SCIENTIFIC REVIEW

Probing early Universe through redshifted 21-cm signal: Modeling and observational challenges

ABINASH KUMAR SHAW[1,2,*] , ARNAB CHAKRABORTY[3,*], MOHD KAMRAN[4], RAGHUNATH GHARA[2], SAMIR CHOUDHURI[5,6], SK. SAIYAD ALI[7], SRIJITA PAL[1], ABHIK GHOSH[8], JAIS KUMAR[9,10], PRASUN DUTTA[9] and ANJAN KUMAR SARKAR[11,12]

[1]Department of Physics, Indian Institute of Technology Kharagpur, Kharagpur 721302, India.
[2]Astrophysics Research Centre, Open University of Israel, 4353701 Ra'anana, Israel.
[3]Department of Physics and McGill Space Institute, McGill University, Montreal, QC H3A 2T8, Canada.
[4]Department of Astronomy, Astrophysics and Space Engineering, Indian Institute of Technology Indore, Indore 453552, India.
[5]School of Physics and Astronomy, Queen Mary University of London, London E1 4NS, UK.
[6]Department of Physics, Indian Institute of Technology Madras, Chennai 600036, India.
[7]Department of Physics, Jadavpur University, Kolkata 700032, India.
[8]Department of Physics, Banwarilal Bhalotia College, Asansol 713303, India.
[9]Department of Physics, Indian Institute of Technology (Banaras Hindu University), Varanasi 221005, India.
[10]Department of Physics, K.N. Government P. G. College, Gyanpur, Bhadohi 221304, India.
[11]Raman Research Institute, Sadashivnagar, Bangalore 560080, India.
[12]National Centre for Radio Astrophysics, TIFR, Pune University Campus, Post Bag 3, Pune 411007, India.
*Corresponding authors. E-mails: abinashkumarshaw@gmail.com; arnab.phy.personal@gmail.com

Abstract. Cosmic dawn (CD) and the epoch of reionization (EoR) are the most important parts of cosmic history during which the first luminous structures emerged. These first objects heated and ionized the neutral atomic hydrogen in the intergalactic medium. The redshifted 21-cm radiation from the atomic hydrogen provides an excellent direct probe to study the evolution of neutral hydrogen (HI) and thus reveal nature of the first luminous objects, their evolution and role in this last phase transition of the Universe and formation and evolution of the structures thereafter. Direct mapping of the HI density during the CD–EoR is rather difficult with the current and forthcoming instruments due to stronger foreground and other observational contamination. The first detection of this redshifted HI signal is expected to be done through statistical estimators. Given the upmost importance of the detection and analysis of the redshifted 21-cm signal, physics of CD–EoR is considered as one of the objective of the upcoming SKA-low telescope. This paper summarizes the collective effort of Indian astronomers to understand the origin of the redshifted 21-cm signal, sources of first ionizing photons, their propagation through the IGM, various cosmological effects on the expected 21-cm signal, various statistical measures of the signal like power spectrum, bispectrum, etc. A collective effort on detection of such signal by developing estimators of the statistical measures with rigorous assessment of their expected uncertainties, various challenges like that of the large foreground emission and calibration issues are also discussed. Various versions of the detection methods discussed here have also been used in practice with the Giant Meterwave Radio Telescope with successful assessment of the foreground contamination and upper limits on the matter density in reionization and post-reionization era. The collective efforts compiled here has been a large part of the global effort to prepare proper observational technique, analysis procedure for the first light of the CD–EoR through the SKA-low.

Keywords. Intergalactic medium—cosmology: theory, observation—dark ages, reionization, first stars—diffuse radiation—large-scale structure of the Universe—methods: analytical, numerical—methods:statistical.

1. Introduction

In the standard model of cosmology, the Universe started with a hot and dense state. As the Universe expands adiabatically, it cooled down to a temperature at which the matter components of the Universe (electrons and protons) combined to form neutral atomic hydrogen (HI). This occurred at redshift $z \approx 1100$ (e.g., Peebles 1993) when radiations got decoupled from matter and traveled freely across the inter-galactic medium (IGM). Afterglow of the surface of last scattering is observed as the cosmic microwave background (CMB). The CMB observations indicate that tiny fluctuations ($\sim 10^{-6}$) were present in the matter density field during matter-radiation decoupling. Afterwards, the Universe remained completely dark and neutral for a long time till these fluctuations have grown sufficiently to form first luminous objects which lit up the Universe at $z \approx 20$ (see e.g., Ghara *et al.* 2015a,b). This marks the cosmic dawn (CD) during which the X-rays from the first objects (stars, quasars, galaxies, etc.) heated up the IGM. Sooner, the ionizing UV photons from these sources also started leaking into the IGM and gradually ionized all the neutral hydrogen (HI) in the IGM. This last phase change in the ionization state of the Universe is coined as the epoch of reionization (EoR). Afterwards, the Universe remained ionized on cosmological scales as we see it today. This last stage is also known as post-reionization epoch. CMB observations give us a clearer picture of how the Universe was in its primary stages and observations like the galaxy-redshift surveys show the present state of the Universe. However, we know only a little about the CD–EoR.

Our present knowledge of CD–EoR is guided by a few indirect observations, such as measurements of Thomson scattering optical depth from CMB observations (e.g., Douspis *et al.* 2015; Planck Collaboration *et al.* 2016, 2020), observation of Gunn–Peterson troughs in the high-z quasars (e.g., Becker *et al.* 2001; Fan *et al.* 2002, 2006) and measurements of luminosity functions from the high-z Ly-α emitters (e.g., Hu *et al.* 2002; Malhotra & Rhoads 2004; Ota *et al.* 2007; Ouchi *et al.* 2010; Choudhury *et al.* 2015). These observations commonly suggest that EoR has started roughly at $z \approx 13$ and ended by $z \approx 6$ (e.g., Robertson *et al.* 2013, 2015; Mondal *et al.* 2015; Mitra *et al.* 2018a,b). However, a deeper insight of these epochs are required to understand these primary stages of the structure formation. HI being the most abundant element in the IGM during these epochs, the 21-cm radiation which originates due to hyperfine transition in HI proves to be most promising probe of CD–EoR. The CD–EoR 21-cm signal is highly sensitive to the properties of the first astrophysical sources and their interactions with the IGM. Therefore, mapping out the intensity distribution of the redshifted 21-cm radiation from HI in the IGM provides a unique and direct way to study CD–EoR (e.g. Sunyaev & Zeldovich 1975; Hogan & Rees 1979; Scott & Rees 1990; Bharadwaj *et al.* 2001; Bharadwaj & Sethi 2001). Given such important information that the observations of redshifted 21-cm signal is expected to reveal, it is in one of the major science goals of the upcoming Square Kilometer Array (SKA-low) (Koopmans *et al.* 2015). It is expected that the direct observations of the CD–EoR 21-cm signal would be able to answer various fundamental questions related to the progress of heating and reionization of the IGM, properties of sources involved and their evolution, etc.

A substantial effort is ongoing with the SKA pathfinder radio-interferometers, such as uGMRT (Swarup *et al.* 1991; Gupta *et al.* 2017), PAPER (Parsons *et al.* 2010), LOFAR (van Haarlem *et al.* 2013), MWA (Tingay *et al.* 2013), HERA (DeBoer *et al.* 2017) and NenuFAR (Mertens *et al.* 2021) to detect the CD–EoR 21-cm signal. However, most of them are not suited for the CD observations due to their limited sensitivity and range of operation. The upcoming SKA-low will be a giant leap in terms of sensitivity as well as it will have a large frequency bandwidth (50–350 MHz) to cover both CD and EoR. Unfortunately, even with first phase of SKA-low, a direct detection of the signal is not possible due to $\sim 10^4$–10^5 times stronger galactic and extra-galactic foregrounds (e.g., Ali *et al.* 2008; Bernardi *et al.* 2009; Ghosh *et al.* 2012). Therefore, the current experiments aim to observe the signal by measuring its statistics, majorly the power spectrum (PS) (e.g., Bharadwaj & Sethi 2001; Bharadwaj & Ali 2004, 2005). However, only few weak upper limits on the PS amplitudes have been reported to date (e.g., GMRT: Paciga *et al.* 2011, 2013; LOFAR: Yatawatta *et al.* 2013; Patil *et al.* 2017; Gehlot *et al.* 2019; Mertens *et al.* 2020; MWA: Li *et al.* 2019; Barry *et al.* 2019; Trott *et al.* 2020; PAPER: Cheng *et al.* 2018; Kolopanis *et al.* 2019; HERA: The HERA Collaboration *et al.* 2021). In its first phase, SKA-low shall be able to measure the 21-cm power spectrum with high precision at different redshifts within relatively less observation time. One would additionally compute higher-order statistics, such as bispectrum (e.g., Majumdar *et al.* 2018, 2020; Kamran *et al.* 2021a,b; Giri *et al.* 2019), trispectrum (Mondal *et al.* 2016, 2017), etc., which can provide additional information of these epochs. Moreover, the second phase of the SKA-low is expected to produce

images of the 21-cm signal maps from CD and EoR. One can then use various image-based statistical tools, such as Minkowski functional (e.g., Kapahtia *et al.* 2018, 2019, 2021) and largest cluster statistics (e.g., Bag *et al.* 2018, 2019; Pathak *et al.* 2022), etc., to extract maximum information out of the signal.

The HI intensity mapping of the post-reionization ($z \leq 6$) 21-cm signal holds the potential to probe the large-scale structures and constrain various cosmological parameters (Bharadwaj *et al.* 2001, 2009; Bharadwaj & Sethi 2001; Loeb & Wyithe 2008; Visbal *et al.* 2009; Villaescusa-Navarro *et al.* 2015). It can independently probe the expansion history of the Universe by measuring the Baryon acoustic oscillation (BAO) in the 21-cm power spectrum (PS) (Chang *et al.* 2008; Wyithe *et al.* 2008; Seo *et al.* 2010).

Several efforts have already been carried out towards detecting the post-reionization 21-cm signal. Pen *et al.* (2009) first detected the signal by cross-correlating the HI Parkes all sky survey (HIPASS) and the six degree field galaxy redshift survey (6dFGRS; Jones *et al.* 2004). At a higher redshift ($z \sim 0.8$), the detection of the cross power spectrum has been presented in Chang *et al.* (2008) using 21-cm intensity maps acquired at the green bank telescope (GBT) and the DEEP2 galaxy survey. Further improvement on these measurements was carried out (Masui *et al.* 2013) by cross-correlating the new intensity mapping data from the WiggleZ dark energy survey (Drinkwater *et al.* 2010). The auto-power spectrum measurement of 21-cm intensity fluctuation maps acquired with GBT has been used to constrain neutral hydrogen fluctuations at $z \sim 0.8$ (Switzer *et al.* 2013). These measurements were conducted using single-dish telescopes in the low-redshift ($z < 1$) regime, where we already have optical surveys. Next-generation intensity mapping surveys with SKA-mid (Bull *et al.* 2015) will have the potential to open up a large cosmological window at the post-reionization epoch, allowing us to detect the 21-cm signal with a high level of accuracy.

Rest of the article is arranged in the following way. We start Section 2, with discussion on analytical and numerical models of the 21-cm signal. Next, we present different statistical tools to quantify the signal in Section 3. We mention observational challenges in detection of the HI 21-cm signal in Section 4. We briefly mention different foreground contributions to the observed signal. In Section 5, we quote the present upper limits on the signal with the current telescopes and also discuss the prospects of measuring signal statistics using SKA-low. We finally summarize the paper in Section 6.

2. Modeling 21-cm signal

The redshifted 21-cm radiation acts as a proxy to the HI distribution in the IGM, which almost follows the underlying matter density field during CD. However during EoR, the HI distribution is largely determined by the ionized regions. The HI 21-cm signal is quantified using the differential brightness temperature (Rybicki & Lightman 1979) observed against CMB at a frequency ν and along a direction $\hat{n}$. This can be written as:

$$T_b(\hat{n}r_\nu, \nu) = \bar{T}(z)\bar{x}_{\mathrm{HI}}(z)[1 + \delta_{\mathrm{HI}}(\hat{n}r_\nu, z)]\left(1 - \frac{T_\gamma}{T_s}\right)$$
$$\times \left[1 - \frac{1+z}{H(z)}\frac{\partial v_{\parallel}(\hat{n}r_\nu, z)}{\partial r_\nu}\right], \quad (1)$$

where ν is the frequency of observation and related to the redshift z as $\nu = 1420/(1 + z)$ MHz. r_ν is the co-moving distance to the redshift z, $\bar{x}_{\mathrm{HI}}$ is the mean neutral hydrogen fraction, δ_{HI} is the HI density contrast, $H(z)$ is the Hubble parameter, $\frac{\partial v_{\parallel}}{\partial r_\nu}$ is the gradient of peculiar velocity along the line-of-sight (LoS) direction and the characteristic HI brightness temperature:

$$\bar{T}(z) = 4.0 \, \mathrm{mK}(1 + z)^2 \left(\frac{1 - Y_P}{0.75}\right)$$
$$\times \left(\frac{\Omega_b h^2}{0.020}\right)\left(\frac{0.7}{h}\right)\left(\frac{H_0}{H(z)}\right), \quad (2)$$

where the symbols have their usual meaning. Note that $\bar{T}(z)$ depends only on the background cosmological model. The CD–EoR 21-cm signal also depends on the factor $(1 - T_\gamma/T_s)$ in Equation (1), where $T_\gamma = T_{\gamma 0}(1 + z)$ is the CMB temperature and T_s is the spin temperature of the HI gas in the IGM. The spin temperature is not a thermodynamical quantity, rather it is defined by the relative population of HI atoms between two hyperfine states, i.e.,

$$\frac{n_1}{n_0} = \frac{g_1}{g_0}\exp\left(\frac{-T_*}{T_s}\right). \quad (3)$$

Here, $g_1 = 3$ and $g_0 = 1$ are the degeneracy factors of the excited states and ground states with n_1 and n_0 being numbers of HI atoms in the respective states. $T_* = 0.068 \, \mathrm{K} = h_P \nu_e/k_B$, where h_P is the Planck constant, $\nu_e = 1420 \, \mathrm{MHz}$ and k_B is the Boltzmann constant. There are three major physical processes which control the level population of HI atoms and couples T_s to either the gas kinetic temperature T_K or T_γ or the Ly-α color temperature T_α during various stages. Computation shows that the T_s depends on the other

three temperatures as (e.g., Pritchard & Loeb 2012):

$$T_s^{-1} = \frac{T_\gamma^{-1} + x_c T_K^{-1} + x_\alpha T_\alpha^{-1}}{1 + x_c + x_\alpha}, \tag{4}$$

where x_c and x_α denote the strength of different processes. During the initial stages of CD, the Ly-α radiations from the sources electronically excited and de-excited the HI atoms due to resonant scattering. This redistributes the atoms between the two hyperfine levels and therefore, couples the spin temperature to $T_\alpha \approx T_K$. This is known as Wouthuysen–Field effect (Wouthuysen 1952; Field 1958; Chen & Miralda-Escudé 2004) makes x_α dominant over all other coupling terms. Later at $z \approx 17$, the X-rays have been started heating IGM and T_s starts rising towards T_γ. Therefore, the CD 21-cm signal is observed in absorption, however, reionization starts as soon as T_s overshoots T_γ and the EoR 21-cm signal is observed in the emission. The evolution of the simulated mean brightness temperature $\bar{T}_b(z)$ during CD–EoR is shown in the bottom panel of Figure 4. There are several experiments which aim to measure this global signal from CD–EoR. However, the radio-interferometric instruments, such as SKA, are expected to observe the spatial fluctuations in the brightness temperature of the CD–EoR signal, which is defined as:

$$\delta T_b(\hat{n}r_\nu, \nu) = T_b(\hat{n}r_\nu, \nu) - \bar{T}_b(\nu), \tag{5}$$

where $\bar{T}_b(\nu) \equiv \bar{T}_b(z)$, as shown in the bottom panel of Figure 4.

The formation and evolution of the first sources (as mentioned in Section 1) is expected to actively control the topology of the fluctuating CD–EoR signal $\delta T_b(\hat{n}r_\nu, \nu)$ (Barkana & Loeb 2001; Furlanetto *et al.* 2006; Pritchard & Loeb 2012). It is widely accepted that the star-forming galaxies during the CD–EoR are the primary source of UV photon production. Having a small mean free path, they cannot photoionize HI much far into the IGM. Some of these galaxies are likely to have accreting supermassive or intermediate-mass black holes at their centers, and they will act as mini-quasars (mini-QSOs). The mini-QSOs can majorly produce high-energy X-ray photons, which have larger mean free path and heat up the IGM. Besides, the galaxies with a high star formation rate are likely to host binary systems, such as high-mass X-ray binaries (HMXBs), which also produce a copious amount of X-ray photons and interact with the IGM differently as compared to the mini-QSOs.

The relative abundances of different sources and their radiation processes can have distinct imprints on the evolution of HI in the IGM during CD–EoR. It is widely understood that the UV photons ionize HI efficiently than the X-ray photons. Hence, for example, if the Universe is dominated by the UV radiations from the galaxies, it would become an so-called 'inside-out' reionization scenario. This is because the UV photons first ionize the dense regions around the sources and then, the ionization fronts advance further into the IGM. On the other hand, if the sources, such as HMXBs producing hard X-ray photons with large mean free paths would have been most abundant in the early Universe, they could cause the so-called 'outside-in' reionization scenario (Choudhury *et al.* 2009). In this scenario, high energy X-ray photons would easily escape from their host galaxies and travel the large distance in the IGM to be redshifted to the lower frequency corresponding to the ionizing photons. Hence, the under-dense regions ionize first, and as time goes on, the reionization slowly progresses into the denser regions. In practice, both of these scenarios contribute to the photo-ionization of IGM gas. However, the inside-out scenario driven by UV photons from the galaxies dominates the photo-ionization. On contrary, the X-ray photons contributes majorly to heat up the IGM thereby modulating T_s of the HI.

We now briefly summarize analytical and numerical methods to simulate $\delta T_b(\hat{n}r_\nu, \nu)$ of the CD–EoR 21-cm signal in the following subsections.

2.1 *Analytical modeling*

According to Equations (1) and (5), one needs a perfect knowledge of $\bar{x}_{HI}(z)$, $\delta_{HI}(\hat{n}r_\nu, z)$, $T_s(\hat{n}r_\nu, z)$ and $\partial v_\parallel(\hat{n}r_\nu, z)/\partial r_\nu$ to accurately model $\delta T_b(\hat{n}r_\nu, \nu)$. Owing to complicated interplay between the cosmological and astrophysical processes, it is not straightforward to write analytical expressions for all of these quantities and hence, for δT_b during CD–EoR. However, modeling of the statistics of the signal is possible under some simplified assumption. Here, we briefly discuss a simplistic but effective model of the EoR 21-cm power spectrum as described in Bharadwaj & Ali (2005). Note that power spectrum is the Fourier transform of the two-point correlation of a fluctuating field. We refer the reader to Section 3.2, for a complete description of the power spectrum statistics.

As mentioned before, assuming that the HI gas is heated well before it is reionized, and that the spin temperature is coupled to the gas temperature with $T_s \gg T_\gamma$ CMB temperature, so that $(1 - T_\gamma/T_s) \rightarrow 1$ (Equation 1). Now, the EoR 21-cm emission signal depends only on the HI density contrast and peculiar velocity. We assume that the hydrogen density and peculiar velocities follow the dark matter with possible bias b_{HI}. Further, it

is assumed that non-overlapping spheres of co-moving radius R are completely ionized, the centers of the spheres being clustered with a bias $b_c \geq 1$ relative to underlying dark matter distribution. One would expect the centers of the ionized spheres to be clustered, given the fact that one can identify them with the locations of the first luminous objects, which are believed to have formed at the peaks of the density fluctuations. Similar kind of models have used in the context of the effect of patchy reionization on the CMB (Gruzinov & Hu 1998; Knox *et al.* 1998) and H I emission in high redshifts (Zaldarriaga *et al.* 2004).

In this simple reionization model, the ionization history is considered as $\bar{x}_{HI}(z) = 1/1 + \exp((z-z_0)/\Delta z)$, where $z_0 = 10$ and $\Delta z = 0.5$, so that 50% of the hydrogen is reionized at a redshift of $z = 10$ (see Figure 1). The fraction of volume ionized f_V, the mean neutral fraction $\bar{x}_{HI}$ and mean co-moving number density of ionized spheres $\bar{n}_{HI}$ are related as $f_V = 1 - \bar{x}_{HI} = (4\pi R^3/3)\bar{n}_{HI}$. The co-moving size of the spheres, R increases with z, so that the mean co-moving number density of the ionized spheres remains constant. Using various assumptions mentioned above, the H I power spectrum of brightness temperature fluctuations have been modeled as (for more details see Bharadwaj & Ali 2005):

$$P(\vec{k}, z) = \bar{T}^2(z) \left[b_{HI} \bar{x}_{HI}(1 + \beta \mu^2) - b_c f_V W(kR) \right]^2$$
$$\times\ P_m(k, z) + \bar{T}^2(z) \frac{f_V^2 W^2(kR)}{\bar{n}_{HI}}, \quad (6)$$

where $W(kR) = 3[\sin(kR) - kR\cos(kR)]/(kR)^3$ is the Fourier transform of the spherical top hat window function, μ is the cosine of the angle between $\vec{k}$ and the LoS vector $\hat{n}$, the term $(1 + \beta \mu^2)$ arises due to the peculiar motion of the H I atoms and β denotes the linear redshift distortion parameter (Bharadwaj *et al.* 2001; Bharadwaj & Ali 2004). The first term, which contains dark matter power spectrum $P_m(k)$ arises from the clustering of the hydrogen and the clustering of the centers of the ionized spheres. The second term, which has $1/\bar{n}_{HI}$ arises due to the discrete nature of the ionized regions.

