MNRAS **000**, 000–000 (2021) Preprint 9 February 2022 Compiled using MNRAS LATEX style file v3.0

The Tianlai dish array low-z surveys forecasts

Olivier Perdereau[1⋆], Réza Ansari[1], Albert Stebbins[4], Peter T. Timbie[2], Xuelei Chen[3,5,6], Fengquan Wu[3], Jixia Li[3,5], John P. Marriner[4], Gregory S. Tucker[12], Zhiping Chen[8], Yanping Cong[3,5], Santanu Das[2], Qizhi Huang[3,5], Yichao Li[9], Tao Liu [8], Yingfeng Liu[3,5], Christophe Magneville[14], Chenhui Niu[3], Calvin Osinga[2], Trevor M. Oxholm[2], Jeffrey B. Peterson[10], Anh Phan[2], Huli Shi[3], Gage Siebert[2], Shijie Sun[3,5], Haijun Tian[11], Qunxiong Wang[11], Rongli Wang[8], Yougang Wang[3], Yanlin Wu[6], Yidong Xu[3], Kaifeng Yu[3,5], Zijie Yu[3,5], Jiao Zhang[13], Juyong Zhang[8], Jialu Zhu[8], Shifan Zuo[3,5,6]

[1] *IJCLab, University of Paris-Saclay, CNRS/IN2P3, Université Paris-Saclay, Orsay, France*
[2] *Department of Physics, University of Wisconsin Madison, 1150 University Ave, Madison WI 53703, USA*
[3] *National Astronomical Observatory, Chinese Academy of Science,20A Datun Road, Beijing 100101, P. R. China*
[4] *Fermi National Accelerator Laboratory, P.O. Box 500, Batavia IL 60510-5011, USA*
[5] *University of Chinese Academy of Sciences Beijing 100049, P. R. China*
[6] *Center of High Energy Physics, Peking University, Beijing 100871, P. R. China*
[7] *Department of Astronomy and Tsinghua Center for Astrophysics, Tsinghua University, Beijing 100084, P.R.China*
[8] *Hangzhou Dianzi University, 115 Wenyi Rd., Hangzhou 310018, P. R. China*
[9] *College of Sciences, Northeastern University. Shenyang Liaoning, P. R. China.*
[10] *Department of Physics, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA*
[11] *China Three Gorges University, Yichang 443002, P. R. China*
[12] *Department of Physics, Brown University, 182 Hope St., Providence, RI 02912, USA*
[13] *College of Physics and Electronic Engineering, Shanxi University, Taiyuan, Shanxi 030006, P. R. China*
[14] *CEA, DSM/IRFU, Centre d'Etudes de Saclay, 91191 Gif-sur-Yvette, France*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

We present the science case of the surveys planned for the Tianlai dish array interferometer, covering the very low redshift range $z \lesssim 0.1$ by tuning the instrument analog electronic system to low frequencies, covering the $[1300, 1400]$ MHz range. Realistic simulations of the surveys, starting from generation of mock visibility data according to the surveys' strategy, then map reconstruction followed by a simple foreground subtraction have been performed. We show that a rather deep survey toward the North Celestial Pole (NCP), covering an area of $\sim 100\text{deg}^2$ over a year, would reach a sensitivity a few mK, and would be marginally impacted by the mode-mixing given the interferometer configuration. Tianlai would then be able to detect nearby, massive H_I clumps, as well as observe a signnificant cross-correlation signal at more than XX-sigma with NCCS optical galaxies. We have also studied the performance of a mid-latitude survey, covering several thousand squared degrees, overlapping the SDSS main survey footprint. Despite a higher noise level as well as significant distortions due to mode mixing, Tianlai would be able to detect the cross-correlation of the 21cm intensity mapping signal with the SDSS low-z spectroscopic galaxy sample. Thanks to these signals within reach, it would be possible to assess precisely the impact of instrument and survey imperfection, such as calibration uncertainties, lack of precise beam knowledge and correlated noise on survey sensitivity and mode mixing.

Key words: galaxies: evolution – large-scale structure – 21-cm

1 INTRODUCTION

21cm Intensity Mapping (IM) is a promising technique to map the cosmological large scale matter distribution

⋆ E-mail: olivier.perdereau@ijclab.in2p3.fr

through the observation of 21cm radio emission/absorption of neutral hydrogen gas (H_I) , while not requiring the detection of individual sources (see e.g (Bharadwaj et al. 2001; Battye et al. 2004)) and has been largely explored in the context of the search for EoR (Epoch of Reionisation) signal (see (Pritchard & Loeb 2008; Morales & Wyithe 2010) for example). Subsequently, it was suggested that 21cm Intensity Mapping surveys could be used to constrain Dark Energy through the measurement of the BAO scale (Chang et al. 2008; Seo et al. 2010; Ansari et al. 2012). Such surveys require instruments with large instantaneous bandwidth and field of view and several groups have built dense interferometric arrays to explore IM, such as CHIME (Bandura et al. 2014) or Tianlai (Chen 2012) in the last decade. Smaller instruments such as PAON4 (Ansari et al. 2020) or BMX (O'Connor et al. 2020) have also been built to explore specific technical aspects of theses arrays, as well as transit mode operation and calibration. CHIME has proved to be a powerful radio burst and pulsar observation machine (The CHIME/FRB Collaboration et al. 2021) and has motivated the design and construction of larger, dish-based, dense interferometric arrays, HIRAX (Newburgh et al. 2016) and CHORD (Vanderlinde et al. 2019). Should we mention HERA, LOFAR, MWA . . . ?

Tianlai is an international collaboration led by NAOC which has built and operates two radio-interferometers dedicated to the 21cm Intensity Mapping since 2015 (Das et al. 2018). A first instrument is composed of three cylindrical reflectors, equiped with a total of 96 dual polarisation feeds (Li et al. 2020) while the second instrument, the Tianlai Dish Pathfinder Array (hereafter T16DPA) features 16, 6 meter diameter, on-axis dishes, equipped with dual polarisation feeds, and arranged in a near hexagonal configuration. The two instruments are located in a radio quiet site in Hongliuxia, Balikun county, in the Xinjiang autonomous region, in northwest China. The two arrays have been observing in the frequency band $[700, 800]$ MHz, corresponding to the redshift range $z \sim [0.775, 1.029]$ and we recently reported on the various aspects of the operation and performances of the Tianlai Dish Pathfinder Array (Wu et al. 2021).

Detection of cosmological H_I signal and the science reach of large instruments to constrain cosmological model and the dark energy equation of state trough IM surveys covering the redshift range $z \lesssim 3 - 6$ has been extensively explored for large dedicated instruments (e.g. Bull et al. (2015) and Cosmic Visions 21 cm Collaboration et al. (2018)), with SKA (Villaescusa-Navarro et al. 2017), or with FAST (Smoot & Debono 2017).

Cross-correlation of 21cm intenisty maps have detected using observations with a large single dish instrument, the Green Bank Telescope (GBT) in 2010 (Chang et al. 2010) and more recently, using eBOSS galaxies (Wolz et al. 2022). CHIME has also published the first detection of a cross correlation of cosmological 21cm signal with eBOSS ELG and LRG galaxies, and quasars, using intensity maps obtained from interferometric observations, around $z \simeq 1$ in 2022 (CHIME Collaboration et al. 2022).

However, the two Tianlai pathfinder instruments, in particular the Dish array, are too small to be sensitive to the cosmological 21cm signal around $z \sim 1$. In this paper, we study the extragalactic H_I signals that could be detected by the Tianlai Dish Array by tuning its frequency band to very low redshifts ($z \lesssim 0.1$), through a detailed simulation of the reconstructed signal, taking into account the instrument response and survey strategy. Such an H_I signal within reach of the instrument would make it possible to assess precisely the instrument and data analysis performance regarding key issues such as gain, phase and band pass calibration, impact of instrument noise, beam and array configuration knowledge on the reconstructed 3D maps and level of residuals after foreground subtraction.

We present an overview of the science targets of the Tianlai Dish Array low-redshift surveys in section 2, while the simulation and analysis method common to the different science cases are discussed in section 3, as well the expected survey sensitivities. Possible direct detection of nearby H_I clumps or galaxies is presented in section 4, and the prospects of detecting the Large Scale structures at low redshifts ($z \lesssim 0.1$) in cross correlation with the SDSS and NCCS optical galaxy surveys is discussed in section 5. Our findings are summarised and further discussed in the last section 6.

Remarks:

- Consider cross-correlation with ALFALFA or FAST HI survey , need survey at lower latitudes to have overlap with theses surveys (Peter)
- There are frequency bands unusable due to strong RFI (from satellites) , around 1380 MHz for example - We should blank these frequency bands which will decrease the statistical significance (Olivier)
- For section 3, evaluate the impact of going from analytical smooth beams to realistic beams from simulations - Peter hopes to have the computed beams soon
- Check whether the stripes observed by SDSS at the highest declinations (80 deg) could be a target area (Albert)

2 LOW REDSHIFT SURVEYS

The Tianlai dish array reflectors are equipped with feeds having frequency bandwidth much larger than the instantaneous 100 MHz bandwidth of the digitisation and correlator system. The instrument observation band is defined by the analog RF filters and this filter system can easily be modified to change the instrument frequency band. It is planned to tune the Tianlai Dish Pathfinder Array (T16DPA) frequency band to observe at low redshift, $[1320, 1420]$ MHz or $[1300, 1400]$ MHz, corresponding to the redshift range $0. \lesssim z \lesssim 0.076$ or $0.014 \lesssim z \lesssim 0.092$.

In addition, T16DPA dishes are fully steerable and equipped with an electronic pointing system. This allows targeted observations, although in transit mode, using T16DPA, in order to increase the integration time, thus the sensitivity, toward specific sky area. The North Celestial Pole (NCP), accessible to the Tianlai Dish Array represent several advantages and is an optimal target to carry deep, high sensitivity observations, as already suggested in (Zhang et al. 2016). A preliminary analysis of long duration observations of the NCP at $z \sim 1$ with T16DPA has also been presented in (Wu et al. 2021).

Low redshift surveys can be considered as a path to prove the effectiveness of the dense interferometric dish array and transit mode observations using the current Tianlai

pathfinder instrument. For an Intensity Mapping survey to succeed, several instrumental and analysis challenges should be overcome, in particular:
- Precise determination of the instrument bandpass response
- Complex gain (amplitude and phase) calibration
- Electronic and environmental induced noise behaviour and its whitening
- Cross-coupling between feeds and correlated noise
- Array configuration, pointing errors
- Individual dish beam response shape knowledge and impact on visibilities and reconstructed 3D maps
- Overall instrument and calibration stability

As we shall discuss it in detail below, there are low-z extragalactic H_I signals, with structuring along the frequency similar to the cosmological LSS signals expected to be seen at higher redshifts, that will be within T16DPA sensitivity reach. The observation of these extragalactic signals would enable Tianlai to assess quantitatively the impact of the above instrumental effects on the recovered signal. We shall also show that it would be possible to determine the residuals from foreground subtraction and impact of individual antenna beam, and bandpass response on these residuals.

The advantage of observing at lower redshifts for T16DPA can be understood in two ways. One one hand, obviously, signals originating from extragalactic sources are much stronger at lower redshifts $z \sim 0.1 - 0.2$ than at redshifts $z \sim 1$, due to the signal strength decrease as $\propto d_L^{-2}(z)$ where $d_L(z)$ is the redshift dependent luminosity distance. However, one might argue that Intensity Mapping does not observe individual sources, but aggregate emissions from neutral hydrogen in $100 - 1000\,\mathrm{Mpc}^3$ voxels. Indeed, for a given setup, instrument angular resolution varies with redshift as $\propto (1+z)$, leading to transverse voxel size evolving as $\propto (1+z)^2 d_M(z)^2$, where d_M stands for the transverse comoving distance with $d_L = (1+z)d_M$ and $d_A = d_M/(1+z)$ (see e.g (Hogg 1999)). So ignoring cosmological evolution of sources, the average per voxel intensity would not vary with redshift. Nevertheless, considering the T16DPA angular resolution of $0.25° - 0.5°$, the voxel transverse size would range from ~ 2Mpc at $z \sim 0.1$ to ~ 10Mpc at $z \sim 0.5$. The voxel size thus exceeds even the cluster size at redshift 0.5, making direct detection of individual structures (galaxies, clusters) by T16PDA, quite unlikley beyond $z \gtrsim 0.1$, as will be shown in section 4.

