

University of Groningen

## Sky-averaged 21-cm signal extraction using multiple antennas with an SVD framework

Saxena, Anchal; Meerburg, P. Daniel; Acedo, Eloy de Lera; Handley, Will; Koopmans, Léon V. E.

*Published in:*  
Monthly Notices of the Royal Astronomical Society

*DOI:*  
[10.1093/mnras/stad1047](https://doi.org/10.1093/mnras/stad1047)

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2023

[Link to publication in University of Groningen/UMCG research database](#)

### *Citation for published version (APA):*

Saxena, A., Meerburg, P. D., Acedo, E. D. L., Handley, W., & Koopmans, L. V. E. (2023). Sky-averaged 21-cm signal extraction using multiple antennas with an SVD framework: the REACH case. *Monthly Notices of the Royal Astronomical Society*, 522(1), 1022–1032. Advance online publication. <https://doi.org/10.1093/mnras/stad1047>

### **Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).






The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

### **Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

# Sky-averaged 21-cm signal extraction using multiple antennas with an SVD framework: the REACH case

Anchal Saxena <sup>1</sup>★, P. Daniel Meerburg <sup>1</sup>★, Eloy de Lera Acedo <sup>2,3</sup>, Will Handley <sup>2,3</sup>  
and Léon V. E. Koopmans <sup>4</sup>

<sup>1</sup>Van Swinderen Institute, University of Groningen, Nijenborgh 4, NL-9747 AG Groningen, The Netherlands

<sup>2</sup>Cavendish Astrophysics, University of Cambridge, Cambridge CB3 0HE, UK

<sup>3</sup>Kavli Institute for Cosmology, Madingley Road, Cambridge CB3 0HA, UK

<sup>4</sup>Kapteyn Astronomical Institute, University of Groningen, P.O. Box 800, NL-9700 AV Groningen, The Netherlands

Accepted 2023 April 3. Received 2023 March 10; in original form 2022 December 22

## ABSTRACT

In a sky-averaged 21-cm signal experiment, the uncertainty of the extracted signal depends mainly on the covariance between the foreground and 21-cm signal models. In this paper, we construct these models using the modes of variation obtained from the singular value decomposition of a set of simulated foreground and 21-cm signals. We present a strategy to reduce this overlap between the 21-cm and foreground modes by simultaneously fitting the spectra from multiple different antennas, which can be used in combination with the method of utilizing the time dependence of foregrounds while fitting multiple drift scan spectra. To demonstrate this idea, we consider two different foreground models (i) a simple foreground model, where we assume a constant spectral index over the sky, and (ii) a more realistic foreground model, with a spatial variation of the spectral index. For the simple foreground model, with just a single antenna design, we are able to extract the signal with good accuracy if we simultaneously fit the data from multiple time slices. The 21-cm signal extraction is further improved when we simultaneously fit the data from different antennas as well. This improvement becomes more pronounced while using the more realistic mock observations generated from the detailed foreground model. We find that even if we fit multiple time slices, the recovered signal is biased and inaccurate for a single antenna. However, simultaneously fitting the data from different antennas reduces the bias and the uncertainty by a factor of 2–3 on the extracted 21-cm signal.

**Key words:** methods: data analysis – dark ages, reionization, first stars.

## 1 INTRODUCTION

The cosmic dawn (CD) and Epoch of Reionization (EoR) are one of the least known eras in the history of the Universe. This was the time when the first sources of light were formed, which emitted high-energy X-ray and UV radiation, which in turn heated and reionized the intergalactic medium (see e.g. Barkana & Loeb 2001; Furlanetto, Oh & Briggs 2006; Pritchard & Loeb 2012, for reviews). Observations from the absorption spectra of high-redshift quasars (Becker et al. 2001; Fan et al. 2003; Boera et al. 2019) and Thomson scattering optical depth from the cosmic microwave background (CMB) (Kaplinghat et al. 2003; Komatsu et al. 2011) suggest that it was an extended process that took place between  $z \sim 50$ –5 (Alvarez et al. 2006; Bouwens et al. 2015; Robertson et al. 2015). However, these indirect probes do not provide a very clear understanding of the physical processes occurring during the CD-EoR, such as the exact redshift range of primordial star formation, star formation efficiency, sources responsible for ionization, and the luminosity of first X-ray sources.

The 21-cm line associated with the spin flip transition in the ground state of the hydrogen atom is the most direct and promising approach to probe these eras. Motivated by the potential of the 21-cm line

to understand the astrophysical processes of the infant universe, a large number of experiments are ongoing to detect and characterize this signal. These experiments can be classified into two different categories. The first are interferometric experiments, including those using the GMRT (Paciga et al. 2013), HERA (DeBoer et al. 2017), LOFAR (Ghara et al. 2020; Mertens et al. 2020), MWA (Barry et al. 2019; Li et al. 2019), PAPER (Kolopanis et al. 2019), and SKA (Koopmans et al. 2015; Mellema et al. 2015), which are designed to measure the spatial brightness temperature fluctuations in the 21-cm signal using various Fourier statistics. Complementary, there are experiments to detect the sky-averaged 21-cm signal. These experiments include BIGHORNS (Sokolowski et al. 2015), EDGES (Bowman et al. 2018), LEDA (Price et al. 2018), PRIZM (Philip et al. 2018), REACH (de Lera Acedo et al. 2022), SARAS (Singh et al. 2018; Singh et al. 2021). Such experiments are simpler to design than interferometric experiments and require shorter integration time, but the calibration is more challenging. The dynamic range in the foreground versus 21-cm signal is much larger in global experiments, by several orders of magnitude. This paper concerns the latter approach for the Radio Experiment for the Analysis of Cosmic Hydrogen (REACH).<sup>1</sup>

\* E-mail: [anchal.saxena.009@gmail.com](mailto:anchal.saxena.009@gmail.com) (AS); [daanmeerburg@gmail.com](mailto:daanmeerburg@gmail.com) (PDM)

<sup>1</sup><https://www.astro.phy.cam.ac.uk/research/research-projects/reach>

Several years ago, the Experiment to Detect the Global EoR Signature (EDGES) reported a detection of an absorption trough centred at 78 MHz (Bowman et al. 2018). However, the shape and depth of the identified signal are quite different from the expectations of standard cosmological and astrophysical models. This unexpected feature either calls for an exotic explanation that could be achieved by enhancing the contrast between the radio background temperature and the spin temperature of the 21-cm signal or there is still some unknown systematic effect in the data. Two physically motivated scenarios for the former explanation include the overcooling of hydrogen gas due to interactions with dark matter (Barkana 2018; Barkana et al. 2018; Berlin et al. 2018; Liu et al. 2019) and the presence of an enhanced radio background other than the CMB (Ewall-Wice et al. 2018; Feng & Holder 2018; Fialkov & Barkana 2019; Ewall-Wice, Chang & Lazio 2020; Reis, Fialkov & Barkana 2020). However, a re-analysis of the EDGES data has shown that unaccounted systematics can distort or mask the signal (Hills et al. 2018; Singh & Subrahmanyan 2019; Sims & Pober 2020; Bevins et al. 2021). There could also be inadequacies in the foreground model used to fit the observations, which could account for the observed trough (Tauscher, Rapetti & Burns 2020a). The measurement of the spectrum of the radio sky in 55–85 MHz band reported by Singh et al. (2021) shows, albeit at  $\sim 2\sigma$  confidence level, that the detection of the signal claimed by EDGES might not be entirely of astrophysical origin.

In order to make a robust detection of the global 21-cm signal, one has to overcome several observational challenges. First, the foregrounds are 4–6 orders of magnitude brighter than the 21-cm signal. Galactic synchrotron emission constitutes the largest contribution to these foregrounds, which is well described by a power law in frequency i.e.  $T_{\text{sync}} \propto \nu^{-\alpha}$  (Shaver et al. 1999). For separating the 21-cm signal from the foregrounds, the key is to exploit the spectral differences between them. For example, foregrounds can be considered spectrally smooth in contrast with the cosmological 21-cm signal. However, the gain of the observing antenna has a spatial as well as a spectral variation, which mixes the spatial distribution of the foreground power into the frequency domain and corrupts the intrinsically smooth foregrounds (Vedantham et al. 2013; Bernardi, McQuinn & Greenhill 2015; Anstey, de Lera Acedo & Handley 2021; Spinelli et al. 2021). These distortions can be sufficiently non-smooth to make them difficult to remove from the data using smooth functions such as polynomials. For this reason, a common goal of all global signal experiments is to strive for as many achromatic beams as possible, thereby minimizing any potential leakage of the foregrounds into the 21-cm signal.

