URSI GASS 2021, Rome, Italy, 28 August - 4 September 2021

CLEAN removes the RFI bias in SKA-1 Low detection of Cosmic Reionization - a view from simulations

Arnab Chakraborty(1) and Abhirup Datta(1)
(1) Discipline of Astronomy, Astrophysics and Space Engineering, Indian Institute of Technology Indore, Indore 453552, India

Abstract

All low-frequency observations using radio telescopes are gravely contaminated by man-made radio frequency interference (RFI). To extract the cosmological HI 21 cm signal from this contaminated data set one has to identify and flag those spurious RFI signal persistent in time or frequency. Flagging of RFI contaminated channels results in non-uniform sampling of the instrumental bandpass response. Hence, the Fourier transformation (FT) of the data along this non-uniformly sampled frequency axis causes for spectral leakage in the Fourier domain. As a result, separating the spectrally fluctuating HI signal from spectrally smooth foregrounds in the Fourier domain becomes challenging and the estimated HI 21 cm power spectrum will have a extra bias due to this spectral leakage. Here we explore the use of a simple one dimensional *CLEAN* algorithm to reduce the spectral leakage and compare the result with other implementations. Using the simulated SKA-1 low observations, we show that *CLEAN* gives a unbiased and robust estimate of the HI 21 cm power spectrum over other techniques in the presence of significant RFI contamination.

1 Introduction

The redshifted HI 21 cm signal is a promising too to understand the evolutionary picture of the Universe starting from high-z Universe (epoch of reionization, EoR: $z \gtrsim 6$) to the present day (post-EoR era: $z \lesssim 6$ [1]. In the post-EoR era, $z \leq 6$, HI is contained within dense clumps, which are self-shielded from the ionizing radiation and intimately connected to matter distribution of the Universe and hence help to understand the large-scale structures at intermediate redshifts in the post-EoR era ($z \lesssim 6$) [2]. Along with that measurement of post-EoR HI 21 cm power spectrum can be used to study Baryon Acoustic Oscillations (BAO) and the equation of state of the dark energy [2].

The major challenge in detecting HI power spectrum is the presence of the bright synchrotron radiation from galactic and extragalactic sources. Several novel techniques have been developed to remove foregrounds [3, 4, 5, 6, 7]. Also foregrounds can be avoided in Fourier space ($\mathbf{k}_\perp, k_\parallel$), where the smooth foregrounds coupled with instrument response are localized in a 'wedge' shape region [8] and there is a 'EoR-window' in this 2D space devoid of foregrounds. This method is being widely used to detect the HI 21 cm signal from EoR [9, 10] and post-EoR regime [11]. However, all of these techniques rely on the fact that foreground is spectrally smooth, whereas HI signal has spectral feature. Hence, any oscillating feature across the frequency introduced by instrumental effects or by any post-processing methods, will pose hindrance in the process of isolating the redshifted HI signal from the foregrounds.

In all low-frequency radio interferometers, there are some flagging of RFI involved in post processing of the data. This results in non-uniform sampling of the instrumental bandpass response. During FT of measured visibilities along frequency, the non-uniform sampling of the bandpass will create fluctuating component in the Fourier conjugate space (η-domain), which is proportional to the Fourier $k_\parallel$ modes. Hence, this missing spectral information due to RFI excision introduces spectral feature across $k_\parallel$ axis and the 'EoR-window' will become contaminated by the leakage of foregrounds.

There are several ways existing in literature to deal with this issue. Recent upper-limit on EoR 21cm power spectrum by PAPER group [9] used Blackman-Harris tapering window function during Fourier transforming to delay domain to reduce foreground leakage during FFT. [12] used least square spectral analysis (LSSA) to perform line-of-sight transform from frequency to delay space instead of Fourier transform to mitigate the issues related to non-uniform bandpass shape for MWA observation. LOFAR-EoR pipeline also used the same LSSA during transformation from frequency to η-space [13, 7].

