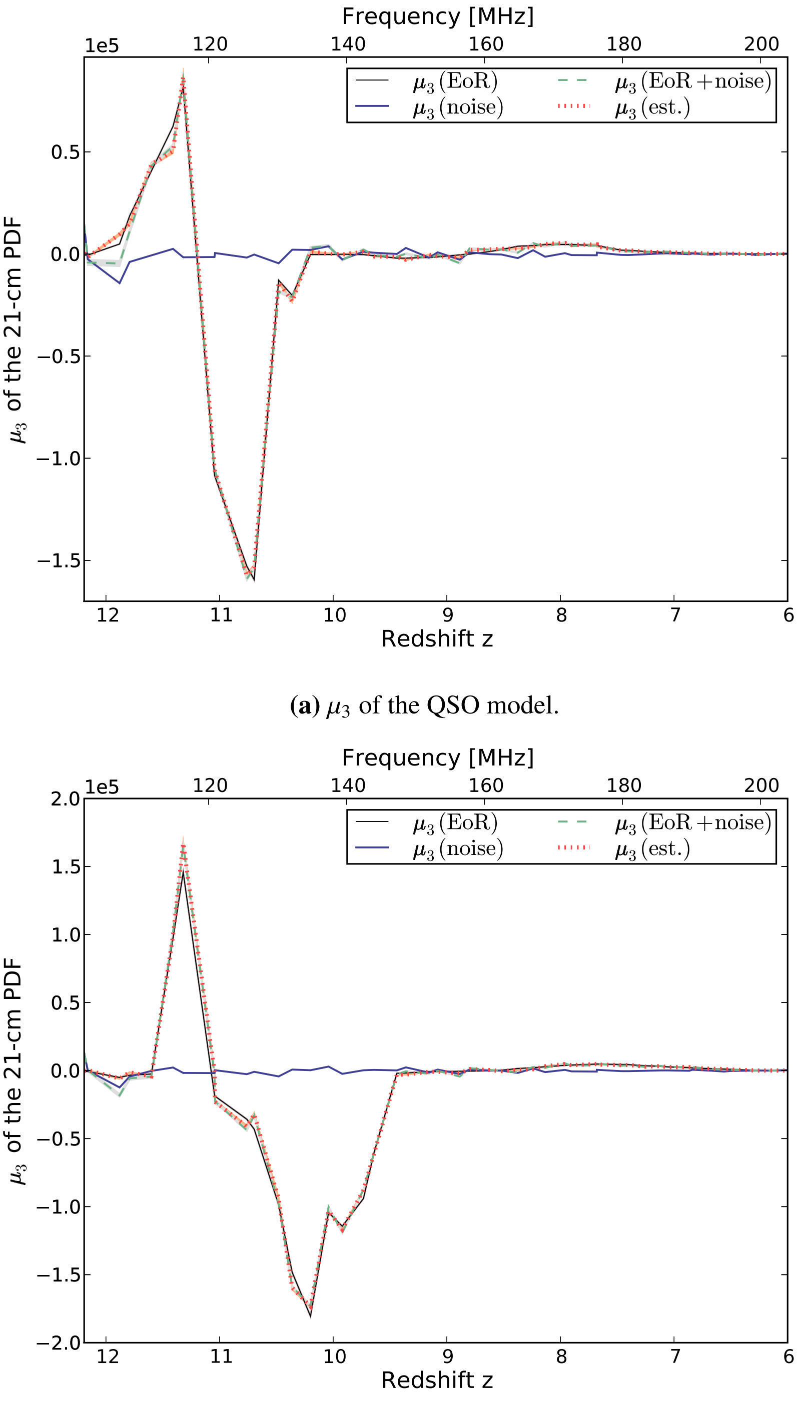
Estimating skewness from noisy data is considerably more difficult: it is a higher order statistic and uncertainties in estimating higher order moments complicate the process. In theory the problem is straight-



(a) μ_2 of the QSO model.

(b) μ_2 of the hybrid model.

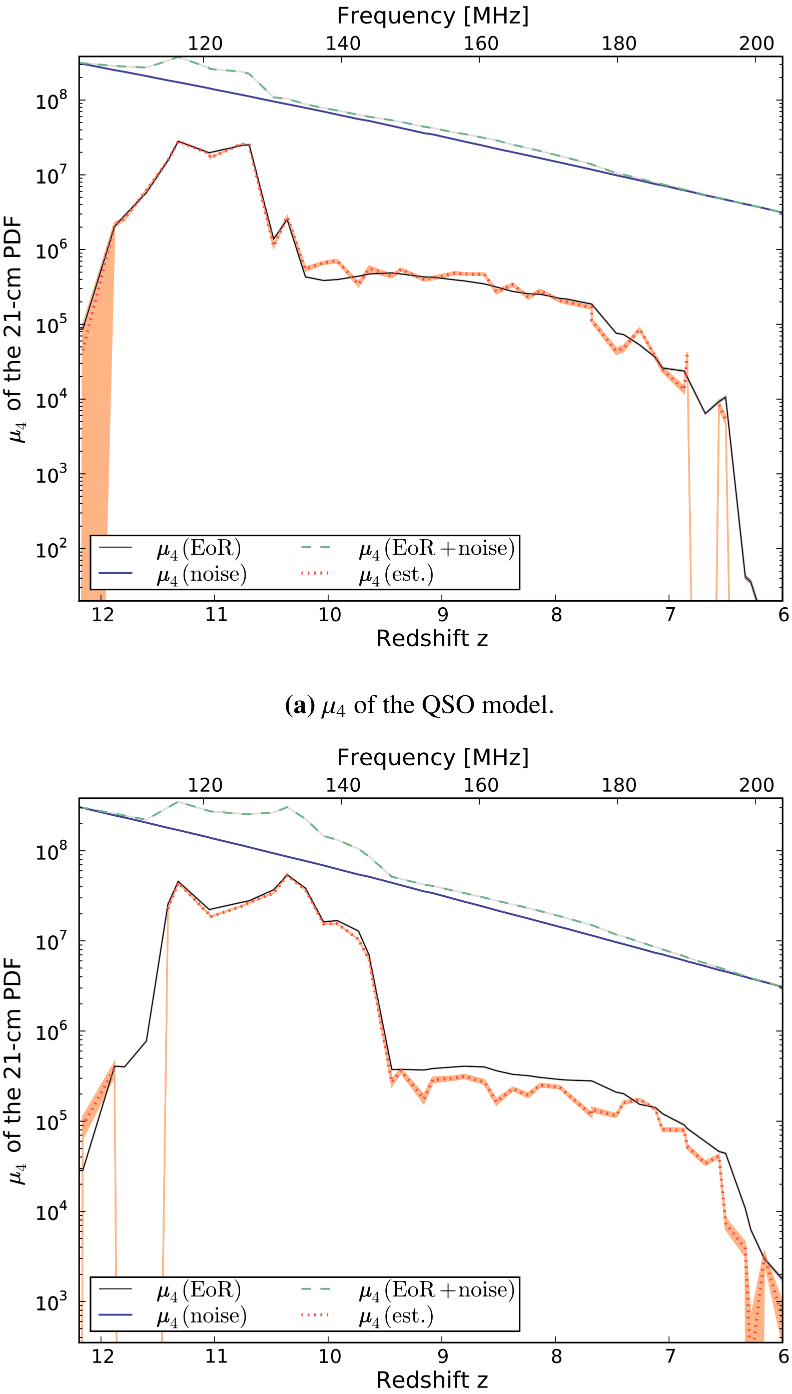
Fig. 8: The solid black line is μ_2(EoR). The solid blue line is μ_2(noise). The dashed green line is μ_2(EoR + noise). The dotted red line is the estimate of μ_2 from the noisy data as given by equation 6. The contour areas represent the 2σ domains of the estimation of the values from the limited number of samples. The light-grey shaded area represents the range of μ_2 values for the noise. The darker grey shaded area represents the range of μ_2 values of the EoR signal. The orange area represents the range of μ_2 values of the estimation of the EoR signal.



