doi:10.1038/nature11177

S1. Description of the simulation code

We developed our own code that implements a hybrid method to produce instances of the expected three-dimensional distribution of the first stars. We first used the known statistical properties of the initial density and velocity perturbations to generate a realistic sample universe on large, linear scales. Specifically, we assumed Gaussian initial conditions and adopted the initial power spectrum corresponding to the currently best-measured cosmological parameters[31]. In a cubic volume consisting of 128^3 cells (each 3 comoving Mpc on a side), we generated as in our previous work[1] a random realization, including the appropriate correlations, of the initial overdensity and relative baryon-dark matter velocity in each cell (with periodic boundary conditions). These values are easily computed at any redshift as long as the scales are sufficiently large to use linear perturbation theory. We then computed analytically the gas fraction in star-forming halos in each cell as a function of these two variables and the redshift, as in our previous papers. Specifically, this gas has density[13]

$$\rho_{\rm gas} = \int_{M_{\rm cool}}^{\infty} \frac{dn}{dM}\, M_{\rm gas}(M)\, dM \ , \tag{1}$$

where dn/dM is the comoving abundance of halos of mass M (i.e., n is the comoving number density), $M_{\rm gas}(M)$ is the gas mass inside a halo of total mass M, and $M_{\rm cool}$ is the minimum halo mass in which the gas can cool efficiently and form stars. In this calculation (whose results are illustrated in Figs. 1 and 2) we included three separate effects of the relative velocity on star formation[14], namely the effect on $M_{\rm cool}$, on dn/dM, and on $M_{\rm gas}(M)$ (see also section S2). The stellar density equals $\rho_{\rm gas}$ multiplied by the star-formation efficiency.

We then used this information to determine the X-ray heating rate in each cell as follows. At each redshift, we smoothed the stellar density field in shells around each cell, by filtering it (using fast Fourier transforms) with two position-space top-hat filters of different radii and taking the difference. We assumed the flux of X-ray photons emitted from each shell to be proportional to the star formation rate, which is in turn proportional to the time derivative of $\rho_{\rm gas}$. We assumed an X-ray efficiency of 1.75×10^{57} photons per solar mass in stars (1.15×10^{57} for the case with no streaming velocity) produced above the minimum energy (assumed to be 200 eV) that allows the photons to escape from the galaxy. The efficiency in each case was chosen so as to get the peak of the cosmic heating transition at $z = 20$, i.e., so that the mean kinetic gas temperature equals the cosmic microwave background (CMB) temperature at that redshift. The actual X-ray efficiency of high-redshift galaxies is highly uncertain, but 10^{57} photons per solar mass along with our adopted power-law spectrum corresponds to observed starbursts at low redshifts[4]. We then computed the heating by integrating over all the shells seen by each cell, as in the 21CMFAST code[17]. In this integral, the radiative contribution of each cell to a given central cell was computed analytically at the time-delayed redshift seen by the central cell, using a pre-computed interpolation grid of star formation versus overdensity, streaming velocity, and redshift. We varied the number and thickness of shells to check for convergence. To estimate the optical depth, we assumed a uniform density and a neutral inter-galactic medium, but did not make a crude step-function approximation as in 21CMFAST. We used photoionization cross sections and energy deposition fractions from atomic physics calculations[32,33].

Given the X-ray heating rate versus redshift at each cell, we integrated as in 21CMFAST to get the gas temperature as a function of time. We interpolated the heating rate between the redshifts where it was explicitly computed, and varied the number of redshifts to ensure convergence. We then assumed that the spin temperature and the gas temperature are coupled to compute the 21cm signal, i.e., that the Lyman-α coupling has already saturated by $z = 20$, as expected (see section S3). Except for the differences noted, in the heating portion of the code we followed 21CMFAST and adopted their fiducial parameters, such as a 10% star-formation efficiency. However, our source distribution was substantially different since they did not include the effect of the streaming velocity. Since we focused on the era well before the peak of cosmic reionization, we did not calculate ionization due to ultra-violet or X-ray radiation. The kinetic temperature T_k and overdensity δ of the gas in each cell gave us the 21-cm brightness temperature (relative to the CMB temperature $T_{\rm CMB}$)[3]

$$\delta T_b = 40(1+\delta)\left(1-\frac{T_{\rm CMB}}{T_{\rm k}}\right)\sqrt{\frac{1+z}{21}}\ {\rm mK}\ , \tag{2}$$

and thus Figs. 3 and 4. Finally, for Fig. S1 (in section S3) we added a calculation of the inhomogeneous Lyman-Werner flux[16] within the box using the halo distribution in the box similarly to our calculation of the inhomogeneous X-ray heating rate.

