# 21-cm foreground removal using AI and the frequency-difference technique

Feng Shi[1,2,*], Haoxiang Chang[1], Le Zhang,[3,4,2] Huanyuan Shan,[5,6] Jiajun Zhang[5,6]
Suiping Zhou,[1] Ming Jiang,[7] and Zitong Wang[1]

[1]*School of Aerospace Science and Technology, Xidian University,
Xi'an 710126, People's Republic of China*
[2]*Peng Cheng Laboratory, No. 2, Xingke 1st Street, Shenzhen 518000, People's Republic of China*
[3]*School of Physics and Astronomy, Sun Yat-sen University,
2 Daxue Road, Tangjia, Zhuhai, 519082, People's Republic of China*
[4]*CSST Science Center for the Guangdong-Hong Kong-Macau Greater Bay Area,
Zhuhai 519082, People's Republic of China*
[5]*Shanghai Astronomical Observatory (SHAO), Nandan Road 80, Shanghai 200030, China*
[6]*University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China*
[7]*National Key Laboratory of Radar Signal Processing, Xidian University,
Xi'an 710126, People's Republic of China*

The deep learning technique has been employed in removing foreground contaminants from 21-cm intensity mapping, but its effectiveness is limited by the large dynamic range of the foreground amplitude. In this study, we develop a novel foreground removal technique grounded in U-Net networks. The essence of this technique lies in introducing an innovative data preprocessing step specifically, utilizing the temperature difference between neighboring frequency bands as input. Combining with the frequency difference, we refer to our method as the UNet-fd (UNet frequency-difference), where the U-Net structure is the same as that in Deep21. Based on our tests, we demonstrate that this frequency-difference preprocessing technique can substantially reduce the dynamic range of foreground amplitudes by approximately two orders of magnitude. This reduction proves to be highly advantageous for the U-Net foreground removal. We observe that the HI signal can be reliably recovered, as indicated by the cross-correlation power spectra showing unity agreement at the scale of $k \lesssim 0.3\ h\mathrm{Mpc}^{-1}$ in the absence of instrumental effects. Moreover, accounting for the systematic beam effects, our reconstruction displays consistent autocorrelation and cross-correlation power spectrum ratios at the $1\sigma$ level across scales $k \lesssim 0.1\ h\mathrm{Mpc}^{-1}$, with only a 10% reduction observed in the cross-correlation power spectrum at $k \simeq 0.2\ h\mathrm{Mpc}^{-1}$. The effects of redshift-space distortion are also reconstructed successfully, as evidenced by the quadrupole power spectra matching with the target truth. In order to test how thermal noise affects the performance of our method, we simulated various white noise levels in the map. This shows the mean cross-correlation ratio $\bar{R}_{\mathrm{cross}} \gtrsim 0.8$ when the level of the thermal noise is smaller than or equal to that of the HI signal. In comparison, our method outperforms the traditional principal component analysis (PCA) method. The PCA-derived cross-correlation ratios are underestimated by around 60%. We conclude that the proposed frequency-difference technique can significantly enhance network performance by reducing the amplitude range of foregrounds and aiding in the prevention of HI loss.

## I. INTRODUCTION

In the next few decades, 21-cm cosmology has the potential to dramatically improve our knowledge of the universe, providing a fully three-dimensional mapping of the HI spectral line from the dark ages [1,2], to the cosmic dawn [e.g., [3,4]], and the postreionization [e.g., [5–8]]. One promising technique for surveying the low-redshift large-scale structure (LSS) of the Universe is 21-cm intensity mapping (IM), which involves measuring the integrated HI emission lines that originate from unresolved sources in large-volume portions of the sky [9–14]. While this is conceptually similar to the traditional galaxy redshift survey, it differs significantly in that 21-cm IM is sensitive to all sources of emission rather than just cataloging the brightest galaxies. This allows for mapping LSS through the redshift desert and beyond (such as $1 \lesssim z \lesssim 6$), where optical spectroscopy is challenging [15].

A large number of interferometric and imaging radio telescopes are currently conducting and will continue to carry out 21-cm IM observations, such as CHIME (Canadian Hydrogen Intensity Mapping Experiment) [16,17],

*fshi@xidian.edu.cn

063509-1

Tianlai [18–21], BINGO (Baryon acoustic oscillations In Neutral Gas Observations) [22,23], MeerKAT (The South African Square Kilometer Array Pathfinder) [24,25], HIRAX (The Hydrogen Intensity and Real-time Analysis eXperiment) [26], and the SKA (The Square Kilometer Array) [27]. So far, most studies have focused on 21-cm IM in cross-correlation with galaxy surveys [28–33]. More recently, there has been a successful detection of the HI autocorrelation power spectrum on small scales using MeerKAT [34].

However, the biggest challenge for upcoming observational data analysis is the presence of foreground contaminants, such as the Galactic synchrotron and free-free emissions and extragalactic radio point sources, which are about 4 orders of magnitude brighter than the HI signal [35–38].

Over the years, a variety of signal separation algorithms have been consistently proposed. These algorithms can be categorized as either "blind" or "nonblind," depending on whether prior knowledge of the signal, foreground, or noise is required for the separation process. A comprehensive overview of current techniques for mitigating 21-cm foreground is provided by [39]. For instance, principal component analysis (PCA) [40], singular value decomposition (SVD) [41,42], independent component analysis (ICA) [43,44], and generalized morphological component analysis (GMCA) [45] represent blind (or "semiblind") methods commonly employed in 21-cm intensity mapping. Conversely, non-blind methods such as generalized needlet internal linear combination (GNILC) [46,47], Gaussian process regression (GPR) [43,48], and the Karhunen-Loeve transform [49], are also extensively cited in the literature. Moreover, an approach known as "foreground avoidance" is utilized, where Fourier modes measured above the so-called "wedge" region in $(k_\perp, k_\parallel)$-space are dominantly associated with the 21-cm signal [50]. In addition, the Bayesian framework [51–54] allows for the simultaneous inference of 21-cm signals and foreground signals under certain *a priori* assumptions.

More recently, there has been a growing interest in exploring deep learning techniques for foreground removal, such as by using deep neural networks with U-Net architecture [55,56]. It has been demonstrated that the U-Net is effective in recovering the HI fluctuation signal, provided that certain preprocessing steps, such as PCA, are undertaken. This initial preprocessing aims to firstly eliminate the dominated foreground components. Subsequently, the U-Net is applied to further eliminate any remaining foreground residuals. However, it should be noted that employing PCA-based preprocessing, which involves subtracting a specific number of PCA modes, may result in an undesired loss of the HI signal.

All of these methods essentially assume that the statistical characteristics of the foreground, as well as the smoothness characteristics of the foreground emissions

across the frequencies, differ from those of the 21-cm signal. As a result, they can effectively remove most of the foreground components [57–59]. However, because the HI signal vector and individual foreground component vectors are not perfectly orthogonal, the process of subtracting the foreground leads to a simultaneous subtraction of the signal itself. Meanwhile, achieving a clean reconstruction of the 21 cm signal is further complicated by the presence of contamination from nonsmoothed foreground components [60–63].

In order to avoid signal loss and to more cleanly eliminate non-smoothed foreground components, we propose a new method—a technique that utilizes the temperature difference of the sky maps in neighboring frequency bands. Since the HI and the foreground have different frequency correlations, this is expected to significantly reduce the dynamic range of the foreground while avoiding HI signal loss. We will demonstrate that the U-Net performs more reliably in recovering large-scale HI information when trained using the frequency-difference map as input.

This paper is organized as follows. In Sec. II we introduce our simulation. In Sec. III, we present our foreground removal method, including the U-Net structure and the frequency-difference preprocessing technique. We verify our method and present the results in Sec. IV, and summarize our main findings in Sec. V.

## II. DATASETS

This section describes our simulated 21-cm intensity maps, which are generated using CRIME[1] [64] in HEALPix pixelization scheme [65]. The code is designed to generate the main components of the 21-cm intensity mapping observations, including the cosmological 21-cm signal, foreground emissions, and instrumental noise. Here, we briefly introduce the simulation method and refer the reader to the references for more details.

### A. Cosmological HI signal

The observed effective brightness temperature of the HI signal, $T_{HI}$, is related to the HI overdensity $\delta_{HI}$ at a given sky position $\hat{\boldsymbol{n}}$ and redshift $z$ as follows [e.g. [66]]:

$$T_{HI}(\hat{\boldsymbol{n}}, z) = \bar{T}_{HI}(z)(1 + \delta_{HI}(\hat{\boldsymbol{n}}, z)). \qquad (1)$$

Here the mean brightness temperature can be expressed as

$$\bar{T}_{HI}(z) = 190.55 \frac{\Omega_b h(1+z)^2 x_{HI}(z)}{\sqrt{\Omega_m(1+z)^3 + \Omega_\Lambda}} \text{ mK}, \qquad (2)$$

where $h$ represents the reduced Hubble parameter, defined as $h = H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1})$, $x_{HI}$ stands for the neutral hydrogen mass fraction relative to the total baryons, and

---

[1]http://intensitymapping.physics.ox.ac.uk/CRIME.html.

$\Omega_b$, $\Omega_m$, and $\Omega_\Lambda$ correspond to the present fractions of baryon, total matter, and dark energy densities, respectively.

In CRIME, the HI overdensity $\delta_{HI}$ and velocity perturbations are first generated by employing a log-normal model on a Gaussian realization of the dark matter density field within a Cartesian box. The results are then projected onto spherical sky shells and pixelated to create maps of the 21-cm brightness temperature across various frequency bands, each corresponding to a specific redshift from the observer. The relationship between the brightness temperature and the underlying HI overdensity is established through Eqs. (1) and (2), assuming $x_{HI}(z) = 0.008(1 + z)$.

Meanwhile, the redshift-space distortion (RSD) is implemented by perturbing the cosmological redshift of each cell with the redshift distortion $\Delta z_{RSD} = (1 + z)v_r/c$, where $v_r$ is the line-of-sight velocity of each cell, and $c$ denotes the speed of light.

In this study, we utilize a map with $N_{side} = 256$, corresponding to an angular resolution of 13.73 arcmin. We consider the frequency range spanning from 1100 to 1164 MHz, which corresponds to a redshift range between 0.29 and 0.22. Subsequently, the simulated full sky is divided into 192 sky patches, each having dimensions of $64^3$. These patches cover approximately 214.86 deg$^2$ of the sky and have a volume equivalent to that of a box with a length of about $187h^{-1}$ Mpc.

The cosmological parameters are set as follows: $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, $\Omega_b = 0.049$, $h = 0.67$, and $\sigma_8 = 0.8$.

## B. Foregrounds

In the simulations, we consider four main foreground components: Galactic synchrotron emission, extragalactic point sources, Galactic and extragalactic free-free emissions. Given the distinct distribution characteristics of anisotropy and isotropy, the simulation for generating the foreground combines two different methods. The Galactic synchrotron emission originates from the accelerated motion of energetic charged particles dispersed in the Galactic magnetic field. It exhibits a highly anisotropic angular structure, with a steep increase in brightness temperature toward the Galactic plane. To produce such a structure, the Galactic synchrotron emission is then simulated by extrapolating the brightness temperature $T_{Haslam}(\hat{n})$ from the Haslam map [67] to the relevant frequencies. This is achieved using a direction-dependent spectral index $\beta(\hat{n})$ [68], expressed as

$$T_{syn,0}(\nu, \hat{n}) = T_{Haslam}(\hat{n})\left(\frac{408 \text{ MHz}}{\nu}\right)^{\beta(\hat{n})}. \quad (3)$$

Because of the low resolution of the Haslam map (approximately 0.85°), the subscript 0 here denotes the interpolation results at larger scales. As Galactic synchrotron anisotropy on small scales cannot be obtained from the low-resolution Haslam sky map, we simulate the small-scale synchrotron component by generating the isotropic structure of synchrotron emission through Gaussian realizations of the angular power spectra, as modeled by [37],

$$C_\ell(\nu_1, \nu_2) = A\left(\frac{\ell_{ref}}{\ell}\right)^\beta \left(\frac{\nu_{ref}^2}{\nu_1 \nu_2}\right)^\alpha \exp\left(-\frac{\log^2(\nu_1/\nu_2)}{2\xi^2}\right), \quad (4)$$

where $A$ represents the amplitude of the power spectrum, and $\xi$ denotes the frequency-space correlation length for a given foreground component. We employ the reference scale $\ell_{ref} = 1000$ and the reference frequency $\nu_{ref} = 130$ MHz. In order to confine the synchrotron simulation to small scales, the random realizations are rescaled to match the resolution scale overlapping with that of the Haslam map before they are incorporated into the final Galactic synchrotron map. Additionally, Eq. (4) is also used to simulate free–free emission and point sources. The model parameter values are drawn from [37] and are provided in Table I. Note that, the Haslam map used in the CRIME contains residual striping and point-source artifacts, which could lead to inadvertent duplication of point sources in the simulations. This concern can be alleviated by the utilization of an improved version of the Haslam map, which was reprocessed by [69].

We also consider the presence of polarized foregrounds, where frequency-dependent Faraday rotation affects synchrotron emission, as discussed in references such as [70]. This phenomenon leads to leakage into the unpolarized signal, which can potentially be a problematic contribution for foreground removal. In CRIME, the influence of polarized synchrotron emission on foregrounds is characterized by two parameters: the Faraday-space correlation length ($\xi_\psi$) and the polarization leakage fraction ($\epsilon_p$). Further details can be found in [64].

Following the approach in CRIME, we adopt a typical value of $\xi_\psi = 0.5$ rad m$^{-2}$ to match simulations obtained using the Hammurabi code [71]. For simplicity, we choose $\epsilon_p = 0.01$, which, importantly, depends entirely on the level of systematic control within the instrument.

In the left panel of Fig. 1, we display the comparison of the angular power spectra. For these calculations,

TABLE I. Model parameters for the fiducial foreground $C_\ell(\nu_1, \nu_2)$ in Eq. (4), adapted from [37], using the reference values of $\ell_{ref} = 1000$ and $\nu_{ref} = 130$ MHz.

