MNRAS **000**, 000–000 (2021) Preprint 15 December 2021 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. A realistic simulation of the survey, starting from generation of mock visibility data according to the survey strategy, then map reconstruction followed by a simple foreground subtraction has 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 a few mK, and would be marginally impacted by the mode-mixing given the interferometer configuration. Tianlai would then be able to detect closeby, massive clumps, as well as clear cross-correlation signal at x-sigma with NCCS optical galaxies. We have also studied the performance of a mid-latitude survey, covering several thousand degrees, overlapping SDSS main survey footprint. Despite a higher noise level as well as significant distorsions 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 Intenisty Mapping is a promising technic to map the cosmological large scale matter distribution through the ob-

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

servation of 21cm radio emission/absorption of neutral hydrogen gas (H_I) , without 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). These 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 features 16, 6 meter diameter on-axis dishes, equipped with dual polarisation feeds, and arranged in a near hexogonal 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) , (Cosmic Visions 21 cm Collaboration et al. 2018) ,) or with SKA (Villaescusa-Navarro et al. 2017) or FAST (Smoot & Debono 2017).

However, the two Tianlai pathfinder instruments, specially 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 the very low redshift $z \lesssim 0.1$, 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.

left from the initial plan - remove once done

- HI intensity mapping
- Tianlai project , reference to the cylinder paper and dish array paper
- Reminding the main challenges : reaching the sensitivity through long integration time, amplitude and phase calibration when observing tn transit, and sperating the cosmological 21cm signal from the foregrounds : references, focus on the high-k analysis

The usual description of the paper structure

- Science reach of Tianlai dish array surveys, at low ($z \sim 0.1$) and medium($z \sim 0.3 - 0.5$) redshifts, targeted toward restricted area
- NCP region , $5-100\text{deg}^2$ area , 2-5 mK visibilities noise level (1MHz x 30s sampling, $\sim$ 1 month observation per declination)
- Mid-latitude (near CasA declination, to overlap with SDSS legacy spectroscopic surver , $1000 - 2000\text{deg}^2$ area, $100 - 200\text{deg}^2$ area overlap with SDSS
- Detection of nearby $z \lesssim 0.05$ HI clumps : reliable estimates of number of detectable clumps (mass & redshift distribution)
- Detection of LSS in cross-correlation with optical survey
- Possible detection of LSS as excess auto-correlation signal ?

Some 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 $\propto d_L(z)^-2$ where $d_L(z)$ is the luminosity distance as a function of redshift. However, one might argue that Intensity mapping do not observe individual sources, but aggregate emissions from neutral hydrogen $100-1000\,\mathrm{Mpc}^3$ 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 $0.25^\circ - 0.5^\circ$, 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 it would 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 quickly as redshift increases, due to the way the radio interferometer noise projects on sky. Indeed, a radio instrument, single dish or interferometer noise power spectrum, projected on sky as $P_{\rm noise}(k)$ scales as: (see for example (Ansari et al. 2012), section 3.2)

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

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

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 $\delta\theta \propto \frac{\lambda}{D_{\rm array}}$ where $D_{\rm array}$ 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 $\delta V = a_\perp^2 \times a_\parallel$:

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

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

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

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

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

Map pixel flutuations due to instrumental noise, denoted σ_T^2 and characterised by the system temperature $T_{\rm sys}$ can be easily related to the noise power $P_{\rm noise}$. 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:

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

$$\sigma_T^2 \simeq P_{\rm noise}\iiint dk_x dk_y dk_z$$

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

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

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

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:

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

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

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 $\nu \sim 1400\,$MHz, corresponding to angular scales $2\pi/\ell$ and transverse wave number (see section 3). 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:

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

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

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

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.

Figure 1 shows the radio sky near the NCP (North Celestial Pole), as it appears at 135 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 `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 radiogalaxy 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.

- Explore also some aspects of component separation (foreground subtraction)

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

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

Table 1. The comoving radial distances, cosmolgical transverse and radial cell size corresponding to angular cell size $0.2^\circ\times0.2^\circ\times$ 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.