Figure 2 shows the behavior of $P(k)$ as a function of k at three different frequencies 110, 120 and 130 MHz, which corresponds to redshifts of $\sim$11.9, 10.8 and 9.9, respectively. At $\nu \leq 110$ MHz, the H I power spectrum traces the dark matter power spectrum as the hydrogen is largely neutral. The results shown here assume $b_{HI} = 2$ and the centers of the ionized bubbles to be clustered with a bias $b_c = 1.5$ with respect to the underlying matter distribution. The presence of these clustered bubbles reduces the signal. At 130 MHz, H II bubbles
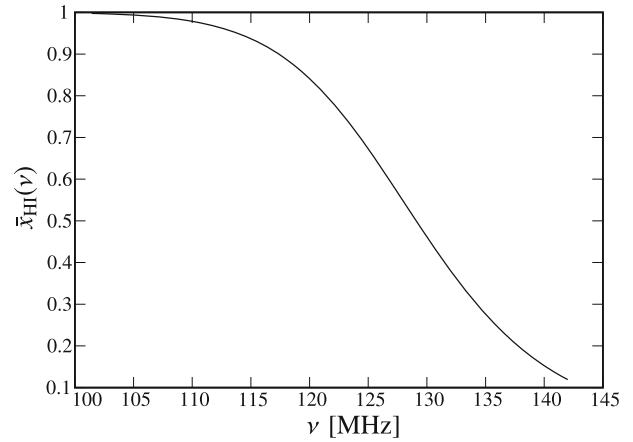

Figure 1. Evolution of the mean neutral fraction $\bar{x}_{HI}$ with the frequency of redshifted H I emission for the reionization model discussed in the text.
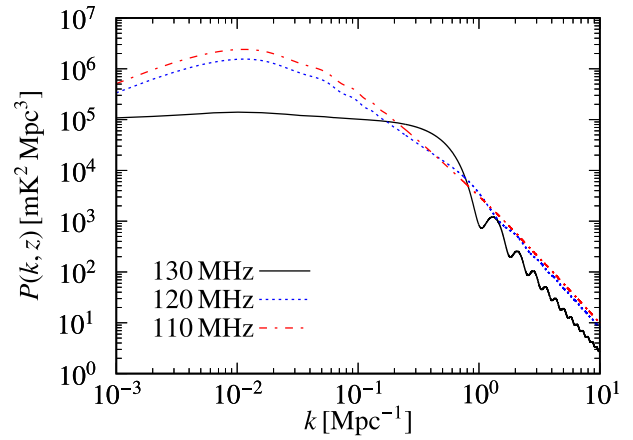

Figure 2. Expected H I power spectrum $P(k, z)$ from the epoch of reionization at different frequencies (as shown in the figure) for models of the H I distribution considered here. Here, the μ dependence has been incorporated by using the average value $\langle \mu^2 \rangle = 1/3$.

occupy around 50% of the Universe. The clustering of the bubbles or the dark matter is no longer important at this stage, and the H I power spectrum is largely governed by the discrete nature (Poisson distribution) of the ionized bubbles. The power spectrum drastically falls towards the large k, which corresponds to the scales smaller than the typical bubble sizes. At lower redshifts (higher frequencies), the H I signal is expected to decrease rapidly because most of the H I would be ionized. The analytical model for the reionization breaks down beyond a stage when a large fraction of the Universe is ionized. This break down occurs beyond 130 MHz for the model presented here. Note that the power spectrum shows oscillations for smaller length-scales ($k > \pi/R$), which arises from the nature of the window function $W(kR)$ (Equation 6). At these

scales, the amplitude of $W^2(kR)$ falls as $(kR)^{-4}$, which is steeper than the matter power spectrum $P_m(k)$. Hence, at sufficiently large k, the EoR power spectrum $P(k)$ is dominated by the dark matter fluctuations and it approaches $P_m(k)$ with the approach being faster for large R. It is important to note that the oscillations in $P(k)$ are due to the fact that all the ionized bubbles are assumed to be spheres of the same size. In reality, the ionized regions will have a spread in the bubble shapes and sizes.

However, this model has a limitation that it cannot be used when a large fraction of the volume is ionized as the ionized spheres start to overlap and the HI density contrast becomes negative in the overlapping regions. Calculating the fraction of the total volume, where the HI density contrast is negative, is found to be $f_V^2/2$. Hence, the model is valid for $z > 10$, where $f_V < 0.5$ and the δ_{HI} is negative in <12.5% of the total volume.

In the post-reionization era ($z \leq 6$), the bulk of the HI resides in the high column density clouds, which produces the damped Lyman-α absorption lines as observed in the quasar spectra (Lanzetta *et al.* 1995; Storrie-Lombardi *et al.* 1996; Péroux *et al.* 2003). The current observations indicate that the co-moving, density of HI expressed as a fraction of the present critical density is nearly constant at a value $\Omega_{HI}(z) \sim 10^{-3}$ for $z \geq 1$ (Péroux *et al.* 2003; Noterdaeme *et al.* 2012; Zafar *et al.* 2013). The damped Lyman-α clouds are believed to be associated with galaxies, which represent highly non-linear overdensities. However, on the cosmological large-scales, it is reasonable to assume that these HI clouds trace the dark matter with a constant linear bias $b_{HI} = 2$ (Sarkar *et al.* 2016).

Converting Ω_{HI} to the mean neutral fraction $\bar{x}_{HI} = \Omega_{HI}/\Omega_b$ gives us $\bar{x}_{HI} = 50\Omega_{HI}\ h^2(0.02/\Omega_b h^2)$ or $\bar{x}_{HI} = 2.45 \times 10^{-2}$. It is also assumed that $T_s \gg T_\gamma$, and hence, one sees the HI in emission. Using these, we have

$$P(\vec{k}, z) = \bar{T}^2(z)\, b_{HI}^2 \bar{x}_{HI}^2 (1 + \beta\mu^2)^2 P_m(k, z). \qquad (7)$$

The fact that the neutral hydrogen is in discrete clouds makes a contribution which is not included here. Another important effect not included here is that the fluctuations become non-linear at low z. Both these effects have been studied using simulations (Bharadwaj & Srikant 2004).

The predictions for post-reionization HI power spectrum $P(k)$ are shown as a function of k for two frequencies 325 MHz ($z \approx 3.36$) and 800 MHz ($z \approx 0.77$) in Figure 3. In this era, the HI is assumed to trace the dark matter field with a bias $b_{HI} = 2$. Therefore, the shape of the HI power spectrum as a function of k
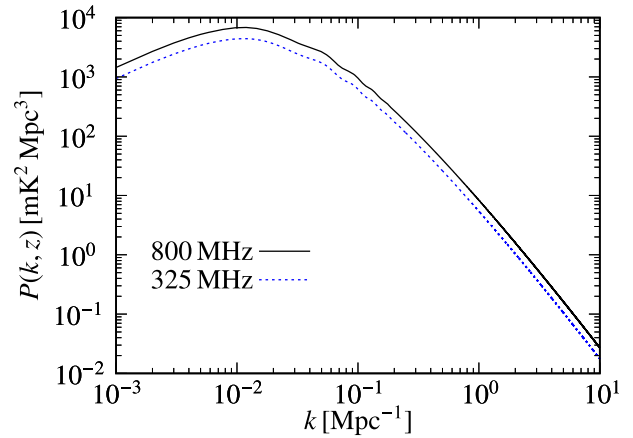

Figure 3. Expected post-reionization HI power spectrum $P(k, z)$ at two frequencies (as shown in the figure).

is decided by the dark matter power spectrum $P_m(k)$ at the relevant Fourier modes. The values of $P(k)$ increase with frequency (decrease with redshift) is a reflection of the fact that $P_m(k)$ grows with time.

2.2 *Numerical modeling*

Even though analytical models provide some approximate pictures of CD–EoR 21-cm signal, it is not possible to incorporate various complex physics into them. It is thus necessary to use numerical simulations to obtain realistic maps of CD–EoR 21-cm signal and other quantities relevant to the observations. The numerical models typically stands on basic assumption that the hydrogen traces the underlying dark matter field and the collapsed halos host the ionizing sources.

2.2.1 Radiative transfer technique   We have already seen that the intensity of the HI 21-cm radiation crucially depends on the ionization fraction, gas temperature and the Lyman-α coupling (see Equations (1) and (4)). Thus, besides knowing how the radiating sources, such as galaxies, high-mass X-ray binaries and quasars are distributed in the simulation volume and their spectral energy distribution (SED), it is also important to know how the emitted UV, X-ray and Lyman-α photons propagate into the clumpy baryonic field. One can simplify the picture of star formation by assuming that the stellar content of a galaxy is proportional to the mass of the host dark matter halo and SEDs of the galaxies per unit stellar mass are the same as one can obtain from a stellar population synthesis code like PEGASE2 (Fioc & Rocca-Volmerange 1997). Note that, in reality, the star formation scenario can be very complex.
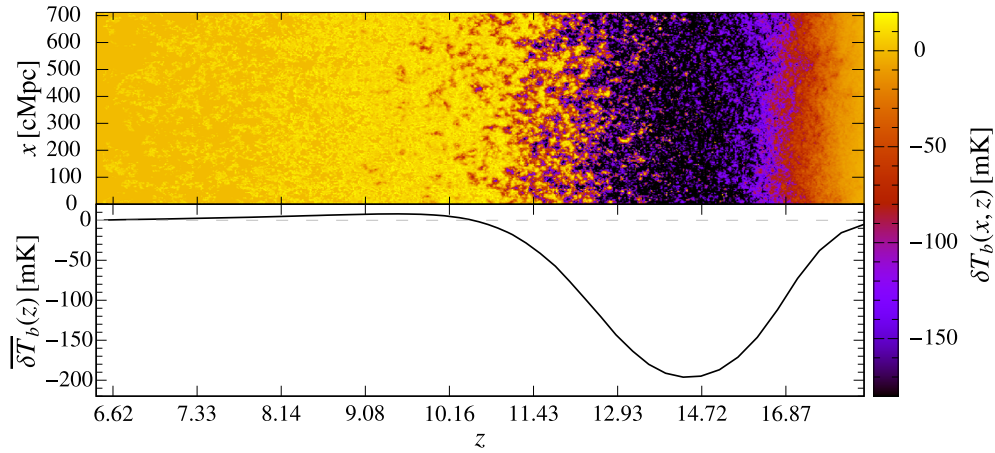
Figure 4. The top panel shows a light-cone demonstrating the redshift evolution of the brightness temperature fluctuations of the Hı 21-cm signal during CD–EoR phase. The bottom panel shows the evolution of the corresponding mean brightness temperature of the 21-cm signal. The details of the dark matter only *N*-body simulation as well as the source model used in GRIZZLY to generate this light-cone can be found in Ghara *et al.* (2015b).

In principle, one needs to solve a seven-dimensional cosmological radiative transfer equation (see e.g., Choudhury *et al.* 2016) at every grid point of the simulation box to include all relevant physical processes. This is computationally very challenging and thus, often approximations are used for the radiation transfer. For example, a 3D radiative transfer code C^2 RAY (Mellema *et al.* 2006) works by tracing rays from all UV emitting sources and iteratively solving for the time evolution of x_{HII}. Even with such approximation, C^2 RAY results match very well with fully numerical radiative transfer Codes (see e.g., Iliev *et al.* 2006). Alternatively, a one-dimensional radiative transfer scheme is used in codes such as GRIZZLY (Ghara *et al.* 2015a, b) and BEARS (Thomas & Zaroubi 2008). This method approximates the transfer of UV, X-rays and Lyman-α photons by assuming that the effect from individual sources is isotropic. Although GRIZZLY employs a range of approximations to correct for overlaps between individual ionized regions around the sources, its results are still in good agreement with those of the 3D radiative transfer code C^2 RAY (see Ghara *et al.* 2018 for details), while being at least 10^5 times faster. In addition, as GRIZZLY also includes X-ray heating and Ly-α coupling, it can probe the cosmic dawn, where the spin temperature fluctuations are expected to dominate the fluctuations in the 21-cm signal. We present one simulated GRIZZLY light-cone of the brightness temperature in Figure 4. Clearly, at the beginning of the cosmic dawn era (around $z \sim 20$ in this case), Lyman-α coupling is the most crucial effect that determined the fluctuations in the 21-cm signal. This was followed by an era, where X-ray heating becomes important. One can see that over time individual emission regions around the X-ray sources overlap together. The gas in the IGM became significantly heated around $z \sim 12$ after which the spin temperature fluctuations became negligible compared to the ionization fluctuations.

As mentioned before, radiative transfer methods are based on post-processing of the outputs of *N*-body simulations (such as dark matter halo catalog, density and velocity fields, etc.), and these miss the hydro-dynamical effects on the small scales. On the other hand, more sophisticated fully coupled radiative-hydrodynamic simulations of galaxy formation, such as Renaissance (Xu *et al.* 2016), CoDa I (Aubert *et al.* 2018), CoDa II (Ocvirk *et al.* 2020) by CPU-GPU code EMMA (Aubert *et al.* 2015) can include several small scale physics and thus are useful to study the stochastic star formation, metal enrichment, radiative feedback, recombination in the IGM, etc. The impact of such relevant processes as found from these studies are often approximately used in other types of simulations, such as semi-numerical and numerical without hydro-dynamics. While these simulations are very important to consider small-scale physical processes, these are often limited by small box size and thus challenging to compare with the 21-cm observations. In addition, resolving very fine scale physical processes, such as a supernova, turbulence, etc., are still beyond the capacity of these methods and often depends on simplified sub-grid prescriptions.

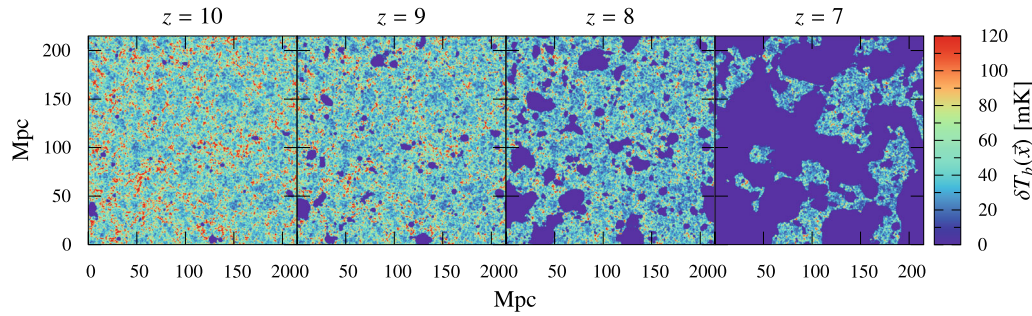*2.2.2 Semi-numerical simulations* Detailed and fast models of reionization are a crucial ingredient for

Figure 5.   One realization of the two-dimensional sections through the simulated H I brightness temperature maps at $z = 10$, 9, 8 and 7 within simulation volume of $[215\ \mathrm{Mpc}]^3$. See Mondal *et al.* (2017) for details.

interpreting the observed data and constraining the physics of the CD–EoR. The aforementioned radiative transfer simulations are required for accurate modeling, however, they are not well-suited for exploring a vast range of parameter space (see e.g., Mondal *et al.* 2020b). Therefore, fast methods are crucial to simulate a high dynamic range large-volume boxes of the H I 21-cm signal. This can be achieved by using semi-numerical techniques, which typically involve three main steps: (a) simulating dark matter distribution at the desired redshifts, (b) identifying the collapsed dark matter halos and (c) generating the reionization map using an excursion set formalism. The first two steps mostly overlap with those in the radiative transfer techniques. It is the excursion set (Furlanetto *et al.* 2004) or similar formalism, which generates ionization maps based on some approximate methodology.

For example, a semi-numerical code REIONYUGA[1] (Mondal *et al.* 2015, 2016, 2017) generates ionization fields following methodologies described in Choudhury *et al.* (2009) and Majumdar *et al.* (2014). The code uses matter density field simulated using a fast, efficient and accurate parallelized particle-mesh N-body code[2] (Mondal *et al.* 2015), and the corresponding halo distribution obtained using a Friends-of-Friends halo finder[3] (Mondal *et al.* 2015). The methodology of REIONYUGA assumes that the amount of ionizing photons produced by a source is directly proportional to its host halo mass above a certain mass cut-off M_{min}, which is a parameter in simulations. The proportionality factor here is quantified through a dimensionless parameter N_{ion}, which is related to several degenerate astrophysical factors (see e.g., Shaw *et al.* 2020). The mean free path of the ionizing photons in the IGM R_{mfp}

is the third parameter of REIONYUGA. The methodology determines whether a grid point to be ionized or not by smoothing the hydrogen density field and the photon density field using spheres of different radii starting from the grid size to R_{mfp}. A grid point is considered to be ionized if for any smoothing radius (within R_{mfp}), the photon density exceeds the hydrogen density at that grid point. Grid points which do not satisfy this condition are assigned with an ionized fraction. Figure 5 shows one realization of the two-dimensional sections through the simulated H I brightness temperature maps at $z = 10, 9$, 8 and 7 for a $[215\ \mathrm{Mpc}]^3$ simulation box. The resultant ionization maps are similar to those simulated using the costlier radiative transfer methods.

There are several semi-numerical codes available publicly such as SCRIPT[4] (Maity & Choudhury 2022), 21cmFAST[5] (Mesinger *et al.* 2011) and SIMFAST21[6] (Santos *et al.* 2010), etc. However, these codes have their own assumptions and physically motivated parameters. One should appropriately choose these codes based on the individual requirement.

3. Statistical measures of 21-cm signal

The CD–EoR 21-cm signal (see Equations (1) and (5)) is a random field of brightness temperature fluctuations. We need statistical estimators to quantify the signal. One can quantify it using the one-point statistics, such as variance (Patil *et al.* 2014), skewness and kurtosis (Harker *et al.* 2009) of δT_b. Despite being easier to compute and interpret, one-point statistics have limited information. Therefore, many-point statistics, such as N-point correlation functions, are needed to quantify the signal completely. We shall discuss about a few

---

[1]https://github.com/rajeshmondal18/ReionYuga.

[2]https://github.com/rajeshmondal18/N-body.

[3]https://github.com/rajeshmondal18/FoF-Halo-finder.

[4]https://bitbucket.org/rctirthankar/script.

[5]https://github.com/21cmfast/21cmFAST.

[6]https://github.com/mariogrs/Simfast21.

many-point statistics in Fourier space, such as power spectrum, bispectrum, etc., in the following sections. One can also utilize other techniques, such as matched filtering, to enhance the detectability of the signal by extracting its important underlying features.

3.1 *Matched filtering technique*

The detectability of the CD–EoR 21-cm signal in terms of different statistical quantities is low as the targeted signal is weak. Thus, one of the major focuses of the theoretical study of the signal has been to design optimum estimators that enhance the detectability of the signal. The application of the matched filtering technique, which applies pre-defined filters to the measurements to enhance the detectability of the observation is one such complementary method.

In the case of EoR 21-cm signal, Datta *et al.* (2007) introduced this method, where spherical top-hat filters in the image-space were applied to the measured visibilities for enhancing the detectability of ionized regions. Follow up studies, such as Datta *et al.* (2008, 2009b, 2012a), Majumdar *et al.* (2011, 2012) considered various scenarios of IGM ionization states and estimated the detectability of ionized bubbles with radio interferometers like GMRT, MWA, LOFAR, SKA-low. For example, Datta *et al.* (2012a) showed that ∼1200 h of observation with LOFAR should be enough to detect ionized regions with a size ∼25 cMpc at redshift 7.57 with mean ionization fraction $\bar{x}_{HI} = 0.5$. As the largest ionized regions in the field-of-view (FoV) is expected to be formed around very bright sources, such as high-z quasars, this technique can also give information about the bright sources (see e.g., Datta *et al.* 2012a; Majumdar *et al.* 2012).

Recently, Ghara & Choudhury (2020) developed a rigorous Bayesian framework based on the same method to constrain the parameters that characterize the ionized regions. This study shows 20 h integration with SKA-low will be enough to constrain the location and size of the ionized regions around typical quasars at $z = 7$ (Figure 6), while it also constrains the difference in the neutral fractions inside and outside the HII bubble. This framework is useful in identifying large ionized regions in the observed field and hence, will be interesting for following up observations with deeper integration times.

In contrast to the distribution of ionized regions in the IGM during the EoR, the CD HI 21-cm signal mostly depends on how emission/absorption regions are distributed in the IGM. In this case, one can also apply the matched filtering technique to enhance the detectability
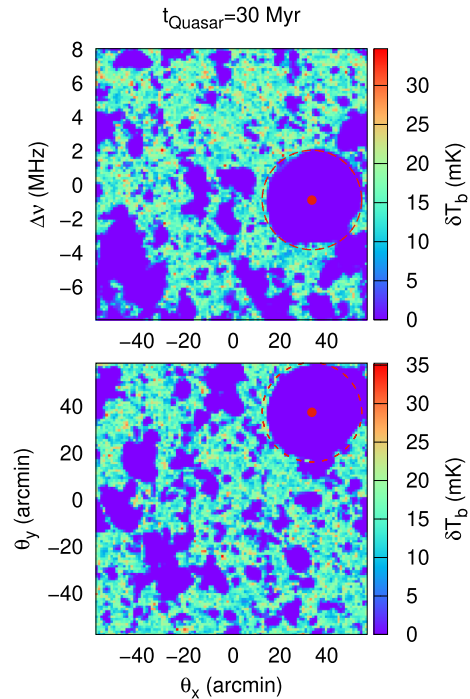

Figure 6. Simulated δT_b maps at $z \approx 7$ using GRIZZLY, the red points indicate the position of a quasar in the field-of-view, the dashed circles represent the best-fit size of matched-filter estimated from a Bayesian analysis as used in Ghara & Choudhury (2020).

of individual emission/absorption regions in the IGM. In a slightly different approach, one can also simply combine the measured visibility from different baselines and frequency channels to enhance the detectability of the signal (see e.g., Ghara *et al.* 2016). We refer the reader to Datta *et al.* (2016) for more details on these methods.