(a) μ_3 of the QSO model.

(b) μ_3 of the hybrid model.

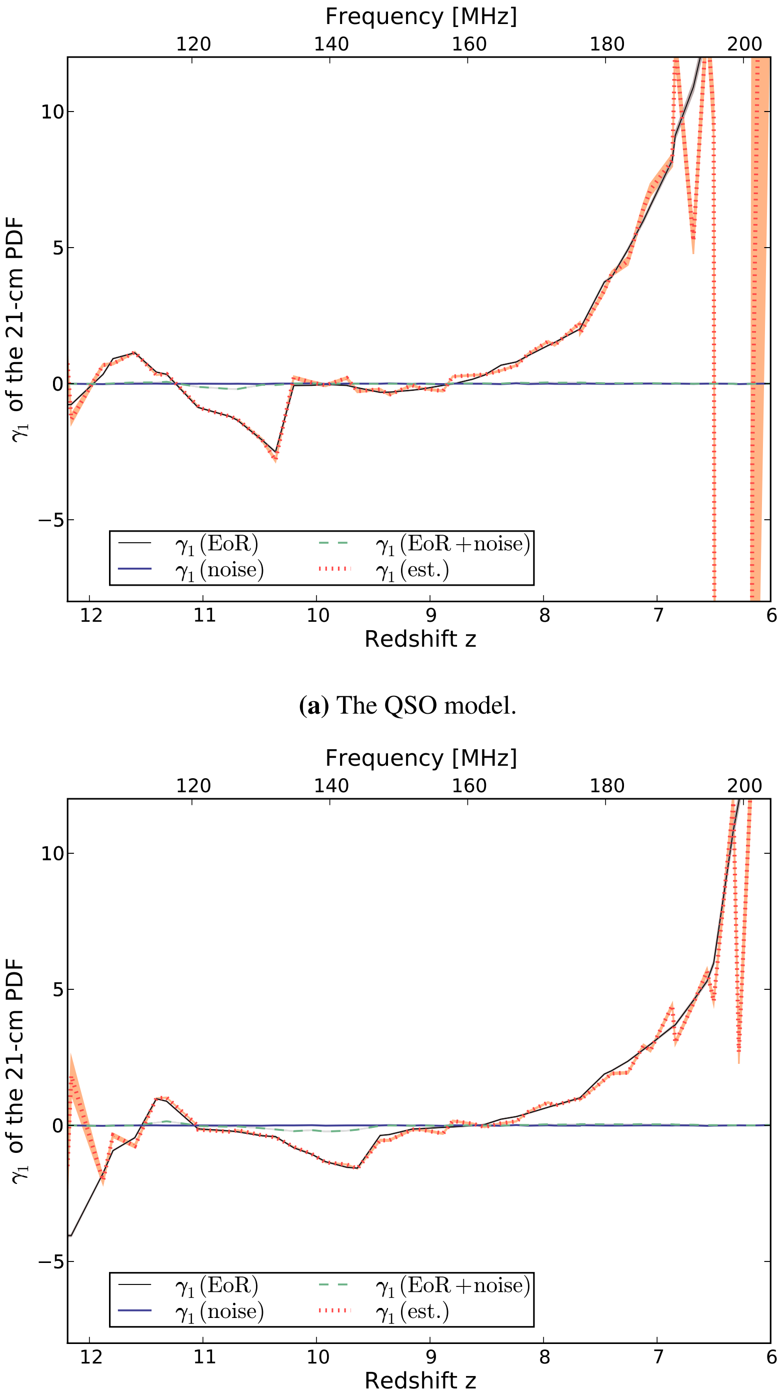
Fig. 9: The solid black line is μ_3(EoR). The solid blue line is μ_3(noise). The dashed green line is μ_3(EoR + noise). The dotted red line is the estimate of μ_3 from the noisy data as given by equation 8. The contour areas represent the 2σ domains of the estimation of the values from the limited number of samples. The light-grey shaded area represents the range of μ_3 values for the noise. The darker grey shaded area represents the range of μ_3 values of the EoR signal. The orange area represents the range of μ_3 values of the estimation of the EoR signal. Note that the ordinate is in units of 10^5.



(a) μ_4 of the QSO model.

(b) μ_4 of the hybrid model.

Fig. 10: The solid black line is μ_4(EoR). The solid blue line is μ_4(noise). The dashed green line is μ_4(EoR + noise). The dotted red line is the estimate of μ_4 from the noisy data as given by equation 9. The contour areas represent the 2σ domains of the estimation of the values from the limited number of samples. The light-gray shaded area represents the range of μ_4 values for the noise. The darker grey shaded area represents the range of μ_4 values of the EoR signal. The orange area represents the range of μ_4 values of the estimation of the EoR signal.



