Research Paper

The FAST Ultra-Deep Survey (FUDS): Observational strategy, calibration and data reduction

Hongwei Xi1,2,3,4, Bo Peng1, Lister Staveley-Smith2,4, Bi-Qing For2,4 and Bin Liu1,4

1CAS Key Laboratory of FAST, National Astronomical Observatories, Chinese Academy of Sciences, Datun Rd., Chaoyang District, Beijing 100101, China, 2International Centre for Radio Astronomy Research (ICRAR), University of Western Australia, 35 Stirling Hwy, Crawley, WA 6009, Australia, 3School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing 100049, China and 4ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Melbourne, Australia

Abstract

The FAST Ultra-Deep Survey (FUDS) is a blind survey that aims for the direct detection of HI in galaxies at redshifts $z < 0.42$. The survey uses the multibeam receiver on the Five-hundred-metre Aperture Spherical Telescope (FAST) to map six regions, each of size 0.72 deg$^2$ at high sensitivity ($\sim 50 \mu$Jy) and high-frequency resolution (23 kHz). The survey will enable studies of the evolution of galaxies and their HI content with an eventual sample size of $\sim 1000$. We present the science goals, observing strategy, the effects of radio frequency interference at the FAST site, our mitigation strategies and the methods for calibration, data reduction and imaging as applied to initial data. The observations and reductions for the first field, FUDS0, are completed, with around 128 HI galaxies detected in a preliminary analysis. Example spectra are given in this paper, including a comparison with data from the overlapping GAL2577 field of Arecibo Ultra-Deep Survey.

Keywords: galaxies: evolution – galaxies: high-redshift – surveys – radio lines: galaxies – methods: observational – telescopes

(Received 2 November 2021; revised 4 April 2022; accepted 4 April 2022)

1. Introduction

Hydrogen is the most abundant element in the Universe. In galaxies, hydrogen is found in three primary forms: neutral atomic (H I), neutral molecular (H$_2$), and ionised (H II). H I is the most common phase of hydrogen found within late-type galaxies (discs or irregulars). H I observations are also sensitive to low-surface brightness, but gas-rich and dark-matter-rich galaxies, which are useful for studying dark matter concentration and constraining models of galaxy feedback and evolution. H I observations, if extended over a significant redshift range, can provide valuable information on how the environmental and intrinsic factors affect galaxy formation and evolution over cosmic time.

Surveys with large sky coverage will detect more galaxies in a given total survey time. Large numbers of detections have therefore been made by the well-known shallow H I surveys, namely the southern H I Parkes All Sky Survey (HIPASS; Barnes et al. 2001; Meyer et al. 2004) and the northern Arecibo Legacy Fast ALFA survey (ALFALFA; Giovannelli et al. 2005). HIPASS yielded 4315 H I galaxies over 21341 deg$^2$ of southern sky by utilising a 13 beam receiver on the Parkes telescope. ALFALFA employed the 7-beam ALFA receiver on the Arecibo telescope to detect nearly 31500 H I galaxies over 70000 deg$^2$ of northern sky. Whilst both surveys provided unprecedentedly large samples for statistical studies of HI in galaxies (Zwaan et al. 2005; Jones et al. 2016b, 2018), they could not provide evolutionary information due to their limited depth (HIPASS, $z < 0.04$; ALFALFA, $z < 0.06$). New H I surveys such as the Widefield ASKAP L-band Legacy All-sky Blind survey (WALLABY; Koribalski et al. 2020) will survey the southern sky, and the Commensal Radio Astronomy FAST survey (CRAFTS, Li et al. 2018; Zhang et al. 2019) which will survey the northern sky, will continue this legacy and dramatically increase the sample sizes in the local Universe (both surveys will, in practice, be sensitivity-limited to $z < 0.1$).

Medium-deep H I surveys such as the ASKAP Deep Investigation of Neutral Gas Origins (DINGO; Meyer, Heald, & Serra 2009) and the MeerKAT International GHz Tiered Extragalactic Exploration (MIGHTEE-HI, Jarvis et al. 2016), have been proposed in order to target areas of 10’s to 100’s of square degrees out to deeper redshifts. However, in order to truly explore the H I evolution of galaxies, smaller fields and higher sensitivity are required. Targeted or pencil beam surveys have been the preferred solution given limited observation time. Recent examples include the HIGHz project which employed the Arecibo telescope to search for HI in 39 late-type galaxies at $z \sim 0.2$ (Catinella & Cortese 2015). They did not find significant evolutionary trends between high-$z$ late-type galaxies and nearby galaxies. The Arecibo Ultra-Deep Survey (AUDS; Freudling et al. 2011; Hoppmann et al. 2015) was carried out to blindly search for H I galaxies at $z < 0.16$ in a limited 1.35 deg$^2$ area. Based on the final AUDS100 analysis, Xi et al. (2021) found evidence for the characteristic H I mass ($< 10^8$ solar mass) increasing with redshift. The ongoing COSMOS H I large extragalactic survey (CHILES; Fernández et al. 2016) employs Very Large Array (VLA) to map $\sim 0.5$ deg$^2$ in the COSMOS

Corresponding author: Bo Peng, email: bpeng@nao.cas.cn.


© The Author(s), 2022. Published by Cambridge University Press on behalf of the Astronomical Society of Australia.
field up to a higher redshift limit, $z = 0.45$. In the near future, the Looking At the Distant Universe with the MeerKAT Array (LADUMA, Blyth et al. 2016; Baker et al. 2018), will map $\sim 2$ deg$^2$ with unprecedented sensitivity and larger redshift coverage ($z < 1.4$).

The world’s largest single-dish telescope, the Five-hundred-metre Aperture Spherical Telescope (FAST, Nan et al. 2011), having been operational for astronomical observations since 2019, is also available to explore the distant Universe with high sensitivity. Although its individual beam size is small, it is equipped with a 19-beam receiver at L-band, which dramatically increases its ability to survey the sky in H I. Moreover, unlike the Parkes and Arecibo multibeam receivers, the FAST multibeam receiver can extend up to a redshift of $z \sim 0.4$. Therefore, a key scientific goal of FAST is to search for H I in galaxies up to this redshift. In this paper, we introduce the FAST Ultra-Deep survey (FUDS), which will map six 0.72 deg$^2$ regions at high sensitivity at Declinations around $+25^\circ$. The approximate on-source observing time is about 100 hr for each target area, which corresponds to a sensitivity of $\sim 50 \, \mu$Jy. We estimate that FUDS will directly detect $\sim 1000$ galaxies, and enable the studies of massive galaxies at high redshift as well as faint objects around Milky Way. The volume is sufficient to robustly study the evolution of the H I mass function and the cosmic H I density, $\Omega_{\text{HI}}$.

In this paper, we introduce the scientific goals of FUDS in Section 2. The FAST telescope, region selection and observation strategy are described in Section 3. We present a study of the RFI at the FAST site in Section 4 and the calibration method in Section 5. Using our pilot observations of the first two surveys fields (FUDS0 and FUDS1), data reduction of FUDS is introduced in Section 6. The impact of confusion is discussed in Section 7. We present the preliminary results from the first target field, FUDS0, in Section 8. Finally, a summary is given in Section 9.

2. Scientific goals

FUDS aims at the direct detection of 'high-redshift' galaxies in H I in a blind survey, and faint low-surface brightness features around nearby galaxies and filaments. The target sensitivity is 50 $\mu$Jy. Based on expectations from the H I mass function, and initial results, FUDS is expected to find around 1 000 H I galaxies and will address the following science questions.

