Foreground simulations for the LOFAR - Epoch of Reionization Experiment

V. Jelić[1][*], S. Zaroubi[1], P. Labropoulos[1], R.M. Thomas[1], G. Bernardi[1],
M. Brentjens[2], A.G. de Bruyn[1,2], B. Ciardi[3], G. Harker[1], L.V.E. Koopmans,[1],
V. Pandey[1], J. Schaye[4], S. Yatawatta[1]

[1]*Kapteyn Astronomical Institute, University of Groningen, P.O. Box 800, 9700 AV Groningen, the Netherlands*
[2]*ASTRON, Postbus 2, 7990 AA Dwingeloo, the Netherlands*
[3]*Max-Planck Institute for Astrophysics, Karl-Schwarzschild-Straße 1, 85748 Garching, Germany*
[4]*Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, the Netherlands*

4 April 2008

ABSTRACT
Future high redshift 21-cm experiments will suffer from a high degree of contamination, due both to astrophysical foregrounds and to non-astrophysical and instrumental effects. In order to reliably extract the cosmological signal from the observed data, it is essential to understand very well all data components and their influence on the extracted signal. Here we present simulated astrophysical foregrounds datacubes and discuss their possible statistical effects on the data. The foreground maps are produced assuming $5° \times 5°$ windows that match those expected to be observed by the LOFAR Epoch-of-Reionization (EoR) key science project. We show that with the expected LOFAR-EoR sky and receiver noise levels, which amount to $\approx 52\mathrm{mK}$ at 150 MHz after 300 hours of total observing time, a simple polynomial fit allows a statistical reconstruction of the signal. We also show that the polynomial fitting will work for maps with realistic yet idealised instrument response, i.e., a response that includes only a uniform uv coverage as a function of frequency and ignores many other uncertainties. Polarized galactic synchrotron maps that include internal polarization and a number of Faraday screens along the line of sight are also simulated. The importance of these stems from the fact that the LOFAR instrument, in common with all current interferometric EoR experiments has an instrumentally polarized response.

Key words: cosmology: theory, diffuse radiation, observation, radio lines: general, instrumentation: interferometers, radio continuum: general

1 INTRODUCTION

The Epoch of Reionization (hereafter, EoR), which marks the end of the Universe's 'Dark Ages', is one of the least explored epochs in cosmic evolution. Currently, there are two main observational constraints on the EoR. The first is the sudden jump in the Lyman-α optical depth in the Gunn-Peterson troughs (Gunn & Peterson 1965) observed in the Sloan Digital Sky Survey quasar spectra (Becker et al. 2001; Fan et al. 2001; Pentericci et al. 2002; White et al.q 2003; Fan et al. 2006), marking a lower limit to the redshift at which the Universe became completely ionized. The second constraint comes from the fifth year WMAP satellite data on the temperature and polarization anisotropies of the cosmic microwave background (CMB) (Spergel et al. 2007; Page et al. 2007) which gives an integral constraint on the Thomson optical depth experienced by the

CMB photons since the EoR. However, both of these observational methods provide limited information on the reionization process.

The redshifted 21-cm hyperfine transition line of neutral hydrogen is the most promising and immediately accessible method for probing the intergalactic medium (IGM) during reionization (e.g. Field 1958, 1959; Scott & Rees 1990; Kumar, Subramanian, & Padmanabhan 1995; Madau, Meiksin, & Rees 1997). Recent years have witnessed a flurry of theoretical activities to predict reionization sources and their impact on the IGM (e.g. Barkana & Loeb 2001; Loeb & Barkana 2001; Ciardi, Ferrara, & White 2003; Ciardi, Stoehr, & White 2003; Bromm & Larson 2004; Iliev et al. 2007; Zaroubi et al. 2007; Thomas & Zaroubi 2007).

Future telescopes like LOFAR[1], MWA[2], 21CMA[3] and SKA[4]

---

[*] E-mail: vjelic@astro.rug.nl

[1] http://www.lofar.org
[2] http://www.haystack.mit.edu/ast/arrays/mwa
[3] http://web.phys.cmu.edu/ past/
[4] http://www.skatelescope.org

are being designed to study the redshifted 21-cm signal from the EoR. A successful detection of this signal will help us derive the nature of the first sources and their impact on the surrounding IGM.

Unfortunately however, the cosmological EoR signal is contaminated by a slew of astrophysical and non-astrophysical components. Typically, the contamination level is orders of magnitude larger than the cosmological 21-cm signal. Thus, the primary challenge of the EoR observations will be the accurate modelling of the various data components – foregrounds, instrumental response, ionospheric disturbances, to name a few – which is essential to develop a robust signal extraction scheme.

For the foregrounds, there are currently no available data in the 115-180 MHz frequency range and 4 arcmin resolution at high Galactic latitude that would allow accurate modelling of the LOFAR-EoR foregrounds. Therefore, one has to rely on the available relevant data and extrapolate, based on theoretical arguments, into the frequency range and resolution observed by LOFAR. This paper focuses on simulating the galactic and extragalactic foregrounds that dominate the sky at frequencies of interest for the LOFAR-EoR experiment (115–215 MHz). The main foreground components are: Galactic synchrotron emission from diffuse and localised sources, Galactic thermal (free-free) emission and integrated emission from extragalactic sources (like radio galaxies and clusters). The dominant component of the foregrounds is the Galactic synchrotron emission ($\sim$70 per cent). The extragalactic emission contributes $\sim$27 per cent and Galactic free-free emission $\sim$1 per cent (Shaver et al. 1999). Although the difference between the mean amplitude of the EoR signal and the foregrounds is expected to be 4-5 orders of magnitude, an interferometer like LOFAR measures only the fluctuations which in this case are expected to be different by 'only' three orders of magnitude.

Various authors have studied the foregrounds in the context the EoR measurements. Shaver et al. (1999) have studied the diffuse synchrotron and free-free emission from our Galaxy and extragalactic sources; Di Matteo et al. (2002) and Di Matteo, Ciardi, & Miniati (2004) have considered emission from unresolved extragalactic sources at low radio frequencies; and Oh & Mack (2003) and Cooray & Furlanetto (2004) studied the effect of free-free emission from extragalactic haloes. Over the years, several methods have been explored to filter out the foregrounds. Most of the methods rely on the relative smoothness in the frequency of the foregrounds, with respect to the signal (Shaver et al. 1999; Di Matteo et al. 2002; Zaldarriaga, Furlanetto, & Hernquist 2004; Gleser, Nusser, & Benson 2007).

Santos, Cooray, & Knox (2005) have studied the foregrounds for the EoR experiment and their influence on the measurement of the 21-cm signal. In their multi-frequency analysis of the power spectra, they considered four types of foregrounds: Galactic diffuse synchrotron emission; Galactic free-free emission; extragalactic free-free emission; and extragalactic point sources. They showed that foregrounds cleaning is aided by the large scale spatial correlation, especially of the extragalactic point sources, which facilitates signal extraction to a level suitable for the EoR experiments.

The current study is part of the general effort undertaken by the LOFAR-EoR key science project to produce simulated data cubes. The pipeline under construction will simulate the LOFAR-EoR data cube that includes the simulated cosmological 21-cm signal, the galactic and extragalactic foregrounds, ionospheric effects, radio frequency interferences (RFIs) and the instrumental response. These datacubes will be used to design the observational strategy and test our signal-processing methods. Our main concern

in this paper is the simulation of the galactic and extragalactic foregrounds.

Recently, a study by Gleser, Nusser, & Benson (2007) has been conducted along lines similar to parts of the current paper. The authors test a certain signal extraction algorithm on simulated foregrounds maps in which they take most of the relevant foregrounds into account. However, there are many important differences between the two papers. First, in the Gleser, Nusser, & Benson (2007) study the assumption for the noise level in the LOFAR-EoR project, as well as the other experiments, is at least an order of magnitude too low. They assume 1 and 5 mK noise models whereas in reality the noise for the LOFAR-EoR experiment is about 50 mK. They also present a simplified model of the Galactic foregrounds that does not take into account all the spatial and frequency correlations of the Galactic diffuse synchrotron emission and underestimates that of the Galactic free-free emission, both of which are very important. In contrast to them, we also present polarized maps and introduce the LOFAR instrumental response and noise in a realistic manner.

In the foregrounds simulations presented in this paper we choose a different approach from previous groups, since our main aim is to produce the simulations that will be part of the LOFAR-EoR data pipeline. In this context our main aim is to produce foregrounds maps in the angular and frequency range of the LOFAR-EoR experiment, i.e. 3D datacubes, and then use those simulations for testing the accuracy of removal of the foregrounds. Section 6 outlines the importance of the polarized character of the foregrounds and how to model the Stokes I, Q and U polarization maps of the Galactic synchrotron emission. Section 7 presents simulated instrumental effects of the LOFAR telescope and their influence on the foregrounds maps, and Section 8 discusses a method to extract the EoR signal from the foregrounds. The paper concludes with a discussion and outlook (Section 9).

2 THE COSMOLOGICAL 21-CM SIGNAL

In radio astronomy, where the Rayleigh-Jeans law is applicable, the radiation intensity, $I(\nu)$ is expressed in terms of the brightness temperature T_b, such that:

$$I(\nu) = \frac{2\nu^2}{c^2} k T_b, \tag{1}$$

where ν is the frequency, c is the speed of light and k is Boltzmann's constant. The predicted differential brightness temperature deviation of the cosmological 21-cm signal from the cosmic microwave background radiation is given by (Field 1958, 1959; Ciardi & Madau 2003):

$$\delta T_{\rm b} = 26 \, {\rm mK} \, x_{\rm HI}(1+\delta)\left(1 - \frac{T_{\rm CMB}}{T_{\rm s}}\right)\left(\frac{\Omega_{\rm b} h^2}{0.02}\right)$$
$$\left[\left(\frac{1+z}{10}\right)\left(\frac{0.3}{\Omega_{\rm m}}\right)\right]^{1/2}. \tag{2}$$

Here T_s is the spin temperature, $x_{\rm HI}$ is the neutral hydrogen fraction, δ is the matter density contrast, $\Omega_{\rm m}$ and $\Omega_{\rm b}$ are the mass and baryon density in units of the critical density and $h = H_0/100$ [5].

In his seminal papers, Field (1958, 1959) used the quasi-static approximation to calculate the spin temperature, T_s, as a weighted

---

[5] We assume a ΛCDM Universe with $\Omega_{\rm b} = 0.04$, $\Omega_{\rm m} = 0.26$, $\Omega_{\Lambda} = 0.738$ and $H_0 = 70.8 \, {\rm k \, m s}^{-1} {\rm Mpc}^{-1}$

average of the CMB, kinetic and colour temperature (Wouthuysen 1952; Field 1958):

$$T_s = \frac{T_{CMB} + y_{kin}T_{kin} + y_\alpha T_\alpha}{1 + y_{kin} + y_\alpha}, \qquad (3)$$

where T_{CMB} is the CMB temperature and y_{kin} and y_α are the kinetic and Lyman-α coupling terms, respectively. We have assumed that the color temperature, T_α, is equal to T_{kin}. The kinetic coupling term increases with the kinetic temperature, whereas the y_α coupling term is due to the Lyman-α pumping, known also as the Wouthuysen-Field effect (Wouthuysen 1952; Field 1958). The two coupling terms are dominant under different conditions and in principle could be used to distinguish between ionization sources, e.g., between first stars, for which Lyman-α pumping is dominant, vs. first mini-quasars for which X-ray photons and therefore heating is dominant (see e.g., Nusser 2005; Kuhlen, Madau, & Montgomery 2006; Zaroubi et al. 2007; Thomas & Zaroubi 2007).

The brightness temperature of the cosmological signal used in this study is produced from a dark-matter-only N-body simulation. This simulation is used to produce a cube of the cosmological signal, i.e., the density as a function of right ascension, declination, and redshift (for more details see Thomas et al., in preparation). We assume that $T_s \gg T_{CMB}$, hence T_b follows the cosmological density and x_{HI}. We further assume that along each sight-line the neutral fraction follows the function $1/(1 + \exp(z - z_{reion}))$, where z_{reion} for each pixel (or line of sight) is set to $8.5 \pm \delta_{z=10}$ and where $\delta_{z=10}$ is the density contrast at redshift 10. We used this approach to randomize the reionization histories along different lines of sight while preserving the spatial correlations of the cosmological signals. In principle, this randomization could be drawn out of a Gaussian distribution function. Redshift 10 here is an arbitrary choice. z_{reion} along each line of sight varies in accordance with the cosmological density along that line-of-sight at z=10 and has a variance of unity centred at 8.5. Fig. 1 shows the signal data cube that we use in order to test our foregrounds filtering procedure.