(a) The QSO model.

(b) The hybrid model.

Fig. 11: Skewness (γ_1) of the δT_b distribution. The solid black line is γ_1(EoR). The solid blue line is γ_1(noise). The dashed green line is γ_1(EoR + noise). The dotted red line is the estimate of γ_1 from the noisy data as given by equation 14. The contour areas represent the 2σ domains of the estimation of the values from the limited number of samples. The light-gray shaded area represents the range of γ_1 values for the noise. The darker grey shaded area represents the range of γ_1 values of the EoR signal. The orange area represents the range of γ_1 values of the estimation of the EoR signal.

forward to solve:

$$\gamma_1(y) = \frac{\mu_3(y)}{\mu_2(y)^{3/2}} = \frac{\mu_3(x+y) - \mu_3(x)}{(\mu_2(x+y) - \mu_2(x))^{3/2}}, \tag{14}$$

where x is shorthand for noise, y is shorthand for the EoR signal. However, even though the error bars grow with respect to the μ_2 case, we still succeed in estimating γ_1 correctly for $11.7 \gtrsim z \gtrsim 7.3$. Global features are captured for the remaining redshifts but outliers are such that the estimate is inaccurate. This is mostly due to the difficulty in estimating μ_3.

Our estimate of the excess kurtosis is shown in Figure 12 for both models. The evolution of the kurtosis will only be discussed for the QSO model. The kurtosis can only be estimated accurately for $11.4 \gtrsim z \gtrsim 10.8$ for the QSO model. The estimation for the hybrid model is inaccurate. Global features are captured roughly for the remaining redshifts but difficulties in estimating μ_4 cause the estimation of the kurtosis to be inaccurate. Similarly as the skewness we can identify six stages in the evolution of the excess kurtosis (see Table 2). The redshifts of the changes in the kurtosis do not coincide in all cases with the redshifts of changes in the skewness. We will discuss them separately. To avoid confusion, we will number them in roman numerals.

Stage I ($z = 12.2 - 11.4$)	Decreasing kurtosis. The differential brightness temperature δT_b is slowly growing increasingly negative in the process of coupling T_s to T_k. As this coupling increases the values of δT_b lie increasingly close together (converging to maximum coupling). The Universe in this stage is in a near uniform state: largely neutral, spin temperature T_s between T_{CMB} and T_k.
Stage II ($z = 11.4 - 10.4$)	Increasing kurtosis. In this regime the mean δT_b shifts from positive to negative. As shown for the skewness, this regime contains the transition from an absorption-dominated Universe to an emission-dominated Universe. Points at negative δT_b now shift towards positive δT_b and in doing so shift through the mean. At $z = 11.0$ the rate of increase in the kurtosis slows down. The rate at which outliers move or are created must then slow down as well. The mechanism that shifts points from negative δT_b to positive is primarily connected to heating rate. The minimum in the kurtosis is located at $z = 11.4$. At this redshift the absorption and emission peak are approximately equal in height, with the mean to the right of the absorption peak. This configuration of peaks causes the lowest value for the kurtosis.

Stage III (z = 10.4 – 8.3)	In this regime the kurtosis decreases. The mean varies between 13 and 23 mK. The data points in absorption completely vanish in this regime and the central peak at $\delta T_b = 0$ continues to grow. The trend for points is to shift from high δT_b values to $\delta T_b = 0$ because of ongoing ionization. This explains the decrease in the mean δT_b from $z = 10.2$ to $z = 8.5$. This can be seen as a decrease in outliers, and therefore decreasing kurtosis. The decrease in the mean sets in when the peak at $\delta T_b = 0$ is larger than the emission peak and continues to grow.
Stage IV (z = 8.3 – 6.0)	In this final regime the kurtosis is strongly increasing. The mean is still shifting towards $\delta T_b = 0$ and the peak at $\delta T_b = 0$ is still growing. This increase in the height of the peak causes the kurtosis to stay high, even though there are comparatively fewer outliers due to the final neutral spots in the Universe becoming ionized (driving $x_{HI} \rightarrow 0$ in equation 4). There are still some spots in the Universe that haven't been ionized yet and these function as outliers, causing the kurtosis to rise.

Globally, the excess kurtosis is positive everywhere. To analyze the kurtosis it makes sense to think in terms of outliers. If the kurtosis is positive, there is a higher probability of outliers (as compared with a Gaussian distribution). Conversely, if it is negative, there is a lower probability of outliers. For all redshifts there are outliers at high values for δT_b. This causes the kurtosis to be positive everywhere.

Trends in the kurtosis are that it drops rapidly to a global minimum in early reionization until $z \simeq 11.4$, after which it picks up again to about $z \simeq 10.4$, followed by a slow decrease until $z \simeq 8.4$. Lastly, the kurtosis rises sharply.

The first drop until $z \sim 11.4$ is a result of the mean shifting towards more negative δT_b values as T_s is being coupled to T_k. As most of the Universe is in the absorption regime the difference between $\langle \delta T_b \rangle$ and points in the absorption regime gets smaller, reducing the kurtosis.

The stage where the kurtosis picks up ($10.4 \lesssim z \lesssim 11.4$)) is the regime where the mean of the 21-cm PDF shifts from the absorption regime towards the emission regime. The transition from absorption to emission happens on a short timescale, causing most of the points in the 21-cm PDF to move quickly towards emission. The change of the mean towards emission happens on a longer timescale so in effect there are more outliers, causing the kurtosis to rise.

The following decrease for $8.4 \lesssim z \lesssim 10.4$ has the same cause as the plateau in this redshift regime for the skewness. The sharp increase in kurtosis for $z \lesssim 8.4$ shares the same cause as the skewness: ongoing ionization of the Universe drives $\langle \delta T_b \rangle \to 0$, but high δT_b points still exist.

Estimating the kurtosis of the EoR signal can be done in theory. Additivity of moments again yields

$$\begin{aligned}\gamma_2 &= \frac{\mu_4(y)}{\mu_2(y)^2} \\ &= \frac{\mu_4(x+y) - 3\mu_2(x+y)^2 - \mu_4(x) + 3\mu_2(x)^2 + 3\mu_2(y)^2}{(\mu_2(x+y) - \mu_2(x))^2}.\end{aligned} \tag{15}$$

Note that this statistic depends on the estimate of the second moment of the EoR signal $\mu_2(y)$. Any uncertainty in this statistic will propagate into the uncertainty of the estimate of the kurtosis. There is no valid estimation within the redshift range $6 \lesssim z \lesssim 7$ and this gives rise to the artifacts in Figure 12.