2.1. The cosmic gas density

The cosmic gas density, $\Omega_{\text{HI}}(z)$, is a measure of how the global supply of cold gas available for star formation changes with redshift or cosmic time. Whilst the cosmic star formation rate is known to have decreased by an order of magnitude from redshift or cosmic time. Whilst the cosmic star formation rate supply of cold gas available for star formation changes with

$$\frac{\Omega_{\text{HI}}}{\Omega_1} \left( 0.1 \right) \text{ remain stubbornly significant.}$$

Recent semi-analytic (Kim et al. 2015) and hydrodynamic (Davé et al. 2017) models have now been adjusted to reflect this slower evolution, but there remain modest differences between models at $z < 1$ and enormous divergence at $z > 1$ (see Figure 19 in Xi et al. 2021). Moreover, the observational and systematic errors in measurements of $\Omega_{\text{HI}}(z > 0.1)$ remain stubbornly significant. The main methods available to measure cosmic H I density are: (1) direct H I spectral-line measurements of individual galaxies in a blind survey; (2) spectral stacking using optical redshifts; (3) intensity mapping; and (4) damped Lyman-$\alpha$ (DLA), or proxy measurements. Of these, the first is strongly preferred but, owing to the weakness of the 21-cm line, it has been difficult to accomplish at $z > 0.1$. Most of $\Omega_{\text{HI}}$ values at $z > 0.1$ (which are much less accurate than for $z < 0.1$) use stacking (Lah et al. 2007; Delhaize et al. 2013; Rhee et al. 2018), intensity mapping (Chang et al. 2010; Masui et al. 2013) or damped Lyman-$\alpha$ (DLA) (Rao & Turnshek 2000; Prochaska & Wolfe 2009; Grasha et al. 2020) methods. With the FUDS, we plan to perform the most accurate measurements to date for $\Omega_{\text{HI}}$ at $z > 0.1$, without reference to optical surveys. Stacking methods will allow accurate measurements in the range $z = 0.2$–0.4, but with a strong dependence on the selection and assumed evolutionary parameters of the corresponding optical input catalogue.

2.2. The evolution of galaxies and their H I content

The study of the cold gas (and stellar) content of individual galaxies, and the effect of environment on gas supply and star-formation rate, is a key to better understand galaxy evolution. Again, due to the difficulties of individual galaxy detection, this has been challenging. Catinella & Cortese (2015) measured the H I content in 39 massive late-type galaxies at $0.17 < z < 0.25$. Compared with scaling relations of galaxies at $z = 0$, they found that these galaxies are the analogues of extreme objects in the local Universe. As above, stacking techniques (Chengalur, Braun, & Wieringa 2001; Zwaan, van Dokkum, & Verheijen 2001) can shed more light on the mean properties of galaxies over a larger redshift range. For example, Rhee et al. (2018) measured a higher H I mass-to-luminosity ratio for distant late-type galaxies as compared to samples at lower redshift. However, the requirement to use optical/IR samples adds extra measurement uncertainties. More importantly, it adds selection effects, such as imaging and spectroscopic incompleteness (Jones et al. 2006), and systematic uncertainties due to redshift-dependent luminosity evolution (McNaught-Roberts et al. 2014) and k-corrections (Lake & Wright 2016), and possible environmental dependence (McNaught-Roberts et al. 2014). There is also a dependence on optical luminosity and/or luminosity density functions which may have been derived from different optical samples. Perhaps the most important of these additional uncertainties is the tendency for optical catalogues to miss the substantial population of H I-rich but optically low-surface brightness galaxies. This tendency is strongly reflected in the steeper low-mass slope in the H I mass function compared with the optical luminosity or stellar mass functions (Butcher et al. 2018).

FUDS is intended to be deep enough to detect individual galaxies right up to the survey limit of $z \sim 0.4$, and in sufficient statistical numbers to allow an accurate study of the evolution of galaxies and their gas content. However, given the large physical size of the FAST beam at higher redshifts, confusion is a possible problem (see e.g. Jones et al. 2016a). We will briefly discuss the impact of confusion on our results in Section 7.

2.3. The evolution of the H I mass and velocity functions

As statistical quantities which relate to the number density of galaxies of different gas and halo mass, the HI Mass Function (HIMF) and velocity width function (HIWF) are important constraints for cosmological models. Thanks to the large sample sizes now available, the form of the HIMF and HIWF and their Schechter parameterisation (Schechter 1976) have been greatly refined (Jones et al. 2018; Oman 2022). However, the HIMF beyond the local Universe is poorly known. Using the Arecibo Ultra-Deep Survey (AUDS), Hoppmann et al. (2015) did not
find any evolutionary trend in the preliminary AUDS60 sample. However for the full sample, Xi et al. (2021) found weak evidence for a ‘knee’ mass which increases with redshift. This conclusion was drawn based on a sample of only 247 galaxies. With its larger sample size, and wider redshift range, FUDS will put more accurate constraints on evolutionary trends.

2.4. Environment dependence

Many studies have explored the environmental dependence of the HIMF. For example, does the density ratio of low-mass and high-mass galaxies increase or decrease with local density? No common trends have yet been established, either for the slope of the low-mass end of the HIMF (α) or the value of the characteristic mass (M*). Rosenberg & Schneider (2002) found a flatter α in an intermediate density region around the Virgo cluster compared with the field. Busekool et al. (2021) also found a flat slope, α = −0.92, in dense regions using H I detections in the Ursa Major cluster. In contrast, Pisano et al. (2011) found flatter α in dense regions based on the full HIPASS sample. None of these studies found any change in M*.

Springob et al. (2005) used ~3 000 AGC galaxies to derive the HIMF and found constant α and decreasing M* in dense regions. Utilising the 40% and 70% ALFALFA samples, Moorman et al. (2014) and Jones et al. (2016b) discovered positive correlation between M* and density, respectively. But, neither found any obvious change in α. Jones et al. (2018) derived a steeper α and higher M* for non-cluster high-density regions from the 100% ALFALFA sample. Said, Kraan-Korteweg, & Staveley-Smith (2019) found steeper α values in dense regions based on the full HIPASS sample. None of these studies found any change in M*.

The Bandwidth of the ROACH2 is configured to be 500 MHz (frequency range 1.0–1.5 GHz), split into a maximum 1 048 576 channels with a maximum frequency resolution of 476.8 Hz. For our observations, we use 65 536 channels, which gives a frequency spacing of 7.63 kHz, equivalent to a velocity spacing of 1.61 km s⁻¹ in the rest frame. No Doppler shift is applied to spectra during observations. This allows us to search for H I in galaxies with cz in the range of −15 908 to +126 034 km s⁻¹. Spectra can be recorded at maximum rate of 10 Hz, but we chose the lower rate of 1 Hz.

For each beam, three linear polarisation signals (XX*, YY* and XY*) are recorded in fits format files, each with a maximum size of 2 GB. No Doppler shift is applied to the raw spectra before the data is recorded. At this stage, the metadata includes the position of

Figure 1. The geometry of the 19 beams of the FAST multibeam receiver is shown on the lower right corner. The black solid circles show the beam size of 2.9 arcmin at 1.4 GHz. The angular distance between nearby beams is 5.9 arcmin, and the spacing between nearby rows is 5.11 arcmin. The index of each beam is given in the circles. The light blue circles indicate the separation of different beams from the central beam. The black star is the position of the FUDS0 (continuum) calibrator source. The desired trace of this example calibration scan is shown with the black dashed line. The real trace of the telescope is shown with the red dotted line.