To circumvent this issue, it has been proposed by Vedantham et al. (2013), Switzer & Liu (2014), Tauscher et al. (2018), Rapetti et al. (2020), Tauscher, Rapetti & Burns (2020b), Bassett et al. (2021) to perform the singular value decomposition (SVD) of the beam-weighted foreground modelling set (created by varying the characteristics of the foreground model) to find the optimal basis with which to fit the foregrounds instead of assuming any specific model. In particular, Hibbard et al. (2020) have shown that in the case of a perfectly achromatic beam, the basis set obtained from the SVD is very similar to polynomials. However, there is a noticeable difference between the two when the beam is chromatic. In the latter case, the basis obtained from SVD provides a better fit to the beam-weighted foregrounds than polynomials.

A similar approach is used to create a 21-cm modelling set and obtain a 21-cm basis set from the SVD. In this approach, observations are jointly fitted by the foreground, and 21-cm basis sets are used to extract the 21-cm signal. The confidence levels

on the extracted signal are determined by the noise level of the data and the covariance between the foreground and 21-cm signal modes. Tauscher et al. (2018) utilized the observations of the four Stokes parameters to reduce this overlap because the foreground modes appear in all polarization channels whereas the 21-cm modes only appear in stokes I. Another strategy for decreasing this overlap could be obtained by utilizing the time dependence of foregrounds by fitting multiple drift-scan spectra simultaneously, as described by Tauscher et al. (2020b). This mitigates the overlap between the modes as the foregrounds change as a function of time, but the global 21-cm signal remains the same across multiple drift-scan spectra. This has also been found in the Bayesian forward modelling approach within REACH (Anstey, Acedo & Handley 2022).

In this paper, our focus is to explore the robustness of the REACH antenna designs for the 21-cm signal extraction. The designs we currently use are not sensitive to the polarization channels, so we cannot utilize this strategy of reducing the overlap between the modes, but we do utilize the time dependence of foregrounds to reduce the overlap. REACH will eventually be observing the sky simultaneously with multiple antennas, including a hexagonal dipole and a conical log spiral antenna (Dyson 1965).

This paper is organized as follows: In Section 2, we briefly outline the mathematical formalism of the signal extraction. In Section 3, we describe the 21-cm signal and foreground simulations and modelling sets for two different foreground models. Section 4 describes the results for both the models. In Section 5, we summarize our findings.

## 2 FORMALISM

This section briefly outlines the mathematical formalism of the 21-cm signal extraction. For more details, we refer the reader to Tauscher et al. (2018), Tauscher et al. (2020b). We consider a mock observation  $\mathbf{y}$ , which is composed of the global 21-cm signal  $\mathbf{y}_{21}$ , beam-weighted foregrounds  $\mathbf{y}_{\text{FG}}$  and a Gaussian random noise  $\mathbf{n}$  with covariance  $\mathbf{C}$ .

$$\mathbf{y} = \Psi_{21} \mathbf{y}_{21} + \mathbf{y}_{\text{FG}} + \mathbf{n}, \quad (1)$$

where  $\Psi_{21}$  is a matrix that expands the 21-cm signal, which is only a function of frequency and not time into the full data vector space. For example, when we simultaneously fit the data across multiple time slices (and antennas), the beam-weighted foregrounds  $\mathbf{y}_{\text{FG}}$  will change from one time slice (and antenna) to another; however,  $\mathbf{y}_{21}$  will remain the same. In that case,  $\Psi_{21}$  expands the 21-cm signal across multiple time bins (and antennas).

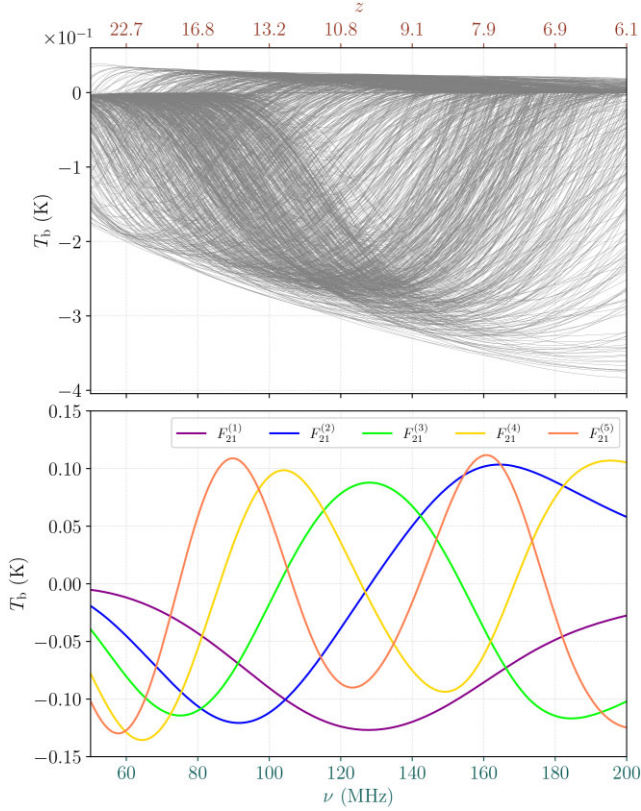
There are also errors introduced in the data due to imperfect calibration such as additive biases through the receiver's gain and antenna noise. In this work, we assume a well calibrated instrument and ignore these sources of uncertainty. The impact of these instrumental errors will be considered in a future work. These effects have been taken into account to some extent in Tauscher et al. (2021).

To extract  $\mathbf{y}_{21}$  from the data  $\mathbf{y}$ , we model the data using the basis vectors (or modes) for the 21-cm signal  $\mathbf{F}_{21}$  and beam-weighted foregrounds  $\mathbf{F}_{\text{FG}}$ , derived from the SVD of their modelling sets. These modes are further normalized such that  $\mathbf{F}_{21}^T \Psi_{21}^T \mathbf{C}^{-1} \Psi_{21} \mathbf{F}_{21} = \mathbf{I}$  and  $\mathbf{F}_{\text{FG}}^T \mathbf{C}^{-1} \mathbf{F}_{\text{FG}} = \mathbf{I}$ , where  $\mathbf{I}$  is the identity matrix. The model of our data then takes the form

$$\mathbf{M}(\mathbf{x}_{21}, \mathbf{x}_{\text{FG}}) = \Psi_{21} \mathbf{F}_{21} \mathbf{x}_{21} + \mathbf{F}_{\text{FG}} \mathbf{x}_{\text{FG}}, \quad (2)$$

where  $\mathbf{x}_{21}$  and  $\mathbf{x}_{\text{FG}}$  are the linear coefficients that scale the basis vector amplitudes contained in  $\mathbf{F}_{21}$  and  $\mathbf{F}_{\text{FG}}$ , respectively. This can be re-written as

$$\mathbf{M}(\mathbf{x}) = \mathbf{F} \mathbf{x}, \quad (3)$$



**Figure 1.** **Top panel:** A thin slice of the 21-cm signal modelling set composed of all the physical signals simulated using *ares*. **Bottom panel:** The first five modes (purple, blue, green, yellow, orange) obtained from the SVD of the modelling set.

where  $\mathbf{F} = [\Psi_{21} \mathbf{F}_{21} \parallel \mathbf{F}_{\text{FG}}]$  and  $\mathbf{x}^T = [\mathbf{x}_{21}^T \parallel \mathbf{x}_{\text{FG}}^T]$ .

To find the parameters  $\mathbf{x}$  of our model  $\mathbf{M}$ , we assume the likelihood to be Gaussian i.e.

$$\mathcal{L}(\mathbf{y}|\mathbf{x}) \propto \exp \left\{ -\frac{1}{2} [\mathbf{y} - \mathbf{F}\mathbf{x}]^T \mathbf{C}^{-1} [\mathbf{y} - \mathbf{F}\mathbf{x}] \right\}. \quad (4)$$

Assuming a flat prior on the scaling parameters<sup>2</sup>, the posterior distribution also turns out to be Gaussian in  $\mathbf{x}$  with mean  $\xi$  and covariance  $\mathbf{S}$  given as

$$\begin{aligned} \xi &= \mathbf{S}\mathbf{F}^T \mathbf{C}^{-1} \mathbf{y}, \\ \mathbf{S} &= (\mathbf{F}^T \mathbf{C}^{-1} \mathbf{F})^{-1}. \end{aligned} \quad (5)$$

The maximum likelihood reconstruction of the 21-cm signal  $\mathbf{y}_{21}$  and its covariance  $\Delta_{21}$  is given as [see equation (3)]