3.5. The Q distribution and its fitting parameters

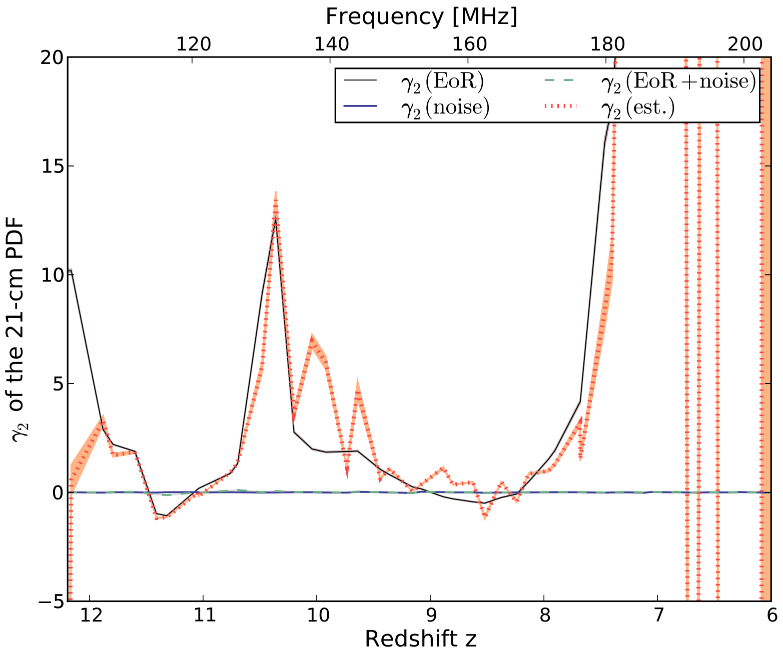
The individual components used in modelling the Q-distribution will be described. Our aim was to use a model with as little parameters as possible. Furthermore, these parameters should be connected to the physics.

The top panel in Figure 7 (with data taken from the QSO model) shows the necessity of having a distribution that looks like an inverted lognormal component in our model. Strictly speaking Q follows a ratio distribution here: T_s is approximately Gaussian at this epoch and the ratio of $(T_s - T_{CMB})/T_s$ looks like an inverted lognormal. However, since the analytical form of the ratio distribution is complicated the lognormal distribution was inverted. How this is done will be detailed later.

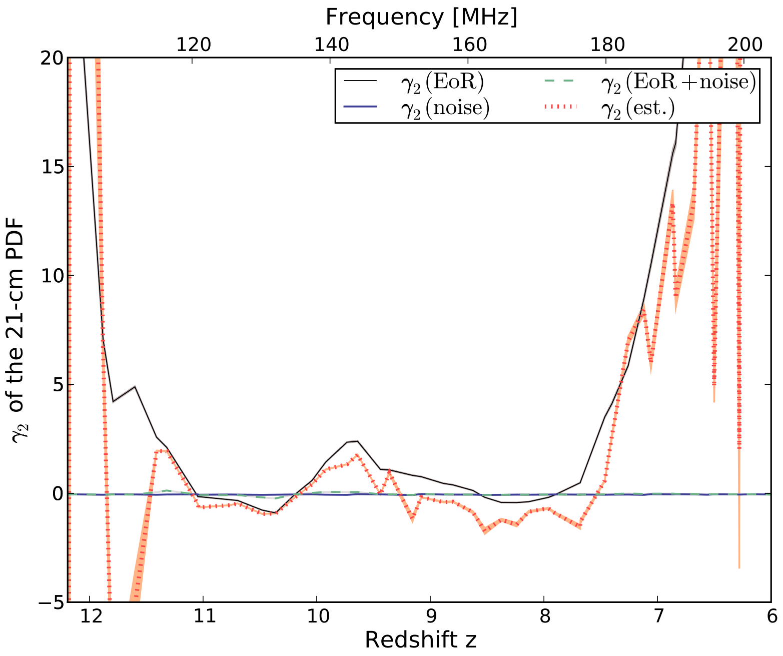
The second panel shows that an inverted lognormal distribution alone is not sufficient to model the absorption part. We have added a Gaussian to model the points that are colder (lower Q-value). Generally this Gaussian component lies to the left of the peak of the inverted lognormal component. As the Universe heats up (see the flat plateau for $Q < 0$ in panels 4 and 5) the Gaussian is suitable to model the last remaining cold spots in the IGM.

In the bottom three panels a bump is seen around $Q = 1/3$. How this bump comes into existence is discussed later in the text, but for now we decide to model this with a Gaussian distribution.

Points that approach $Q = 1$ are undergoing heating (T_s is certainly coupled to T_k in this regime) and the shape of points moving from lower Q to $Q = 1$ is such that it can be modeled via an exponential distribution. The temperature difference in Q-bins increases as one approaches $Q = 1$. The distribution will then naturally tend to an exponential.



(a) Excess kurtosis of the QSO model.



(b) Excess kurtosis of the hybrid model.

Fig. 12: Excess kurtosis (γ_2) of the δT_b distribution. The solid black line is γ_2(EoR). The solid blue line is γ_2(noise). The dashed green line is γ_2(EoR + noise). The dotted red line is the estimate of γ_2 from the noisy data as given by equation 15. The contour areas represent the 2σ domains of the estimation of the values from the limited number of samples. The light-gray shaded area represents the range of γ_2 values for the noise. The darker grey shaded area represents the range of γ_2 values of the EoR signal. The orange area represents the range of γ_2 values of the estimation of the EoR signal.

To summarize: the absorption part is modelled with a Gaussian distribution and an inverted lognormal distribution. This inverted lognormal distribution $L(x)$ is defined as follows: for $x < 0, L(x) = f(-x)$ and for $x \geq 0, f(x) = 0$, where $f(x)$ is the usual lognormal distribution with parameters μ and σ. The parameters μ and σ are always such that modelling the lognormal in this way does not give rise to discontinuities.