- Discuss the way instrument noise (radiometer equation) project on sky - when observations are discussed as cosmolgical power spectrum $P(k)$: variations with reshift - Discuss also the accessible k-range (wave-number) - depending on the survey area and instrument configuration ($k_\perp k_\parallel$).
- Present the the three cosmological signals we might aim for : direct detection of HI clumps at very low redshifts ($z < 0.05$), cross-corerlation with optical surveys and possible LSS detection in auto-correlation at $z \sim 0.4$.

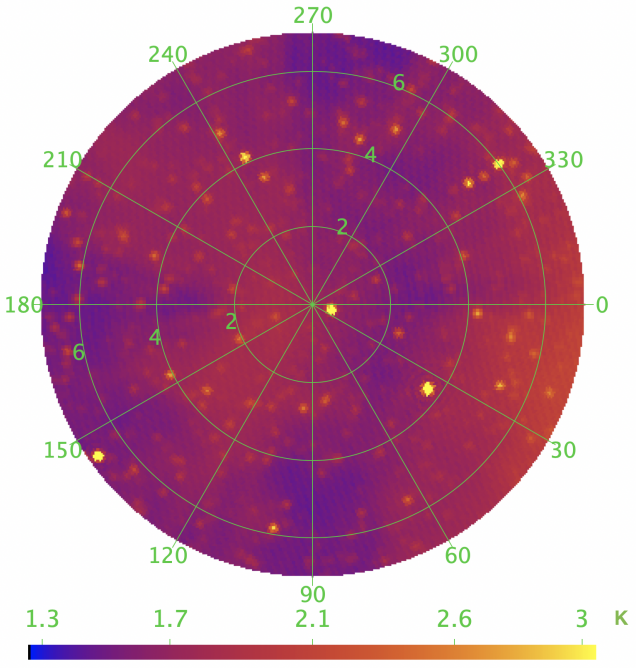

Figure 1. Foreground map of a 7 deg. radius region around NCP at 1350 MHz, smoothed with a 15arcmin resolution gaussian beam. Haslam map of diffuse emission at 408 MHz as well as NVSS radisources, extrapolated to 1350 MHz with a spectral index $\beta = -2$. Color scale corresponds tp temperature in Kelvin.

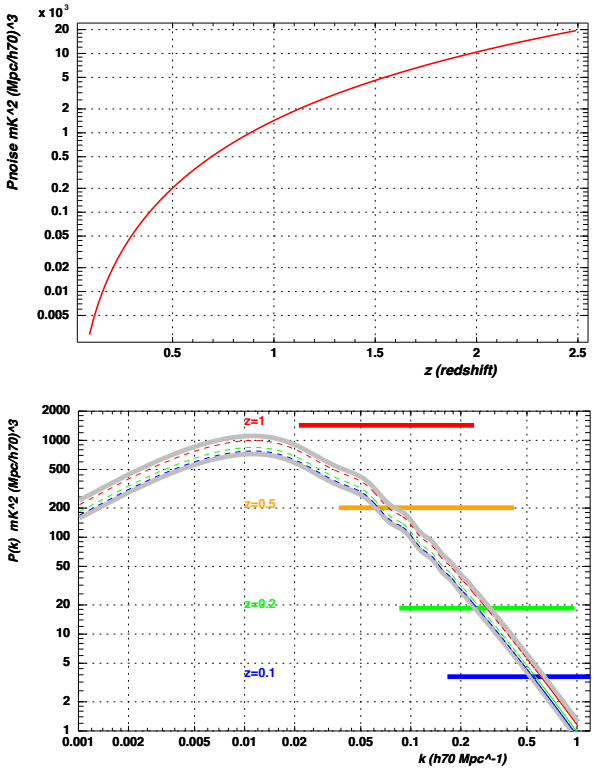

Figure 2. Top: White noise power spectrum $P_{\rm noise}(z)$ level as a function of redshift for a survey similar to T16DPA with an angular map pixel size of 0.2° and 1MHz, and for per pixel noise level $\sigma_T^2 = 10{\rm mK}^2$. Bottom: Projected noise power spectrum $P_{\rm noise}(k)$ and the accessible 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 simulated H_I 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 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 $\delta = 90°, 88°, 86°, 84°$, 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, $\delta > 83°$, 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 86 constant declination scans at $\delta = 49°, 51°, 53°, 55°, 57°, , 59°$, covering a band in declination $48° \leq \delta \leq 60°$, representing about $\sim 12\%$ of the sky or $\sim 2500{\rm deg}^2$. 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 $\sim 10{\rm days}$ (more precisely $8.5 \times 24{\rm hours}$ spent on each constant declination scan, corresponding to a total integration time $t_{int} = 8.5 \times 30 \simeq (256 = 16^2){\rm s}$ per visibility samples, leading to a a noise level :