What about statistical detection of LSS through the 3D map auto-correlation power spectrum? The LSS power spectrum changes slowly with redshift, contrary to distances. One might then expect that an IM instrument's ability to measure the LSS power spectrum would not change significantly with redshift. Unfortunately, the sensitivity to observe the cosmological LSS power spectrum decreases very sharply as redshift increases, due to the way the radio interferometer's noise projects on sky. Indeed, a radio instrument, single dish or interferometer noise power spectrum, projected on sky as $P_{\mathrm{noise}}(k)$ scales as (see for example (Ansari et al. 2012), section 3.2):

$$P_{\mathrm{noise}}(k,z) \propto d_A^2(z)\frac{c}{H(z)}(1+z)^4 \quad (1)$$

$$\propto d_M^2(z)\frac{c}{H(z)}(1+z)^2 \quad (2)$$

This trend is defined by the mapping from instrument coordinates, the two angles defining a direction on sky and the frequency to a 3D position in a cosmological volume. We justify below in a slightly different way the noise power dependence with redshift. Let's consider brightness temperature sky maps $T_b(\alpha,\delta)$ with angular resolution $\delta\theta$ and frequency resolution $\delta\nu$. Instrument angular resolution $\delta\theta$ varies with wavelength $\delta\theta \propto \frac{\lambda}{D_{\mathrm{array}}}$ where D_{array} is the array spatial extent and $\lambda = c/\nu$ the observation wavelength. Projecting such a map on a cosmological volume at redshift z, determined by the observation frequency ν, we obtain voxels with transverse $a_\perp$ and radial $a_\parallel$ comoving sizes, corresponding to a comoving volume $\delta V = a_\perp^2 \times a_\parallel$:

$$\frac{\nu_{21}}{\nu} = (1+z) \qquad \nu_{21} = 1420.4\,\mathrm{MHz} \quad (3)$$

$$\delta\theta = (1+z)\delta\theta_0 \qquad \delta\theta_0 = \delta\theta\,(\nu=\nu_{21}) \quad (4)$$

$$a_\parallel = (1+z)\frac{c}{H(z)}\frac{\delta\nu}{\nu} = (1+z)^2\frac{c}{H(z)}\frac{\delta\nu}{\nu_{21}} \quad (5)$$

$$a_\perp = (1+z)d_A(z)\delta\theta = d_M(z)\delta\theta \quad (6)$$

$$\delta V(z) = d_M^2(z)\frac{c}{H(z)}\frac{\delta\nu}{\nu_{21}}(1+z)^4\,(\delta\theta_0)^2 \quad (7)$$

Map pixel value flutuations due to instrumental noise, denoted σ_T^2 and characterised by the system temperature T_{sys} can be easily related to the noise power P_{noise}. Cosmological volume cell size $(a_\perp, a_\parallel)$ determines the maximum accessible wave numbers $(k_\perp, k_\parallel)$. Assuming white noise and ignoring damping due to per cell averaging, we can write the Plancherel-Parseval identity:

$$\sigma_T^2 = \sum_{k_x,k_y,k_z} |F(k_x,k_y,k_z)|^2$$

$$\sigma_T^2 \simeq P_{\mathrm{noise}} \iiint^{k^{\mathrm{max}}} dk_x dk_y dk_z$$

$$k_{\perp,\parallel}^{\mathrm{max}} = \frac{2\pi}{2a_{\perp,\parallel}}$$

$$\sigma_T^2 \simeq P_{\mathrm{noise}} k_x^{\mathrm{max}} k_y^{\mathrm{max}} k_z^{\mathrm{max}}$$

$$\sigma_T^2 \simeq P_{\mathrm{noise}}\left(\frac{2\pi}{2}\right)^3 \frac{1}{a_\perp^2 a_\parallel}$$

Check - some factor 2 might be missing, as the integral on k should go over positive and negative frequencies We obtain then then the redshift dependence of cosmologically projected instrumental noise, which increases drastically with redshift:

$$P_{\mathrm{noise}}(z) \simeq \frac{1}{\pi^3}\sigma_T^2\left(a_\perp^2 a_\parallel\right) \quad (8)$$

$$P_{\mathrm{noise}}(z) \simeq \frac{1}{\pi^3}(1+z)^2 d_M^2(z)\frac{c}{H(z)}\frac{\delta\nu}{\nu_{21}}(\delta\theta_0)^2 \times \sigma_T^2 \quad (9)$$

An survey of NCP by T16DPA would be sensitive to spherical harmonics $Y_{\ell,m}$ order ℓ in the range $75 \lesssim \ell \lesssim 850$ at $\nu \sim 1400\,\mathrm{MHz}$ (see section 3), corresponding to angular scales $2\pi/\ell$. Taking into account evolution of the instrument angular scale range with redshift ($\ell \propto 1/(1+z)$), we obtain the survey transverse wave number sensitivity range:

$$k_\perp(z) = \frac{\ell(z)}{d_M(z)} \quad (10)$$

$$\ell^{\mathrm{min}}(z=0) \simeq 75 \qquad \ell^{\mathrm{max}}(z=0) \simeq 850 \quad (11)$$

$$k_\perp^{\mathrm{min,max}} = \frac{1}{(1+z)\,d_M(z)} \times \ell^{\mathrm{min,max}}(z=0) \quad (12)$$

We have gathered in the table 1 the cosmological volume cell size, and the accessible transverse $k_\perp$ range for an survey with angular scale sensitivities similar to T16DPA, map pixels with angular size 0.2° at $\nu \sim 1400$ MHz and frequency resolution 1MHz. The projected noise level as a function of redshift is shown in figure 2 as well as the accessible transverse $k_\perp$ range for a T16DPA survey. The maximum value of the radial wave number $k_\parallel^{\max}$ is also listed assuming voxels with $\delta\nu = 1$MHz resolution. However, Tianlai dish array correlator computes visbilities with $\simeq 244$kHz frequency resolution, so the survey could reach a maximum $k_\parallel$ four times higher than the values listed in the table. Unfortunately, all the foreground subtraction methods rely of the smoothness of synchrotron emission with frequency and thus remove the signal modes with low $k_\parallel$. The simulations we have carried out here suggest a low cut-off value $k_\parallel^{\min} \sim 0.15 k_\parallel^{\max}$ (see section 3.3).

Figure 1 shows the radio sky near the NCP (North Celestial Pole), as it appears at 1350 MHz through the combination of Haslam synchrotron map (Haslam et al. 1981) and the sources from the NVSS catalog (Condon et al. 1998). The source closest to NCP, correspond to `NVSS 011732+892848` with J2000 coordinates ($\alpha = 01h17m32.82s, \delta = +89d28m48.7s$) with a flux $\sim$ 2Jy at 1.4 GHz. It is likely associated to the `6C B004713+891245` source identified in the Sixth Cambridge catalog (Baldwin et al. 1985) with a flux of $\sim 7-8$Jy at 152 MHz. The brightest source visible in the map is the `3C 061.1` FR-II radio-galaxy with J2000 coordinates ($\alpha = 02h22m35.046s, \delta = +86d19m06.17s$), and redshift $z = 0.18781$[1] resolved into three sources in NVSS, with a total flux exceeding 6Jy at 1.4 GHz, and $8-10$Jy at 750 MHz.

The visibility simulation and 3D map reconstruction is briefly described in the next section, as well as the simple foreground subtraction methods we have used. We will then show that it is possible to detect individual galaxies or group of galaxies at very low redshifts $z \lesssim 0.05$ in the NCP region. We have also studied the statistical detection of the LSS through cross-correlation with optical surveys, as discussed in section 5. A mid latitude survey, covering a larger area would be less sensitive due to higher noise level, but even more so due to much larger residuals from imperfect foreground subtraction, as discussed in the next section. However, thanks to the larger area, it is also possible to detect the cross-correlation signal.

[1] NED query for object name 3C 061.1 - https://ned.ipac.caltech.edu/

z	d_M	$a_\perp$	$a_\parallel$	$k_\perp^{\min}$	$k_\perp^{\max}$	$k_\parallel^{\max}$	$P_{\rm noise}$
0.1	451	1.7	3.7	0.16	1.9	0.85	4
0.2	880	3.7	4.2	0.08	0.96	0.75	20
0.5	2028	10.6	5.5	0.037	0.42	0.57	200
1.0	3536	24.7	7.3	0.021	0.24	0.43	1430
2.0	5521	57.8	9.7	0.013	0.15	0.32	10400

Table 1. The comoving radial distances, cosmolgical transverse and radial cell size corresponding to angular cell size $0.2^\circ \times 0.2^\circ \times$ 1MHz at $z = 0$. Distances in units of Mpc/h$_{70}$ and k in h$_{70}$Mpc^{-1} and $P_{\rm noise}$ in mK^2/(Mpc/h$_{70}$)^3, assuming a per pixel noise of $\sigma_T^2 \sim 10$mK^2. The values listed for $k_\parallel^{\max}$ assume 1MHz frequency resolution. The foreground subtraction introduces a low cut-off for $k_\parallel$ - see text.

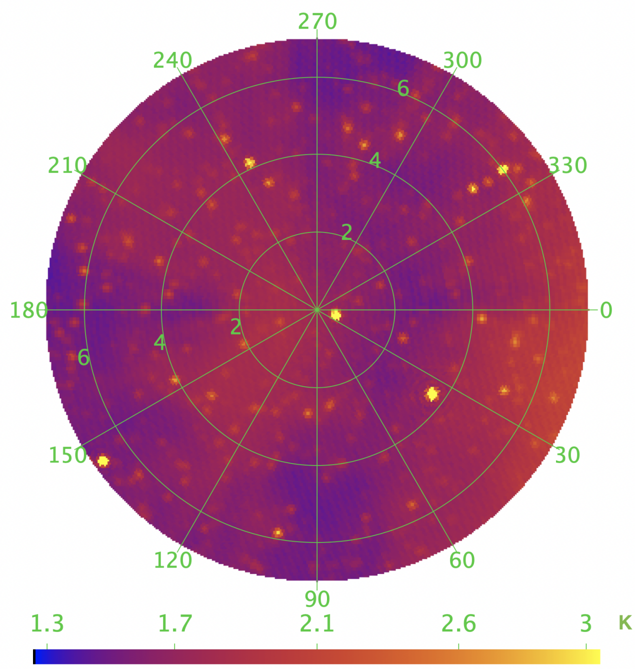

Figure 1. Foreground map of a 7 deg. radius region around NCP at 1350 MHz, smoothed with a 15arcmin resolution gaussian beam. Haslam map of diffuse emission at 408 MHz as well as NVSS radisources, extrapolated to 1350 MHz with a spectral index $\beta = -2$. Color scale corresponds tp temperature in Kelvin.

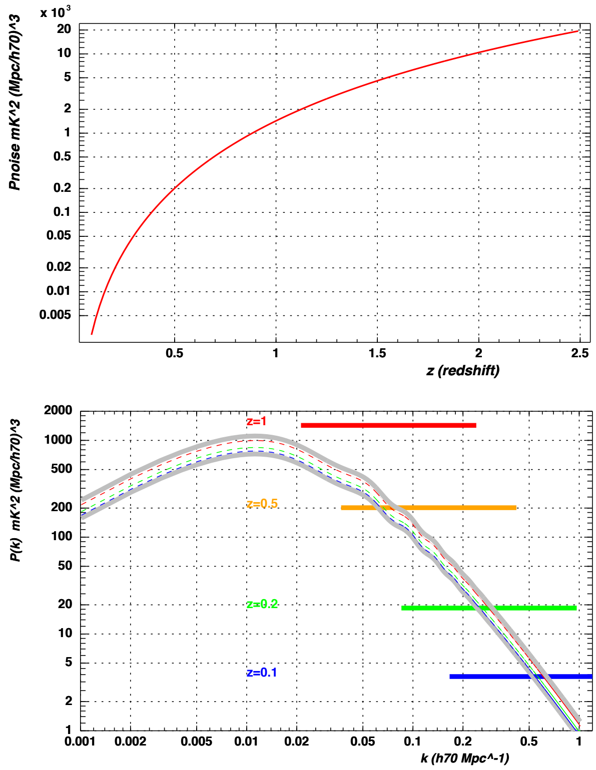

Figure 2. Top: White noise power spectrum $P_{\rm noise}(z)$ level as a function of redshift for a survey similar to T16DPA with an angular map pixel size of 0.2° and 1MHz, and for per pixel noise level $\sigma_T^2 = 10{\rm mK}^2$. Bottom: Projected noise power spectrum $P_{\rm noise}(k)$ and the accessible transverse $k_\perp$ range for a survey of the NCP region by T16DPA.