The emission part is modelled with an exponential and Gaussian distribution. The full distribution is given as follows:

$$P(x) = \alpha G(x|\mu_1, \sigma_1) + \beta L(x|\mu_2, \sigma_2) + \gamma G(x|\mu_3, \sigma_3) + (1 - \alpha - \beta - \gamma) E(x|\lambda), \tag{16}$$

where $E(x|\lambda) = \lambda \exp(\lambda(x-1))$. The parameters α, β and γ are the normalized fractions of each component of the model. The inequality $\alpha + \beta + \gamma \leq 1$ has to be satisfied in the fitting. Furthermore, the distributions themselves need to be properly normalized (i.e. $\int f(x)\mathrm{dx} = 1$).

3.6. Particular feature: $Q = 1/3$

One interesting feature of the Q-distribution is the bump around $Q = 1/3$. It is clearly visible in the data in the redshift range $10.70 \lesssim z \lesssim 11.88$. The question is how this peak comes into existence. It can be explained by looking at the definition of the spin temperature. Using equation 2 and writing

$$T_k = \eta T_{CMB}, \tag{17}$$

assuming that $y_c = 0$ (collisions unimportant which is a valid assumption on the scales we are looking at) we can derive

$$y_\alpha = \frac{-\eta Q}{1 - \eta + \eta Q}. \tag{18}$$

y_α is the Ly-α coupling strength coefficient and η is a measure of the temperature of the IGM. Equation 18 has two limiting cases for $Q = 1/3$: $y_\alpha \to \infty$ for $\eta \to 3/2$ and $y_\alpha \to 1/2$ for $\eta \to \infty$. We discard the last possibility because it is unphysical, leaving us with the situation where we have lots of Ly-α photons in a relatively cold region (the y_α coupling strength is proportional to the local flux density of Ly-α photons). This points at the $Q = 1/3$ feature being connected to having lots of Ly-α photons in a spherical shell around the source.

In order to explain the bump there are two components to be analyzed: the “rise” of the bump towards $Q = 1/3$ and the “fall” onwards from $Q = 1/3$. Points do not preferentially stay around $Q = 1/3$ because of the ongoing heating of the Universe. The bump is the result of cold IGM spots getting heated to $\sim \frac{3}{2} T_{CMB}$ and flowing “in” to the bump minus the spots in the bump getting heated to temperatures

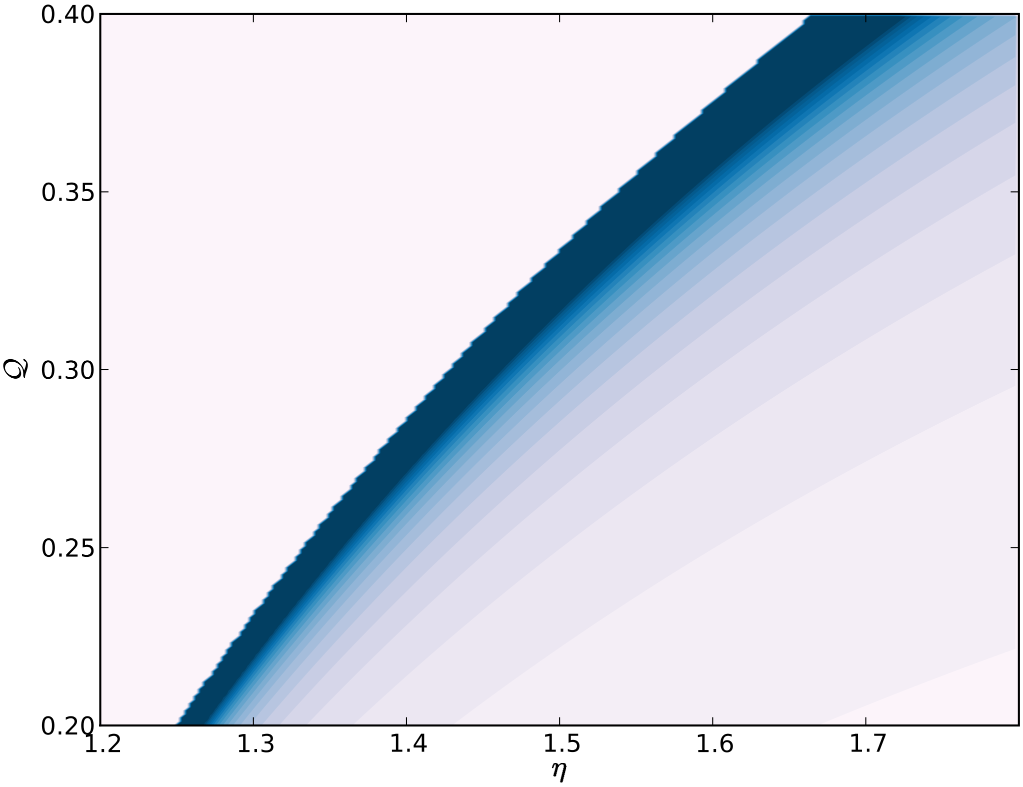


Fig. 13: Contourplot of y_α versus Q and η. The top left corner can be ignored: those contain values outside the domain of validity of equation 18. Light values denote small values, the dark blue values denote large values.

beyond $\sim \frac{3}{2}T_{CMB}$ and thus flowing "out". The bump disappears when most points in the Universe are hotter than $\frac{3}{2}T_{CMB}$ (i.e. no more inflow and the phenomenon disappears).

The answer is of a geometrical nature: the fact that a bump is seen means that there are many points around $Q = 1/3$. Assuming $T_s = T_k$ this corresponds to a zone where T_k is around $\frac{3}{2}T_{CMB}$. Because of the spherical symmetry employed in BEARS, every radiative source has a shell around it where the temperature is on the order of $\frac{3}{2}T_{CMB}$. As time proceeds this shell will expand radially outwards until it disappears when most of the Universe is heated to temperatures exceeding $\frac{3}{2}T_{CMB}$ and these shells start to merge with those of other nearby sources.

The fall of the bump after $Q = 1/3$ naturally corresponds with the volume decrease in shell size of those points with temperatures exceeding $T_k = \frac{3}{2}T_{CMB}$. The increase seen in the number of points with $Q > \frac{1}{2}$ (beyond the bump) is explained by the fact that the size of the temperature bins for the largest Q values is much larger and more points fall within that region.