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

Tianlai day time data is contaminated by the sun signal leaking through far side lobes, it is therefore planned to only use night time data. T16DPA should thus be able to reach this 5mK by spending 2 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 spending a whole year on the NCP, to get 2×40 days per declination observation time, so a total of $2 \times 4 \times 40 = 320$ days.

For each frequency plane, a spherical map is reconstructed through m-mode map making, or rather, the map making equation 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 through parameters which defining ratio $r_{\rm PSI}$ of

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

smallest to largest eigenvalue for each inversion, as well as a absolute threshold on the minimal eigenvalue $\lambda_{\rm PSI}$. The two values has been set to $r_{\rm PSI} = 0.02$ and $\lambda_{\rm PSI} = 0.001$, which can be considered as medium level.

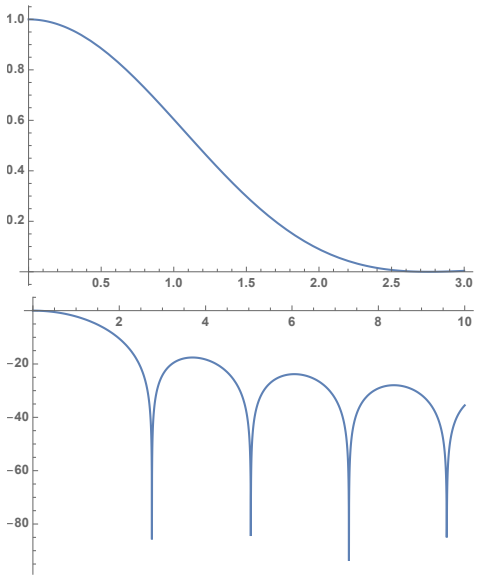

Figure 3. Beam response (Bessel J_1) as a function of the polar axis θ in degrees, with respect to the beam axis (antenna pointing direction) at $f = 1350$ MHz. top: main beam, in linear scale; bottom: main beam and side lobes, in logarithmic scale (dB).

We have used spherical maps with a resolution of 5arcmin for the reconstructed map, 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 haven used a the `SphereThetaPhi` pixelisation scheme, which have nearly 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 perform also an additional 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 qaulity and hence the noise level increases for angular resolution corresponding to the size, for $\ell \gtrsim 850$ for the NCP survey. We have smoothly damped a_{ellm} coefficients for $\ell \lesssim 75$ and $\ell \gtrsim 875$. A gaussian filter with $\sigma_\ell = 750$ have also been applied, and all $m = 0$ modes have been put to zero. This last filter is intended to remove wiggles with near perfect asymmetry which appears due to the partial sky coverage combined with limited sensitivity range in ℓ.

Figure 4 shows an example of 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 $(2-4$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.

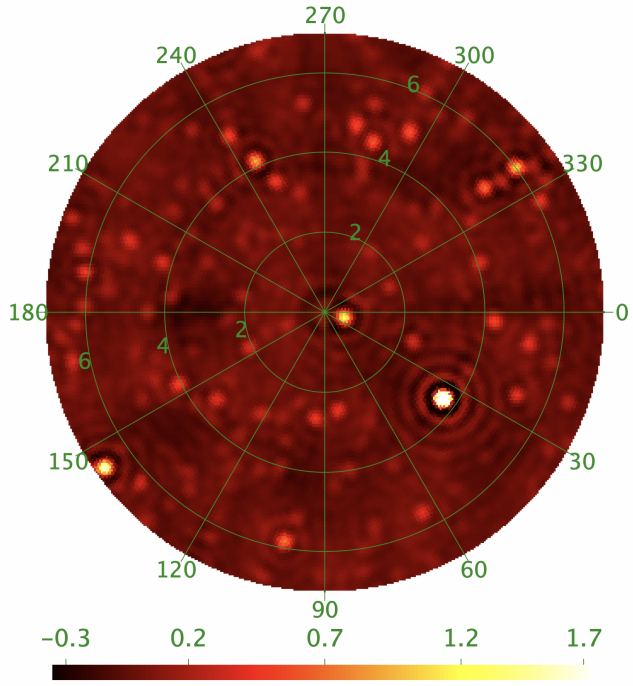

Figure 4. Reconstructed map of the NCP region, as observed by T16DPA at $f = 1350$MHz. This is a the 7 deg. radius around $\delta = 90°$, extracted from the reconstructed spherical map using m-mode map making and after (ℓ, m) space filtering.