3 THE LOFAR EOR KEY SCIENCE PROJECT

Currently, a number of experiments (e.g., LOFAR, 21CMA, MWA and SKA) are being designed to directly measure δT_b of the HI 21-cm hyperfine line and probe the physics of the reionization process by observing the neutral fraction of the IGM as a function of redshift. In this study, we focus on predictions for LOFAR, but our conclusions could be easily applied to the other telescopes.

Here we give some details about the LOFAR telescope and the LOFAR-EoR key science project which are relevant to the following few sections. A more detailed account will be given later in the paper when the instrumental effects are discussed (Section 7). The LOFAR array consists geographically of a compact core area and about 25 remote stations. Each remote station will be equipped with 48 High Band and 48 Low Band tiles each with 16 antennas. In the core area, with 2 km diameter, there will probably be 25 Low Band substations with 2 High Band substations attached to each of them, i.e, LOFAR will roughly have a 50 High Band substations in the core area. For the astronomy application, there will be a total of 19200 High Band antennas in the core area. Currently, the planned maximum baseline between the remote stations is roughly 100 km. The Low Band antenna will be optimized for the 30–80 MHz range while the High Band antenna will be optimized for the 115–240 MHz range.

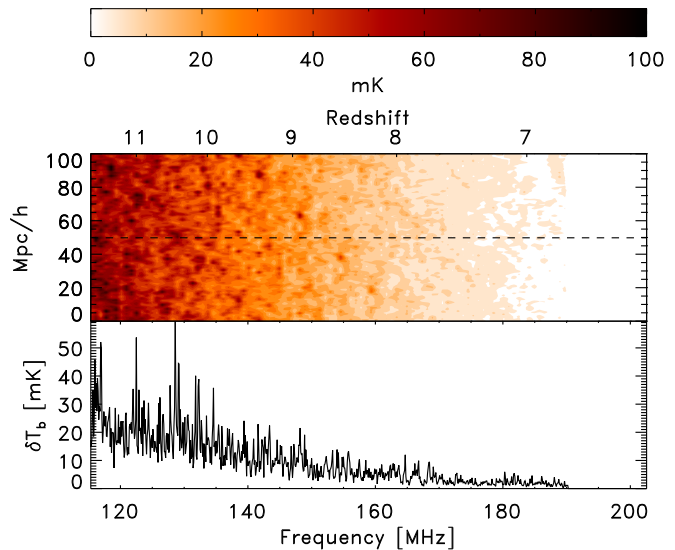The High Band antennas are sensitive enough to allow for the

Figure 1. *Simulated EoR signal assuming no reionization and $T_s \gg T_{CMB}$. The simulation box is 100 Mpc h^{-1} (comoving) a side. The upper panel shows the differential brightness temperature in a slice along the redshift/frequency direction and another spatial direction. The lower panel shows the brightness temperature as a function of redshift/frequency along a certain sight line (the dashed line in the upper panel). The resolution along the frequency direction is 10 kHz.*

detection of the brightness temperature produced by the high redshift 21 cm transitions. In this band, the LOFAR-EoR key project plans to measure the brightness fluctuations in the frequency range of 115–190 MHz, corresponding to redshift range 6-11.5 with spectral resolution of ≈ 1 MHz and spatial resolution of about ≈ 4 arcmin. As currently configured, LOFAR will be sensitive to the 21 cm emission at redshift 6.4–11.5 over 3–5 fields of 25 sq degrees each. Full operation of the LOFAR core is foreseen towards the middle of 2009.[6]

4 GALACTIC FOREGROUNDS

The Galactic foregrounds have three main contributions. The first and largest component is the Galactic diffuse synchrotron emission (GDSE), which is the dominant foreground component in the frequency range of the LOFAR-EoR experiment. The second component is radio synchrotron emission from discrete sources, mostly supernova remnants (SNRs). The third and last component is the free-free radio emission from diffuse ionized gas. This component is the weakest of the three, yet it still dominates over the cosmological component. Moreover, it has a different spectral dependence, making it very imporant in testing the signal extraction schemes that we have. In this section we describe how we simulate the contribution of each of these components to the total intensity. The polarized intensity simulations are described later on.

4.1 Galactic diffuse synchrotron emission (GDSE)

The GDSE originates from the interaction between the free electrons in the interstellar medium and the Galactic magnetic field.

---

[6] For more information, see the LOFAR web site: www.lofar.org and the LOFAR-EoR web site: www.astro.rug.nl/ LofarEoR.

Therefore the observed GDSE intensity as a function of frequency, $I(\nu)$, depends on the number density of emitting electrons, N_e, and the Galactic magnetic field component perpendicular to the line of sight, $B_\perp$:

$$I(\nu) \sim N_e B_\perp^{(\gamma+1)/2} \nu^{-(\gamma-1)/2} \qquad (4)$$

where γ is the electron spectral energy distribution power law index (Pacholczyk 1970). The intensity of the synchrotron emission as expressed in terms of the brightness temperature varies with position and frequency and its spectrum is close to a featureless power law $T_b \sim \nu^\beta$, where β is the brightness temperature spectral index, related to γ by $\beta = 2 + (\gamma - 1)/2$.

Observational data that are relevant to the LOFAR-EoR project are scarce. Landecker & Wielebinski (1970) have produced an all sky map of the total intensity of the GSDE at low radio frequencies at 150 MHz with $5°$ resolution. The other Galactic survey relevant to the LOFAR-EoR experiment is the 408 MHz survey of Haslam et al. (1982) with a resolution of $0.85°$ and of Reich & Reich (1988) at 1420 MHz with $0.95°$ resolution. In the Reich & Reich (1988) paper the authors also assume a smooth power law change in the intensity as a function of frequency which they calculate from their 1420 MHz and 408 MHz maps.

At high Galactic latitudes the minimum brightness temperature of the GDSE is about 20 K at 325 MHz with variations of the order of 2 per cent on scales from 5–30 arcmin across the sky (de Bruyn et. al 1998). At the same Galactic latitudes, the temperature spectral index β of the GDSE is about -2.55 at 100 MHz and steepens towards higher frequencies (e.g. Reich & Reich 1988; Platania et al. 1998). Furthermore, the spectral index gradually changes with position on the sky. This change appears to be caused by a variation in the spectral index along the line of sight. An appropriate standard deviation in the power law index, σ_β, in the frequency range 100–200 MHz appears to be of the order of ~ 0.1 (Shaver et al. 1999). Recent data, collected around a galaxy cluster Abell 2255 using the WSRT telescope at 350 MHz, indicate that the rms of the brightness temperature at 3 arcmin resolution could be as low as 0.1–0.3 K (Pizzo and de Bruyn, private communication). If extrapolated to 150 MHz this result implies that the rms in that region could be 1–2 mK, which is an order of magnitude smaller than the low resolution data suggest.

For the purpose of this paper we assume that the GDSE as a function of frequency is well approximated by a power law within the limited frequency range of 115–180 MHz. This is a central assumption in our simulation which is consistent with the general trend shown by the available data, namely that the change in the frequency power law index is gradual. The values we choose for the power law index are based on the high Galactic latitude regions in the Haslam et al. (1982) and Reich & Reich (1988) maps. The second assumption we make is that both the intensity and power law index of the GDSE can be spatially modelled as Gaussian random fields (GRFs). For the power spectrum of GRFs we assume a power law with 2D index $\alpha = -2.7$. The standard deviation of the GRFs is normalized to 0.4. This is consistent with the value adopted by Tegmark et al. (2000), Giardino et al. (2002) and Santos, Cooray, & Knox (2005) for the angular power spectrum index α, where $C_l \sim l^\alpha$, α varies from -2.4 to -3, and l is the harmonic number.

In contrast to the previous authors (Tegmark et al. 2000; Giardino et al. 2002; Santos, Cooray, & Knox 2005) who directly used the angular and frequency power spectrum of the GDSE for their analysis, we simulate GDSE in four dimensions (three spatial and one frequency), produce maps at each frequency and then do our
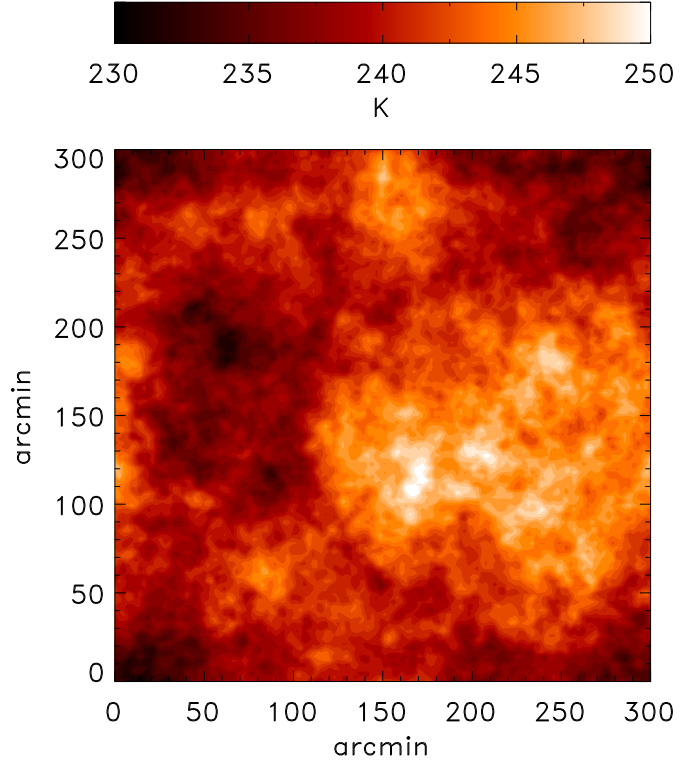

Figure 2. *Simulated map of total intensity emission of Galactic diffuse synchrotron emission with spatial spectral index $\alpha = -2.7$ and frequency spectral index $\beta = -2.55$. The angular size of the map is $5° \times 5°$, with $\sim 0.6'$ resolution. The colour bar represents the brightness temperature T_b of the Galactic diffuse synchrotron emission in Kelvin at 120 MHz.*

analysis on them. The four dimensional realisation approach has the added benefit of enabling us to account for the amplitude and temperature spectral index variations of the GDSE along the line of sight (z-coordinate). We obtain the final map of the GDSE at each frequency, ν, by integrating the GDSE amplitude ($A(x,y,z,\nu)$) along the z-coordinate:

$$T_b(x,y,\nu) = C \int A(x,y,z,\nu)\mathrm{d}z \qquad (5)$$

where $T_b(x,y,\nu)$ is the brightness temperature of the GDSE as a function of position and frequency and C is a normalization constant. $A(x,y,z,\nu)$ is dimensionless and at each frequency is defined by power law:

$$A(x,y,z,\nu) = A(x,y,z,\nu_0)\left(\frac{\nu}{\nu_0}\right)^{\beta(x,y,z,\nu)} \qquad (6)$$

where ν_0 is the reference frequency at which the normalisation is done and $\beta(x,y,z,\nu)$ is the temperature spectral index as a function of 3D position and frequency ν. The power law index β has a weak frequency dependence, also as a power law.

$A(x,y,z,\nu_0)$ and $\beta(x,y,z,\nu_0)$ of the GDSE at the normalization frequency ν_0 are modelled spatially as two Gaussian random fields with 3D power law spectrum $P(k) \sim k^\delta$. Note that the absolute value of the 3D power law index δ is $|\delta| = |\alpha| + 1$ where α is the 2D power law index mentioned above. $A(x,y,z,\nu_0)$ and $\beta(x,y,z,\nu_0)$ are normalized according to observations (the Galactic surveys mentioned above).