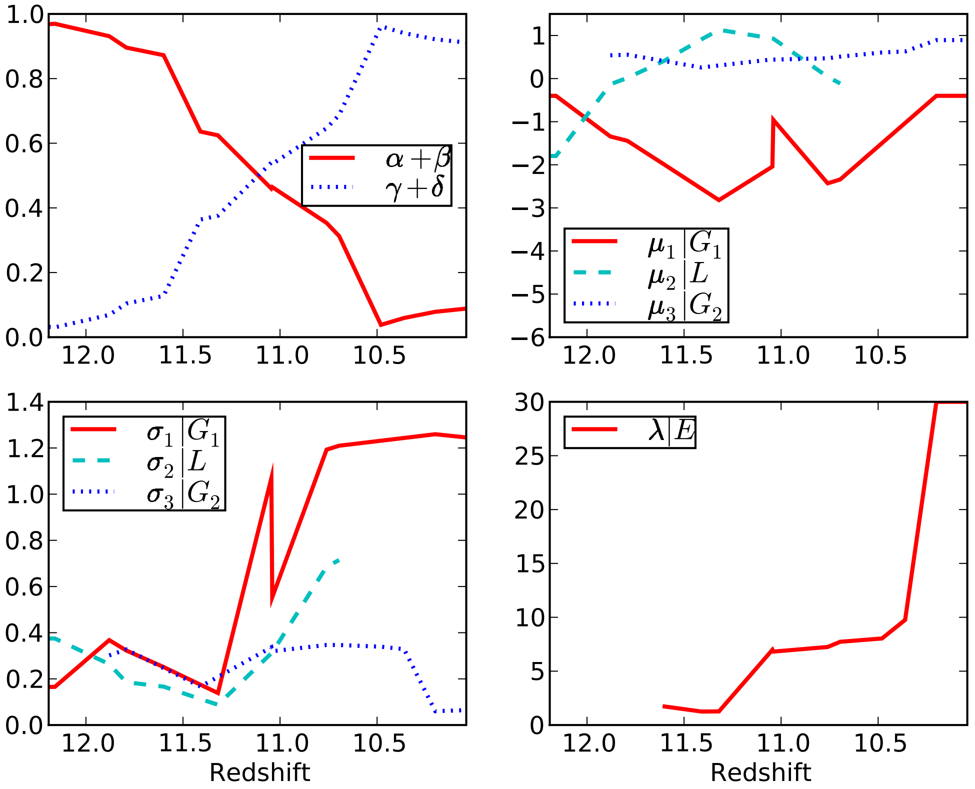


Fig. 14: All panels correspond to the corresponding parameters in equation 16. The legend is in the format PARAMETER $|$ FUNCTION, where function can be G_1 for the Gaussian defined in the absorption regime, G_2 for the Gaussian in the emission regime, L for the inverse lognormal function or E for the exponential function. Only those points of the parameters have been plotted where the probability of their component exceeds 0.05.

3.7. Evolution of the fitting parameters of the Q distribution with redshift

In Figure 14 we present the evolution of the fitting parameters of equation 16 with redshift. Only points whose components have weights larger than 0.05 are plotted in the Figure. The top left panel shows the sum of the probabilities of the absorption $(\alpha + \beta)$ and the emission $(\gamma + \delta)$ components. For clarity, $\delta = 1 - \alpha - \beta - \gamma$ and it thus models the fractional probability of the exponential component. While the absorption component is dominant in the early Universe it decreases almost monotonically with a corresponding rise in the emission component.

The rest of the panels are best interpreted alongside with the help of Figure 7. We see that the parameters μ_1 and σ_1, describing the location and shape of the first Gaussian, partially model the absorption peak in early reionization. However, later on the fitting algorithm uses this component to model the much smaller remains of the cold IGM in the Universe. It fills two roles and one should be careful not to mix the two stages. μ_1 starts at -0.5, decreases towards -3 and then increases to -0.5. σ_1 is small at first and grows gradually broader. This change in parameters corresponds with a decrease in importance: α goes to zero as these parameters change. In all cases the first Gaussian can be thought of as modelling the points in the Universe that are "lagging behind" on heating, i.e. points that are not subject to the dominant physical process at work. In early reionization because it is left of the inverted log normal peak and in late reionization because it models the low Q values while the majority of points is at high Q. The inverted lognormal distribution parametrizes the ratio distribution where Gaussians are involved. The additional Gaussian can then be thought of as modelling the nongaussianity in T_s.

Similarly, μ_2 and σ_2 are the inverse lognormal component to the absorption peak, which (as we can see in Figure 7) almost completely vanishes after some time. This is why the contribution of the lognormal is negligible for lower redshifts ($\beta \to 0$). It represents the majority of the points in the absorption regime: first the lognormal component models most points cooling down with the expansion of the Universe, followed by modelling most points heating up from the absorption to the emission regime. After this last phase its contribution is essentially zero. It is the contribution of those T_s points which follow a Gaussian or nearly Gaussian distribution.

For the emission parameters μ_3, σ_3 and λ the situation is much simpler. These model one physical mechanism at work: they represent the ongoing heating of points in the Universe (as opposed to heating and cooling for the earlier case). The Gaussian is required to model the bump seen around $Q = 1/3$ and the exponential the increase of points seen towards $Q = 1$. Looking at the data in detail there are several bumps between the absorption peak and before Q. Each of these smaller bumps can be connected to a heating bubble around a source as the Universe is being reheated. However, the bump at $Q = 1/3$ is the most pronounced.

3.8. Influence of changing parameters on the results

BEARS is well suited to explore a wide variety of source properties such as their lifetime, radiation spectrum and clumping factor of their surroundings. Of these properties the radiation spectrum is the most influential.

The largest influence is expected to be found in the distribution of Q: changes in Q are governed by changes in the efficiency of the coupling mechanisms as well as the evolution of the temperature of the IGM. The low end of the radiation spectrum is mostly responsible for coupling T_s to T_k (which could be constrained by the speed of the initial shift from negative Q to even more negative Q), while the high end of the radiation spectrum is mostly responsible for heating the IGM (which could be constrained by the speed of points shifting from negative Q to higher Q). Photons at 13.6 eV are the most efficient at ionizing the IGM. The growth in size of ionizing bubbles might constrain this particular part of the radiation spectrum.