Average reconstructed sky angular power spectrum is shown in figure 5, corresponding to temperature fluctuations about ~ 1K. The sky power spectrum is larger at larger angular scales. Effet of the instrument and map making ℓ-space response is visible $\ell \lesssim 50$ 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 sample. The very large increase of the noise level above $\ell \gtrsim 800$ due to the instrument response and map reconstruction, as well as the effect of $\ell, m)$ filetring on the $C_{\rm noise}(\ell)$ can be seen on the orange (before filtering) and red (after filetring) power spectra.

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

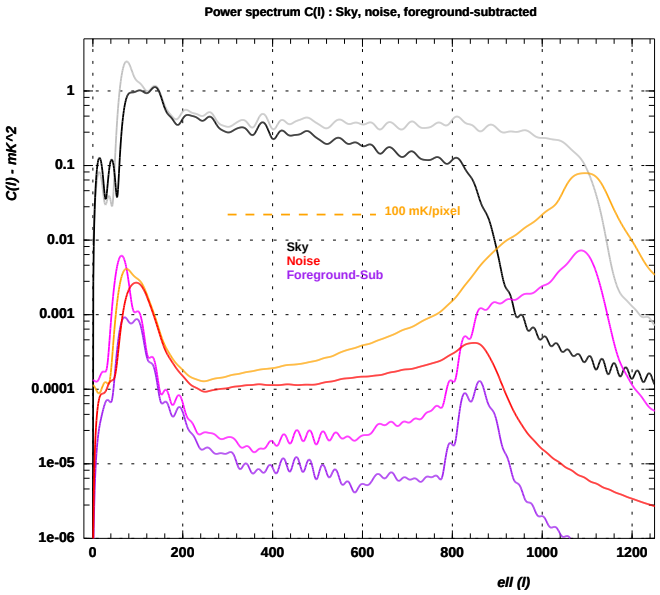

Figure 5. Average angular power spectrum $C(\ell)$ from the a data cube of 100 reconstrcuted maps of the NCP region, covering an area with 7 deg. radius around NCP, and the frequency range 1300-1400 MHz. The reconstructed sky power spectrum is shown in black, noise power spectrum in red and residual 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.

3.2 Foreground subtraction

Our aim here is not to devise the best foreground subtraction method, but rather characterise the instrument ans 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:

$$T^{\rm fgnd-P}_{\alpha,\delta}(\nu) = A\alpha,\delta\nu^2 + B\alpha,\delta\nu + C\alpha,\delta \quad (13)$$

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

The second method (**DF**) is a very simple differential 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.

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

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

We have represented on figure 7 the angular power spectrum of the residual signal $C_{\rm res}(\ell)$, after foreground subtraction for the two methods (P,DF), using reconstructed maps and (ℓ, m) plane filetring for the NCP case. Compared to the input sky (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 differential along frequency (DF) method, corresponding to a factor ~ 150 (P) and ~ 250 (DF) damping in amplitude for temperature fluctuations due to foreground. While this is not enough for cosmological 21cm signal, the foreground residuals from foreground subtraction due to mode mixing would be well below the instrumental noise level for the NCP survey by Tianlai. However, as it can be seen from figure 7 the residual from foreground subtraction for the mid latitude survey reach a level comparable to the instrumental noise, despite a significantly higher noise level.