The front-end used for our observations is the 19-beam receiver (Jiang et al. 2020). According to the angular distance from the centre beam, the receiver has one central beam (01), 6 beams (02–07) evenly spread around an inner circle, 6 beams (09, 11, 13, 15, 17, 19) around a second concentric circle, and 6 beams (08, 10, 12, 14, 16, 18) around the third circle. The inner, second, and third circles are shown by light blue circles in Figure 1. The angular distance between the closest beams is 5.9 arcmin, and the spacing between parallel rows in the geometry shown in Figure 1 is 5.11 arcmin. Each beam has two orthogonal linear polarisation channels (X and Y). The receiver can rotate to track the parallactic angle. The allowable rotation of the line joining beams 08, 02, 01, 05, 14 with a line of constant declination (current epoch) is ±80°.

A calibration noise diode is installed at 45° relative to two linear polarisation probes in each horn. The noise diode operates at two temperatures (high at 10 K or low at 1 K) for different requirements. Since the structure of FAST main reflector depends on pointing direction, calibration of main reflector is frequently required.

The FAST backend consists of 12 Re-configurable Open Architecture Computing Hardware version 2 (ROACH2) units developed by the Collaboration for Astronomy Signal Processing and Electronics Research (CASPER; Hickish et al. 2016; Jiang et al. 2019). Each ROACH2 unit records two pairs of polarised signals simultaneously. Hence, 10 are used to record the signals from two polarisations of 19 beams, and 2 are spare in case of breakdown.

The bandwidth of the ROACH2 is configured to be 500 MHz (frequency range 1.0–1.5 GHz), split into a maximum 1 048 576 channels with a maximum frequency resolution of 476.8 Hz. For our observations, we use 65 536 channels, which gives a frequency spacing of 7.63 kHz, equivalent to a velocity spacing of 1.61 km s⁻¹ in the rest frame. No Doppler shift is applied to spectra during observations. This allows us to search for H I in galaxies with cz in the range of −15 908 to +126 034 km s⁻¹. Spectra can be recorded at maximum rate of 10 Hz, but we chose the lower rate of 1 Hz.

For each beam, three linear polarisation signals (XX*, YY* and XY*) are recorded in fits format files, each with a maximum size of 2 GB. No Doppler shift is applied to the raw spectra before the data is recorded. At this stage, the metadata includes the position of
3.2. Field selection

3.2.1. Target fields

In order to achieve our scientific goals, we apply the following criteria to select target fields:

1. The Declination of the target fields should be close to $+25^\circ$ in order for FAST to utilise its full gain for the maximum length of observing time per day.
2. The target fields should be evenly distributed in RA to minimise the influence of cosmic variance and maximise the ease of scheduling night observations at any time of year.
3. Target fields should avoid strong radio continuum sources ($S_{1.4\,\text{GHz}} > 50\,\text{mJy}$) to avoid reduction in sensitivity due to imperfect spectral subtraction. These are normally background sources, so there is no implicit selection bias.
4. The target fields should overlap with SDSS and other imaging and spectroscopic surveys to provide corresponding optical identifications and galaxy parameters.

For the pilot FUDS0 field, we deliberately select a field which overlaps with the GAL2577 field of AUDS, which slightly compromises the continuum source criterion. In FUDS0, there are four continuum sources with flux densities larger than 50 mJy. For the remaining five target fields, all selection criteria are met. Currently, observations for two fields are in hand: the pilot FUDS0 field, and FUDS1. Details are listed in Table 1 and the field positions are illustrated in Figure 2.

3.2.2. Calibrators

Due to the deformable primary, the FAST gain barely changes at different zenith angles (within $26.4^\circ$) (Jiang et al. 2020). We only use one flux calibrator for each target field in FUDS using the following selection criteria:

1. Strong (> 1.0 Jy at 1.4 GHz) and stable flux density (marked as ‘P’ in the VLA calibrator list\(^a\)) for the VLA D primary focus relative to the spherical centre, the skew of the lower platform, and the rotation angle of 19-beam receiver, which are all recorded in an independent spreadsheet. The sky coordinates and Doppler shifts are computed during the data reduction procedure by using the python package, ASTROPY (Astropy Collaboration et al. 2013, 2018).

As an example, we show the parameters for the FUDS0 calibrator, 0603+219, in Table 2 and show the fit to its continuum spectrum in Figure 3. The VLA flux density is 2.75 Jy at 1.4 GHz configuration, which has the beam size most similar to FAST).

2. In order to avoid confusion, no other strong continuum sources close to the selected calibrator are listed in the NVSS catalogue (Condon et al. 1998).

3. Calibrators are close to target fields, and can be observed within a zenith angle of $+26.4^\circ$.

Flux densities at frequencies near $1.0–1.5\,\text{GHz}$ and their uncertainties were extracted from NED.\(^b\) To derive the frequency dependence of the calibrator flux densities, we use a power law to fit the measurements and uncertainties by using an optimised $\chi^2$ method:

$$S_\nu = S_{\nu_0} \left( \frac{\nu}{\nu_0} \right)^{\alpha}, \quad (1)$$

where the $\nu$ is frequency, $\nu_0$ is the reference frequency (1.4 GHz), and $\alpha$ is the spectral index. For FUDS, we assume the variability of compact sources is negligible during the course of our observations of a particular field.

As an example, we show the parameters for the FUDS0 calibrator, 0603+219, in Table 2 and show the fit to its continuum spectrum in Figure 3. The VLA flux density is 2.75 Jy at 1.4 GHz.

\(^a\)https://science.nrao.edu/facilities/vla/observing/callist.
\(^b\)http://ned.ipac.caltech.edu/.
and the computed spectral index is $\alpha = 0.565 \pm 0.034$. For computing frequency-dependent system parameters, we anchor the flux density with the accurate VLA measurement, and use our computed spectral index.

### 3.3. Observational strategy

#### 3.3.1. Target fields

As shown in Figure 4 for the FUDS0 field, an on-the-fly (OTF) observation mode is employed. Each target field is $30' \times 30'$, as scanned by the central beam, with a rotation angle of $0^\circ$. However, due to the multiple beam pattern, this results in a field size of $50' \times 50'$ with good sensitivity. The spacing between scans is 1 arcmin, which is smaller than the Nyquist sampling interval for an individual 2.9 arcmin beam at 1.4 GHz. The target field is scanned in both RA and Dec direction with a rate of 15 arcsec s$^{-1}$. Each scan takes 120 s and spectra are recorded every second. We use the high-temperature calibration noise diode (10 K) to inject noise for 1 s every 60 s. The total observing time per set of 31 scans is around 1.2 h (62 min scanning, 10 min overhead), or 2.4 h for the combined RA and Dec scans. Multiple sets of similar observations are then combined into the final data cube.

#### 3.3.2. Calibrators

The normal FAST calibration procedure is to point each beam in turn at a calibration source for $\sim$1 min. However Jiang et al. (2020) use an alternative technique of scanning across calibrators which has the advantage of also measuring beam size and pointing offsets. We take this technique, but extend it to the full range of frequencies from 1.0 to 1.5 GHz, as FUDS has good sensitivity to high-redshift galaxies. As mentioned before, OTF mode is used to scan the calibrator with a receiver rotation angle of $0^\circ$, but along scan lines of 1° in length (see Figure 1), which allows all the beams in the same row pass the calibrator in a single scan with adequate baseline either side. The spacing between scan lines is 5.11 arcmin.

Since the rotation angle is relative to the Dec at the current epoch, a slight tilt in the scan is introduced except for the central beam. The maximum deviation is $\sim$1.5 arcsec for the outer beams and $\sim$0.7 arcsec for the inner beams, which is much smaller than the pointing accuracy ($\sim$8 arcsec). As for the target observations, the scan speed is 15 arcsec s$^{-1}$, and the spectra are recorded every 1 s. The high-temperature 10 K noise diode is injected in a square wave pattern with an alternating pattern of 1 s on and 1 s off. The total observation time is around 30 min (20 min scanning, 10 min overhead).