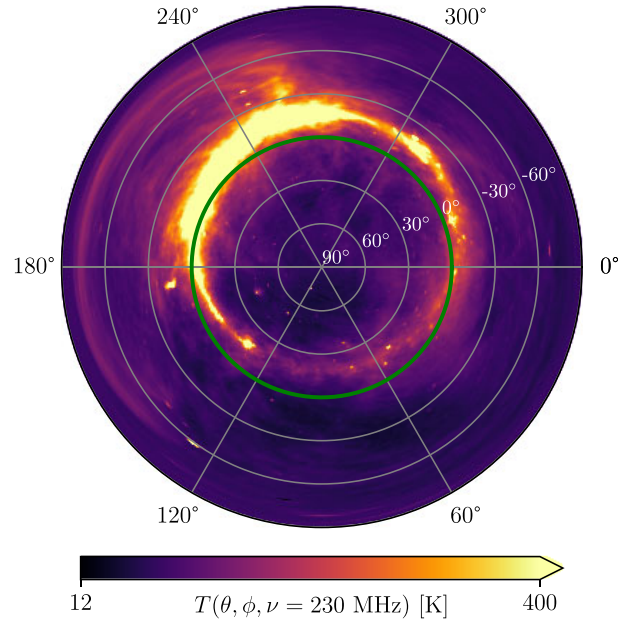
$$\begin{aligned} \mathbf{y}_{21} &= \mathbf{F}_{21} \xi_{21}, \\ \Delta_{21} &= \mathbf{F}_{21} \mathbf{S}_{21} \mathbf{F}_{21}^T, \end{aligned} \quad (6)$$

where  $\xi_{21}$  and  $\mathbf{S}_{21}$  are part of the  $\xi$  and  $\mathbf{S}$  matrices, respectively, corresponding to the 21-cm component. The 1- $\sigma$  root-mean-square (RMS) uncertainty on the reconstructed signal is then given by

$$\text{RMS}_{21}^{1\sigma} = \sqrt{\text{Tr}(\Delta_{21})/n_v}, \quad (7)$$

where  $n_v$  is the number of frequency channels.

<sup>2</sup>In principle, one can derive informative priors from the modelling sets; however, it is based on a few key assumptions which limit its applicability (see Appendix C of Tauscher et al. (2018)).



**Figure 2.** Radio sky (de Oliveira-Costa et al. 2008) at 230 MHz for the simulated observation on 2019-10-01 at 00:00:00 UTC when the galactic disc remains below the horizon ( $\theta < 0^\circ$ ) for an antenna located in the Karoo radio reserve. This is shown in altitude ( $\theta$ ) and azimuth ( $\phi$ ) coordinates of the antenna's frame with the zenith in the centre, where  $\theta = 0^\circ$  (green line) marks the horizon.

While fitting the data, we select the number of modes for each component by minimizing the deviance information criterion (DIC) (Spiegelhalter et al. 2002), which is given as

$$\text{DIC} = \delta^T \mathbf{C}^{-1} \delta + 2(n_b^{21} + n_b^{\text{FG}}), \quad (8)$$

where  $\delta = \mathbf{F}\xi - \mathbf{y}$ ,  $n_b^{21}$  and  $n_b^{\text{FG}}$  are the number of 21-cm and foreground modes used in the fit. The first term in equation (8) represents the goodness of fit and the second term quantifies the model complexity. We also compare its performance with the Bayesian Information Criterion (BIC) and find that minimizing DIC results in more unbiased fits compared to the BIC. This is consistent with Tauscher et al. (2018).

## 2.1 Statistical measures

To examine and compare the performance of different combinations of antennas, we estimate a few statistical measures. First, we estimate the signal bias statistic defined in Tauscher et al. (2018) as

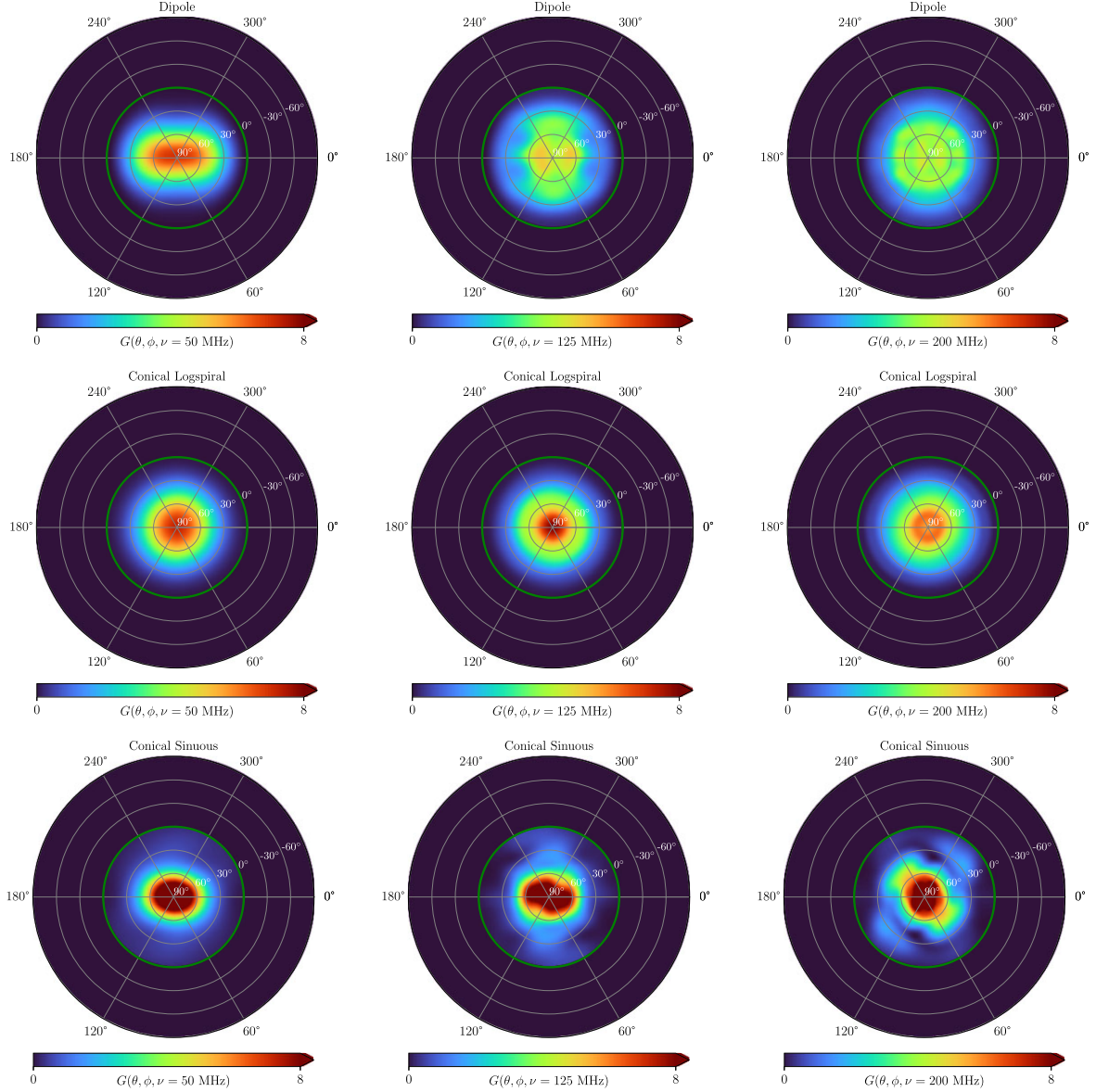
$$\varepsilon = \sqrt{\frac{1}{n_v} \sum_{i=1}^{n_v} \frac{(\mathbf{y}_{21} - \mathbf{y}_{21})^2}{(\Delta_{21})_{ii}}}, \quad (9)$$

where  $n_v$  is the number of frequency channels,  $\mathbf{y}_{21}$  is the input signal, and  $\mathbf{y}_{21}$  and  $\Delta_{21}$  are the reconstructed signal and its channel covariance respectively from equation (6). It determines the number of  $\sigma$  at which the uncertainty interval on the extracted signal includes the input 21-cm signal. Now, we can define the uncertainty interval, which includes the input signal as

$$\text{RMS}_{21} = \varepsilon \text{RMS}_{21}^{1\sigma}, \quad (10)$$

where  $\varepsilon$  and  $\text{RMS}_{21}^{1\sigma}$  are given from equations (9) and (7), respectively.





**Figure 3.** Beam patterns  $G(\theta, \phi, \nu)$  for the dipole (first row), conical log spiral (second row), and conical sinuous (third row) antenna at  $\nu = 50$  MHz (first column), 125 MHz (second column) and 200 MHz (third column). These are shown in the altitude ( $\theta$ ) and azimuth ( $\phi$ ) coordinates of the antenna’s reference frame with the zenith in the centre. The green line ( $\theta = 0^\circ$ ) marks the horizon.

The signal bias statistic and RMS uncertainty of the extracted 21-cm signal can only be estimated if the input 21-cm signal is known, which is not the case in practice. So, in order to quantify the overall quality of the fit to the full data  $\mathbf{y}$ , we calculate the normalized chi-squared statistic as

$$\chi^2 = \frac{\delta^T \mathbf{C}^{-1} \delta}{N_c - N_p}, \quad (11)$$

where  $N_c$  is the number of data channels, and  $N_p$  is the number of parameters (or modes) used in the fit.