3.2 *3D power spectrum*

The spatial fluctuations in the HI differential brightness temperature $\delta T_b(\vec{x})$ within a 3D co-moving cube of volume V can be decomposed into different Fourier modes in $\vec{k}$-space as:

$$\Delta \tilde{T}_b(\vec{k}) = V^{-1} \sum_{\vec{x}} \exp[-i\vec{k} \cdot \vec{x}] \delta T_b(\vec{x}), \qquad (8)$$

where $\Delta \tilde{T}_b(\vec{k})$ is the Fourier conjugate of $\delta T_b(\vec{x})$ and the summation is over all real space points within the volume. Note that the wave vector $\vec{k}$ can be both positive and negative, however, we have $\Delta \tilde{T}_b^*(\vec{k}) = \Delta \tilde{T}_b(-\vec{k})$ as $\delta T_b(\vec{x})$ is a real field. Now, the HI 21-cm 3D power spectrum $P(\vec{k})$ can be defined as follows:

$$\langle \Delta \tilde{T}_b(\vec{k}_1) \Delta \tilde{T}_b(\vec{k}_2) \rangle = V \delta^K(\vec{k}_1 + \vec{k}_2) P(\vec{k}_1), \qquad (9)$$
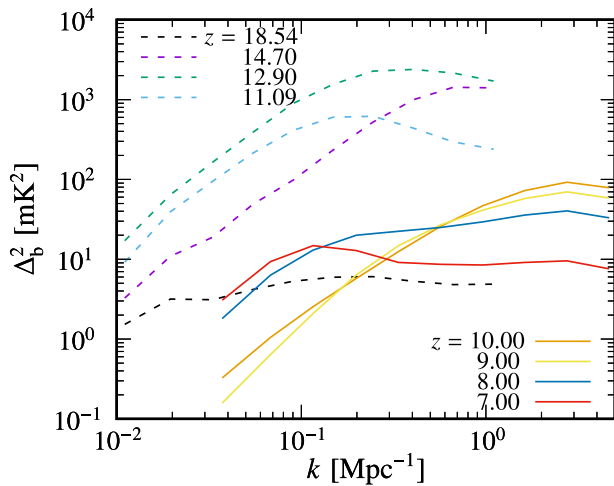
Figure 7. Dimensionless form of CD (dashed lines) and EoR (solid lines) 21-cm 3D power spectrum $\Delta_b^2(k) = k^3 P(k)/(2\pi^2)$. The CD power spectra are obtained from the radiative transfer simulations GRIZZLY, whereas the EoR power spectra are computed using the semi-numerical simulations REIONYUGA. See Ghara *et al.* (2015b) and Mondal *et al.* (2017) for details.

where δ^K is the 3D Kronecker's delta function. The power spectrum is the Fourier pair to the two-point correlation function, which measures excess probability as a function of two points in real space. We now discuss a few important properties of the $P(\vec{k})$. The 3D Kronecker's delta function δ^K appearing in above Equation (9) is a consequence of the fact that the HI fluctuations are assumed to be statistically homogeneous or ergodic in nature along all three spatial directions. It signifies that the HI signal in the different Fourier modes are uncorrelated. Assuming statistical isotropy in the signal removes the direction dependence of the 3D power spectrum, and we can express power spectrum as a function of only k instead of $\vec{k}$, i.e., $P(\vec{k}) = P(k)$. In reality, the CD–EoR 21-cm signal is affected by several LoS effects, such as the redshift-space distortion, light-cone effect and Alcock–Paczynski effect. However, the statistical isotropy is always valid for the monopole component of the 3D power spectrum $P(k)$. Figure 7 shows dimensionless 3D power spectrum $\Delta_b^2(k) = k^3 P(k)/(2\pi^2)$ of the EoR 21-cm signal (solid lines) obtained from an ensemble of simulated signals at four redshifts as shown in Figure 5. In Figure 7, we also show the CD 21-cm 3D power spectrum (dashed lines), which have been simulated using the 1D radiative transfer simulation GRIZZLY as mentioned in Section 2.2.1.

One of the primary goals of the ongoing radio-interferometric experiments is to measure the 3D power spectrum of the EoR 21-cm signal. However, these interferometers are limited in terms of sensitivity. The SKA-low in its first phase is expected to provide EoR 21-cm power spectrum at a considerably high SNR level within 100 h of observations (Shaw *et al.* 2019). Owing to its larger frequency bandwidth, SKA-low will be able to provide measurements of CD 2-cm power spectrum at a reasonably high SNR.

3.3 *Multifrequency angular power spectrum*

The CD–EoR 21-cm signal is observed across a frequency bandwidth for a range of cosmic times, as each wavelength corresponds to a different look-back time. The observed data sets are thus three-dimensional (3D) with the two directions on the sky-plane and the frequency (wavelength) constituting the third dimension. The light-cone effect imprints the cosmological evolution of the 21-cm signal along the LoS direction (or frequency axis) (Datta *et al.* 2012b). The effect is particularly pronounced when the mean 21-cm signal changes rapidly as the Universe evolves. Mondal *et al.* (2018) have developed a method to properly incorporate the light-cone effect in simulations of the 21-cm signal that also includes the effects of peculiar velocities. They showed that the 3D power spectrum $P(k)$ fails to quantify the entire two-point statistical information as it inherently assumes the signal to be ergodic and periodic in all three directions (Figure 8), whereas the light-cone effect breaks these conditions along the LoS direction (Mondal *et al.* 2018). Therefore, the issue is how to analyse the statistics of the 21-cm signal in the presence of the light-cone effect. Datta *et al.* (2012b) and Mondal *et al.* (2018) have proposed an unique statistical estimator of the signal, the multifrequency angular power spectrum (MAPS), which quantifies the entire second-order statistics of the 21-cm signal without requiring the signal to be ergodic and periodic along the LoS direction.

The MAPS is defined as follows:

$$C_\ell(\nu_1, \nu_2) = \Omega^{-1} \langle \Delta \tilde{T}_{b2}(\vec{U}, \nu_1) \, \Delta \tilde{T}_{b2}(-\vec{U}, \nu_2) \rangle, \tag{10}$$

for a given multipole $\ell = 2\pi|\vec{U}|$ and at a pair of frequencies (ν_1, ν_2). In the above equation, Ω is the solid angle subtended by the observation volume to the observer, $\Delta \tilde{T}_{b2}(\vec{U}, \nu)$ is the Fourier conjugate of the HI 21-cm brightness temperature fluctuations $\delta T_b(\vec{\theta}, \nu) \equiv \delta T_b(\vec{x})$ with respect to the two-dimensional angle $\vec{\theta}$ defined on the sky-plane, and $\vec{U}$ is the Fourier conjugate of $\vec{\theta}$. For the case, where the signal is assumed to be both
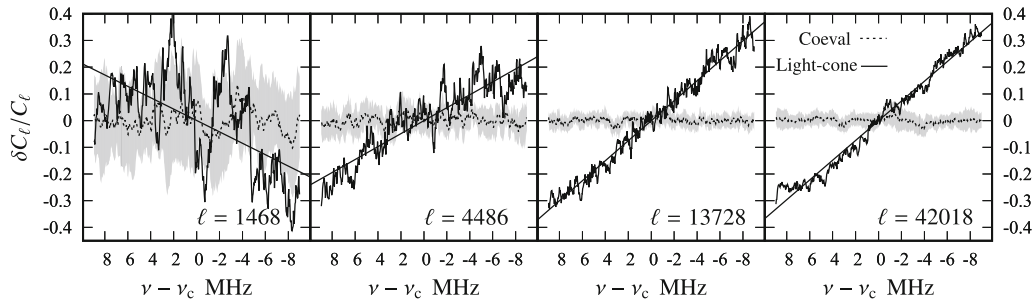
Figure 8. This plot shows information loss in the assumption of statistical homogeneity and periodicity along the frequency direction. Dotted curves are for the power spectrum, which assumes the signal is ergodic and periodic along the frequency direction and the soiled curves are for MAPS, which does not require these assumptions. See Mondal *et al.* (2018) for details.

ergodic and periodic, the MAPS bears a direct relation to the three-dimensional power spectrum $P(\vec{k})$ (see Section 3.2) i.e.,

$$C_\ell^{\mathrm{EP}}(\Delta\nu) = (r_c^2 r_c' B)^{-1} \sum_{k_\parallel} \mathrm{e}^{ik_\parallel r_c' \Delta\nu} P(\vec{k}_\perp, k_\parallel), \quad (11)$$

where $\Delta\nu = |\nu_1 - \nu_2|$, r_c and r_c' denote the distance between the observer and the center of the observation volume and its derivative regarding ν and B gives the observing bandwidth. In Equation (11), $\vec{k}_\perp$ and $k_\parallel$, respectively, denote the perpendicular and parallel components of the 3D wavevector $\vec{k}$ regarding the LoS direction of the observer.

Mondal *et al.* (2020a) have applied MAPS to quantify the statistics of light-cone EoR 21-cm signals and study the prospects for measuring it, using the upcoming SKA observations. In a previous study (Mondal *et al.* 2019), we have also found that it is possible to recover the cosmic history across the entire observational bandwidth using MAPS statistics. Furthermore, by construction, the MAPS is more natural to use for radio-interferometric observations than the 3D power spectrum $P(k)$, which provides biased estimates of the signal (Mondal *et al.* 2020a).

3.4 *Higher order statistics*

The two-point Fourier space statistics (i.e., power spectrum) can provide a complete statistical description of a pure Gaussian random field. However, the fluctuations in the CD–EoR 21-cm signal $\delta T_b(\vec{x})$ are highly non-Gaussian (Bharadwaj & Pandey 2005; Mellema *et al.* 2006). This non-Gaussianity arises due to the non-random distribution of the first luminous sources in the IGM and the interaction of radiation from these sources with IGM gas through various complex astrophysical processes, such as the heating and ionization during

CD–EoR. Furthermore, this intrinsic non-Gaussianity evolves (either increases or decreases) with time as the heating and the ionization fronts grow and percolate in the IGM. It is necessary to consider the non-Gaussian statistics of the signal, while interpreting it thoroughly via statistical inference methods. As discussed in Section 3.2, the power spectrum $P(k)$ is a resultant of signal correlation at a single Fourier mode, and therefore, it is not sensitive to the non-Gaussianity present in the CD–EoR 21-cm signal. To capture this inherent non-Gaussianity present at various length scales, one needs the higher-order statistics, such as the bispectrum (three-point), trispectrum (four-point), etc. The bispectrum, the Fourier transform of the three-point correlation function, is the lowest order statistic that can directly probe the non-Gaussianity present in the signal. We define bispectrum $B(\vec{k}_1, \vec{k}_2, \vec{k}_3)$ as

$$\langle \Delta\tilde{T}_b(\vec{k}_1) \Delta\tilde{T}_b(\vec{k}_2) \Delta\tilde{T}_b(\vec{k}_3) \rangle = V\delta^{\mathrm{K}}(\vec{k}_1 + \vec{k}_2 + \vec{k}_3)$$
$$\times B(\vec{k}_1, \vec{k}_2, \vec{k}_3), \quad (12)$$

where $\Delta\tilde{T}_b(\vec{k})$ is the Fourier transform of $\delta T_b(\vec{x})$, and $\delta^{\mathrm{K}}(\vec{k}_1 + \vec{k}_2 + \vec{k}_3)$ is the Kronecker's delta function. The occurrence of Kronecker's delta function is due to statistical homogeneity of the signal. This indicates that bispectrum is defined only when the three vectors $(\vec{k}_1, \vec{k}_2, \vec{k}_3)$ form a closed triangle in $\vec{k}$-space, i.e., $\vec{k}_1 + \vec{k}_2 + \vec{k}_3 = 0$, shown in Figure 9. The bispectrum, hence, provides us with the correlations among the signal at different Fourier modes and thus is capable to capture the non-Gaussianity.

Recently, Trott *et al.* (2019) for the first time tried to put an upper limit on the 21-cm bispectrum using the observational data from the MWA phase II array. The detection of the CD–EoR 21-cm bispectrum will require more sensitive observations than needed for power spectrum detection. The next-generation radio interferometers, such as HERA and upcoming SKA-low

are expected to have enough sensitivity for bispectrum measurements. It is thus more relevant now to predict the characteristics of the CD–EoR 21-cm bispectrum in view of the construction of SKA-low.

Several studies on the CD–EoR 21-cm bispectrum have been performed so far. Bharadwaj & Pandey (2005) and Ali *et al.* (2005) made the first theoretical attempts to characterize the EoR 21-cm bispectrum using their analytical models. One of the most crucial features of the bispectrum predicted in their studies is its sign that can attain both positive and negative values. Later on, a number of independent studies, such as Watkinson *et al.* (2017, 2019), Majumdar *et al.* (2018, 2020), Hutter *et al.* (2020) and Kamran *et al.* (2021a, b), confirmed the sign of the CD–EoR 21-cm bispectrum using the detailed simulations of the 21-cm signal from the CD–EoR and for a variety of k-triangles (equilateral, isosceles and scalene). The sign of the bispectrum
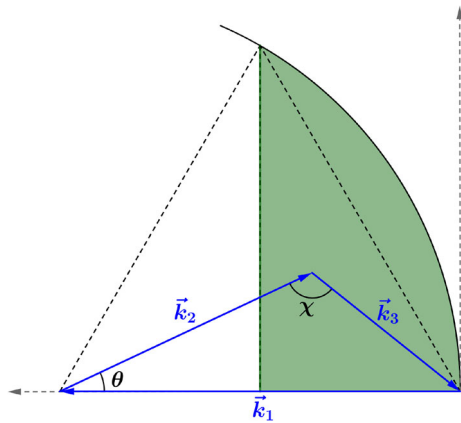
is an important feature of this statistic. Figure 10 clearly demonstrates that the redshift evolution of the bispectrum during CD, for instance, has way more features in its magnitude and sign than the power spectrum evolution. The sign and magnitude of the bispectrum uniquely identify which one among the dominant physical processes, such as the Ly-α coupling, X-ray heating, and photo-ionization is driving the non-Gaussianity at a particular cosmic time. For further insights, we refer the interested reader to (b). Furthermore, when analyzed, the sign change of the bispectrum's evolution with triangle configurations and the redshift can be used as a confirmatory test of the CD–EoR 21-cm mean signal detection.

A comprehensive view of the non-Gaussianity requires the study of the bispectra for all possible triangles in the Fourier space. Bharadwaj *et al.* (2020) for the first time, proposed a new parameterization for the bispectrum of triangles of all possible unique shapes in the Fourier space, shown by the shaded region in Figure 9, while the following conditions are satisfied.

$$k_1 \geq k_2 \geq k_3, \tag{13}$$

$$0.5 \leq n \leq 1.0, \tag{14}$$

$$n \cos \theta \geq 0.5, \tag{15}$$

where $n = k_2/k_1$ and $\cos \theta = -\vec{k}_1 \cdot \vec{k}_2/k_1 k_2$.

Majumdar *et al.* (2020) and Kamran *et al.* (2021a) followed the same prescription and presented a comprehensive and correct interpretation of the CD–EoR 21-cm signal by studying the bispectrum of all possible unique triangles using the simulated signal. These studies conclude that the squeezed limit bispectrum typically attains the largest magnitude among all possible unique triangles in the Fourier space. The squeezed-limit bispectrum is, thus, expected to have the largest detection probability.
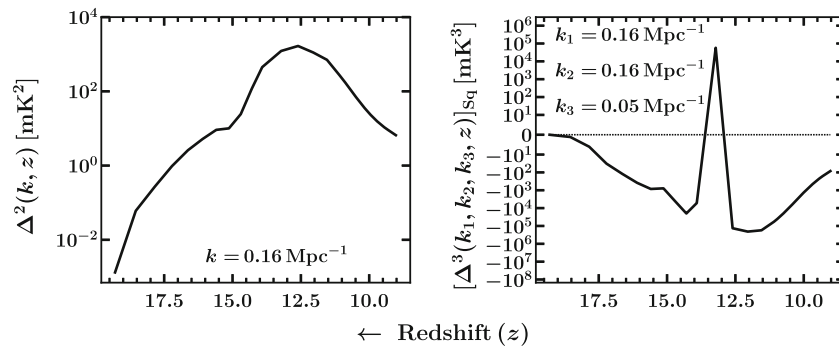

Figure 9. Definition of triangles in $\vec{k}$-space. The shaded region shows the positions of all the unique points of intersections of $\vec{k}_1$ and $\vec{k}_2$ vectors, while satisfying Equations (13)–(15).

Figure 10. Left panel: Redshift evolution of the 21-cm power spectrum during CD at $k = 0.16$ Mpc^{-1}. Right panel: Redshift evolution of the 21-cm bispectrum for squeezed-limit k-triangle during CD at $k_1 = 0.16$ Mpc^{-1} ($k_2 = 0.16$ and $k_3 = 0.05$ Mpc^{-1}).

3.5 *Intrinsic errors in statistical estimators*

Statistical errors which are inherent to the measurements of the CD–EoR 21-cm signal have two components: (1) cosmic variance, which is inherent to the signal itself and (2) system noise, which is observationally inevitable. The cosmic variance error arises from the finite volume of the Universe accessible to a measurement, whereas the system noise is observed by the receivers even looking at a hypothetically blank sky. There have been several works to predict the errors on the EoR 21-cm power spectrum $P(k)$ measurements (see e.g., Parsons *et al.* 2012; Zaroubi *et al.* 2012; Jensen *et al.* 2013; Pober *et al.* 2014; Ewall-Wice *et al.* 2016a). Most of these earlier works have commonly assumed the signal to be a Gaussian random field, for simplicity. However, the EoR 21-cm signal is a highly non-Gaussian field as already discussed in Section 3.4. This non-Gaussianity is expected to affect the total error in the measurement of any statistical estimator of the CD–EoR 21-cm signal. The system noise, on the other hand, is considered to be a Gaussian random field, whose impact in the error co-variance would be competing with the non-Gaussianity of the signal.

Considering the 3D power spectrum $P(\vec{k})$ (Equation 9), the corresponding statistical error covariance between the measurements at $\vec{k}_i$ and $\vec{k}_j$ can be written as (see e.g., Mondal *et al.* 2016; Shaw *et al.* 2019):

$$\mathbf{C}_{ij} = [P_b^2(\vec{k}_i) + P_N^2(\vec{k}_i)]\delta_{ij}^{\mathrm{K}} + \frac{T(\vec{k}_i, -\vec{k}_i, \vec{k}_j, -\vec{k}_j)}{V}.$$

(16)

Here, the P_b and P_N, respectively, denote the 21-cm PS and the system noise power spectrum. Note that $\mathbf{C}_{ij}$ has contribution from the four-point statistics, trispectrum $T(\vec{k}_i, -\vec{k}_i, \vec{k}_j, -\vec{k}_j)$, which would have been zero if the signals were a Gaussian random field.

The trispectrum contribution in Equation (16) is majorly a result of the formation of the ionized bubbles in the IGM. The non-zero trispectrum indicates that the ionized bubble formation is not a random process, therefore, different k modes have correlated information. If the EoR 21-cm signals were a Gaussian random field then the different k mode would have been uncorrelated, and we expect the signal-to-noise ratio (SNR) for $P(k)$ to scale as the square-root of the number of Fourier modes N_k that are being averaged over. However, the study of Mondal *et al.* (2015) show that the SNR saturates to a limiting value, even if one increases the number of Fourier modes (Figure 11). In a follow-up work, Mondal *et al.* (2016) presented a detailed and generic theoretical framework to interpret and quantify
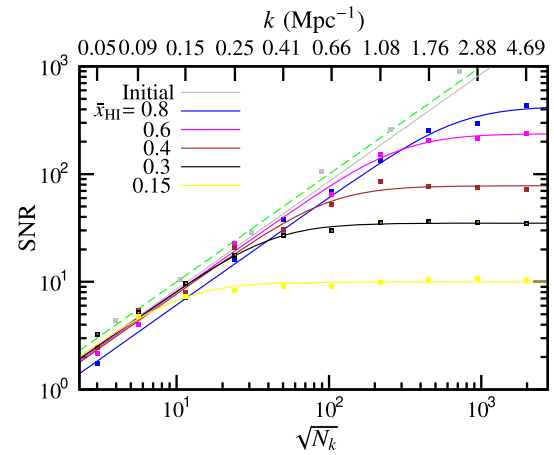

Figure 11. SNR as a function of $\sqrt{N_k}$. The 45° dashed line shows the SNR expected for a Gaussian random field, and 'Initial' refers to the input linear density fluctuations used for the dark matter *N*-body simulations. The other data points are the SNR estimated for the different stages of reionization as mentioned by $\bar{x}_{\mathrm{HI}}$ in the legend. Credit: Mondal *et al.* (2015).

the effect of non-Gaussianity on the error estimates for $P(k)$ through the full error (cosmic) covariance matrix. Notably, Mondal *et al.* (2017) studied how the effect of non-Gaussianity on the reionization power spectrum cosmic variance evolves as the reionization progresses. These works collectively conclude that the effect of non-Gaussianity is dominant on the large k modes (small length-scales) and also towards the later stages of reionization ($\bar{x}_{\mathrm{HI}} \leq 0.5$) as seen in Figure 11. It is important to note that all these three works do not included system noise contributions in their error analysis.

The large system noise contribution to the signal can wash out these non-Gaussian features in the observed CD–EoR 21-cm signal. The system noise power spectrum P_N (see Equation (1) of Shaw *et al.* 2019) scales inversely proportional to the observation time t_{obs} as the Gaussian system noise will reduce down, if averaged over larger number of time stamps. P_N scales inversely also with the distribution of baselines $\vec{U} \propto \vec{k}_\perp$, i.e., the $\vec{k}$ grids at which more baselines contribute will have smaller P_N value. Recently, Shaw *et al.* (2019) presented a methodology for incorporating the non-Gaussianity with the antenna baseline distributions and system noise to make more realistic error predictions for the future SKA-low observations. They have also studied effectiveness of the inherent non-Gaussianity in presence of the system noise by comparing the Gaussian and the non-Gaussian error predictions. Figure 12 shows the percentage deviation of the non-Gaussian error with respect to the Gaussian predictions for early (top panel) and late (bottom panel) stages of reionization for different t_{obs} values. Note that the deviation Δ
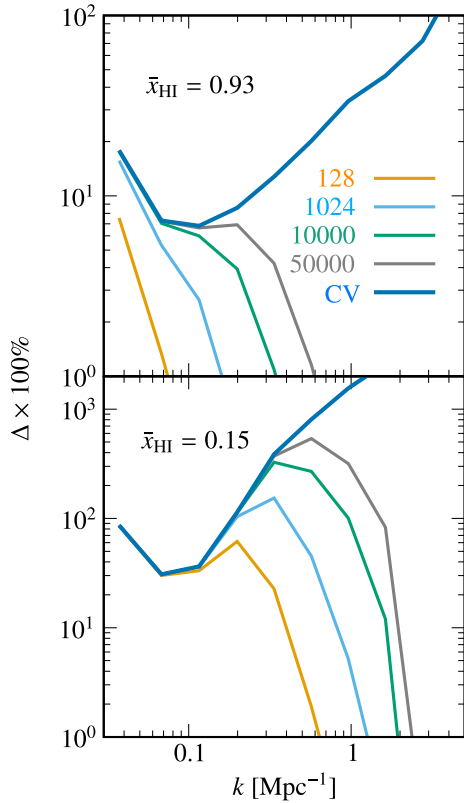
Figure 12. The % deviation of non-Gaussian error variance in observed EoR 21-cm power spectrum with respect to the corresponding Gaussian predictions. Credit: Shaw *et al.* (2019).

is the largest for the cosmic variance (CV) only and it increases with k. However, system noise contribution in variance prevails towards the larger k modes as the baseline density decreases for larger $\vec{U}$ values. Hence, the deviation Δ sharply decreases after the intermediate k values in both the panels. It can also be noted that increasing t_{obs}, decreases the system noise and the trispectrum contribution becomes more prominent. Note that Δ is larger towards the later stages of reionization ($\bar{x}_{HI} = 0.15$) for any particular t_{obs} as the non-Gaussian effects are stronger (CV lines) and also P_N decreases towards smaller redshifts. The impact of non-Gaussianity is significant at $k \lesssim 0.1 \, \mathrm{Mpc}^{-1}$ and $k \lesssim 1 \, \mathrm{Mpc}^{-1}$ during the early (top panel) and late stages (bottom panel), respectively. In a follow-up work, Shaw *et al.* (2020) have studied the effect of the non-Gaussianity on the errors in the inferred REIONYUGA model parameters.

For other statistics, such as MAPS (Mondal *et al.* 2020a) and bispectrum (Mondal *et al.* 2021), one also expects similar qualitative behavior of non-Gaussian effects in presence of system noise in their error covariances.