S2. Comparison with previous work

In this section we briefly summarize previous work on the streaming velocity and note the differences with our work.

It is now known that the relative motion between the baryons and dark matter has three effects on halos: (1) suppressed halo numbers, i.e., the abundance of halos as a function of total mass M and redshift z; (2) suppressed gas content of each halo, i.e., the gas mass within a halo of a given M and z; and (3) boosted minimum halo mass needed for cooling, i.e., the minimum total mass M of halos at each redshift z in which catastrophic collapse due to cooling, and thus star formation, can occur. Note that this separation into three distinct effects is natural within our model (see Eq. 1 in section S1), but this does not preclude the possibility that they are physically correlated or mutually dependent.

The original paper in which the importance of the relative motion was discovered[1] included only the impact on the halo abundance (effect #1). This was sufficient for them to deduce the important implication of enhanced large-scale fluctuations, but quantitatively the effect was underestimated. Also, their calculations had a number of simplifying assumptions: they calculated the baryon perturbations under the approximation of a uniform sound speed (which has a big impact on the no-streaming-velocity case which is still relevant in regions where the streaming velocity is low), and used the old (and relatively inaccurate) Press-Schechter halo mass function.

The effect of the relative velocity on suppressing the gas content of halos (effect #2) was the next to be demonstrated[2]. These authors predicted significant fluctuations on large scales, with prominent baryon acoustic oscillations. However, they made a number of simplifying approximations (detailed previously[14]). Most important were two limitations: they included only effect #2 (i.e., they left out the already-known #1), and they scaled star formation according to the total gas content in halos, without including a cooling criterion for star formation. The vast majority of the gas is in minihalos that cannot cool, and because of their low circular velocities their ability to collect baryons is much more affected by the streaming velocity than the star-forming halos. Even more importantly, they only considered fluctuations in the Lyman-α radiation, which yielded a prediction at $z = 20$ of a large-scale power spectrum peak of amplitude 5 mK^2 (see their Fig. 4). In comparison, in our Fig. 4 the large-scale peak (due to X-ray heating fluctuations) is more than 20 times higher, at around 110 mK^2. We also note that they assumed a particularly low Lyman-α efficiency in order to get significant Lyman-α fluctuations at a redshift as low as 20, while such fluctuations are actually expected to be significant only at a much higher redshift (see section S3 below), where the observational noise is much higher.

In a subsequent paper[13] we calculated the consequences of the combination of effects #1 and #2 on the distribution of star-forming halos as well as on star-less gas minihalos. At this point there were indications from numerical simulations[5,7,8] that the minimum halo mass needed for cooling also changed as a result of the streaming velocity (effect #3). Recently, numerical simulations have also been used for a more robust and detailed look at effect #1[6]. We have studied the three effects on halos and shown[14] that the effect on star-forming halos, and thus also on the various radiation fields, is mainly due to effects #1 and #3, while the smaller gas minihalos are mainly affected by effects #1 and #2.

In summary, the existence and correct determination of the various effects of the streaming velocity on star formation have been worked out gradually. The present paper fully incorporates that understanding in order to study the implications for X-ray heating fluctuations, resulting in a solid prediction of strong large-scale 21-cm fluctuations around redshift 20.

S3. Timing of feedback transitions

In the main text, we noted that three radiative transitions are expected to occur at high redshift: Lyman-α coupling, X-ray heating, and Lyman-Werner suppression. In our results in Fig. 4, we assumed that Lyman-α coupling occurs early, while the other two transitions occur later and may overlap. In this section we explain why this relative timing of the feedback transitions is expected.

It has been previously shown[4] that the heating transition is expected to occur significantly later than Lyman-α coupling. Specifically, the scenarios considered by these authors showed a clear period of observable 21-cm absorption before heating (see their Fig. 1). They, however, considered scenarios in which only large (atomic cooling) halos are included. In our calculations, we included also the highly abundant molecular-cooling halos, and these help produce the various transitions at higher redshifts, and with a larger gap between the Lyman-α coupling and the heating transition. Specifically, we find that in our model the coupling transition (which is also when Lyman-α fluctuations are maximal) is expected to occur at redshift 27.8 (compared to 30.1 without the streaming velocity effect). Note that the result without velocities is in good agreement with a similar previous calculation[16]. For the heating transition, we adopted redshift 20 in the paper, but allowing for a range of uncertainty of an order of magnitude in the X-ray efficiency (centered around the efficiency of observed starbursts) gives a transition redshift within $z = 17 - 21$, well after the peak of the Lyman-α coupling transition (the range is $z = 17 - 23$ without the velocity effect).