### 3 SIMULATIONS AND MODELLING SETS

We begin by creating the modelling sets for the 21-cm signal and the beam weighted foregrounds. For the 21-cm signal, we form a modelling set from all the physical signals simulated using the

code Accelerated Reionization Era Simulations (*ares*)<sup>3</sup> (Mirocha, Harker & Burns 2015; Mirocha, Furlanetto & Sun 2017). In the top panel of Fig. 1, we show a subset of the entire 21-cm modelling set. We then perform the SVD of this modelling set to derive the modes of variation, which form an orthonormal basis. The first five modes are shown in the bottom panel of Fig. 1. These modes are numbered (from violet to orange) according to their ability to capture the variation in the modelling set.

A foreground model is derived from the following features: (i) a spatial brightness temperature distribution, (ii) a spatial spectral index distribution, and (iii) the frequency-dependent beam pattern of the observing antenna. For simulating such beam-weighted foregrounds, we consider two different models based on their complexity.

<sup>3</sup><https://ares.readthedocs.io/en/v0.8/>

### 3.1 A simple foreground model

In our simplest foreground model, we assume the spectral index to be constant across the sky. We generate our sky model as

$$T_{\text{sky}}(\theta, \phi, \nu, t) = (T_{230}(\theta, \phi, t) - T_{\text{CMB}}) \left( \frac{\nu}{230} \right)^{-\beta} + T_{\text{CMB}}, \quad (12)$$

where  $T_{230}(\theta, \phi, t)$  is the spatial brightness temperature distribution derived from the Global Sky Map at 230 MHz (de Oliveira-Costa et al. 2008), which we use as our base foreground map, and  $T_{\text{CMB}} = 2.725$  K. The choice of our foreground map follows from an independent Bayesian analysis of REACH (Anstey et al. 2022); however, a physical sky model such as GMOSS (Rao et al. 2016) can also be used which encapsulates the effects of radiative processes from different components of the radio sky. Our simulated observations start at 2019-10-01 00:00:00 UTC and are integrated for 6 h. In Fig. 2, we show the overhead sky at 230 MHz on 2019-10-01 00:00:00 UTC for an antenna located in the Karoo radio reserve in South Africa.

We convolve this sky model with the beam pattern of the antenna to form the beam-weighted foregrounds as

$$T_{\text{FG}}(\nu, t) = \frac{1}{4\pi} \int_0^{4\pi} G(\theta, \phi, \nu) T_{\text{sky}}(\theta, \phi, \nu, t) d\Omega, \quad (13)$$

where  $G(\theta, \phi, \nu)$  is the beam pattern of the antenna with the beam centred on the zenith. Within REACH, we consider here three different antenna designs: a dipole, a conical sinusoidal, and a conical log spiral antenna. In Fig. 3, we show the beam patterns  $G(\theta, \phi)$  of these antennas at three different frequencies  $\nu = 50, 125,$  and  $200$  MHz in the antenna's reference frame. The distortions caused by the chromaticity of these antennas are sufficiently non-smooth to prevent proper fitting by smooth polynomial-based foreground models, thereby masking the 21-cm signal (see fig. 5 of Anstey et al. (2021)). Our analysis is based on the observations during which the galactic plane remains below the horizon, as these distortions are more pronounced when the galaxy is above the horizon.

To form the beam-weighted foreground modelling set for each antenna, we use the spectral index  $\beta$  as the free parameter, which is varied in the range (2.45, 3.15). In the top panel of Fig. 4, we show a subset of the beam-weighted foreground modelling set for the dipole antenna. The first five modes obtained from the SVD are shown in the bottom panel.

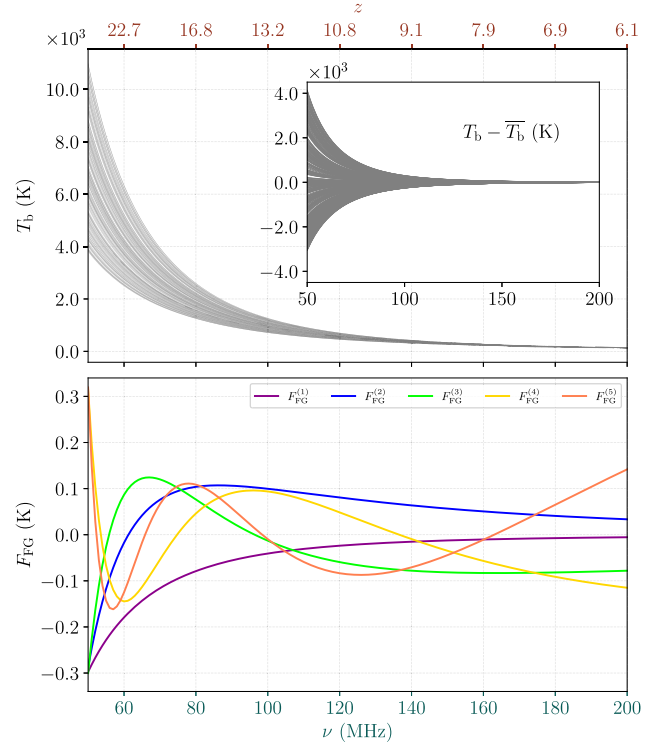
### 3.2 A detailed foreground model

So far, for simplicity, we assumed that the spectral index is constant. In general, a spatial variation of the spectral index across the sky is expected. This spectral index variation map is derived by calculating the spectral index required to map each pixel of an instance of 2008 Global Sky Map (GSM) at 408 MHz to the corresponding pixel of an instance at 230 MHz as

$$\beta(\theta, \phi) = \frac{\log \left( \frac{T_{230}(\theta, \phi) - T_{\text{CMB}}}{T_{408}(\theta, \phi) - T_{\text{CMB}}} \right)}{\log \left( \frac{230}{408} \right)}. \quad (14)$$

In the left-hand panel of Fig. 5, we show the resulting spectral index map. We use this spectral index map to render our mock observations alongside the GSM as the base foreground map with the beam patterns shown in Fig. 3.

To model the foregrounds in a parametrized way, we divide the sky into 30 regions and assign a constant spectral index to each region (Anstey et al. 2021). To form these regions, we divide the full range of spectral indices shown in the left-hand panel of Fig. 5 into 30



**Figure 4.** **Top panel:** A thin slice of the simple beam-weighted foreground modelling set for the dipole antenna. The inset plot shows the same slice of the modelling set with its mean subtracted. **Bottom panel:** The first five modes (shown by different colours) obtained from the SVD. The sign of the basis function is irrelevant since it is absorbed in the weights.

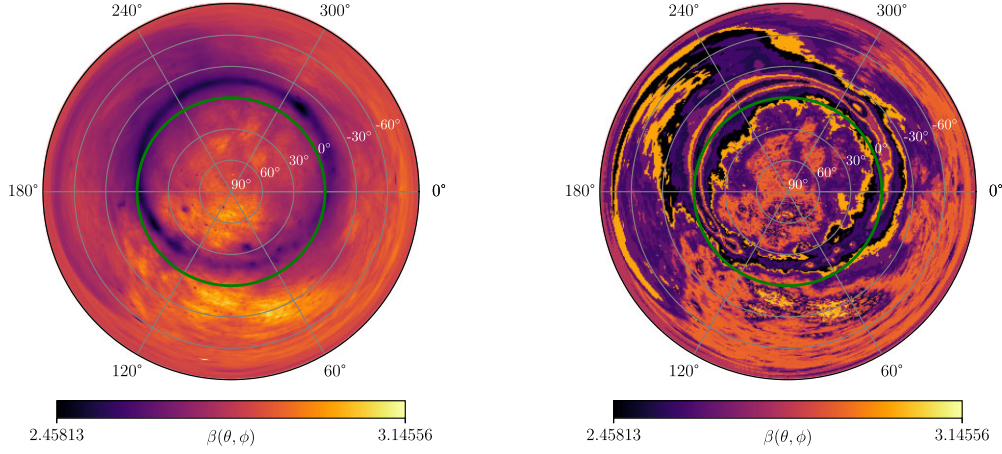
equal intervals, and each region is defined as the patch of the sky with spectral indices within these intervals. We show the division of the sky in the right-hand panel of Fig. 5.

To form the beam-weighted foregrounds, we use equations (12) and (13) with  $\beta(\theta, \phi)$  as the spectral index model. For creating a modelling set, we randomly sample the spectral index for each region  $\beta_i \in (2.45813, 3.14556)$  for  $1 \leq i \leq 30$ , which results in different spectral index maps. One of these sample maps is shown in the right-hand panel of Fig. 5. We take 50 000 samples to form the beam-weighted foreground modelling set. In the top panel of Fig. 6, we show a subset of the modelling set for the dipole antenna. The first five modes obtained from the SVD are shown in the bottom panel.