3 SURVEY SENSITIVITY

3.1 Simulation and analysis pipeline

The JSkyMap [2] package has been used for computing visibilities for the Tianlai dish array setup and the survey strategies studied in this paper. The package provides also several tools for reconstructing maps from transit visibilities. Here, we have used the m-mode visibility computation and map making tools, which operates in the spherical harmonics space $Y_{\ell,m}$ as described in (Zhang et al. 2016) and (Shaw et al. 2015). The simulation and analysis pipeline includes several other C++ or python software modules, which handles the preparation of the input data, such as the generation of HI sources from optical catalogs, foreground subtraction, source detection, power spectrum computation and optical - radio cross-correlation computation.

The study presented here uses only intensity maps, ignoring polarisation. Foreground has been modeled through a combination of the diffuse synchrotron emission, represented by the reprocessed Haslam map at 408 MHz (Remazeilles et al. 2015) and the radiosources from the NVSS catalog (Condon et al. 1998). In practice, for each simulated observation frequency, diffuse synchrotron emission and radiosources have been extrapolated from their reference frequencies, (408 MHz and 1400 MHz) using a fixed value of the spectral index $\beta \sim -2 \ldots -2.5$. All sources with flux larger than 0.05Jy and $\delta > 15^\circ$ have been included in the simulation.

The array configuration corresponds to the actual positions of the antennae in the array, and we have used a frequency dependent Bessel J_1 single dish response, with azimuthal symmetry and an effective dish diameter $D_{\rm eff} = 5.6m$. The beam response used here is represented for $f = 1350\,{\rm MHz}$ in the figure 3. There are 120 different baselines, excluding auto-correlations and ignoring polarisation. Visibilities have been computed with a time sampling $\delta t = 30\,{\rm s}$ and we have generally considered a $\delta\nu = 1{\rm MHz}$ frequency resolution. Two surveys have been studied here, spanning a total duration of several months, up to two years.

(i) A survey of the NCP region with 4 constant declination scans $\delta = 90^\circ, 88^\circ, 86^\circ, 84^\circ$, and covering an area of about $100\,{\rm deg}^2$ around the north pole. We have generally used a fiducial area within 7 deg. from the north pole, $\delta > 83^\circ$, which would yield a surveyed area $\sim 150{\rm deg}^2$. The simulated visibility data set for each simulated frequency plane and for each case represents thus a total number of time samples: $\sim 4(\delta) \times 120({\rm visi}) \times 2800({\rm time}) \simeq 1.35\,10^6$

(ii) A survey in mid-latitude area, covering a much larger portion of sky, using 6 constant declination scans at $\delta = 49^\circ, 51^\circ, 53^\circ, 55^\circ, 57^\circ, , 59^\circ$, covering a 12° band in declination $48^\circ \le \delta \le 60^\circ$, representing about $\sim 12\%$ of the sky or $\sim 2500{\rm deg}^2$. However, we have excluded a region in right ascension contaminated by the galactic plane and bright sources such as CasA and CygA when computing noise power spectrum and mode mixing residuals. The fiducial area used $40^\circ < \alpha < 260^\circ$ represents about $1500{\rm deg}^2$. The visibility data set represents about $2.\,10^6$ time samples par frequency plane and for each simulated case.

The T16DPA noise system temperature has been determined to be $T_{\rm sys} \sim 80\,{\rm K}$ (Wu et al. 2021). The simulations here have been carried out with a fiducial noise level of 5 mK per $\delta t = 30{\rm s}$ visibility samples, and for a $\delta\nu = 1{\rm MHz}$ frequency band. Such a noise level should indeed be reached after $\sim$ 10days (more precisely 8.5×24hours) spent on each constant declination scan, corresponding to a total integration time $t_{int} = 8.5 \times 30 \simeq (256 = 16^2)$s per visibility samples, leading to a a noise level :

$$\sigma_{V_{ij}} = \frac{T_{\rm sys}}{\sqrt{t_{int}\delta\nu}} = \frac{80\,{\rm mK}}{16} = 5\,{\rm mK}$$

AS shown in Wu et al. (2021), the Tianlai dish array day time data is contaminated by the Sun signal leaking through far side lobes. It is therefore planned to use only night time data. T16DPA should thus be able to reach a 5mK per sample noise level by surveying the NCP region during two periods of 1.5 month each, separated by 6 months. 10 days would be spent on each of the four declinations, corresponding to a total of 40 days, or about 1.5 month, and the operation would be repeated six months later, to get full night time coverage of the 24 hours RA range. A noise level of 2.5mK would also be reachable by observing the NCP area over a full year, spending 2×40 days per declination observation, so in total $2 \times 4 \times 40 = 320$ days.

[2] JSkyMap: https://gitlab.in2p3.fr/SCosmoTools/JSkyMap (check the wiki pages)

For each frequency plane, a spherical map is reconstructed through m-mode map making, or rather, the linear system of equations relating visibilities to the input sky is solved in spherical harmonics $a_{\ell,m}$ space. A pseudo-inverse method is used in JSkyMap package and the numerical stabiliy of the inversion process, as well as the noise level are controlled through parameters which define the ratio r_{PSI} of smallest to largest eigenvalue for each inversion, as well as an absolute threshold on the minimal eigenvalue λ_{PSI}. The two values has been set to $r_{\mathrm{PSI}} = 0.02$ and $\lambda_{\mathrm{PSI}} = 0.001$ for the analysis presented here, which can be considered as a medium level.

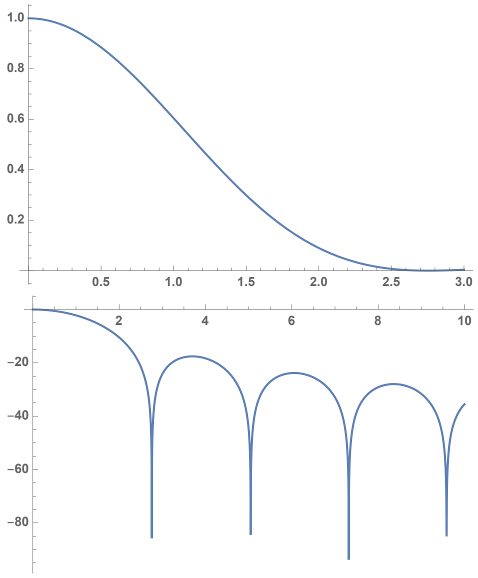

Figure 3. Beam response (Bessel J_1) as a function of the polar axis θ in degrees, with respect to the beam axis (antenna pointing direction) at $f = 1350$ MHz. top: main beam, in linear scale; bottom: main beam and side lobes, in logarithmic scale (dB).

We have used spherical maps with a resolution of 5 arcmin for the reconstructed maps, although the array angular resolution is closer to $10 - 15$ arcmin, as stated in section 2. The reconstructed map pixels have thus a certain level of redundancy, with pixel to pixel noise values being correlated for neighboring pixels, but these higher resolution maps showed a slight advantage for source detection and foreground removal. We have used the `SphereThetaPhi` pixelisation scheme, which features almost square and equal area pixels along θ, ϕ directions, instead of the more standard HEALPix scheme. This pixelisation scheme, available in the SOPHYA[3] library, sometimes called IGLOO pixelisation, preserves to some extent the symmetry around a pixel located exactly at the pole $\theta = 0$ and has also the advantage of being fully flexible in terms of angular resolution or the pixel size.

We also perform an optional filtering step in the spherical harmonics space $a_{\ell,m}$, before map reconstruction and foreground subtraction. The quality of the reconstruction degrades at the two ends of the T16DPA ℓ sensitivity range. At low ℓ, this is explained by the absence of the autocorrelation signal which is not used in map reconstruction, and the minimal baseline length, about 8.8m limits the sensitivity below $\ell \lesssim 75$ for the NCP survey. At the other end, map reconstruction quality and hence the noise level increases for angular resolution corresponding to the array size, for $\ell \gtrsim 850$ for the NCP survey. We have smoothly damped $a_{\ell m}$ coefficients for $\ell \lesssim 75$ and $\ell \gtrsim 875$. A gaussian filter with $\sigma_\ell = 750$ has also been applied, and all $m = 0$ modes have been put to zero. This last filter is intended to remove wiggles with near perfect azimuthal symmetry which appears due to the partial sky coverage combined with limited sensitivity range in ℓ.

Figure 4 shows an example of a reconstructed map, after (ℓ, m) space filtering at a frequency of $f = 1350$MHz. Sources present in the *true* sky map (figure 1), as well as larger structures, are clearly visible, while the noise level $(2 - 4\,\mathrm{mK})$ is too low to be noticable. Some artefacts, such as rings around bright sources can easily be seen and are due to incomplete (ℓ, m) plane coverage and filtering. A patch of sky, as reconstructed from Tianlai dish array survey of a mid-latitude area is shown in the top panel of figure 5. Despite significantly higher noise level (~ 15 mK), it is not noticeable on this reconstructed map, where brightest sources are a few Jy.

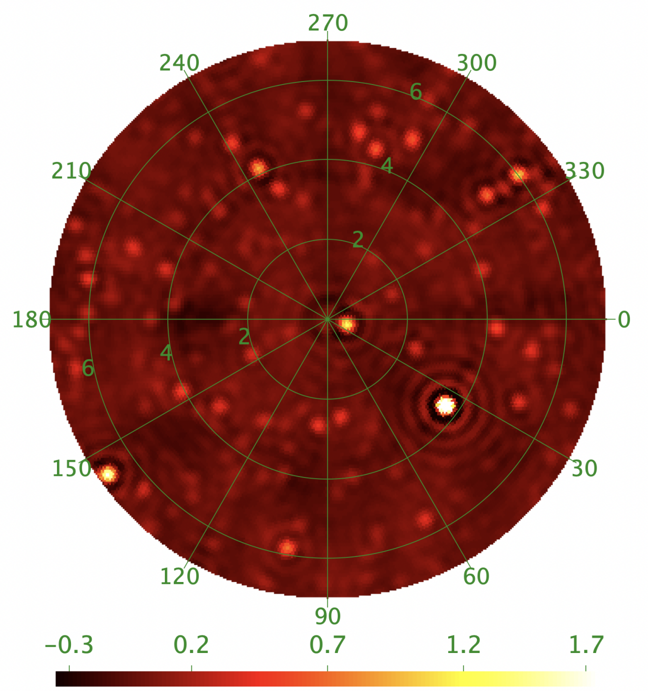

Figure 4. Reconstructed map of the NCP region, as observed by T16DPA at $f = 1350$MHz. This is a the 7 deg. radius around $\delta = 90°$, extracted from the reconstructed spherical map using m-mode map making and after (ℓ, m) space filtering.

Average reconstructed sky angular power spectra of the NCP region is shown in figure 6. The sky power spectrum is higher at larger angular scales, with an overall level about $\sim$ 1K. Effect of the instrument and map making response in the

[3] SOPHYA c++ class library http://www.sophya.org

ℓ space is visible at the two ends, $\ell \lesssim 50$ and $\ell \gtrsim 1150$ (grey curve). The additional effect of (ℓ, m) can clearly be seen comparing the sky power spectrum after filtering (in black) and before filtering (grey curve). The projected noise angular power spectrum $C_{\rm noise}(\ell)$ is also shown on this figure. These $C_{\rm noise}(\ell)$ have been computed from maps reconstructed from white noise-only visibilities, with a RMS fluctuation level of 5mK per $\delta\alpha = 30$s visibility samples. As expected, the noise spectrum increases significantly toward the high-ℓ end of the spectral sensitivity range, above $\ell \gtrsim 800$. This is due first to the decrease of the baselines' redundancy with $< ell$, and then, to the incomplete coverage of wave modes in the (ℓ, m) plane at the high-ℓ end. The effect of (ℓ, m) filtering on the noise power spectrum $C_{\rm noise}(\ell)$ can be seen comparing the orange curve, before filtering, with the red curve, after filtering.

