Peering towards Cosmic Dawn
Proceedings IAU Symposium No. 333, 2018
V. Jelić & T. van der Hulst, eds.

doi:10.1017/S1743921318000716

Analysis of 21-cm tomographic data

Garrelt Mellema, Sambit Giri and Raghunath Ghara

Department of astronomy and Oskar Klein Centre, AlbaNova, Stockholm University, SE-10691 Stockholm, Sweden
email: garrelt.mellema@astro.su.se, sambit.giri@astro.su.se, raghunath.ghara@astro.su.se

Abstract. The future SKA1-Low radio telescope will be powerful enough to produce tomographic images of the 21-cm signal from the Epoch of Reionization. Here we address how to identify ionized regions in such data sets, taking into account the resolution and noise levels associated with SKA1-Low. We describe three methods of which one, superpixel oversegmentation, consistently performs best.

Keywords. techniques: image processing, interferometric, (cosmology:) diffuse radiation, intergalactic medium, radio lines: general

1. Introduction

The two science priorities related to the Cosmic Dawn and Epoch of Reionization which guide the design of the SKA1-Low telescope are the determination of the power spectrum of the 21-cm signal *and* imaging this signal. The latter makes SKA1-Low a unique instrument as no existing or other planned telescope is expecting to produce sky images of the 21-cm signal. The reasons why it is essential to have images as well as power spectra have been explained in detail in Mellema *et al.* (2015) and Wyithe *et al.* (2015). Apart from the ability to connect directly to observations at other wavelengths, e.g. galaxy surveys, the main reason is that the 21-cm signal during both the Cosmic Dawn and EoR is strongly non-Gaussian, which means that the power spectrum does not provide a full characterization of it.

However, for interpreting the image data in terms of (astro)physical and cosmological parameters we need to develop statistical metrics which can be compared to simulation results. One strategy is to explore the higher order versions of the power spectrum, such as the bispectrum and the trispectrum, which are sensitive to non-Gaussian signals. The contributions of Majumdar *et al.* and Hoffman *et al.* in these proceedings describe this approach in more detail. Here we take a different approach in which we develop techniques based on image analysis.

Although many different types of image analysis techniques could be considered, it is most attractive to have one which has a clear physical interpretation. When one considers the process of reionization, the most obvious choice is to try to establish the distribution of ionized regions. These of course provide a direct measurement of the progress of reionization. In addition, as we know from simulations, the sizes and shapes of ionized regions also depend on the properties of the sources and small scale absorbers. Furthermore, the distribution of ionized regions is the main source of non-Gaussianity in the 21-cm signal as the density fluctuations remain fairly close to Gaussian.

As the differential brightness temperature $\delta T_{\rm b}$ is directly proportional to the neutral hydrogen density, ionized regions are characterised by $\delta T_{\rm b} = 0$†. However, in images

† Here we assume the high spin temperature limit which removes the dependency on the spin temperature

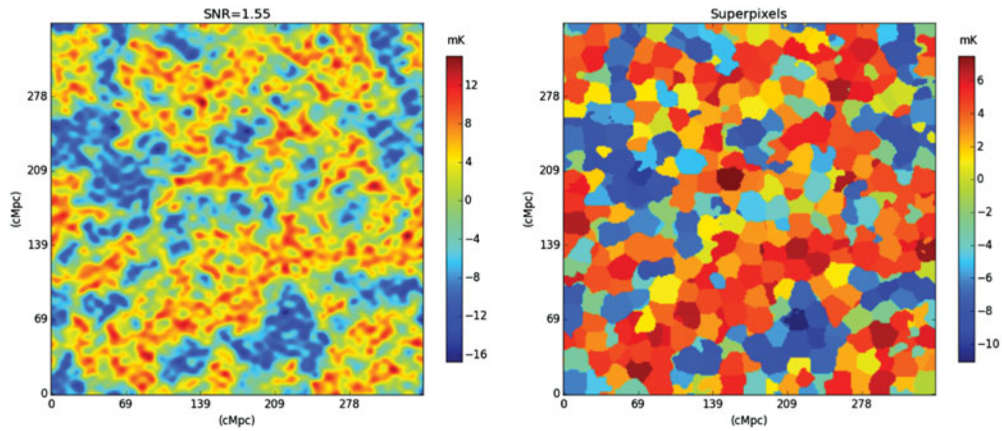

Figure 1. *Left panel*: The 21-cm signal with noise smoothed with a kernel corresponding to a maximum baseline of 2 km. The features of the ionized regions become visible at this resolution. The rms of this data set is 5.98 mK. *Right panel*: The same image after processing with the superpixel oversegmentation algorithm.