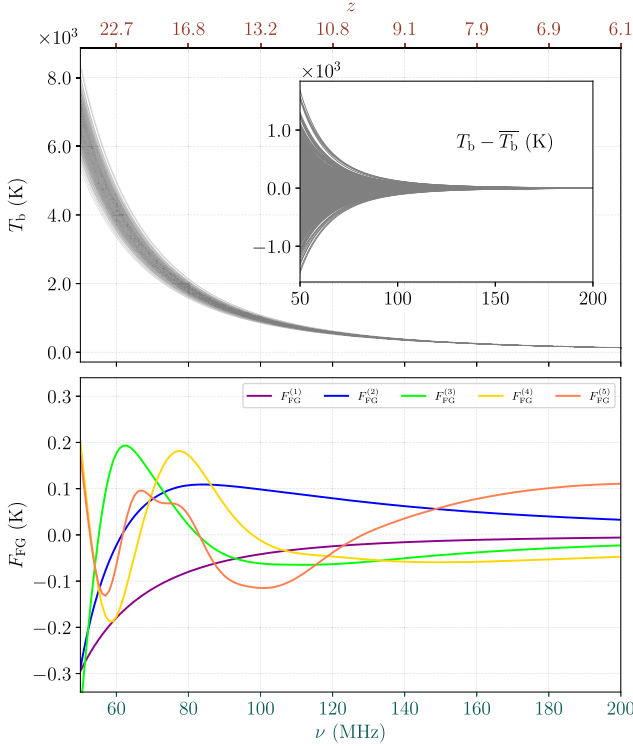
Once we obtain the modes for the 21-cm signal and the foregrounds, we fit the mock observation by selecting the number of modes that minimize the DIC. In Fig. 7, we show this minimization for a typical observation, where we estimate the DIC over a grid of 21-cm and foreground modes. In this case, we select four foreground modes and eleven signal modes, as shown with the blue squared box in the figure.

## 4 RESULTS

In this section, we discuss the application of this formalism in the 21-cm signal extraction with different antenna designs for the two foreground models that we consider.



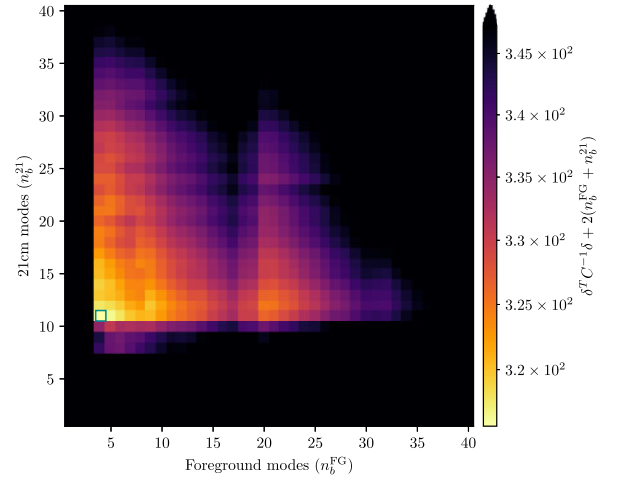
**Figure 5.** **Left-hand panel:** Spectral index distribution on 2019-10-01 at 00:00:00 UTC in the altitude ( $\theta$ ) and azimuth ( $\phi$ ) coordinates of the antenna's reference frame, derived from the pixel-wise tracing between two instances of 2008 Global Sky Map at 408 and 230 MHz, **Right-hand panel:** The division of the sky into 30 regions, where the spectral index in each region is randomly sampled from  $\beta_i \in (2.45813, 3.14556)$  for  $1 \leq i \leq 30$ . The green line ( $\theta = 0^\circ$ ) marks the horizon.



**Figure 6.** **Top panel:** A thin slice of the detailed beam-weighted foreground modelling set for the dipole antenna. The inset plot shows the same slice of the modelling set with its mean subtracted. **Bottom panel:** The first five modes (shown by different colours) obtained from the SVD.

#### 4.1 Simple foreground model

In Fig. 8, we show the 21-cm signal extraction for different antennas and the different number of time slices. In each case, the simulated mock observation is generated by taking a 21-cm signal from its modelling set and a foreground component from the simple foreground modelling set (see Section 3.1). In the first row, we show the extracted signal when we fit the data for a single time slice. The

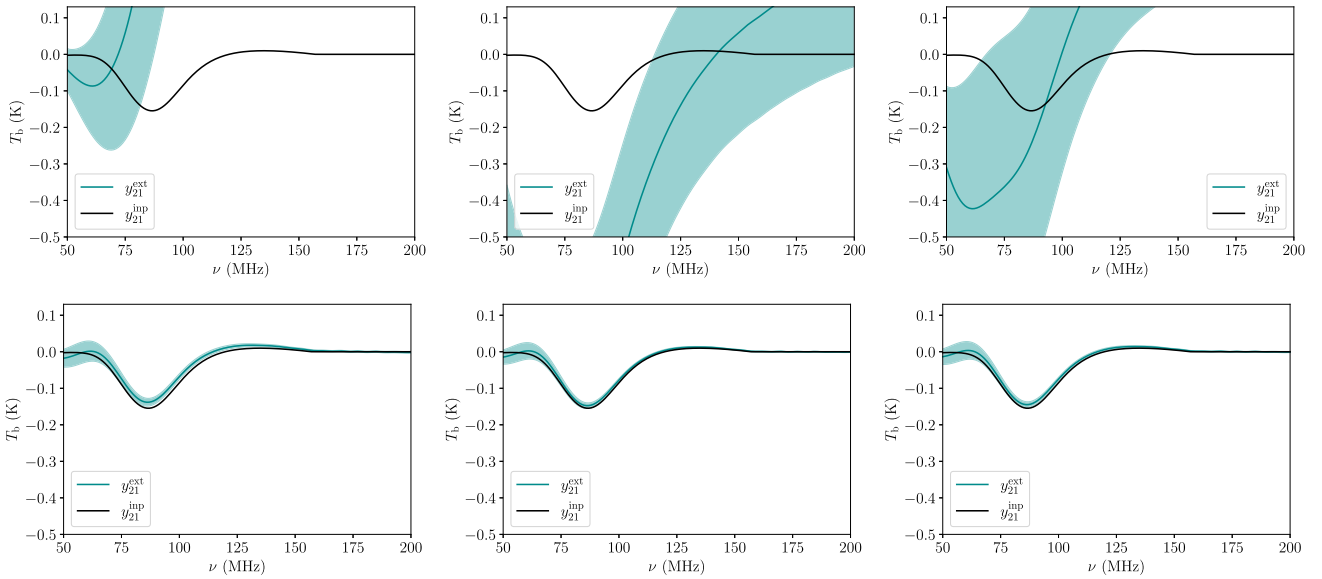


**Figure 7.** DIC minimization: The colourbar represents the DIC value estimated over a grid of the number of foreground and 21-cm modes. In this case, DIC is minimized with  $(n_b^{\text{FG}}, n_b^{21}) = (4, 11)$ .

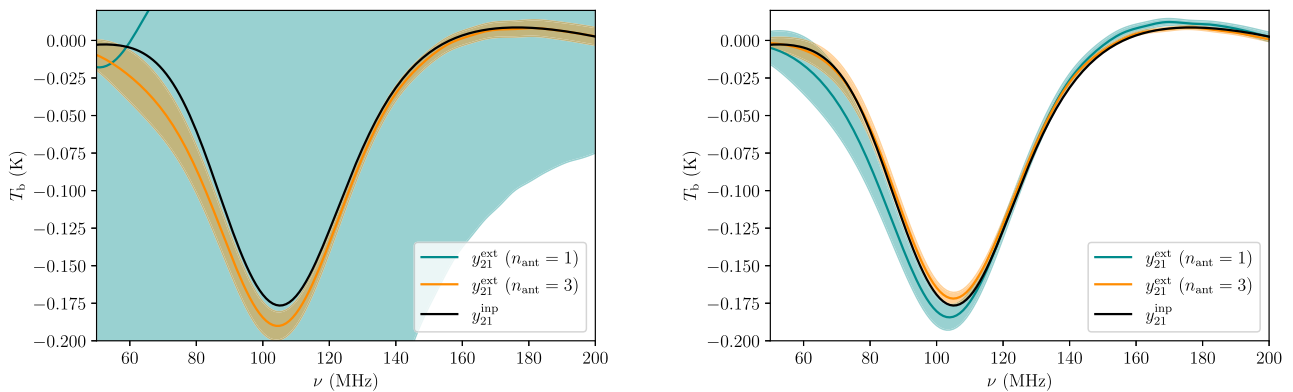
shaded regions represent the  $1\sigma$  confidence intervals. In this case, the pipeline fails to extract the 21-cm signal as the RMS uncertainty for any antenna design remains around 1000 mK. This happens due to a significant covariance between the foreground and 21-cm signal modes. However, we note that in each case, the  $\chi^2 \approx 1$ , which indicates that the full data gets a good fit by choice of basis vectors from the minimization of the DIC.