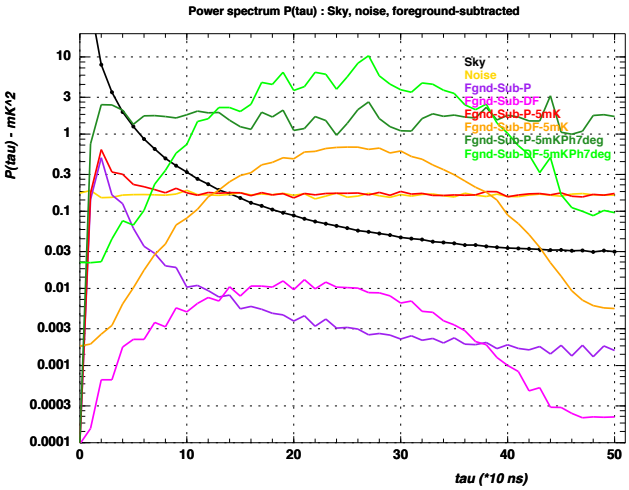

Figure 8. Angular power spectrum $C(\ell)$ computed from a set of three frequency maps, using the DF method for the mid-latitude survey.

left from the original paper plan - remove once done

- Describe the visbility simulation and map reconstruction process
- Describe and discuss the two foreground subtraction method used here
- per pixel noise level (visibility space and map space)
- Impact of imperfect calibration - (phase and amplitude calibration errors)
- illustrate for the NCP case, as well as lower latitude case

4 H_I CLUMP DETECTION

mass distribution and effective detection thresholds (based on simulation including foregrounds and radio-sources , followed by map-making , or in visibility space)

The aim of this analysis is to assess the number of direct detection of 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 HI clumps detection efficiency as a function of their apparent flux for

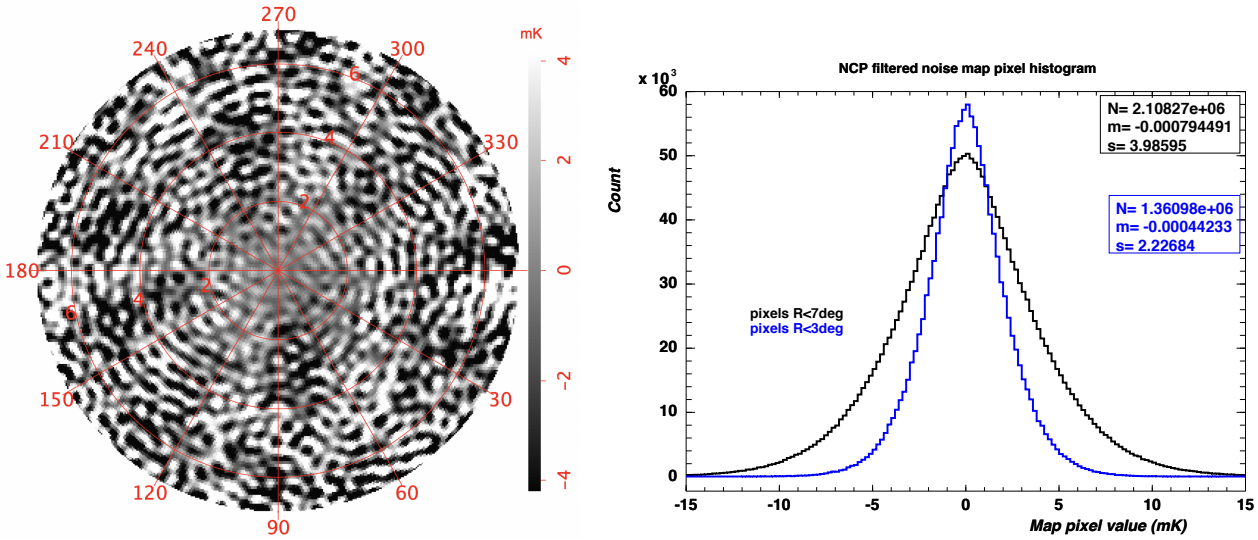

Figure 6. Left: Noise map after reconstruction with (ℓ, m) filetring of the NCP region covering an area with 7 deg. radius at $f = 1350$MHz (map scale in mK). Right: Noise map pixel value distribution, in black for the full 7 deg. radius map around NCP, and blue, restricted to the central 3 deg. radius, covering $\sim 30\mathrm{dg}^2$. 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