### 4. Radio frequency interference (RFI)

RFI reduces the signal-to-noise ratio of observational data. A list of RFI-affected frequencies is useful for pre-flagging when reducing target field data. As a newly built radio telescope, the FAST list is yet incomplete. From the initial FUDS data, we identify four types of RFI:

- Signals from the global navigation satellite system (GNSS), which include Galileo, GLONASS, GPS, and Beidou. The intensity of these signals depends on the number of satellites in the sky. Within our observational window, they transmit in the frequency range 1.15–1.30 GHz.
- Signals from synchronous communication satellite, AsiaStar. These signals are strong and range from 1.48 to 1.49 GHz.
- Signals from the civil aviation radar. They occur at 1.03 and 1.09 GHz. An occasional strong pulse radar signal appears at 1.105 and 1.131 GHz.
- Signals generated by the refrigerating dewar in the compressor. These cover the entire 500 MHz bandwidth. The signals have a temperature of a few K, and a linewidth of about 1 MHz. In order to solve the issue, the compressor was screened in Aug 2021. Raw FUDS data taken since then do not show this RFI, even in spectra averaged over 20 min. However, initial FUDS0 and some FUDS1 data are contaminated.

#### 4.1. RFI detection and baseline fitting

The GNSS signals are strong (having comparable intensity of the system temperature of FAST) and highly variable in time. They also occupy a large fraction of the FAST bandwidth (37.5%).
signals can affect the system temperature and its measurement even at non-RFI frequencies. The RFI from the compressor also adds uncertainty for flux intensity measurements by biasing the bandpass calibration. In order to mitigate these effects, we identify the RFI and fit an underlying spectral baseline using the following steps:

1. The spectrum is divided into several segments with an overlap of 15 MHz. The width of each segment varies from 90 to 240 MHz, depending on the number of channels that are free of RFI.

2. For each segment, a Gaussian weighted polynomial function is employed to fit the baseline. The standard deviation of the Gaussian weight is 0.17 times the segment width. The order of the fit is iterated from 0 to 6. For each order, a pseudo-RMS is calculated using $1.4826 \times \text{MAD}$ (Median Absolute Deviation). The channels with absolute values exceeding $7 \times \text{RMS}$ and immediate neighbouring channels with absolute values exceeding $3 \times \text{RMS}$ are marked as RFI.

3. A temporary baseline for the whole spectrum is determined by combining the baselines of each segment with Gaussian weights.

4. The above procedures are iterated on a smoothed baseline-removed spectrum with increasing smoothing length (25, 50, 100, and 200 kHz). The intermediate baseline is generated by summing together the temporary baselines from each iteration.

5. The final baseline is the intermediate baseline generated in the final iteration. All the channels ever marked as RFI in each iteration are recorded as RFI in the final spectrum.

We note that RFI flagging in raw data is time consuming. Since the RFI is stable in frequency in a short period (around 20 min, see Section 4.2), we only perform this procedure on one spectrum per 60 spectra (the same frequency as the noise injection in the target fields). The template for flagged RFI is therefore used for the next 59 spectra.

In Figure 5, we show a total power spectrum (XX-polari- sation) from a single 1 s integration from the FUDS0 central beam. The fitted baseline and the identified regions with RFI are demonstrated. About 41.7% of the spectrum contains RFI at some level. This procedure may miss weak RFI, but the baseline fitting is relatively robust.

4.2. Compressor RFI

To characterise the signals generated by the refrigerating dewar in the compressor, we analyse the calibrator data taken on 2019 August 25. In Figure 6, we show a waterfall image of the calibrated spectra between frequencies of 1.31 and 1.41 GHz. This frequency range excludes the GNSS RFI and H I emission from the Milky Way. The compressor signals (the vertical stripes) generally have an intensity of about 0.3 Jy and a linewidth of $\sim 1$ MHz. The signals are variable in amplitude and frequency (middle panel of Figure 6) and also shift in frequency when comparing data on different days (see Figure 7). The signals do not depend on azimuth or elevation as the compressor is housed in the focus cabin.

Calibrated spectra of all 19 beams from the same observation are shown in the left panel of Figure 8. There is a similarity in frequency and intensity of the compressor RFI across all 19 beams. We create a smoothed median spectrum from all beams and use it as template to subtract from all spectra. The residual spectra after RFI removal are shown in the right panel of Figure 8. Removal of the compressor signals is only performed after the calibration due to large power differences between the beams. Large sources which extend across several beams will potentially

---

**Figure 5.** The upper panel shows a total power spectrum (polarisation XX*) from the central beam (black line) and the underlying fitted baseline (red line). The units are arbitrary. The spectrum around 1.23 GHz is zoomed in to show the standing wave. The frequencies of the detected RFI are indicated by the blue bars in the lower panel.

**Figure 6.** In the left panel, calibrated data between 1.31 and 1.41 GHz is shown for the XX* polarisation channel of beam 1 as a function of frequency (in GHz) and time (in sec). The right panel is a zoom (in frequency space) of the region indicated by the red rectangle. The signals present are due to RFI from a compressor in FAST data taken prior to 2021 July.
have their flux density reduced. We note that further flagging is required to remove residual RFI.

5. Calibration

5.1. Gain stability

The FAST calibration method involves switching on and off a noise diode, which is assumed to have a constant amplitude. Once the spectrum of noise diode has been calibrated against a radio continuum source, the diode noise can then be used to measure the system equivalent flux density (SEFD) in Jy, $S_{\text{sys}}$ and its stability. Because the high-temperature noise diode elevates $T_{\text{sys}}$ by nearly 50%, it is impractical to use with a 50% duty cycle for target field observations. We therefore needed to confirm that the receiver calibration was stable in a mode where the calibration ‘on’ or off at a rate of 1 Hz while performing a calibrator observation, and the mean value is nearly 50%, it is impractical to use with a 50% duty cycle for target field observations. We therefore needed to confirm that the receiver calibration was stable in a mode where the calibration ‘on’ spectrum was only taken for 1 s every 60 s.

To confirm this, we first removed the identified RFI channels (see Section 4.1), and channels that are outside of the official band-pass ($< 1.05$, or $> 1.45$ GHz). The noise diode was switched on and off at a rate of 1 Hz while performing a calibrator observation, and the difference in the mean power in each pair of on/off spectra is calculated. Figure 9 shows the result, which indicates that no significant variation occurs over any time frame less than 1 500 s. This confirms that both the noise diode and the receiver gain are adequately constant within this period.

5.2. Instrumental characterisation

For a radio telescope, a Gaussian function is a good approximation for angular response of the beam (the point spread function). The following function can therefore be used to fit a 1-dimensional scan of a calibrator where the noise diode (CAL) is either on or off:

$$f(d) = \begin{cases} \frac{(d - d_0)^2}{2\sigma^2} + C_1 & \text{(CAL on)} \\ A e^{-\frac{(d - d_0)^2}{2\sigma^2}} & \text{(CAL off)} \end{cases}$$

where $d$ is the angular distance offset from the calibrator, $d_0$ is the calibrator position, $A$ is the amplitude, $\sigma$ is the standard deviation, and $C_1$ and $C_2$ are the CAL on and off baselines, respectively. The flux density of the noise diode is then $S_{\text{noise}} = \frac{C_1-C_2}{7}\times S_{\text{calibrator}}$, the SEFD $S_{\text{sys}} = \frac{C_2}{7}\times S_{\text{calibrator}}$, and the full width half power (FWHP) beamwidth is $\theta = 2\sqrt{\ln 2}\sigma$, where $S_{\text{calibrator}}$ is the flux density of the calibrator. We also measure the pointing error, $\sigma_{\text{pointing}}$ from the RMS of the deviation of $d$, from the catalogued position.

