Prospects of detection of the first sources with SKA using matched filters

Raghunath Ghara[1][*]**, T. Roy Choudhury**[2]**, Kanan K. Datta**[3]**, Garrelt Mellema**[1]**, Samir Choudhuri**[2]**, Suman Majumdar**[4] **and Sambit K. Giri**[1]

[1]Department of Astronomy & Oskar Klein Centre, AlbaNova, Stockholm University, SE-106 91 Stockholm, Sweden
[*] email: ghara.raghunath@gmail.com

[2]National Centre for Radio Astrophysics, TIFR, Post Bag 3, Ganeshkhind, Pune 411007, India

[3]Department of Physics, Presidency University, 86/1 College Street, Kolkata - 700073, India

[4]Department of Physics, Blackett Laboratory, Imperial College, London SW7 2AZ, UK

Abstract. The matched filtering technique is an efficient method to detect H II bubbles and absorption regions in radio interferometric observations of the redshifted 21-cm signal from the epoch of reionization and the Cosmic Dawn. Here, we present an implementation of this technique to the upcoming observations such as the SKA1-low for a blind search of absorption regions at the Cosmic Dawn. The pipeline explores four dimensional parameter space on the simulated mock visibilities using a MCMC algorithm. The framework is able to efficiently determine the positions and sizes of the absorption/H II regions in the field of view.

Keywords. radiative transfer; methods: numerical; galaxies: high-redshift, formation, evolution, intergalactic medium

1. Introduction

The formation of the first sources of radiation in the Universe during the "Cosmic Dawn" and their impact on the inter-galactic medium (IGM) during the Epoch of reionization (EoR) are currently not well understood. Although recently a significant numbers of galaxies, quasars have been detected through various surveys at redshift $z \gtrsim 6$, the very first sources are yet to be detected. Detection of these early sources is one of the prime aims of the modern astronomical observations. The near future space based observations such as the JWST aims to detect such high-redshift faint sources.

In addition, 21 cm signal from the neutral hydrogen in the IGM can be used as an indirect probe to detect these very first sources. Although the fact that this signal is very weak compared to the foregrounds and instrumental noise, many radio observations such as LOFAR, PAPER, MWA, GMRT have dedicated their resources to detect the signal from the EoR. The future radio telescope such as the SKA aims to observe this signal from the Cosmic Dawn as well.

During last few decades, the detectability of this H I signal is investigated in terms of various statistical quantities such as the power spectrum (Greig and Mesinger 2015; Shimabukuro and Semelin 2017), H II bubble sizes distribution (Giri *et al.* 2018) etc. using various simulations. The detectability of the signal in images has also been investigated (Mellema *et al.* 2015; Ghara *et al.* 2017a). The detectability of the signal critically depends on the properties of the sources responsible for reionization. The expected signal from the Cosmic Dawn crucially depends on the emissivities of the Lyα and X-ray photons. In general, one expect relatively stronger absorption regions around the individual/clustered sources during the Cosmic Dawn (see e.g., Ghara *et al.* 2015).
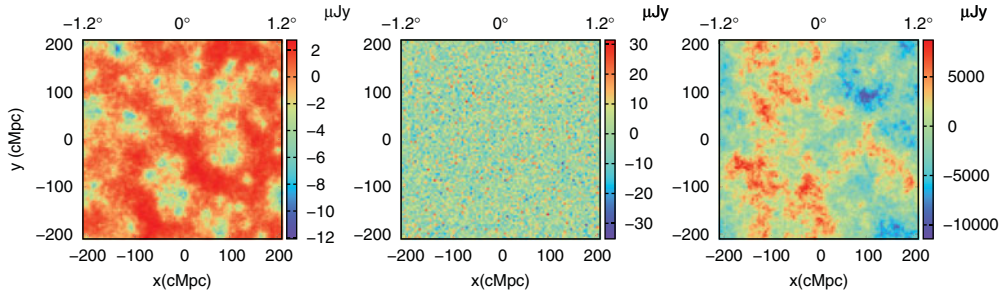
Figure 1. *Left panel*: The mean subtracted 21-cm signal map at the cosmic dawn ($z = 15$ in presence of the spin temperature fluctuations. The size of the simulation box is ~ 450 comoving megaparsec (cMpc). The resolution of the image is $2'$ (length scale of ~ 6.1 cMpc). We used a mini-QSO source model to generate the δT_b maps. *Middle panel*: System noise which correspond to 1000 h of integration with SKA1-low with a frequency resolution of 100 kHz. The rms of the noise map is 19 μJy. *Right panel*: Foregrounds from the galactic synchrotron radiation and extra-galactic point sources. The details of the simulations can be found in Ghara *et al.* (2017a).