The third (LW) transition should occur significantly later than previously estimated in the literature. Both simulations of individual halos[35] and full cosmological simulations[34,36,37] that investigated halo formation under the influence of an external LW background used an artificially input *fixed* LW flux during the entire halo formation process. In reality the LW flux rises exponentially with time (along with the cosmic star formation rate) at high redshifts. Taking the final, highest value reached by the LW flux when the halo forms, and assuming that this value had been there from the beginning, greatly overestimates the effect of the LW feedback. In fact, a change in LW flux takes some time to affect the halo. The flux changes the formation rate of molecular hydrogen, but it then takes some time for this to affect the collapse. For instance, if the halo core has already cooled and is collapsing to a star, changing the LW flux will not suddenly stop or reverse the collapse. Another indication for the gradual process involved is that the simulation results can be approximately matched[34] by comparing the cooling time in halo cores to the Hubble time (which is a relatively long timescale). Thus, estimates[35,16] of the LW feedback based on the LW flux at halo virialization overestimate the transition redshift.

In order to better estimate the effect of LW feedback, we have calculated the mean LW intensity in our simulated volume, and compared it to a critical threshold for significant suppression of halos. We adopt a threshold intensity of $J = 10^{-22}$ erg s^{-1} cm^{-2} Hz^{-1} sr^{-1} as defining the center of the LW transition. The above-mentioned cosmological simulations indicate that at this intensity, the minimum halo mass for cooling (in the absence of streaming velocities) is raised to $\sim 2 \times 10^6 M_\odot$ due to the LW feedback. This is a useful fiducial mass scale, roughly intermediate (logarithmically) between the cooling masses obtained with no LW flux or with saturated LW flux, and characteristic of the scale at which the streaming velocity effect is significantly but not overwhelmingly suppressed (e.g., the velocity effect on the halo abundance is maximized at this mass scale[1]). Thus, at this level of LW suppression we would expect the 21-cm power spectrum (in the case of an X-ray heating transition at $z = 20$) to be approximately the average of the two top curves in Fig. 4. Note that even in the case of a fully saturated LW feedback, a minor ($5 - 10\%$) effect remains for the velocities on the 21-cm power spectrum.

A key point is that the critical feedback threshold must be compared not to the LW intensity when the halo virializes, but to its typical or average value during the entire process of halo formation. Another important feature is that the LW transition is very gradual. Adopting a reasonable range of uncertainty, we find (Fig. S1) that the LW transition, for our adopted parameters, should be centered somewhere in the range $z = 21 - 28$, with its main portion extending over a $\Delta z \sim 6-8$ (note that the center is expected in the range $z = 24-31$ without the velocity effect). In fact, the feedback itself will delay the heating transition to lower redshift, so that in general we expect redshift 20 to show a significant velocity signature.

We conclude that the Lyman-α transition occurs well before the X-ray heating transition, while the latter likely overlaps in redshift with the LW transition. Note that the prediction for the Lyman-α transition is more secure (for a given star formation efficiency) than the others, since the Lyman-α radiation comes directly from stars (unlike the more uncertain X-ray emission associated with stellar remnants), and it directly affects the low-density intergalactic gas (unlike the more uncertain LW feedback which occurs within the non-linear cores of collapsing halos). In particular, the LW feedback may be further delayed by complex local feedback effects that can oppose the suppression effect[39,40].

S4. Observational considerations

In the main text we argued that there are good prospects for observing the 21-cm power spectrum that we predict at redshift 20. In this section we briefly elaborate on the experimental sensitivity that we adopted and on the observational challenges.