3.6 *Impact of first sources on statistical estimators*

Several studies have been performed so far to study the impact of different kinds of first sources on the IGM via its impact on the different statistical measures of 21-cm signal from the CD–EoR. Depending on the length-scales of observations, the statistical measures, such as variance, skewness, and power spectrum can distinguish 'inside-out' and 'outside-in' reionization scenarios (Watkinson & Pritchard 2014; Majumdar *et al.* 2016). These statistics can also distinguish the isolated sources, such as Pop III stars, galaxies, mini-QSOs and HMXBs by measuring their signature on the 21-cm signal (Ghara *et al.* 2016; Ross *et al.* 2017, 2019; Islam *et al.* 2019). It is possible to constrain these sources with high SNR using $\sim$1000 h of SKA-low observations in future (Ghara *et al.* 2016). The formation, growth and percolation of HII bubbles in the IGM during EoR carries imprints of ionizing sources. However, the power spectrum, being a two-point statistics, is unable to capture these important non-Gaussian features in the EoR 21-cm signal. The power spectrum can react to different source properties only up to a certain extent by showing the variation in its amplitude, which arises due to the change in the Gaussian component of the signal.

Bispectrum is able to quantify some of the non-Gaussian imprints of the first sources in the EoR 21-cm field. It can respond to the signatures of different sources and processes more efficiently than the power spectrum by reflecting changes in both its sign and magnitude, as illustrated in Figure 10 (Watkinson *et al.* 2019; Kamran *et al.* 2021a). Further, the 21-cm bispectrum is also able to probe the IGM physics during CD by quantifying the non-Gaussianity invoked due to different astrophysical processes (b). Apart from these, the 21-cm bispectrum is also a robust way of distinguishing the dark matter models and can put a better constraint on the nature of dark matter in the future EoR observations as compared to the power spectrum alone (Saxena *et al.* 2020).

3.7 *Line-of-sight anisotropy effects*

The observations of the redshifted CD–EoR 21-cm signal along a particular line-of-sight (LoS) direction ($\hat{n}$) provide additional information related to different unique local effects at the point of emission compared to the other cosmological signals. As already stated in Section 3.2, the major LoS effects are the redshift space distortion (RSD) (Kaiser 1987; Hamilton 1998), light-cone (LC) (Matsubara *et al.* 1997; Barkana & Loeb 2006), and Alcock–Paczynski (AP) (Alcock & Paczynski 1979) effects. The RSD comes into the picture due

to the non-linear peculiar velocities of the gas particles that cause additional redshift or blueshift on top of the cosmological redshift, and hence, distorts the signal along the LoS direction. The LC effect results from the redshift evolution of the signal along the LoS, as the 21-cm signal coming from different redshifts essentially belongs to different frequencies. The AP effect is quite different from these two effects as it is truly related to the geometry of the space-time, which is non-Euclidean.

These LoS effects imprint their signature on the statistical measures of the 21-cm signal by making the signal anisotropic along the LoS axis. In the view of the future SKA-low interferometric observations of the CD–EoR, it is important to understand the impacts of these crucial effects on the signal for its accurate detection and correct interpretation. Therefore, one needs to consider these effects, while estimating the CD–EoR parameters using various statistical estimates, such as the power spectrum and bispectrum, etc.

Bharadwaj & Ali (2004, 2005), for the first time, analytically pointed out the RSD effects in the context of the CD–EoR 21-cm signal. Their study showed that the peculiar velocities significantly change the magnitude and shape of the 21-cm power spectrum when computed using visibilities measured in a radio-interferometric observation. Using the inherent anisotropy in the 21-cm power spectrum, it is possible to extract the matter power spectrum during CD–EoR (e.g., Barkana & Loeb 2005; Shapiro *et al.* 2013). Majumdar *et al.* (2013) are the first to properly implement RSD effects in the simulations of the EoR 21-cm signal. They have quantified the effect of RSD by estimating the monopole and quadrupole moments of the 3D power spectrum. The evolution in the quadrupole moment of the power spectrum with the mean neutral fraction $\bar{x}_{HI}$ at large length-scales ($k \sim 0.12$ Mpc^{-1}) can be used to track the reionization history to a great degree (Majumdar *et al.* 2016). Ghara *et al.* (2015a) and Ross *et al.* (2021) quantified the effect of RSD on 21-cm power spectrum during CD, when the fluctuations in the T_s control the 21-cm fluctuations. They have reported that the effect of RSD on the CD 21-cm power spectrum is not too high as compared to its effect on EoR 21-cm power spectrum.

Using a set of simulated 21-cm signal, Datta *et al.* (2012b, 2014) have performed the first numerical investigation on the LC effect on the 21-cm power spectrum from the EoR. They have found that the LC effect significantly enhances the large-scale power spectrum and suppresses the small-scale power spectrum. They have also found that during the EoR, the LC effect has the largest impact on the 21-cm power spectrum when reionization is ∼20% and another when it is ∼80% com-

pleted. However, they did not find that the LC effect introduces any significant LoS anisotropy to the power spectrum. Ghara *et al.* (2015b) made the first numerical study on the impact of the LC effect on the redshifted 21-cm power spectrum from CD. They have found that the impact of the LC effect is more dramatic when one considers the spin temperature fluctuations in the signal compared to the case when it is ignored.

The Alcock–Paczynski effect is another anisotropy in the signal. This makes any shape that is intrinsically spherical in nature appear elongated along the LoS due to the non-Euclidean geometry of the space-time. It is not much significant at low redshifts, but at high redshifts ($z \gtrsim 0.1$), this causes a substantial distortion in the signal, making the CD–EoR 21-cm power spectrum anisotropic along the LoS. Ali *et al.* (2005) was first to consider the AP effect in the context of the EoR 21-cm signal. They have quantified the relative contribution in anisotropy due to the AP effect when compared with the anisotropy due to the RSD and how they differ in their nature.

Apart from the power spectrum, the LoS anisotropy can affect the higher-order statistics too. Majumdar *et al.* (2020) and Kamran *et al.* (2021a) for the first time quantified the impact of the RSD on the CD–EoR 21-cm bispectrum from the simulated 21-cm signal. They have found that depending on the length scales of the observation, RSD significantly impacts the magnitude, sign and shape of the bispectrum. The LC effect also shows its significant impact on the 21-cm bispectrum during the EoR (Mondal *et al.* 2021). The RSD and LC effects are substantial enough and make them necessary to account for the correct interpretation of the signal statistics.

3.8 *Parameter estimation using models*

We have seen in Section 2 that the astrophysical information is embedded in the 21-cm signal through x_{HI} and T_s terms in the brightness temperature fluctuations (Equation 1). However, the extraction of this information from a measurement of the signal is not trivial and an exploration of many theoretical models of the expected HI signal is necessary to interpret the measurements from radio observations.

First of all, one should realize that it is not feasible to perform a full pixel-by-pixel comparison between the observed signal and theoretical models. Instead, the observed signal is first characterized by one or a combination of statistical estimators, such as mean signal (see e.g., Singh *et al.* 2017; Ghara *et al.* 2022), variance (e.g., Patil *et al.* 2014), power spectrum (e.g., Mertens *et al.* 2020; Greig *et al.* 2021), bispectrum (e.g., Kamran

et al. 2021a), bubble size distribution (e.g., Giri *et al.* 2018), topological measurements (e.g., Kapahtia *et al.* 2021), pattern recognition in 21-cm signal images (e.g., Gillet *et al.* 2019), etc., to perform a comparison with theoretical models (see discussion in Section 3). Among these, the 3D power spectrum has been frequently used in many of the ongoing 21-cm observations. The preferred statistical measures often depend on instrumental sensitivity and might differ from the optimal methods that provide maximum information about the observed redshifts.

The next thing required for the parameter estimation is a fast method (see Section 2) that provides the above mentioned statistical quantities of the signal for a set of model parameters of our interest (e.g., the astrophysical parameters). To obtain the probability distribution function (PDF) in the simulation model parameter space, hundreds of thousands of models may be needed. In case, running such a large number of models are computationally expensive, alternative approaches based on emulators (e.g., Ghara *et al.* 2020) neural networks (Shimabukuro & Semelin 2017) etc., might be taken for parameter inference. For example Ghara *et al.* (2020) build emulators based on Gaussian process regression, which generate power spectrum for a set of source model parameters namely, ionization efficiency (ζ), minimum mass of the UV emitting halos ($M_{\min}$), minimum mass of X-ray emitting halos ($M_{\min,X}$) and X-ray heating efficiency (f_X).

Finally, one needs to perform a robust probabilistic exploration of the source model parameters to unlock astrophysical information from the 21-cm measurements. To explore and constrain the reionization parameters, the signal prediction algorithms are often combined with a Bayesian inference framework. The likelihood used in the Bayesian analysis should take into account all possible uncertainties, both from observational and theoretical sides. It should also distinguish differences between an upper limit and a detection of the signal. Ghara *et al.* (2020) combined GRIZZLY emulators with the Monte Carlo Markov chains (MCMC) to constrain the source parameters at $z \approx 9.1$ using the LOFAR upper limits on the 3D power spectrum of the 21-cm signal (Mertens *et al.* 2020). Figure 13 shows the constraints on the astrophysical parameters used in GRIZZLY as obtained from such an analysis. In addition to constraining the simulation astrophysical model parameters, one can also constraint the IGM quantities, such as average ionization fraction, average gas temperature of the partially ionized IGM that characterize size of the emission regions in the IGM, etc., (see e.g., Ghara *et al.* 2020, 2021).

3.9 *Image-based statistics*

Contrary to the currently operating radio interferometers, which mainly focus to detect the signal statistically, the second phase of upcoming SKA-low will be able to produce tomographic images of the CD–EoR 21-cm signal (Mellema *et al.* 2015; Ghara *et al.* 2017). Thus, use of the CD–EoR tomographic images has been the focus of several theoretical studies, which aim to develop methods for extracting information from such maps. Here, we briefly mention some of these methods.

The expected 21-cm signal from the CD–EoR can be imagined as distributions of complex structures of ionized/neutral or hot/cold regions in space (as can be seen in Figure 5). Such complex topology is often studied in terms of the excursion set $F(\vec{x}, u_{th})$, which represents the part of a given field with values larger than a certain field threshold u_{th}. Such excursion sets of the EoR 21-cm signal have been studied using four Minkowski functionals (MFs) to characterize the geometry of the 3D maps of reionization for different thresholds (see e.g., Gleser *et al.* 2006; Friedrich *et al.* 2011; Chen *et al.* 2019). Among the four Minkowski functionals, three $V_{i=0,1,2}(u_{th})$ stands for the total volume, total surface and mean curvature of $F(\vec{x}, u_t h)$, respectively. The fourth functional $V_3(u_{th})$ is the integrated Gaussian curvature over the surface of $F(\vec{x}, u_{th})$. Studies such as Yoshiura *et al.* (2016) show that these MFs are sensitive to the variation in reionization scenarios due to variation of the background source model. This shows the potential of using MFs in addition to the PS to tighten the constraints on the reionization parameters.

These MFs calculations can be used to characterize the shapes of compact surfaces. Shape-finder algorithms, such as SURFGEN2 (e.g., Bag *et al.* 2019) use these MFs to define thickness $= 3V_0/V_1$, breadth $= V_1/V_2$, length $= V_3/4\pi$ to tract the evolution of the ionized/heated regions. While the thickness and breadth of the largest clustered region grow slowly during the entire reionization process, the length evolves rapidly showing the percolation of ionization regions into a filamentary structure (Bag *et al.* 2018). Recently, Pathak *et al.* (2022) studied the evolution of difference reionization scenarios using the largest cluster statistics (LCS), which can be defined as the ratio of the volume of the largest ionized region and the total volume of all the ionized regions. This study shows that the percolation transition in the ionized regions, as probed by studying the evolution of the LCS, can robustly distinguish the inside-out and outside-in reionization scenarios.

Another generalization of the MFs is a rank-2 contour Minkowski tensor, which can provide alignment
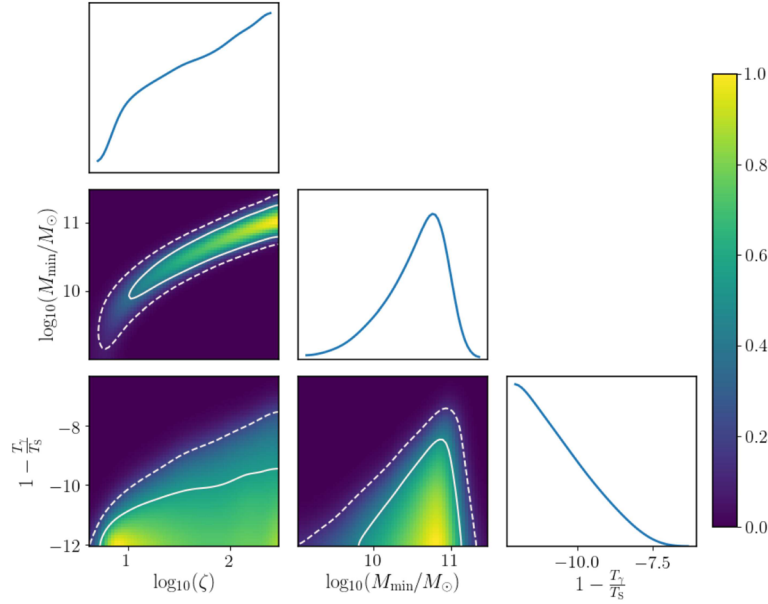
Figure 13. Constraints on the GRIZZLY model parameters from the MCMC analysis using the LOFAR upper limit for $z \approx 9.1$. For this analysis, we assume uniform T_s for the neutral gas in the IGM. The color bar represents the probability that models are ruled out by the upper limit. The 68(95)% credible intervals of the ruled out models are shown by the solid (dashed) curves. The marginalized probability distributions are shown in the diagonal panels. See Ghara *et al.* (2020) for details.

of 2D structures in $F(\vec{x}, u_{th})$ and their anisotropy. The two eigenvalues of the contour Minkowski tensor λ_1 and λ_2 can be combined to construct morphological descriptors, such as the shape anisotropy parameter as $\beta = \lambda_1/\lambda_2$, an effective radius of the enclosed curve $r = (\lambda_1 + \lambda_2)/2\pi$. A region formed by a bounded curve enclosing a set of values larger than u_{th} in the excursion set is known as a connected region. On the other hand, a hole represents a bound curve with values smaller than u_{th}. The count of these connected regions and holes are Betti numbers. Kapahtia *et al.* (2021) explored the prospects of constraining EoR scenarios using these morphological descriptors, Betti numbers, and the areas of structures in the connected regions and holes. The study shows that these morphological descriptors are sensitive to reionization parameters (Figure 14) and can be used to constrain the source parameters as well as can put a strong bound on the ionization fraction at the probed reionization stage.

The complex structures in the ionization fraction maps and 21-cm maps of the EoR are often characterized in terms of distributions of spherical regions (often named as bubbles). Several studies using mean-free path (Mesinger & Furlanetto 2007), granulometry (Kakiichi *et al.* 2017), Watershed methods (Lin *et al.* 2016), image segmentation method (Giri *et al.* 2018) have aimed to study the complex morphology of the signal in terms of the size distribution of these bubbles and their

evolution with time. As the size distribution of the bubbles vary significantly with change in reionization scenario, this in principle can be used for reionization parameter inference.

In addition to these, study of the tomographic images of the 21-cm signal using fractal dimension analysis can also provide distinguishable information between outside-in and inside-out reionization scenarios (see e.g., Bandyopadhyay *et al.* 2017). In addition, application of convolutional neural network constructed using 2D or 3D maps as training set is useful to distinguish between reionization scenarios and constrain astrophysical/cosmological parameters directly using a measured 21-cm input image (see e.g., Gillet *et al.* 2019; Hassan *et al.* 2019).

4. Detection and challenges

Detection of the 21-cm signal and extraction of the information content of it, thereafter will bring us a much clearer picture of the evolution of the cosmic matter since the epoch of recombination. In this section, we discuss different challenges identified in the detection of faint 21-cm signal along with some of the developments in literature that aim to resolve these problems. The majority of these problems come from the fact that the expected 21-cm signal is much fainter than other
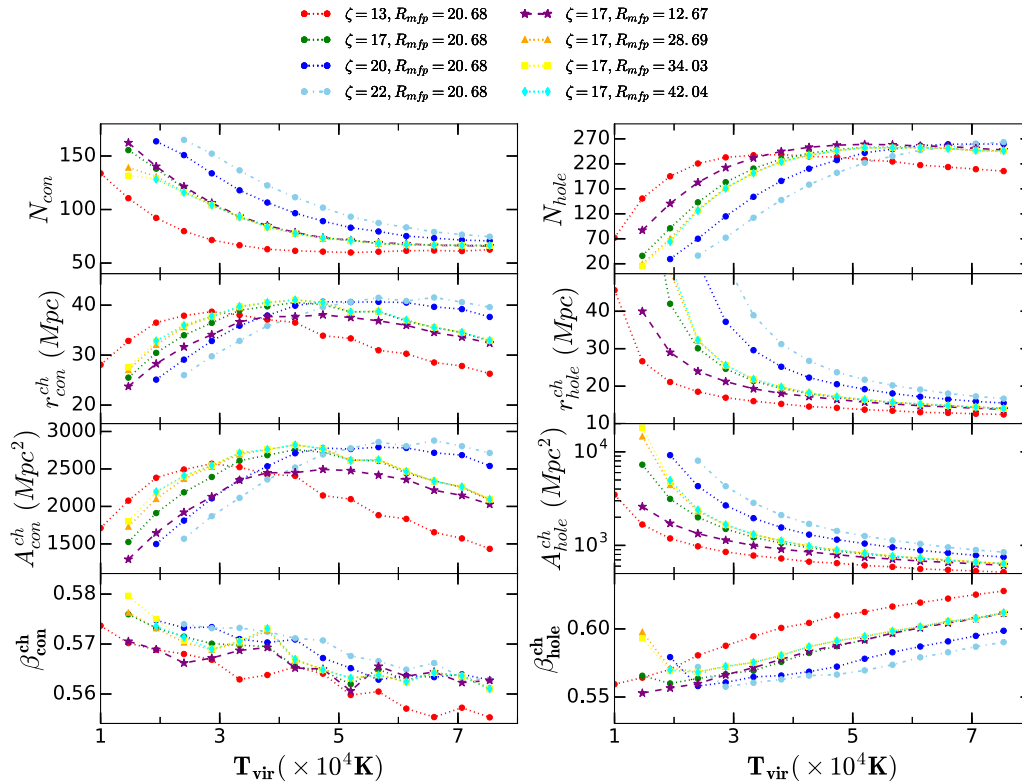
Figure 14. Variation of hole (right panels) and connected region (left panels) statistics with virial temperature T_{vir} for different fixed values of the ionization efficiency ζ and mean-free path R_{mfp}. The simulations are done using 21CMFAST. The top to bottom panels represent the threshold integrated number of distinct curves N_x, characteristic scale r_x^{ch}, area A_x^{ch} and shape anisotropy parameter β_x^{ch}, respectively. Suffix $x = $ con stands for connected regions, while $x = $ hole refers holes. Credit: Kapahtia *et al.* (2021).

signals present in continuum in the same observing frequencies, termed collectively as foreground. Furthermore, the SKA-low and mid telescopes that would have expected sensitivity to measure the 21-cm signals would be able to only estimate statistical measure of the signals in its first phase. Considering the fact that these telescopes measure visibilities with incomplete baseline coverage, different unbiased estimators of the statistical properties of the signal are designed and then tuned to address the problems of mitigating the unwanted signals in these frequencies. Recently, the importance of accurate calibration and the effect of uncalibrated part of the visibilities in detection of 21-cm signals are also being studied. We shall discuss these topics in this section. This section is organized such that the reader is informed about any challenge in the signal and the steps taken to find a resolution.

4.1 *Foregrounds*

We refer to the radiation from different astrophysical sources, other than the cosmological HI signal,

collectively as foregrounds. Foregrounds include extra-galactic point sources, diffuse synchrotron radiation from our galaxy and low redshift galaxy clusters; free–free emission from our galaxy (GFF) and external galaxies (EGFF). Extra-galactic point sources and the diffuse synchrotron radiation from our galaxy largely dominate the foreground radiation at 150 MHz and their strength is ~4–5 orders of magnitude larger than the ~20–30 mK cosmological 21-cm signal (Ali *et al.* 2008; Ghosh *et al.* 2012). The free–free emissions from our galaxy and external galaxies make much smaller contributions though each of these is individually larger than the HI signal.

All the foreground components mentioned earlier are continuum sources. It is well accepted that the frequency dependence of the various continuum foreground components can be modeled by power laws (Santos *et al.* 2005), and we model the multi-frequency angular power spectrum (Datta *et al.* 2007) for each foreground component as:

$$C_\ell(\nu_1, \nu_2) = A \left(\frac{1000}{\ell}\right)^\beta \left(\frac{\nu_f}{\nu_1}\right)^\alpha \left(\frac{\nu_f}{\nu_2}\right)^\alpha, \qquad (17)$$

where A is the amplitude and β and α (spectral index) are the power law indices for the ℓ and the ν dependence, respectively. In general, we are interested in a situation where $\nu_2 = \nu_1 + \Delta\nu$ with $\Delta\nu \ll \nu_1$, and we have

$$C_\ell(\Delta\nu) \equiv C_\ell(\nu_1, \nu_1 + \Delta\nu)$$
$$\approx A \left(\frac{1000}{\ell}\right)^\beta \left(\frac{\nu_f}{\nu_1}\right)^{2\alpha} \left(1 - \frac{\alpha\Delta\nu}{\nu_1}\right), \qquad (18)$$

which varies slowly with $\Delta\nu$. For the foregrounds, we expect $C_\ell(\Delta\nu)$ to fall by $<10\%$, if $\Delta\nu$ is varied from 0 to 3 MHz, in contrast to the $\sim 90\%$ decline predicted for the HI signal (Bharadwaj & Ali 2005; Ali & Bharadwaj 2014). The frequency spectral index α is expected to have a scatter $\Delta\alpha$ in the range of 0.1–0.5 for different foreground components in different directions causing $<2\%$ additional deviation in the frequency band of our interest. We only refer the mean spectral indices for the purpose of the foreground presented here. In a nutshell, the $\Delta\nu$ dependence of $C_\ell(\Delta\nu)$ is markedly different for the foregrounds as compared to the HI signal and would be very useful to separate the foregrounds from the HI signal (Ghosh *et al.* 2012).

Extragalactic point sources are expected to dominate the 150 MHz sky at the angular scales relevant for redshifted 21-cm observations. The contribution from extragalactic point sources is mostly due to the emission from normal galaxies, radio galaxies, star forming galaxies and active galactic nuclei (Santos *et al.* 2005; Singal *et al.* 2010; Condon *et al.* 2012).