We firstly calculate mean instrumental parameters, $\bar{S}_{\text{noise}}, \bar{S}_{\text{sys}}, \bar{\sigma}_{\text{pointing}},$ and $\bar{\theta}$ averaged over the frequency range 1.05–1.45 GHz, using the VLA calibrator flux density at 1.4 GHz.

However, we also derive frequency-dependent parameters, $S_{\text{noise}}(v), S_{\text{sys}}(v), \sigma_{\text{pointing}}(v),$ and $\theta(v)$ using the frequency-dependent flux density of calibrator, $S_{\text{calibrator}}(v)$ derived in Section 3.2.2. Some uncertainty is introduced at frequencies affected by RFI, so stable results can only be derived using the baseline of raw spectrum provided in Section 4.1. In some cases (e.g. for beam size), the frequency dependence is strong, as expected. In other cases (e.g. pointing error), there is virtually no frequency dependence. An example Gaussian fit to the beam response for polarisation XX, beam 1 is shown in Figure 10, and the derived instrumental parameters are shown in Figure 11. Due to strong GNSS RFI the measurements between 1.15 and 1.30 GHz (the grey area Figure 11) are less accurate.

5.2.1. Noise diode flux density

In the upper left panel of Figure 11, we show the flux density of noise diode. The mean flux density is $\bar{S}_{\text{noise}} = 0.72 \pm 0.08$ Jy, and the frequency dependence, $S_{\text{noise}}(v)$ is small, rising between 1.08 and 1.12 GHz, and decreasing above 1.4 GHz. A similar variation is also found by Jiang et al. (2020).

5.2.2. System equivalent flux density (SEFD)

In the upper right panel of Figure 11, we show the SEFD. The mean value is $\bar{S}_{\text{sys}} = 1.24 \pm 0.08$ Jy, but $S_{\text{sys}}(v)$ increases by 15% below 1.05 GHz, which is outside of the nominal frequency range. A bump is also seen between 1.08 and 1.12 GHz. Similar deviations are found in other beams and in the YY polarisation. Since it also seems to correspond to a change in the SEFD and beamwidth, it is possibly an instrumental effect in the feed or the feed-LNA transition.

In Figure 12, we derive a median frequency-dependent correction factor for the mean SEFD correction by using data from both polarisation on all 19 beams. The correction factor is parameterised according to:

$$f_{\text{sys}}(v) = A \times e^{-\frac{(v-\nu_0)^2}{2\sigma^2}} + C \times e^{-\frac{(v-\nu_1)^2}{2\sigma^2}} + E,$$  

where the first term is the exponential rise at low frequency ($< 1.05$ GHz), the second term is the Gaussian component at 1.1 GHz, and the third term is the deviation at high frequency ($> 1.3$ GHz). We use the optimised $\chi^2$ method to derive the best fit parameters for the median line, excluding the RFI affected frequency range (1.15–1.30 GHz). The derived best fit parameters are given in Table 3 and the best fit line is shown by the red solid line. The uncertainty for each of the parameters is estimated using a jackknife method, that is removing one spectrum from 19 beams $\times 2$ polarisation channels each time.

To evaluate the effects of solar interference on the SEFD, we compare the $\bar{S}_{\text{sys}}$ of each beam for the day and night observations (see Figure 13). We find values for $\bar{S}_{\text{sys}}$ which are about 16% higher in the day. The mean $\bar{S}_{\text{sys}}$ values are 1.34 and 1.15 Jy in the day and night, respectively.
For daytime observations, we list the mean \( \bar{S}_{\text{sys}} \) and its uncertainty for different combinations of beams in Table 4. The central beam (beam 1) has the lowest \( \bar{S}_{\text{sys}} \) value of 1.23 Jy. The inner circle of beams (beams 2–7) has a similar mean value (\( \bar{S}_{\text{sys}} = 1.29 \) Jy). The second circle (beams 9, 11, 13, 15, 17 and 19) has a significantly higher mean value (\( \bar{S}_{\text{sys}} = 1.36 \) Jy), and the final circle (beams 8, 10, 12, 14, 16 and 18) have the highest mean value (\( \bar{S}_{\text{sys}} = 1.40 \) Jy). Table 4 shows that this is not due to a rise in system temperature \( T_{\text{sys}} \), which is fairly constant (here we adopt the gain at 1.35 GHz cited by Jiang et al. 2020 to convert SEFD to \( T_{\text{sys}} \)). Instead, it is due to the classical drop in off-axis antenna efficiency for paraboloid antennas (Ruize 1965).

5.2.3. Pointing error

In the lower left panel of Figure 11, we show the measured pointing error. It clearly shows that \( \sigma_{\text{pointing}}(\nu) \) is nearly constant, and consistent with the mean value, \( \bar{\sigma}_{\text{pointing}} \). We also show the histogram of mean pointing errors across all beams and polarisations over two days of observations (76 measurements) in Figure 14. The mean and median errors are 6.3 and 8.1 arcsec, respectively. Considering the 2.8 arcmin beam size of FAST at 1.5 GHz, the pointing accuracy is better than \( \frac{2}{3} \) of the FAST beam.

5.2.4. Beam size

The lower right panel of Figure 11 shows the beam size measurements. The mean beam size \( \bar{\theta} \), is 3.26 arcmin. The beam size for beams at different radii from the optical axis show a weakly increasing trend (see the final column in Table 4). Our findings are consistent with the results from Jiang et al. (2020).

As expected, the beam size \( \theta(\nu) \), decreases with frequency. Across all beams and both polarisations, the median frequency-dependent beam size and its 1-\( \sigma \) uncertainty is shown in Figure 15. The beam size is dependent on wavelength and telescope diameter. For an unblocked circular aperture with uniform illumination, the FWHP beam size is \( \theta = 1.02\lambda/D \) in radians (Wilson, Rohlfs, & Hüttemeister 2013), where \( \lambda \) is the wavelength and \( D \) is the illuminated diameter. Normal radio telescopes with tapered illumination patterns will typically have beam sizes which are larger by 20%, but this depends on the taper (Wilson et al. 2013) and the offset from the optical axis. For the similar prime-focus Parkes multibeam system (Staveley-Smith et al. 1996), the average FWHP beam size is \( \theta = 1.225\lambda/D \) in radians. Hence, for the multiple beams on FAST, we parameterise the beam size as follows:

\[
\theta = 1.225 \frac{\lambda}{A + B \left( \frac{\lambda}{0.21} \right) + C \left( \frac{\lambda}{0.21} \right)^2},
\]

where \( A, B \) and \( C \) are functions of \( \lambda \). We use an optimised \( \chi^2 \) method to find the best fit (see Figure 15). The best fit parameters and the parameters for all the 19 beams are listed in Tables 5 and Tables A.1, respectively.

6. Data reduction

6.1. System equivalent flux density

For target field flux density calibration, we use the flux density of noise diode as calibrated by the above calibrator observations. We first remove channels affected by the known RFI and channels outside the FAST nominal bandpass (< 1.05 or > 1.45 GHz) (for calibration purposes only). Subsequently, we calculate the mean power for all the spectra. We derive the CAL step by subtracting the CAL off spectrum from the CAL on spectrum that it follows. The value for \( \bar{S}_{\text{noise}} \) derived in Section 5.2 is used as the equivalent flux density of the CAL step. The derived value for \( \bar{S}_{\text{sys}} \) is then used for the subsequent 59 spectra. The frequency-dependent scale factor derived in Section 5.2 is applied after flagging.