In this work, we implement the 1D *CLEAN* [14] algorithm in the η-domain to reduce the spectral leakage into the 'EoR' window. Note that, previous PAPER analysis used wide-band iterative deconvolution algorithm (WIDA) to remove the foreground components in the delay space [15]. They fit and subtract the brightest delay modes within the Horizon limit (or with some additional buffer) and use the residual delay spectrum of each baseline to estimate the HI 21cm power spectrum [16]. But, our analysis does not attempt such kind of filtering in the delay domain. Instead we use the *CLEAN* algorithm to deconvolve the delay PSF and reconstruct the entire visibility spectrum in the η-domain and then use the modes outside the Horizon

limit to estimate the HI power spectrum. This conscious choice of avoiding the subtraction of bright delay modes iteratively ensures no loss of HI signal. In Sec.2 we present the methodology of our analysis and in Sec.3 we show our findings. Finally we draw conclusion in Sec.4.

2 Methodology

To investigate the effect of RFI flagging in HI 21 cm power spectrum, we have simulated a data for the SKA-1 low array configuration (296 tiles) using the OSKAR [1] package [17]. We simulate a data set with 8 MHz bandwidth around 142 MHz ($z \sim 9$) sampled by 125 channels resulting into 64 KHz frequency resolution. For the foreground and EoR HI skymodel simulation, we follow the procedure given in [18], which incorporates detailed and realistic structure of the sky. The foreground sky model consists of diffuse galactic synchrotron radiation, free-free emission and a point source distribution. The EoR sky signal is being simulated using 21cmFAST [2] [19]. Using the foreground and EoR skymaps, we simulate visibility data set with OSKAR for each frequency channel. We then use WSCLEAN [3] to make a $5^\circ \times 5^\circ$ dirty image cube, $I^D(l,m,\nu)$ with natural weighting containing 125 channels and use this for further processing.

To estimate the power spectrum from $I^D(l,m,\nu)$, we follow the procedure as described in [20, 8, 13]. We first spatially FT the image cube back to *uv*-domain given as,

$$V(\mathbf{U},\nu) = B(\nu) \iint A(\hat{\mathbf{s}},\nu) I^D(l,m,\nu) e^{-i2\pi\nu \mathbf{U}.\hat{\mathbf{s}}} d\Omega, \quad (1)$$

,

where, $V(\mathbf{U},\nu)$ is the measured visibility of a baseline $\mathbf{U}$, $A(\hat{\mathbf{s}},\nu)$ is antenna's beam pattern assumed to be Gaussian here and B(ν) is the instrumental bandpass response. The direction on the sky is denoted by the unit vector as, $\hat{\mathbf{s}} \equiv (l,m,n)$, where l,m,n are the direction cosines towards east, north and zenith respectively with $n = \sqrt{1-l^2-m^2}$ and $d\Omega = \frac{dldm}{\sqrt{1-l^2-m^2}}$. In this analysis, we assume a flat instrumental bandpass (B(ν)) for all visibilities.

We next transformed $V(\mathbf{U},\nu)$ to the η-domain using inverse FT along frequency given as [20]:

$$V(\mathbf{U},\eta) = \int V(\mathbf{U},\nu) B(\nu) S(\nu) e^{i2\pi\nu\eta} d\nu, \quad (2)$$

where, $S(\nu)$ encodes frequency dependent sample weights that result from RFI flagging given by:

$$\begin{aligned} S(\nu) &= 0, \forall flagged\ frequency\ channel \\ &= 1, \forall channels\ without\ flagging \end{aligned} \quad (3)$$

Using the Fourier convolution theorem we can write the Eqn.2 as:

$$V(\mathbf{U},\eta) = F[V(\mathbf{U},\nu] * F[B(\nu)S(\nu)], \quad (4)$$

where F denotes the Fourier transform operator and $*$ denotes the convolution. We refer $F[B(\nu)S(\nu)]$ as the delay space point spread function (PSF) and denote it as D_{psf}. The side lobes of D_{psf} due to RFI flagging will be large and convolution of that with $F[V(\mathbf{U},\nu]$ will result in the leakage of foregrounds beyond the horizon limit into the EoR window.