To reduce the overlap between the foreground and 21-cm signal modes, we simultaneously fit the data from multiple time slices. While doing so, we utilize the correlations between different time slices and enforce that the 21-cm signal does not change between these slices. In the second row of Fig. 8, we show the signal estimates while simultaneously fitting four time bins. We see that the signal extraction significantly improves with an RMS uncertainty of around 10 mK for any antenna design. These findings are consistent with the earlier studies by Tauscher et al. (2020b).

Next, we demonstrate how we can utilize the multiple antenna designs within REACH to improve the extraction of the 21-cm signal



**Figure 8.** Extracted 21-cm signals for the mock observation were generated from a physical 21-cm signal and a foreground model taken from the simple foreground modelling set for dipole (left), conical sinuous (middle), and conical log spiral (right) antenna while simultaneously fitting one time slice (top) and four time slices (bottom). The shaded region represents the  $1\sigma$  confidence region, and the black line shows the input 21-cm signal.



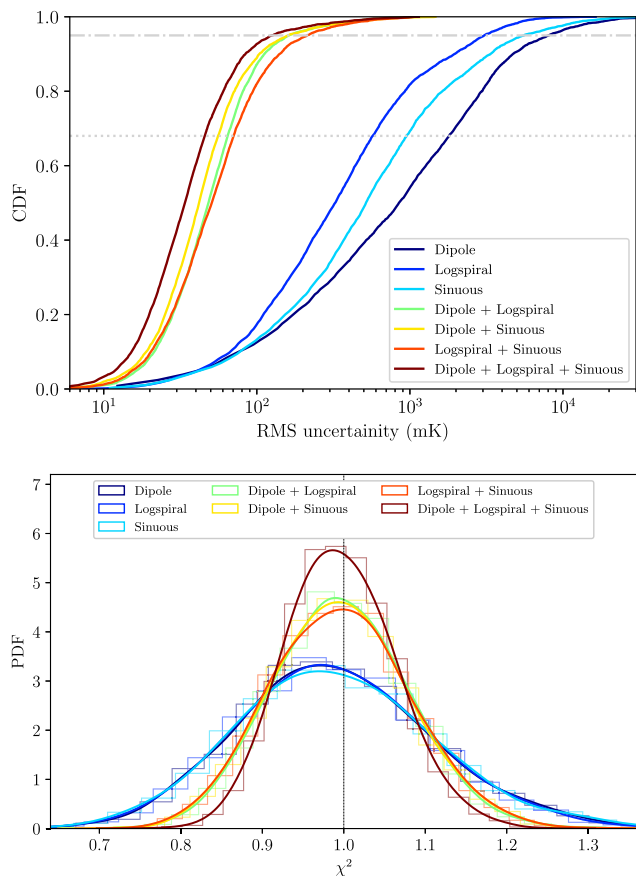
**Figure 9.** Extracted a 21-cm signal for a single antenna (cyan) and multiple antennas (orange) while fitting a single time bin (left-hand panel) and simultaneously fitting two time bins (right-hand panel). The shaded regions represent the  $1\sigma$  confidence region for each case, and the black line shows the input 21-cm signal.

further. For this, we compare the signal extraction from a single antenna with the extraction from multiple antennas in Fig. 9. Note that we assumed a total integration time of 24 h for both scenarios. In the left-hand panel of Fig. 9, we fit the data for a single time slice, where we can clearly see the advantage of simultaneously fitting the data from multiple antennas. In this case, the RMS uncertainty of the single antenna fit (cyan) is 747 mK, which reduces to 23 mK for multiple antennas (orange). This happens because simultaneously fitting multiple antennas utilizes the correlations between different antennas and enforces that the global 21-cm signal does not depend on the antenna design, thereby reducing the covariance between the foreground and 21-cm modes. In the right-hand panel of Fig. 9, we simultaneously fit the data from two time slices. In this case, the RMS uncertainty of the single antenna fit (cyan) is 32 mK, which reduces to 11 mK for three antennas (orange). Also, in this case, we find that fitting the data from multiple antennas results in a lower uncertainty on the extracted 21-cm signal with a slightly reduced bias, although this improvement is not as pronounced, as it is for a single time slice (the left-hand panel). This is because fitting multiple

time slices (the right-hand panel) even for a single antenna already removes a significant overlap between the foreground and 21-cm modes.

To better understand the extent to which multiple antennas tend to improve the fit, we quantify the statistical measures of the extracted 21-cm signal that we discussed in Section 2.1. First, we estimate the RMS uncertainty on the extracted signal for an ensemble of 5000 different mock observations, which are generated by randomly sampling a 21-cm and foreground component from their modelling sets. In the top panel of Fig. 10, we show the cumulative distribution function (CDF) of the RMS uncertainty for seven different configurations that can be formed from the three antenna designs we considered. In each case, we fit the data from a single time slice and assume the total integration time to be 6 h, which is distributed evenly in the analysis of multiple antennas. The advantage of having multiple antennas can be clearly noticed as the distribution for multiple antennas tends towards lower RMS uncertainty levels. For a single antenna, 68 per cent (and 95 per cent) of the 5000 different mock observations are extracted within 954 mK





**Figure 10.** **Top panel:** CDF of RMS uncertainty level on the extracted signal estimated from 5000 different mock observations for the simple foreground model. Different colours in the plot represent different combinations of antennas used in the analysis. The uncertainty levels at 68 and 95 per cent for each combination are shown in Table 1. **Bottom panel:** Probability distribution function (PDF) of the normalized  $\chi^2$  statistic estimated from 5000 different mock observations for different combinations of antennas.

**Table 1.** This tabulates the RMS uncertainty level (in mK) on the extracted 21-cm signal for 68 and 95 per cent of the mock observations, and the variance of the PDF of  $\chi^2$  for different antenna configurations shown in Fig. 10 for the simple foreground model.

Configuration	RMS  <sub>68 per cent</sub>	RMS  <sub>95 per cent</sub>	$\sigma(\chi^2)$
Dipole	1807.64	8101.69	0.119
Logspiral	572.70	3054.16	0.118
Sinuous	954.46	5568.83	0.122
Dipole + Logspiral	64.61	161.84	0.083
Dipole + Sinuous	55.96	159.41	0.083
Logspiral + Sinuous	70.52	212.96	0.085
Dipole + Logspiral + Sinuous	45.83	125.49	0.067

$< \text{RMS} < 1808 \text{ mK}$  (and  $5568 \text{ mK} < \text{RMS} < 8102 \text{ mK}$ ). For a pair of antennas, 68 per cent (and 95 per cent) of the mock observations are extracted within  $55 \text{ mK} < \text{RMS} < 71 \text{ mK}$  (and  $159 \text{ mK} < \text{RMS} < 213 \text{ mK}$ ). For three antennas, 68 per cent (and 95 per cent) of the mock observations are extracted with  $\text{RMS} < 46 \text{ mK}$  (and  $\text{RMS} < 126 \text{ mK}$ ). These values are tabulated in Table 1 for different antenna configurations.

In the bottom panel of Fig. 10, we show the probability distribution function (PDF) (from 5000 different mock observations) of the normalized  $\chi^2$ , which quantifies how well the set of foreground and 21-cm modes fit the data for all seven antenna configurations. As we include more antennas in the fit, the quality of the fit to the data improves as the variance of the  $\chi^2$  distribution decreases, and it becomes more centred around  $\chi^2 = 1$ . For a single antenna, the variance of the distribution  $\sigma(\chi^2) \sim 0.12$ . For a pair of antennas,  $\sigma(\chi^2) \sim 0.083$ , which reduces to  $\sigma(\chi^2) \sim 0.067$  for three antennas. These values for different configurations are tabulated in Table 1.

#### 4.2 Detailed foreground model

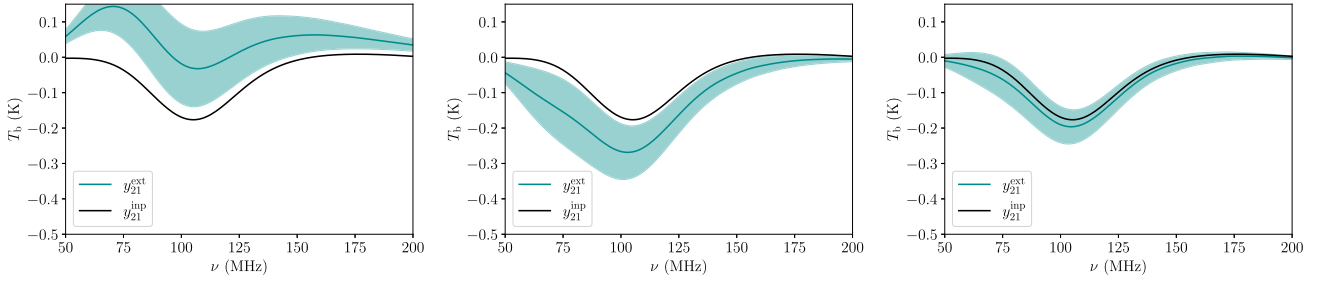
Next, we discuss the results for a more realistic foreground model (see Section 3.2). In Fig. 11, we show the extracted 21-cm signal for three different antenna configurations: (i) dipole (left), (ii) dipole + conical log spiral (middle), and (iii) dipole + conical sinuous + conical log spiral (right). In each case, we simultaneously fit the data from four time bins, and the mock observation is generated by taking a 21-cm signal and a foreground component from their modelling sets.