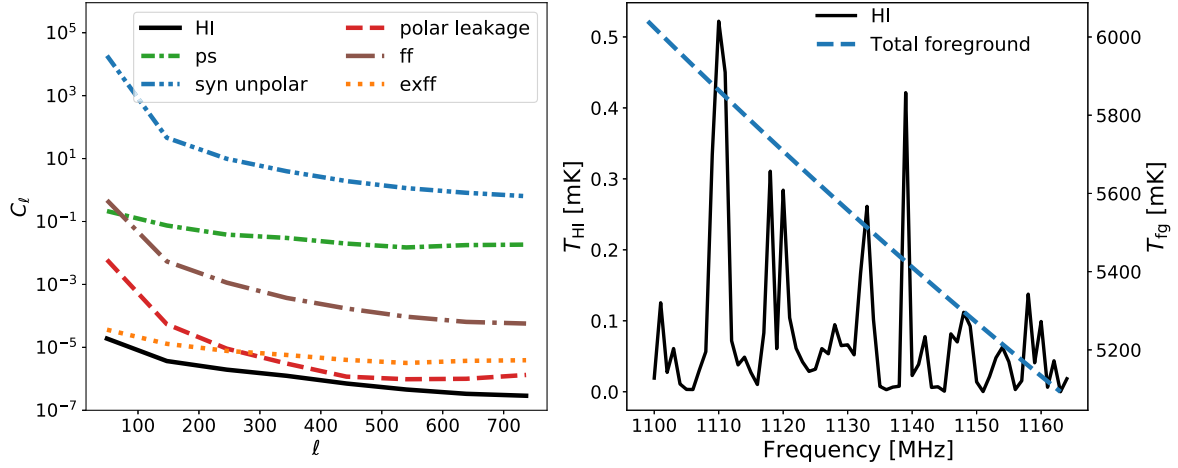| Foreground | $A[\text{mK}^2]$ | $\beta$ | $\alpha$ | $\xi$ |
|---|---|---|---|---|
| Galactic synchrotron | 700 | 2.4 | 2.80 | 4.0 |
| Point sources | 57 | 1.1 | 2.07 | 1.0 |
| Galactic free-free | 0.088 | 3.0 | 2.15 | 35 |
| Extragalactic free-free | 0.014 | 1.0 | 2.10 | 35 |

FIG. 1. Left: a comparison of the angular power spectrum among different simulated components at a frequency of $\nu = 1110$ MHz. Results are computed for a randomly selected patch from the full sky, which has been divided into 192 patches. The cosmological HI is represented by the black solid line, while other components of the foreground are depicted by colored dashed or dotted lines, as indicated. Right: the brightness temperature as a function of frequency for a randomly selected sky pixel. The black solid line corresponds to the cosmological HI, while the blue dashed line represents the total foreground. For a clear comparison, note that the amplitude scales are different for the foreground and HI components.
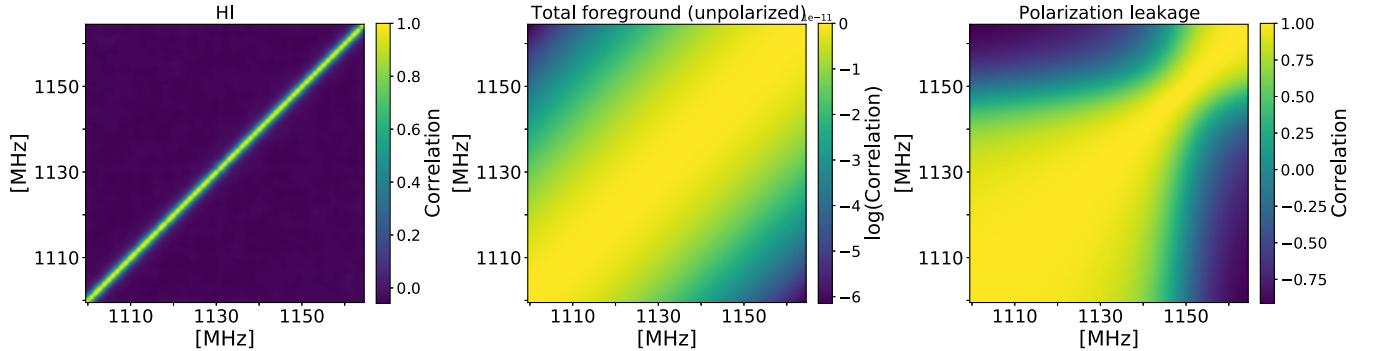


FIG. 2. Frequency-frequency correlation matrices for cosmological HI (left), total unpolarized foreground (center), and polarization leakage (right). Please note that the correlation coefficients of the unpolarized foreground are displayed in logarithmic values.

we employed the publicly available code NaMaster[2] [72], which was used to calculate the spectra among different components at a frequency of $\nu = 1110$ MHz. Black solid lines are plotted for the cosmological HI, while other colored dashed or dotted lines represent different foreground components. It is evident that the foregrounds dominate the HI signal by 4-5 orders of magnitude in terms of $\sqrt{C_\ell}$.

The right panel of Fig. 1 shows a randomly selected sky pixel as an example to demonstrate the dependence of brightness temperature on frequency. The black solid and blue dashed lines correspond to the cosmological HI and total foreground, respectively. As expected, the foreground temperature is smoothly distributed, whereas the HI exhibits significantly more fluctuations in temperature.

Figure 2 displays the respective frequency-frequency correlation matrices of the cosmological HI, the total unpolarized foreground, and the polarized leakage, providing statistical evidence regarding the frequency coherence scales of the different components. Notably, the frequency correlation of the unpolarized foreground is much stronger than that of the HI signal. However, for the polarized synchrotron, which arises due to frequency-dependent Faraday rotation, the correlation lengths are significantly reduced. This non-smoothed component substantially increases the presence of foreground residuals in foreground-cleaned maps [73].

## C. Instrument effects

To simulate the effect of the beam of a radio telescope, we employ a Gaussian beam model for simplicity.
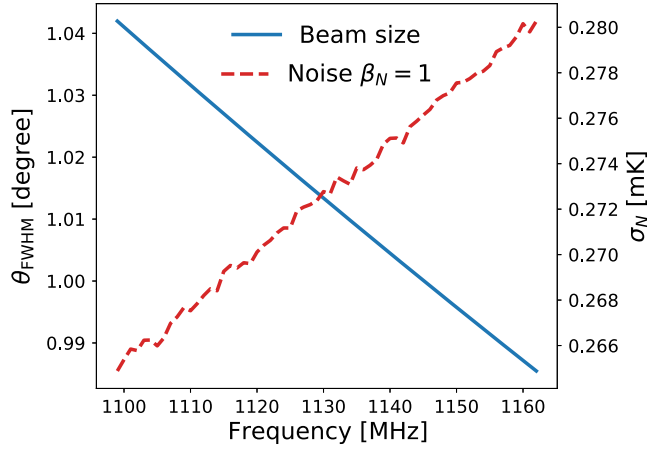
FIG. 3. Beam size (blue solid) and thermal noise (red dashed) as a function of frequency, which includes 64 channels. The noise is plotted for the case with $\beta_N = 1$ according to Eq. (7).

This model approximately represents the primary beam response of a telescope as

$$B(\nu, \theta) = \exp\left[-4\ln 2\left(\frac{\theta}{\theta_{\mathrm{FWHM}}(\nu)}\right)^2\right], \qquad (5)$$

where $\theta_{\mathrm{FWHM}}$ represents the full width at half maximum (FWHM) of the primary beam. In Fig. 3, the blue solid line shows the beam size as a function of frequency across a total of 64 channels, spanning from 1100 to 1164 MHz. Ideally, we would anticipate a smooth and gradual change in the beam size as the frequency varies. We then convolve our simulated sky maps (as described in previous sections) with such a beam model to incorporate the beam effect. However, in real observation, there would be frequency-dependent fluctuations in the beam size, which is called the ripple structure, e.g., [74]. We plan to conduct a detailed analysis of this in a future study.

In the telescope receiver noise model, we solely consider uncorrelated white noise. We incorporate a Gaussian random field with a width of $\sigma_N$ into our simulated sky map, given by

$$\sigma_N = T_{\mathrm{sys}}\sqrt{\frac{4\pi f_{\mathrm{sky}}}{\Omega_{\mathrm{beam}} N_{\mathrm{dish}} t_{\mathrm{obs}} \Delta\nu}}, \qquad (6)$$

where $T_{\mathrm{sys}}$ is the total system temperature, which is the sum of the sky and receiver noise; $\Omega_{\mathrm{beam}} \approx 1.133\theta_{\mathrm{FWHM}}^2$ is the solid angle of the primary beam; $f_{\mathrm{sky}}$ is the fraction of sky coverage; $N_{\mathrm{dish}}$ denotes the number of dishes in the observation; $t_{\mathrm{obs}}$ indicates the integration time and $\Delta\nu$ is the bandwidth of each frequency channel.

In this study, to test the performance of our signal recovery method on thermal noise of varying amplitude, we rescale this white noise model by introducing a free parameter $\beta_N$, given by

$$\beta_N = \frac{\sigma_N}{\sigma_{\mathrm{HI}}}, \qquad (7)$$

where $\sigma_{\mathrm{HI}}$ denotes the standard deviation of the HI signal over the simulation cube. Thus, $\beta_N$ is inversely proportional to the signal-to-noise ratio (S/N). In our analysis, we varied $\beta_N$ in the range of $0.0 < \beta_N < 1.0$, which is consistent with the noise amplitude desired from potential present and future intensity mapping configurations, e.g., [55]. In Fig. 3, the red dashed line represents the noise as a function of frequency, specifically for the case of $\beta_N = 1$.

## III. FOREGROUND REMOVAL METHOD

In this section, we will introduce our novel foreground removal method. By implementing the designed frequency-difference technique and utilizing the existing U-Net convolutional neural network (CNN) [75], we find that this strategy can efficiently recover cosmological HI signals with high fidelity by removing both smooth and nonsmooth foregrounds.

### A. Frequency-difference technique

One of the key ingredients of our method is to reduce the dynamic range of foreground amplitudes before feeding them into U-Net. Based on the characterization of the frequency correlation discussed in Sec. II B, *we use the difference in the sky map temperature between two neighboring frequency bands as the input, which we call the frequency-difference technique.* In this case, the foreground temperatures of the neighboring bands can largely cancel each other out due to the strong correlation in frequency. Conversely, The HI signal can be preserved in the frequency-difference map because the HI signal becomes uncorrelated between adjacent frequency bands when the bandwidth $\Delta\nu \gtrsim 2$ MHz (as indicted in Fig. 5).

Figure 4 presents a visual comparison of temperature maps and their differences between neighboring frequency bands. The first and second panel correspond to the pure-HI temperature map, $T_{\mathrm{HI}}(\nu_0)$, and the HI plus foreground map, $T_{\mathrm{HI,fg}}(\nu_0)$ respectively, where $\nu_0 = 1110$ MHz. The third one shows the temperature difference between neighboring bands, $T_{\mathrm{HI,fg}}(\nu_0) - T_{\mathrm{HI,fg}}(\nu_1)$, where $\nu_1 = 1111$ MHz. It is clear that due to the strong correlation of the foreground in frequency, the amplitude range is significantly reduced by 2–3 orders of magnitude by means of this frequency-difference technique. Furthermore, the last panel shows the frequency-difference map while considering the beam. In this regard, it is important to consider that the size of the beam ($\theta_{\mathrm{FWHM}}$) varies with frequency. To ensure consistency in the beam size level, we have smoothed each adjacent-band map to match their relatively lower angular resolution. This was done by convolving the higher frequency map with a Gaussian beam characterized by a beam size of $\Delta\theta_{\mathrm{FWHM}}$, denoted as,
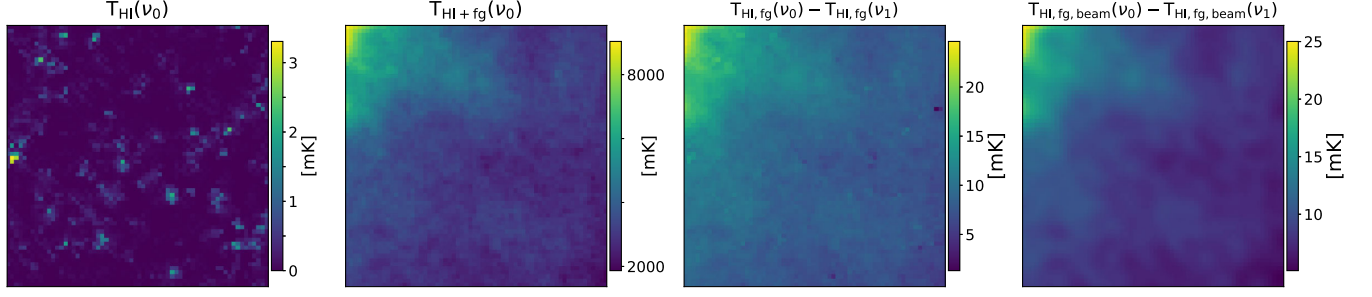
FIG. 4. Visualization of the temperature difference between neighboring frequency bands. The first and second panels correspond to the pure HI temperature map $T_{\mathrm{HI}}(\nu_0)$ and the foreground added map $T_{\mathrm{HI,fg}}(\nu_0)$, at $\nu_0 = 1110$ MHz, respectively. The third panel displays the frequency-difference map $\Delta T = T_{\mathrm{HI,fg}}(\nu_0) - T_{\mathrm{HI,fg}}(\nu_1)$, where $\nu_1 = 1111$ MHz. The last panel shows the same frequency-difference map, but taking into account the beam effect, i.e., $\Delta T_{\mathrm{beam}} = B(\nu_0) \otimes T_{\mathrm{HI,fg}}(\nu_0) - B(\nu_1) \otimes T_{\mathrm{HI,fg}}(\nu_1)$.

$$\Delta\theta_{\mathrm{FWHM}} = \sqrt{\theta_{\mathrm{FWHM,low}}^2 - \theta_{\mathrm{FWHM,high}}^2}. \qquad (8)$$

Here $\theta_{\mathrm{FWHM,low}}$ and $\theta_{\mathrm{FWHM,high}}$ correspond to the beam size of low and high frequency, respectively. It reveals that the frequency-difference map yields a similar outcome, with a significant reduction in the range of amplitude.