One can use a window function $W(\nu)$ (Blackman-Harris window is being used here) during FT to the η-domain to reduce the spectral leakage. Here we deconvolve the D_{psf} using a 1D Hogbom *CLEAN* algorithm across the delay axis [14]. The *CLEAN* algorithm iteratively searches for maximum peaks in the delay spectrum of a visibility and then subtract a fraction of the delay space PSF (D_{psf}) from that and store the peak values as the *CLEAN* components. This process continues till a relative threshold with respect to the maximum is reached in the residual. Then we add the residual to the *CLEAN*ed components. This delay space *CLEAN*ing algorithm effectively deconvolves the delay space PSF and reduces the side lobes of the delay spectrum due to non-uniform sampling of the bandpass and gives the final delay spectra.

The cylindrically averaged 2D power spectrum is estimated from the delay spectrum following [20] :

$$P(\mathbf{k}_\perp, k_\parallel) = \left(\frac{\lambda^2}{2k_B}\right)^2 \left(\frac{X^2 Y}{\Omega B}\right) |V(\mathbf{U},\eta)|^2, \quad (5)$$

where λ is the wavelength corresponding to the bandcenter, K_B is the Boltzmann constant, Ω is sky-integral of the squared antenna primary beam response, B is the bandwidth and, X and Y are the conversion factors from angle and frequency to transverse co-moving distance (D(z)) and the co-moving depth along the line of sight, respectively ([20]). The power spectrum $P(\mathbf{k}_\perp, k_\parallel)$ is in units of $K^2(Mpc)^3$. The 3D power spectrum is estimated by spherically averaging the modes given by [8]:

$$\Delta^2(k) = \frac{K^3}{2\pi^2} < P(\mathbf{k}) >_k, \quad (6)$$

where $k = \sqrt{k_\perp^2 + k_\parallel^2}$.

[1] https://github.com/OxfordSKA/OSKAR
[2] https://21cmfast.readthedocs.io/en/latest/installation.html
[3] https://gitlab.com/aroffringa/wsclean

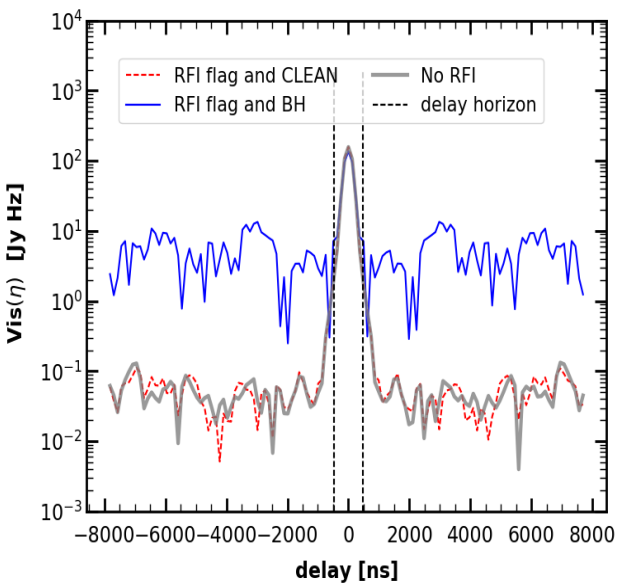

Figure 1. Delay spectrum of a single baseline. The top blue curve is the spectrum with 15% channels are being flagged and after applying BH window. The red points is the spectrum of the baseline after using 1D *CLEAN* across delay axis. The gray curve is the delay spectrum of the baseline, when there is no flagging.