produced by an interferometer such as SKA1-Low, the absolute flux cannot be measured. Furthermore, the finite resolution of the telescope will mix ionization and density fluctuations and in addition there will be noise. All these effects complicate the identification of ionized regions in 21-cm images. The left panel of Figure 1 shows an example of a mock 21-cm image which includes these effects. The challenge is to identify the ionized regions in such an image.

2. Methods

When trying to identify ionized regions several methods have been proposed. The most straightforward one is to pick a threshold value below which an image resolution element ('pixel') is labelled as ionized. As the ionized regions have the lowest values, one choice for that threshold is 0 mK, which in an interferometric image corresponds to the mean signal of the image. This choice was proposed by Kakiichi *et al.* (2017). However, as shown in that paper this choice only works once the global ionization fraction $\langle x_{\rm HII} \rangle$ is higher than a certain value as below that neutral low density regions are identified as ionized regions. This method is also sensitive to the noise level as noise can easily put a single pixel in the wrong category.

A slightly more sophisticated method is to pick a threshold based on the Probability Distribution Function (PDF) of pixel values. From simulations we know that this PDF is bimodal with one peak corresponding to ionized regions and the other to neutral regions (see e.g. Mellema *et al.* 2006). Selecting a threshold value in between the two peaks should therefore separate ionized from neutral regions. This method was introduced in Giri *et al.* (2018a) and was found to work well as long as the PDF is clearly bimodal which is not the case during early reionization. As this method also uses pixel values, it is sensitive to the noise level.

The previous two methods can be classified as intensity-based segmentation methods as they use the values of the pixels to label them as ionized or not. A different class of segmentation methods is known as region-based as they use information from groups of pixels to achieve the same. Here we propose to use a technique known as superpixel oversegmentation which divides the data set into larger superpixels and iteratively adjusts

https://doi.org/10.1017/S1743921318000716 Published online by Cambridge University Press

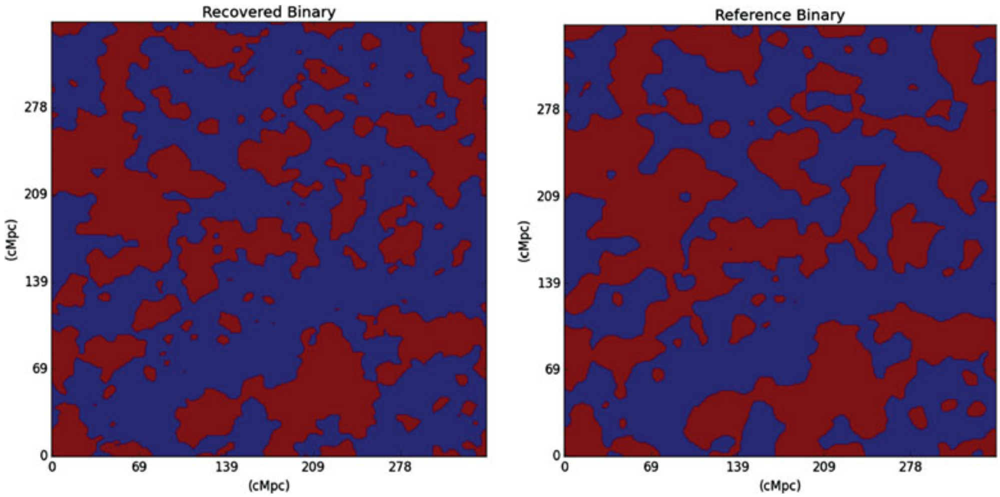

Figure 2. Ionized regions identified at $z = 6.905$. *Left panel*: from the superpixel segmentation of the 21-cm signal data shown in Fig. 1. *Right panel*: from the ionization fraction data taken from the simuation.

the boundaries of these superpixels to group pixels which have similar intensity values. The algorithm for finding these boundaries is known as Simple Linear Iterative Clustering (SLIC, Achanta *et al.* 2012). Since the ionized regions all have the same intensity value, we can expect this method to work well. After the superpixels have been identified, their average values can be determined and used to label a superpixel as ionized or not, this time using the PDF of superpixel values. As superpixels consist of several pixels, their average values are less affected by noise. The right panel of Fig. 1 shows a slice through the superpixel oversegmentation of the 21-cm image shown in the left panel. The main disadvantage of this method is that ionized regions which are only a few pixels in size will most likely not be identified. The superpixel oversegmentation method is described in more detail in Giri *et al.* (2018b).