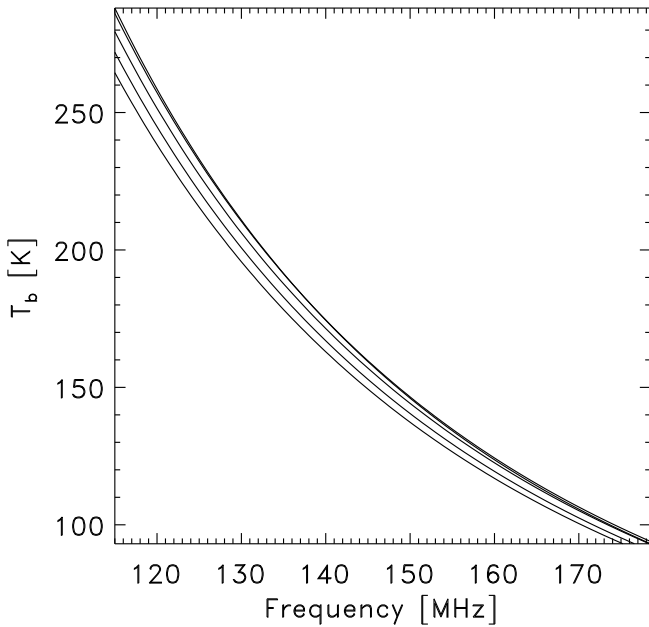For clarity, the steps we followed to produce the GDSE maps are listed below:

Figure 3. *Brightness temperature of Galactic diffuse synchrotron emission as a function of frequency, for five different lines of sight. Each line of sight has a slightly different power law index along the frequency direction as a result of the spatial and frequency variations in the temperature spectral index.*

Table 1. Angular size, flux density at 150 MHz ($S_{150\ \mathrm{MHz}}$) and spectral index (α) of the two simulated supernova remnants. Values are calculated according to the data in Green (1998) catalogue.

	angular size [arcmin]	$S_{150\ \mathrm{MHz}}$	α
SNR$_\mathrm{I}$	14×11	7.91	-0.65
SNR$_\mathrm{II}$	5×6	3.02	-0.7

4.2 Emission from SNRs

Supernova remnants are composed of expanding shells that have strong magnetic fields which are able to produce cosmic rays. As the particles escape the expanding shell, their energy decreases due to synchrotron cooling and we detect them at radio frequencies. The majority of the Galactic SNRs are within the Galactic plane but their distribution exponentially decreases with distance from the Galactic plane, z, (e.g. Caswell & Lerche 1979; Xu, Zhang, & Han 2005), that is, $N \sim \mathrm{e}^{-z}$. Moreover, due to the interaction of SNRs with the interstellar medium their radio brightness Σ decreases with an increase of their diameter D and with an increase of their height z, (e.g. Caswell & Lerche 1979), namely $\Sigma \sim D^{-3}\mathrm{e}^{-z}$.

Our goal is to calculate the expected number of known SNRs within a LOFAR-EoR observational window at high Galactic latitudes, using the known number of observed radio SNRs from the Green (1998) catalogue and assuming that their distribution follows $N \sim \mathrm{e}^{-z}$. On average, we obtain between one and two known SNRs in each $5° \times 5°$ observational window. Given the extended nature of the SNRs we include two of them in each window in order to examine the influence of bright extended sources on the calibration process and foreground removal.

The simulated SNRs assume a power law spectrum:

$$S_\nu = S_0 \left(\frac{\nu}{\nu_0} \right)^\alpha, \tag{7}$$

where S_ν is the flux density of a SNR at frequency ν, S_0 is its value at normalization frequency ν_0, and α is the spectral index.

The simulated SNRs are placed randomly on the map and their angular size, flux density and spectral index are arbitrary chosen from the Green (1998) catalogue.

Properties of the two SNRs included in our foreground simulations are shown in Table 1.

4.3 Diffuse free-free emission

The diffuse thermal (free-free) emission contributes only ~ 1 per cent of the total foregrounds within the frequency range of the LOFAR-EoR experiment (Shaver et al. 1999). It arises due to bremsstrahlung radiation in very diffuse ionized gas, with a total emission measure of about 5 pc cm^{-6} at $T_e = 8000$ K (Reynolds 1990). This gas is optically thin at frequencies above a few MHz, so its spectrum is well determined and has a temperature spectral index of $\beta = -2.1$.

At high Galactic latitudes, Hα and free-free emission of the diffuse ionized gas are both proportional to the emission measure. Therefore, the Galactic Hα survey is generally used as a tracer of the Galactic diffuse free-free emission (Smoot 1998). However, some groups also find significant correlation between free-free emission and dust emission (Kogut et al. 1996; de Oliveira-Costa et al. 1997) which can also be used as another independent tracer of the Galactic free-free emission.

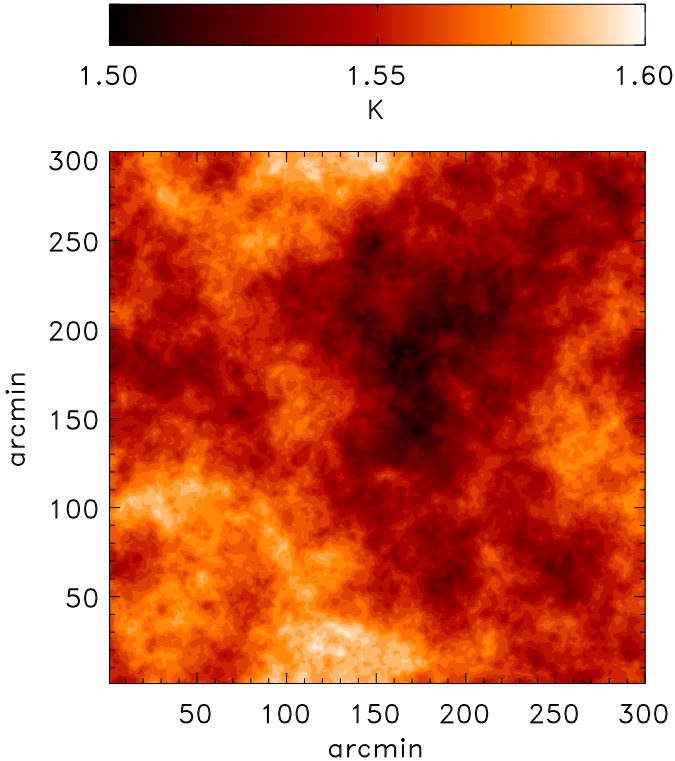In our simulations we followed Tegmark et al. (2000) and

(i) Generate the same 3D Gaussian random field for both A and β. The assumption here is that both fields have a correlated spatial distribution, which is supported by visual inspection of the high Galactic latitude portions of the Reich & Reich (1988) maps. We have also explored the possibility that A and β are independent; this has led to results very similar to the correlated case, and therefore we show only maps in which A and β are correlated.

(ii) Normalize the mean and standard deviation of $A(x, y, z, \nu_0)$ and $\beta(x, y, z, \nu_0)$ by integrating along the z direction and comparing with observations (the Galactic surveys mentioned above).

(iii) Use Eq. 6 to calculate A at each frequency.

(iv) Integrate along the z-coordinate to get the two-dimensional maps of the GDSE brightness temperature at each frequency ν (Eq. 5).

Fig. 2 shows a simulated map of the Galactic diffuse synchrotron emission according to the procedure described above, at a frequency of 120 MHz with an angular size of $5° \times 5°$ on a 512^2 grid. The mean brightness temperature of the map is $T_b = 253$K with $\sigma = 1.3$K.

In contrast to Fig. 2, which shows the spatial variations of the GDSE at one frequency, Fig. 3 shows the amplitude variations of GDSE as a function of frequency for a number of lines of sight. Each line of sight has a slightly different power law index along the frequency direction as a result of the spatial variations in the temperature spectral index. Furthermore, the brightness temperature variation for one line of sight is not a single power law but superposition of many power laws, due to the spectral index variations both spatially and in the frequency direction. Note that T_b is still a very smooth function of frequency.

Figure 4. *Simulated map of total intensity emission of Galactic diffuse free–free emission with spatial spectral index $\alpha = -3$ and frequency spectral index $\beta = -2.15$. The angular size of the map is $5° \times 5°$, with $\sim 0.6'$ resolution. The colour bar represents the brightness temperature T_b of the Galactic diffuse free–free emission in Kelvin at* 120 MHz.

Santos, Cooray, & Knox (2005) who included the Galactic diffuse free-free emission as a separate component of the Galactic foregrounds with an angular power spectrum $C_l \sim l^{-3.0}$ and frequency $T_b \sim \nu^{-2.15}$. Despite its small contribution to the foregrounds, the free-free emission is important for two reasons. Firstly, the amplitude of its spatial fluctuations is much larger than that of the EoR signal. Secondly, and more importantly, its spectral index along the frequency direction is quite different from the other foreground components and could be important in testing the algorithms for the EoR signal extraction.

To obtain the Galactic free-free emission maps we followed the same procedure as for the Galactic synchrotron emission with the additional simplification of fixing the power law index β to -2.15 across the map. The normalization of T_b was done according to observations reviewed by Smoot (1998).

Fig. 4 shows a simulated map of Galactic diffuse free-free emission at 120 MHz. The angular size of the map is $5° \times 5°$ on 512^2 grid, with the mean brightness temperature of $T_b = 1.6$K and $\sigma = 0.05$K.

5 EXTRAGALACTIC FOREGROUNDS

5.1 Radio galaxies

At the frequency range of the LOFAR-EoR experiment, bright radio sources are dominated by radio-loud galaxies, quasars and BL Lac objects (an AGN class of objects). However, at sub-mJy flux densities the contribution of late-type (star forming) galaxies, whose radio synchrotron emission originates from super-

novae rather than AGN, becomes significant (Prandoni et al. 2001; Magliocchetti et al. 2002; Sadler et al. 2002)

The bright extragalactic radio sources are normally divided into two classes based on the relative physical position of their high and low surface brightness area within the lobes. These two classes are called FRI and FRII radio sources (Fanaroff & Riley 1974; Jackson & Wall 1999). Our simulations of radio galaxies are based on the tables by Jackson (2005) of extragalactic radio source counts at 151 MHz. Jackson (2005) has used ΛCDM based models to calculate the evolution of the radio luminosity function of these sources, from which was predicted the source distributions and their number densities.

In obtaining these tables, Jackson (2005) assumed that the radio sky consists of three population types of radio sources: FRI, FRII and star forming galaxies. Moreover, it is assumed that the local radio luminosity function of star forming galaxies can be determined from the 2dFGRS-NVSS (Sadler et al. 2002) galaxy sample at 1.4 GHz. The parameterized number density and luminosity evolution of star forming galaxies is adopted from Haarsma et al. (2000). For the local radio luminosity function of FRI and FRII radio galaxies, Jackson (2005) assumed that it can be determined by exponential fitting of luminosity-dependent density evolution to the observed source counts at 151 MHz (Jackson & Wall 1999). Given the three evolving radio luminosity functions, Jackson (2005) simulated the sky at different frequencies by randomly positioning and orientating each source on the sky. The intrinsic size is also selected randomly, assuming redshift independence. The FRI and FRII sources were modelled as double-lobe structures, and the star forming galaxies as circular discs.

In our simulations of radio galaxies we adopt the three types of radio sources from Jackson (2005) and use the predicted source surface densities per $\deg^2$ for 10, 5, 2, 1 and 0.1 mJy flux density limit in order to obtain the number of sources with certain flux density per $\deg^2$. However, and in contrast to the simulations by Jackson where each source is randomly positioned, we introduce a spatial clustering of the sources. The clustering is motivated by the results of Di Matteo, Ciardi, & Miniati (2004) in which they showed that the contribution of the spatial clustering of extragalactic radio sources to the spatial fluctuations of the foregrounds, at scales $\gtrsim 1$ arcmin, is dominated by bright sources. Hence, in order to detect angular fluctuations in the cosmological 21-cm emission, efficient source removal $S \gtrsim 0.1$mJy should be carried out.

For spatial clustering of the radio galaxies we used the particularly elegant procedure of Rayleigh-Lèvy random walk proposed by Mandelbrot (1975, 1977). Starting from any arbitrary position, one places the next galaxy in a randomly chosen direction at angular distance θ, drawn from the distribution:

$$P(> \theta) = \begin{cases} (\theta/\theta_0)^\gamma & \text{for } \theta \geqslant \theta_0 \\ 1 & \text{for } \theta < \theta_0, \ \gamma > 0. \end{cases} \quad (8)$$

This is repeated many times until the correlation function of the distribution converges to the one desired. However, to compare the introduced correlation with obervational results and set the right values of γ and θ_0, the two point correlation function needs to be calculated.

The two point correlation function, $w(\theta)$, of the radio galaxy population is defined as the excess probability, over that expected for a Poisson distribution, of finding a galaxy at an angular distance θ from a given galaxy (e.g. Peebles 1980):

$$\delta P = n[1 + w(\theta)]\delta\Omega, \quad (9)$$

where δP is probability, n the mean surface density and $\delta\Omega$ a sur-

Table 2. The flux density distribution of the simulated radio galaxies on the $5° × 5°$ map (Fig. 5). The corresponding frequency is 150 MHz.

	number of sources with flux density limit				
	10 mJy	5 mJy	2 mJy	1 mJy	100 μJy
FRᵢ	20	55	122	177	1001
FRᵢᵢ	25	39	48	54	89
SF	4	38	210	419	18379

face area element. Given the very large number of galaxies that can be simulated, we adopted the simplest form for estimating $w(\theta)$, defined as:

$$w(\theta) = \frac{N_D(\theta)}{N_R(\theta)} - 1, \tag{10}$$

where N_D is the number of pairs of galaxies with separation θ in the correlated sample of galaxies and N_R is the number of pairs with the same separation θ but in a randomly distributed uncorrelated sample of galaxies. The total number of galaxies of the two samples must be the same.