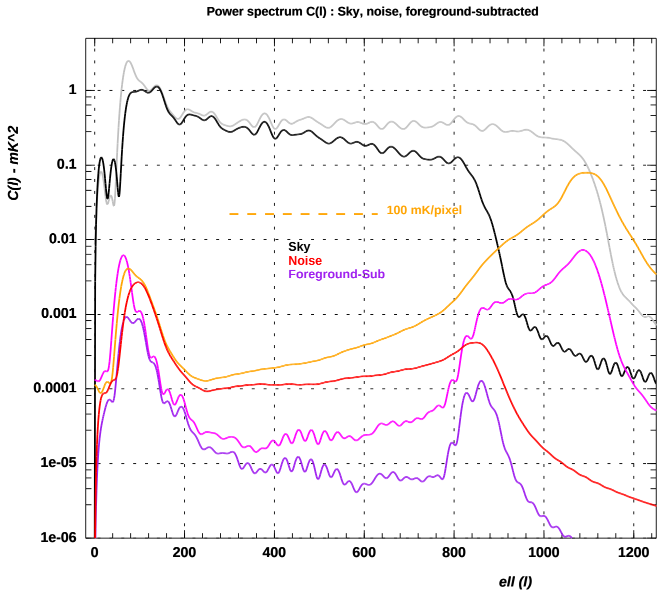

Figure 6. Average angular power spectrum $C(\ell)$ from the a data cube of 100 reconstructed maps of the NCP region, covering an area with 7 deg. radius around NCP, and the frequency range 1300-1400 MHz. The reconstructed sky power spectrum is shown in black, noise power spectrum in red and residual power spectrum after foreground subtraction in purple, after filtering in (ℓ, m) plane. Curves in lighter colors (grey, orange and light purple) show the power spectrum from maps without (ℓ, m) plane filter.

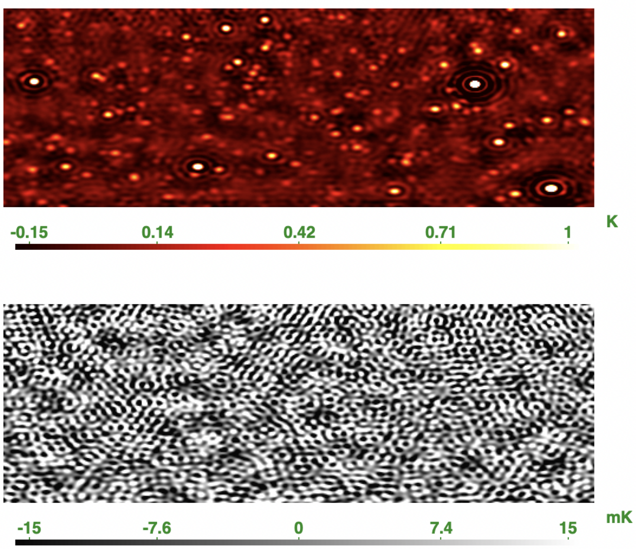

Figure 5. Reconstructed map (top) and noise map (bottom) of the mid-latitude region, after (ℓ, m) filtering, as observed by T16DPA at $f = 1350$MHz. The patch of sky shown covers the declination range $43° < \delta < 57°$ and the right ascension range $90° < \alpha < 130°$, with 5 arcmin pixel size.

3.2 Foreground subtraction

Our aim here is not to devise the best foreground subtraction method, but rather characterise the instrument and survey strategy sensitivity and the corresponding performance for mitigating the mode mixing. Two simple foreground subtraction methods have been used here which exploits the synchrotron dominated foreground smoothness with frequency. The first method (**P**) represents the synchrotron emission frequency dependence as a second degree polynomial in frequency. The coefficients are determined for each direction through a linear χ^2 fit to the measured temperatures, and the resulting fitted foreground $T^{\rm fgnd-P}_{\alpha,\delta}(\nu)$ is then from the 3D temperature map:

$$T^{\rm fgnd-P}_{\alpha,\delta}(\nu) = A_{(\alpha,\delta)}\, \nu^2 + B_{(\alpha,\delta)}\, \nu + C_{(\alpha,\delta)} \quad (13)$$

$$T^{\bf P}(\alpha,\delta,\nu) = T(\alpha,\delta,\nu) - T^{\rm fgnd-P}_{\alpha,\delta}(\nu) \quad (14)$$

The second method (**DF**) is a very simple difference filter along the frequency. For each frequency plane, we subtract the average of two nearby frequency planes at ν_-, ν_+, with a specified frequency gap $\Delta\nu$, $\nu_- = \nu - \Delta\nu$ and $\nu_+ = \nu + \Delta\nu$. We have used $\Delta\nu = 2$MHz throughout this paper.

$$T^{\rm fgnd-DF}_{\alpha,\delta}(\nu) = \frac{1}{2}\left(T(\alpha,\delta,\nu_-) + T(\alpha,\delta,\nu_+)\right) \quad (15)$$

$$T^{\bf DF}(\alpha,\delta,\nu) = T(\alpha,\delta,\nu) - T^{\rm fgnd-DF}_{\alpha,\delta}(\nu) \quad (16)$$

We have represented the average angular power spectrum of the residual signal $C_{\rm res}(\ell)$, after foreground subtraction for the two methods (P,DF), for the NCP survey on the left panel of figure 8. A set of 100 sky maps, reconstructed from mock visibilities, corresponding to four constant declination scans at or near the NCP, and after (ℓ, m) plane filtering have been used. Compared to the input sky angular power spectrum $C_{\rm sky}(\ell)$ shown as the black curve, one can see that the foreground angular power spectrum is suppressed by a factor $\gtrsim 20000$ for the polynomial subtracted foreground (P), and $\gtrsim 60000$ for the difference along frequency filter (DF method). These values correspond to a factor ~ 150 (P) and ~ 250 (DF) damping in amplitude for temperature fluctuations due to foreground. While this might not be sufficient for the direct detection of the cosmological 21cm signal, the foreground residuals due to mode mixing and imperfect subtraction would be well below the instrumental noise level for the NCP survey by Tianlai.

However, T16DPA becomes less efficient to fight mode mixing for a mi-latitude survey. We have shown the angular power spectrum of the residual after foreground subtraction using the DF method, for a mid-latitude survey, on figure 9. A set of three reconstructed maps at $f_{-2} = 1348$MHz, $f_0 = 1350$MHz and $f_{-2} = 1352$MHz have been used for the power spectra shown on this figure. The fiducial area used to compute the power spectra excludes right ascension ranges corresponding to the Galactic plane or contaminated by CasA and CygA and represents about 1500deg^2. Comparing the black curve, which represents the reconstructed $C_{\rm sky}(\ell)$ of the diffuse synchrotron and radio sources and violet curve $C_{\rm res}(\ell)$, corresponding to the residuals after foreground subtraction through the DF method, we see that the $C(\ell)$ power spectrum has been damped by a factor ~ 1200, or about ~ 35 for the temperature fluctuation amplitude. The residual after foreground subtraction reaches a level similar to the noise on the map, with RMS fluctuations ~ 15mK.

This dampling factor, about 50 times lower, or 7 times lower in amplitude, is explainded by a higher level of mode-mixing for a mid-latitude survey by Tianlai, compared to the NCP case. Indeed, one can consider that for observations toward NCP, the projected baselines changes with the sky rotation, improving map making performance in terms of individual mode reconstruction. The circular confiuration of T16DPA was optimised for a good coverage of the angular sky modes or the (u, v), minimising the redundant baselines, compared to a regular rectangular grid configuration for example. Although arrays with redundant baselines offer advantages for the gain and phase calibration, they should exhibit higher level of mode mixing. For very large arrays, with several hundred or several thousand elements, a combination of redundant and non redundant baselines should be used to mitigate both mode mixing and calibration issues.

3.3 Noise level and survey sensitivity

The left panel of figure 7 shows an example of noise map for the Tianlai NCP survey, while the histogram of the corresponding pixel value distribution is shown on the right panel. The pixel to pixel temperature fluctuation is bout 4mK, and even about 2.2mK for the central 3° radius area, and for a 5mK noise level per 30s visibility samples. The bottom panel of figure 5 shows a similar noise map for the mid-latitude survey, with an RMS pixel fluctuation level of ~ 16mK. The noise level scales slightly faster than the square root of the ratio of the surveyed sky area, $(2500{\rm deg}^2/150{\rm deg}^2 \simeq 16)$, as the mid latitude survey discussed here requires 6 constant declination scans, hence 50% more observing time.

As mentionned already, the maps with 5arcmin pixels used here have higher resolution than the effective instrument and reconstruction angular resolution, limited to about $\ell^{\rm max} \sim 850$ or 12arcmin. The noise correlation between neighbouring pixels is visible on the noise maps, and compatible with the noise angular power spectrum $C_{\rm noise}(\ell)$. For the NCP survey, the noise power spectrum is nearly flat for $200 < \ell < 800$, and 15-40 higher than the foreground subtraction residuals, as shown in the left panel of figure 8. Tianlai should be able to get noise dominated foreground subtracted maps for the NCP region, even for a deep NCP survey, reaching ~ 2mK per pixel noise level, for a survey where each of the 4 declinations would be observed during $2 \times 40 \times 24$hours.

The right panel of figure 8 shows the sky and residual after foreground subtraction radial mode power spectra $P(\tau)$ for the NCP survey. Theses have been obtained computing the average spectrum obtained through a Fourier Transform (FFT) along the frequency direction, for each direction of sky. The Fourier wave modes along the frequency correspond to a time and are sometimes referred to as the delay τ. Given the 100 MHz bandwidth, with 100 frequency planes, the frequency waves modes, or delay, cover the range from $\tau_1 = 10$ns to $\tau_{50} = 500$ns. The black curve represents the average reconstructed sky $P_{\rm sky}(\tau)$, with the power highly concentrated at very low delay modes ($\tau \leq 20 - 30$ns), but with still significant power up to $\tau \lesssim 100$ns. The effect of the two foreground subtraction methods, and their The $\tau-$response can be understood by looking at the shape of the average delay-spectrum of the residual maps without noise (purple curves), or the ones with noise (orange/red). The latter are noise dominated and it can be seen that the polynomial foreground subtraction (P) removes delay-modes below $\tau \lesssim 50 - 60$ns, while the differential filter along frequency (DF) can be considered as a band pass filter, removing $\tau \lesssim 100$ns and $\tau \gtrsim 400$ns. The (DF) method is more effective at removing foreground modes at low delay, but leads to noisier maps. In addition to removing low-delay modes, the polynomial subtraction method (P) damps the power $P(\tau)$ by a factor about 30 for all modes $\tau > 100$ns.

So far, we have assumed a prefect knowledge of the instrument response, specially the individual antenna angular response, instrument frequency and relative gain and phase calibration. Discussion of impact of an imperfect knowledge of the instrument response on the survey performance is beyond the scope of this paper. However, preliminary studies suggest that one should be able to cope with lack of knowledge of individual antenna side lobes for the NCP survey, if the main lobe is well modeled. On the other hand, phase calibration errors can significantly degrade the instrument ability to remove foregrounds. To illustrate this, we have included the green curves in figure 8 which show the average power spectrum of the foreground subtracted maps, in the presence of phase calibration errors To obtain these curves, we have applied a relative phase error, drawn independently for each baseline and each frequency, according to a zero-mean normal distribution with an RMS of 7°. In this case, the residual maps would be dominated by foreground contamination, with a power spectrum $C_{\rm res}(\ell)$ 10 times larger than the one due to the instrument noise.

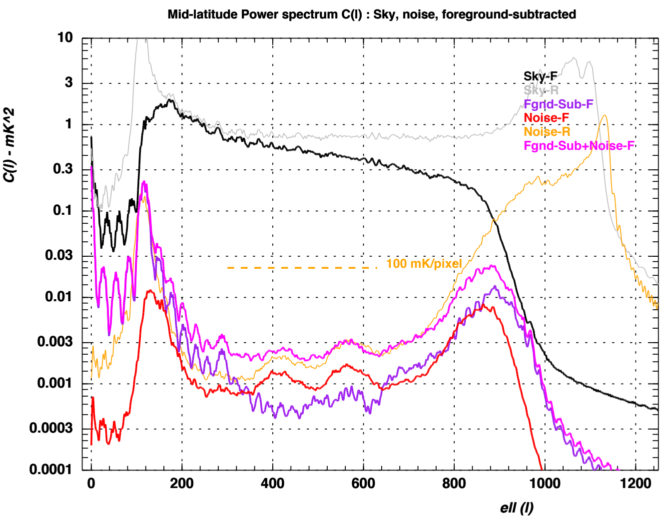

Figure 9. Sky, noise and residual after foreground subtraction angular power spectrum $C(\ell)$ for the mid-latitude survey. These power spectra have been computed from a set of three frequency maps at 1348, 1350, 1352 MHz and the differential filter along the frequency (DF) foreground subtraction method has been used for this plot.