In the left panel of Fig. 5, we show the angular cross-correlation power spectrum, $C^{\nu_0 \times (\nu_0 + \Delta\nu)}$ between the pure HI at $\nu_0$ and that at $\nu_0 + \Delta\nu$, which is normalized to the autocorrelation power spectrum of the pure HI at $\nu_0$. Different lines correspond to the results of different $\Delta\nu$, as indicated. We can see a clear positive correlation signal at $\Delta\nu = 1$. This is due to the fact that this bandwidth corresponds to $\Delta z \simeq 0.0012$, and hence the comoving distance is $\Delta r \simeq 4.5~h^{-1}\mathrm{Mpc}$. Thus, the presence of HI large-scale structures (e.g., filaments, voids, and sheets) allows the maps between the two frequencies to remain correlated. Therefore, temperature differences between neighboring bands ($\Delta\nu = 1$) might also lead to a loss of

the HI signal. However, when $\Delta\nu \geq 2$ MHz, there is no longer a correlation of the HI signals between the sky maps, resulting in the corresponding cross-correlation signals decaying rapidly to zero.

In the second panel of Fig. 5, which aims to demonstrate the correlation of the frequency-difference map with the pure HI signal for different $\Delta\nu$ values, we present the ratios obtained by varying $\Delta\nu$. Here, the ratio is defined as $R_{T_{\mathrm{HI}} \times \Delta T_{\mathrm{HI,fg}}} = C_\ell^{T_{\mathrm{HI}} \times \Delta T_{\mathrm{HI,fg}}} / C_\ell^{T_{\mathrm{HI}}}$, where $C_\ell^{T_{\mathrm{HI}} \times \Delta T_{\mathrm{HI,fg}}}$ is the angular cross-correlation power spectra between $T_{\mathrm{HI}}(\nu_0)$ and $\Delta T_{\mathrm{HI,fg}} = T_{\mathrm{HI,fg}}(\nu_0) - T_{\mathrm{HI,fg}}(\nu_0 + \Delta\nu_0)$, and $C_\ell^{T_{\mathrm{HI}}}$ is the angular autocorrelation power spectra of $T_{\mathrm{HI}}(\nu_0)$. The blue solid, red dashed, and green dashed-dot lines correspond to $\Delta\nu = 1, 2$, and 6 MHz, respectively. The error bars represent the standard deviation estimated across the 192 sky patches, as mentioned in Sec. II.

In the cases of $\Delta\nu = 2$ and 6 MHz, the ratios are positive and close to unity, proving that the frequency-difference helps to reduce the foreground contamination and keep the HI signal visible.
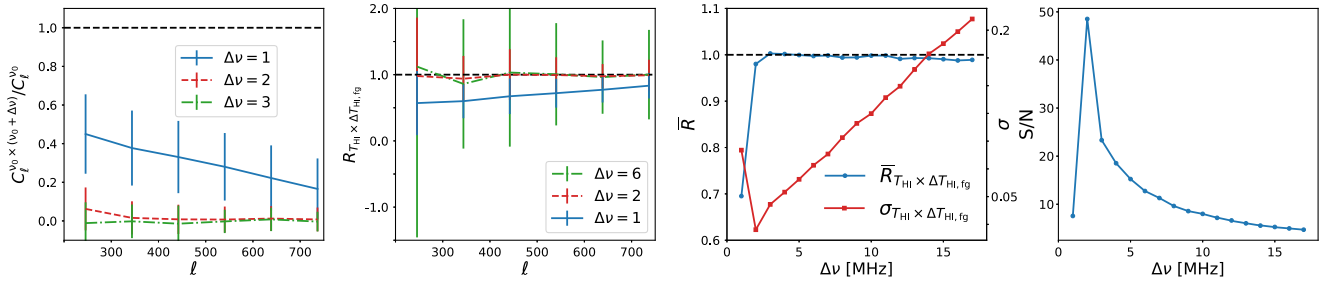


FIG. 5. First: angular cross-correlation power spectrum between the pure HI at $\nu_0$ and that at $\nu_0 + \Delta\nu$, $C^{\nu_0 \times (\nu_0 + \Delta\nu)}$, which is normalized to the autocorrelation power spectrum of the pure HI at $\nu_0$. Different lines correspond to the results of different $\Delta\nu$, as indicated. Second: ratios of the angular power spectra for different $\Delta\nu$, $R_{T_{\mathrm{HI}} \times \Delta T_{\mathrm{HI,fg}}} = C_\ell^{T_{\mathrm{HI}} \times \Delta T_{\mathrm{HI,fg}}} / C_\ell^{T_{\mathrm{HI}}}$, where $C_\ell^{T_{\mathrm{HI}} \times \Delta T_{\mathrm{HI,fg}}}$ is the cross-correlation power spectrum between HI map, $T_{\mathrm{HI}}(\nu_0)$, and frequency-difference map, $\Delta T_{\mathrm{HI,fg}} = T_{\mathrm{HI,fg}}(\nu_0) - T_{\mathrm{HI,fg}}(\nu_0 + \Delta\nu_0)$, and $C^{T_{\mathrm{HI}}}$ denotes the autocorrelation power spectrum of $T_{\mathrm{HI}}(\nu_0)$. The blue solid, red dashed, and green dashed-dot lines correspond to $\Delta\nu = 1, 2$, and 6 MHz, respectively. Here, the error bars correspond to the $1\sigma$ statistical uncertainty estimated from among the all 192 sky patches. Third: mean ratio (averaged over all sky patches for the $C_\ell$ values in the range of $\ell = 200$ to 800), $\bar{R}_{T_{\mathrm{HI}} \times \Delta T_{\mathrm{HI,fg}}}$ (blue line, and its value is related to the left y-axis), and the RMS error, $\sigma_{T_{\mathrm{HI}} \times \Delta T_{\mathrm{HI,fg}}}$ (red line, and its value is related to the right y-axis), as a function of $\Delta\nu$. Fourth: S/N as a function of $\Delta\nu$ according to Eq. (9).

When $\Delta\nu = 1$ MHz, however, the ratios are slightly underestimated. This indicates a loss of the HI signal due to large-scale HI correlations.[3] On the other hand, the error bars for $\Delta\nu = 6$ are considerably larger than those for $\Delta\nu = 2$. This is due to the reduced correlation between the foreground sky maps at wider frequency intervals, leading to more substantial residual foregrounds on the frequency-difference maps.

In the third panel of Fig. 5, by varying $\Delta\nu$ from 2 to 17 MHz, we present the mean ratios averaged over all sky patches, $\bar{R}_{T_{HI} \times \Delta T_{HI,fg}}$ (blue line), and the RMS error, $\sigma_{T_{HI} \times \Delta T_{HI,fg}}$ (red line). The results are computed based on $C_\ell$ values within the range of $\ell = 200$ to 800. The ratios quickly converge to unity when $\Delta\nu > 1$ MHz. However, the uncertainties first reach their minimum at $\Delta\nu = 2$ MHz and then rapidly increase beyond that. Therefore, the width $\Delta\nu$ needs to be optimized to minimize the HI correlation while maximizing the foreground correlation between two frequency bands. To do so, we then compute the signal-to-noise ratios (S/N) as a function of $\Delta\nu$ by,

$$\text{S/N}(\Delta\nu) = \sqrt{\frac{\bar{R}_{T_{HI} \times \Delta T_{HI,fg}}}{\sigma_{T_{HI} \times \Delta T_{HI,fg}}}}. \tag{9}$$

The S/N result are depicted on the right panel of Fig. 5. When $\Delta\nu = 2$, a peak occurs, indicating the potential increase in sensitivity for detecting HI signals.

So, the frequency-difference technique has the potential to significantly reduce foreground contamination. By selecting an appropriate bandwidth, it becomes possible to efficiently diminish foreground residuals in the resulting maps while enhancing the S/N ratio. Additionally, when considering the effect of beam smoothing, the frequency-difference maps at small-scales are smeared out (as seen in the last panel of Fig. 4), which can lead to a challenge in recovering the HI signal.

### B. U-Net network

To further mitigate foreground contamination while recovering the HI signal, we employ a deep neural network based on the U-Net architecture [75]. We will demonstrate that the combination of the frequency-difference technique and U-Net can yield an outperformed performance in recovering HI signals. U-Net is a versatile model for image-to-image translation, which incorporates structural modifications into the convolutional neural network (CNN) framework. These modifications enable U-Net to operate effectively even with limited training images, resulting in more accurate translations.

---

[3]The HI correlation is expected to be enhanced by the small-scale redshift space distortion (RSD) effect (the so-called finger of god effect.). However, in this paper, we do not consider the impact of this effect when using the linear velocity simulations in CRIME.

Figure 6 provides a visualization of our model architecture, which we refer to as "UNet-fd" (UNet frequency-difference). The input and output cubes, both containing $64^3$ grids, represent the frequency-difference maps and the reconstructed HI temperature maps, respectively. The U-Net maps the input cube to the output cube through a symmetric encoder-decoder convolution scheme. The encoder part, which is comprised of 6 convolutional layers, downsamples the input to the bottleneck, where the spatial information is reduced and feature information is increased. Skip connections are used to connect layers in the encoder with corresponding layers in the decoder, allowing to preserve the spatial information during downsampling. The decoder part then upsamples the bottleneck to the required output size by using 6 transposed convolutional layers. Following [55], we perform 3 convolutions at each layer, and apply batch normalization and ReLU activation between convolutional layers, except for the final convolutional block on the output side.

To train the networks, we use the Logcosh loss function,

$$\mathcal{L}(T_p, T_t) = \mathbb{E}[\log \cosh(T_p - T_t)], \tag{10}$$

where $T_p$ denotes the predicted temperature and $T_t$ denotes target truth. The Logcosh loss function has been tested to be more robust and less sensitive to outliers [55]. We train the network from scratch and use the AdamW [76] optimizer, which provides weight decay regularization. The hyperparameters are fine-tuned to optimize the network, following the configuration of Deep21 [55]. The network processes 140 training samples each epoch, followed by 12 validation samples used solely for model evaluation during training. Finally, we save our best model based on the minimum validation-set loss function.

In the training phase, the UNet model is tasked with extracting the targeted HI signal at a given frequency when provided with the frequency-difference map as input. This map combines the signals from two neighboring frequency bands into a single channel. To understand how the UNet model process in extracting the targeted single-band HI from the difference map, a detailed explanation is provided in the Appendix.

On the other hand, we found that without the frequency-difference preprocessing, directly using the U-Net performs poorly in recovering the HI signal. To demonstrate the threshold limit for foreground cleaning by the U-Net, we ran the network by feeding the HI cubes with different-level foregrounds added. To do so, we vary the CRIME-simulated foreground temperature (which is the original output foreground of CRIME as described in Sec. II B) to the desired amplitude level, which is parametrized by

$$\beta_{fg} = \sqrt{\frac{\langle (T_{fg} - \bar{T}_{fg})^2 \rangle}{\langle (T_{HI} - \bar{T}_{HI})^2 \rangle}}, \tag{11}$$
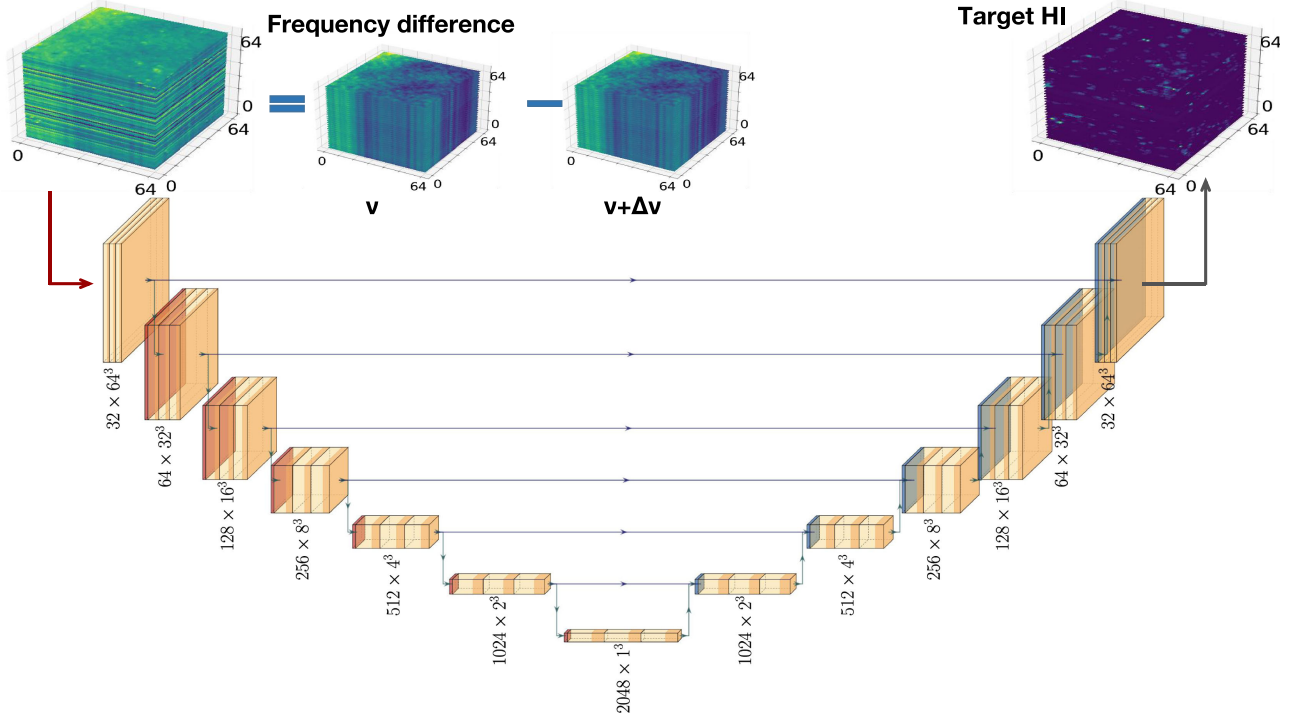
FIG. 6.    Visualization of the UNet-fd architecture. The input and output cubes, each containing grids of size $64^3$, correspond to the frequency-difference maps, where the difference between two frequencies is denoted as $\Delta\nu$, and the target HI temperature maps, respectively. The architecture comprises a total of 13 layers, depicted by the orange box. These layers consist of 6 in the encoder and decoder each, with an additional layer for the bottleneck. Each layer consists of 3 convolutional operations, incorporating batch normalization and activation functions (except for the final convolutional block on the output side). The dark red box and the dark blue box indicate the input and output of each layer. The bottom of each box illustrates the number of channels and the size of the output. The down, right, and up arrows indicate maximum pooling, skip connection, and transpose convolutions, respectively. This visualization is made with the PlotNeuralNet library.

representing the square root of the ratio between the foreground temperature variance and the HI signal variance averaged over pixels.