As expected, when considering a more realistic foreground model, the signal extraction is worse than it was for the simple foreground model, particularly for single antenna analysis. For example, if we compare the extracted signal from a dipole for the detailed foreground model (left-hand panel of Fig. 11) with the extracted signal for the simple foreground model (bottom left-hand panel of Fig. 8), the RMS uncertainty on the extracted signal is larger (123 mK in comparison to 21 mK). In addition, the extracted signal is also significantly biased. This happens because of a larger covariance between the foreground and the 21-cm signal modes, where the realistic foregrounds introduce an additional frequency-dependent structure in the higher-order foreground modes (see the bottom panel of Fig. 6). Evidently, it is suboptimal to use just a single antenna for 21-cm signal extraction with such foreground models.

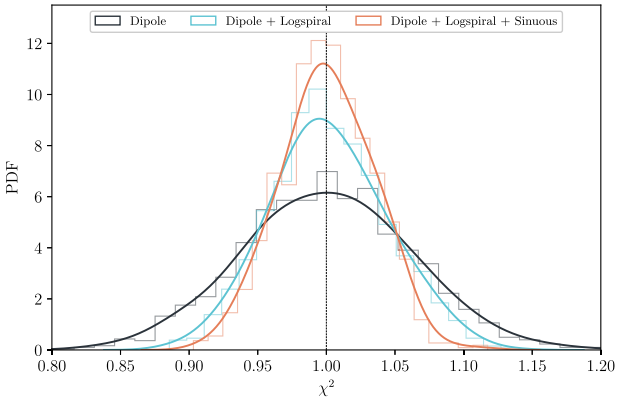
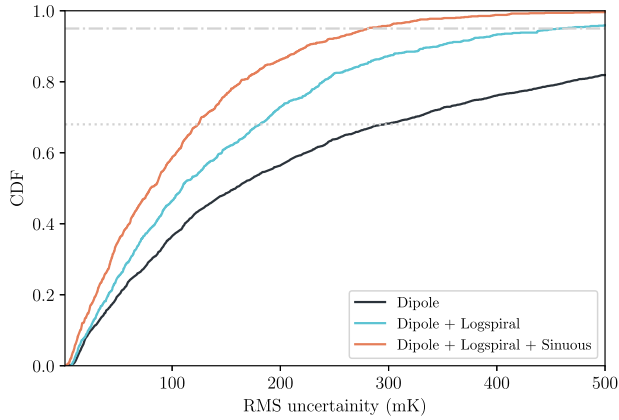
We can improve this extraction if we utilize different antenna designs. For example, when we simultaneously fit the observations from a dipole and a conical log spiral antenna, as shown in the middle panel of Fig. 11, the uncertainty on the extracted signal decreases from 123 to 69 mK, and it also removes some of the bias in the extracted signal. This can be further improved by including all three antennas in the fitting, as shown in the right-hand panel of the figure, where the RMS uncertainty on the extracted signal is 37 mK.

We also estimate the CDF of the RMS uncertainty on the extracted signal from 2000 different mock observations formed from the detailed foreground model in the top panel of Fig. 12 for three different antenna configurations. The distribution for multiple antennas tends towards lower RMS uncertainty levels. For three antennas, 68 per cent (and 95 per cent) of the mock observations are extracted with  $\text{RMS} < 124 \text{ mK}$  (and  $< 281 \text{ mK}$ ) in comparison to  $\text{RMS} < 297 \text{ mK}$  (and  $< 971 \text{ mK}$ ) for a single antenna. In the bottom panel, we show the PDF of the normalized  $\chi^2$  statistic of the fit for the detailed foreground model. The variance of the  $\chi^2$  distribution decreases from  $\sigma(\chi^2) = 0.061$  to 0.034, as we include more antennas in our fit. These values are tabulated in Table 2. These results suggest that utilizing different antenna designs is quite powerful for extracting cosmological signals when considering more complicated foreground models.

However, the performance of the signal extraction depends on how the 21-cm signal and foregrounds are simulated in their modelling sets. For example, in Fig. 13, we show the extraction of the 21-cm signal when we take an EDGES like profile to render the mock



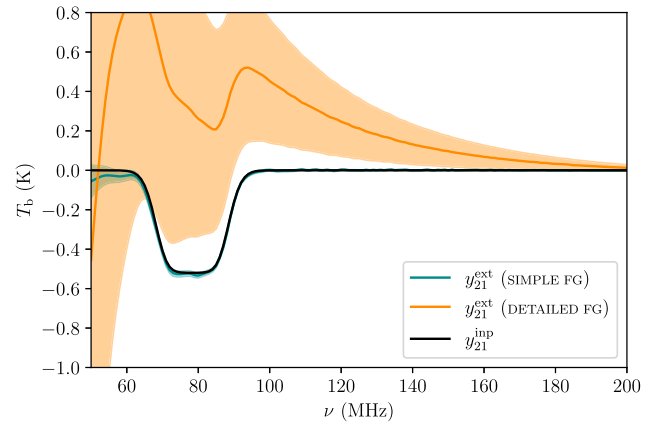
**Figure 11.** Extracted 21-cm signals for the mock observation generated from a physical 21-cm signal and a foreground model taken from the detailed foreground modelling set for dipole (left), dipole + conical log spiral (middle), and dipole + conical sinuous + conical log spiral (right) antennas configuration while simultaneously fitting four time bins.



**Figure 12.** **Top panel:** CDF of RMS uncertainty level on the extracted signal estimated from 2000 different mock observations for the detailed foreground model. Different colours represent different antenna configurations used in the analysis. The uncertainty level at 68 and 95 per cent level are tabulated in Table 2. **Bottom panel:** PDF of the normalized  $\chi^2$  statistic estimated from 2000 different mock observations for each configuration.

**Table 2.** This tabulates the RMS uncertainty level (in mK) on the extracted 21-cm signal for 68 and 95 per cent of the mock observations, and the variance of the PDF of  $\chi^2$  for different antenna configurations shown in Fig. 12 for the detailed foreground model.

Configuration	RMS  <sub>68 per cent</sub>	RMS  <sub>95 per cent</sub>	$\sigma(\chi^2)$
Dipole	296.07	970.55	0.061
Dipole + Logspiral	191.24	464.83	0.042
Dipole + Logspiral + Sinuous	124.03	280.64	0.034



**Figure 13.** Extracted 21-cm signal for the simple (cyan) and detailed (orange) foreground model. The input 21-cm signal (black line) follows the EDGES profile and is not a part of the 21-cm modelling set.

observation while the 21-cm modelling set is simulated with only the physical signals. We simultaneously fit the data for all three antennas from four time slices. In this scenario, we can recover the 21-cm signal with the simple foreground model (cyan) with reasonable accuracy. The minimization of DIC, in this case, requires more 21-cm modes to fit the data. However, the signal extraction is significantly biased for our detailed foreground model (orange) because of the more significant covariance between the foreground and 21-cm modes. This could become substantially more challenging when one includes the uncertainties in the foreground model as well.

## 5 SUMMARY

In this article, we have shown a method to utilize multiple different antenna designs in the global 21-cm signal extraction. Our analysis is based on the current approach of the REACH experiment, but it can easily be generalized to any global 21-cm signal experiment. Our formalism for extracting the 21-cm signal is based on fitting the modes derived from the SVD of the 21-cm signal, and beam-weighted foreground modelling sets (Tauscher et al. 2018; Tauscher et al. 2020b).

We demonstrated the impact of utilizing multiple antennas to better constrain the 21-cm signal by considering two different foreground models (i) A simple foreground model, where we assumed the spectral index to be constant across the sky and used that spectral index to create a modelling set, and (ii) A more realistic foreground

model, where we considered a variation of the spectral index across the sky, divided into 30 regions.

We find that for the simple foreground model, we can extract the signal with any antenna design when we simultaneously fit the data for multiple time slices. This extraction is further improved when we include multiple different antennas in our fitting. To better understand this improvement, we quantify some statistical measures of the extracted 21-cm signal for an ensemble of different mock observations. For multiple antennas, the CDF of the RMS uncertainties tends towards lower RMS uncertainty levels. In addition, the variance of the PDF decreases as we include more antennas, and it becomes more centred around  $\chi^2 = 1$ .

The impact of including multiple antennas in the fitting becomes more pronounced when we consider the more realistic foreground model in the mock observation. In that case, with only a single antenna, even when we fit the data for multiple time bins, we cannot precisely recover the 21-cm signal. Both the large RMS of the reconstructed signal and the significant bias show that only using a single antenna is likely, but not sufficient, to recover the cosmological signal. Instead, when we use multiple antennas of different types, we reduce the bias in the extracted signal and also the RMS uncertainty level by a factor of 2–3.