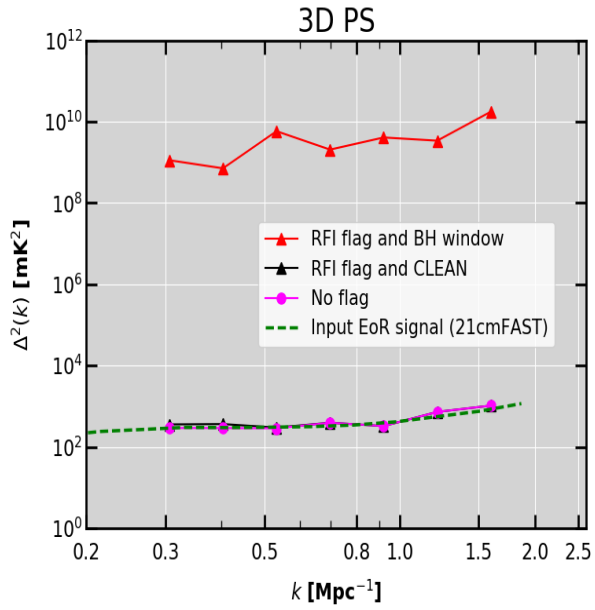

Figure 2. The estimated 3D power spectrum after spherically averaged the Fourier modes above the foreground wedge. The top red curve is for 15% random RFI flagging of the data and after the use of BH window, the black curve is after using delay space 1D *CLEAN* algorithm to the flagged data and the magenta curve is for no RFI flagging. The green dashed line is the input HI signal power spectrum estimated from 21cmFAST.

3 Results

We have considered three different scenarios - **A.** No RFI flagging, **B.** 15% channels are flagged for each visibilities and using BH window during FT along frequency axis and **C.** we use 1D *CLEAN* algorithm to the RFI flagged data set as described above.

In Fig.1 we show the delay spectrum of a single baseline. The top blue curve shows that there is significant spectral leakage beyond the horizon limit of the baseline in presence of 15% channels flag, although we have used BH spectral window. However, when we use 1D *CLEAN* in the delay space, we can restrict this spillover of the foregrounds due to RFI flagging beyond the horizon line and the resultant spectrum is nearly identical to the scenario when there is no flagging (gray line). This example with a single baseline demonstrates that in presence of RFI flagging, a simple tapering window (BH) is unable to mitigate the foreground spillover beyond the horizon limit and the 'EoR window' (above the wedge) gets contaminated significantly. However, use of *CLEAN* has the ability to restrict the spectral leakage and one can reconstruct the delay spectrum identical to the no flagging case.

We next estimate the 3D power spectrum by choosing the modes outside the horizon limit, i.e, using the modes inside the 'EoR window' and shown in Fig. 2. We compare our results with the 3D power spectrum estimated for different scenarios. We find that in presence of RFI flagging and the use of a simple BH window gives a large bias to the estimated power spectrum in comparison with the input HI 21 cm power spectrum. However, application of *CLEAN* reduces the spectral leakage due to non-uniformly sampled bandpass and the estimated power spectrum is close to the input HI signal.

This signifies that, the foreground contaminated Fourier modes due to RFI flagging can be recovered with the application of *CLEAN* algorithm in delay space and the modes within the EoR window available again for detection of the cosmological HI power spectrum.

4 Conclusion

Here, we show that flagging of RFI contaminated channels can cause for spectral leakage beyond the horizon limit of a single baseline. Hence the estimation of HI power spectrum using the Fourier modes above the horizon will have extra bias due to this foreground spillover. The reason behind is that the non-uniformly sampled bandpass due to RFI excision creates large sidelobes of the delay space PSF (point spread function) which gets convolved with the visibilities in the delay space. We show one can deconvolve this non-ideal delay psf using simple 1D *CLEAN* and restricts the spectral leakage beyond the horizon limit. We compare our result with the use of window during FT along non-uniformly sampled frequency axis. The result shows that windowing is unable to mitigate the spectral leakage

and there is a huge extra bias in the estimated 3D HI power spectrum.

This comparative analysis presented here demonstrates the effectiveness and robustness of 1D *CLEAN* to estimate the HI 21 cm power spectrum in presence of significant RFI flagging. In our future work, we plan to study the use of other algorithms to reduce the spectral leakage and compare the result with 1D *CLEAN* technique.

5 Acknowledgements

AC would like to acknowledge DST for providing INSPIRE fellowship. AD would like to acknowledge the support of EMR-II under CSIR No. 03(1461)/19. AM would like to thank Indian Institute of Technology Indore for supporting this research with Teaching Assistantship.

References

[1] Furlanetto S.R., Oh S.P., Briggs F.H. (2006), Cosmology at low frequencies: The 21 cm transition and the high-redshift Universe, , 433(4-6), 181–301, doi: 10.1016/j.physrep.2006.08.002.