6.2. Bandpass removal

We follow the robust bandpass calibration procedure as used for HIPASS (Barnes et al. 2001). In order to calibrate a raw spectrum \( P_{\nu}(\nu) \) according to its measured SEFD \( S_{\nu} \), we select neighbouring spectra as a reference spectrum set, \( C_{\nu} \). The reference SEFD \( S_{\text{sys},r} \), and the reference spectrum \( P_{\nu}(\nu) \), are computed from the median of the spectra in \( C_{\nu} \). Subsequently, the raw spectrum is flattened using the reference spectrum and scaled to units of \( S_{\text{sys},r} \). The system temperature contribution, including sky continuum, is then removed by subtracting \( S_{\text{sys},r} \). The procedure is summarised by:

\[
S_{\nu}(\nu) = S_{\text{sys},r} \times \frac{P_{\nu}(\nu)}{P_{\nu}(\nu)} - S_{\text{sys},r},
\]
Table 3. The best fit parameters for the normalised frequency dependence of the SEFD using Equation (3). The uncertainty is from the jackknife method.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value (jackknife uncertainty)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\nu_0$ (GHz)</td>
<td>0.98 (0.01)</td>
</tr>
<tr>
<td>$\nu_1$ (GHz)</td>
<td>1.1108 (0.0002)</td>
</tr>
<tr>
<td>$A$</td>
<td>0.28 (0.06)</td>
</tr>
<tr>
<td>$B$</td>
<td>0.060 (0.002)</td>
</tr>
<tr>
<td>$C$</td>
<td>0.094 (0.001)</td>
</tr>
<tr>
<td>$D$</td>
<td>0.0306 (0.0006)</td>
</tr>
<tr>
<td>$E$</td>
<td>0.9828 (0.0006)</td>
</tr>
</tbody>
</table>

Table 4. The mean SEFD, system temperature and mean beam size for beams at different offsets from the optical axis (beam 1) for observations of 2019 Aug 25.

<table>
<thead>
<tr>
<th>Position</th>
<th>$\bar{S}_{\text{sys}}$ (Jy)</th>
<th>$\bar{T}_{\text{sys}}$ (K)</th>
<th>$\bar{\theta}$ (arcmin)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Center</td>
<td>1.23 (0.08)</td>
<td>19.7 (1.4)</td>
<td>3.26 (0.01)</td>
</tr>
<tr>
<td>1st circle</td>
<td>1.29 (0.08)</td>
<td>19.6 (1.3)</td>
<td>3.28 (0.01)</td>
</tr>
<tr>
<td>2nd circle</td>
<td>1.36 (0.08)</td>
<td>19.7 (1.3)</td>
<td>3.29 (0.02)</td>
</tr>
<tr>
<td>3rd circle</td>
<td>1.40 (0.08)</td>
<td>20.0 (1.3)</td>
<td>3.32 (0.03)</td>
</tr>
<tr>
<td>Any</td>
<td>1.34 (0.08)</td>
<td>19.8 (1.3)</td>
<td>3.30 (0.02)</td>
</tr>
</tbody>
</table>

Table 5. Parameters of Equation (4) for the beam size of FAST, and beam size in our final cube.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$A$ (m)</th>
<th>$B$ (m)</th>
<th>$C$ (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>295.69 (0.05)</td>
<td>89.42 (0.67)</td>
<td>−76.47 (1.70)</td>
</tr>
<tr>
<td>Final cube</td>
<td>271.27 (0.04)</td>
<td>111.44 (0.53)</td>
<td>−81.83 (1.37)</td>
</tr>
</tbody>
</table>

where $S_i(\nu)$ is the calibrated spectrum in Jy, $S_{\text{sys},i}$ and $S_{\text{sys},r}$ are in Jy, and $P_i(\nu)$ and $P_r(\nu)$ are in (arbitrary) units of power. For each beam, we include the nearest 260 spectra in the same beam as the reference spectrum set, giving spatial coverage of about two scan lines. As explained in Calabretta, Staveley-Smith, & Barnes (2014), the form of Equation (5) flattens the bandpass, converts the flux scale to Jy and removes continuum to zeroth-order. Due to residual spectral artefacts (spectral index and ripple) in strong continuum sources (> 10 mJy), spectra within 3 arcmin of these sources are excluded in the reference spectrum set. To reduce sidelobes caused by any strong HI emission, we use a 5-channel Hanning window (0.25, 0.75, 1, 0.75, 0.25) to convolve with the calibrated spectra. Therefore, the frequency resolution is lowered to 22.9 kHz. This procedure is performed for each polarisation channel and for all 19 beams. We use a cubic polynomial function to fit the calibrated spectra to further flatten the baseline. Figure 16 shows 60 calibrated spectra for the XX* polarisation of beam 1. The residual compressor RFI, particularly the vertical strips between 1.31 and 1.41 GHz, are still visible (upper left panel). The use of the median spectrum for all 19 beams suppresses the compressor RFI (upper right panel) but also the extended Galactic H I emission.6

6Raw data are archived, so re-processing is possible for future Galactic H I studies.

**6.3. RFI removal**

To remove residual RFI from all target spectra, we adopt the following flagging steps:

1. Pixels without extreme values (not deviating more than ±0.1 Jy) are used for deriving noise. The theoretical sensitivity for the given frequency and time resolution is 15.3 mJy, assuming a SEFD of 1.3 Jy (corresponding to a gain of 15 K Jy$^{-1}$ and a system temperature of 20 K).

2. Channels with high noise levels, $\sigma_c > 1.2\sigma_g$, are noted as bad, where $\sigma_c$ is computed as 1.4826 × MAD for each channel in all scans in one dimension (either RA or Dec) over one day, and $\sigma_g$ is the global average calculated in the same way over all channels.

3. Pixels in bad channels with high noise levels in the time-frequency domain are flagged, that is $\sigma_p > 2\sigma_g$, where $\sigma_p$ is defined as the RMS in the surrounding $5 \times 5$ box. This step also detects regions where the mean flux density deviates...
The variation of noise diode flux density (top left), SEFD (top right), pointing error (bottom left) and beam size (bottom right) as a function of frequency is shown in black for beam 1, polarisation XX. The hatched areas show the 1-σ uncertainties. The grey area shows the frequency range from 1.15 to 1.3 GHz which is affected by RFI (GNSS satellites). The mean values, which exclude RFI-affected regions and edge channels, are shown in red.

from zero, so bright galaxies in RFI-affected channels may be eliminated.

4. Any flagged pixel is grown into a flagged 5 × 5 pixel region.

5. The spectra take with CAL on, which have higher $S_{\text{sys}}$, are also flagged.

The third panel of Figure 16 shows the effect of removing the flagged pixels. The flagged data accounts for 25.6% of the calibrated data, which is smaller than the 41.7% of flagged RFI data in the raw spectral data (Figure 5). This noise-based flagging method therefore recovers 15% of the data that would otherwise have been flagged. This is partly helped by the compressor RFI template. As already noted, the frequency-dependent flux density correction factor, $f_{\text{sys}}(\nu)$ is applied to the data after flagging, and also shown in Figure 16.

6.4. Gridding

We employ the robust gridding algorithm developed for HIPASS to create our final FUDS data cube. The final cube has a size of 1° × 1° and a pixel size of 1 arcmin × 1 arcmin. The frequency spacing in the final data cube is the same as the raw spectra, 7.63 kHz, but the resolution after Hanning smoothing is 22.9 kHz. For each pixel, we use a radius of 1.3 arcmin as a cutoff to select the contributing spectra. Considering the variation of beam size with frequency, up to 40% across the whole bandwidth, the median frequency-dependent beam size derived in Section 5.2 is used for all beams. The spectrum for a given position in the cube is then derived as follows:

$$S(\nu) = \frac{\text{median}(S_i(\nu))}{\text{median}(w_i(\nu))},$$  

where $S_i(\nu)$ are the spectra within 1.3 arcmin and $w_i(\nu)$ are their corresponding Gaussian weights.