Recent results on the spatial clustering of radio sources in the NRAO VLA Sky Survey (NVSS) and Faint Images of the Radio Sky at Twenty centimetres (FIRST) (Overzier et al. 2003) showed that the two point correlation function is best fitted by a double power law $w(\theta) = B\theta^{1-\gamma_B} + A\theta^{1-\gamma_A}$ with slopes of $\gamma_B = 4.4$, $\gamma_A = 1.8$ and amplitudes $B = (1.5 \pm 0.2) \times 10^{-6}$, $A = (1.0 \pm 0.2) \times 10^{-3}$. However, in this study we adopt the simpler single power law correlation function with only two parameters, $\gamma = 1.8$ and $A = 0.002$.

Fig. 5 shows a simulated map of radio galaxies with angular power law distribution. All sources are point like as the angular resolution of the LOFAR-EoR project will not be sufficient to resolve most of them. The angular extent of the map is $5° × 5°$. On the map there are in total 20690 radio galaxies: 92 per cent SF radio galaxies, 7 per cent FRI and 1 per cent FRII radio galaxies. The flux density distribution at 150 MHz of the simulated radio galaxies are shown in the Table 2, while the two point correlation fuction is $w(\theta) = 0.002\theta^{-0.8}$ ($A = 0.002$, $\gamma = 1.8$). The simulated radio galaxies assume a power law spectrum (see Eq. 7) with spectral index -0.7 (e.g. Jackson 2005).

5.2 Radio clusters

Galaxy clusters as radio sources are classified into cluster radio haloes and cluster radio relics. The former are morphologically regular diffuse sources, typically centred inside the cluster and mostly unpolarized, whereas the radio relics are typically irregular, located at the periphery of the cluster and consist mostly of polarized radio diffuse sources. Both types of cluster radio source have steep frequency spectra with $\beta \sim -3$ (see e.g. Feretti 2002, for review).

The emission in radio haloes is due to synchrotron radiation by relativistic electrons with energies of ~ 10 GeV in μG magnetic fields. The distribution of the radio haloes seems to follow closely the large scale distribution of the free-free driven X-ray emission of clusters (Govoni et al. 2001). This association is also supported by a strong correlation between the radio halo luminosity and the host cluster X-ray luminosity (e.g. Enßlin & Röttgering 2002). However, not all clusters host radio haloes. Statistically, it is found that roughly 30–40 per cent of galaxy clusters with X-ray luminosity $L_X \geqslant 10^{45}$ erg s^{-1} do host radio haloes.

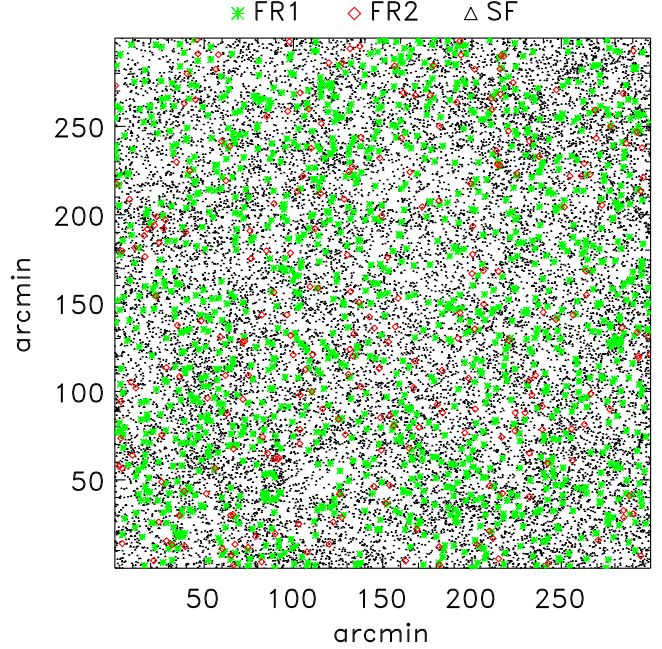In our simulations of extragalactic foregrounds maps, as a

Figure 5. *Simulated 2D spatial clustering of radio galaxies (92 per cent - star forming (SF), 7 per cent - FRI and 1 per cent - FRII radio galaxies) with angular correlation function of the form $w(\theta) = 0.002\theta^{-0.8}$ ($A = 0.002$, $\gamma = 1.8$). The flux density distribution of the galaxies is shown in the Table 2. The angular size of the map is $5° × 5°$.*

starting point for simulating radio clusters, we used the ΛCDM deep wedge cluster catalogue of The Hubble Volume Project[7]. The catalogue was obtained from an N-body simulation with one billion dark matter particles. This catalogue provides a list of clusters up to redshift $z \leqslant 4.37$ with angular coverage of 100 deg^2 (Colberg et al. 2000; Jenkins et al. 2001; Evrard et al. 2002).

In order to translate the cluster mass into X-ray luminosity L_X and then into radio luminosity L_r, we used the empirical mass–X-ray luminosity relation of Reiprich & Böhringer (2002):

$$L_X = a_X 10^{45} h_{50}^{-2} \text{ erg s}^{-1} \left(\frac{M_{R\&B}}{10^{15} h_{50}^{-1} M_\odot} \right)^{b_X} \tag{11}$$

and the X-ray–radio luminosity relation of Enßlin & Röttgering (2002):

$$L_{r,1.4 \text{ GHz}} = a_r 10^{24} h_{50}^{-2} \text{ W Hz}^{-1} \left(\frac{L_X}{10^{45} h_{50}^{-2} \text{ erg s}^{-1}} \right)^{b_r} \tag{12}$$

where $a_X = 0.449$, $b_X = 1.9$, $a_r = 2.78$ and $b_r = 1.94$ (Enßlin & Röttgering 2002). It is important to note that Eq. 12 is derived for the radio frequency at 1.4 GHz. Note also that in Eq. 11 the mass, $M_{R\&B}$, is related to the cluster real mass, M, by $M_{R\&B} \approx M\Omega_m^{1/2}$.

Since L_r is derived for 1.4 GHz, we extrapolate the luminosity of each cluster to lower frequencies, relevant to the LOFAR-EoR experiment, according to:

$$L_r(\nu) = L_{r,1.4 \text{ GHz}} \left(\frac{\nu[MHz]}{1400} \right)^\alpha, \tag{13}$$

---

[7] The cluster catalogue is part of simulations carried out by the Virgo Supercomputing Consortium using computers based at the Computing Centre of the Max-Planck Society in Garching and at the Edinburgh parallel Computing Centre. The data are publicly available at http://www.mpa-garching.mpg.de/galform/virgo/hubble.
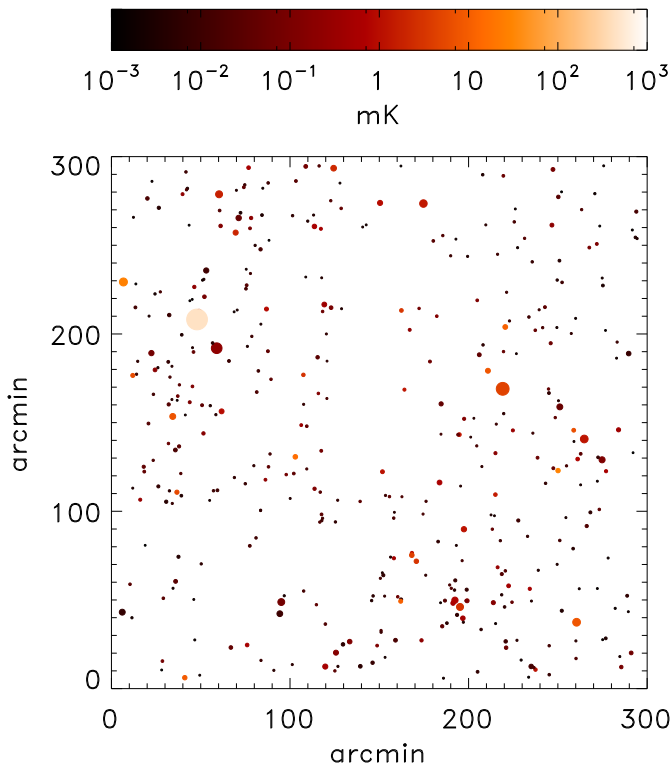
Figure 6. *Simulated 120 MHz map of radio clusters of galaxies. The angular size of the map is $5° \times 5°$. Note that the colour bar represents the brightness temperature of the clusters in logarithmic units and that the size of each cluster has been scaled by a factor of 10 for visual clarity.*

where $\alpha = -1.2$ (Kempner et al. 2004).

The angular size of the radio clusters was estimated from their physical radius and redshift. For the physical radius we used the virial radius r_{vir} calculated from the cluster mass according to (Bryan & Norman 1998),

$$M = 4\pi r_{vir}^3 \rho_{crit} \Delta_c / 3 \tag{14}$$

where Δ_c is the mean density and ρ_{crit} is the critical density at redshift z.

In order to obtain the maps of radio clusters at the angular and frequency ranges desired we first randomly choose 30 per cent of the catalogue's clusters (note that observations show that only 30–40 per cent of clusters have radio properties). The cluster masses are then used to estimate the radio luminosity of each cluster (Eq. 11, 12 & 13) and its virial radius (Eq. 14). Finally, the cluster is projected onto the simulated map according to its position and estimated size.

Fig. 6 shows a $5° \times 5°$ map of radio clusters simulated at 120 MHz. The colour bar represents the brightness temperature of the clusters in logarithmic units. The size of each cluster has been scaled by a factor of 10 for visual clarity.

6 POLARIZATION

The need for understanding the polarization response stems from two main factors. One is the geometry of the LOFAR configuration, and the other is the cross-talk contamination between the two dipoles of a LOFAR antenna. In order to detect the EoR signal, which has at best a signal to noise ratio (SNR) of ≈ 0.2 per beam

on most angular scales – assuming 300 hours of observation under the configuration specified in Section 7 – we need to fully understand the response of the LOFAR system in total intensity and polarization. As discussed before, this is vital for us in order to be able to span a dynamic range of 4–5 orders of magnitude.

Since the LOFAR antennae are fixed to the ground, the sources are tracked only by beam-forming and not by steering the antennae as in traditional radio astronomy. This implies that, depending on the position of the source on the sky, only a certain projection of the two dipoles is visible. This projection changes as the source is tracked over time. Therefore, at most times the sources in the sky see different projections of each of these dipoles. Now, if the sources/foregrounds are polarized, we immediately see that the power output from the pair of dipoles (which is the sum of the two polarized components) will vary dramatically. This has to be taken in account almost exactly during the inversion and calibration processes in order to achieve the desired dynamic range.

On the other hand, a leakage in the electronics can cause the power that is supposed to go through one of the dipoles to show up in the other (referred to as cross-talk). Although the cross-talk is small compared to the effect due to projection, we need to take it into account to eliminate any systematics.

In this paper we show a first order simulation of the Galactic diffuse polarized emission and defer a more advanced discussion on the topic to a future paper.

There are several polarization surveys of the Galactic synchrotron emission between 327 MHz and 2.7 GHz (see review, Reich 2006). The most recent one was done with the Westerbork telescope at 327 MHz, with arc minute angular resolution (Wieringa et al. 1993; Haverkorn, Katgert, & de Bruyn 2000, 2002, 2003). Its low frequency maps reveal a large amount of unusually shaped polarized small-scale structures, which have no counterpart in the total intensity. These structures are normally attributed to the coexistence of magnetic fields and thermal gas in the interstellar medium, which produce Faraday rotation at each line of sight.

The Faraday rotation depends on the observing frequency and rotation measure (RM) of the structure and it is defined along the line of sight. In order to measure Faraday rotation, observations at two or more frequencies are required. However, full understanding of the observed results could be quite difficult due to the possibility of multiple Faraday rotation layers (screens) along the line of sight and depolarization effects. Brentjens & de Bruyn (2005) introduced a new method (Faraday Rotation Measure Synthesis) that is able to cope with multiple screens and analyzes the contribution of each of these screens separately.

The Galactic diffuse synchrotron emission is linearly polarized and its polarized intensity I_P can be expressed in terms of Stokes parameters U and Q:

$$I_P = \sqrt{U^2 + Q^2} \tag{15}$$

$$\theta = \frac{1}{2} \arctan \frac{U}{Q} \tag{16}$$

where θ is the polarization angle.

In order to simulate the polarization maps of the Galactic synchrotron emission at low radio frequencies, we use a simple model of the Galactic synchrotron polarization at high frequencies (Giardino et al. 2002) in combination with Faraday screens that are introduced to account for the effects of rotation and depolarization at low frequencies. However, it is important to note that the assumptions going into Galactic synchrotron polarization model at high frequencies are not valid at low frequencies, but can be acceptable for a first estimate. In the future, we will improve on the

model itself and use results from real polarization data obtained by the LFFE[8].