We then rescale the foreground to the 17 levels of $\beta_{\mathrm{fg}}$: 1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, and 4096. By adding the rescaled foregrounds on the pure HI map, we run U-Net repeatedly for each of the 17 cases and obtain the best-fitting models based on the training samples described above. We then apply the resulted models on the 40 test samples and compare the recovered temperatures with the target HI. In the left panel of Fig. 7, we show the temperature distribution as a function of $T$. The back solid line represents the target HI. The red dashed and blue dash-dotted lines are shown for the U-Net-reconstructed results of $\beta_{\mathrm{fg}} = 32$ and $\beta_{\mathrm{fg}} = 1600$, respectively. The results are averaged over the test samples. As can be seen, the target truth and the reconstructed result of $\beta_{\mathrm{fg}} = 32$ coincide nicely, however, the result of $\beta_{\mathrm{fg}} = 1600$ is systematically underestimated. Figure 7 shows the mean ratios of temperature distribution, $\langle N(T_{\mathrm{rec}})/N(T_{\mathrm{HI}})\rangle$, between the U-Net-reconstructed $N(T_{\mathrm{rec}})$ and the target $N(T_{\mathrm{HI}})$. The red dotted solid line represents the U-Net-reconstructed results as a

function of $\beta_{\mathrm{fg}}$. The blue and green vertical dashed lines are shown for the $\beta_{\mathrm{fg}}$ level of the frequency-difference ($\Delta\nu = 2$) case and the CRIME-simulated foreground case respectively, where the former has $\beta_{\mathrm{fg}} \sim 30$ and the latter has $\beta_{\mathrm{fg}} \sim 5700$. It is evident that the ratios keep values consistent to unity for $\beta_{\mathrm{fg}} < 100$ and 10% reduction at $\beta_{\mathrm{fg}} = 1000$; however, for $\beta_{\mathrm{fg}} > 1000$, the ratios systematically decreases. Therefore, it is virtually impossible to remove the CRIME-simulated foreground contamination using U-Net alone. This is primarily due to the large dynamic range of the map amplitude [35,38,77], which makes extracting correct information with the U-Net network difficult [55,56]. Fortunately, the frequency-difference $\beta_{\mathrm{fg}}$ value falls within the U-Net's working range. We then expect to have a recovered HI temperature that was close to the target truth by applying the U-Net on the frequency-difference map.

## IV. RESULTS

In this section, we present our results of the foreground removal. In order to verify the impact of various
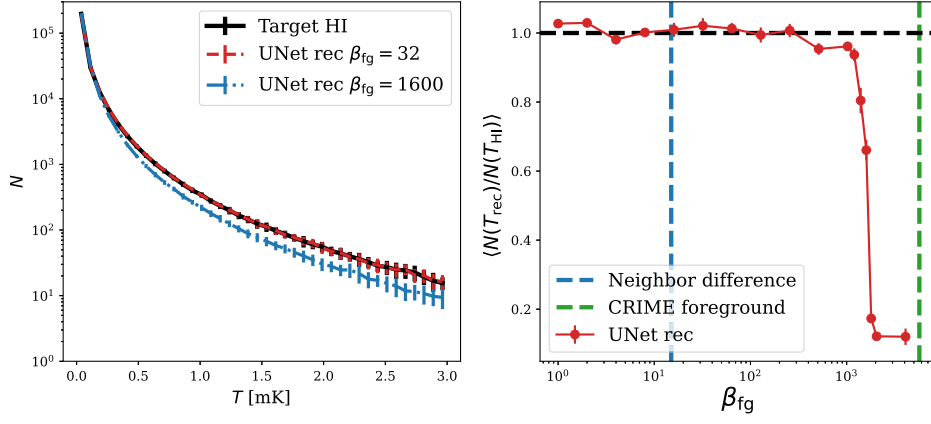
FIG. 7.    Testing UNet-fd cleaning in different-level foreground cases. Left: comparisons of temperature distribution as a function of $T$. The back solid line represents the target HI. The red dashed (blue dash-dotted) line is shown for the U-Net-reconstructed results of $\beta_{\mathrm{fg}} = 32$ ($\beta_{\mathrm{fg}} = 1600$), where $\beta_{\mathrm{fg}}$ is the foreground strength parameter as defined in Eq. (11). Right: mean ratios of temperature distribution, $\langle N(T_{\mathrm{rec}})/N(T_{\mathrm{HI}})\rangle$, between the U-Net-reconstructed $N(T_{\mathrm{rec}})$ and the target $N(T_{\mathrm{HI}})$. The red dotted solid line represents the U-Net-reconstructed results as a function of $\beta_{\mathrm{fg}}$. The blue and green vertical dashed lines are shown for the $\beta_{\mathrm{fg}}$ level of the frequency-difference ($\Delta\nu = 2$) case and the CRIME-simulated foreground case (which is the original output foreground of CRIME as described in Sec. II B), respectively.

contaminations, we implement the tests considering the following three cases:

   (i) Case I: add the cosmological HI signal and the foregrounds without the presence of the beam effect and thermal noise. In other words, the observed data ($d$) is simply the sum of the foreground ($f$) and HI signal ($s$), such that $d = s + f$.
  (ii) Case II: in comparison to Case I, we further consider the beam ($A$) convolution effect, which can be expressed as $d = A \otimes (s + f)$.
 (iii) Case III: to mimic a realistic observation, we introduce white noise ($n$) into the data. In this case, the data can be represented as $d = A \otimes (s + f) + n$.

For each case, we run the U-Net training independently. We also compare our results with those from the PCA method. In our analysis, we subtract the first two and three modes for Case I and Case II, separately. Hence, the notation PCA-$N$ corresponds to the reconstruction for which the first $N$ principal components have been removed.

## A. Case I: Adding HI and foreground

We start with the test for Case I. Figure 8 shows the comparisons in a HEALPix pixel slice randomly drawn from the test set at frequency channel $\nu = 1130$ MHz. Different panels correspond to the target HI, the UNet-fd reconstruction, and the PCA-2 reconstruction, as indicated at the top of each panel. The UNet-fd reconstruction exhibits recognizable structures that are in excellent agreement with the target signals over a range of temperature regions. In contrast, PCA underperforms and over-subtracts the signal, especially in low-temperature regions.

Figure 9 exhibits quantitative comparisons. The left panel shows the temperature distributions. The error bars reflect the $\pm 1\sigma$ variance among the 40 test slices. Black solid, red dashed, and blue dash-dotted lines correspond to the results of the target, UNet-fd, and PCA-2, respectively. As we can see, the UNet-fd reconstruction consistently agrees with the target, but the PCA reconstruction shows a
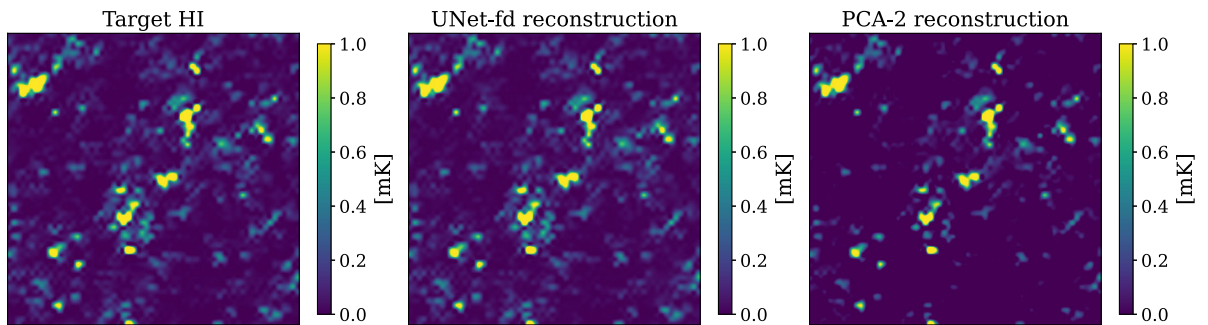


FIG. 8.    Comparison of UNet-fd and PCA reconstructed HI maps for Case I. From left to right, we show the target HI, the UNet-fd reconstruction, and the PCA reconstruction, respectively. The data cube is randomly selected from the test set, and the slice shown here corresponds to a frequency of $\nu = 1130$ MHz.
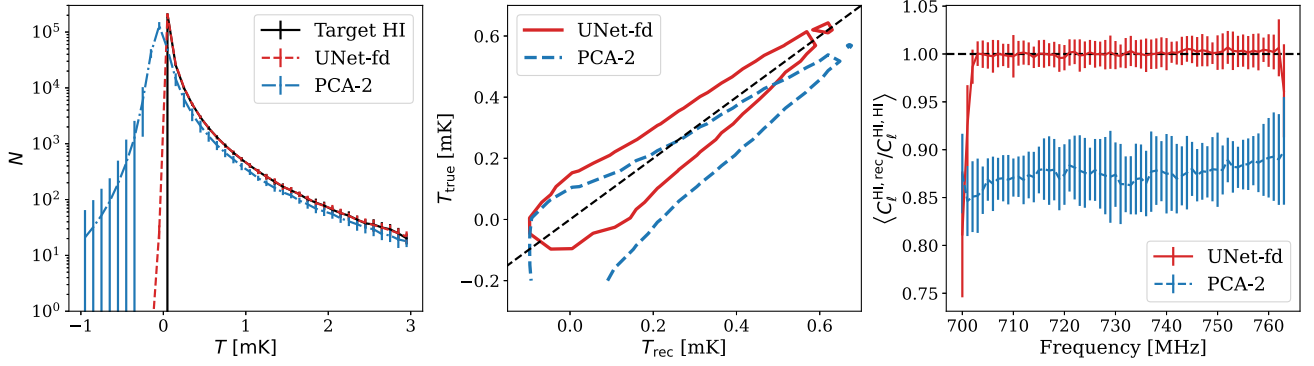
FIG. 9.    The results for Case I. Left: the comparison of the temperature distributions among the truth (black), PCA (blue) and UNet-fd (red). Middle: the reconstructed $T_{\rm rec}$ versus the true $T_{\rm true}$. The results from UNet-fd and PCA are respectively shown with red solid and blue dashed contours, which include about 95% grid cells. Right: the mean ratio of angular power spectrum between $C_{\ell}^{\rm HI, rec}$ and $C_{\ell}^{\rm HI, HI}$ as a function of the frequency channel. Red solid and blue dashed lines correspond to the results of UNet-fd and PCA, respectively. Error bars correspond to $1\sigma$ statistical uncertainty, estimated from the test sets.
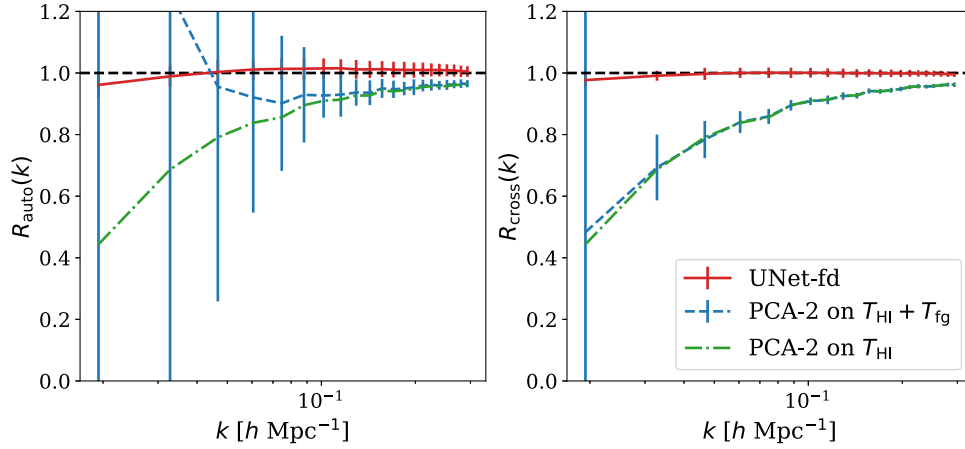


FIG. 10.    For Case I, ratios of autocorrelation (left panel) and cross-correlation (right panel) $P(k)$ between the reconstruction and the true HI as defined in Eqs. (12) and (13). Red solid and blue dashed lines correspond to UNet-fd and PCA results reconstructed from the map of $T_{\rm HI} + T_{\rm fg}$. Green dash-dotted lines show the results obtained by subtracting PCA modes on the pure HI map. Error bars indicate the $\pm$ sigma variance among the 40 test sets.

systematical underestimation and reproduces nearly half of the negative-temperature values due to over-subtraction. The middle panel shows the temperature relation between the reconstructed and the target grid cells. The results from UNet-fd and PCA-2 are respectively shown with red solid and blue dashed contours, which encompass about 95% grid cells. The UNet-fd reconstruction is linearly correlated with the true field and does not exhibit any significant bias, while the PCA result shows a highly biased relation.