4 $\mathrm{H_I}$ CLUMP DETECTION

The aim of this analysis is to assess the number of direct detection of $\mathrm{H_I}$ clumps in a low-z survey of either a mid-latitude band or a circular region around the North Celestial Pole with T16DPA. We first estimate the $\mathrm{H_I}$ clumps detection efficiency as a function of their flux for the NCP and mid-latitude cases. In a second step we combine these detection efficiencies with 21cm flux derived from the ALFALFA $\mathrm{H_I}$ clump mass function to determine the expected number of detected clumps, assuming a spatial uniform random distribution, for the NCP and mid-latitude cases.

4.1 $\mathrm{H_I}$ clumps flux detection thresholds

To assess the detection efficiency for point-like $\mathrm{H_I}$ sources we have used a pipeline sharing most of the components described in section 3.1. We simulate observations of the NCP and mid-lattitude surveys as described there, for only three frequencies : 1348, 1350 and 1342 MHz.

To the generic astrophysical components (diffuse synchrotron Galactic emission , and continuum NVSS sources) we add, for the central frequency a set of uniformly distributed point-like sources of a given flux (in Jy). For each frequency we compute simulated visibilities, and then reconstruct sky maps and apply angular mode filter, as explained in 3.1. The noise level per visibility sample (30s integration time) used in the following is 5mK. In order to account for the impact of foregrounds, we have used the difference filter along the frequency (**DF**) described above, which correspond to subtract from the central frequency map the average of the two outer frequency ones. Finally, we reproject the obtained difference into rectangular (mid-latitude case) or square (NCP case) maps. An additional high-pass filter in spacial frequency domain has been applied in the mid-latitude case to reduce foreground subtraction residuals.

The final step is the source detection. We use a basic scheme based on the `DAOStarFinder` class from the `photutils` Python package (Bradley et al. 2021). Loose sphericity criteria for the source detector has been set, to compensate for remaining artefacts due to map reconstruction ans foreground subtraction. We set the detection threshold, expressed as multiples of the map pixel to pixel RMS fluctuation level to 7 and 10 for the NCP and mid-latitude cases respectively, to avoid spurious detections. times this value The detection efficiency by is determined by the number of detected sources within 2 pixels (CHECK THIS) of the simulated ones. In the NCP case, we simulated 5 sources over the 7 (TBC) degrees circular observed region, but repeat this operation 20 times to reach a statistical accuracy of a few percents. In the mid-latitude case, the rectangular surveyed area is much larger so that this repetition is not needed more than 2 or 3 times.

The detection efficiencies we measure in the NCP and mid-latitude simulations are reported on figure 4.1. Thanks to the higher integration time per map pixel in the NCP case, these results show that the detection threshold S_*^{th}, defined as the flux limit with a detection efficiency $\geq 50\%$ is much lower in the NCP case than the mid-latitude one, with $S_*^{th} \simeq 0.08$Jy for the NCP case, compared to $S_*^{th} \simeq 0.9$Jy for the mid-latitude case. We have in each case fitted the source detection efficiency as a function of the flux an error function $\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$. The fiited shapes have then been used in the computation of the expected number of $\mathrm{H_I}$ clump direct detections in the low-z Tianlai survey.

4.2 Number of expected $\mathrm{H_I}$ clumps observations

We assume the $\mathrm{H_I}$ clumps population to have a random spatial distributed and to follow the characteristics measured using ALFALFA survey data in Jones et al. (2018). As shown in this paper, the mass function (MF) of the $\mathrm{H_I}$ clumps - the number density of clumps in a logarithmic HI mass bins, is well described by a Schechter function :

$$\Phi(M) = \frac{dN_{\mathrm{H_I}}}{dV \, d\log_{10}(M)} = \log(10)\Phi^* \left(\frac{M}{M^*}\right)^{\alpha+1} \exp\left(-\frac{M}{M_*}\right) \quad (17)$$

where Φ^* corresponds to the normalisation, M^* the knee mass and α, the low mass slope. Jones et al. (2018) fit these parameters using several subsets of the ALFALFA clump dataset ; these results show some spatial dependence. We will retain here the parameters fitted with the whole dataset (ALFALFA 100%) and its 'near' subset ($v_{CMB} <$ 4000km/s), listed in Table 2. The difference between the "full" and "near" parameter may give an indication of the systematics linked to the HI mass function. ALFALFA also observed some variation of these parameters in different regions on the sky but we do consider different areas in this study, therefore we stick to this global variation with observed distance in the following.

We use the following expression to relate the $\mathrm{H_I}$ mass

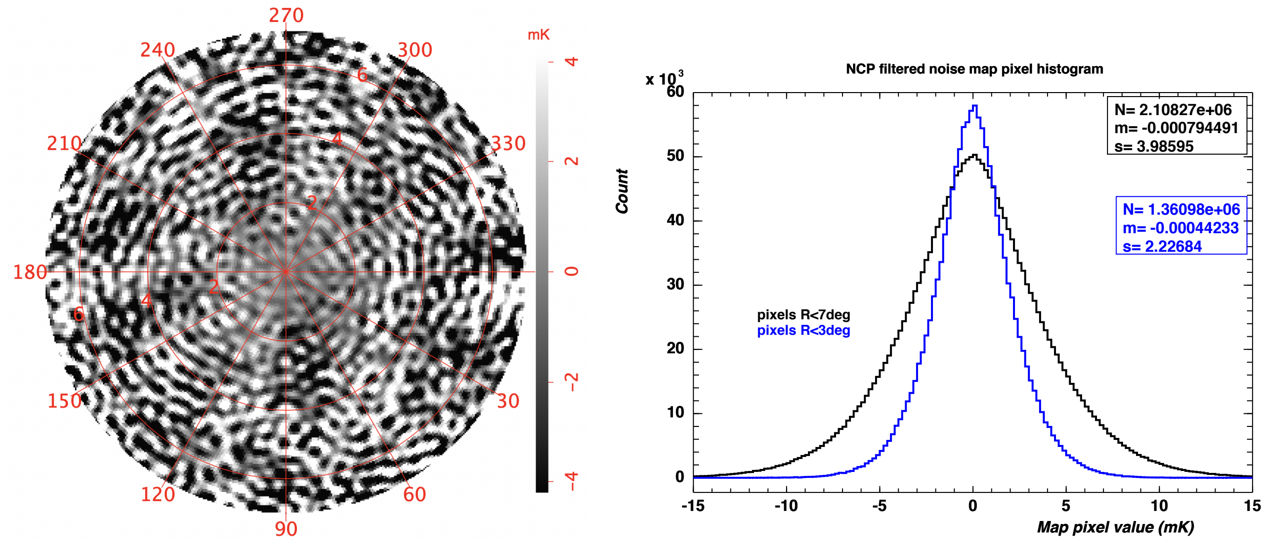

Figure 7. Left: Noise map after reconstruction with (ℓ, m) filtering of the NCP region covering an area with 7 deg. radius at $f = 1350$MHz (map scale in mK). Right: Noise map pixel value distribution, in black for the full 7 deg. radius map around NCP, and blue, restricted to the central 3 deg. radius, covering $\sim 30\text{deg}^2$. Note that restricted area corresponding to the blue histogram correspond to $\sim 18\%$ of the full area; The blue histogram has been rescaled to enhance the figure readability

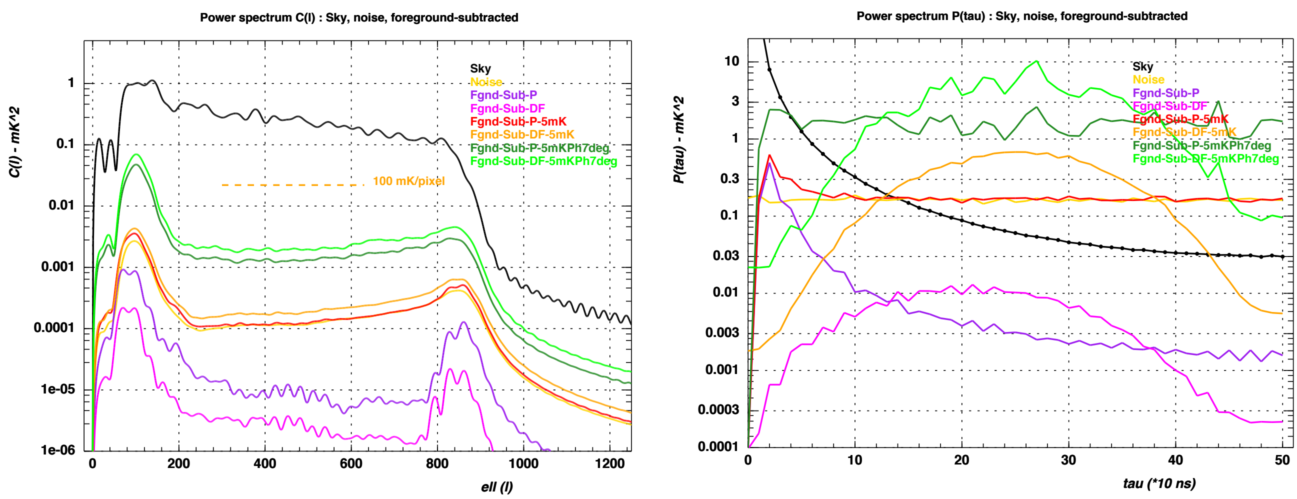

Figure 8. Left: average angular power spectrum $C(\ell)$ from the a data cube of 100 reconstructed maps, with (ℓ, m) filtering of the NCP region covering an area with 7 deg. radius around NCP, and the frequency range 1300-1400 MHz. Right: average power spectrum along the frequency axis $P(\tau)$. The reconstructed sky power spectrum is shown in black, noise power spectrum in gold and residual power spectrum after foreground subtraction in purple without noise, and in red with noise. The green curve shows the effect of phase calibration errors (7 degree RMS gaussian phase erros) on the residuals after foreground subtraction.

Dataset	α	m_*	ϕ_*
Full	-1.25 ± 0.02	9.94 ± 0.01	.0045 ± .0002
Near	-1.22 ± 0.02	9.76 ± 0.04	.0062 ± .0005

Table 2. HI mass function parameters determined using either the whole ALFALFA dataset and its near subset

and the total 21cm flux S_{21}, also quoted in Jones et al. (2018) :

$$\frac{M_{HI}}{M_\odot} = 2.356 \times 10^5 \left(\frac{D}{1\text{Mpc}}\right)^2 S_{21} \tag{18}$$

where D is the source distance in Mpc and S_{21} the integrated 21cm flux in Jy.km/s. For each redshift value, we compute D using a fiducial cosmology (Planck 2015). Using this distance and assuming a $\sim$ 210km/s velocity width (corresponding to ~ 1 MHz in the frequency domain) we can translate the flux limits or detection efficiencies in Jy into H_I mass limits or detection efficiencies at each redshift. The integral of the H_I mass function convolved with the detection efficiency gives the expected number density of clumps detectable of Tianlai at any given redshift, and integrating over the redshifts, taking into account volume element evolution with redshift, we obtain the expected total number of H_I clump detections.

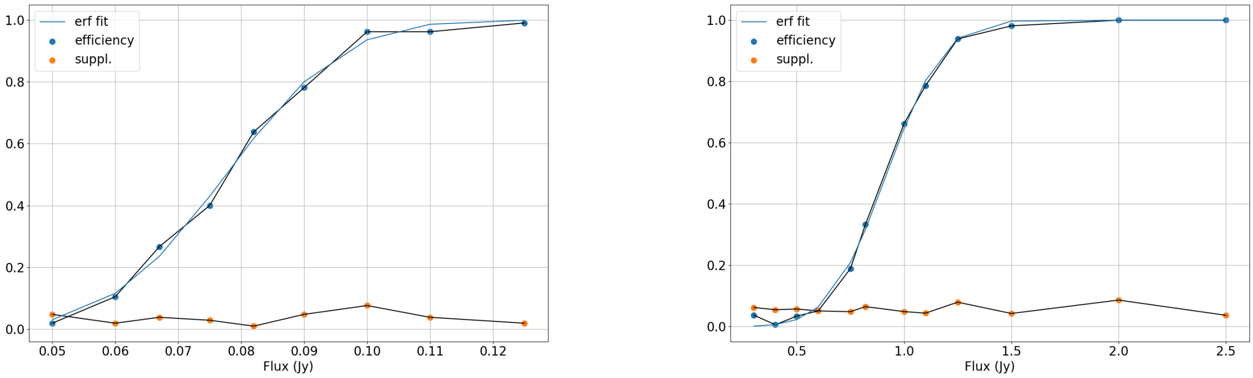

Figure 10. HI clump detection efficiences as a function of flux, for the NCP (left-) and mid-latitude (right hand side) measured by our simulations. On each part we represent as blue dots the efficiences measured at each simulated flux. the read dots correspond to the number of spurious detection (detections located farther than 2 pixels from the simulated clumps positions). The cyan curve is a fit of the efficiencies values with an error function.