The signal extraction presented in this analysis could be further improved by measuring all four Stokes parameters from different antennas instead of only Stokes I (Tauscher et al. 2020b). This could further reduce the covariance between the signal and foreground modes and could be crucial for analysing more detailed foreground models, which overlap more significantly with the 21-cm models. This will be considered in a future work. We also plan to utilize the spectral constraints derived in this paper to calculate the full posterior probability distribution of any signal parameter space of choice while marginalizing over the linear foreground modes (Rapetti et al. 2020), which results in a tighter constraint on the 21-cm signal. To derive these constraints, one can quickly scan the parameter space by using the sky-averaged 21-cm signal emulators (Cohen et al. 2020; Bye, Portillo & Fialkov 2022).

## ACKNOWLEDGEMENTS

We thank the Center for Information Technology of the University of Groningen for their support and for providing access to the Peregrine high performance computing cluster. P.D.M acknowledges support from the Netherlands organization for scientific research (NWO) VIDI grant (dossier 639.042.730). LVEK acknowledges the financial support from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (Grant agreement No. 884760, ‘CoDEX’).

## DATA AVAILABILITY

Accompanying code is available at <https://github.com/anchal-009/SAVED21cm>. The data underlying this article will be shared on reasonable request to the corresponding author.

## REFERENCES

Alvarez M. A., Shapiro P. R., Ahn K., Iliev I. T., 2006, *ApJ*, 644, L101  
 Anstey D., de Lera Acedo E., Handley W., 2021, *MNRAS*, 506, 2041  
 Anstey D., Acedo E. D. L., Handley W., 2022, *MNRAS*, 520, 850, <https://arxiv.org/abs/2210.04707>  
 Barkana R., 2018, *Nature*, 555, 71

Barkana R., Loeb A., 2001, *Phys. Rep.*, 349, 125  
 Barkana R., Outmezguine N. J., Redigol D., Volansky T., 2018, *Phys. Rev. D.*, 98, 103005  
 Barry N. et al., 2019, *ApJ*, 884, 1  
 Bassett N., Rapetti D., Tauscher K., Burns J. O., Hibbard J. J., 2021, *ApJ*, 908, 189  
 Becker R. H. et al., 2001, *AJ*, 122, 2850  
 Berlin A., Hooper D., Krnjaic G., McDermott S. D., 2018, *Severely Constraining Dark-Matter Interpretations of the 21-cm Anomaly*, <https://link.aps.org/doi/10.1103/PhysRevLett.121.011102>  
 Bernardi G., McQuinn M., Greenhill L. J., 2015, *ApJ*, 799, 90  
 Bevins H. T. J., Handley W. J., Fialkov A., de Lera Acedo E., Greenhill L. J., Price D. C., 2021, *MNRAS*, 502, 4405  
 Boera E., Becker G. D., Bolton J. S., Nasir F., 2019, *ApJ*, 872, 101  
 Bouwens R. J., Illingworth G. D., Oesch P. A., Caruana J., Holwerda B., Smit R., Wilkins S., 2015, *ApJ*, 811, 140  
 Bowman J. D., Rogers A. E. E., Monsalve R. A., Mozdzen T. J., Mahesh N., 2018, *Nature*, 555, 67  
 Bye C. H., Portillo S. K. N., Fialkov A., 2022, *ApJ*, 930, 79  
 Cohen A., Fialkov A., Barkana R., Monsalve R. A., 2020, *MNRAS*, 495, 4845  
 DeBoer D. R. et al., 2017, *PASP*, 129, 045001  
 de Lera Acedo E. et al., 2022, *Nature Astron.*, 6, 984  
 de Oliveira-Costa A., Tegmark M., Gaensler B. M., Jonas J., Landecker T. L., Reich P., 2008, *MNRAS*, 388, 247  
 Dyson J., 1965, *IEEE Trans. Antennas Propag.*, 13, 488  
 Ewall-Wice A., Chang T. C., Lazio J., Doré O., Seiffert M., Monsalve R. A., 2018, *ApJ*, 868, 63  
 Ewall-Wice A., Chang T.-C., Lazio T. J. W., 2020, *MNRAS*, 492, 6086  
 Fan X. et al., 2003, *AJ*, 125, 1649  
 Feng C., Holder G., 2018, *ApJ*, 858, L17  
 Fialkov A., Barkana R., 2019, *MNRAS*, 486, 1763  
 Furlanetto S. R., Oh S. P., Briggs F. H., 2006, *Phys. Rep.*, 433, 181  
 Ghara R. et al., 2020, *MNRAS*, 493, 4728  
 Hibbard J. J., Tauscher K., Rapetti D., Burns J. O., 2020, *ApJ*, 905, 113  
 Hills R., Kulkarni G., Meerburg P. D., Puchwein E., 2018, *Nature*, 564, E32  
 Kaplinghat M., Chu M., Haiman Z., Holder G. P., Knox L., Skordis C., 2003, *ApJ*, 583, 24  
 Kolopanis M. et al., 2019, *ApJ*, 883, 133  
 Komatsu E. et al., 2011, *ApJS*, 192, 18  
 Koopmans L. et al., 2015, in Bourke Tyler L., Proceedings of Advancing Astrophysics with the Square Kilometre Array (AASKA14). 9 -13 June, 2014. PoS, AASKA14, Giardini Naxos, Italy  
 Li W. et al., 2019, *ApJ*, 887, 141  
 Liu H., Outmezguine N. J., Redigolo D., Volansky T., 2019, *Phys. Rev. D.*, 100, 123011, <https://link.aps.org/doi/10.1103/PhysRevD.100.123011>  
 Mellema G., Koopmans L., Shukla H., Datta K. K., Mesinger A., Majumdar S., 2015, in Bourke Tyler L., Proceedings of Advancing Astrophysics with the Square Kilometre Array (AASKA14). 9 -13 June, 2014. HI tomographic imaging of the Cosmic Dawn and Epoch of Reionization with SKA, PoS, AASKA14, Giardini Naxos, Italy  
 Mertens F. G. et al., 2020, *MNRAS*, 493, 1662  
 Mirocha J., Harker G. J. A., Burns J. O., 2015, *ApJ*, 813, 11  
 Mirocha J., Furlanetto S. R., Sun G., 2017, *MNRAS*, 464, 1365  
 Paciga G. et al., 2013, *MNRAS*, 433, 639  
 Philip L. et al., 2018, *J. Astron. Instrum.*, 8, 1950004, <https://arxiv.org/abs/1806.09531>  
 Price D. C. et al., 2018, *MNRAS*, 478, 4193–4213  
 Pritchard J. R., Loeb A., 2012, *Rep. Prog. Phys.*, 75, 086901  
 Rao M. S., Subrahmanyan R., Shankar N. U., Chluba J., 2016, *AJ*, 153, 26  
 Rapetti D., Tauscher K., Mirocha J., Burns J. O., 2020, *ApJ*, 897, 174  
 Reis I., Fialkov A., Barkana R., 2020, *MNRAS*, 499, 5993  
 Robertson B. E., Ellis R. S., Furlanetto S. R., Dunlop J. S., 2015, *ApJ*, 802, L19  
 Shaver P. A., Windhorst R. A., Madau P., de Bruyn A. G., 1999, *A&A*, 345, 380

- Sims P. H., Pober J. C., 2020, *MNRAS*, 492, 22  
Singh S., Subrahmanyan R., 2019, *ApJ*, 880, 26  
Singh S. et al., 2018, *ApJ*, 858, 54  
Singh S. et al., 2021, *Nature Astron.*, 6, 607, <https://arxiv.org/abs/2112.06778>  
Sokolowski M. et al., 2015, *PASA*, 32, e004  
Spiegelhalter D. J., Best N. G., Carlin B. P., Van Der Linde A., 2002, *J. R. Stat. Soc. B*, 64, 583  
Spinelli M., Bernardi G., Garsden H., Greenhill L. J., Fialkov A., Dowell J., Price D. C., 2021, *MNRAS*, 505, 1575  
Switzer E. R., Liu A., 2014, *ApJ*, 793, 102  
Tauscher K., Rapetti D., Burns J. O., Switzer E., 2018, *ApJ*, 853, 187  
Tauscher K., Rapetti D., Burns J. O., 2020a, *ApJ*, 897, 132  
Tauscher K., Rapetti D., Burns J. O., 2020b, *ApJ*, 897, 175  
Tauscher K. et al., 2021, *ApJ*, 915, 66  
Vedantham H. K., Koopmans L. V. E., de Bruyn A. G., Wijnholds S. J., Ciardi B., Brentjens M. A., 2013, *MNRAS*, 437, 1056

This paper has been typeset from a  $\text{\TeX/L\AA\TeX}$  file prepared by the author.