The basic assumptions of the Giardino et al. (2002) model of Galactic synchrotron polarization at high frequencies are:

(i) The polarized component of Galactic synchrotron emission is proportional to the unpolarized intensity, which in terms of brightness temperature T_b is:

$$Q = fT_b \cos(2\theta) \qquad (17)$$

$$U = fT_b \sin(2\theta)$$

where f is the fraction of polarized emission (or polarization degree) and θ is the polarization angle;

(ii) The fraction of polarized radiation f is related to the temperature spectral index β (Cortiglioni & Spoelstra 1995):

$$f = \frac{3\beta - 3}{3\beta - 1} \qquad (18)$$

(iii) The polarization angle θ is given by:

$$\theta = \frac{1}{2} \arctan(x/n, y/n) \qquad (19)$$

where x, y are 2D random Gaussian fields characterized by a power law spectrum, and $n = \sqrt{x^2 + y^2}$. The power law spectral index is $\alpha = -1.7$ and its value is driven by observations, e.g. the Parkes 2.4 GHz polarimetric survey (Duncan et al. 1995).

The Faraday screens are modelled as 2D fields of rotation angles $\Delta\theta$ defined by (Rybicki & Lightman 1986):

$$\Delta\theta = RM\lambda^2 \qquad (20)$$

where λ is the wavelength of radiation and RM is the rotation measure modelled as a 2D Gaussian random field with a power law spectrum of spectral index α. Note that for the demonstrative purpose of this simulation we introduced only two Faraday screens and arbitrarily set the value of α to -2.

Therefore, in order to generate polarization maps of Galactic synchrotron emission at given frequencies, first we take the GDSE maps of total intensity (T) and temperature spectral index (β) from Sec. 4.1 and calculate the fraction of polarized radiation according to Eq. 18. Then, using Eqs. 19 & 20 we obtain polarization angle θ and Faraday rotation angle $\Delta\theta$. Finally, we use Eq. 17 to get polarization maps Q and U. The angle in Eq. 17 is the sum of $\Delta\theta$ over all Faraday screens and θ.

Fig. 7 shows the simulated 120 MHz map of polarized intensity (I_p) of diffuse Galactic synchrotron emission. The polarization angles are shown as white lines. The Stokes Q map of simulated Galactic polarized emission is shown in Fig. 8. The related Stokes U map looks very similar to the Q map.

7   INSTRUMENTAL EFFECTS

In this section we give a basic overview of the simulations of LOFAR antenna response and show how the foreground maps are seen by LOFAR. More detailed discussion on the LOFAR response and data model for the LOFAR-EoR experiment will appear in Labropoulos et al., in prep.. For the EoR experiment we plan to
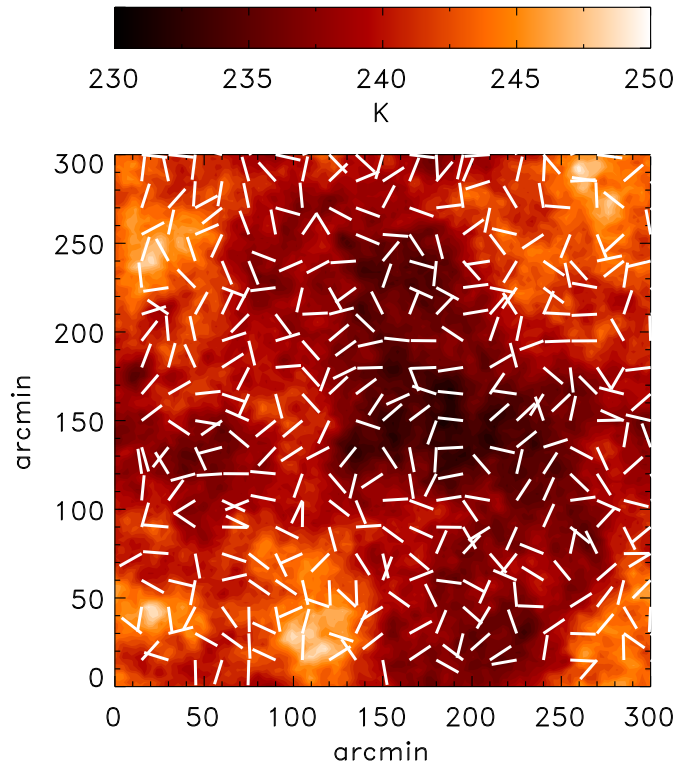
Figure 7. *Simulated* 120 MHz *map of polarized intensity (I_p) of diffuse Galactic synchrotron emission, and polarization angle (white lines). The angular size of the map is* $5° \times 5°$, *with* $\sim 0.6'$ *resolution.*

use the LOFAR core, which will consist of approximately 25 stations. Each station is further split into two substations which are separated by a few tens of metres. Each substation consists of 24 tiles, with each tile having 4×4 crossed dipoles. For our goals we assume that each of the forty-eight substations is a circular array with a radius of thirty-five metres. The stations are distributed in a randomized spiral layout and span a baseline coverage from 40 to 2000 metres.

In order to compute the true underlying visibilities, we make some simplifying assumptions. We assume that the narrow bandwidth condition holds and that the image plane effects have been calibrated to a satisfactory level. This includes station complex gain calibration, a stable primary beam, and adequate compensation for the ionospheric effects, such that the ionospheric phase introduced during the propagation of electromagnetic waves in the ionosphere and the ionospheric Faraday rotation are corrected for. For an interferometer, the measured spatial correlation of the electric fields between two antennae is called 'visibility' and is approximately given by Taylor, Carilli & Perley (1999):

$$V_f(u, v, w) = \int A(l, m, n) I_f(l, m, n)$$
$$e^{i(ul+vm+wn)} dl\,dm\,dn \qquad (21)$$

where A is the primary beam, I_f is the intensity map corresponding to frequency f, (u, v, w) are the coordinates, as seen from the source, of the tracks followed by an interferometer as the Earth rotates, and (l, m, n) are the direction cosines.

We further treat each pixel of the map as a point source with the intensity corresponding to intensity of the pixel. Note that the
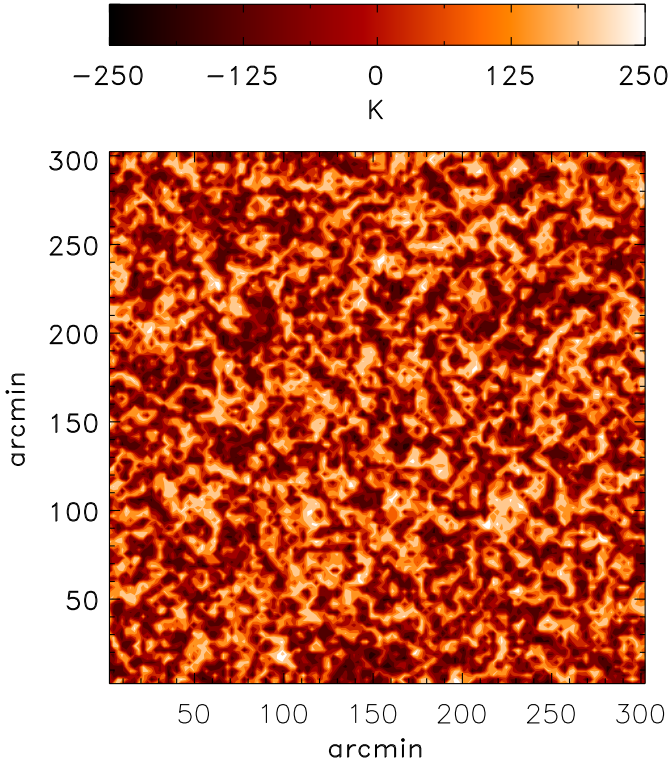
Figure 8. *Simulated* 120 MHz *Stokes Q map of polarized diffuse Galactic synchrotron emission (DGSE). The angular size of the map is* $5° \times 5°$, *with* $\sim 0.6'$ *resolution. Simulated Stokes U map of polarized GDSE looks very similar to Q map.*

equation above takes into account the sky curvature. The visibilities are sampled for all station pairs and also at different pair positions, as the Earth rotates.

We calculate the Fourier transform of the foreground model for each frequency in the above range. For every baseline and frequency, the uv tracks sample different scales of the Fourier transform of the sky at that frequency. Thus, the sampling function S becomes

$$S(u,v) = \sum_k \delta(u - u_k)\delta(v - v_k), \qquad (22)$$

where the summation is carried over all the pixels k.

We compute those tracks for each interferometer pair for 6 hours of synthesis with an averaging interval of 100 s and we then grid them on a regular grid in the uv plane. After counting how many track points fall within each grid cell, we end up with a matrix representing the naturally weighted sampling function in the uv plane. By multiplying this sampling matrix with the Fourier transform of our model sky we get the visibilities on that grid with appropriate weights. This procedure is done for each baseline pair.

$$V_f(u,v) = S \cdot \mathcal{F}I \qquad (23)$$

where $\mathcal{F}I$ is the Fourier transform of the input image and S is the sampling function.

The LOFAR visibility densities per resolution element at 150 MHz, for the LOFAR-EoR experiment, are shown in Fig. 9. The total integration time is 400 hours (100×4 h night^{-1}) with averaging time of 10 s and observing declinations $\delta = 90°$ (left panel) and $\delta = 52°$ (right panel).

The inverse Fourier transform of the sampled visibilities is

called the 'dirty' map. It is actually the sky map convolved with the Fourier transform of the sampling function, which is called the 'dirty' beam or the 'PSF'. This is a simple-minded approach to estimating the sky brightness as it uses linear operations. The approximation of the underlying brightness with the 'dirty' map is not always satisfactory, as side lobes from bright features will obscure fainter ones. In cases of low signal to noise, however – such as during the observation of the redshifted 21-cm transition line of HI – one might choose not to proceed further than this first approximation. To go beyond that we need extra information like the positivity of the intensity and compact support. The discussion of such issues is beyond the scope of this paper. This incomplete sampling of the uv plane also means that we do not measure the complete power at all scales, due to the holes in the uv coverage and its finite extent.

An example of a 'dirty' map of the diffuse components in the foregrounds is shown in the Fig. 10. The corresponding total integration time is 400 hours, with an averaging time of 100 s at 150 MHz.

The ultimate sensitivity of a receiving system is determined principally by the system noise. The discussion of the noise properties of a complex receiving system like LOFAR can be lengthy, so we concentrate for our purposes on some basic principles. The theoretical rms noise level in terms of flux density on the final image is given by

$$\sigma_{noise} = \frac{1}{\eta_s} \times \frac{SEFD}{\sqrt{N \times (N-1) \times \Delta\nu \times t_{\text{int}}}} \qquad (24)$$

where η_s is the system efficiency that accounts for electronic, digital losses, N is the number of substations, $\Delta\nu$ is the frequency bandwidth and t_{int} is the total integration time. SEFD is the System Equivalent Flux Density in mJy. The system noise we assume has two contributions. The first comes from the sky and is frequency dependent ($\approx \nu^{-2.55}$) and the second from receivers. We also assume that the distribution of noise over the map at one frequency is Gaussian. For the LOFAR core the SEFD will be around 1000 K, depending on the final design (de Bruyn et. al 2007). This means that we can reach a sensitivity of 520 mK at 150 MHz with 1 MHz bandwidth in one night of observations. Accumulating data from a hundred nights of observations brings this number down to 52 mK.

As the uv coverage varies linearly with frequency, one has to be careful in using the 'dirty' map for the extraction steps. This is because of the additional difficulties of the mixing of spatial scales introduced by the incomplete sampling. What we will show in the following section is that once the uv coverage has been made the same for each frequency – by only choosing the uv points present at each frequency – then the extraction of the signal from the foregrounds is possible. We will not, however, be taking into account calibration residuals and other systematics that might influence the data. Those issues are going to be addressed in a follow-up paper. We just use the correct level of noise in order to demonstrate our ability to statistically detect the signal in the perfect scenario of extremely precise calibration to the level of the system noise. This is done on total intensity maps only.