Dataset	NCP	mid-latitude
Surface (deg^2)	150	1500
H_I clumps/deg^2 (full MF)	.048	.0012
H_I clumps/deg^2 (near MF)	.035	.0009
Total N-clumps (full MF)	7.2	1.8
Total N-clumps (near MF)	5.25	1.35

Table 3. Number of expected HI clump discoveries per square degree for the NCP and mid-latitude surveys, for the two parametrizations of the HI mass function given in table 2. The sky area covered for each survey is given, as well as the total number of expected detections.

We present the expected number of H_I clump detections per square degree and per redshift bin for the NCP and mid-latitude cases (right hand side) as a function of redshift, assuming the H_I mass function parameters from ALFALFA full sample on figure 4.2. The total number of detections per square degrees are reported on this figure and also in table 3. As can be seen from the plots on this figure 4.2, Tianlai would only be able to detect very nearby clumps of galaxies, below $z \lesssim .02$ for the NCP survey, and $z \lesssim 0.005$ for the mid-latitude survey, and nearly independent of the H_I mass function parameters used. At these very low redshifts it may therefore be justified to use the *near* H_I mass fuction parameters from ALFALFA. As the knee mass is somewhat lower in that case, this results in lower number of expected detections, as reported in table 3.

In the mid-latitude case, on the sky results in a higher detection threshold, hence a lower expected redshift and number of detections than in the NCP case. This is not totally compensated by the much larger surface covered by this survey, ,making the NCP survey the most promizing in terms of expected HI clump discovery rate. As mentionned in section 3 a lower noise per visibility sample may be achieved by observing over longer period, less than a year for the full survey. In addition, the **DF** foreground subtraction method used for determining the source detection efficiency increases the noise level of the resulting difference map. The noise level which is the major limitation of the detection efficiency for NCP case can be reduced by a factor about 2-3, combining lower visibility noise from the long survey duration and lower noise impact using the (**P**) foreground subtraction. A detection threshold $S_*^{th} \lesssim 0.05$Jy could then reachead for the NCP survey, leading to an expected number of detectable clumps about 10-15.

RA: Maybe we should computed the number of detectable sources assuming that the detection efficiency curve for the NCP case would be shifted to 0.05 or 0.04 Jy, which should be conservative, and add the two curves to figure 11-left, and add also the numbers to table 3

5 CROSS-CORRELATION WITH OPTICAL GALAXY CATALOGS

In this section, we assess the prospects of detecting the cross-correlation signal between the Tianlai low redshift surveys intensity maps with optical galaxy catalogs, SDSS for the mid-latitude and NCCS for the polar cap survey. we explore the prospects for extracting the cross-correlation signal between the Tianlai low redshift survey. We study first the most straightforward case, estimating the strength of the cross-correlation signal between a Tianlai mid-latitude low redshift survey and the SDSS catalog RA: need a reference here for SDSS, which overlaps its sky footprint.

The SDSS catalog does not cover the NCP, that's why we are carrying a spectroscopic survey using the WIYN telescop to obtain the spectroscopic redshifts for the brightest galaxies of the photometric NCCS catalog (Gorbikov & Brosch 2014). To evaluate the cross correlation signal for the NCP case, we have used an artificial catalog built by rotating the coordinates of the objects in the SDSS catalog to get an overlap with observations towards NCP. The respective footprints of the SDSS and NCCS are shown on figure 5.

Starting from the optical galaxy catalog, we create a catalog of 21cm H_I sources using a two step procedure. We derive first a stellar mass from optical galaxy properties, following (Taylor et al. 2011). The stellar mass is then converted into an H_I mass using the relations determined by (Brown et al. 2015), derived from the study of a combined

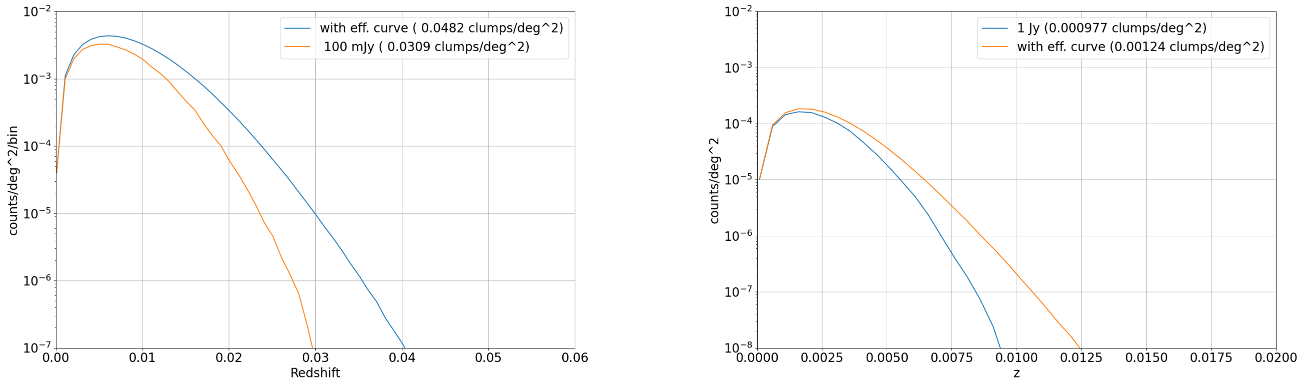

Figure 11. Expected number of detections vs redshit for the NCP (left-) and mid-latitude (right hand side) cases. The integral numbers of detections are reported in the caption of the figures. In both cases we indicate the results we obtain with a detection threshold on the flux, in Jy, and with the efficiency curves shown on figure 4.1.

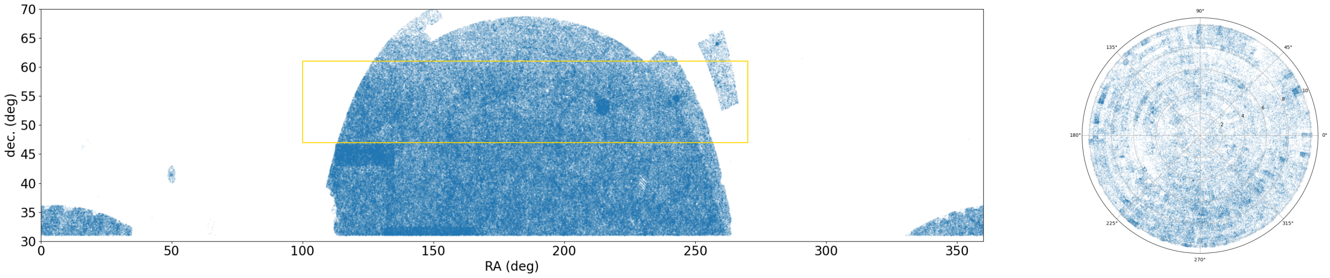

Figure 12. Footprints of the SDSS (left) and NCCS (right) catalogs used in this paper. We selected galaxies above $\delta = 30$ deg in the SDSS catalogs, and objects with PESS (point vs extended source identification) score greater than 2 in the NCCS. The rectangular area outlined on the SDSS footprint is the area where cross-correlation with Tianlai low-z simulated observations have been computed.

ALFALFA-SDSS catalog. A more detailed description of the procedure converting the optical catalog of galaxies into a list of 21cm source properties can be found in the appendix A

The procedure used to determine the cross-correlation of Tianlai low-z observations with the SDSS or NCCS optical catalogs uses the pipeline described in section 3.1 , with few additional components:

(i) creation of the 21 cm source catalog from SDSS galaxies with their redshifts, or the rotated SDSS catalog for the NCP case

(ii) simulation of visibility that one one observe with the T16DPA setup, combining signal from the different components of the sky, diffuse synchrotron emission, radio sources noise and redshifted 21 cm sources. Instrument noise, is added to visibility samples, as white noise. level,

(iii) Sky reconstruction, independently for each frequency, using the m-mode decomposition approach

(iv) Computation of spherical sky map, after filtering in spherical harmonic (ℓ, m) space

(v) Foregound removal using either of the two approaches presented in section 3.2

(vi) re-projection of the filtered maps in a equatorial band around the central latitude of the simulated observations and selection of relevant portion of this band for cross-correlation studies, as outlined on the left panel of figure 5. For the NCP case, the portion of spherical maps around the celestial pole are projected into square maps, using tangent projection. A sky cube is then made out of all projected frequency plane.

(vii) A source sky cube is also constructed from the optical catalog, using only the galaxies angular positions, and their redshifts, ignoring photometric information. all galaxies brighter than mag RA: specify the magnitude cut are included. We apply a gaussian filter to smear source positions, with RA: specify the gaussian sigma in the transverse-angular plane and along the frequency axis

(viii) For each frequency plane, we compute the cross-correlation of the sky reconstructed from the visibilities with the corresponding plane from the source cube. A correlation coefficient is computed, as well as a cross spectrum, in spherical harmonics $C_\nu^\times(\ell)$ for the mi-latitude case, and in Fourier domain $P_\nu^\times(k_\perp)$.

The source cube to be correlated with the reconstructed sky cube has the same angular and frequency resolution RA: specify pixel resolution as the sky cube. Each galaxy in the optical catalog RA: is there a magnitude cut is assigned to a pixel in the cube. The frequency plane is determined from the source redshift, and the position in the plane from the angular coordinates of the galaxy. All galaxies have the

Bin	$\nu_{\rm min}$ (GHz)	$\nu_{\rm max}$ (GHz)	$z_{\rm center}$
1	1250	1300	0.103
2	1300	1350	0.068
3	1350	1400	0.032

Table 4. Frequency/redshift bins used fro the mid-latitude Tianlai-SDSS cross-correlation analysis.

same weight, equal to one, regardless of their photometric magnitudes. A gaussian smearing with $\sigma_{\rm freq} = xx$MHz along the frequency direction is then applied, as well as a 2D gaussian filter to each plane, with a fixed angular width $\sigma_{\perp} = yy$arcmin.

We build also shuffled source cube to determine the level of residual cross correlation signal, due to imperfect foreground subtraction and instrumental noise. The source positions are randomised in the input catalog through shuffling to create the shuffled or null cubes. We expect a null correlation with shuffled cubes, and about a hundred shuffled cube has been made and correlated with the reconstructed sky cube to estimate the cross correlation signal dispersion

We ran the pipleine and analysed the results for different combinations of sources in the sky, as lsited below, and two noise configurations, with or without 5mK noise added to visibility samples.

- no signal (noise only simulation)
- continuum sources only (Haslam-based synchrotron map and NVSS sources)
- $\rm H_I$ simulated sources only
- all, which include contribution from all the above source (diffuse synchrotron, radio sources, $\rm H_I$ sources and noise)

5.1 Mid-latitude survey cross-correlation with the SDSS catalog

We ran this pipeline for all 1 MHz frequency planes between 1250 and 1400 MHz. We computed auto- and cross C_ℓ for each frequency plane, between the reconstructed sky cube and the source cube, built from the optical catalog. The cross correlation is also computed with each of the shuffled source cubes. One expects the cross power spectra between the randomized source cube and the simulated maps to be null in average, and their dispersion around their average will give an estimate of the uncertainty of the computed cross-correlation coefficient or power spectrum. To get a more synthetic view we average these cross correlations over frequency bands and we define three frequency (or redshift) intervals as described in Table 4.

We present on figure 13 examples of auto-correlation power spectra we obtain after such averaging over frequencies between 1350 and 1375 MHz, without visibility noise. The spectrum prior to foreground subtraction seems to be truncated below $\ell \sim 100$ and above $\ell \gtrsim 750$, mainly as a result of the map-making and filtering procedure as shown in zection3.1. In between these two ℓ values, the effect of the foreground subtraction (polynomial fit) decreases the auto-spectrum by 3 to 4 orders. For comparision we also show the variations of the power spectrum from the 'reference plane'.