Changes are expected for the 21-cm PDF as well: sources that are efficient at heating will shorten the duration of the absorption phase of the Universe, while sources that are inefficient at heating (such as stars as the only reionizing sources) will prolong the duration of the absorption phase. This will show in the statistics as well: changes of skewness and kurtosis will happen on shorter (longer) timescales depending on how fast (slow) sources heat up the IGM.

4. Observational effects

4.1. LOFAR properties

It is unlikely that the next generation of radio telescopes will be able to image the EoR signal directly. This is because the observations of the EoR will suffer from instrumental noise, beam smearing, RFI, calibration errors and so on. Furthermore, the sensitivity, resolution and noise properties of these telescopes are not good enough. The EoR signal will therefore be detected statistically at first. In order to detect the cosmological signal as accurately as possible the foregrounds (among other effects) have to be removed. In this paper we will assume that this removal is feasible (see for example Jelić et al., 2008). Also calibration errors are assumed to be negligible (see Lampropoulos et al., 2010, in prep). The RFI errors can be taken out of the data via flagging it. The noise contribution that we do take into account is frequency-dependent non-RFI errors from the sky. The limited resolution of LOFAR causes the images to be smeared as well (commonly called beam smearing).

4.2. Noise calculation and addition

A standard LOFAR setup is assumed: the size of a LOFAR station is assumed to be 30 m, and we take 2 km for the maximum baseline length (corresponding to an approximate size of the LOFAR core). The level of noise expected is on the order of 100 mK. The field of view is roughly $5° \times 5°$. The contribution from non-RFI noise is taken from a Gaussian distribution. As a first-order approximation we generate 512^3 pixels with values corresponding to a Gaussian distribution with the appropriate noise parameter (the frequency-dependent sky noise). However, the noise is also subject to the LOFAR resolution. Therefore this distribution is smoothed with a kernel with the same size as the LOFAR resolution. Since smoothing causes the standard deviation of the distribution to change the values are renormalized such that the smoothed noise map has the appropriate standard deviation.

The expected standard deviation for the noise (at 150 MHz, with 400 hours of observation and 1 MHz frequency bandwidth) is expected to be on the order of 52 mK. For this reason we model it as

$$\sigma_{noise} = 52 \text{ mK} \left(\frac{\nu}{150 \text{ MHz}} \right)^{-2.55}. \tag{19}$$

The maps of the differential brightness temperature are smoothed with the same kernel. To create a corrupted 21-cm map, the smoothed noise is added to the smoothed map of the 21-cm signal.

In the following sections it is assumed that the noise properties are known. Estimating the noise with LOFAR is in principle possible by, for example, doing observations on an extended calibrated source. The difference between the observed temperatures along the different line of sights is then a sample from the noise distribution. Repeating this should allow one to reconstruct the parameter σ of the Gaussian distribution.

4.3. Measuring higher order statistics with LOFAR

Measuring higher order statistics with LOFAR can be done. However, the problem is that each successive statistic can be thought of as representing fluctuations in the lower-order statistic. The higher order the statistic to be estimated is, the more samples are required to obtain an accurate estimation of the statistic.

A few factors influence how easily data can be obtained from LOFAR. An example is its field of view. The field of view represents the area in the sky that the telescope can collect data from. The larger the field of view, the more data samples can be taken (with constant resolution). The resolution is another important metric: it represents the number of independent elements that a telescope can detect from the sky. The resolution of an interferometer like LOFAR depends on the baseline length and the wavelength of the radiation to be detected. The larger the baseline, the better the resolution and the more samples that can be taken. However, for larger baseline lengths it becomes more difficult to remove foregrounds and do proper calibration. The EoR signal might get under- or overfitted out of the signal. For this reason the LOFAR properties are taken to be that of the LOFAR core. Another important quantity is the frequency resolution. The frequency direction is similar to the time direction. Extra samples can be obtained by binning in the frequency direction provided that changes in the 21-cm signal do not change strongly with time. We assume a frequency bandwidth of 1 MHz for LOFAR, combined with a frequency resolution of 0.8 kHz. The simulation snapshots in which significant changes take place are more than 1 MHz apart, justifying this assumption of binning in the frequency direction.

4.4. Finding Q from 21-cm maps

In principle it is possible to find the distribution of Q when one has the distribution of δT_b, x_{HI} and $(1+\delta)$. The problem is that there are multiple distributions Q that give rise to the same δT_b when direct inversion methods are used. It is straightforward to see why: the continous integral of equation 13 can be reduced to a matrix equation by writing the integral as a Riemann sum:

$$\begin{aligned} P(\delta T_b) &= \int \frac{1}{\omega} P_1(\omega) P_2\Big(\frac{\delta T_b}{\omega}\Big) d\omega \\ &\approx \sum_{i=1}^{N} \frac{1}{\omega_i} P_1(\omega_i) P_2\Big(\frac{\delta T_b}{\omega_i}\Big) \Delta\omega_i. \end{aligned} \tag{20}$$

For $N \to \infty$ they become the same. The discretization allows us to reduce this to a simple $\bar{\bar{A}}\bar{x} = \bar{b}$ equation. Suppose that $\bar{\bar{A}}$ contains the $\frac{1}{\omega_i} P_1(\omega_i)\Delta\omega_i$ components, $\bar{x}$ the $P_2(\frac{\delta T_b}{\omega_i})$ components and $\bar{b}$ all $P(\delta T_b)$ that need to be solved for. Then, to find $\bar{x}$ (i.e. Q), an inverse of $\bar{\bar{A}}$ is required such that $\bar{\bar{A}}^{-1}\bar{b} = \bar{x}$. However, $\bar{\bar{A}}$ is only invertible if and only if its rank is N. It fails to meet this criterion (there is only one linearly independent column in $\bar{\bar{A}}$) and such there is no unique inverse.

5. Conclusions

In this report we have shown how to create 21-cm PDFs from integration of the component distributions, or via samples of the component distributions and lastly, computing the 21-cm PDF using the simulation data.

From these PDFs we have shown that it is possible to compute moments. This has been done for the pure and uncorrupted data and also for the data which had noise added to it. Via the Monte Carlo approach employed it was shown that it is possible to estimate a few of the higher order moments accurately from noisy data as the estimate from the pure case agrees with the estimate from the noisy case for the second and third moment. Estimating the fourth moment proves to be difficult.