To compare the reconstructed HI distribution at each frequency, we estimate the ratio of the angular power spectra between the cross-correlation, $C_{\ell}^{\rm HI, rec}$, and the autocorrelation, $C_{\ell}^{\rm HI, HI}$ as a function of the frequency channel in the right panel of Fig. 9. Red solid and blue dashed lines correspond to the results of UNet-fd and PCA-2 respectively, which are averaged from $C_{\ell}$ within

the range of $0 < \ell < 760$. Error bars indicate the $\pm 1\sigma$ variance among the test sets. The UNet-fd ratios agree very well with the unity over middle-frequency channels, proving that the HI signal was successfully recovered overall. The cross-correlations are underestimated at the boundary-frequency channels, showing the existence of the boundary effect in our method. As a comparison, the PCA results have cross-correlations with about 13% reduction over all frequencies.

To further verify the clustering of the reconstructed signal in 3D space, we measure the power spectra $P(k)$, 2D power spectra $P(k_{\perp}, k_{\parallel})$, and quadrupole $P_2(k)$ in the cubes of test sets, where we used the PYLIANS[4] code [78] for calculating. Figure 10 shows the ratios of auto and

---

[4]https://github.com/franciscovillaescusa/Pylians3/blob/masterz/docs/source/citation.rst.

cross-correlation $P(k)$ between the reconstruction and the target HI, which are respectively defined as,

$$R_{\text{auto}}(k) = \frac{P_{\text{rec,rec}}(k)}{P_{\text{HI,HI}}(k)}, \qquad (12)$$

and,

$$R_{\text{cross}}(k) = \frac{P_{\text{rec,HI}}(k)}{P_{\text{HI,HI}}(k)}. \qquad (13)$$

Here $P_{\text{HI,HI}}(k)$ and $P_{\text{rec,rec}}(k)$ are the autocorrelation power spectrum of the target HI and the reconstructed HI respectively, and $P_{\text{rec,HI}}(k)$ is their cross-correlation power spectrum. Red solid and blue dashed lines correspond to the UNet-fd and the PCA-2 reconstructions, respectively. Error bars indicate the 1-$\sigma$ uncertainty among the 40 test sets. Both autocorrelation and cross-correlation ratios of UNet-fd agree excellently with unity over the ranges of scales $k < 0.3 \ h\text{Mpc}^{-1}$, indicating a successful recovery of the HI clustering and remarkable concordance between the reconstructed and the true HI temperature field. It is significant

that UNet-fd outperforms PCA, which, especially at large scale, suffers from residual foregrounds and signal loss. On the other hand, the method of Makinen *et al.* [55] uses PCA to subtract a few modes before feeding to the UNet. Although the method can potentially improve performance, the reduced HI signal that is eliminated by PCA is virtually impossible to be recovered. To show the level of signal loss subtracted by PCA, the green dash-dotted lines show the results obtained by subtracting the first two modes on the pure HI map. There are both underestimations of $R_{\text{auto}}(k)$ and $R_{\text{cross}}(k)$ due to signal loss, which is unexpected to be avoided by the PCA-UNet strategy. As a comparison, it is reassuring that our UNet-fd method is immune to the signal loss issue.

In order to gauge the accuracy of the reconstruction of the redshift-space distortions (RSD), we now compare $P(k_\perp, k_\parallel)$, as shown in the upper panels of Fig. 11. The left and right panels correspond to the autocorrelation and cross-correlation power spectrum, respectively. Black solid, red dashed, and blue dotted contours correspond to the true HI, the UNet-fd reconstruction, and the PCA-2 reconstruction, respectively. Note that the results are
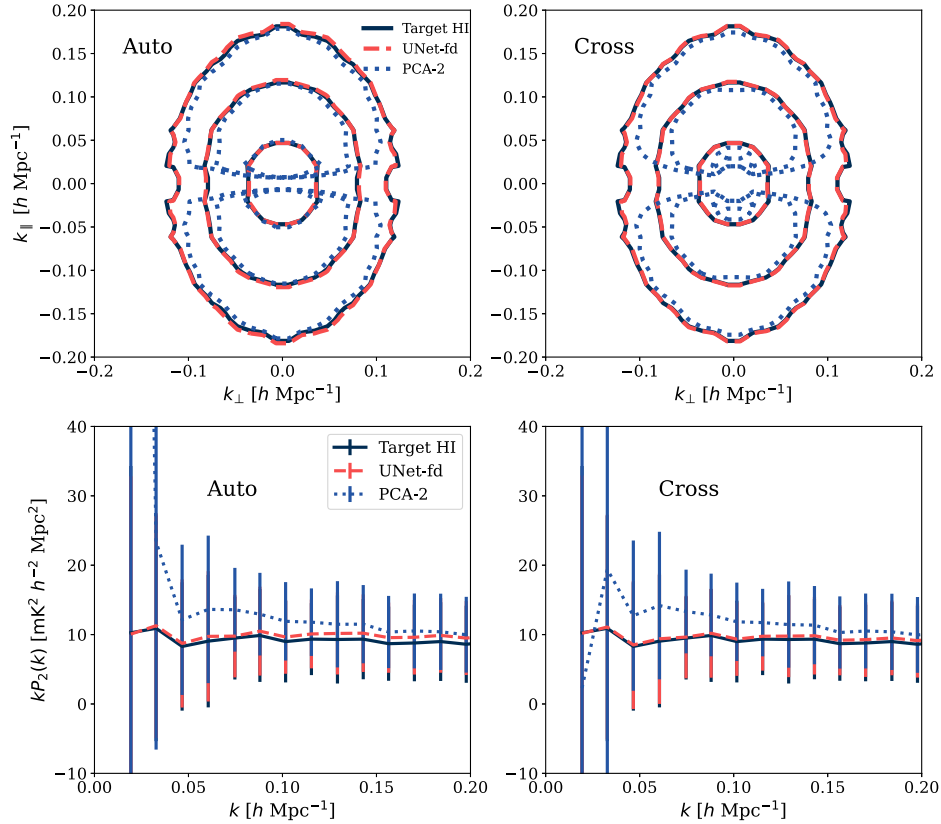


FIG. 11. Same as in Fig. 10, but for the 2D power spectra for Case I. Upper: comparison of $P(k_\perp, k_\parallel)$, where black solid, red dashed, and blue dotted contours correspond to the true HI, the UNet-fd reconstruction, and the PCA reconstruction, respectively. The contour level corresponds to $P(k) = 900, 300,$ and $180 \ h^{-3}\text{Mpc}^3$. The results are averaged from the 40 test sets. Lower: comparison of $P_2(k)$. Black solid, red dashed, and blue dotted contours correspond to the true HI, the UNet-fd reconstruction, and the PCA reconstruction, respectively. Error bars indicate the $\pm 1$ variance among the 40 test sets. The lower-left and lower-right panels correspond to autocorrelation and cross-correlation power spectrum, respectively.
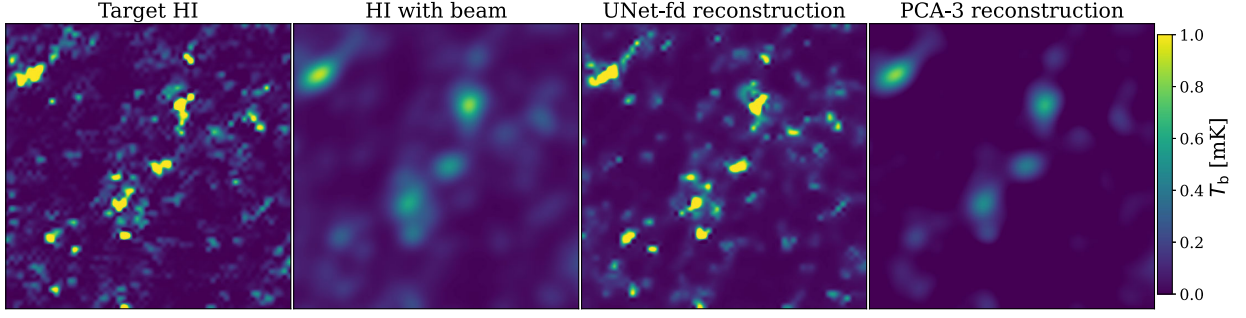
FIG. 12.   Comparison of the brightness temperature field for Case II (considering the beam effect). From left to right, we show the target HI, the target HI convolved with the beam [according to Eq. (5)], the UNet-fd reconstruction, and the PCA reconstruction, respectively. The target HI is selected to be same as the one in Fig. 8. As shown, our neural network not only removes strong foregrounds but also automatically performs beam deconvolution with high accuracy.

averaged from the 40 test cubes. In the case of UNet-fd, both autocorrelation and cross-correlation $P(k_\perp, k_\parallel)$ are clearly anisotropic with the "elongated" feature in the central region, revealing the impact of the Kaiser effect on large scales (small k values). A comparison with the true $P(k_\perp, k_\parallel)$ shows that the UNet-fd reconstructed RSDs are overall very successful. In contrast, the PCA reconstruction shows significant signal reduction at small $|k_\parallel|$, which is always removed during the PCA foreground subtraction [60]. However, such signal reduction is negligible in our method.

The Kaiser effect cause the quadrupole $P_2(k)$ to deviate significantly from zero, which is usually used to constrain the growth rate of the cosmic structure. The lower panels of Fig. 11 compare $P_2(k)$ for both autocorrelation (left panel) and cross-correlation (right panel). Black solid, red dashed, and blue dotted lines correspond to the true HI, the UNet-fd reconstruction, and the PCA-2 reconstruction, respectively. Error bars indicate the $\pm 1$ variance among the 40 test sets. As shown, the UNet-fd reconstructed $P_2(k)$ has deviations from zero, which is consistent with the target truth. Although there is a large uncertainty due to the cosmic variance in our small-volume samples, we can

still observe a discrepancy between the PCA reconstructed and the true $P_2(k)$.

As a result, while just taking into account the foreground-contaminated 21-cm signal, our UNet-fd method can reliably remove the foreground and consistently reconstruct the redshift-space brightness temperature field of the cosmic HI.

### B. Case II: Considering beam effect

It is shown that the foreground contamination can be removed efficiently depending on the feature of frequency correlation. However, the systematic beam effect reduces the correlation feature and causes nonsmooth foreground residual, which challenges the detection of the HI signal. In this subsection, we validate our method for Case II.

Figure 12 shows the visual comparison of the brightness temperature field. From left to right, the first panel shows the target HI field, which is the same as the one in Fig. 8. To demonstrate the effectiveness of the deconvolution, the second panel presents the target HI map convolved with the same beam as the corresponding foreground-contaminated map according to Eq. (5). For UNet-fd, the recovered temperature fluctuations nicely match the correspondence
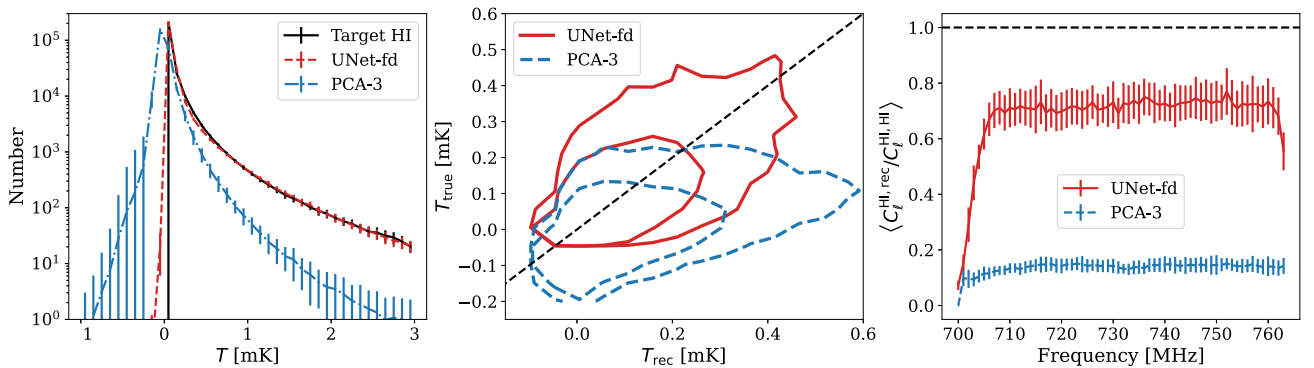


FIG. 13.   Same as in Fig. 9, but for testing the beam effect for Case II. In the middle panel, the two contours of each case encompass 90% and 95% of the grid cells, respectively.
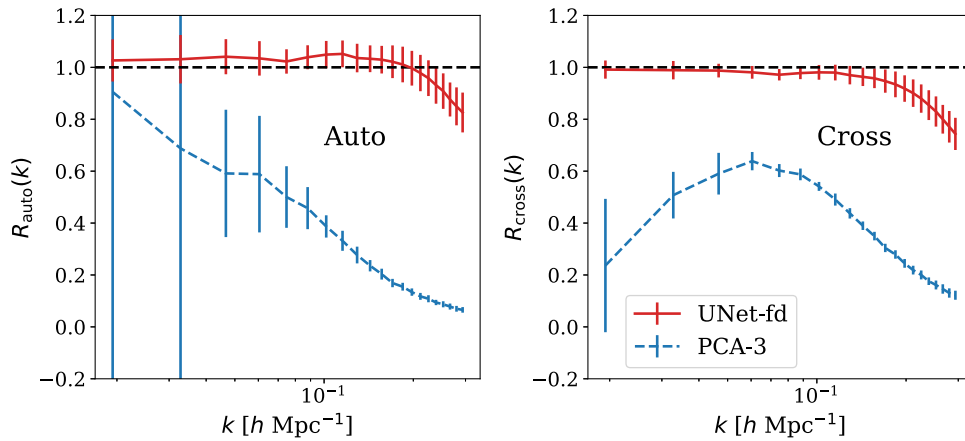
FIG. 14.   Same as in Fig. 10, but for Case II.

in the target field. Additionally, a noticeable deconvolution effect can be observed when comparing it to the second panel, indicating that the U-Net has learned to remove the beam effect. However, it has limited performance in predicting the detailed shape, suggesting that the beam effect is not perfectly eliminated. Comparatively, the PCA performs substantially worse, which always results in the signal being subtracted, especially in the beam case.