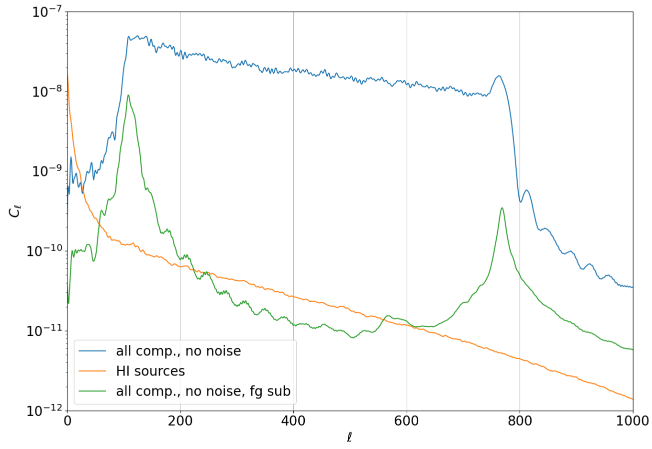

Figure 13. Average of auto-correlation spectra (C_ℓ) from frequency bin 5 RA: should not be 5, give the frequncy range here. Blue and green curves correspond to the sky cubes with all components, with no noise added on visibilities, before (blue) and after(green) foreground subtraction (using polynomial fit in frequency domain). The orange curge is the auto-correlation spectrum from the corresponding source cube

From the same frequency interval, we present in figure 14 averaged cross-power spectra between the 'reference planes' or sources planes and sky cube reconstructed from visibilities, for the same cases: all astrophysical components with and without foreground subtraction (blue and orange, respectively). The improvement brought by the foreground subtraction for extracting a non sero cross-correlation signal is evident. In that case, the cross-power spectrum amplitudes are indeed much decreased but stay above 0 for a broad range of multipoles, in the ℓ interval less affected map-making, filtering and foreground subtraction procedures. The same stays true, at least on average, with noise at visibility level (green curve). showing that a significant cross-correlation signal could be extracted from a future Tianlai mid-latitude low-z survey. We also note that at low ℓ ($\ell \lesssim 150$), reconstruction residuals and mode mixing due to incomplete coverage induce large biases on the cross C_ℓ spectra. RA: Should update the figure 14 below with a smoothed version of the figure, and show the cross correlation signal, without and without noise, with and without foreground subtraction

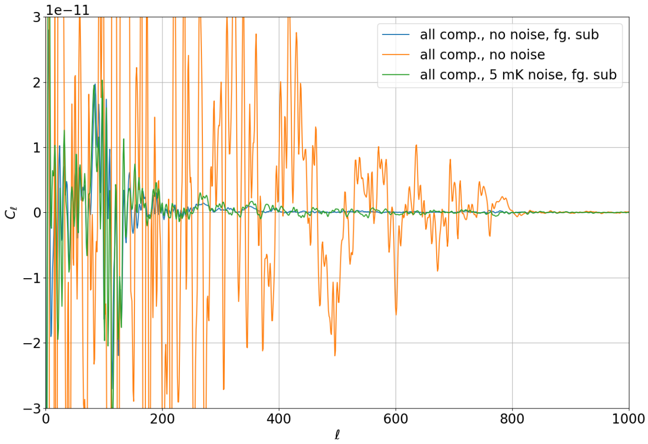

Figure 14. Average of cross-spectra coefficients (C_ℓ) from frequencies between 1350 and 1375 MHz. The orange and blue curve represent the cross-spectrum obtained between the 'reference plane' and data simulated with all astrophysical components but no noise, respectively before (orange) and after(blue) foreground subtraction (polynomial fit). The green curve shows the result obtained with 5 mK noise added, after foreground subtraction.

Due to (very) incomplete coverage, these power spectra modes are correlated with each other. We then smoothed the power spectra with a $\Delta\ell = 15$ gaussian to mitigate this effect. We present in figure 15 a set of power spectra from the ideal situation of a simulation including only $\mathrm{H_I}$ sources, without noise to the complete simulation of all component with noise. We can observe the impact of adding noise or including more components which leaves systematic residuals in the maps after foreground subtraction, due e.g. to mode mixing. We observe that these two effects have roughly similar impacts in terms of the cross-power spectra between simulated planes and the data cube, as could be expected from the analysis reported in section 3.3. We also note that in a broad ℓ range, the averaged cross power spectrum for the complete simulation (in blue) stays positive, and well outside the dispersion from the 100 shuffled cubes. This reinforces the indication that Tianlai could observe a significant cross-correlation with the SDDSS catalog, when performing a mid-latitude low-z survey as analysed in this work.

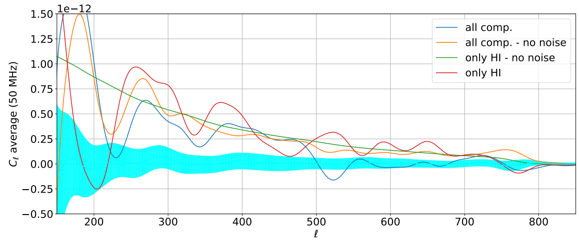

Figure 15. Smoothed cross power spectra averaged over frequencies from bin 3, for different simulation combinations, after foreground subtraction. The cyan band outlines the dispersion around central values of the average of the cross power spectra between maps from the simulation combining all components and noise and the 100 shuffled data cubes.

To get a more synthetic view of the cross-correlation detection perspective, we compute the average of the cross-spectra amplitudes for $\ell \in [250, 500]$, well within the ℓ range unperturbned by out map-making and filtering procedures. We present results of this averaging procedure for the three redshift bins on figure 16.

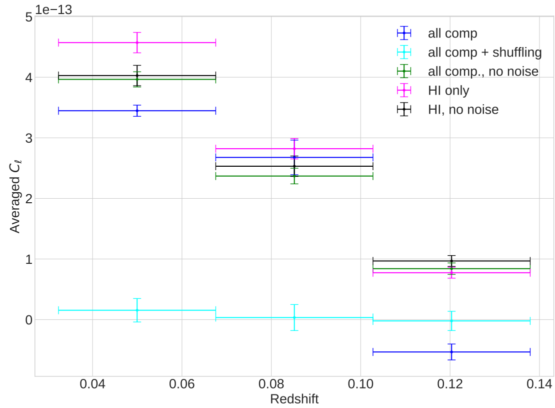

Figure 16. Average of the cross-correlation C_ℓ in the interval $\ell \in [250, 500]$ for each of the redshift bins defined in table 4

We can observe that in the two lowest redshift bins the averaged C_ℓ of the complete simulation case is positive and at a distance from zero much larger than the dispersion expected from the null tests. Taking these differences at face value this would amount to $\sim$ 15 and $\sim$ 10 σ deviations. While this uncertainty estimation is very crude, this shows that a low-redshift mid-latitude survey operated in the conditions described in this paper by Tianlai with the T16DPA array would observe a very significant cross-correlation with the sources listed in the SDSS catalog.

5.2 Cross-correlation towards NCP

In this case, we cannot rely on any existing spectroscopic extragalactic source catalog. We first forecast the cross-correlation sensitivity of a survey towards the NCP with the

T16DPA instrument within an idealized case, in which we rotate thet SDSS source's coordinates in order to cover the NCP area. in a second step, we will use this ideal simulation to evaluate what we will obtain when cross-correlating the T16DPA observations with the spectroscopic catalog we are preparing, starting from the NCCS source catalog [ref]. RA: we should clarify the texte - SDSS is not an idealized case, we assume that the statistical distribution of galaxies near NCP is not very different from that of SDSS at mi-latitudes. We have then to check what is the fraction / and magnitude limits for which we get the redshifts from WYIN and see how much the cross correlation degrades

5.2.1 *With a rotated SDSS catalog*

For this first analysis, we start with the catalog described in section A. We rotate the source's sky coordinates in order to bring the $(\alpha, \delta) = (180, 45)$ deg direction in the original frame towards the NCP, thus getting an artificial covrerage of the NCP survey by the catalog. We use a pipeline similar to that used in the mid-latitude cross-correlaton study presented in section 5.1, with exception of the last stages. Given the range of declinations we simulated here (betgween 84 and 90 degrees) we restrict the skycube to projection to a circular area around the NCP, within a radius of 7 degrees. The spherical maps are projected into small square flat maps (169 pixels with $\sim$ 5arcmin resolution) through a gnomonic projection. A source cube with identical resolution is constructed from the rotated catalog, with the same prescription as for the midlatitude case (smearing in angular space and redshift spread according to velocity width) and as previously generate 100 shuffled source cubes for systematics checks.

To evaluate the cross-correlation between the simulated maps and the catalog derived maps, we compute 2D FFTs for each pair of maps of a given frequency plane. We average the amplitudes of the FFT modes in bins of angular wavemodes (azimutal average), which results in 1D amplitude vectors (one per frequency plane). We present in figure 17 averages of such spectra, in the second and third frequency intervals of table 4, 1300-1350 MHz, on the left panel and band 3, 1350-1400 MHz, on the right panel. We restrict the mode number range in these plots since, on the one hand, at low mode number (i.e. larger angular scales) maps are dominated by reconstructions systematics and at the other end, the signal is limited by the simulated instrumental angular resolution.

We compare the power spectra reconstructed for different simulation cases compared with the ideal case with only H_I sources and no noise (in red). In that case, we get a smooth and positive spectrum indicating a very significant (as judged from the dispersion obtained in the 100 shuffled realisations, shown in cyan) positive cross-correlation between the reconstructed map and those based on the catalog. We can observe that adding noise (blue curves) significantly degrades this at large scales (mode numbers below $\sim$ 5). Nevertheless, thew spectra amplitudes stay on the positive side, more significantly in the highest frequency case. The foreground residuals, observed when simulating all components wqithout adding noise in visibilities also affect the reconstructged spectra (shown as green curves in figure 17). This effect seems less strong that the addition of noise, which was also observed in the analysis of the surveys sensitivities presented in 3.3 : in the case of NCP, the dominant distorsion wrt to the ideal case comes from noise. Finally, with noise added, the spectra (shown in magenta) in both frequency intervals seem to stay significantly positive.

Figure 18 represents the evolution of a raw correlation coefficient, computed from the product of the two maps, as a function of frequency. Large values of the correlation coefficients computed for the ideal case (no foreground, no noise, in blue), appearing as peaks in the distribution, near 1387 MHz or 1370 MHz for example, might be interpreted as the sign of presence of non linear clustering, maybe sheets or filaments appearing in the galaxy distribution. Many of these features remain in the realistic case where all sky components and noise are included (orange circles), at leat for frequencies above $\gtrsim$ 1360 MHz.

Taking the dispersions extracted from the shuffled simulations at face value we can estimate the distance from a constant null value of the cross-correlation power spectrum. Doing this in each of the 3 frequency bins of table 4 we find a 10, 4 and $\lesssim$ 1 sigmas, for mode numbers between 5 and 12 or 15. We can therefore conclude that we would detect with a high significance a cross-correlation between a catalog with similar characteristics (completeness and sensitivity) than the SDSS covering the region surrounding the NCP, in at least the lowest redshift intervals we consider. While significant, the S/N seemds a bit less favorable than that of the cross-correlation we forecast for a mid-latitude survey. However, we note that as shown in section 3.3 noise is the sensitivity's main limitation and the noise per visibility sample we have used here is conservative, hence so is our forecast. On the other hand, the sky area covered by the NCP survey is much smaller than the mid-latitude case, which could explain part if not all the comparatively lower significance of the cross-correlation we forecast.

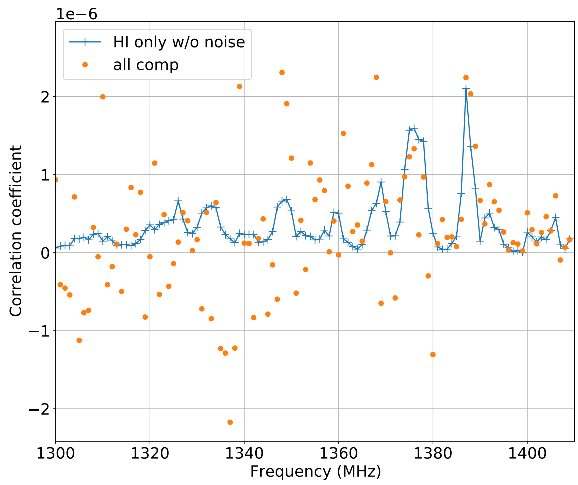

Figure 18. Raw correlation coefficient computed $\langle RecSky(\nu) \times Src(\nu) \rangle$, for each frequency plane. Blue curve (and crosses) correspond to the case where only H_I sources were contributing to the simulated visibilities, while orange circles correspond to the case where visibilities were computed from all sky components (foregrounds and H_I sources) and include 5mK noise.