3. Results

In Giri *et al.* (2018b) we compared the segmentation performance of the three methods from the previous section on mock observations of 1000 h with SKA1-Low. The main conclusion is that the superpixel oversegmentation method provides results which consistently follow the actual distribution of ionized regions. Figure 2 shows an example. The left panel shows in red the ionized regions identified with superpixels from the data shown in the left panel of Fig. 1 and the right panel shows the actual distribution of ionized regions as determined from the ionization fraction data from the simulation. Clearly this reconstruction is very close. The correlation coefficient of these two segmentation data sets is around 0.8. Once the ionized regions have been identified, their properties can be analysed, for example their size distribution (see e.g. Giri *et al.* 2018a) or other properties such as the their Euler characteristic (see e.g. Friedrich *et al.* 2011).

Since the superpixel method is much less sensitive to noise, it can even deal with lower noise levels than the one presented here. Tests show that even at noise levels corresponding to 300 h of observation with SKA1-Low, a reasonable reconstruction of ionized regions is possible. Obviously, for higher noise levels only the larger ionized regions can be reliably found as the signal to noise within a superpixel needs to be larger than one.

In addition the segmentation with superpixels can be used for other applications than to identify ionized regions. Several examples are presented in Giri *et al.* (2018b). As shown in that paper there can be tight correlation between the value of the average signal inside a superpixel and the production rate of ionizing photon inside the superpixel. If this relation holds for a wide range of models it can be used to reconstruct, at least partially, the distribution of sources of reionization. Other applications are estimates for the global 21-cm signal and improved measurements of the 21-cm PDF and quantities derived from that.

The use of the superpixel oversegmentation technique on 21-cm tomographic data sets therefore seems to enable a range of applications, not in the least a mapping of the shapes and sizes of ionized regions and therefore charting the progress of reionization.

References

Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, & Sabine Süsstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence*, 34 (11): 2274–2282, 2012.

M. M. Friedrich, G. Mellema, M. A. Alvarez, P. R. Shapiro, & I. T. Iliev. Topology and sizes of H II regions during cosmic reionization. *MNRAS*, 413: 1353–1372, May 2011.

Sambit K. Giri, Garrelt Mellema, Keri L. Dixon, & Ilian T. Iliev. Bubble size statistics during reionization from 21-cm tomography. *MNRAS*, 473 (3): 2949–2964, 2018a. URL + http://dx.doi.org/10.1093/mnras/stx2539.

Sambit K. Giri, Garrelt Mellema, & Raghunath Ghara. Optimal identification of hii regions during reionization in 21-cm observations. *ArXiv*, 1801.06550, 2018b.

K. Kakiichi, S. Majumdar, G. Mellema, B. Ciardi, K. L. Dixon, I. T. Iliev, V. Jelić, L. V. E. Koopmans, S. Zaroubi, & P. Busch. Recovering the H II region size statistics from 21-cm tomography. *MNRAS*, 471: 1936–1954, October 2017.

G. Mellema, I. T. Iliev, U.-L. Pen, & P. R. Shapiro. Simulating cosmic reionization at large scales - II. The 21-cm emission features and statistical signals. *MNRAS*, 372: 679–692, October 2006.

G. Mellema, L. Koopmans, H. Shukla, K. K. Datta, A. Mesinger, & S. Majumdar. HI tomographic imaging of the Cosmic Dawn and Epoch of Reionization with SKA. *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, art. 10, April 2015.

S. Wyithe, P. Geil, & H. Kim. Imaging HII Regions from Galaxies and Quasars During Reionisation with SKA. *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, art. 15, April 2015.

https://doi.org/10.1017/S1743921318000716 Published online by Cambridge University Press