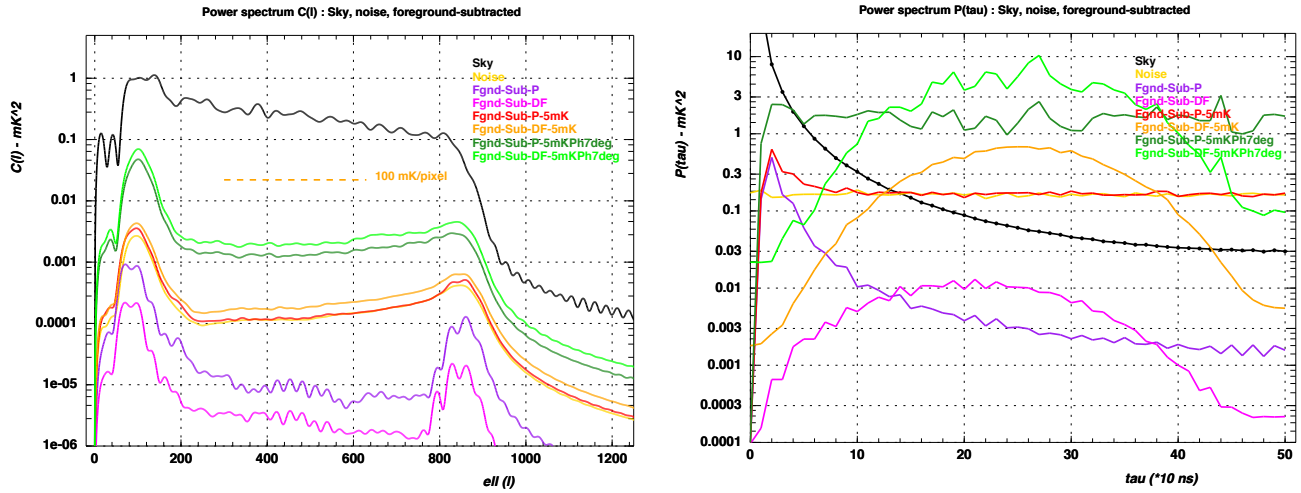

Figure 7. Left: average angular power spectrum $C(\ell)$ from the a data cube of 100 reconstructed maps, with (ℓ, m) filetring of the NCP region covering an area with 7 deg. radius around NCP, and the frequency range 1300-1400 MHz. Right: average power spectrum along the frequency axis $P(\tau)$. The 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.

the NCP and mid-latitude cases. In a second step we combine these detection thresholds in the flux domain with the HI mass function to extract the number of HI that could be detected, assuming a spatial uniform random distribution, for the NCP and mid-latitude cases.

4.1 H$_\mathrm{I}$ clumps flux detection thresholds

To assess the detection efficiency for point-like HI sources we have used a pipeline sharing most of the ingredients 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. For each frequency we compute simulated visibilities, project them into sky maps and then filter them, as explained in 3.1. The noise level per visibility sample (30s integration time) used in the following is 5mK. To roughly account for foreground subtraction, we subtract from the central frequency map the average of the two outer frequency ones. Finally, we reproject the difference into rectangular (mid-latitude case) or square (NCP case) maps using simple projections. For the mid-latitude difference maps, we include in the pipeline a

loose high-pass filter in spacial frequency domain to reduce some foreground subtraction and projection residuals.

The final step is the sources detection. We use a basic scheme based on the `DAOStarFinder` class from the `photutils` Python package (Bradley et al. 2021). As the map foreground subtraction and map reconstruction and projection leave some artefacts in the residual maps, we set loose sphericity criteria to this detector. We set the detection threshold relatively to the map's RMS, at 7 (NCP) and 10 (mid-latitude)times this value to avoid spurious detections on reconstruction-induced high spacial frequency residuals. To assess the deetection efficiency, we count 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 an 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. Thansk to the higher integration time per map pixel in the NCP case, these results show that the detection threshold is much lower in the NCP case wrt the mid-latitude one : about .1 and 1.2 Jy, respectively. We have in each case fitted the detection efficiency vs flux variations with an error function ; these fits will be used in the computation of the expected number of HI 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 be randomly spatially 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 HI clumps - the number density of galaxies falling in a logarithmic HI mass bin, is well described by a Schechter function :

$$\phi(M_{HI}) = ln(10)\phi^* \left(\frac{M_{HI}}{M_*}\right)^{\alpha+1} \exp\left(-\frac{M_{HI}}{M_*}\right) = \frac{dN_{ga}}{dV\, d\log_{10}(M_{HI})} \quad (17)$$

The parameters in this equation are ϕ^* (normalisation), M_* ('knee' mass ; we use in the following the reduced parameter $m_* = \log(M_*/M_\odot)$ and α , defining the low-mass slope of this distribution. 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 ($v_{CMB} < 4000 km/s$), 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.