Figure 13 shows the temperature distribution (left panel), pixel-to-pixel temperature comparison (middle panel), and

the angular cross-correlation ratio as a function frequency channel (right panel). All panels are the same as those in Fig. 9 but for testing the additional beam effect. In the middle panel, the two contours of each case encompass 90% and 95% of the grid cells, respectively. The UNet-fd recovered temperature distribution still follows consistently with the target and shows an unbiased pixel-to-pixel relation. However, the beam effect causes a larger scattering in pixel-to-pixel relationships and a 30% reduction in angular cross-correlation ratios. The PCA reconstruction,
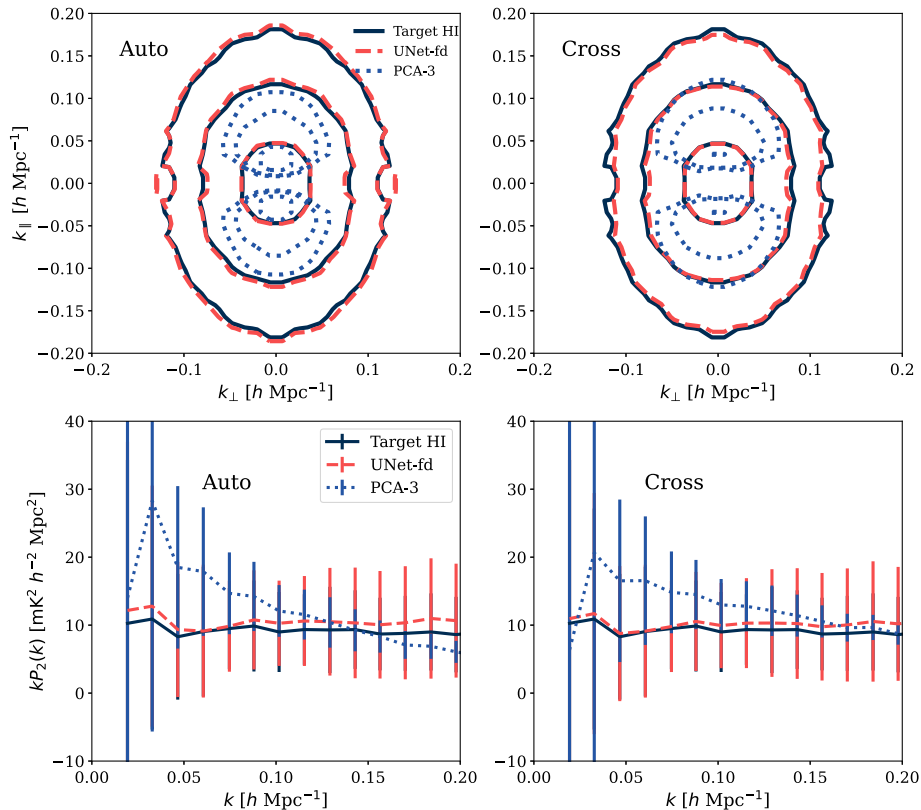


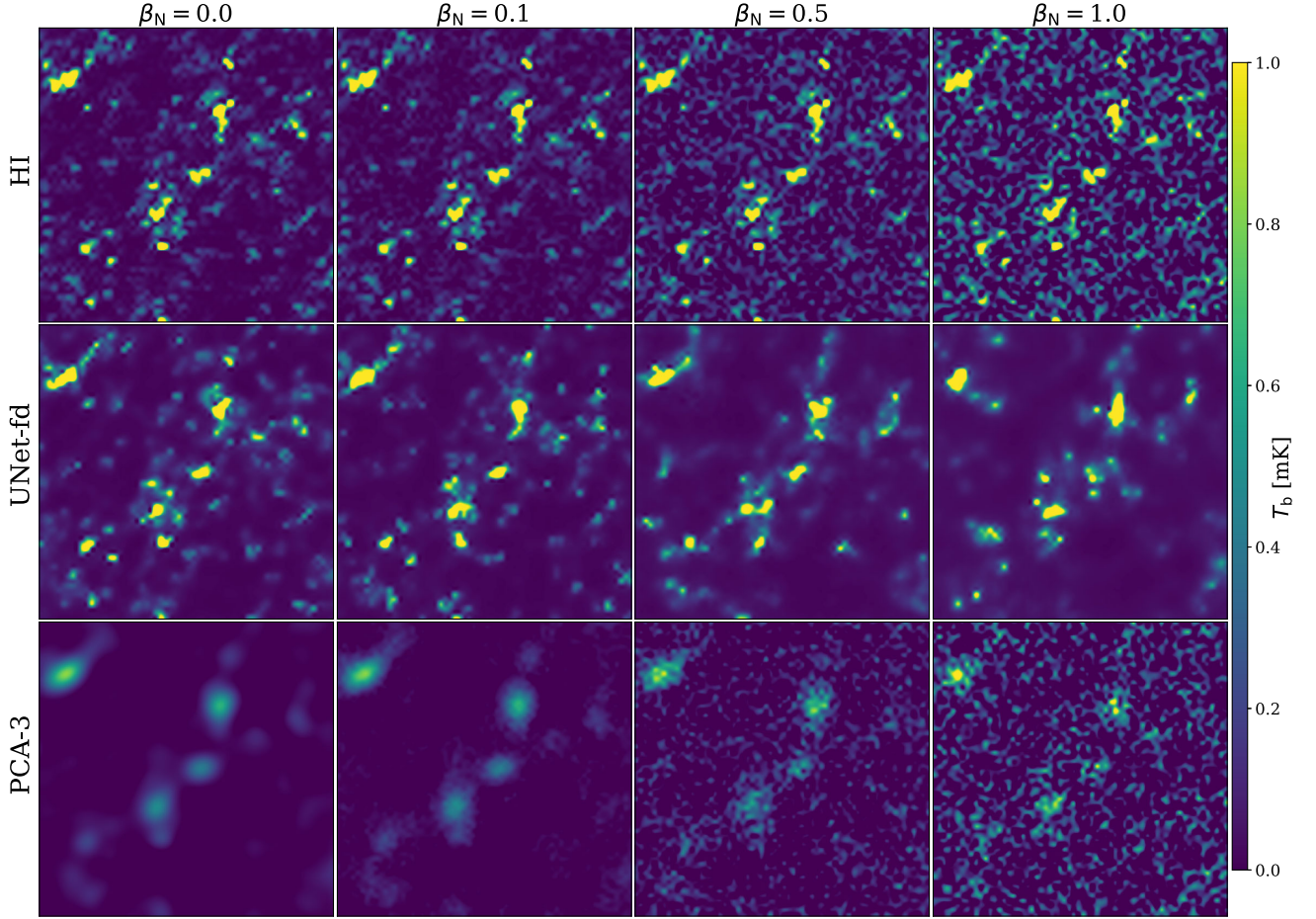FIG. 15.   Same as in Fig. 11, but for Case II.

FIG. 16. Comparison of the brightness temperature field for Case III in introducing thermal noise. Different columns correspond to different noise levels, as indicated at the top of each column. The first row presents the target HI with different levels of added noise (where the HI with $\beta_N = 0.0$ is the same target HI as in Figs. 8 and 12). The second and third rows correspond to the UNet-fd and PCA-3 reconstructions respectively, as indicated to the left of each row.

in comparison, has a drastically reduced temperature distribution, biased pixel-to-pixel relationships, and cross-correlation ratios underestimated by around 60%.

Figure 14 shows $R_{\text{auto}}(k)$ (left panel) and $R_{\text{cross}}(k)$ (right panel). Red solid and blue dashed lines correspond to UNet-fd and PCA reconstruction, respectively. In UNet-fd reconstruction, both values of $R_{\text{auto}}(k)$ and $R_{\text{cross}}(k)$ are consistent with unity at the $1\sigma$ level over the scales $k < 0.1\ h\text{Mpc}^{-1}$. The underestimation of $R_{\text{cross}}(k)$ at scale $k > 0.1\ h\text{Mpc}^{-1}$ is caused by the beam effect in comparison to the beam-free result of Fig. 10. Though, the overall reduction of the cross-correlation power spectrum at $k = 0.2\ h\text{Mpc}^{-1}$ and $0.3\ h\text{Mpc}^{-1}$ are at 10% and 30% levels, respectively. The PCA, in contrast, exhibits large fluctuations in $R_{\text{auto}}(k)$ and a reduction in $R_{\text{auto}}(k)$, indicating an uncleaned recovery with lingering residual foreground and signal loss.

To test the RSD reconstruction, Fig. 15 shows the comparisons of $P(k_{\perp}, k_{\parallel})$ (upper panels) and $P_2(k)$ (lower panels). All panels are the same as those in Fig. 11 but

for considering additional beam effect. In this case, both UNet-fd's autocorrelation and cross-correlation $P(k_{\perp}, k_{\parallel})$ are in good agreement with the target truth, except for a few discrepancies of auto-$P(k_{\perp}, k_{\parallel})$ at large $|k_{\parallel}|$. The RSDs are also recovered successfully with the anisotropic Kaiser features apparent. As expected, the quadrupole $P_2(k)$ deviates from zero and agrees with the target truth. For both the auto and cross-correlation, the PCA results show a highly atypical shape of $P(k_{\perp}, k_{\parallel})$ and unjustifiable amplitude of $P_2(k)$.

Therefore, even when taking into account the systematic beam effect, UNet-fd is still effective in removing the foregrounds.

### C. Case III: Introducing thermal noise

Along with foregrounds and beam effects, we also take into account the instrumental thermal white noise as the Case III. As described in Sec. II C, we are not referring to any specific experiment for HI but rather simulating various
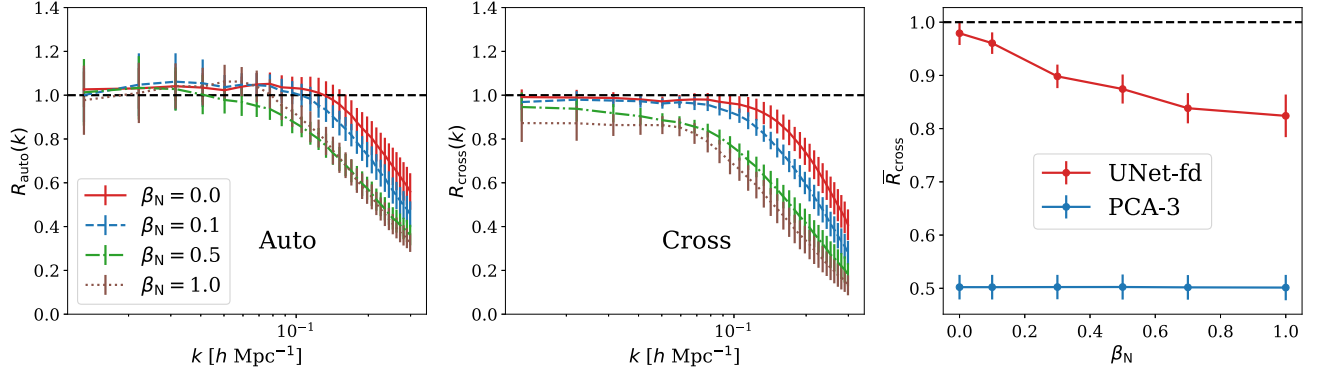
FIG. 17.    Results of $R_{\mathrm{auto}}(k)$ (left) and $R_{\mathrm{cross}}(k)$ (middle) for Case III are shown using the UNet-fd method, with varying noise levels as indicated. The values for $\beta_{\mathrm{ns}} = 0$ are the same as the UNet-fd results in Fig. 14. The right panel shows a comparison of the mean cross-correlation ratio, $\bar{R}_{\mathrm{cross}}(\beta_{\mathrm{N}})$, between UNet-fd (red) and PCA-3 (blue). The value of $\bar{R}_{\mathrm{cross}}$ is calculated by averaging $R_{\mathrm{cross}}(k)$ from the values over scales $0 < k < 0.1 \ h\mathrm{Mpc}^{-1}$.

levels of uncorrelated Gaussian noise. In this subsection, we verify how the noise affects our cleaning procedure, and look for a guideline threshold level for possible future configurations of intensity mapping experiments.

Figure 16 shows the visual comparison of the brightness temperature field. Different columns correspond to different noise levels: $\beta_{\mathrm{N}} = 0.0$, $\beta_{\mathrm{N}} = 0.1$, $\beta_{\mathrm{N}} = 0.5$, and $\beta_{\mathrm{N}} = 1.0$, as indicated at the top of each column. To demonstrate the effectiveness of noise elimination, the first row presents the target HI with different levels of added noise (where the HI with $\beta_{\mathrm{N}} = 0.0$ is the same target HI as in Figs. 8 and 12). The second and third rows correspond to the UNet-fd and PCA-3 reconstructions respectively, as indicated to the left of each row. Upon comparing the first two rows, it becomes evident that UNet-fd successfully restores the high-temperature structures while eliminating random noise. This highlights its ability to effectively separate the signal from Gaussian noise. The expected outcome from UNet is based on the distinct spatial distribution patterns of HI and noise. The HI demonstrates scale-dependent clustering, while the noise is dispersed in an uncorrelated random manner. However, when it comes to recovering the low-temperature structure, especially under high levels of noise, U-Net's capability to separate the signal is somewhat limited. Additionally, the noise variance of the difference maps is twice that of the original frequency maps, further compromising the accuracy of the U-Net reconstruction of the 21-cm signal. Nonetheless, the reconstructions obtained through PCA yield even poorer results, as the method is unable to effectively subtract the noise.

Figure 17 shows $R_{\mathrm{auto}}(k)$ (left panel) and $R_{\mathrm{cross}}(k)$ (middle panel) for the UNet-fd reconstructions. Different color lines are shown for different-level noise as indicated, where lines of $\beta_{\mathrm{N}} = 0$ are the same as the UNet-fd results in Fig. 14. The results of $R_{\mathrm{auto}}(k)$ are all near to unity at $1\sigma$ level over large scales, but have reductions at small scales (e.g. $k > 0.1 \ h\mathrm{Mpc}^{-1}$). For the cross-correlation case, it reveals underestimates throughout all scales.