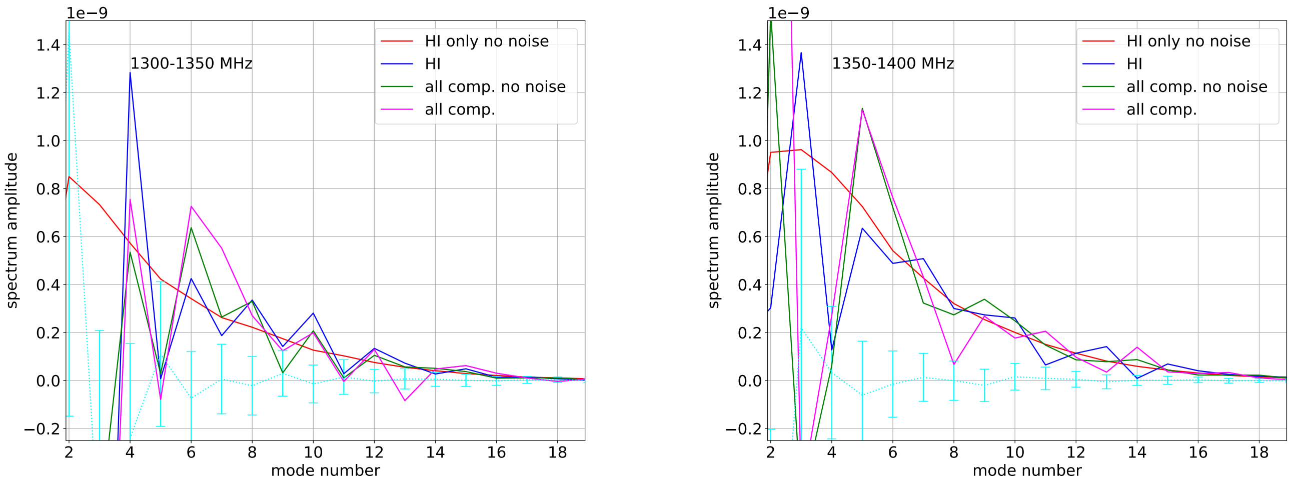

Figure 17. Reconstructed averaged spectra in the two frequency intervals corresponding to bins 2 (left) and 3 (right) from table 4. We compare results obtained from various simulation parameters, from the idelal (only HI sources, no noise) to the complete case. On both sides, the cyan results represent the averages and dispersions from the power spectra from the cross-correlations with a set of 100 shuffled catalogs.

5.2.2 What can we expect with the future Tianlai NCCS-based spectroscopic catalog ?

6 DISCUSSION

Discussion / Optimal strategy : NCP area coverage and redshift ranges, mi- latitude coverage and redshift ranges

Is there anything to say about direct detection at redshift about 0.4

APPENDIX A: FROM OPTICAL PHOTOMETRY TO RADIO PARAMETERS

In order to be able to simulate observations of a sky composed of diffuse and point-like continuum sources and an optical (of IR) galaxy 3D catalog we have to estimate the radio properties of the latter. We follow a two step procedure to achieve this. First, following Taylor et al. (2011) we estimate the stellar mass of each galaxy from their photometric properties using their equation 8 :

$$\log(M_\star/M_\odot) = -0.68 + 0.70(M_g - M_i) - 0.4M_i \qquad \text{(A1)}$$

This equation can easily be applied for SDSS objects, using their Photoz table that includes absolute magnitudes. From a cross-match between the ALFALFA and SDSS, Brown et al. (2015) investigated the relation between the HI and stellar masses of galaxies. From their figure 3, where results of stacking ALFALFA observations for the full SDSS sample are reported, we extract a simple relation between the HI and stellar masses :

$$\log(M_{HI}/M_\star) = 0.179 = 0.66(\log M_\star - 9.21) \qquad \text{(A2)}$$

Combining equations A1 and A2 we can now estimate the HI mass for each galaxy. Moorman et al. (2014) (their figure 7) show the relation between the HI mass and 50% flux velocity dispersion (W50), from which we infer a simple linear relation between these two parameters.

Using equation this and 18 we finally can estimate the 21cm peak flux and frequency width for each galaxy, after accounting for their distance as determined by their redshift within a fiducial cosmology.

We prepared a catalog for our simulations from the SDSS photo- and spectroscopic catalog, selecting sources above $\delta = 30$ deg satisfying basic photometric parameters to select galaxies. Redshift and HI mass distribution of these sources are sown on figure A.

The variations of their peak flux and frequency width is shown on figure

ACKNOWLEDGEMENTS

This research made use of Photutils, an Astropy package for detection and photometry of astronomical sources (Bradley et al. (2021)).

REFERENCES

Ansari R., et al., 2012, A&A, 540, A129
Ansari R., et al., 2020, MNRAS, 493, 2965
Baldwin J. E., Boysen R. C., Hales S. E. G., Jennings J. E., Waggett P. C., Warner P. J., Wilson D. M. A., 1985, MNRAS, 217, 717
Bandura K., et al., 2014, in Stepp L. M., Gilmozzi R., Hall H. J., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 9145, Ground-based and Airborne Telescopes V. p. 914522 (arXiv:1406.2288), doi:10.1117/12.2054950
Battye R. A., Davies R. D., Weller J., 2004, MNRAS, 355, 1339
Bharadwaj S., Nath B. B., Sethi S. K., 2001, Journal of Astrophysics and Astronomy, 22, 21
Bradley L., et al., 2021, astropy/photutils: 1.2.0, doi:10.5281/zenodo.5525286, https://doi.org/10.5281/zenodo.5525286

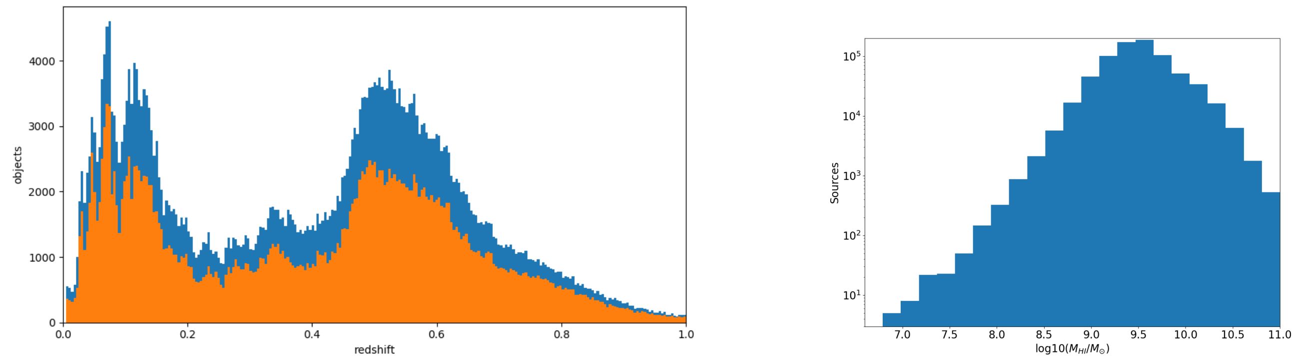

Figure A1. Redshift (left) and HI Mass (extrapolated from optical photometry) distributions from the SDSS catalog sources use in our simulations. The orange distribution correspond to our final source selection.

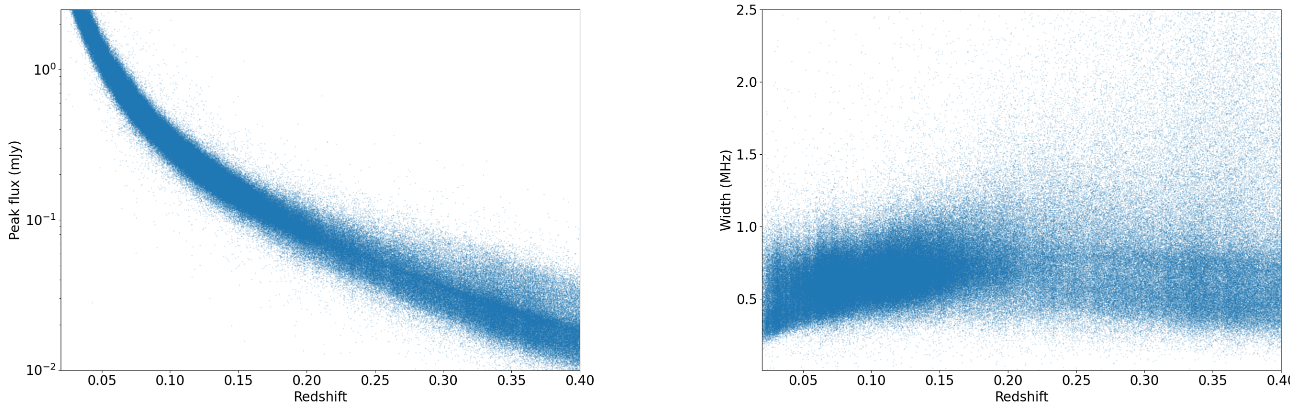

Figure A2. 21 cm flux and frequency width as a function of redshift for our simulated HI source set extrapolated from the SDSS catalogs.

Brown T., Catinella B., Cortese L., Kilborn V., Haynes M. P., Giovanelli R., 2015, Monthly Notices of the Royal Astronomical Society, 452, 2479âĂŞ2489

Bull P., Ferreira P. G., Patel P., Santos M. G., 2015, ApJ, 803, 21

CHIME Collaboration et al., 2022, arXiv e-prints, p. arXiv:2202.01242

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

Chang T.-C., Pen U.-L., Bandura K., Peterson J. B., 2010, Nature, 466, 463

Chen X., 2012, in International Journal of Modern Physics Conference Series. pp 256–263 (arXiv:1212.6278), doi:10.1142/S2010194512006459

Condon J. J., Cotton W. D., Greisen E. W., Yin Q. F., Perley R. A., Taylor G. B., Broderick J. J., 1998, AJ, 115, 1693

Cosmic Visions 21 cm Collaboration et al., 2018, arXiv e-prints, p. arXiv:1810.09572

Das S., et al., 2018, in Zmuidzinas J., Gao J.-R., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 10708, Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy IX. p. 1070836 (arXiv:1806.04698), doi:10.1117/12.2313031

Gorbikov E., Brosch N., 2014, Mon. Not. Roy. Astron. Soc., 443, 725

Haslam C. G. T., Klein U., Salter C. J., Stoffel H., Wilson W. E., Cleary M. N., Cooke D. J., Thomasson P., 1981, A&A, 100, 209

Hogg D. W., 1999, arXiv e-prints, pp astro–ph/9905116

Jones M. G., Haynes M. P., Giovanelli R., Moorman C., 2018, Monthly Notices of the Royal Astronomical Society, 477, 2âĂŞ17

Li J., et al., 2020, Science China Physics, Mechanics, and Astronomy, 63, 129862

Moorman C. M., Vogeley M. S., Hoyle F., Pan D. C., Haynes M. P., Giovanelli R., 2014, Monthly Notices of the Royal Astronomical Society, 444, 3559âĂŞ3570

Morales M. F., Wyithe J. S. B., 2010, ARA&A, 48, 127

Newburgh L., et al., 2016, in Ground-based and Airborne Telescopes VI. p. 99065X

O'Connor P., et al., 2020, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. p. 114457C (arXiv:2011.08695), doi:10.1117/12.2576250

Pritchard J. R., Loeb A., 2008, Phys. Rev. D, 78, 103511

Remazeilles M., Dickinson C., Banday A. J., Bigot-Sazy M. A., Ghosh T., 2015, MNRAS, 451, 4311

Seo H.-J., Dodelson S., Marriner J., Mcginnis D., Stebbins A., Stoughton C., Vallinotto A., 2010, ApJ, 721, 164

Shaw J. R., Sigurdson K., Sitwell M., Stebbins A., Pen U.-L.,

2015, Phys. Rev., D91, 083514
Smoot G. F., Debono I., 2017, A&A, 597, A136
Taylor E. N., et al., 2011, Monthly Notices of the Royal Astronomical Society, 418, 1587âĂŞ1620
The CHIME/FRB Collaboration et al., 2021, arXiv e-prints, p. arXiv:2106.04352
Vanderlinde K., et al., 2019, in Canadian Long Range Plan for Astronomy and Astrophysics White Papers. p. 28 (arXiv:1911.01777), doi:10.5281/zenodo.3765414
Villaescusa-Navarro F., Alonso D., Viel M., 2017, MNRAS, 466, 2736
Wolz L., et al., 2022, MNRAS, 510, 3495
Wu F., et al., 2021, MNRAS, 506, 3455
Zhang J., Ansari R., Chen X., Campagne J.-E., Magneville C., Wu F., 2016, MNRAS, 461, 1950

This paper has been typeset from a TEX/LATEX file prepared by the author.