8   DETECTION OF THE EOR SIGNAL FROM THE FG

This section presents the results on the statistical detection of the EoR signal from 'original' and 'dirty' LOFAR-EoR data maps that include the cosmological 21cm signal, diffuse components of the foregrounds and realistic noise. By 'original' maps we mean maps before inversion or in other words maps with no calibration errors
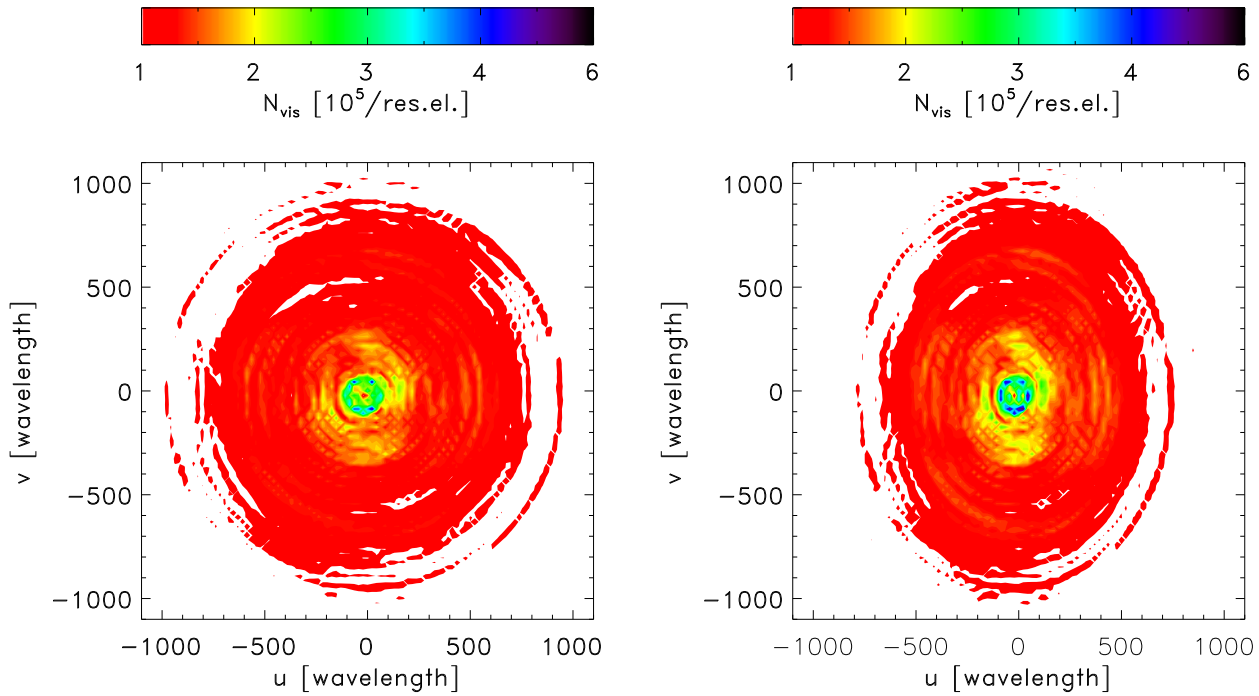
Figure 9. *The expected LOFAR visibility densities per resolution element at* 150 *MHz for* 400 *h of total integration time* (100×4 h night^{-1}) *with averaging time of* 10 s *and for observing declinations* $\delta = 90°$ *(left panel) and* $\delta = 52°$ *(right panel).*
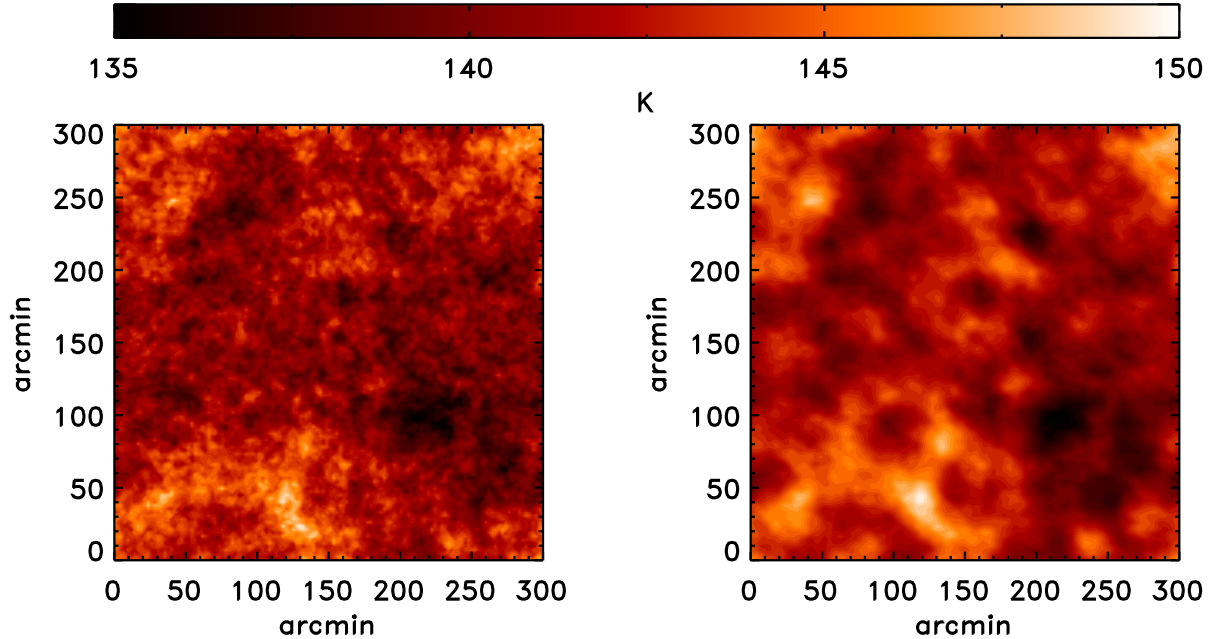

Figure 10. *Total intensity map of the simulated diffuse components in the foregrounds and its corresponding 'dirty' map after* 400 *h of total integration time with averaging of* 100 s *at* 150 MHz.

or interferometric effects. 'Dirty' maps include simplified interferometric effects but no calibration errors.

By using only diffuse components of the foregrounds (Galactic diffuse synchrotron and free-free emission and integrated emission from unresolved extragalactic sources) we assume that all resolved discrete and extended sources have been successfully re-

moved from the observed maps, without any subtraction residuals. Also note that our analysis is done on total intensity maps only. The polarized case will be considered in the follow-up paper.

The foreground and noise maps are simulated in the frequency range between 80 MHz and 200 MHz in steps of 0.5 MHz. The original maps simulated for a $5° \times 5°$ field on a 512^2 grid are re-
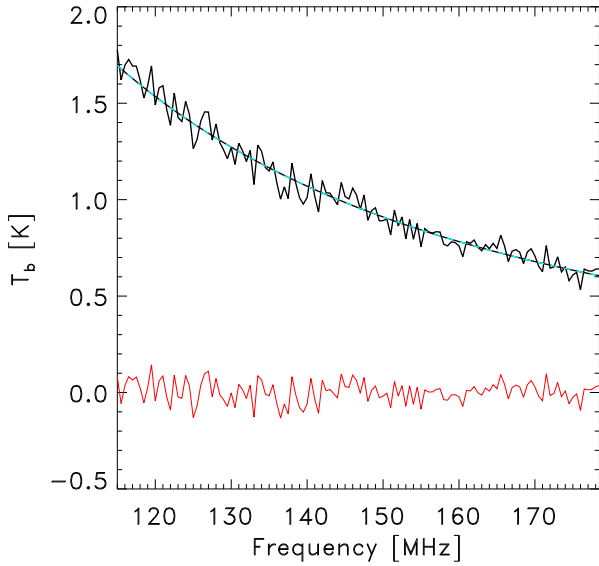
Figure 11. *One line of sight (one pixel along frequency) of the LOFAR-EoR data maps (solid black line), smooth component of the foregrounds (dashed black line), fitted foregrounds (dashed cyan line) and residuals (solid red line) after taking out the foregrounds. Note that the residuals are the sum of the EoR signal and the noise.*
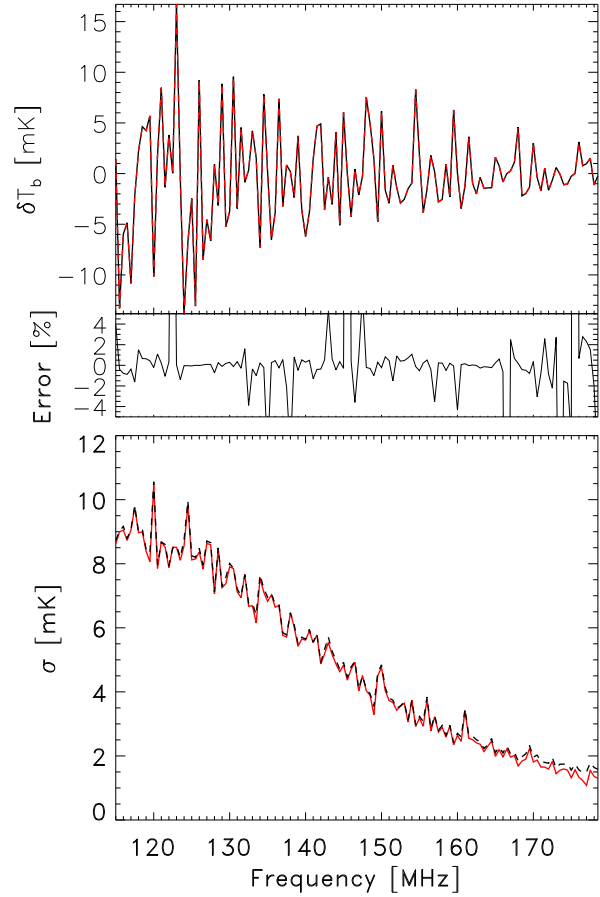

Figure 12. *Detection of the EoR signal from the simulated foreground maps, without noise: for a single random line of sight (top panel) and as a standard deviation over all lines of sight (bottom panel). The red line represents the original simulated EoR signal, and the dashed black line the extracted EoR signal.*

binned to a 128^2 grid, so that each pixel corresponds to $2.3'$ which is the resolution attained by the core of the LOFAR telescope.

The EoR maps are simulated between the frequencies of 115 MHz and 178.5 MHz in steps of 0.5 MHz, corresponding to redshifts between 11.5 and 6.5.

The mean of the EoR signal, foreground and noise maps at each frequency are set to zero since LOFAR is an interferometric instrument and measures only fluctuations around the mean. The typical variations, σ, over the map at 150 MHz are ~ 5 mK for the EoR signal, ~ 2 K for the foregrounds and ~ 52 mK for noise. Hereafter, these values are considered fiducial values for the EoR signal, foregrounds and noise.

The analysis on the LOFAR-EoR data maps can be done in two ways: firstly along the frequency direction where the foregrounds are assumed to be smooth in contrast to the EoR signal; and secondly in the spatial domain where the EoR signal and some components of the foregrounds are spatially correlated, but the noise is not. In this paper we will demonstrate statistical detection of the signal by analysis along the frequency direction, taking lines of sight (map pixels) one by one.

Fig. 11 shows one line of sight (one pixel along frequency) of the LOFAR-EoR data cube (black solid line). The first step in the extraction of the EoR signal is to take out the smooth foreground component (black dashed line). It is important to note, however, that the smooth component of the foregrounds is not a simple power law but a superposition of three power laws (Galactic synchrotron and free-free emission and integrated emission from unresolved extragalactic sources) including the fact that one of the power law indices β (Galactic synchrotron emission) changes slightly with frequency.

The simplest method for foreground removal is a polynomial fitting in logarithmic scale $(\log(T_b) - \log(\nu))$. The dashed cyan line on Fig. 11 represents the foregrounds fitted with a $3^{\rm rd}$ order polynomial in the logarithmic scale.

Fig. 12 shows a comparison between the detected and original EoR signal for randomly chosen lines of sight in the case of the fiducial foreground level and without noise. As one can see, there is an almost exact agreement: this confirms that our approach when applied to noiseless data does not introduce any systematic biases.

After taking out the fitted foregrounds from the original data, the residuals should contain only the noise and the EoR signal (red solid line on Fig. 11). However, the assumption here is that we have fitted well enough such that the residuals between the fitted and the 'real' foregrounds are smaller than the EoR signal. Otherwise the EoR signal could be fitted out if we are over-fitting, or be dominated by the foreground fitting residuals if we are under-fitting the foregrounds.

The extraction of the EoR signal from the residuals along one line of sight is an impossible task, since the level of the noise is order of magnitudes larger than EoR signal and its value is unknown for a certain pixel. However, general statistical properties of the noise (standard deviation as a function of frequency) might be determined from the experiment and be used to statistically detect the EoR signal. By statistical detection we mean determination of the standard deviation of the EoR signal over the entire map as an excess variance over the variance of the noise.