Assuming all the 19 beams have same analytic beam size, $\theta(\nu)$, derived in Section 5.2, we perform the same gridding procedure above to derive the beam size in our final cube by using a Monte Carlo (MC) method. We then use Equation (4) to fit the gridded beam. The best fit line is shown in Figure 15, and the best fit parameters of Equation (4) are given in Table 5. This beam size is required when deriving precise flux densities for extended source in the gridded cubes.

6.5. Strong continuum sources

We employ the method of Barnes et al. (2001) to remove residual emission near strong continuum sources. We use the spectra toward continuum sources with flux densities stronger than 10 mJy to generate a template spectrum, given by:

$$S_T(\nu) = \Sigma_i w_i S_i(\nu),$$  

where $S_i(\nu)$ is the central spectrum of ith continuum source, and $w_i$ is the weight. The spectrum associated with the strongest
continuum source has a weight set to unity. The weight of other sources is determined by the slope of the linear regression between two spectra $S_i(\nu)$ and $S_j(\nu)$ in the frequency ranges $1.015$–$1.137$ GHz, $1.30$–$1.416$ GHz and $1.423$–$1.463$ GHz. For each spectrum in the final data cube, we perform a linear regression with the template in the same manner to derive the weight $w$, then subtract $w S_T(\nu)$ from the spectrum.

7. Confusion

Confusion happens when multiple galaxies cannot be distinguished in both space and velocity due to the limited resolution of a telescope. There are two main effects on a survey: (1) an underestimate of the number of observable galaxies, and (2) an overestimate of the HI mass due to the contribution of unresolved galaxies. Jones et al. (2016a) shows an example of the increase in apparent HI mass with increasing beam size. Based on the estimates from stacking, they predict a confused mass of $\sim 2 \times 10^9 h^{-2} M_\odot$ within a FAST-like beam size (3 arcmin) for a galaxy with velocity width of $600$ km s$^{-1}$ at $z = 0.1$. Zhang et al. (2019) employed the same method to estimate the confused mass within the FAST beam for the CRAFTS survey, and found a similar confused mass of $\sim 2 \times 10^9 h^{-2} M_\odot$ at a similar redshift.

To investigate the confusion in FUDS, we perform a simulation following the method described in Jones et al. (2015). In the simulation, we assume a FUDS sensitivity of $50 \mu$Jy beam$^{-1}$. First, we sample the galaxies based on the HI mass derived from the HIPASS HIMF and the velocity widths sampled from the conditional velocity function (Martin et al. 2010). We limit the galaxy masses between $10^6$ and $10^{12} h^{-2} M_\odot$ and the velocity widths between $15$ and $10^3$ km s$^{-1}$. The correlation function is used to distribute the mock galaxies in $50^3 h^{-1}$ Mpc co-moving volumes at each redshift, starting with one galaxy located at each centre. The beam size in our final cube (see Section 5.2.4) is used as the confusion limit for angular separation. The distance along the line-of-sight is converted to a velocity separation using the Hubble-Lemaître law. Subsequently, we perform the simulation $10^4$ times to calculate mean values, independent of cosmic variance.

Here, we inspect the impact of confusion only for the detected galaxies. We define $f_{\text{Det-All}}(N)$ and $f_{\text{Det-De}}(N)$ as the fraction of detected galaxies which are confused with other simulated

---

**Figure 12.** The median normalised SEFD across all beams and polarisation channels is shown as a function of frequency (black dashed line), with corresponding 1-$\sigma$ uncertainty (hatched area). A rise is apparent at low frequency ($< 1.15$ GHz) in all beams. We use Equation (3) to fit the median, and derive the best fit line shown by the solid red line. Data affected by RFI (grey area) are excluded in the fit.

**Figure 13.** The SEFD for all 19 beams is shown. The red pluses represent day time data, while the blue dots represent night time data. The beams in outer circles of the receiver have higher SEFD. The mean daytime SEFD is $1.34$ Jy, and $1.15$ Jy at night.

**Figure 14.** A histogram of the pointing errors measured in both polarisations, all beams, and in both observing days. The bins have a width of 5 arcsec. The median and mean values are $6.30$ and $8.05$ arcsec, and are shown by red dotted line and red dashed line, respectively.

**Figure 15.** The median FWHP beam size for all 19 beams is shown by the black solid line, with the corresponding 1-$\sigma$ uncertainty shown by the hatched area. The best fit line by using Equation (4) is shown by the red dashed line. The beam size in our final cube (post-gridding, described in Section 6.4) is shown by the red solid line.
where \( N_{\text{Det--All}} \) is the number of detected galaxies confused with any other galaxy, \( N_{\text{Det--Det}} \) is the number of detected galaxies confused with any other detected galaxy, and \( N_{\text{Det}} \) is the number of detected galaxies. The latter ratio, \( f_{\text{Det--Det}}(N) \), permits an estimate of the impact of confusion on the number of detected galaxies in FUDS. In Figure 17, we use green and red circles to show predicted values for \( f_{\text{Det--All}}(N) \) and \( f_{\text{Det--Det}}(N) \) in each redshift bin, respectively. \( f_{\text{Det--All}}(N) \) increases quickly to 100% at \( z \sim 0.1 \). However, \( f_{\text{Det--Det}}(N) \) is much lower. At low redshift, it is negligible, but reaches a peak value of 40% at \( z = 0.1 \). For \( z > 0.1 \), it drops sharply with \( f_{\text{Det--Det}}(N) \) being below 15% at \( z > 0.25 \). Therefore, at all redshifts, confusion will reduce the number of detections by less than 20% (40/2%) . Confusion will also increase the apparent \( \text{HI} \) mass by merging with confused galaxies. Considering the galaxies below the detection limit, we find that the average total confused mass in a beam for our simulated central galaxies is \( \sim 2 \times 10^9 \ h^{-2}_{70} \ M_\odot \) at \( z = 0.1 \), and \( 10^{10} \ h^{-2}_{70} \ M_\odot \) at \( z = 0.4 \). This is similar to the results of Zhang et al. (2019), and slightly lower than the values predicted in Jones et al. (2016a), probably due to taking into account the velocity distribution.

In FUDS, we are interested in the corresponding ratio of the confused mass to the mass of the largest galaxy in each beam. We introduce \( f_{\text{Det--All}}(M) \) and \( f_{\text{Det--Det}}(M) \) to represent the fraction of mass in the largest detected galaxy which is confused with mass in other simulated galaxies and the fraction confused with only mass in other detected galaxies, respectively. The definitions are:

\[
f_{\text{Det--All}}(M) = \frac{M_{\text{Det--All}}}{M_{\text{Det}}},
\]

(9)

where \( M_{\text{Det--All}} \) is the mass in all the confused galaxies except the largest one, \( M_{\text{Det--Det}} \) is the mass in detected confused galaxies except the largest one, and \( M_{\text{Det}} \) is the mass of the largest detected confused galaxy. The two parameters show us the deviation of measured \( \text{HI} \) mass caused by confusion. In Figure 17, we use green and red triangles to show the average values \( f_{\text{Det--All}}(M) \) and \( f_{\text{Det--Det}}(M) \) across all our simulations as a function of redshift. We find that \( f_{\text{Det--All}}(M) \) slowly increases with redshift, and reaches about 70% at \( z \sim 0.15 \). The fluctuation beyond \( z \sim 0.25 \) is the result of less samples (< 20) in each of these redshift bins. The mean value \( f_{\text{Det--All}}(M) \) is about 95% at high redshifts. However, confusion from undetected galaxies tends to elevate the spectral...
baseline, so does not necessarily result in an increase in the measure HI mass, $f_{\text{Det} - \text{Det}}(M)$ may therefore be a more appropriate statistic to indicate the impact of confusion on measured HI mass. $f_{\text{Det} - \text{Det}}(M)$ is about 15% at low redshift ($0.04 < z < 0.15$), and almost zero at higher redshift ($z > 0.25$). Note that both $f_{\text{Det} - \text{All}}(M)$ and $f_{\text{Det} - \text{Det}}(M)$ become large ($\sim 30\%$) at very low redshift ($z < 0.03$). That is caused by the dramatic increase in the numbers of low-mass galaxies which are detected. However, the cosmic volume at low redshift in FUDS is small. In reality, the contribution from confused mass will likely lie somewhere between $f_{\text{Det} - \text{All}}(M)$ and $f_{\text{Det} - \text{Det}}(M)$.