Studies such as Datta *et al.* (2007, 2012) and Majumdar *et al.* (2012) have attempted to understand the detectability of the H II regions during the EoR for observations such as LOFAR, GMRT, MWA using matched filtering techniques. Detectability of the signal from the cosmic dawn has also been studied using similar techniques with simplified filters (Ghara *et al.* 2016). These studies show that the method is effective enough to detect the signal from the measured visibilities. While these studies have considered targeted observations where the positions of the sources are known beforehand from other observations, we intend to extend this for a blind search.

Here, we aim to develop a framework to detect the absorption regions during the Cosmic Dawn in a blind search using the observed visibilities. We use simulated visibilities which contain the signal, foregrounds and system noise and study the detectability using the previously introduced matched filtering technique. The whole machinery is coupled to a Markov chain Monte Carlo (MCMC) algorithm to explore the parameter space and to determine the position and size of the absorption regions. The same pipeline can also be used for detecting the H II bubbles during the EoR.

2. Methods

For generating gridded mock visibilities, we have used the code GRIZZLY (Ghara *et al.* 2015, 2017b) which uses one-dimensional radiative transfer scheme to generate the 21-cm brightness temperature (δT_b) maps from the outputs of a N-body simulation. A randomly chosen slice from the simulated signal at redshift 15 is shown in the left panel of Fig.1. The noise map (as shown in the middle panel of the figure) corresponds to 1000 hours of observation with SKA1-low at redshift 15 with a frequency resolution of 100 kHz. The foregrounds in the right panel of the figure are contributed from the extra-galactic point sources and galactic diffused synchrotron radiation.

One can easily notice in Fig.1 that there are isolated absorption regions in the 21-cm signal map. This expected signal is several orders of magnitude weaker than the system noise and foregrounds. However, the foregrounds are smooth along the frequency direction and thus, can in principle be subtracted out/separated. On the other hand, the system noise can be reduced using long integration time. This opens up the possibility of recovering the 21-cm signal from the observation.

Parameters	Explored range	Best fit	Standard deviation
R (cMpc)	10.0, 100.0	12.1	18.3
θ_x (arcmin)	-5.0, 40.0	31.6	10.0
θ_y (arcmin)	-40.0, -10.0	-11.6	7.8
$\Delta\nu$ (MHz)	-1.4, 1.4	0.16	0.54

Table 1. This shows the parameters of the model and the outcome from the MCMC analysis. θ_x, θ_y are the angular positions of the spherical absorption region of radius R used as filter signal. $\Delta\nu = \nu - \nu_c$ is the frequency deviation from the central channel ν_c.

The matched filtering technique as introduced in Datta *et al.* (2007) used above mentioned properties of foregrounds and system noise to construct matched filters and applied these to the observed visibilities. In this technique, the calculated estimator and corresponding errors (also the signal to noise ratio, SNR) include contributions from all baselines of the radio interferometer and frequency channels of the observation. While the previous studies have focused at targeted search of ionized regions, we apply this method to a blind search. We use the SNR as the likelihood in a MCMC algorithm and explore the parameter space to constrain the positions and size of the absorption/H II regions. As we do not have any prior information about the shape of the absorption regions in the field of view (FOV), we chose simple filter signal which in this case is a uniform spherical absorption region with radius R. Thus, we have four parameters: three position coordinates of the filter signal ($\theta_x, \theta_y, \Delta\nu$) and the radius ($R$) for the filter.

3. Results and discussions

We have tested the pipeline for an absorption region around an isolated source at redshift 15 with stellar mass $\sim 10^7$ M$_\odot$ (which correspond to a dark matter halo of mass $\sim 6 \times 10^8$ M$_\odot$) in the FOV with unknown position and size. The framework correctly determines the position of the source, while the maximum SNR achieved is ~ 9 for 1000 h of observation with SKA1-low. For example, the best fit position estimated by the technique is $\theta_x = 13.0' \pm 5.5', \theta_y = 25.3' \pm 5.5', \Delta\nu = -0.006 \pm 0.17$ MHz (1σ error) for a test source with position ($13.1', -26.2', 0.0$ MHz). Unlike the H II bubbles, the isolated absorption regions do not have sharp boundaries (see e.g., Ghara *et al.* 2016). Thus, the physical interpretation of the best fit radius (20.1 cMpc for the test source), which we will denote as the characteristic size, is not easy and needs further investigation.

In the more realistic scenario (as presented in section 2), we chose a random part from the observed field and explore the parameter space to determine the absorption regions. The details of our choice and the outcomes from the MCMC analysis are shown in Table 1. The best fit angular positions of the absorption region is denoted by the filled square in the top right panel of Fig.2. One can easily notice that this correctly corresponds to the position of the strongest absorption region in the explored region. While the best fit values of the parameters θ_x, θ_y and $\Delta\nu$ agree quite well with the position of the absorption region, there is a significant spread for the best fit value of the parameter R.