There are different radio surveys that have been conducted at various frequencies ranging from 151 MHz to 8.5 GHz, and these have a wide range of angular resolutions ranging from $1''$ to $5'$ (e.g., Singal *et al.* 2010 and references therein). There is a clear consistency among the differential source count functions ($dN/dS \propto S^{-\epsilon}$) at 1.4 GHz for sources with flux $S > 1$ mJy. The source counts are poorly constrained at $S < 1$ mJy. Based on the various radio observations (Singal *et al.* 2010) we have identified four different regimes for the 1.4 GHz source counts: (i) $\gtrsim 1$ Jy, which are the brightest sources in the catalogs. These are relatively nearby objects and they have a steep, Euclidean source count with $\epsilon \sim 2.5$; (2) 1 mJy–1 Jy, where the observed differential source counts decline more gradually with $\epsilon \sim 1.7$, which is caused by redshift effects; (3) 15 μJy–1 mJy, where the source counts are again steeper with $\epsilon > 2$, which is closer to Euclidean, and there is considerable scatter from field to field; and (4) ~ 15 μJy, the source counts must eventually flatten ($\epsilon < 2$) at low S to avoid an infinite integrated flux. The cut-off lower flux, where

the power law index ϵ falls below 2 is not well established, and deeper radio observations are required. The first turnover or flattening flux in the 1.4 GHz differential source count has been reported at ~ 1 mJy (Condon 1989; Hopkins *et al.* 2003; Owen & Morrison 2008) and it is equivalent to ~ 5 mJy at 150 MHz assuming a spectral index of 0.7 (Blake *et al.* 2004; Randall *et al.* 2012).

The analysis of large samples of nearby radio-galaxies has shown that the point sources are clustered (Cress *et al.* 1996; Wilman *et al.* 2003; Blake *et al.* 2004). The measured two point correlation function can be well-fitted with a single power law $w(\theta) = (\theta/\theta_0)^{-\beta}$, where θ_0 is the correlation length. We can calculate w_ℓ, which is $\propto \ell^{\beta-2}$ and the Legendre transform of $w(\theta)$. Using this and source counts, we have modeled the angular power spectrum due to clustering and Poisson contribution of point sources (Ali *et al.* 2008; Ali & Bharadwaj 2014).

The galactic diffuse synchrotron radiation is believed to be produced by cosmic ray electrons propagating in the magnetic field of the galaxy (Rybicki & Lightman 1979). Angular structure in the diffuse Galactic emission has been shown to be well described by a power-law spectrum in Fourier space over a large range of scales. La Porta *et al.* (2008) have determined the angular power spectra of the Galactic synchrotron emission at angular scales $>0.5°$ using total intensity all sky maps at 408 MHz (Haslam *et al.* 1982) and 1.42 GHz (Reich 1982; Reich & Reich 1986). However, there are only a few observations, which have directly measured C_ℓ at the frequencies and angular scales, which are relevant to the EoR studies (Bernardi *et al.* 2009; Parsons *et al.* 2010; Ghosh *et al.* 2012; Iacobelli *et al.* 2013; Choudhuri *et al.* 2020). They found that the angular power spectrum of synchrotron emission is well described by a power law (Equation 17) where the value of β varies in the range of 1.8–3.0 depending on the galactic latitude. La Porta *et al.* (2008) have analyzed the frequency dependence to find $A \propto \nu^{-2\alpha}$ with α varying in the range of 2.8–3.2. There is a modest variation in the spectral index as a function of direction and frequency. The mean spectral index of the synchrotron emission at high Galactic latitude has been recently constrained to be $\alpha \approx 2.5$ in the 150–408 MHz frequency range (Rogers & Bowman 2008) using single dish observations. In general, it is steeper at high Galactic latitudes than toward the Galactic plane. The GFF and EGFF components, which are relatively much weaker foregrounds as compared to earlier two. But both are stronger than the redshifted 21-cm signal. Their α and β are presented at 130 MHz (Santos *et al.* 2005).

4.2 *Statistical detection of 21-cm signal*

Most of the radio interferometers measure visibility function sampled at certain baseline positions. Here, we discuss development of various visibility-based estimators of the power spectrum of the sky signal. These estimators are unbiased within the scope of the signal considered. Different estimators discussed here also address and solve various challenges that one face for 21-cm detection.

4.2.1 Bare estimator   The Bare estimator measures the angular power spectrum (C_ℓ), which quantifies the intensity fluctuations in the two-dimensional sky plane. It uses individual visibilities to measure C_ℓ. As the visibility $\mathcal{V}_i$ at a baseline $\vec{U}_i$ corresponds to a Fourier mode in the sky, the two visibility correlation straight away gives the angular power spectrum, which can be written as:

$$\langle \mathcal{V}_i \mathcal{V}_j^* \rangle = V_0 e^{-|\Delta\vec{U}_{ij}|^2/\sigma_0^2} C_{\ell_i} + \delta_{ij} 2\sigma_n^2, \qquad (19)$$

where $V_0 = (\pi\theta_0^2/2)(\partial B/\partial T)^2$, $\Delta\vec{U}_{ij} = \vec{U}_i - \vec{U}_j$ and the Kronecker delta δ_{ij} is nonzero, only if we correlate a visibility with itself. For the Gaussian approximation of the primary beam $\theta_0 = 0.6\theta_{\text{FWHM}}$, B is the Planck function and $(\partial B/\partial T) = 2k_B/\lambda^2$ in the Raleigh–Jeans limit, which is valid at the frequencies considered here. The σ_n is the rms of the real and imaginary parts of noise in the measured visibilities. We thus see that the visibilities at two different baselines $\vec{U}_i$ and $\vec{U}_j$ are correlated, only if the separation is small ($|\Delta U| \leq \sigma_0$), and correlation falls as the separation is beyond a disk of radius σ_0.

To avoid the positive noise bias in the second term of Equation (19), we define the Bare[7] estimator as:

$$\hat{E}_B(a) = \frac{\sum_{i,j} w_{ij} \mathcal{V}_i \mathcal{V}_j^*}{\sum_{i,j} w_{ij} V_0 e^{-|\Delta\vec{U}_{ij}|^2/\sigma_0^2}} = \frac{Tr(\mathbf{w}\mathbf{V}_2)}{Tr(\mathbf{w}\mathbf{I}_2)}. \qquad (20)$$

The weight $w_{ij} = (1 - \delta_{ij})$ is chosen, such that it is zero when we correlate a visibility with itself, thereby avoiding the positive noise bias. We have the matrices: $\mathbf{w} \equiv w_{ij}$, $\mathbf{V}_2 \equiv \mathcal{V}_i \mathcal{V}_j^*$, $\mathbf{I}_2 = V_0 e^{-|\Delta\vec{U}_{ij}|^2/\sigma_0^2}$ and $Tr(\mathbf{A})$ denotes the trace of a matrix $\mathbf{A}$.

The variance of the Bare estimator can be simplified to

$$\sigma_{E_B}^2(a) = \frac{\sum_{i,j,k,l} w_{ij} w_{kl} V_{2il} V_{2kj}}{[\text{Tr}(\mathbf{w}\mathbf{I}_2)]^2} = \frac{\text{Tr}(\mathbf{w}\mathbf{V}_2\mathbf{w}\mathbf{V}_2)}{[\text{Tr}(\mathbf{w}\mathbf{I}_2)]^2}, \qquad (21)$$

under the assumptions that $\mathbf{w}$ is symmetric. The detailed formalism of the Bare estimator and the validation with realistic simulations are given in Section 4 of Choudhuri *et al.* (2014).

4.2.2 Tapered gridded estimator   Point sources are the most dominant foreground components at angular scale $\leq 4°$ (Ali *et al.* 2008). Due to the highly frequency-dependent primary beam, it is very difficult to remove the point sources from the outer edge of the primary beam. These outer point sources create a oscillation along the frequency axis (Ghosh *et al.* 2011) and make it difficult to remove under the assumption of smoothness along with frequency. This issue is not addressed in the bare estimator. The tapered gridded estimator (TGE[8]) incorporates three novel features: First, the estimator uses the gridded visibilities, which makes it computationally much faster for large data volume. Second, the noise bias is removed by subtracting the auto-correlation of the visibilities from each grid point. Third, the estimator also taper the FoV to restrict the contribution from the sources in the outer regions and the sidelobes. The mathematical formalism and the variance of the TGE are given in Equations (17) and (25) of Choudhuri *et al.* (2016).

In Figure 15, we show the validation of the TGE using realistic simulation of GMRT observations at 150 MHz . The red solid line shows the model power-law power spectrum $C_\ell^M = A_{150} (1000/\ell)^\beta$ used for the simulation. Here, we have used $A_{150} = 513 \, \text{mK}^2$ and $\beta = 2.34$. The different color points are for different types of the tapering values $f = 10, 2$ and 0.6. The value of $f = 10$ essentially corresponds to a situation with no tapering, and the sky response gets confined to a progressively smaller region as the value of f is reduced to $f = 2.0$ and 0.6, respectively. The points here indicates the estimated values of the C_ℓ with 1σ error bar after applying the TGE. We see that the TGE is able to recover the input model C_ℓ^M quite accurately for all values of tapering function f.

4.2.3 Image-based tapered gridded estimator   The effect of sources outside the field-of-view of an interferometric observations for the estimation of the power spectrum of redshifted 21-cm emission is addressed in TGE by introducing an analytical tapering function in the visibility plane (Choudhuri *et al.* 2016). In such observations of 21-cm signal, significant contribution from localized foreground emission from within the field-of-view also can overwhelms the HI signal. These
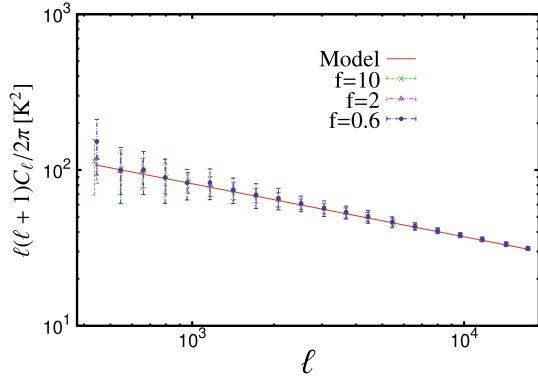
---

[7]https://github.com/samirchoudhuri/BareEstimator.

[8]https://github.com/samirchoudhuri/TGE.

Figure 15. Estimated values of C_ℓ from the simulated visibilities using the TGE for different tapering of values $f = 10$, 2 and 0.6, with $1 - \sigma$ error bars estimated from $N_r = 128$ realizations of the simulations. We also show the input model with red solid line for comparioson. This figure is adapted from Figure 2 of Choudhuri *et al.* (2016).

localized emissions are often compact and can be subtracted from the visibilities with a reasonable model by the method of UVSUB. However, it is rather difficult to sufficiently accurately model localized, but extended sources and a simple UVSUB may not work. Addressing this issue with the TGE is also difficult owing to the complicated nature of window that one need to model analytically to reduce the foreground effects. It is possible to introduce an image-based tapering algorithm, where a wide variety of non-trivial and non-analytical tapering functions can be used. This method of estimating the power spectrum directly from visibilities with arbitrary tapering window implementation in the image plane is implemented in the image-based tapered gridded estimator (ITGE). A detailed discussion of the algorithm along with diagnostics tests are discussed in Choudhuri *et al.* (2019), and we give a brief description here.

As for the case of TGE, the basic expression that coins the estimation of power spectrum at a grid is given in Equation (17) of Choudhuri *et al.* (2016). To use an image-based tapering window, the following steps are followed. First, the visibilities are gridded in baseline plane with suitable choice of grid size and weighing schemes. A dirty image is made from the gridded visibilities and multiplied with the chosen tapering window $W(\theta)$ in the image plane. An inverse Fourier transform of this provides the tapered and convolved visibilities in a grid. Apparently, these gridded visibilities then can be squared to estimate the power spectrum. However, as discussed for the TGE, this would introduce a large bias from the noise contribution in the self visibility correlations. To avoid the noise bias in the estimated power
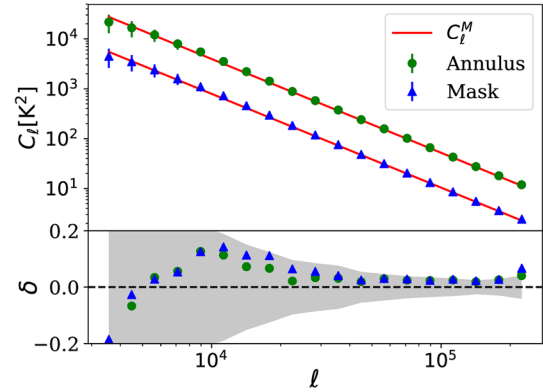

Figure 16. Estimated and model C_ℓ using ITGE (Section 4.2.3) for two different type of windows applied in the image plane: (a) 2D annulus window and (b) mask window with three holes present within the main lobe of the primary beam. We also show the model C_ℓ^M with red solid lines. The lower panel shows the fractional deviation between the estimated and model angular power spectrum. The shaded region in the lower panels show the expected statistical fluctuations σ/C_ℓ^M. For more details, we refer the reader to Figure 4 of Choudhuri *et al.* (2019).

spectrum, the TGE algorithm needs to estimate the self correlations of tapered visibilities in each grid. In ITGE algorithm, correlations of the visibilities in each grid are estimated. In parallel to this, the Fourier transform of the tapering window $\tilde{w}_g$ is evaluated at the grids of the baseline plane. Two dirty images, one corresponds to the visibility self correlation and other of $|\tilde{w}(\vec{U})|^2$ are estimated and multiplied in the image plane. Inverse Fourier transform of the later gives the self correlations of tapered visibilities in each baseline grid. Rest of the algorithm as well as the error estimation follows directly from the original TGE.

Choudhuri *et al.* (2019) have also validated the estimator for two window functions, namely, annulus window and mask window. The annulus window allows us to estimate C_ℓ only from the outer annular region, where as the mask window has three holes within the main lobe of the primary beam. The upper panel of Figure 16 shows the estimated C_ℓ along with the input model C_ℓ for these two different windows applied in the image plane. The results have been arbitrarily scaled for clarity of presentation. The lower panel of Figure 16 shows the fractional deviation between the estimated and model angular power spectrum. The shaded region in the lower panels show the expected statistical fluctuations σ/C_ℓ^M. We see that the fractional deviation is $<15\%$ for whole ℓ range, also within 1σ of the statistical fluctuations.

4.2.4 3D tapered gridded estimator The power spectrum estimator addressed so far estimates the power spectrum on the sky plane at the same frequency channel. To measure the fluctuations along the line of sight or, equivalently, along frequency directions, we have further extended the two-dimensional TGE, to three-dimensional TGE which measures the fluctuations from a 3D cube. Here, we first multiply the data along the frequency with a Blackman–Nuttall (Nuttall 1981) window function to make data periodic, and then use fastest Fourier transform in the west (FFTW) (Frigo & Johnson 2005) algorithm along the frequency axis to convert the visibilities in delay space (τ), the Fourier conjugate of frequency (ν). After conversion in delay space, we apply the same 2D TGE (earlier section) in each slice of τ, and further, we spherically bin the data to measure the power spectrum as a function of k. The mathematical expression for the 3D TGE (Equation 44 of Choudhuri *et al.* 2016) is

$$\hat{P}_g(\tau_m) = \left(\frac{M_g B_{bw}}{r^2 r'}\right)^{-1} \left(|v_{cg}(\tau_m)|^2 - \sum_i |\tilde{w}(\vec{U}_g - \vec{U}_i)|^2 |v_i(\tau_m)|^2\right), \qquad (22)$$

where $v_i(\tau_m)$ and $v_{cg}(\tau_m)$ are the individual and convolved-gridded visibilities in delay (τ) space, respectively, $\tilde{w}(\vec{U})$ is the Fourier transform of the tapering window function used to suppress the primary beam side lobes in the sky plane, M_g is the normalization constant and it depends on the baseline distribution and the form of the tapering function. We used simulated visibilities corresponding to an unit angular power spectrum (UAPS) $C_\ell = 1$ to estimate M_g. The B_{bw} is the bandwidth of the observation, r is the co-moving distance corresponding to the redshifted 21-cm radiation at the observing frequency ν, $r' = |dr/d\nu|$.

Figure 17 shows the dimensionless spherically-averaged power spectrum for the model $P(k) = k^{-3}$. The results are shown for the three different values of tapering windows parametrized by $f = 10, 2$ and 0.6 to demonstrate the effect of varying the tapering. For all the values of f, we found that the estimated power spectrum using the 3D TGE is within the $1 - \sigma$ error bars of the model prediction for the entire k range considered here.

4.2.5 MAPS and power spectrum Nearly all the estimators for $P(k_\perp, k_\parallel)$ (Morales 2005; Parsons *et al.* 2012; Liu *et al.* 2014a, b; Trott *et al.* 2016; Mertens *et al.* 2020), including the 3D-TGE (see earlier sub-section), consider a Fourier transform (FT) of the measured visibilities $\mathcal{V}_i(\nu)$ along the frequency ν to obtain the
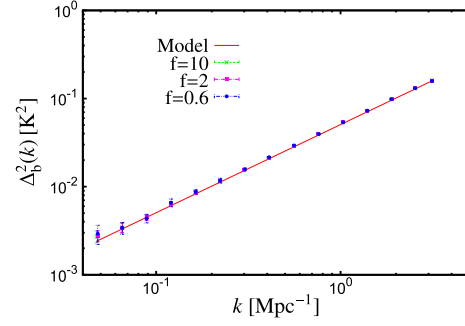

Figure 17. Dimensionless power spectrum $\Delta_b^2(k)$ for different values of tapering function f. The values obtained using the 3D TGE are compared with model power spectrum shown with red solid line. The $1 - \sigma$ error bars are estimated using 16 different realizations of the simulated visibilities. This figure is adapted from Figure 7 of Choudhuri *et al.* 2016.

visibilities $\mathcal{V}_i(\tau)$ in the delay space τ, which is then used to estimate the cylindrical power spectrum $P(k_\perp, k_\parallel)$. In a real observation, particularly at low frequencies, quite often many of the frequency channels are flagged. As the quantity $\mathcal{V}_i(\tau)$ is estimated by a direct Fourier transform of the frequency axes of the visibilities, this gives rise to a convolution with the FT of the frequency sampling function, and thus introduces artifacts in the measured $P(k_\perp, k_\parallel)$. To mitigate this problem, Bharadwaj *et al.* (2019) further developed the MAPS-based TGE estimator that estimates the power spectrum of the cosmological 21-cm signal by first estimating the MAPS $C_\ell(\Delta\nu)$ and subsequently, by taking a FT of the $C_\ell(\Delta\nu)$ along the $\Delta\nu$. Note that the MAPS-based TGE estimator preserves all the salient features of the 2D and 3D TGE mentioned in the earlier subsection, while also providing a way to eliminate the challenges presented by flagging, usually prevalent in the observed radio data.

The details of the MAPS-based estimator can be found in Bharadwaj *et al.* (2019) and Pal *et al.* (2021). To summarize, we first estimate the gridded visibilities $\mathcal{V}_{cg}(\nu_a)$ at a grid point $\vec{U}_g$ on a rectangular grid in the UV plane, by convolving the visibilities $\mathcal{V}_i(\nu_a)$ with the window function $\tilde{w}(\vec{U})$ at all the frequency channels ν_a, where $1 \leq a \leq N_c$, N_c being the total number of channels. The convolution of the visibilities in the UV plane modulates the PB pattern by a multiplication in the image domain with the FT of the convolving function $\tilde{w}(\vec{U})$, which is suitably chosen to taper the sky response away from phase center and especially, outside the main lobe of the PB. We then have:

$$\mathcal{V}_{cg}(\nu_a) = \sum_i \tilde{w}(\vec{U}_g - \vec{U}_i)\mathcal{V}_i(\nu_a)F_i(\nu_a), \qquad (23)$$

where the variable $F_i(\nu_a)$ incorporates the flagging, and has a value '0' for the flagged data and '1' otherwise. The TGE for $C_\ell(\nu_a, \nu_b)$ is then defined as:

$$\hat{E}_g(\nu_a, \nu_b) = M_g^{-1}(\nu_a, \nu_b)\mathcal{R}e\left[\mathcal{V}_{cg}(\nu_a)\mathcal{V}_{cg}^*(\nu_b)\right.$$

$$\left. - \sum_i F_i(\nu_a)F_i(\nu_b)|\tilde{w}(\vec{U}_g - \vec{U}_i)|^2\mathcal{V}_i(\nu_a)\mathcal{V}_i^*(\nu_b)\right],$$
(24)

where $\mathcal{R}e[..]$ denotes the real part of the terms within [..] and $M_g(\nu_a, \nu_b)$ are the normalization factors determined using simulations. Note that the second term in the brackets denotes the self-correlation of the visibilities, and are subtracted out to eliminate the noise bias from the estimator considering that the noise in the visibility measurements at different baselines and timestamps are uncorrelated. The TGE described in Equation (24) provides an unbiased estimate of the $C_{\ell g}(\nu_a, \nu_b)$ at the angular multipole $\ell_g = 2\pi U_g$, i.e.,

$$\langle\hat{E}_g(\nu_a, \nu_b)\rangle = C_{\ell g}(\nu_a, \nu_b). \quad (25)$$

The MAPS $C_\ell(\nu_a, \nu_b)$ does not assume any ergodicity of the sky signal along the LoS and can be used to quantify the second order statistics of the cosmological 21-cm signal including the light-cone effect for wide-bandwidth observations. Conversely, considering a small observation bandwidth, the redshifted 21-cm signal can be assumed to be statistically homogeneous (ergodic) along the LoS (e.g., Mondal *et al.* 2018), allowing us to express $C_\ell(\nu_a, \nu_b)$ (Equations 24 and 25) in terms of $C_\ell(\Delta\nu)$, where $\Delta\nu = |\nu_b - \nu_a|$. This transformation implies that the statistical properties of the signal depends only on the frequency separation and not the individual frequencies. In the flat sky approximation, the power spectrum $P(k_\perp, k_\parallel)$ of the brightness temperature fluctuations of the redshifted 21-cm signal is then given by Datta *et al.* (2007):

$$P(k_\perp, k_\parallel) = r^2 r' \int_{-\infty}^{\infty} d(\Delta\nu)e^{-ik_\parallel r'\Delta\nu}C_\ell(\Delta\nu), \quad (26)$$

where $k_\parallel$ and $k_\perp = \ell/r$ are the components of $\mathbf{k}$, respectively, parallel and perpendicular to the line of sight, r is the co-moving distance and r' ($= dr/d\nu$), both being evaluated at the central observation frequency ν_c. Starting from Equation (26), Pal *et al.* (2021) have shown that given the estimates of $C_\ell(n\Delta\nu_c)$, $\Delta\nu_c$ being the smallest channel separation with $0 \leq n \leq N_c - 1$, the maximum likelihood estimate (MLE) of the cylindrical power spectrum $\bar{P}(k_\perp, k_{\parallel m})$ is given by,

$$\bar{P}(k_\perp, k_{\parallel m}) = \sum_n \{[\mathbf{A}^\dagger\mathbf{N}^{-1}\mathbf{A}]^{-1}\mathbf{A}^\dagger\mathbf{N}^{-1}\}_{mn}C_\ell(n\Delta\nu_c),$$
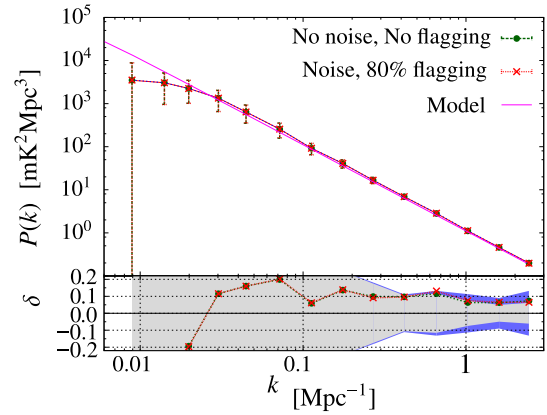(27)

Figure 18. Upper panel shows the estimated spherically-binned power spectrum $P(k)$ along with $1 - \sigma$ error-bars obtained from the simulations with no noise and flagging and also with noise and 80% flagging. The input model $P^m(k)$ is shown by the solid magenta line for comparison. The bottom panel shows the fractional error $\delta = [P(k) - P^m(k)]/P^m(k)$ (shown by the data points) and the relative statistical fluctuation $\sigma/P^m(k)$ (shown by the shaded regions). See Bharadwaj *et al.* (2019) for details.

where $\mathbf{A}$ refers to the $N_c \times N_c$ Hermitian matrix corresponding to the FT coefficients, $\mathbf{N}$ refers to the noise covariance matrix and '$\dagger$' denotes the Hermitian conjugate.