The relationship between the HI mass and flux is also reported in Jones et al. (2018) :

$$\frac{M_{HI}}{M_\odot} = 2.356\ 10^5 D^2_{Mpc} S_{21} \quad (18)$$

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 fitted using the whole ALFALFA dataset and its near subset

Dataset	NCP	mid-latitude
Surface (deg^2)	100	1500
HI clumps/deg^2 (full MF)	.048	.0012
HI clumps/deg^2 (near MF)	.035	.0009

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. We also give an order of magnitude of the sky surface covered in each of these surveys.

where D_{Mpc} is the distance to the source 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 $\sim$ $210 km/s$ velocity width (corresponding to $\sim$ 1 MHz in the frequency domain) we can translate the flux limits or detection efficiencies in Jy into HI mass limits or detection efficiencies a each redshift. The integral of the HI mass function convolved with such detection efficiency gives the number of expected detections at any given redshift ; the sum of these contributions finally gives the total number of expected detections.

We present the expected number of HI clump detections per squared degree for the NCP (left-) and mid-latitude case (right hand side) as a function of redshift, assuming the HI 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 shown on figure 4.2 the expected redshifts of the HI clumps detected are very low : below .02 for the NCP case, and .005 in the mid-latitude case. This is at first order irrespective of the HI mass function parameters used. At these very low redshifts it may therefore be justified to use the 'near' HI 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, the larger dilution of the observations 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 achived by observing over longer periods. This would increase the number of expected HI clumps, but only in the NCP case where the noise dominates the reconstruction artefacts in the reconstructed maps.

5 CROSS-CORRELATION WITH OPTICAL GALAXY CATALOGS

discuss LSS in cross correlation with optical surveys , at mid-latitude, with SDSS , NCSS for NCP :

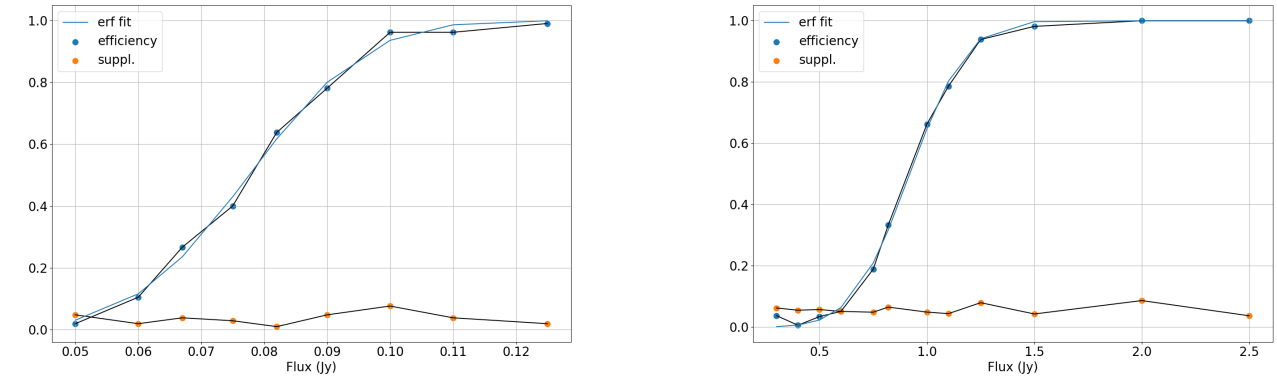

Figure 9. HI clump detection efficiences as a function of flux, for the NCP (left-) and mid-latitude (right hand side) measured by our simulations. On each part we represent as blue dots the efficiences measured at each simulated flux. the read dots correspond to the number of spurious detection (detections located farther than 2 pixels from the simulated clumps positions). The cyan curve is a fit of the efficiencies values with an error function.