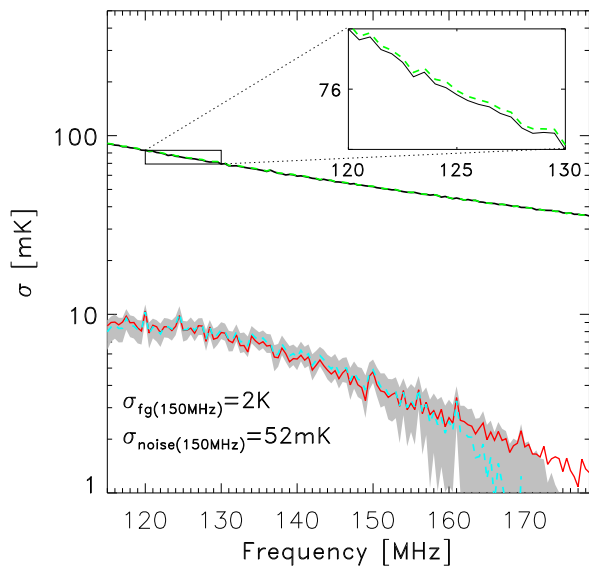Fig. 13 shows the standard deviation of residuals as a func-

Figure 13. *Statistical detection of the EoR signal from the LOFAR-EoR data maps that include diffuse components of the foregrounds and realistic noise ($\sigma_{\rm noise}$(150 MHz) = 52 mK). The solid black line represents the standard deviation (σ) of the noise as a function of frequency, the green dashed line the σ of the residuals after taking out the smooth foreground component, and the red solid line the σ of the original EoR signal. The grey shaded zone shows the 2σ detection, whereas the cyan dashed line shows the mean of the detection. Note that the y-axis is in logarithmic scale.*

Table 3. Five different sets of values for standard deviation of foregrounds (σ_{fg}[K] at 150 MHz) and of noise (σ_{noise}[mK] at 150 MHz), used for testing the EoR extraction and detection scheme. Note that case (a) represents the fiducial case.

	case (a)	case (b)	case (c)	case (d)	case (e)
σ_{fg}[K]	**2**	4	8	2	2
σ_{noise}[mK]	**52**	52	52	36	0

for the four times bigger foreground level we are able to detect the EoR signal up to 150 MHz.

Comparing Figs. 12, 13 & 14, we can see that a lower noise level increases the quality of the EoR detection, as expected. Better precision in the EoR detection with lower noise level also confirms that our foreground removal procedure works well.

Finally, in Fig. 16, we show the statistical detection of the EoR signal from the 'dirty' foregrounds + EoR signal maps without any noise. Note that the 'dirty' maps are produced with the same uv coverage for all frequencies, in order to overcome additional difficulties from the mixing of spatial scales in the frequency direction introduced by the linear frequency variation of the uv coverage and incomplete sampling. The smooth component of the foregrounds is removed by polynomial fitting to each line of sight. The most satisfactory result we get with a $6^{\rm th}$ order polynomial. A different order of polynomial from the case of the 'original' maps is required due to the spatial scale mixing over each map introduced by convolution of the map with a 'dirty' beam.

Fig. 15 shows a comparison along the frequency direction of the same pixel from the 'original' (black solid line) and 'dirty' (red solid line) foregrounds + signal maps. Note that the foregrounds are still smooth along the frequency direction of the 'dirty' maps, but the shape of the function is slightly different. The difference is due to incomplete uv coverage sampling and its finite extent, determined by the shortest and longest baselines.

The cyan dashed line in Fig. 16 shows the standard deviation, as a function of frequency, of the extracted EoR signal from 'dirty' foregrounds + EoR signal maps. The black solid line shows the standard deviation of the 'original' EoR signal maps, while the red solid line shows the standard deviation of the 'dirty' EoR signal maps. The agreement between the standard deviations of the extracted and 'dirty' EoR signals is satisfactory, while their slightly lower levels than the 'original' signal are due to the smoothing effect of the instrumental response.

tion of frequency (green dashed line), after taking out the smooth component of the foregrounds, by polynomial fitting in logarithmic scale to each line of sight. The most satisfactory result we get with a $3^{\rm rd}$ order polynomial. The black solid line represents the standard deviation of the noise. By subtracting (in quadrature) the $\sigma_{\rm noise}$ from $\sigma_{\rm residuals}$, we get the excess variance ($\sigma_{\rm EoR}$) of the EoR signal. However, in order to determine the error on the detection of the EoR signal, we conducted a Monte-Carlo simulation of the extraction of the signal. We made 1000 independent noise cube realisations and applied the signal extraction algorithm to each. The results of the simulation are shown in Fig. 13. The grey shaded zone shows the 2σ detection, whereas the cyan dashed line shows the mean of the detection. As one can see the mean of the detected EoR signal is in good agreement with the original (red solid line) up to 165 MHz. The disagreement for higher frequencies is due to over-fitting and low EoR signal level. Remember that for most of the sightlines our simulated Universe has already been ionized at this frequency, corresponding to $z \approx 7.5$ (see section 2).

In order to see the influence of the foreground and noise level on the EoR extraction and detection scheme, we repeated the same analysis on four different models. The first model has a foreground level two times bigger than fiducial and the second four times; the third has the fiducial foreground level but smaller noise level by $\sqrt{2}$; and the last one has a normal foreground level and no noise (see Table 3). Note that by fiducial foreground level and noise level we mean σ_{fg}(150 MHz) $\sim$ 2 K and $\sigma_{\rm noise}$(150 MHz) $\sim$ 52 mK. The results are shown in Figs. 12 & 14.

Comparing Figs. 13 & 14, we see the higher foreground levels decrease the quality of the EoR detection. Lower quality in the EoR detection is due to over-fitting at higher frequencies. However, even

9 DISCUSSION AND OUTLOOK

This paper presents foreground simulations tailored for the LOFAR-EoR experiment that is set to study the redshifted 21-cm hyperfine line of neutral hydrogen from the Epoch of Reionization. The foreground simulations include Galactic diffuse synchrotron and free-free emission, synchrotron emission from Galactic supernova remnants and extragalactic emission from radio galaxies and clusters. For each of the foreground components, we generate the $5° \times 5°$ field in the frequency range approximately between 115 and 180 MHz pertaining to the LOFAR-EoR.

Since the diffuse Galactic synchrotron emission is the dominant component ($\sim$ 70 per cent) we include all its observed characteristics: spatial and frequency variations of brightness temperature and its spectral index, and brightness temperature variations along
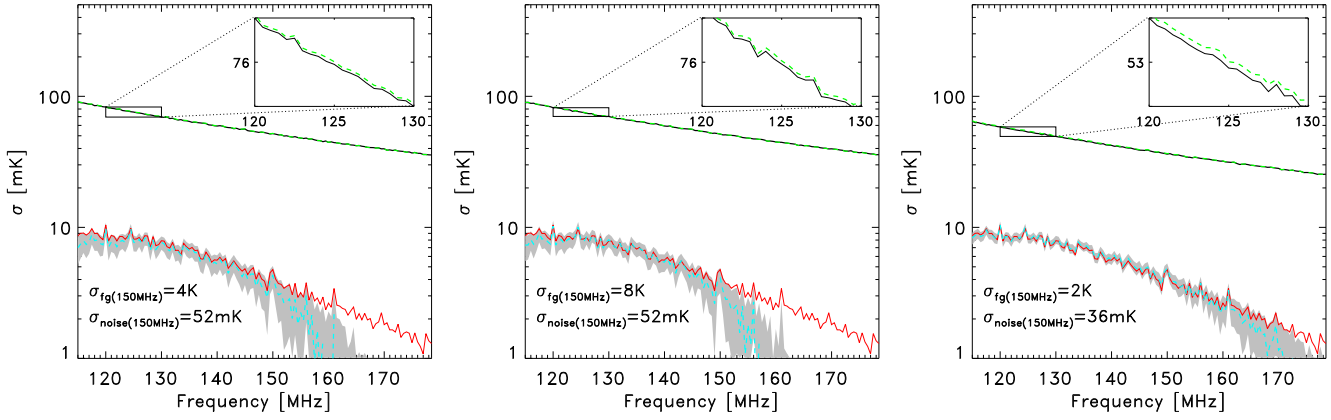
Figure 14. *Statistical detection of the EoR signal from the LOFAR-EoR data maps with foreground level two (left panel) and four (middle panel) times bigger than the fiducial foreground level, and with noise level smaller by $\sqrt{2}$ (right panel) than the fiducial noise level. Colours and line coding are the same as in Fig. 13. Note that the y-axis is in logarithmic scale.*
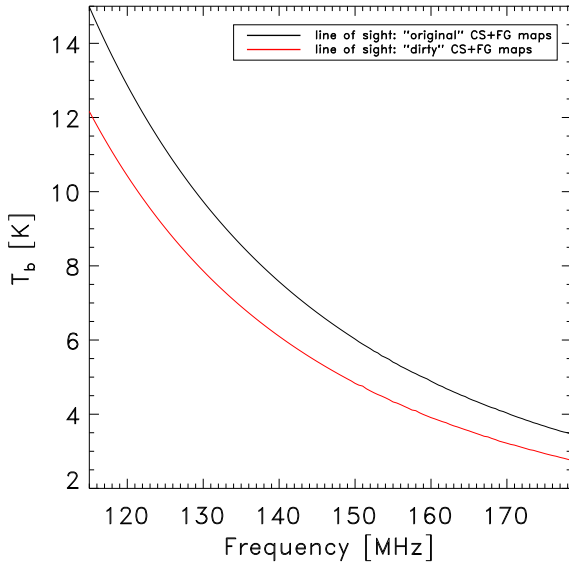

Figure 15. *One line of sight (one pixel along frequency) of the 'dirty' foreground (FG) + cosmological 21-cm signal (CS) maps (red solid line) in comparison with the same pixel along the frequency of the 'original' FG+CS maps (black solid line). The difference between these two lines is due to incomplete uv coverage and its finite extent.*

Figure 16. *Detection of the EoR signal from the simulated simplified 'dirty' foreground maps, without noise, as a standard deviation σ over all lines of sight (red solid line). The black solid line represents the σ over all lines of sight of 'original' EoR signal, while the cyan dashed line the σ over all lines of sight of 'dirty' EoR signal maps.*

the line of sight. Discrete sources of Galactic synchrotron emission are included as observed emission from supernovae remnants.

Despite the minor contribution of the Galactic free-free emission (~ 1 per cent), it is included in our simulations of the foregrounds as an individual component. It has a different temperature spectral index from Galactic synchrotron emission.

Integrated emission from extragalactic sources is decomposed into two components: emission from radio galaxies and from radio clusters. Simulations of radio galaxies are based on the source count functions at low radio frequency by Jackson (2005), for three different types of radio galaxies, namely FRI, FRII and star forming galaxies. Correlations obtained by radio galaxy surveys are used for their spatial distribution. Simulations of radio clusters are based on
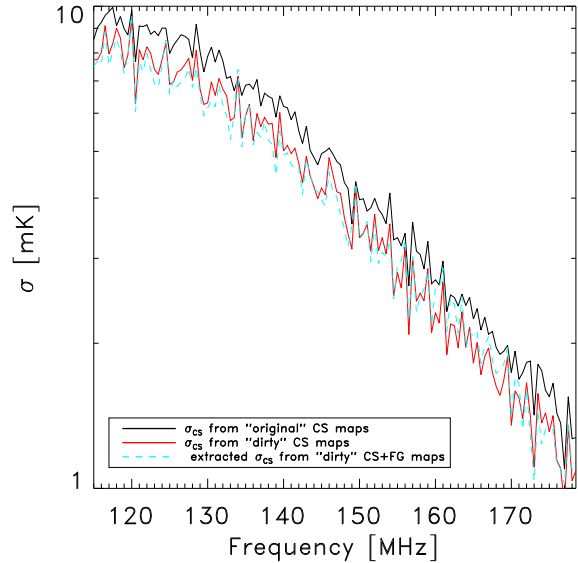
a cluster catalogue from the Virgo consortium and observed mass–X-ray luminosity and X-ray–radio luminosity relations.

Under the assumption of perfect calibration, LOFAR-EoR data maps that include the simulated cosmological 21cm signal ($\sigma_{\mathrm{EoR}}(150\ \mathrm{MHz}) \sim 5$ mK), diffuse components of the foregrounds ($\sigma_{FGs}(150\ \mathrm{MHz}) \sim 2$ K) and realistic noise ($\sigma_{\mathrm{noise}}(150\ \mathrm{MHz}) \sim 52$ mK) are produced. We refer to this set of parameters as our fiducial model. For noise we assume it has two components, the sky noise and receiver noise. The former varies with frequency as $\nu^{-2.55}$ whereas the latter is roughly frequency independent.

The extraction of the EoR signal is performed along the frequency direction, taking lines of sight (map pixels) one by one. The first step in the EoR extraction is removal of the smooth fore-

grounds component for each line of sight (see Fig. 11). In our analysis we fit a 3rd order polynomial in the logarithmic scale. However, one should be careful in choosing the order of the polynomial to perform the fitting. If the order of the polynomial is too small, the foregrounds will be under-fitted and the EoR signal could be dominated and corrupted by the fitting residuals, while if the order of the polynomial is too big, the EoR signal could be fitted out.