The MAPS-based TGE estimator has been validated using realistic simulations of GMRT observations at 150 MHz. Using a power-law model power spectrum $P^M(k)$, Bharadwaj *et al.* (2019) have shown that the TGE can recover the input MAPS as well as the $P^M(k)$ with an accuracy of 5–20% over a large k range, even in the presence of the system noise and when 80% of the data, chosen randomly, are flagged (Figure 18). The signal in the visibility measurements $\mathcal{V}_i(\nu_a)$ at different baselines $\vec{U}_i$ are not independent, and gets correlated at baselines within a scale of D/λ due to the telescope's PB pattern, D being the antenna diameter. Similarly, the visibility measurements $\mathcal{V}_i(\nu_a)$ at different ν_a are not independent and the signal gets correlated across different channels, the width of the correlation depending on the value of $\vec{U}_i$ (see Bharadwaj & Pandey 2003; Bharadwaj & Ali 2005 for more details). The signal contained in the flagged data, which is lost due to the flagging thus also gets contained in the valid data, which is available to us for the power spectrum estimation, and consequently the MAPS-based TGE is able to recover the power spectrum equally well even if 80% of the data is flagged.

4.3 *Observations with GMRT*

The Giant Metrewave Radio Telescope (GMRT) is one of the largest and most sensitive fully operational low-frequency radio telescopes in the world today (Swarup *et al.* 1991) with 30 fully steerable 45-m diameter dishes spread with maximum antenna separation of 25 km. Recently GMRT has been upgraded to uGMRT with some extra features, such as: (i) huge frequency coverage, from 120 to 1500 MHz; (ii) maximum bandwidth available is 400 MHz instead of 32 MHz bandwidth of original GMRT design; (iii) digital backend correlator catering to 400 MHz bandwidth; (iv) improved receiver systems with higher G/T_{sys} and better dynamic range (Gupta *et al.* 2017). The uGMRT is considered as one of the pathfinder of the SKA telescopes.

The expected brightness temperature of redshifted HI signal from the EoR and post-reionization epoch is many orders of magnitude fainter than the radio emissions from different galactic and extragalactic foregrounds (Zaldarriaga *et al.* 2004; Bharadwaj & Ali 2005; Santos *et al.* 2005). The challenges are nearly identical for both EoR and post-reionization experiments. So the knowledge of foregrounds at post-reionization epoch can also help us to understand the intricacies involved in detection of the HI signal coming from EoR. Accuracy of extraction of the cosmological HI signal strongly depends on the ability to characterize and remove the foregrounds from observational data sets at the frequency of redshifted HI 21-cm line. There have been a considerable amount of efforts put forward over the last decade in characterizing the statistical nature of foregrounds, calibration issues, instrumental systematics and putting upper limits on HI signal both from EoR and post-EoR epochs using sensitive low-frequency observations with GMRT. Here, we will briefly mention different findings using low-frequency radio observations so far with GMRT.

Ali *et al.* (2008) observed a field centered on Upsilon Andromedae, an extra-solar planetary system, at $\alpha_{2000} = 1^h 36^m 48^s$, $\delta_{2000} = 41°24'23''$ for a total of 14 h using GMRT software backend (GSB) system. The observation was conducted at 153 MHz, corresponding to the redshift $z \sim 8.28$, with 8 MHz bandwidth subdivided into 128 frequency channels, giving a spectral resolution of 62.5 KHz. They have studied, for the first time, the statistical nature of foregrounds on sub-degree angular scales using this observation at a redshift corresponding to the epoch of reionization. They have used the correlations among measured visibilities to directly determine the multi-frequency angular power spectrum $C_\ell (\Delta \nu)$ (MAPS, Datta *et al.* 2007). The visibility cor-

relation as a function of baseline distance or the angular scale l is shown in Figure 19. They have found that the measured $C_\ell (\Delta \nu)$ before point source subtraction has a value around 10^4 mK^2. This is seven order of magnitude stronger than the expected redshifted HI signal. After subtracting the point-source model from the data, they have estimated the angular power spectrum of the residual data, which shows a power-law-like feature corresponding to the diffuse galactic synchroton emisson (DGSE). They have found that the DGSE is dominated at short baselines (large angular scales), $U < 150$ and the clustering of point sources dominate in the range $150 < U < 2 \times 10^3$ and the residual point source Poisson fluctuations dominate at $U > 2 \times 10^3$. For a given angular multipole (ℓ), the foreground signal is expected to be smooth across frequency. However, Ali *et al.* (2008) reported a oscillatory feature across frequency, which is most likely due to calibration error.

Paciga *et al.* (2011) put the first upper limit on the 21-cm power spectrum at redshift $z \sim 8.6$, using ~ 40 h of observation with the legacy GMRT. The chosen field was centered around a pulsar B0823 + 26, having a period of 0.53 s and an average flux of 350 mJy at 150 MHz, whose on-pulse flux is about 6 Jy, brighter than all other sources in the field, making it a good calibrator. After calibration and RFI removal, they fit a piece-wise linear function to each baseline at each time stamp and subtract it to remove the contamination by bright spectrally smooth foregrounds. After subtracting the foregrounds, they cross-correlate multiple night's data set and estimate the power spectrum. The 2σ upper limit on HI 21-cm power spectrum at redshift $z \sim 8.6$ is $(70$ mK$)^2$ at $k = 0.65$ h Mpc^{-1} (Paciga *et al.* 2011).

A revised upper limit using the same data was reported by Paciga *et al.* (2013). In the previous work, they did not account the signal loss associated with their foreground subtraction algorithm. In this work, they used singular value decomposition (SVD). After removing the first 4 SVD modes, which are mostly dominant by foregrounds, they quantified the signal loss by quantifying the transfer function between observed power after SVD modes subtraction and the fiducial 21-cm signal model. They put a revised 2σ upper limit about $(248$ mK$)^2$ at $k = 0.5$ h Mpc^{-1} (Paciga *et al.* 2013).

Ghosh *et al.* (2012) have analyzed four different fields: FIELD I ($\alpha_{2000} = 5^h 30^m 00^s$, $\delta_{2000} = 60°00'00''$), FIELD II ($\alpha_{2000} = 6^h 00^m 00^s$, $\delta_{2000} = 62°12'58''$), FIELD III ($\alpha_{2000} = 12^h 36^m 49^s$, $\delta_{2000} = 62°12'58''$), FIELD IV ($\alpha_{2000} = 1^h 36^m 48^s$, $\delta_{2000} = 41°24'23''$) using GMRT observations at 150 MHz ($z \sim 8.4$). They found that after subtracting the point
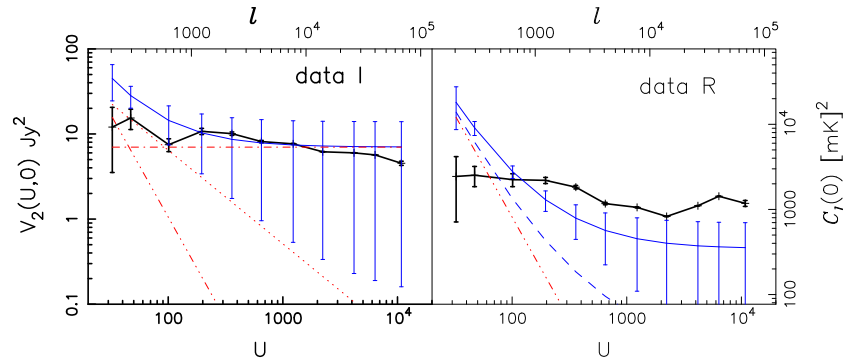
Figure 19. Visibility correlation as a function of baseline distance for total data set (I; left panel) and the residual data set (R; right panel). For data I, the thin solid line shows the total model prediction for $S_c = 900$ mJy. Also shown are the contributions from point source Poisson distribution (dash–dotted), point source clustering (dot) and galactic synchrotron (dash–dot–dot–dot). For data R, the thin solid line shows the total model predictions for $S_c = 100$ mJy and the long dashed line for 10 mJy. The dash–dot–dot–dot curve shows the galactic synchrotron contribution.

sources from FIELD I data, the residual data is dominated by galactic synchrotron radiation at ℓ range $253 \leqslant \ell \leqslant 800$, which they modeled as $C_\ell \propto \ell^{-\beta}$. The estimated fluctuations in the galactic synchrotron emission is $\sqrt{\ell(\ell+1)C_\ell/(2\pi)} \sim 10$ K at $\ell = 800$, corresponding to $\theta > 10'$. The residual point sources and artifacts dominate C_ℓ at smaller angular scales, where $C_\ell \sim 10^3$ mK^2 for $\ell > 800$. For other three fields, they found that the data is dominated by residual artifacts and they did not quantify the statistical properties of foregrounds for these fields. The point source subtracted residual map is shown in the top panel of Figure 20. Ghosh *et al.* (2012) also estimated the Euclidean normalized differential source counts as a function of flux density bins and found that the radio source population in the field is dominated by radio-loud AGNs and there is no signature of flattening in the source count at low flux density regime. They also quantified the spectral variation of angular power spectrum of foregrounds at a particular ℓ, $C_\ell (\Delta \nu)$, using MAPS (Datta *et al.* 2007). The measured $C_\ell(\Delta\nu)$ as a function of l for FIELD I is shown in the bottom panel of Figure 20. They found that $C_\ell (\Delta \nu)$ shows a smooth frequency dependence over 2.5 MHz $\Delta \nu$ range with 20–40% variation. However, in addition to this, they also found a oscillatory feature, similar to Ali *et al.* (2008), in the $\Delta \nu$ dependence of the measured $C_\ell (\Delta \nu)$. In contrast to Ali *et al.* (2008), where point source model were subtracted above a flux cut off, Ghosh *et al.* (2012) used a polynomial fit across frequency to model and subtract the foregrounds. However, if this oscillatory feature across frequency remained in the data due to calibration error or unknown systematics, then it will significantly affect any attempt to detect cosmological Hɪ signal.

Choudhuri *et al.* (2017) have analyzed two different pointings of TIFR all sky survey (TGSS; Intema *et al.* 2017) at 150 MHz near the galactic plane $(9°, +10°)$ and $(15°, -11°)$ to characterize the statistical properties of DGSE. They applied the novel 2-dimensional TGE to the low-frequency observation (150 MHz) with GMRT to quantify the angular power spectrum of the diffuse foreground from the measured visibility data set. They found that the measured angular power spectrum shows a power law behavior in the ℓ range $240 \leqslant \ell \leqslant 580$ and $240 \leqslant \ell \leqslant 440$ and the best fitted values of (A, β) are $(356, 2.8)$ and $(54, 2.2)$ for two different fields, respectively. Recently, Choudhuri *et al.* (2020) analyzed the whole TGSS all sky data to find out the fluctuations in brightness temperature of galactic synchrotron radiation and to quantify the variation of amplitude of fluctuation at different galactic latitudes and longitudes. The rms fluctuation of the brightness temperature δT_b for the whole sky at angular scale $l = 384$, before and after the point source subtraction is shown in the Figure 21. This will essentially help us to understand the relative contribution of synchrotron radiation at different parts of the sky. They found that C_ℓ decreases significantly as we move away from the galactic plane and C_ℓ is nearly symmetric in the northern and southern hemispheres except in the latitude range of 15–30°, which is the transition region from the disc dominated to the diffuse halo-dominated region.

Pal *et al.* (2021) have applied the MAPS-based TGE to the FIELD I data at 150 MHz $(z \sim 8.2)$ of Ghosh *et al.* (2012). They first estimated the MAPS $C_\ell (\Delta \nu)$ and from that they estimated the 21-cm power spectrum in $P(k_\perp, k_\parallel)$ domain. They obtained the 2σ upper limit on the dimensionless 21-cm power spectrum at $k =$

1.59 Mpc^{-1}, $\Delta^2(k) = (72.66)^2$ K^2. The spherically averaged 3D power spectrum is shown in Figure 22. The estimated power spectrum is consistent with the expected noise, which shows that the analysis pipeline correctly estimates and subtracts the positive noise bias and giving a robust un-biased estimate of cosmological HI 21 cm power spectrum. They also showed that using
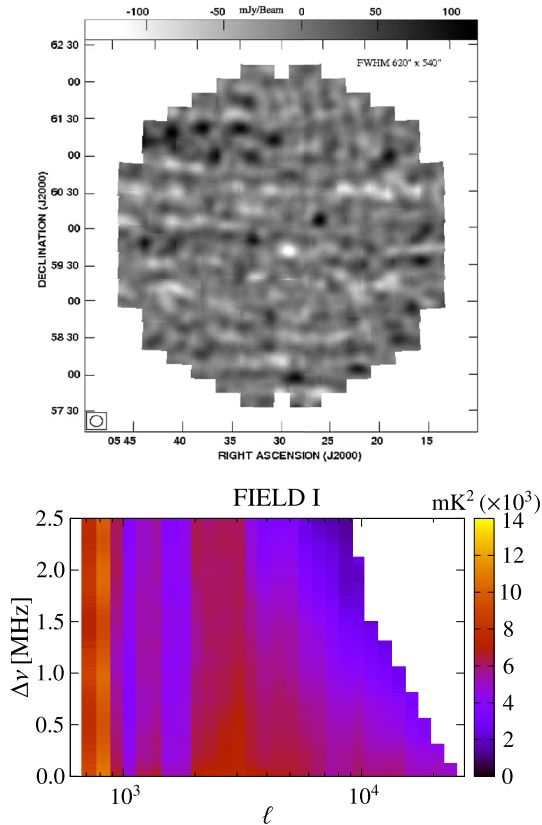
higher tapering of FoV, foreground contribution can be suppressed at large angular separation from the phase center.

In addition to these observations corresponding to EoR redshift, quite a significant amount of work has been done using GMRT at frequencies corresponding to the post-EoR epoch. Ghosh *et al.* (2011) used GMRT observation of the Hubble deep field north at 618 MHz to measure the fluctuations in the faint HI 21-cm background. They, for the first time, put upper limit on $[\bar{x}_{HI}b_{HI}] \leq 2.9$ ($[\Omega_{HI}b_{HI}] \leq 0.11$) at $z \sim 1.32$, where $\bar{x}_{HI}$ is the mean neutral fraction. Chakraborty *et al.* (2019a, b) observed the ELAIS-N1 field using upgraded wideband correlator of GMRT (uGMRT) at 300–500 MHz. They have characterized the spatial and spectral variation of DGSE and point-sources using this sensitive observation. Chakraborty *et al.* (2019b) showed that how direction-dependent and -independent calibration affects the estimation of the angular power spectrum of DGSE in presence of bright point sources far away from the phase center. After applying different calibration techniques to the data, they showed that using large tapering of the field-of-view (FoV), TGE can still reconstruct the DGSE power spectrum even in the presence of direction-dependent calibration errors. This study proves the robustness and effectiveness of TGE in presence of systematic calibration. For the first time, Chakraborty *et al.* (2019a) have estimated the spectral characteristics of the angular power spectrum of DGSE over the wide frequency bandwidth of 300–500 MHz. The estimated spectral index of DGSE is consistent with previous all-sky monopole measurements, however, they also showed that there may be a possible break in the DGSE spectrum at ~400 MHz. But more sensitive observations using much wider bandwidth data is required to infer conclusively. Finally, Chakraborty *et al.* (2021) divided the data in 8 MHz chunk and chose four such relatively less RFI-dominated sub-bands to estimate the HI power spectrum. They found that the
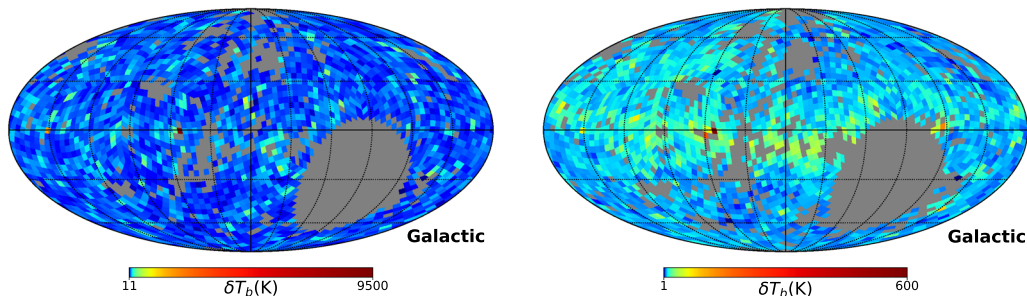

Figure 20. Top panel: The image of FIELD I ($\alpha_{2000} = 5^h30^m00^s$, $\delta_{2000} = 60°00'00''$) after subtracting the point sources and using a uv-taper at $|\vec{U}| = 200$. The synthesized beam width is 620×540 arcsec^2. The diffuse structure is visible on >10 arcmin scale. Bottom panel: The measured C_ℓ ($\Delta \nu$) of FIELD I.

Figure 21. The rms fluctuations of the brightness temperature (δT_b) all over the sky at angular scale $l = 384$ (0.47 deg). The left and right panels show the values of δT_b before and after subtracting the point sources from the data, respectively.
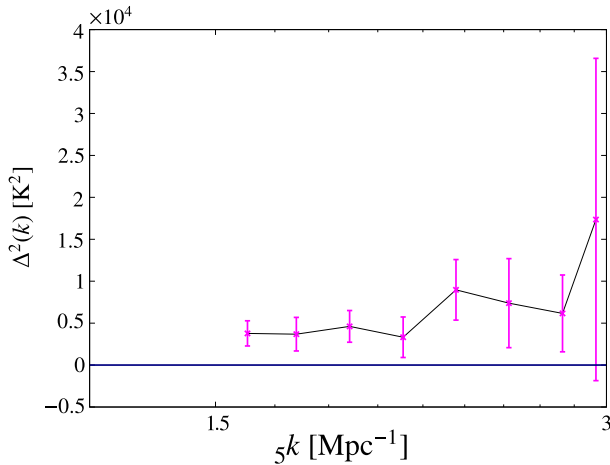
Figure 22. Mean square brightness temperature fluctuations $\Delta^2(k)$ shown as a function of k along with 2σ error bars.

spherically averaged power spectrum is close to the theoretical thermal noise power, which shows that the estimated HI power is thermal noise limited. The lowest limits, at $k \sim 1.0$ Mpc^{-1}, on spherically averaged HI 21-cm power spectrum are $(58.87$ mK$)^2$, $(61.49$ mK$)^2$, $(60.89$ mK$)^2$ and $(105.85$ mK$)^2$ at $z = 1.96, 2.19, 2.62$ and 3.58, respectively. Chakraborty *et al.* (2021) also constrained the product of neutral HI mass density (Ω_{HI}) and HI bias (b_{HI}) to the underlying dark matter density field using the best limit on the HI power spectrum. The upper limits on [$\Omega_{HI}b_{HI}$] are $0.09, 0.11, 0.12, 0.24$ at $z = 1.96, 2.19, 2.62, 3.58$, respectively.

4.4 *Calibration error and residual gain effects*

In an interferometric observation, the observed voltages hides the sky signal over the instrumental and line of sight effects. Various techniques with known sky models are used (Pearson & Readhead 1984; Wieringa 1992; van der Tol *et al.* 2007; Wijnholds & van der Veen 2009) to estimate the modification in the signal and hence, it is calibrated. However, estimation of the instrumental and ionospheric effects, and the gains, is subjected to the accuracy of the sky model and the sensitivity of the telescope. The residual gains are usually small and can be ignored. For the detection of the cosmological 21-cm signal, however, the residual gains can overwhelm the signal due to the presence of much larger foreground. Various calibration effects for LOFAR-EoR, HERA and other experiments are discussed in Liu *et al.* (2010), Vedantham & Koopmans (2015), Ali *et al.* (2016), Mevius *et al.* (2016), Patil *et al.* (2016), Vedantham & Koopmans (2016), Gehlot *et al.* (2018), Dillon *et al.* (2020), Choudhuri *et al.* (2021). The effect of
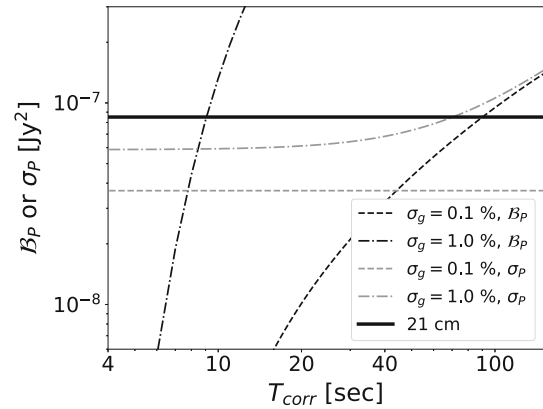

Figure 23. Variation of $\mathcal{B}_P$ (thin black) and σ_P (thin grey) against different correlation times with $\sigma_g = 0.1$ and 1.0%. This considers a bandwidth of 16 MHz, and an observation time of 400 days. The plots are made at a baseline of 300 λ. The thick black line corresponds to the expected EoR power spectrum (Bharadwaj & Ali 2005).

inaccurate models for sky based calibrations and its limitations are discussed in Barry *et al.* (2016), Ewall-Wice *et al.* (2017) and Byrne *et al.* (2019). Other instrumental effects like antenna position errors and variations in telescope beam with time and frequency on calibration solutions are discussed in Joseph *et al.* (2018), Orosz *et al.* (2019) and Choudhuri *et al.* (2021). Extragalactic point source contamination due to position errors in the sky model for bright sources, as well as the frequency-independent residual gain errors in interferometric calibration are studied in Datta *et al.* (2009a, 2010).