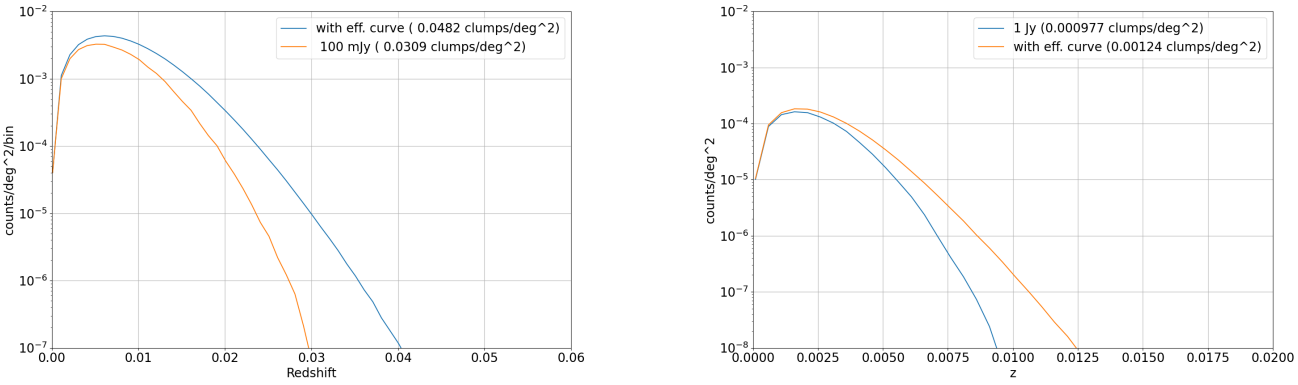

Figure 10. Expected number of detections vs redshit for the NCP (left-) and mid-latitude (right hand side) cases. The integral numbers of detections are reported in the caption of the figures. In both cases we indicate the results we obtain with a detection threshold on the flux, in Jy, and with the efficiency curves shown on figure 4.1.

- *discuss the possible scenarions : NCP, mid latitude , cross correlation with SDSS ,*
- *show optical catalog redshift distribution.*
- *effect of incomplete spectroscopic catalogs*
- *effect of redshift errors*

In this section we explore the prospects for extracting the cross-correlation signal between the Tianlai low redshift survey. We begin by outlining the path from the galaxy parameters as recorded in the optical catalogs to their expected radio parameters. We then turn to the most straightforward case, extracting the cross-correlation between a Tianlai mid-latitude low redshift survey and the SDSS catalog, which overlaps its sky footprint. For the NCP case, the SDSS catalog does not cover this area. We are in the process to construct a spectroscopic catalog based on the NCCS catalog (Gorbikov & Brosch 2014), and we first try tu use it to forecast its cross-correlation with Tianlai observations towards NCP. To complement this, we also built an artificial catalog by rotating the coordinates from the SDSS catalog to get an overlap with observations towards NCP, which would consist in a more complete cata;log than the current state of the NCCS.

5.1 From optical photometry to radio parameters

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

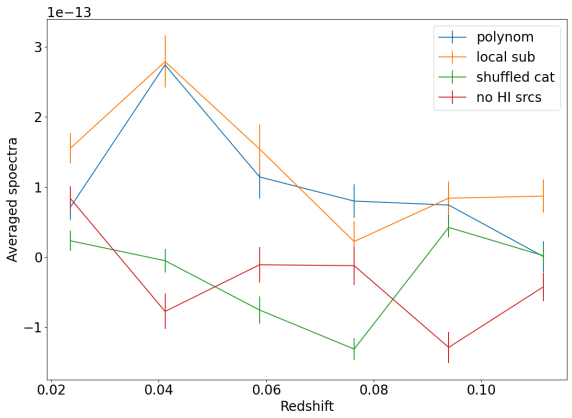

Figure 11. Average of the cross-correlation C_ℓ in the interval [X,Y] for ...

5.3 Cross-correlation towards NCP

5.3.1 With NCCS

5.3.2 With a rotated SDSS 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

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
Bull P., Ferreira P. G., Patel P., Santos M. G., 2015, ApJ, 803, 21
Chang T.-C., Pen U.-L., Peterson J. B., McDonald P., 2008, Phys. Rev. Lett., 100, 091303
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
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
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
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.