In Fig. 4 we showed the projected 1-σ sensitivity of one-year observations with an instrument like the first-generation MWA and LOFAR experiments. Specifically, we adopted the projected sensitivity of the MWA from a detailed analysis of the sensitivity to the power spectrum[29]. The parameters of the actual instruments have changed somewhat, but in any case no

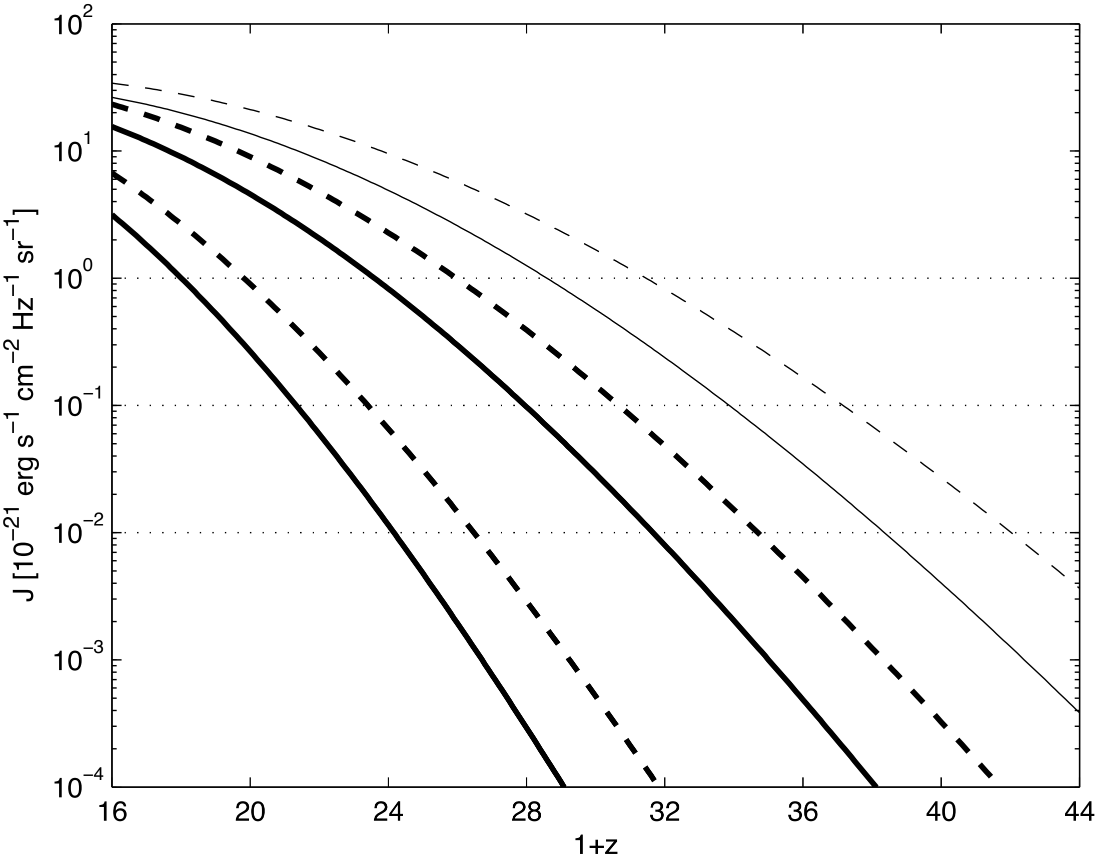

Figure S1: **The expected timing of the Lyman-Werner feedback.** We show the mean Lyman-Werner intensity J in our simulation box as a function of redshift, with (solid) and without (dashed) the relative velocity effect. In each case, we show the actual intensity (top, thin curve), and a range of effective intensities for halo feedback (bottom, thick curves). Specifically, for the effective intensity we adopt the intensity that was in place at the midpoint of halo formation. This is a reasonable estimate of the characteristic value during halo formation since, during the formation process, half the time J was below this value, and half the time above it. We estimate the midpoint of halo formation (in terms of cosmic age) based on the standard spherical collapse model[38]. To obtain a plausible range of uncertainty, we consider the start of halo formation to be either the beginning of the universe, or the start of the actual collapse (i.e., the moment of turnaround); the former yields an earlier characteristic time and corresponds to the bottom curve in each case. Also shown (horizontal lines) are critical values of LW intensity (to be compared with the effective intensities) that correspond to the central portion of the LW transition, during which the minimum halo mass for cooling (in the absence of streaming velocities) is raised by LW feedback to[36] $8 \times 10^{5} M_{\odot}$, $2 \times 10^{6} M_{\odot}$, and $5 \times 10^{6} M_{\odot}$, respectively.

current instrument is designed for observations at $z = 20$; we considered instruments in the same class of capabilities but designed to operate at 50–100 MHz. Specifically, we assumed an instrument with 500 antennas, a field of view of 800 deg^2, and an effective collecting area at $z = 20$ of 23,000 m^2, and scaled the noise power spectrum from redshift 12 to redshift 20 up by a factor of 12 [proportional to $(1+z)^{5.2}$] due to the brighter foreground[29]. The sensitivity in Fig. 4 is calculated for an 8 MHz band and bin sizes of $\Delta k = 0.5k$. It assumes a 1000 hr integration in a single field of view, i.e., it allows for a selection (out of an 8800 hr year) of night-time observations with favorable conditions.