[2] Bull P., Ferreira P.G., Patel P., Santos M.G. (2015), Late-time Cosmology with 21 cm Intensity Mapping Experiments, , 803(1), 21, doi:10.1088/0004-637X/803/1/21.

[3] Harker G., et. al. (2009), Non-parametric foreground subtraction for 21-cm epoch of reionization experiments, , 397(2), 1138–1152, doi:10.1111/j.1365-2966.2009.15081.x.

[4] Ghosh A., Bharadwaj S., Ali S.S., Chengalur J.N. (2011), GMRT observation towards detecting the post-reionization 21-cm signal, , 411(4), 2426–2438, doi:10.1111/j.1365-2966.2010.17853.x.

[5] Liu A., Tegmark M. (2012), How well can we measure and understand foregrounds with 21-cm experiments?, , 419(4), 3491–3504, doi:10.1111/j.1365-2966.2011.19989.x.

[6] Chapman E., et. al. (2013), The scale of the problem: recovering images of reionization with Generalized Morphological Component Analysis, , 429(1), 165–176, doi:10.1093/mnras/sts333.

[7] Mertens F.G., Ghosh A., Koopmans L.V.E. (2018), Statistical 21-cm signal separation via Gaussian Process Regression analysis, , 478(3), 3640–3652, doi: 10.1093/mnras/sty1207.

[8] Datta A., Bowman J.D., Carilli C.L. (2010), Bright Source Subtraction Requirements for Redshifted 21 cm Measurements, , 724(1), 526–538, doi: 10.1088/0004-637X/724/1/526.

[9] Kolopanis M., et al (2019), A Simplified, Lossless Reanalysis of PAPER-64, , 883(2), 133, doi: 10.3847/1538-4357/ab3e3a.

[10] Trott C.M., et. al. (2020), Deep multiredshift limits on Epoch of Reionization 21 cm power spectra from four seasons of Murchison Widefield Array observations, , 493(4), 4711–4727, doi:10.1093/mnras/staa414.

[11] Chakraborty A., et. al. (2020), First multi-redshift limits on post-Epoch of Reionization (post-EoR) 21 cm signal from z = 1.96 - 3.58 using uGMRT, ApJL, arXiv:2012.04674.

[12] Trott C.M., et. al. (2016), CHIPS: The Cosmological H I Power Spectrum Estimator, , 818(2), 139, doi: 10.3847/0004-637X/818/2/139.

[13] Patil A.H., et. al. (2017), Upper Limits on the 21 cm Epoch of Reionization Power Spectrum from One Night with LOFAR, , 838(1), 65, doi:10.3847/1538-4357/aa63e7.

[14] Högbom J.A. (1974), Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines, , 15, 417.

[15] Parsons A.R., et. al. (2014), New Limits on 21 cm Epoch of Reionization from PAPER-32 Consistent with an X-Ray Heated Intergalactic Medium at z = 7.7, , 788(2), 106, doi:10.1088/0004-637X/788/2/106.

[16] Kerrigan J.R., et. al. (2018), Improved 21 cm Epoch of Reionization Power Spectrum Measurements with a Hybrid Foreground Subtraction and Avoidance Technique, , 864(2), 131, doi:10.3847/1538-4357/aad8bb.

[17] Mort B., Dulwich F., Razavi-Ghods N., de Lera Acedo E., Grainge K. (2017), Analysing the impact of far-out sidelobes on the imaging performance of the SKA-LOW telescope, , 465(3), 3680–3692, doi: 10.1093/mnras/stw2814.

[18] Li W., et. al. (2019), Separating the EoR signal with a convolutional denoising autoencoder: a deep-learning-based method, , 485(2), 2628–2637, doi: 10.1093/mnras/stz582.

[19] Mesinger A., Furlanetto S., Cen R. (2011), 21CMFAST: a fast, seminumerical simulation of the high-redshift 21-cm signal, , 411(2), 955–972, doi: 10.1111/j.1365-2966.2010.17731.x.

[20] Morales M.F., Hewitt J. (2004), Toward Epoch of Reionization Measurements with Wide-Field Radio Observations, , 615(1), 7–18, doi:10.1086/424437.