Because this estimation of the second and third moment was possible we have shown that it is also possible to estimate more complicated statistics from the noisy data, such as the skewness. The kurtosis depends on the estimate of the 4th moment and is accurate only in a limited redshift range.

Lastly, in examining simulation data a quantity Q was found, which is an excellent tracer of early reionization. It contains information about two components: propagation of the very first Ly-α photons throughout the Universe and the second one being the heating from these first photons.

What remains to be done however is to find a statistic that enables us to extract the size of the ionized bubbles as a function of redshift. Knowing this would put constraints on the proceedings of reionization.

This requires statistics of a different type: so far we have used one-point PDFs and interpreted these, but the next step is to try and estimate correlation functions from (noisy) data. Estimating the extent of ionized bubbles requires spatial data and therefore two-point PDFs.

We have shown that reconstructing the Q distribution is not possible using direct inversion methods. However, our parametrization of the Q distribution combined with our 21-cm mock PDF creation methods might aid in finding the distribution of Q using maps of the 21-cm signal only, combined with known functions for the density field and ionized fraction distributions. Reconstructing the distribution of Q would greatly aid in constraining source properties as we have shown.

In principle simulations could also be used to investigate the influence of cosmological parameters (such as σ_8 (a measure of fluctuations in the power spectrum on scales of $8h^{-1}$ Mpc) or n_s (the slope of the power spectrum)) on the 21-cm signal.

Bibliography

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

G. Bruzual and S. Charlot. Stellar population synthesis at the resolution of 2003. *MNRAS*, 344:1000–1028, October 2003. .

H. M. P. Couchman. Reheating of the intergalactic medium at Z greater than 10. *MNRAS*, 214:137–159, May 1985.

M. Davis, G. Efstathiou, C. S. Frenk, and S. D. M. White. The evolution of large-scale structure in a universe dominated by cold dark matter. *ApJ*, 292:371–394, May 1985. .

J. Dunkley, E. Komatsu, M. R. Nolta, D. N. Spergel, D. Larson, G. Hinshaw, L. Page, C. L. Bennett, B. Gold, N. Jarosik, J. L. Weiland, M. Halpern, R. S. Hill, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, E. Wollack, and E. L. Wright. Five-Year Wilkinson Microwave Anisotropy Probe Observations: Likelihoods and Parameters from the WMAP Data. *ApJS*, 180:306–329, February 2009. .

X. Fan, M. A. Strauss, D. P. Schneider, R. H. Becker, R. L. White, Z. Haiman, M. Gregg, L. Pentericci, E. K. Grebel, V. K. Narayanan, Y.-S. Loh, G. T. Richards, J. E. Gunn, R. H. Lupton, G. R. Knapp, Ž. Ivezić, W. N. Brandt, M. Collinge, L. Hao, D. Harbeck, F. Prada, J. Schaye, I. Strateva, N. Zakamska, S. Anderson, J. Brinkmann, N. A. Bahcall, D. Q. Lamb, S. Okamura, A. Szalay, and D. G. York. A Survey of z¿5.7 Quasars in the Sloan Digital Sky Survey. II. Discovery of Three Additional Quasars at z¿6. *AJ*, 125:1649–1659, April 2003. .

X. Fan, M. A. Strauss, G. T. Richards, J. F. Hennawi, R. H. Becker, R. L. White, A. M. Diamond-Stanic, J. L. Donley, L. Jiang, J. S. Kim, M. Vestergaard, J. E. Young, J. E. Gunn, R. H. Lupton, G. R. Knapp, D. P. Schneider, W. N. Brandt, N. A. Bahcall, J. C. Barentine, J. Brinkmann, H. J. Brewington, M. Fukugita, M. Harvanek, S. J. Kleinman, J. Krzesinski, D. Long, E. H. Neilsen, Jr., A. Nitta, S. A. Snedden, and W. Voges. A Survey of z¿5.7 Quasars in the Sloan Digital Sky Survey. IV. Discovery of Seven Additional Quasars. *AJ*, 131:1203–1209, March 2006. .

GB Field. Excitation of the hydrogen 21-cm line. *Proceedings of the Institute Radio Engineers*, 46:240–250, 1958.

V. Gluscevic and R. Barkana. Statistics of 21-cm Fluctuations in Cosmic Reionization Simulations: PDFs and Difference PDFs. *ArXiv e-prints*, May 2010.

J. E. Gunn and B. A. Peterson. On the Density of Neutral Hydrogen in Intergalactic Space. *ApJ*, 142:1633–1641, November 1965. .

V. Jelić, S. Zaroubi, P. Labropoulos, R. M. Thomas, G. Bernardi, M. A. Brentjens, A. G. de Bruyn, B. Ciardi, G. Harker, L. V. E. Koopmans, V. N. Pandey, J. Schaye, and S. Yatawatta. Foreground simulations for the LOFAR-epoch of reionization experiment. *MNRAS*, 389:1319–1335, September 2008. .

J. R. Pritchard and S. R. Furlanetto. 21-cm fluctuations from inhomogeneous X-ray heating before reionization. *MNRAS*, 376:1680–1694, April 2007. .

U. Seljak and M. Zaldarriaga. A Line-of-Sight Integration Approach to Cosmic Microwave Background Anisotropies. *ApJ*, 469:437–+, October 1996. .

J. M. Shull and M. E. van Steenberg. X-ray secondary heating and ionization in quasar emission-line clouds. *ApJ*, 298:268–274, November 1985. .

V. Springel, S. D. M. White, A. Jenkins, C. S. Frenk, N. Yoshida, L. Gao, J. Navarro, R. Thacker, D. Croton, J. Helly, J. A. Peacock, S. Cole, P. Thomas, H. Couchman, A. Evrard, J. Colberg, and F. Pearce. Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature*, 435:629–636, June 2005. .

R. M. Thomas and S. Zaroubi. On the spin-temperature evolution during the epoch of reionization. *MNRAS*, pages 1559–+, October 2010. .

R. M. Thomas, S. Zaroubi, B. Ciardi, A. H. Pawlik, P. Labropoulos, V. Jelić, G. Bernardi, M. A. Brentjens, A. G. de Bruyn, G. J. A. Harker, L. V. E. Koopmans, G. Mellema, V. N. Pandey, J. Schaye, and S. Yatawatta. Fast large-scale reionization simulations. *MNRAS*, 393:32–48, February 2009. .