Although the framework works efficiently for determining the position and the characteristic size of the absorption/H II regions, there are some drawbacks in this method as well. First, the spherical model of the absorption region used in the filter may not be the optimal one. However, it is difficult to determine the optimal shape of the filter as we do not have prior knowledge about the signal. Second, we have used SNR as the likelihood of the MCMC algorithm. A better choice for the likelihood may improve the efficiency of this method. These issues warrant further investigations.
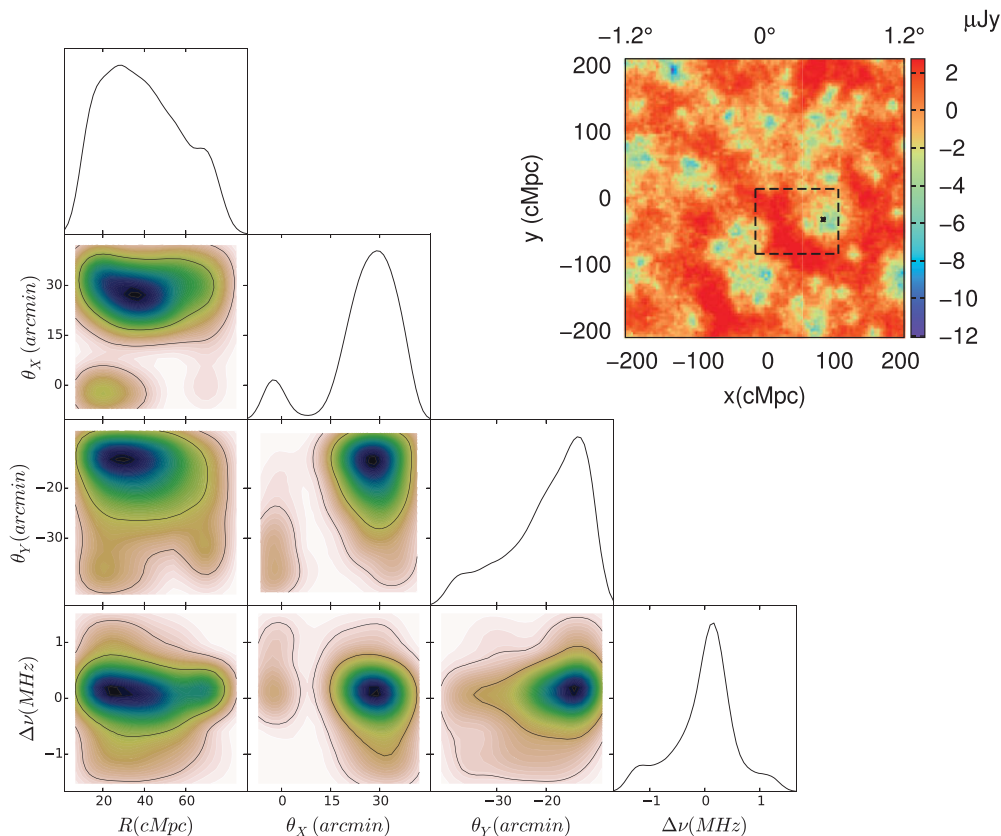
Figure 2. Constraints from the MCMC analysis over the parameter space as listed in Table 1. Two contours in the panels represent 1σ and 2σ confidence levels between the parameters respectively. The diagonal panels represent the probability distribution or the marginal distributions of each parameter. The dashed box in the right panel shows the randomly chosen region where the MCMC analysis is done. The best fit of the two angular positions have been shown by the filled square inside the box.

References

K. K. Datta, S. Bharadwaj, & T. R. Choudhury. *MNRAS*, 382: 809–818, December 2007.

K. K. Datta, M. M. Friedrich, G. Mellema, I. T. Iliev, & P. R. Shapiro. *MNRAS*, 424: 762–778, July 2012.

R. Ghara, T. R. Choudhury, & K. K. Datta. *MNRAS*, 447: 1806–1825, February 2015.

R. Ghara, T. R. Choudhury, & K. K. Datta. *MNRAS*, 460: 827–843, July 2016.

R. Ghara, T. R. Choudhury, K. K. Datta, & S. Choudhuri. *MNRAS*, 464: 2234–2248, January 2017a.

R. Ghara, G. Mellema, S. K. Giri, T. R. Choudhury, K. K. Datta, & S. Majumdar. *arXiv:1710.09397*, October 2017b.

S. K. Giri, G. Mellema, & R. Ghara. *arXiv:1801.06550*, January 2018.

B. Greig & A. Mesinger. *MNRAS*, 449: 4246–4263, June 2015.

S. Majumdar, S. Bharadwaj, & T. R. Choudhury. *MNRAS*, 426: 3178–3194, November 2012.

G. Mellema, L. Koopmans, H. Shukla, K. K. Datta, A. Mesinger, & S. Majumdar. *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, art. 10, 2015.

H. Shimabukuro & B. Semelin. *MNRAS*, 468: 3869–3877, July 2017.