In the right panel of Fig. 17, we show the mean cross-correlation ratio, $\bar{R}_{\mathrm{cross}}(\beta_{\mathrm{N}})$, for a comparison of UNet-fd (red) versus the PCA-3 (blue). We compute $\bar{R}_{\mathrm{cross}}$ by averaging $R_{\mathrm{cross}}(k)$ over scales $0 < k < 0.1 \ h\mathrm{Mpc}^{-1}$. The UNet-fd result shows an inverse relation between $\bar{R}_{\mathrm{cross}}$ and $\beta_{\mathrm{N}}$, providing a practical guideline. It shows that the mean cross-correlation ratio remains above 0.8 when the variance of the thermal noise is smaller than or equal to that of the HI signal. In this case, the UNet-fd method reaches a limitation, leading to a systematic loss of the 21-cm signal. On the other hand, the PCA results exhibit significantly lower cross-correlation ratios, with values being around 0.5. This highlights the superior performance of the UNet-fd method compared to PCA in terms of cross-correlation ratios.

## V. SUMMARY AND CONCLUSION

In this study, we propose the frequency-difference technique for pre-processing the data in U-Net deep learning for 21-cm foreground removal. We find that this technique dramatically suppresses the dynamic range of the foreground amplitudes by 2-3 orders of magnitude. This improvement significantly enhances the performance of U-Net in foreground removal. Additionally, it circumvents the uncertainty associated with the traditional PCA pre-processing method, which involves selecting the number of eigenmodes and avoids the 21-cm signal loss often associated with PCA. Our tests reveal that the frequency-difference technique and U-Net are primarily employed to eliminate smooth and non-smooth foreground components, respectively.

We first verify our method using the 21-cm temperature including HI and foregrounds only. By conducting a series of comparison tests, including image comparison, pixel-to-pixel temperature relation, temperature distribution, and power spectrum, we find that our method can reliably remove the foreground. Both autocorrelation and cross-correlation power spectrum ratios agree excellently with

unity over the ranges of scales $k < 0.3\ h\mathrm{Mpc}^{-1}$, indicating a successful recovery of the HI clustering and remarkable concordance between the reconstructed and the true HI temperature field. Further considering the systematic beam effects, our reconstruction has autocorrelation and cross-correlation power spectrum ratios consistently at the $1\sigma$ level over the scales $k < 0.1\ h\mathrm{Mpc}^{-1}$. Beam effects smear out the small-scale clustering with 10% and 30% reduction of the cross-correlation power spectrum at $k = 0.2\ h\mathrm{Mpc}^{-1}$ and $0.3\ h\mathrm{Mpc}^{-1}$, respectively. Moreover, the RSD effects are also reconstructed overall very successfully with the anisotropic Kaiser features apparent.

As a comparison, our method outperforms PCA, which has a drastically reduced temperature distribution, biased pixel-to-pixel temperature relationships, and cross-correlation ratios underestimated by around 60%. PCA also fails to reconstruct the RSD structure, which has $P(k_\perp, k_\parallel)$ with a highly atypical shape and $P_2(k)$ with unjustifiable amplitude.

We eventually investigate the impact of instrumental thermal noise on the cleaning procedure and find that, with the help of the U-Net network, our approach can separate the signal from the noise. By simulating various levels of Gaussian uncorrelated noise, we present an inverse relationship between the mean cross-correlation ratio $\bar{R}_\mathrm{cross}$ and the noise level $\beta_\mathrm{N}$. It suggests that if the variance level of the thermal noise is smaller than or equal to the HI signal, our method should be capable of obtaining a signal recovery with a cross-correlation ratio greater than approximately 0.8, surpassing the PCA result of 0.5. This information could be used as a reference for applying our method in practical experiments.

The large dynamic range of the foreground amplitude limits the deep learning method, which is now being applied in the 21-cm foreground removal. PCA preprocessing could be a solution, but it cannot solve the issue of HI signal loss. As a result, the frequency difference can considerably enhance network performance by reducing the amplitude range of the smooth foreground components and helping in the prevention of HI loss.

## APPENDIX: EXPLANATION ON USING UNet TO EXTRACT THE SINGLE-BAND HI FROM THE DIFFERENCE MAP

As an illustration, Fig. 18 shows the pure HI temperature map at $\nu_1 = 1130$ MHz (left panel), $\nu_2 = 1132$ MHz (middle panel), and their corresponding difference map, $T_\mathrm{HI}(\nu_1) - T_\mathrm{HI}(\nu_2)$ (right panel). The difference map retains a significant number of structures from both frequency bands, with most positive values corresponding to $\nu_1$ signals and most negative values corresponding to $\nu_2$ signals. It should be noted that the pure HI signals have a minimum value of zero in our dataset. Figure 19 shows the pixel-to-pixel temperature relation between $T_\mathrm{HI}(\nu_1)$ and $T_\mathrm{HI}(\nu_1) - T_\mathrm{HI}(\nu_2)$. The blue circles reveals a significant positive correlation between $T_\mathrm{HI}(\nu_1)$ and
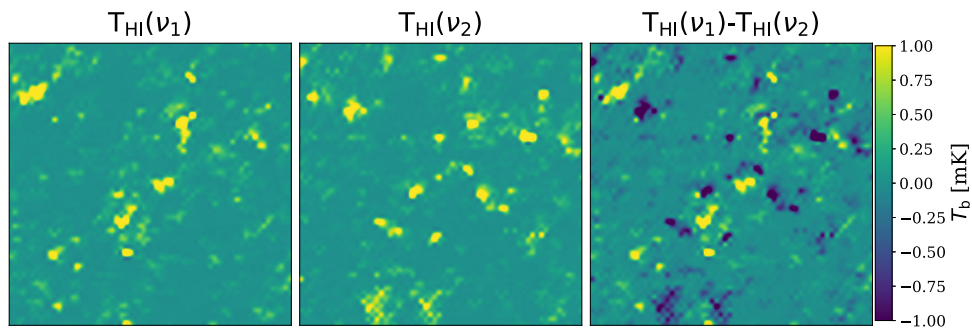


FIG. 18. The pure HI temperature map at $\nu_1 = 1130$ MHz (left panel), $\nu_2 = 1132$ MHz (middle panel), and their corresponding difference map, $T_\mathrm{HI}(\nu_1) - T_\mathrm{HI}(\nu_2)$ (right panel). The $\nu_1$ slice is taken to be same as the target HI in Fig. 8, which is randomly selected from the dataset.
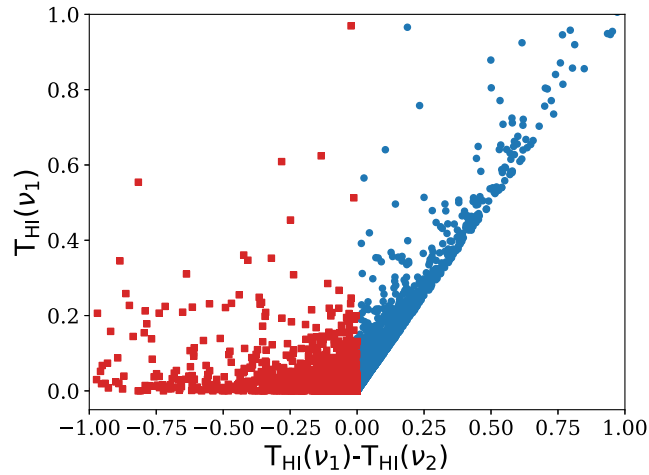
FIG. 19. Pixel-to-pixel temperature relation, $T_{\mathrm{HI}}(\nu_1)$ versus $T_{\mathrm{HI}}(\nu_1) - T_{\mathrm{HI}}(\nu_2)$. The pixels are selected from the slices shown in Fig. 18. The blue circles and red squares indicate positive and negative values on the $x$-axis, respectively.

$T_{\mathrm{HI}}(\nu_1) - T_{\mathrm{HI}}(\nu_2)$, indicating the pixels associated with the $T_{\mathrm{HI}}(\nu_1)$ map. In contrast, the red squares exhibit a much lower level of correlation, suggesting the pixels from the

$T_{\mathrm{HI}}(\nu_2)$ map. It is expected that such relationships will enable the UNet model to distinguish between $T_{\mathrm{HI}}(\nu_1)$ and $T_{\mathrm{HI}}(\nu_2)$, as well as extract the targeted single-band HI signals from the difference map. However, the presence of biased pixels within the blue circles and red squares is a result of mutual cancellation by the overlapping region in the difference map between $T_{\mathrm{HI}}(\nu_1)$ and $T_{\mathrm{HI}}(\nu_2)$. This may introduce some error when attempting to recover the targeted HI signals.

From another perspective, if the signals in the frequency-difference maps are purely random and lack any specific structure or features, such as white noise, it would be impossible to recover either of the signal maps from their difference maps using any method, including AI. This is because there are far more degrees of freedom that need to be recovered than the number of dimensions in the data. However, in our case, the HI signal does exhibit some structure, and the number of degrees of freedom for major features (eigenmodes) contained in the signal map is less than the dimension of the frequency-difference map. Therefore, this is the reason why we were able to effectively recover the HI maps for each frequency from the frequency-difference maps.

[1] D. Scott and M. J. Rees, The 21-cm line at high redshift: A diagnostic for the origin of large scale structure, Mon. Not. R. Astron. Soc. **247**, 510 (1990).

[2] A. Loeb and M. Zaldarriaga, Measuring the small-scale power spectrum of cosmic density fluctuations through 21 cm tomography prior to the epoch of structure formation, Phys. Rev. Lett. **92**, 211301 (2004).

[3] A. Lidz, O. Zahn, M. McQuinn, M. Zaldarriaga, and L. Hernquist, Detecting the rise and fall of 21 cm fluctuations with the Murchison widefield array, Astrophys. J. **680**, 962 (2008).

[4] P. Christian and A. Loeb, Measuring the x-ray background in the reionization era with first generation 21 cm experiments, J. Cosmol. Astropart. Phys. 09 (2013) 014.

[5] T. G. Sarkar, K. K. Datta, A. K. Pal, T. R. Choudhury, and S. Bharadwaj, Redshifted HI 21-cm signal from the post-reionization epoch: Cross-correlations with other cosmological probes, J. Astrophys. Astron. **37**, 26 (2016).

[6] S. Choudhuri, S. Bharadwaj, N. Roy, A. Ghosh, and S. S. Ali, Tapering the sky response for angular power spectrum estimation from low-frequency radio-interferometric data, Mon. Not. R. Astron. Soc. **459**, 151 (2016).

[7] I. P. Carucci, F. Villaescusa-Navarro, and M. Viel, The cross-correlation between 21 cm intensity mapping maps and the Lyα forest in the post-reionization era, J. Cosmol. Astropart. Phys. 04 (2017) 001.

[8] V. R. Marthi, S. Chatterjee, J. N. Chengalur, and S. Bharadwaj, Simulated predictions for HI at z = 3.35 with

the Ooty wide field array—I. Instrument and the foregrounds, Mon. Not. R. Astron. Soc. **471**, 3112 (2017).

[9] S. Bharadwaj, B. B. Nath, and S. K. Sethi, Using HI to probe large scale structures at z ~ 3, J. Astrophys. Astron. **22**, 21 (2001).

[10] R. A. Battye, R. D. Davies, and J. Weller, Neutral hydrogen surveys for high-redshift galaxy clusters and protoclusters, Mon. Not. R. Astron. Soc. **355**, 1339 (2004).

[11] T.-C. Chang, U.-L. Pen, J. B. Peterson, and P. McDonald, Baryon acoustic oscillation intensity mapping of dark energy, Phys. Rev. Lett. **100**, 091303 (2008).

[12] J. B. Peterson et al., 21-cm intensity mapping, in Astro2010: The Astronomy and Astrophysics Decadal Survey (2009), Vol. 2010, p. 234, arXiv:0902.3091.

[13] J. S. B. Wyithe and A. Loeb, The 21-cm power spectrum after reionization, Mon. Not. R. Astron. Soc. **397**, 1926 (2009).

[14] H.-J. Seo, S. Dodelson, J. Marriner, D. Mcginnis, A. Stebbins, C. Stoughton, and A. Vallinotto, A ground-based 21 cm baryon acoustic oscillation survey, Astrophys. J. **721**, 164 (2010).

[15] A. Liu and J. R. Shaw, Data analysis for precision 21 cm cosmology, Publ. Astron. Soc. Pac. **132**, 062001 (2020).

[16] M. Amiri et al. (CHIME Collaboration), Detection of cosmological 21 cm emission with the Canadian hydrogen intensity mapping experiment, Astrophys. J. **947**, 16 (2023).

[17] M. Amiri et al. (CHIME Collaboration), An overview of CHIME, the Canadian hydrogen intensity mapping experiment, Astrophys. J. Suppl. Ser. **261**, 29 (2022).

[18] X. Chen, Radio detection of dark energy—the Tianlai project, Sci. Sin. Phys. Mech. Astron. **41,** 1358 (2011).

[19] Y. Xu, X. Wang, and X. Chen, Forecasts on the dark energy and primordial non-Gaussianity observations with the Tianlai cylinder array, Astrophys. J. **798,** 40 (2015).

[20] F. Wu *et al.*, The Tianlai dish pathfinder array: Design, operation, and performance of a prototype transit radio interferometer, Mon. Not. R. Astron. Soc. **506,** 3455 (2021).

[21] O. Perdereau *et al.*, The Tianlai dish array low-z surveys forecasts, Mon. Not. R. Astron. Soc. **517,** 4637 (2022).

[22] R. A. Battye, I. W. A. Browne, C. Dickinson, G. Heron, B. Maffei, and A. Pourtsidou, HI intensity mapping: A single dish approach, Mon. Not. R. Astron. Soc. **434,** 1239 (2013).

[23] E. Abdalla *et al.*, The BINGO project. I. Baryon acoustic oscillations from integrated neutral gas observations, Astron. Astrophys. **664,** A14 (2022).

[24] M. Santos *et al.*, A large sky survey with MeerKAT, in *MeerKAT Science: On the Pathway to the SKA* (Proceeding of Science, 2016), p. 32, arXiv:1709.06099.