An instrument like LOFAR – with 64 antennas, a field of view of 50 deg^2, and a collecting area at $z = 20$ of 190,000 m^2 – should have a slightly better power spectrum sensitivity, i.e., lower noise by about a factor of two[29]. A second-generation instrument should reach a substantially better sensitivity, e.g., by an order of magnitude for the SKA or a 5000-antenna MWA[29].

A possible concern, especially with large-scale modes in the 21-cm signal, is the degeneracy with the foregrounds. At each point on the sky, or at each point in the Fourier (u, v)-plane, the intensity spectrum of synchrotron and free-free foregrounds is smooth. The fitting and removal of these foregrounds also removes some of the signal at small radial wavenumbers $k_{\|}$, which means that the power spectrum of the cosmological 21-cm signal at sufficiently small k is not measurable.

The range of wavenumbers k that are affected by foregrounds follows from geometrical considerations as well as the complexity of the foreground model that must be removed. The first issue is that template projection removes a range of $k_{\|} = k \cos\theta$ rather than a range of k, where θ is the angle between the wave vector and the line of sight. Therefore if we must cut at some $k_{\|,\mathrm{min}}$ then all values of $k < k_{\|,\mathrm{min}}$ are rejected, and at larger values of k a fraction $1 - k_{\|,\mathrm{min}}/k$ survive. The foreground model consists of a smooth function such as a low-order polynomial (as well as Galactic radio recombination lines confined to specific frequencies[41]).

The relation between the foreground model complexity and the range of suppressed $k_{\|}$ is more complex[30]. The simplest argument to derive $k_{\|}$ is via mode counting: at each pixel in the (u, v)-plane of size $\Delta u\,\Delta v = \Omega^{-1}$ (where Ω is the solid angle of the survey), if one removes a polynomial of order $N - 1$ (i.e., with N independent coefficients) then one has removed the lowest N radial modes. Since the number of modes per unit radial wavenumber (including both positive and negative $k_{\|}$) is $\Delta r/(2\pi)$, where Δr is the radial width of the survey, mode-counting would suggest that radial wavenumbers from $-k_{\|,\rm min} < k < k_{\|,\rm min}$ are lost in the projection, giving $k_{\|,\rm min} = \pi N/\Delta r$. Despite its simplicity, the mode-counting argument holds up well against much more detailed studies. A good example[11] is a simulated foreground subtraction in the frequency range 142–174 MHz, using the subtraction of a cubic polynomial ($N = 4$). This corresponds to a radial shell of width $\Delta r = 551\,$Mpc. Mode-counting suggests that subtraction of real signal should become an issue at $k_{\|,\rm min} = 0.023\,{\rm Mpc}^{-1}$, and in fact Fig. 13 of these authors[11] shows that the 21-cm signal remains intact over the entire range of scales investigated (0.03–1.0 $\rm Mpc^{-1}$). Larger values of $k_{\|,\rm min}$ occur in calculations with narrower bandwidths[42].

For our $z = 20$ case, assuming a bandwidth of 60–80 MHz, the same mode-counting argument leads to $k_{\|,\rm min} = 0.03\,{\rm Mpc}^{-1}$ for $N = 5$. Thus we would expect that if the foregrounds can be described by the lowest 5 modes over a factor of 1.33 in frequency, that they are distinguishable from our signal. In Fig. 4 we have included the estimated degradation of the observational sensitivity for these parameters, with a $1/\sqrt{1 - k_{\|,\rm min}/k}$ factor.

This of course leaves open the issue of how many foreground modes actually need to be removed. A previous study[30] suggests 3–4 modes might be sufficient, but they considered higher frequencies (where the foreground:signal ratio is smaller) and used principal components of their foreground spectra (which given their assumptions must work better than polynomials, although after rescaling by an overall power law their eigenfunctions are – unsurprisingly –

very similar to polynomials). Fortunately, if the foreground spectrum is analytic (as expected for synchrotron and free-free emission), polynomial fits are expected to converge exponentially fast to the true foreground spectrum as N is increased. The true value of N that will be required for future 21 cm experiments (and hence the required $k_{\|,\mathrm{min}}$) will likely be determined by how well such smooth functions can really describe the foreground.