A detailed investigation on the effect of time-correlated residual gain errors in presence of strong foreground is presented in Kumar et al. (2020a; 2020b, 2022). Using simulated observations for GMRT baseline configuration, Kumar *et al.* (2020a) have shown that even for a perfect foreground removal, the presence of time-correlated residual gain errors introduces a bias and enhances the uncertainty in the 21-cm power spectrum estimates. Further, with an analytical investigation of the problem, in presence of strong foregrounds and thermal noise, Kumar *et al.* (2022) explored the importance to address this problem, while designing the visibility-based power spectrum estimators.

Figure 23 show the bias $\mathcal{B}_P$ and uncertainty σ_P in the power spectrum at a fixed baseline of 0.3 $k\lambda$ for 400 days observation at 150 MHz for various levels of residual gains. The solid black line gives the expected 21-cm signal (Bharadwaj & Ali 2005). We see that for small values of gain errors σ_g, σ_P is rather small and does not vary much with T_{corr}. For $\sigma_g = 1\%$, we see that both $\mathcal{B}_P$ and σ_P increase with correlation time and
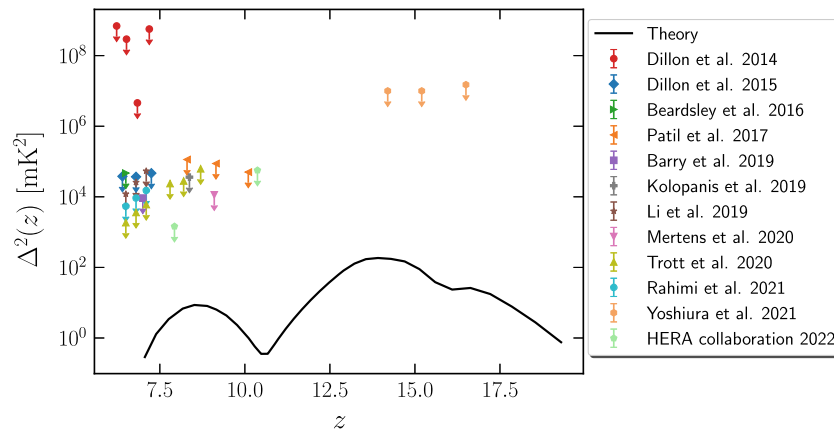
Figure 24. Summery plot of the upper limits (in points) at $k \sim 0.1$ Mpc^{-1}, available to date, measured with the current instruments. The solid black line is a theoretical power spectrum estimated from a typical GRIZZLY simulation.

$\mathcal{B}_P$ dominates σ_P. This figure demonstrate that in case, even if the error in the power spectrum is significantly low, the bias $\mathcal{B}_P$ can still be higher. Ignorance of the bias originated from the residual gains, then can confuse the scientific inference with a biased detection of the signal.

In a nutshell, given the advancements in design of estimators for the signal statistics, it is important to realize that the calibration issues are expected to provide further challenge in the detection of the 21-cm signal. Various hybrid calibration approaches, to minimize the effect of calibration errors, are also being developed and applied in different EoR experiments. Byrne *et al.* (2021) have presented a hybrid calibration framework that unifies both sky-based and redundant calibrations and they show an improvement in calibration performance through simulations. A hybrid correlation calibration (CorrCal) scheme, to address the issue of sky-model errors and imperfect array redundancy, is also presented and applied on PAPER experiments data (Gogo *et al.* 2022). A similar calibration approach is also discussed and applied on the MWA phase II data in Zhang *et al.* (2020). A precision bandpass calibration method namely CALAMITY is presented in Ewall-Wice *et al.* (2021). On the other hand, it seems also viable to design the statistical estimators of the 21-cm signal, such that the calibration bias can be directly avoided. We are working on these aspects.

5. Current upper limits and future prospects with SKA-low

There are several sensitive upper limits (Figure 24) placed on CD–EoR 21-cm power spectrum by current radio interferometers, such as MWA, LOFAR, HERA, using advanced data analysis algorithms and dedicated observation campaigns. The knowledge acquired with these observations will be used in future SKA-low to detect the 21-cm power spectrum. Here, we briefly discuss the current status of these telescopes.

LOFAR: Low Frequency Array (LOFAR)[9] is a phased aperture array with tiles or stations spread over nine countries in Europe, centered in the Netherlands and operates in the range of 30–240 MHz (van Haarlem *et al.* 2013). Patil *et al.* (2017) placed the first upper limit on EoR using 13 h of data on north celestial pole (NCP) with LOFAR high band antenna (HBA) at redshift range $z \sim 7.9$–10.6. The 2σ upper limit is (79.6 mK)^2 at $k = 0.053$ h cMpc^{-1}. Gehlot *et al.* (2019) put the first upper limit on the power spectrum of HI 21-cm brightness fluctuation from CD using 14 h of LOFAR low band antenna (LBA). They used data on two different fields centered around NCP and 3C220.3 radio galaxy. The 2σ upper limits on HI power spectrum at redshift range $z \sim 19.8$–25.2 are (14561 mK)^2 and (14886 mK)^2 at $k = 0.038$ h cMpc^{-1} for the 3C220 and NCP field, respectively. A more stringent and deep upper limit at $z \sim 9.1$ was placed by Mertens *et al.* (2020) using 141 h of data obtained with LOFAR HBA centered on NCP field. The best 2σ upper limit at $k = 0.075$ h cMpc^{-1} is (73 mK)^2. However, all these results have also shown excess power above theoretical thermal noise due to residual foreground emission, polarization leakage, chromatic calibration errors, ionosphere, or low-level radio-frequency interference, etc., (Patil *et al.* 2017; Gehlot *et al.* 2019; Mertens *et al.* 2020; Gan *et al.* 2022; Mevius *et al.* 2022).

---

[9]https://www.astron.nl/telescopes/lofar/.

MWA: Murchison Widefield Array (MWA)[10] is also an aperture array, operational between 80 and 300 MHz, consisted of 128 square 'tiles' of 4 m × 4 m, distributed over ∼3 km, which later got upgraded to 256 tiles over ∼5 km (Tingay *et al.* 2013; Wayth *et al.* 2018). It is a precursor to the SKA-low located in the Murchison radio observatory in western Australia. Dillon *et al.* (2015) set a upper limit of 3.7×10^4 mK^2 at $k = 0.18$ h Mpc^{-1} around $z \sim 6.8$ using 3 h of data with 128-tile MWA. Ewall-Wice *et al.* (2016b) put upper limits on 21-cm brightness temperature fluctuation between the redshift range $z \sim 11.6$–17.9 using 3 h of observation with MWA-128 tiles. They achieved an upper limit of 10^4 mK on co-moving scales $k < 0.5$ h Mpc^{-1}. However, they found that their result is limited by calibration systematic (Ewall-Wice *et al.* 2016b). Beardsley *et al.* (2016) have improved their analysis pipeline and put more stringent upper limit using 32 h of data at $z \sim 7.1$. They found that their estimated power spectrum is systematic limited and put a limit of 2.7×10^4 mK^2 at $k = 0.27$ h Mpc^{-1}. Barry *et al.* (2019) used the same data set as Beardsley *et al.* (2016) and presented much improved results, lowering the systematic effects by a factor of 2.8 in power. Using newly developed analysis technique and RFI removal algorithm, they put a upper limit of 3.9×10^3 mK^2 at $k = 0.2$ h Mpc^{-1} around $z \sim 7$ using 21 h of MWA-128 data set. This improved the upper limit by almost an order of magnitude from their previous analysis. Li *et al.* (2019) improved their calibration algorithm and RFI removal strategy and put the best upper limit of 2.39×10^3 mK^2 at $k = 0.59$ h Mpc^{-1} around $z \sim 6.5$ using 40 h of data set. Trott *et al.* (2020) presented multi-redshift upper limits using four seasons of data obtained with MWA. Their best measurement yields 1.8×10^3 mK^2 at $k = 0.14$ h Mpc^{-1} around $z \sim 6.5$ using 110 h of data set. Yoshiura *et al.* (2021) put an upper limit at higher redshifts, $z \sim 13$–17, using 15 h of MWA data set. The best upper limit they got about 6.6×10^6 mK^2 at $k = 0.14$ h Mpc^{-1} around $z \sim 15.2$.

HERA: Hydrogen Epoch of Reionization Array (HERA)[11] is an interferometric array of zenith pointing fixed dishes of diameter 14 m located in Karoo desert, South Africa. The dishes are packed in a hexagonal configuration with nearly continuous core of 300 m across. HERA will operate in the range of 50–250 MHz and will be built in a series of phases with simultaneous construction and observations. The full HERA will consist of 350 dishes spread out to ∼850 m, once fully operational. Recently, The HERA Collaboration *et al.* (2021) reported results with 18 nights of data (∼36 h) obtained with phase-I of HERA, which commenced with 50 antennas. They have put upper limits on the 21-cm power spectrum of $(30.76$ mK$)^2$ at $k = 0.192$ h Mpc^{-1} at $z \sim 7.9$ and also $(95.74$ mK$)^2$ at $k = 0.256$ h Mpc^{-1} at $z \sim 10.4$ (The HERA Collaboration *et al.* 2021).

In terms of sensitivity and resolution, SKA-low is going to surpass its precursor telescopes, which are mentioned earlier. It is planned to have ∼130,000 log-periodic dipole antennas, distributed over 512 aperture array stations, each having 256 dipole antennas. Construction of the dipoles have already been started in the Boolardy site in western Australia. Roughly, half of the stations will be located in the dense core within 1 km diameter and the rest would be distributed on the three spiral arms around the central core, allowing the maximum baseline to extend up to 65 km. Each station will be 35 m in diameter and the dipoles will operate within frequency bandwidth of 50–350 MHz, which cover the range of 21-cm signal from CD–EoR. Numerical predictions (Figure 25) indicate that ∼100 h of SKA-low observations will be sufficient to measure the EoR 21-cm power spectrum at $k \lesssim 0.3$ Mpc^{-1} with $>5\sigma$. The SNR will increase rapidly for a longer observation (∼1000 h) thereby allowing $>5\sigma$ detection within a broader k range. The results in Figure 25 considers the system noise contribution, assuming that foregrounds have been completely removed. The foreground avoidance decreases the overall SNR (see e.g., Shaw *et al.* 2019). A qualitatively similar prospect can be seen for the EoR 21-cm MAPS (Mondal *et al.* 2020a). We refer the readers to these articles for details of the prospects with SKA-low. Based on an older configuration of SKA-low, Koopmans *et al.* (2015) have shown that ∼ 1000 h of observations can detect the 21-cm power spectrum during mid and later stages of CD ($z \lesssim 15$) with SNR as high as ∼70 at $k \sim 0.1$ Mpc^{-1}. The SNR drops drastically towards the beginning of CD as the system noise increases as well as the signal decreases (see bottom panel of Figure 4) towards the larger redshifts.

Although bispectrum carries more information of the CD–EoR 21-cm signal as compared to the power spectrum, the measurement of bispectrum comparatively suffers more from the statistical errors (Mondal *et al.* 2021; Watkinson *et al.* 2021). This is evident from Figure 26, which shows the SNR predictions of EoR 21-cm bispectrum for all unique triangles of three different sizes (k_1), considering 1024 h of SKA-low observations.

---

[10]https://www.mwatelescope.org/.
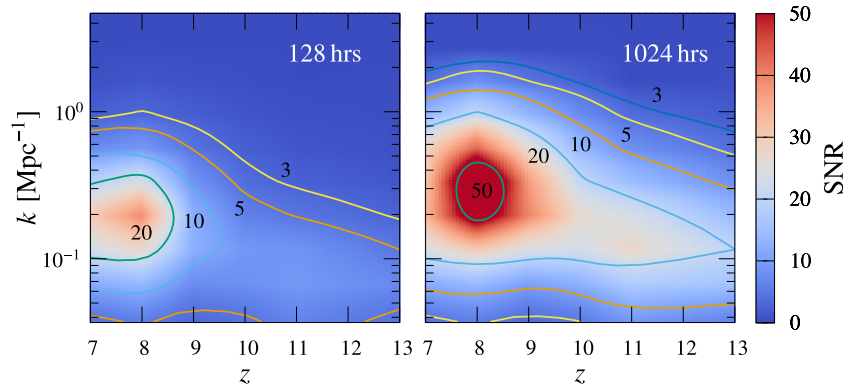
[11]https://reionization.org/.

Figure 25. SNR predictions for measuring 3D EoR 21-cm power spectrum $\Delta^2(k, z)$ using 128 h (left panel) and 1024 h (right panel) of SKA-low observations. This predictions include the system noise contributions, however, ignores the foreground effects. Credit: Shaw *et al.* (2019).
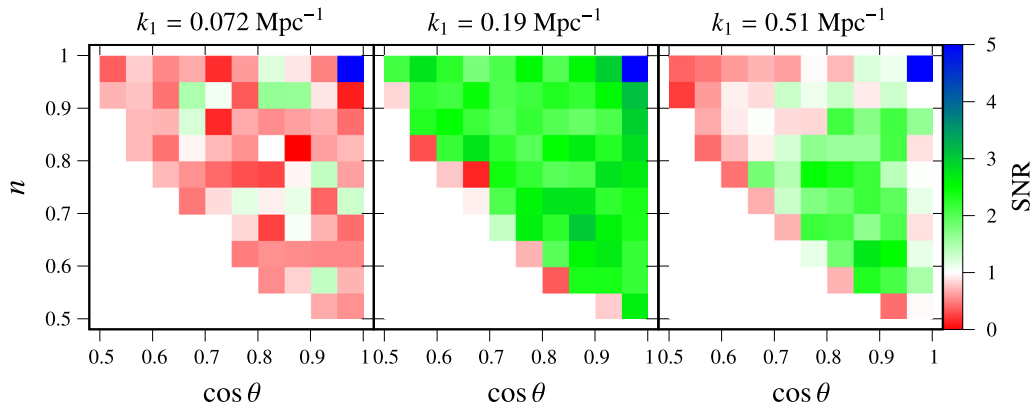

Figure 26. SNR for detecting the reionization spherically-averaged bispectrum for all unique triangle configurations for three different triangle sizes. The predictions incorporate the system noise contribution of $t_{obs} = 1024$ h of SKA-low observations. Credit: Mondal *et al.* (2021).

A significant ($>5\sigma$) detection will is possible only for the squeezed limit ($n \rightarrow 1, \cos\theta \rightarrow 1$) bispectrum, where the signal value is itself larger. A less significant ($>3\sigma$) is expected to be possible for all unique triangles of intermediate sizes ($k_1 \sim 0.1\,\mathrm{Mpc}^{-1}$). However, the SNR drops for both the smaller ($k_1 \lesssim 0.1\,\mathrm{Mpc}^{-1}$) and ($k_1 \gtrsim 1\,\mathrm{Mpc}^{-1}$), respectively, due to prevailing cosmic variance and the system noise. Hence, either a high-resolution or a longer observation or a combination of both will be needed for a reliable measurement of the CD–EoR 21-cm bispectrum from the observations.

6. Summary and future scope

The most successful theory of structure formation states that the Universe underwent changes in its thermal and ionized phases during cosmic evolution. The adiabatic expansion of the Universe made it cold and neutral from a hot dense state at $z \approx 1100$ for the first time. Afterwards, the matter perturbations grew under self-gravity and eventually collapsed to form the first structures in the Universe during the CD. The X-rays from the first stars, quasars and galaxies propagated out and heated the IGM during CD. This is followed immediately by the EoR during which the UV photons from the first sources traveled into the IGM and ionized the cold atomic hydrogen (HI). CD and EoR together marks the last transition in the thermal and ionization states of the Universe. The CD–EoR is also enriched with information of the first sources, which evolved into the structures, which we see today. Studying these epochs will help us to bridge the gap in our understanding of structure formation history. In this review, we discuss the

importance of these epochs and understand the physical processes and the sources which drive the phase change. We also discuss the observational probes of CD–EoR, challenges and remedies.

Our current understanding of CD–EoR is only limited by few indirect observations which provide some loose constraints over the beginning and the end of these epochs. A dedicated direct observation is mandatory to answer various fundamental questions regarding these epochs, such as the exact timing of these epochs, the properties of the sources involved and their evolution, how did reionization and heating progressed and changed the topology of the IGM, etc. The redshifted 21-cm radiation emitted due to hyperfine transition in the ground state of HI, acts as the most promising direct probe to CD–EoR. We aim to observe the fluctuations in 21-cm brightness temperature and its evolution as CD–EoR signal. The first sources leave their imprint on the CD–EoR 21-cm signal on cosmological scales by heating and ionizing the IGM. Several current telescopes, such as uGMRT, LOFAR, MWA and HERA as well as the future SKA-low aim to map this cosmological signal from CD–EoR. However, the CD–EoR 21-cm signal is inherently 3–4 orders of magnitude weaker as compared to the contamination from the galactic and extra-galactic foregrounds. A state-of-art methodology and precise knowledge of foreground are required to separate the signal. Also, the system noise, ionospheric turbulence and other instrumental systematics adds up to the foreground contamination making direct mapping of the CD–EoR 21-cm signal very difficult for the current instruments we have. Therefore, the primary goal of the current CD–EoR experiments, including first phase of SKA-low, is to measure the signal using statistical estimators, such as variance, power spectrum, bispectrum, etc.

A beforehand knowledge of the CD–EoR 21-cm signal and its hindrances are crucial for interpreting the observed data. There have been several theoretical studies, which use analytical or numerical prescriptions to simulate and understand the nature of the signal and its dependence on the different source properties. The state-of-art radiative transfer simulations model the signal accurately by incorporating the detailed physics during CD–EoR, but they are computationally expensive. The approximate but faster semi-numerical simulations provide a helpful alternative for exploring vast range of CD–EoR models. These different models are based on our current understanding of CD–EoR guided by indirect observations. Several previous works have used simulations to make sensitivity predictions to measure the CD–EoR statistics in the context of present

and future telescopes. These prediction studies are helpful in strategizing future observation.

Although the properties of cosmological 21-cm signal are believed to be well understood through the state-of the art large-scale simulations, however, the astrophysical foregrounds are poorly constrained. The major challenge for ongoing experiments and upcoming SKA-low is separating the cosmological signal from foregrounds, which are 10^3–10^5 times larger. Almost all foreground separation techniques use the fact that foregrounds are smooth as a function of frequency, whereas cosmological 21-cm signal decorrelates rapidly as each frequency corresponds to a different radial slice of the Universe. However, real-world problems, such as gain calibration errors, ionospheric fluctuations, variation of beam, instrumental systematics, etc., make it difficult to separate smooth foreground components from the cosmological signal of interest. Hence, it is imperative to characterize foregrounds and understand systematics through deep low-frequency observations with SKA pathfinder telescopes, such as GMRT.

In addition, it is essential to develop a robust estimator, which will give an unbiased estimate of the power spectrum of the signal. We have developed different visibility-based estimators over the past decade, which can address most of the real-world challenges, such as (1) reducing the noise bias by correlating the visibilities in nearby baselines, (2) reducing the computation cost by gridding the visibilities, (3) reducing the effect of the compact sources outside the field-of-view by tapering the response of the antenna, (4) a novel algorithm to subtract effect of resolved compact sources in the field-of-view by implementing complex tapering function in image plane and (5) using a MAPS-based algorithm to mitigate effect of flagged frequency channels. All of these estimators are tested against simulated observations and have been used to estimate the power spectrum of compact sources and galactic synchrotron radiation at $\lesssim 600$ MHz frequencies.

Even if we are able to remove the foregrounds perfectly from the data set, the gain calibration error can still pose a hindrance to detect the cosmological signal. It is extremely challenging to calibrate the instrumental and ionospheric effects at low-frequencies, the precision of which depends on the sky model and instrument sensitivity. Since the dynamic range required to detect the cosmological signal is huge, i.e., $\sim 10^5$, any residual gain errors can overwhelm the 21-cm signal. We have shown that time-correlated gain error, at the level of a few percent, can give rise to a bias and enhances the uncertainty in the 21-cm power spectrum estimates. This signifies the fact that to have a statistically robust

detection of the cosmological HI signal, we should be able to precisely calibrate all instrumental effects.

The SKA-low is going to be a giant leap in terms of sensitivity as compared to the current radio interferometers and its precursors. With its improved sensitivity, SKA-low is expected to push the field of CD–EoR cosmology towards a new edge in future. Also, a parallel development of efficient algorithms/estimators are required to extract the CD–EoR signal statistics from the contaminated data. One can also do a cross-correlation study with other line emissions, such as Ly-α, Hα, CO and CII, etc. This can provide additional constraints over the different reionization models.