[25] J. Wang, M. G. Santos, P. Bull, K. Grainge, S. Cunnington, J. Fonseca, M. O. Irfan, Y. Li, A. Pourtsidou, P. S. Soares, M. Spinelli, G. Bernardi, and B. Engelbrecht, HI intensity mapping with MeerKAT: Calibration pipeline for multidish autocorrelation observations, Mon. Not. R. Astron. Soc. **505,** 3698 (2021).

[26] L. B. Newburgh *et al.*, HIRAX: A probe of dark energy and radio transients, in *Proceedings of the SPIE*, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 9906 (2016), p. 99065X, 10.1117/12.2234286.

[27] M. Santos, P. Bull, D. Alonso, S. Camera, P. Ferreira, G. Bernardi, R. Maartens, M. Viel, F. Villaescusa-Navarro, F. B. Abdalla, M. Jarvis, R. B. Metcalf, A. Pourtsidou, and L. Wolz, Cosmology from a SKA HI intensity mapping survey, Proc. Sci. AASKA14 (**2015**) 019 [arXiv:1501.03989].

[28] T.-C. Chang, U.-L. Pen, K. Bandura, and J. B. Peterson, An intensity map of hydrogen 21-cm emission at redshift $z \approx 0.8$, Nature (London) **466,** 463 (2010).

[29] K. W. Masui, E. R. Switzer, N. Banavar, K. Bandura, C. Blake, L. M. Calin, T. C. Chang, X. Chen, Y. C. Li, Y. W. Liao, A. Natarajan, U. L. Pen, J. B. Peterson, J. R. Shaw, and T. C. Voytek, Measurement of 21 cm brightness fluctuations at $z \approx 0.8$ in cross-correlation, Astrophys. J. Lett. **763,** L20 (2013).

[30] E. R. Switzer, K. W. Masui, K. Bandura, L. M. Calin, T. C. Chang, X. L. Chen, Y. C. Li, Y. W. Liao, A. Natarajan, U. L. Pen, J. B. Peterson, J. R. Shaw, and T. C. Voytek, Determination of $z \approx 0.8$ neutral hydrogen fluctuations using the 21 cm intensity mapping autocorrelation, Mon. Not. R. Astron. Soc. **434,** L46 (2013).

[31] C. J. Anderson, N. J. Luciw, Y. C. Li, C. Y. Kuo, J. Yadav, K. W. Masui, T. C. Chang, X. Chen, N. Oppermann, Y. W. Liao, U. L. Pen, D. C. Price, L. Staveley-Smith, E. R. Switzer, P. T. Timbie, and L. Wolz, Low-amplitude clustering in low-redshift 21-cm intensity maps cross-correlated with 2dF galaxy densities, Mon. Not. R. Astron. Soc. **476,** 3382 (2018).

[32] L. Wolz *et al.*, HI constraints from the cross-correlation of eBOSS galaxies and Green Bank Telescope intensity maps, Mon. Not. R. Astron. Soc. **510,** 3495 (2022).

[33] S. Cunnington, Y. Li, M. G. Santos, J. Wang, I. P. Carucci, M. O. Irfan, A. Pourtsidou, M. Spinelli, L. Wolz, P. S. Soares, C. Blake, P. Bull, B. Engelbrecht, J. Fonseca, K. Grainge, and Y.-Z. Ma, HI intensity mapping with Meer-KAT: Power spectrum detection in cross-correlation with WiggleZ galaxies, Mon. Not. R. Astron. Soc. **518,** 6262 (2023).

[34] S. Paul, M. G. Santos, Z. Chen, and L. Wolz, A first detection of neutral hydrogen intensity mapping on Mpc scales at $z \approx 0.32$ and $z \approx 0.44$, arXiv:2301.11943.

[35] T. Di Matteo, R. Perna, T. Abel, and M. J. Rees, Radio foregrounds for the 21 centimeter tomography of the neutral intergalactic medium at high redshifts, Astrophys. J. **564,** 576 (2002).

[36] S. P. Oh and K. J. Mack, Foregrounds for 21-cm observations of neutral gas at high redshift, Mon. Not. R. Astron. Soc. **346,** 871 (2003).

[37] M. G. Santos, A. Cooray, and L. Knox, Multifrequency analysis of 21 centimeter fluctuations from the era of reionization, Astrophys. J. **625,** 575 (2005).

[38] A. de Oliveira-Costa, M. Tegmark, B. M. Gaensler, J. Jonas, T. L. Landecker, and P. Reich, A model of diffuse Galactic radio emission from 10 MHz to 100 GHz, Mon. Not. R. Astron. Soc. **388,** 247 (2008).

[39] A. Liu and J. R. Shaw, Data analysis for precision 21 cm cosmology, Publ. Astron. Soc. Pac. **132,** 062001 (2020).

[40] A. Liu and M. Tegmark, A method for 21 cm power spectrum estimation in the presence of foregrounds, Phys. Rev. D **83,** 103006 (2011).

[41] E. R. Switzer, T.-C. Chang, K. W. Masui, U.-L. Pen, and T. C. Voytek, Interpreting the unresolved intensity of cosmologically redshifted line radiation, Astrophys. J. **815,** 51 (2015).

[42] S. Zuo, X. Chen, and Y. Mao, A semiblind PCA-based foreground subtraction method for 21 cm intensity mapping, Astrophys. J. **945,** 38 (2023).

[43] E. Chapman, F. B. Abdalla, G. Harker, V. Jelić, P. Labropoulos, S. Zaroubi, M. A. Brentjens, A. G. de Bruyn, and L. V. E. Koopmans, Foreground removal using FASTICA: A showcase of LOFAR-EoR, Mon. Not. R. Astron. Soc. **423,** 2518 (2012).

[44] S. Cunnington, M. O. Irfan, I. P. Carucci, A. Pourtsidou, and J. Bobin, 21-cm foregrounds and polarization leakage: Cleaning and mitigation strategies, Mon. Not. R. Astron. Soc. **504,** 208 (2021).

[45] E. Chapman, F. B. Abdalla, J. Bobin, J. L. Starck, G. Harker, V. Jelić, P. Labropoulos, S. Zaroubi, M. A. Brentjens, A. G. de Bruyn, and L. V. E. Koopmans, The scale of the problem: Recovering images of reionization with generalized morphological component analysis, Mon. Not. R. Astron. Soc. **429,** 165 (2013).

[46] L. C. Olivari, M. Remazeilles, and C. Dickinson, Extracting HI cosmological signal with generalized needlet internal linear combination, Mon. Not. R. Astron. Soc. **456,** 2749 (2016).

[47] A. Marins, F. B. Abdalla, K. S. F. Fornazier, E. Abdalla, L. H. F. Assis, M. Remazeilles, C. Alexandre Wuensche, L. Barosi, A. R. Queiroz, T. Villela, B. Wang, C. Feng, R. Landim, V. Liccardo, C. P. Novaes, L. Santos, M. V. dos Santos, and J. Zhang, Foreground removal and 21 cm signal

estimates: Comparing different blind methods for the BINGO Telescope, arXiv:2209.11701.

[48] N. Kern and A. Liu, Gaussian process foreground subtraction and power spectrum estimation for 21 cm cosmology, Mon. Not. R. Astron. Soc. **501,** 1463 (2021).

[49] J. R. Shaw, K. Sigurdson, U.-L. Pen, A. Stebbins, and M. Sitwell, All-sky interferometry with spherical harmonic transit telescopes, Astrophys. J. **781,** 57 (2014).

[50] M. F. Morales, B. Hazelton, I. Sullivan, and A. Beardsley, Four fundamental foreground power spectrum shapes for 21 cm cosmology observations, Astrophys. J. **752,** 137 (2012).

[51] A. Ghosh, L. V. E. Koopmans, E. Chapman, and V. Jelić, A Bayesian analysis of redshifted 21-cm HI signal and foregrounds: Simulations for LOFAR, Mon. Not. R. Astron. Soc. **452,** 1587 (2015).

[52] L. Zhang, E. F. Bunn, A. Karakci, A. Korotkov, P. M. Sutter, P. T. Timbie, G. S. Tucker, and B. D. Wandelt, Bayesian semi-blind component separation for foreground removal in interferometric 21 cm observations, Astrophys. J. Suppl. Ser. **222,** 3 (2016).

[53] P. H. Sims and J. C. Pober, Joint estimation of the epoch of reionization power spectrum and foregrounds, Mon. Not. R. Astron. Soc. **488,** 2904 (2019).

[54] J. Burba, P. H. Sims, and J. C. Pober, All-sky modelling requirements for Bayesian 21 cm power spectrum estimation with BAYESEOR, Mon. Not. R. Astron. Soc. **520,** 4443 (2023).

[55] T. L. Makinen, L. Lancaster, F. Villaescusa-Navarro, P. Melchior, S. Ho, L. Perreault-Levasseur, and D. N. Spergel, Deep21: A deep learning method for 21 cm foreground removal, J. Cosmol. Astropart. Phys. 04 (2021) 081.

[56] S. Ni, Y. Li, L.-Y. Gao, and X. Zhang, Eliminating primary beam effect in foreground subtraction of neutral hydrogen intensity mapping survey with deep learning, Astrophys. J. **934,** 83 (2022).

[57] X. Wang, M. Tegmark, M. G. Santos, and L. Knox, 21 cm tomography with foregrounds, Astrophys. J. **650,** 529 (2006).

[58] L. Gleser, A. Nusser, and A. J. Benson, Decontamination of cosmological 21-cm maps, Mon. Not. R. Astron. Soc. **391,** 383 (2008).

[59] V. Jelić, S. Zaroubi, P. Labropoulos, R. M. Thomas, G. Bernardi, M. A. Brentjens, A. G. de Bruyn, B. Ciardi, G. Harker, L. V. E. Koopmans, V. N. Pandey, J. Schaye, and S. Yatawatta, Foreground simulations for the LOFAR-epoch of reionization experiment, Mon. Not. R. Astron. Soc. **389,** 1319 (2008).

[60] E. R. Switzer, T. C. Chang, K. W. Masui, U. L. Pen, and T. C. Voytek, Interpreting the unresolved intensity of cosmologically redshifted line radiation, Astrophys. J. **815,** 51 (2015).

[61] F. Shi, Y.-S. Song, J. Asorey, D. Parkinson, K. Ahn, J. Yao, L. Zhang, and S. Zuo, HIR4: Cosmological signatures imprinted on the cross-correlation between a 21-cm map and galaxy clustering, Mon. Not. R. Astron. Soc. **499,** 4613 (2020).

[62] J. Asorey, D. Parkinson, F. Shi, Y.-S. Song, K. Ahn, J. Kim, J. Yao, L. Zhang, and S. Zuo, HIR4: Cosmology from a simulated neutral hydrogen full sky using Horizon Run 4, Mon. Not. R. Astron. Soc. **495,** 1788 (2020).

[63] S. Cunnington, L. Wolz, P. Bull, I. P. Carucci, K. Grainge, M. O. Irfan, Y. Li, A. Pourtsidou, M. G. Santos, M. Spinelli, and J. Wang, The foreground transfer function for HI intensity mapping signal reconstruction: MeerKLASS and precision cosmology applications, Mon. Not. R. Astron. Soc. **523,** 2453 (2023).

[64] D. Alonso, P. G. Ferreira, and M. G. Santos, Fast simulations for intensity mapping experiments, Mon. Not. R. Astron. Soc. **444,** 3183 (2014).

[65] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann, HEALPix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere, Astrophys. J. **622,** 759 (2005).

[66] P. Bull, P. G. Ferreira, P. Patel, and M. G. Santos, Late-time cosmology with 21 cm intensity mapping experiments, Astrophys. J. **803,** 21 (2015).

[67] C. G. T. Haslam, C. J. Salter, H. Stoffel, and W. E. Wilson, A 408-MHZ all-sky continuum survey. II. The ATLAS of contour maps, Astron. Astrophys. Suppl. Ser. **47,** 1 (1982).

[68] J. Delabrouille et al., The pre-launch Planck sky model: A model of sky emission at submillimetre to centimetre wavelengths, Astron. Astrophys. **553,** A96 (2013).

[69] M. Remazeilles, C. Dickinson, A. J. Banday, M. A. Bigot-Sazy, and T. Ghosh, An improved source-subtracted and destriped 408-MHz all-sky map, Mon. Not. R. Astron. Soc. **451,** 4311 (2015).

[70] G. B. Rybicki and A. P. Lightman, Radiative Processes in Astrophysics (Wiley, 1986).

[71] A. Waelkens, T. Jaffe, M. Reinecke, F. S. Kitaura, and T. A. Enßlin, Simulating polarized Galactic synchrotron emission at all frequencies. The Hammurabi code, Astron. Astrophys. **495,** 697 (2009).

[72] D. Alonso, J. Sanchez, and A. Slosar, A unified pseudo-$C_\ell$ framework, Mon. Not. R. Astron. Soc. **484,** 4127 (2019).

[73] S. De and H. Tashiro, Galactic Faraday rotation effect on polarization of 21 cm lines from the epoch of reionization, Phys. Rev. D **89,** 123002 (2014).

[74] S. D. Matshawule, M. Spinelli, M. G. Santos, and S. Ngobese, HI intensity mapping with MeerKAT: Primary beam effects on foreground cleaning, Mon. Not. R. Astron. Soc. **506,** 5075 (2021).

[75] O. Ronneberger, P. Fischer, and T. Brox, U-Net: Convolutional networks for biomedical image segmentation, arXiv:1505.04597.

[76] I. Loshchilov and F. Hutter, Decoupled weight decay regularization, arXiv:1711.05101.

[77] A. Liu, A. R. Parsons, and C. M. Trott, Epoch of reionization window. I. Mathematical formalism, Phys. Rev. D **90,** 023018 (2014).

[78] F. Villaescusa-Navarro, PYLIANS: Python libraries for the analysis of numerical simulations, Astrophysics Source Code Library, record ascl:1811.008 (2018).

[79] http://www.blsc.cn/.