After foreground removal, the residuals are dominated by instrumental noise. Since the noise is unknown for each line of sight and is an order of magnitude larger than the EoR signal, it is an impossible task to directly extract the EoR signal for each line of sight. However, assuming that the statistical properties of the noise (σ_{noise}) will be known, we can use it to statistically detect the EoR signal. The statistical detection of the EoR signal is the measure of the excess variance over the entire map, σ_{EoR}^2, that should be obtained by subtracting the variance of the noise, $\sigma_{\mathrm{noise}}^2$, from that of the residuals, $\sigma_{\mathrm{residuals}}^2$.

Fig. 13 shows the results of a successful statistical detection of the EoR signal in the fiducial model of the foregrounds and noise. The detected standard deviation of the signal is in a good agreement with original signal up to 165 MHz. The disagreement for higher frequencies is due to over-fitting caused by the very weak cosmological signal at these frequencies.

In order to see the influence of the foreground and noise level on the EoR extraction and detection, the same analysis was done for models with two and four times bigger foreground levels than in the fiducial model, for a model where the noise is smaller by $\sqrt{2}$, and for a model without noise (see Tabel 3). The results are shown in Figs. 14 & 12.

In the case of higher foreground levels, the EoR signal detection is hampered by over-fitting. In the case of lower noise levels, however, the proposed EoR detection algorithm performs extremely well.

For the diffuse components of simulated foregrounds, a 'dirty' map with realistic but idealised instrumental response of LOFAR is produced (see Fig. 10). However, the signal extraction scheme we apply only to the 'dirty' maps that have been produced with a uniform uv coverage as a function of frequency and have no noise. This is due to the additional difficulties introduced by mixing of spatial scales in the frequency direction. Those issues will be discussed in a follow-up paper.

In addition to the simulations of the total brightness temperature, the polarized Galactic synchrotron emission maps are produced. Here, we follow a simple model that includes multiple Faraday screens along the line of sight (see Figs. 7 & 8). The motivation behind these simulations is that improper polarization calibration could severely contaminate the EoR signal, so future robust extraction algorithms have to take this into account.

Fig. 17 shows the angular power spectra of the simulated EoR signal (solid red line), simulated diffuse component of the foregrounds (solid black line) and three levels of noise (cyan lines) at 150 MHz. The lines are drawn as the best fit to the corresponding points. The solid cyan line represents the level of the noise ($\sigma_{\mathrm{noise}}(150\ \mathrm{MHz}) = 52\ \mathrm{mK}$) after one year of the LOFAR-EoR experiment (300 h of total observing time) with a single beam. For this case of instrumental noise and inclusion of realistic diffuse foregrounds we have shown that we are able to statistically detect the EoR signal despite the small signal to noise ratio. However, the observing plan of the LOFAR-EoR experiment (de Bruyn et. al 2007) is to observe with five independent beams, which reduces the σ_{noise} by a factor of $\sqrt{5}$ (cyan dashed-dotted line). After four years of observations (4 × 300 h) with five beams the σ_{noise}
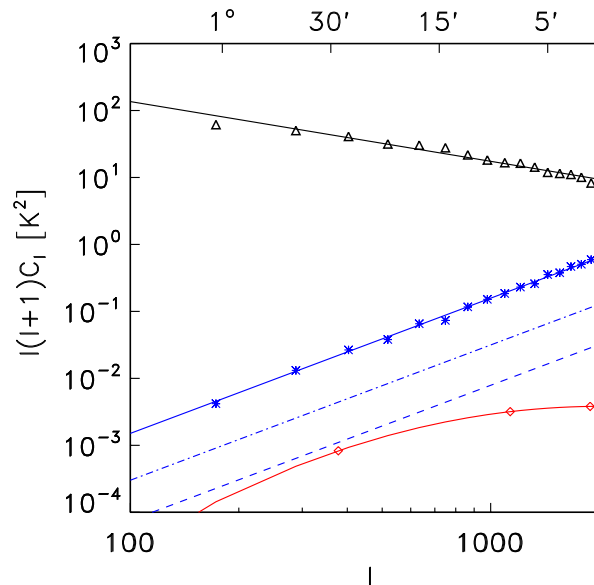

Figure 17. *Angular power spectra of the simulated EoR signal (red solid line), simulated dominant component of the foregrounds (black solid line) and three levels of noise: the solid cyan line represents noise for a single beam after one year of integration, the dashed-dotted line for five beams after one year of integration and the dashed cyan line for five beams and four years of integration. The lines are drawn as the best fit to the corresponding points.*

is reduced by a factor of $\sqrt{20}$ (cyan dashed line), which means that the signal to noise ratio is roughly two. If this is indeed the case we could expect to retrieve maps of the 21-cm EoR signal from different redshifts.

Finally we would like to emphasize that this paper is just a first step in testing and developing the analysis tools for the LOFAR-EoR experiment. Future papers will use the foregrounds simulations developed in this paper together with simulations of the EoR signal (Thomas et al., in prep.), instrumental response (Labropoulos et al., in prep.), and other non-astrophysical effects (e.g. ionosphere, RFIs, ...) in order to test all aspects of the pipeline in the LOFAR-EoR experiment.

REFERENCES

Barkana R., Loeb A., 2001, PhR, 349, 125
Becker R. H., et al., 2001, AJ, 122, 2850
Brentjens M. A., de Bruyn A. G., 2005, A&A, 441, 1217
Bromm V., Larson R. B., 2004, ARA&A, 42, 79
Bryan G. L., Norman M. L., 1998, ApJ, 495, 80
Caswell J. L., Lerche I., 1979, MNRAS, 187, 201
Ciardi B., Madau P., 2003, ApJ, 596, 1
Ciardi B., Ferrara A., White S. D. M., 2003, MNRAS, 344, L7
Ciardi B., Stoehr F., White S. D. M., 2003, MNRAS, 343, 1101

Colberg J. M., et al., 2000, MNRAS, 319, 209

Cooray A., Furlanetto S. R., 2004, ApJ, 606, L5

Cortiglioni S., Spoelstra T. A. T., 1995, A&A, 302, 1

de Bruyn G., Miley G., Rengelink R., et al., 1998, http://www.strw.LeidenUniv.nl/wenss

de Bruyn G., Zaroubi S. & Koopmans L., 2007, LOFAR-EoR Project Plan

de Oliveira-Costa A., Kogut A., Devlin M. J., Netterfield C. B., Page L. A., Wollack E. J., 1997, ApJ, 482, L17

Di Matteo T., Perna R., Abel T., Rees M. J., 2002, ApJ, 564, 576

Di Matteo T., Ciardi B., Miniati F., 2004, MNRAS, 355, 1053

Duncan A. R., Stewart R. T., Haynes R. F., Jones K. L., 1995, MNRAS, 277, 36

Enßlin T. A., Röttgering H., 2002, A&A, 396, 83

Evrard A. E., et al., 2002, ApJ, 573, 7

Fan X., et al., 2001, AJ, 122, 2833

Fan X., et al., 2006, AJ, 132, 117

Fanaroff B. L., Riley J. M., 1974, MNRAS, 167, 31P

Feretti L., 2002, IAUS, 199, 133

Field G. B., 1958, Proc, IRE, 46, 240.

Field G. B., 1959, ApJ, 129, 536

Giardino G., Banday A. J., Górski K. M., Bennett K., Jonas J. L., Tauber J., 2002, A&A, 387, 82

Gleser L., Nusser A., Benson A. J., 2007, arXiv, 712, arXiv:0712.0497

Govoni F., Enßlin T. A., Feretti L., Giovannini G., 2001, A&A, 369, 441

Green D. A., 1998, yCat, 7211, 0

Gunn J. E., Peterson B. A., 1965, ApJ, 142, 1633

Haarsma D. B., Partridge R. B., Windhorst R. A., Richards E. A., 2000, ApJ, 544, 641

Haslam C. G. T., Salter C. J., Stoffel H., Wilson W. E., 1982, A&AS, 47, 1

Haverkorn M., Katgert P., de Bruyn A. G., 2000, A&A, 356, L13

Haverkorn M., Katgert P., de Bruyn A. G., 2002, AIPC, 609, 72

Haverkorn M., Katgert P., de Bruyn A. G., 2003, A&A, 403, 1045

Iliev I. T., Mellema G., Shapiro P. R., McDonald P., Pen U.-L., 2007, ASPC, 380, 3

Jackson C. A., Wall J. V., 1999, MNRAS, 304, 160

Jackson C., 2005, PASA, 22, 36

Jenkins A., Frenk C. S., White S. D. M., Colberg J. M., Cole S., Evrard A. E., Couchman H. M. P., Yoshida N., 2001, MNRAS, 321, 372

Kogut A., Banday A. J., Bennett C. L., Gorski K. M., Hinshaw G., Reach W. T., 1996, ApJ, 460, 1

Kempner J. C., Blanton E. L., Clarke T. E., Enßlin T. A., Johnston-Hollitt M., Rudnick L., 2004, rcfg.proc, 335

Kuhlen M., Madau P., Montgomery R., 2006, ApJ, 637, L1

Kumar A., Subramanian K., Padmanabhan T., 1995, JApAS, 16, 83

Labropoulos, P., et. al, 2008, in preparation

Landecker T. L., Wielebinski R., 1970, AuJPA, 16, 1

Loeb A., Barkana R., 2001, ARA&A, 39, 19

Magliocchetti M., et al., 2002, MNRAS, 333, 100

Madau P., Meiksin A., Rees M. J., 1997, ApJ, 475, 429

Mandelbrot B., 1975, CRASM, 280, 1551

Mandelbrot B. B., 1977, fgn..book

Nusser A., 2005, MNRAS, 359, 183

Oh S. P., Mack K. J., 2003, MNRAS, 346, 871

Overzier R. A., Röttgering H. J. A., Rengelink R. B., Wilman R. J., 2003, A&A, 405, 53

Pacholczyk A. G., 1970, ranp.book,

Page L., et al., 2007, ApJS, 170, 335

Peebles P. J. E., 1980, lssu.book

Pentericci L., et al., 2002, AJ, 123, 2151

Platania P., Bensadoun M., Bersanelli M., de Amici G., Kogut A., Levin S., Maino D., Smoot G. F., 1998, ApJ, 505, 473

Prandoni I., Gregorini L., Parma P., de Ruiter H. R., Vettolani G., Zanichelli A., Wieringa M. H., Ekers R. D., 2001, A&A, 369, 787

Reich P., Reich W., 1988, A&AS, 74, 7

Reich W., 2006, astro, arXiv:astro-ph/0603465

Reiprich T. H., Böhringer H., 2002, ApJ, 567, 716

Reynolds R.J., 1990, In: Bowyer S., Leinert C. (eds.) The Galactic and Extragalactic Background Radiation. Kluwer, Dordrecht, p. 157

Rybicki G. B., Lightman A. P., 1986, rpa..book

Sadler E. M., et al., 2002, MNRAS, 329, 227

Santos M. G., Cooray A., Knox L., 2005, ApJ, 625, 575

Scott D., Rees M. J., 1990, MNRAS, 247, 510

Shaver P. A., Windhorst R. A., Madau P., de Bruyn A. G., 1999, A&A, 345, 380

Smoot G. F., 1998, astro, arXiv:astro-ph/9801121

Spergel D. N., et al., 2007, ApJS, 170, 377

G. B. Taylor, C. L. Carilli, R. A. Perley, 1999, rpa..book

Tegmark M., Eisenstein D. J., Hu W., de Oliveira-Costa A., 2000, ApJ, 530, 133

Thomas R. M., Zaroubi S., 2007, MNRAS, accepted. arXiv, 709, arXiv:0709.1657

Thomas R. M., et al., 2008, in preparation

White R. L., Becker R. H., Fan X., Strauss M. A., 2003, AJ, 126, 1

Wieringa M. H., de Bruyn A. G., Jansen D., Brouw W. N., Katgert P., 1993, A&A, 268, 215

Wild J. P., 1952, ApJ, 115, 206

Wouthuysen S. A., 1952, AJ, 57, 31

Xu J.-W., Zhang X.-Z., Han J.-L., 2005, ChJAA, 5, 442

Zaldarriaga M., Furlanetto S. R., Hernquist L., 2004, ApJ, 608, 622

Zaroubi S., Thomas R. M., Sugiyama N., Silk J., 2007, MNRAS, 375, 1269

Zygelman B., 2005, ApJ, 622, 1356

This paper has been typeset from a TeX/ LaTeX file prepared by the author.