The most difficult part of the foreground removal has been the calibration problem: even if the foreground frequency spectrum is smooth, frequency-dependent calibration errors will beat against the bright foreground and produce spurious frequency-dependent fluctuations. The problem is made more difficult by the nature of interferometry: a baseline measuring a particular Fourier mode in the (u, v)-plane at one frequency ν actually measures a different Fourier mode, $(\nu'/\nu)(u, v)$, at a neighboring ν'. Thus each pixel in (u, v)-space is actually made up from different pairs of antennas as the frequency varies, which means that the relative calibration of the gains and beams of all antennas must be known very accurately[11,43,44]. Note that the relevant gain and beam are those projected onto the sky, including phase and (of particular importance at lower frequencies) amplitude shifts induced by the ionosphere. The polarization calibration is also important: Faraday rotation is expected to produce rapidly varying structure in the polarized Stokes parameters Q and U of the Galactic synchrotron radiation, which has been observed at high Galactic latitudes at frequencies as low as 150 MHz[45] (albeit with some nondetections[46], which may be the result of lower sensitivity). The proper extrapolation of this signal to the $z \sim 20$ band is not clear, as it depends in detail on the small-scale structure of the emitting and rotating regions, but it seems likely that leakage into the Stokes I map will need to be carefully controlled. The current ($z \sim 10$) 21-cm experiments are working to achieve the required accuracy in calibration and it is hoped that they will succeed in laying the groundwork for similar efforts at higher redshift.

References

31. Komatsu, E., et al., Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation. *Astrophys. J. Supp.* **192**, 18 (2011)

32. Verner, D. A., Ferland, G. J., Korista, K. T., Yakovlev, D. G. Atomic Data for Astrophysics. II. New Analytic FITS for Photoionization Cross Sections of Atoms and Ions. *Astrophys. J.* **465**, 487 (1996)

33. Furlanetto, S. R., Stoever, S. J. Secondary ionization and heating by fast electrons. *Mon. Not. R. Astron. Soc.* **404**, 1869 (2010)

34. Machacek, M. E., Bryan, G. L., Abel, T. Simulations of Pregalactic Structure Formation with Radiative Feedback. *Astrophys. J.* **548**, 509 (2001)

35. Haiman, Z., Abel, T., Rees, M. J. The Radiative Feedback of the First Cosmological Objects. *Astrophys. J.* **534**, 11 (2000)

36. Wise, J. H., Abel, T. Suppression of H_2 Cooling in the Ultraviolet Background. *Astrophys. J.* **671**, 1559 (2007)

37. O'Shea, B. W., Norman, M. L. Population III Star Formation in a ΛCDM Universe. II. Effects of a Photodissociating Background. *Astrophys. J.* **673**, 14 (2008)

38. Gunn, J. E., Gott, J. R. On the Infall of Matter Into Clusters of Galaxies and Some Effects on Their Evolution. *Astrophys. J.* **176**, 1 (1972)

39. Ahn, K., Shapiro, P. R. Does radiative feedback by the first stars promote or prevent second generation star formation? *Mon. Not. R. Astron. Soc.* **375**, 881 (2007)

40. Johnson, J. L., Greif, T. H., Bromm, V. Local Radiative Feedback in the Formation of the First Protogalaxies. *Astrophys. J.* **665**, 85 (2007)

41. Oh, S. P., Mack, K. J. Foregrounds for 21-cm observations of neutral gas at high redshift. *Mon. Not. Roy. Astron. Soc.* **346**, 871 (2003)

42. Petrovic, N., Oh, S. P. Systematic effects of foreground removal in 21-cm surveys of reionization. *Mon. Not. R. Astron. Soc.* **413**, 2103 (2011)

43. Datta, A., Bowman, J. D., Carilli, C. L. Bright Source Subtraction Requirements for Redshifted 21 cm Measurements. *Astrophys. J.* **724**, 526 (2010)

44. Morales, M. F., Hazelton, B., Sullivan, I., Beardsley, A. Four Fundamental Foreground Power Spectrum Shapes for 21 cm Cosmology Observations. arXiv:1202.3830 (2012)

45. Bernardi, G., et al. Foregrounds for observations of the cosmological 21 cm line. II. Westerbork observations of the fields around 3C 196 and the North Celestial Pole. *Astron. Astrophys.* **522**, A67 (2010)

46. Pen, U.-L., et al., The GMRT EoR experiment: limits on polarized sky brightness at 150 MHz. *Mon. Not. R. Astron. Soc.* **399**, 181 (2009)