Acknowledgements

References

Alcock C., Paczynski B. 1979, Nature, 281, 358

Ali S. S., Bharadwaj S. 2014, JoAA, 35, 157

Ali S. S., Bharadwaj S., Chengalur J. N. 2008, MNRAS, 385, 2166

Ali S. S., Bharadwaj S., Choudhuri S., Ghosh A., Roy N. 2016, JoAA, 37, 35

Ali S. S., Bharadwaj S., Pandey B. 2005, MNRAS, 363, 251

Aubert D., Deparis N., Ocvirk P. 2015, MNRAS, 454, 1012

Aubert D., Deparis N., Ocvirk P., *et al.* 2018, ApJL, 856, L22

Bag S., Mondal R., Sarkar P., *et al.* 2019, MNRAS, 485, 2235

Bag S., Mondal R., Sarkar P., Bharadwaj S., Sahni V. 2018, MNRAS, 477, 1984

Bandyopadhyay B., Choudhury T. R., Seshadri T. R. 2017, MNRAS, 466, 2302

Barkana R., Loeb A. 2001, Phys. Rep., 349, 125

Barkana R., Loeb A. 2005, ApJL, 624, L65

Barkana R., Loeb A. 2006, MNRAS Lett., 372, L43

Barry N., Hazelton B., Sullivan I., Morales M. F., Pober J. C. 2016, MNRAS, 461, 3135

Barry N., Wilensky M., Trott C. M., *et al.* 2019, ApJ, 884, 1

Beardsley A. P., Hazelton B. J., Sullivan I. S., *et al.* 2016, ApJ, 833, 102

Becker R. H., Fan X., White R. L., *et al.* 2001, AJ, 122, 2850

Bernardi G., de Bruyn A. G., Brentjens M. A., *et al.* 2009, A&A, 500, 965

Bharadwaj S., Ali S. S. 2004, MNRAS, 352, 142

Bharadwaj S., Ali S. S. 2005, MNRAS, 356, 1519

Bharadwaj S., Mazumdar A., Sarkar D. 2020, MNRAS, 493, 594

Bharadwaj S., Nath B. B., Sethi S. K. 2001, JoAA, 22, 21

Bharadwaj S., Pal S., Choudhuri S., Dutta P. 2019, MNRAS, 483, 5694

Bharadwaj S., Pandey S. K. 2003, JoAA, 24, 23

Bharadwaj S., Pandey S. K. 2005, MNRAS, 358, 968

Bharadwaj S., Sethi S. K. 2001, JoAA, 22, 293

Bharadwaj S., Sethi S. K., Saini T. D. 2009, Phys. Rev. D, 79, 083538

Bharadwaj S., Srikant P. S. 2004, JoAA, 25, 67

Blake C., Ferreira P. G., Borrill J. 2004, MNRAS, 351, 923

Bull P., Camera S., Raccanelli A., *et al.* 2015, AASKA14, 24

Byrne R., Morales M. F., Hazelton B. J., Wilensky M. 2021, MNRAS, 503, 2457

Byrne R., Morales M. F., Hazelton B., *et al.* 2019, ApJ, 875, 70

Chakraborty A., Roy N., Datta A., *et al.* 2019a, MNRAS, 490, 243

Chakraborty A., Datta A., Choudhuri S., *et al.* 2019b, MNRAS, 487, 4102

Chakraborty A., Datta A., Roy N., *et al.* 2021, ApJL, 907, L7

Chang T.-C., Pen U.-L., Peterson J. B., McDonald P. 2008, PRL, 100, 091303

Chen X., Miralda-Escudé J. 2004, ApJ, 602, 1

Chen Z., Xu Y., Wang Y., Chen X. 2019, ApJ, 885, 23

Cheng C., Parsons A. R., Kolopanis M., *et al.* 2018, ApJ, 868, 26

Choudhuri S., Bharadwaj S., Ali S. S., *et al.* 2017, MNRAS, 470, L11

Choudhuri S., Bharadwaj S., Chatterjee S., *et al.* 2016, MNRAS, 463, 4093

Choudhuri S., Bharadwaj S., Ghosh A., Ali S. S. 2014, MNRAS, 445, 4351

Choudhuri S., Bull P., Garsden H. 2021, MNRAS, 506, 2066

Choudhuri S., Dutta P., Bharadwaj S. 2019, MNRAS, 483, 3910

Choudhuri S., Ghosh A., Roy N., *et al.* 2020, MNRAS, 494, 1936

Choudhury T. R., Datta K., Majumdar S., *et al.* 2016, JoAA, 37, 29

Choudhury T. R., Haehnelt M. G., Regan J. 2009, MNRAS, 394, 960

Choudhury T. R., Puchwein E., Haehnelt M. G., Bolton J. S. 2015, MNRAS, 452, 261

Condon J. J. 1989, ApJ, 338, 13

Condon J. J., Cotton W. D., Fomalont E. B., *et al.* 2012, ApJ, 758, 23

Cress C. M., Helfand D. J., Becker R. H., Gregg M. D., White R. L. 1996, ApJ, 473, 7

Datta K. K., Bharadwaj S., Choudhury T. R. 2007, MNRAS, 382, 809

Datta K. K., Majumdar S., Bharadwaj S., Choudhury T. R. 2008, MNRAS, 391, 1900

Datta A., Bhatnagar S., Carilli C. L. 2009a, ApJ, 703, 1851

Datta K. K., Bharadwaj S., Choudhury T. R. 2009b, MNRAS, 399, L132

Datta A., Bowman J. D., Carilli C. L. 2010, ApJ, 724, 526

Datta K. K., Friedrich M. M., Mellema G., Iliev I. T., Shapiro P. R. 2012a, MNRAS, 424, 762

Datta K. K., Mellema G., Mao Y., *et al.* 2012b, MNRAS, 424, 1877

Datta K. K., Jensen H., Majumdar S., *et al.* 2014, MNRAS, 442, 1491

Datta K. K., Ghara R., Majumdar S., *et al.* 2016, JoAA, 37, 27

DeBoer D. R., Parsons A. R., Aguirre J. E., *et al.* 2017, PASP, 129, 045001

Dillon J. S., Neben A. R., Hewitt J. N., *et al.* 2015, PRD, 91, 123011

Dillon J. S., Lee M., Ali Z. S., *et al.* 2020, MNRAS, 499, 5840

Douspis M., Aghanim N., Ilić S., Langer M. 2015, A&A, 580, L4

Drinkwater M. J., Jurek R. J., Blake C., *et al.* 2010, MNRAS, 401, 1429

Ewall-Wice A., Hewitt J., Mesinger A., *et al.* 2016a, MNRAS, 458, 2710

Ewall-Wice A., Dillon J. S., Hewitt J. N., *et al.* 2016b, MNRAS, 460, 4320

Ewall-Wice A., Dillon J. S., Liu A., Hewitt J. 2017, MNRAS, 470, 1849

Ewall-Wice A., Dillon J. S., Gehlot B., *et al.* 2021, arXiv e-prints, arXiv:2110.11994

Fan X., Narayanan V. K., Strauss M. A., *et al.* 2002, AJ, 123, 1247

Fan X., Strauss M. A., Becker R. H., *et al.* 2006, AJ, 132, 117

Field G. B. 1958, Proceedings of the IRE, 46, 240

Fioc M., Rocca-Volmerange B. 1997, A&A, 326, 950

Friedrich M. M., Mellema G., Alvarez M. A., Shapiro P. R., Iliev I. T. 2011, MNRAS, 413, 1353

Frigo M., Johnson S. 2005, Proc. IEEE, 93, 216

Furlanetto S. R., Oh S. P., Briggs F. H. 2006, Phys. Rep., 433, 181

Furlanetto S. R., Zaldarriaga M., Hernquist L. 2004, ApJ, 613, 1

Gan H., E Koopmans L. V., Mertens F. G., *et al.* 2022, arXiv e-prints, arXiv:2203.02345

Gehlot B. K., Koopmans L. V. E., de Bruyn A. G., *et al.* 2018, MNRAS, 478, 1484

Gehlot B. K., Mertens F. G., Koopmans L. V. E., *et al.* 2019, MNRAS, 488, 4271

Ghara R., Choudhury T. R. 2020, MNRAS, 496, 739

Ghara R., Choudhury T. R., Datta K. K. 2015a, MNRAS, 447, 1806

Ghara R., Choudhury T. R., Datta K. K. 2016, MNRAS, 460, 827

Ghara R., Choudhury T. R., Datta K. K., Choudhuri S. 2017, MNRAS, 464, 2234

Ghara R., Datta K. K., Choudhury T. R. 2015b, MNRAS, 453, 3143

Ghara R., Giri S. K., Ciardi B., Mellema G., Zaroubi S. 2021, MNRAS, 503, 4551

Ghara R., Mellema G., Giri S. K., *et al.* 2018, MNRAS, 476, 1741

Ghara R., Mellema G., Zaroubi S. 2022, JCAP, 2022, 055

Ghara R., Giri S. K., Mellema G., *et al.* 2020, MNRAS, 493, 4728

Ghosh A., Bharadwaj S., Ali S. S., Chengalur J. N. 2011, MNRAS, 418, 2584

Ghosh A., Prasad J., Bharadwaj S., Ali S. S., Chengalur J. N. 2012, MNRAS, 426, 3295

Gillet N., Mesinger A., Greig B., Liu A., Ucci G. 2019, MNRAS, 484, 282

Giri S. K., D'Aloisio A., Mellema G., *et al.* 2019, JCAP, 2019, 058

Giri S. K., Mellema G., Ghara R. 2018, MNRAS, 479, 5596

Gleser L., Nusser A., Ciardi B., Desjacques V. 2006, MNRAS, 370, 1329

Gogo T. G., Ma Y.-Z., Kittiwisit P., *et al.* 2022, MNRAS, 510, 1680

Greig B., Mesinger A., Koopmans L. V. E., *et al.* 2021, MNRAS, 501, 1

Gruzinov A., Hu W. 1998, ApJ, 508, 435

Gupta Y., Ajithkumar B., Kale H. S., *et al.* 2017, Curr. Sci., 113, 707

Hamilton A. J. S. 1998, in Astrophysics and Space Science Library, Vol. 231, The Evolving Universe, ed Hamilton D., p. 185

Harker G. J. A., Zaroubi S., Thomas R. M., *et al.* 2009, MNRAS, 393, 1449

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

Hassan S., Liu A., Kohn S., La Plante P. 2019, MNRAS, 483, 2524

Hogan C. J., Rees M. J. 1979, MNRAS, 188, 791

Hopkins A. M., Afonso J., Chan B., *et al.* 2003, AJ, 125, 465

Hu E. M., Cowie L. L., McMahon R. G., *et al.* 2002, ApJL, 568, L75

Hutter A., Watkinson C. A., Seiler J., *et al.* 2020, MNRAS, 492, 653

Iacobelli M., Haverkorn M., Orrú E., *et al.* 2013, A&A, 558, A72

Iliev I. T., Ciardi B., Alvarez M. A., *et al.* 2006, MNRAS, 371, 1057

Intema H. T., Jagannathan P., Mooley K. P., Frail D. A. 2017, A&A, 598, A78

Islam N., Ghara R., Paul B., Choudhury T. R., Nath B. B. 2019, MNRAS, 487, 2785

Jensen H., Datta K. K., Mellema G., *et al.* 2013, MNRAS, 435, 460

Jones D. H., Saunders W., Colless M., *et al.* 2004, MNRAS, 355, 747

Joseph R. C., Trott C. M., Wayth R. B. 2018, AJ, 156, 285

Kaiser N. 1987, MNRAS, 227, 1

Kakiichi K., Majumdar S., Mellema G., *et al.* 2017, MNRAS, 471, 1936

Kamran M., Ghara R., Majumdar S., *et al.* 2021a, MNRAS, 502, 3800

Kamran M., Majumdar S., Ghara R., *et al.* 2021b, arXiv e-prints, arXiv:2108.08201

Kapahtia A., Chingangbam P., Appleby S. 2019, JCAP, 2019, 053

Kapahtia A., Chingangbam P., Appleby S., Park C. 2018, JCAP, 2018, 011

Kapahtia A., Chingangbam P., Ghara R., Appleby S., Choudhury T. R. 2021, JCAP, 2021, 026

Knox L., Scoccimarro R., Dodelson S. 1998, PRL, 81, 2004

Kolopanis M., Jacobs D. C., Cheng C., *et al.* 2019, ApJ, 883, 133

Koopmans L., Pritchard J., Mellema G., *et al.* 2015, in Advancing Astrophysics with the Square Kilometre Array (AASKA14), 1

Kumar J., Dutta P., Choudhuri S., Roy N. 2022, MNRAS, 512, 186

Kumar J., Dutta P., Roy N. 2020a, MNRAS, 495, 3683

Kumar J., Dutta P., Das S., Roy N. 2020b, in 2020 URSI Regional Conference on Radio Science (URSI-RCRS), p. 1

La Porta L., Burigana C., Reich W., Reich P. 2008, A&A, 479, 641

Lanzetta K. M., Wolfe A. M., Turnshek D. A. 1995, ApJ, 440, 435

Li W., Pober J. C., Barry N., *et al.* 2019, ApJ, 887, 141

Lin Y., Oh S. P., Furlanetto S. R., Sutter P. M. 2016, MNRAS, 461, 3361

Liu A., Parsons A. R., Trott C. M. 2014a, PRD, 90, 023018

Liu A., Parsons A. R., Trott C. M. 2014b, PRD, 90, 023019

Liu A., Tegmark M., Morrison S., Lutomirski A., Zaldarriaga M. 2010, MNRAS, 408, 1029

Loeb A., Wyithe J. S. B. 2008, Phys. Rev. Lett., 100, 161301

Maity B., Choudhury T. R. 2022, MNRAS, 511, 2239

Majumdar S., Bharadwaj S., Choudhury T. R. 2012, MNRAS, 426, 3178

Majumdar S., Bharadwaj S., Choudhury T. R. 2013, MNRAS, 434, 1978

Majumdar S., Bharadwaj S., Datta K. K., Choudhury T. R. 2011, MNRAS, 413, 1409

Majumdar S., Kamran M., Pritchard J. R., *et al.* 2020, MNRAS, 499, 5090

Majumdar S., Mellema G., Datta K. K., *et al.* 2014, MNRAS, 443, 2843

Majumdar S., Pritchard J. R., Mondal R., *et al.* 2018, MNRAS, 476, 4007

Majumdar S., Jensen H., Mellema G., *et al.* 2016, MNRAS, 456, 2080

Malhotra S., Rhoads J. E. 2004, ApJL, 617, L5

Masui K. W., Switzer E. R., Banavar N., *et al.* 2013, ApJL, 763, L20

Matsubara T., Suto Y., Szapudi I. 1997, ApJ, 491, L1

Mellema G., Iliev I. T., Pen U.-L., Shapiro P. R. 2006, MNRAS, 372, 679

Mellema G., Koopmans L., Shukla H., *et al.* 2015, Advancing Astrophysics with the Square Kilometre Array (AASKA14), 10

Mertens F. G., Semelin B., Koopmans L. V. E. 2021, in SF2A-2021: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics, eds Siebert A., Baillié K., Lagadec E., *et al.* p. 211

Mertens F. G., Mevius M., Koopmans L. V. E., *et al.* 2020, MNRAS, 493, 1662

Mesinger A., Furlanetto S. 2007, ApJ, 669, 663

Mesinger A., Furlanetto S., Cen R. 2011, MNRAS, 411, 955

Mevius M., van der Tol S., Pandey V. N., *et al.* 2016, Radio Science, 51, 927

Mevius M., Mertens F., Koopmans L. V. E., *et al.* 2022, MNRAS, 509, 3693

Mitra S., Choudhury T. R., Ferrara A. 2018a, MNRAS, 473, 1416

Mitra S., Choudhury T. R., Ratra B. 2018b, MNRAS, 479, 4566

Mondal R., Bharadwaj S., Datta K. K. 2018, MNRAS, 474, 1390

Mondal R., Bharadwaj S., Iliev I. T., *et al.* 2019, MNRAS, 483, L109

Mondal R., Bharadwaj S., Majumdar S. 2016, MNRAS, 456, 1936

Mondal R., Bharadwaj S., Majumdar S. 2017, MNRAS, 464, 2992

Mondal R., Bharadwaj S., Majumdar S., Bera A., Acharyya A. 2015, MNRAS, 449, L41

Mondal R., Mellema G., Shaw A. K., Kamran M., Majumdar S. 2021, MNRAS, 508, 3848

Mondal R., Shaw A. K., Iliev I. T., *et al.* 2020a, MNRAS, 494, 4043

Mondal R., Fialkov A., Fling C., *et al.* 2020b, MNRAS, 498, 4178

Morales M. F. 2005, ApJ, 619, 678

Noterdaeme P., Petitjean P., Carithers W. C., *et al.* 2012, A&A, 547, L1

Nuttall A. H. 1981, IEEE Trans. Acoust. Speech Signal Process., 29, 84

Ocvirk P., Aubert D., Sorce J. G., *et al.* 2020, MNRAS, 496, 4087

Orosz N., Dillon J. S., Ewall-Wice A., Parsons A. R., Thyagarajan N. 2019, MNRAS, 487, 537

Ota K., Iye M., Kashikawa N., *et al.* 2007, Nuovo Cimento B Serie, 122, 1015

Ouchi M., Shimasaku K., Furusawa H., *et al.* 2010, ApJ, 723, 869

Owen F. N., Morrison G. E. 2008, AJ, 136, 1889

Paciga G., Chang T.-C., Gupta Y., *et al.* 2011, MNRAS, 413, 1174

Paciga G., Albert J. G., Bandura K., *et al.* 2013, MNRAS, 433, 639

Pal S., Bharadwaj S., Ghosh A., Choudhuri S. 2021, MNRAS, 501, 3378

Parsons A. R., Pober J. C., Aguirre J. E., *et al.* 2012, ApJ, 756, 165

Parsons A. R., Backer D. C., Foster G. S., *et al.* 2010, AJ, 139, 1468

Pathak A., Bag S., Majumdar S., *et al.* 2022, arXiv e-prints, arXiv:2202.03701

Patil A. H., Zaroubi S., Chapman E., *et al.* 2014, MNRAS, 443, 1113

Patil A. H., Yatawatta S., Zaroubi S., *et al.* 2016, MNRAS, 463, 4317

Patil A. H., Yatawatta S., Koopmans L. V. E., *et al.* 2017, ApJ, 838, 65

Pearson T. J., Readhead A. C. S. 1984, ARAA, 22, 97

Peebles P. J. E. 1993, Principles of Physical Cosmology

Pen U.-L., Chang T.-C., Hirata C. M., *et al.* 2009, MNRAS, 399, 181

Péroux C., McMahon R. G., Storrie-Lombardi L. J., Irwin M. J. 2003, MNRAS, 346, 1103

Planck Collaboration, Adam R., Aghanim N., *et al.* 2016, A&A, 596, A108

Planck Collaboration, Aghanim N., Akrami Y., *et al.* 2020, A&A, 641, A6

Pober J. C., Liu A., Dillon J. S., *et al.* 2014, ApJ, 782, 66

Pritchard J. R., Loeb A. 2012, Rep. Prog. Phys., 75, 086901

Randall K. E., Hopkins A. M., Norris R. P., *et al.* 2012, MNRAS, 421, 1644

Reich P., Reich W. 1986, A&AS, 63, 205

Reich W. 1982, AAPS, 48, 219

Robertson B. E., Ellis R. S., Furlanetto S. R., Dunlop J. S. 2015, ApJL, 802, L19

Robertson B. E., Furlanetto S. R., Schneider E., *et al.* 2013, ApJ, 768, 71

Rogers A. E. E., Bowman J. D. 2008, AJ, 136, 641

Ross H. E., Dixon K. L., Ghara R., Iliev I. T., Mellema G. 2019, MNRAS, 487, 1101

Ross H. E., Dixon K. L., Iliev I. T., Mellema G. 2017, MNRAS, 468, 3785

Ross H. E., Giri S. K., Mellema G., *et al.* 2021, MNRAS, 506, 3717

Rybicki G. B., Lightman A. P. 1979, Radiative processes in Astrophysics

Santos M., Ferramacho L., Silva M., Amblard A., Cooray A. 2010, SimFast21: Simulation of the Cosmological 21cm Signal ascl:1010.025

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

Sarkar D., Bharadwaj S., Anathpindika S. 2016, MNRAS, 460, 4310

Saxena A., Majumdar S., Kamran M., Viel M. 2020, MNRAS, 497, 2941

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

Seo H.-J., Dodelson S., Marriner J., *et al.* 2010, ApJ, 721, 164

Shapiro P. R., Mao Y., Iliev I. T., *et al.* 2013, Phys. Rev. Lett., 110, 151301

Shaw A. K., Bharadwaj S., Mondal R. 2019, MNRAS, 487, 4951

Shaw A. K., Bharadwaj S., Mondal R. 2020, MNRAS, 498, 1480

Shimabukuro H., Semelin B. 2017, MNRAS, 468, 3869

Singal J., Stawarz Ł., Lawrence A., Petrosian V. 2010, MNRAS, 409, 1172

Singh S., Subrahmanyan R., Udaya Shankar N., *et al.* 2017, ApJ, 845, L12

Storrie-Lombardi L. J., McMahon R. G., Irwin M. J. 1996, MNRAS, 283, L79

Sunyaev R. A., Zeldovich I. B. 1975, MNRAS, 171, 375

Swarup G., Ananthakrishnan S., Kapahi V. K., *et al.* 1991, Curr. Sci., 60, 95

Switzer E. R., Masui K. W., Bandura K., *et al.* 2013, MNRAS Lett., 434, L46

The HERA Collaboration, Abdurashidova Z., Aguirre J. E., *et al.* 2021, arXiv e-prints, arXiv:2108.02263

Thomas R. M., Zaroubi S. 2008, MNRAS, 384, 1080

Tingay S. J., Goeke R., Bowman J. D., *et al.* 2013, PASA, 30, e007

Trott C. M., Pindor B., Procopio P., *et al.* 2016, ApJ, 818, 139

Trott C. M., Watkinson C. A., Jordan C. H., *et al.* 2019, PASA, 36, e023

Trott C. M., Jordan C. H., Midgley S., *et al.* 2020, MNRAS, 493, 4711

van der Tol S., Jeffs B. D., van der Veen A. J. 2007, IEEE Trans. Signal Process., 55, 4497

van Haarlem M. P., Wise M. W., Gunst A. W., *et al.* 2013, A&A, 556, A2

Vedantham H. K., Koopmans L. V. 2015, MNRAS, 453, 925

Vedantham H. K., Koopmans L. V. 2016, MNRAS, 458, 3099

Villaescusa-Navarro F., Viel M., Alonso D., *et al.* 2015, JCAP, 2015, 034

Visbal E., Loeb A., Wyithe S. 2009, JCAP, 2009, 030

Watkinson C. A., Giri S. K., Ross H. E., *et al.* 2019, MNRAS, 482, 2653

Watkinson C. A., Majumdar S., Pritchard J. R., Mondal R. 2017, MNRAS, 472, 2436

Watkinson C. A., Pritchard J. R. 2014, MNRAS, 443, 3090

Watkinson C. A., Trott C. M., Hothi I. 2021, MNRAS, 501, 367

Wayth R. B., Tingay S. J., Trott C. M., *et al.* 2018, PASA, 35, e033

Wieringa M. H. 1992, Exp. Astron., 2, 203

Wijnholds S. J., van der Veen A.-J. 2009, IEEE Trans. Signal Process., 57, 3512

Wilman R. J., Röttgering H. J. A., Overzier R. A., Jarvis M. J. 2003, MNRAS, 339, 695

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

Wyithe J. S. B., Loeb A., Geil P. M. 2008, MNRAS, 383, 1195

Xu H., Wise J. H., Norman M. L., Ahn K., O'Shea B. W. 2016, ApJ, 833, 84

Yatawatta S., de Bruyn A. G., Brentjens M. A., *et al.* 2013, A&A, 550, A136

Yoshiura S., Shimabukuro H., Takahashi K., Matsubara T. 2016, MNRAS, 465, 394

Yoshiura S., Pindor B., Line J. L. B., *et al.* 2021, MNRAS, 505, 4775

Zafar T., Péroux C., Popping A., *et al.* 2013, A&A, 556, A141

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

Zaroubi S., de Bruyn A. G., Harker G., *et al.* 2012, MNRAS, 425, 2964

Zhang Z., Pober J. C., Li W., *et al.* 2020, PASA, 37, e045