In order to understand the impact of confusion on the measured HIMF, we recover the HIMF from the simulated catalogue using the $\Sigma_1/V_{\text{max}}$ method (Schmidt 1968). As mentioned above, the contribution of confusing mass should be between $f_{\text{Det} - \text{All}}(M)$ and $f_{\text{Det} - \text{Det}}(M)$. Hence, we calculate the HIMF in three cases: (1) no confusion; (2) confusing mass from all the confused galaxies; (3) confusing mass only from the detected confused galaxies. The first is a reality check of the $\Sigma_1/V_{\text{max}}$ method. The second and third cases give upper and lower limits to the contribution of confused mass. For simplicity, we assume that the confusion only increases flux, without changing the velocity width. Thanks to the large simulated sample, the uncertainties in each mass bin are low. The recovered HIMFs are shown in Figure 18. We use the Schechter function (Schechter 1976) to parameterise the HIMF. All($\log(\frac{M_{\text{HI}}}{M_\odot}) < 0.03$) at very low redshift ($z < 0.03$). That is caused by the dramatic increase in the numbers of low-mass galaxies which are detected. However, the cosmic volume at low redshift in FUDS is small. In reality, the contribution from confused mass will likely lie somewhere between $f_{\text{Det} - \text{All}}(M)$ and $f_{\text{Det} - \text{Det}}(M)$.

In order to understand the impact of confusion on the measured HIMF, we recover the HIMF from the simulated catalogue using the $\Sigma_1/V_{\text{max}}$ method (Schmidt 1968). As mentioned above, the contribution of confusing mass should be between $f_{\text{Det} - \text{All}}(M)$ and $f_{\text{Det} - \text{Det}}(M)$. Hence, we calculate the HIMF in three cases: (1) no confusion; (2) confusing mass from all the confused galaxies; (3) confusing mass only from the detected confused galaxies. The first is a reality check of the $\Sigma_1/V_{\text{max}}$ method. The second and third cases give upper and lower limits to the contribution of confused mass. For simplicity, we assume that the confusion only increases flux, without changing the velocity width. Thanks to the large simulated sample, the uncertainties in each mass bin are low. The recovered HIMFs are shown in Figure 18. We use the Schechter function (Schechter 1976) to parameterise the HIMF. All($\log(\frac{M_{\text{HI}}}{M_\odot}) < 0.03$) at very low redshift ($z < 0.03$). That is caused by the dramatic increase in the numbers of low-mass galaxies which are detected. However, the cosmic volume at low redshift in FUDS is small. In reality, the contribution from confused mass will likely lie somewhere between $f_{\text{Det} - \text{All}}(M)$ and $f_{\text{Det} - \text{Det}}(M)$.

In order to understand the impact of confusion on the measured HIMF, we recover the HIMF from the simulated catalogue using the $\Sigma_1/V_{\text{max}}$ method (Schmidt 1968). As mentioned above, the contribution of confusing mass should be between $f_{\text{Det} - \text{All}}(M)$ and $f_{\text{Det} - \text{Det}}(M)$. Hence, we calculate the HIMF in three cases: (1) no confusion; (2) confusing mass from all the confused galaxies; (3) confusing mass only from the detected confused galaxies. The first is a reality check of the $\Sigma_1/V_{\text{max}}$ method. The second and third cases give upper and lower limits to the contribution of confused mass. For simplicity, we assume that the confusion only increases flux, without changing the velocity width. Thanks to the large simulated sample, the uncertainties in each mass bin are low. The recovered HIMFs are shown in Figure 18. We use the Schechter function (Schechter 1976) to parameterise the HIMF. All($\log(\frac{M_{\text{HI}}}{M_\odot}) < 0.03$) at very low redshift ($z < 0.03$). That is caused by the dramatic increase in the numbers of low-mass galaxies which are detected. However, the cosmic volume at low redshift in FUDS is small. In reality, the contribution from confused mass will likely lie somewhere between $f_{\text{Det} - \text{All}}(M)$ and $f_{\text{Det} - \text{Det}}(M)$.

8. Preliminary results for the FUDS0 field

Observations of the FUDS0 field have been concluded in our pilot survey. The on-source integration time is about 90 h. The calibration, data reduction methods described above are employed to process the data. The rms for the final data cube is $\sim 50 \mu$Jy in the RFI-free regions and at the centre of the field. We detect 128 galaxies in HI with redshifts in the range of $0 < z < 0.4$ and HI masses in the range of $6.8 < \log(M_{\text{HI}}/M_\odot) < 11.0$.

In Figure 19, we show the spectra of two of the detected galaxies, with a Busy Function fit (Westmeier et al. 2014) overlaid. Since the FUDS0 field partially overlaps with GAL2577 field in the Arecibo Ultra-Deep Survey (AUDS) (Xi et al. 2021), 33 galaxies with $z < 0.16$ (the AUDS limit) are detected in both surveys. A comparison of integrated flux and line width was performed between the two surveys. And we did not find any systematic errors. FUDS 040, in the upper panel is one of these. The AUDS spectrum is overlaid (cyan line), and its AUDS100 ID is shown in the upper left corner. In the lower panel, we show an example spectrum of a galaxy at $z = 0.27$, which is beyond the AUDS redshift limit of 0.16.
is beyond the redshift limit of AUDS. Detailed results from the FUDS0 field will be presented in a separate paper.

9. Summary

We introduce a new deep extragalactic H\textsc{i} survey on the FAST telescope, the FAST Ultra-Deep Survey (FUDS). The high sensitivity of FAST and wide frequency range of its multibeam receiver enables studies of high redshift galaxies, faint nearby objects, and the evolution of the H\textsc{i} mass function and cosmic H\textsc{i} density. In this paper, we present the science goals, the observing technique, the calibration and data reduction methods as applied to its initial data. In summary:

1. In order to quickly calibrate all 19 beams, we employ an on-the-fly (OTF) observation mode to scan a continuum calibrator source, which allows us to measure the flux density of the calibration noise diode, the system equivalent flux density, the pointing accuracy and the beam size.

2. The high-temperature (10 K) noise diode was found to be stable, with a mean flux density of 0.72 ± 0.08 Jy. The mean system equivalent flux density was found to be 1.24 ± 0.08 Jy. The median and mean value of pointing errors were measured to be 6.3 and 8.1 arcsec, which is smaller than 1/3 of FAST beam size at 1.5 GHz.

3. The wide bandwidth of the multibeam receiver necessitated calculating the frequency dependence of calibration parameters in a manner that was robust against the considerable GNSS RFI present from 1.15 to 1.30 GHz. The main departures from frequency independence of calibration parameters were measured to be: a decrease in the high-frequency response of the noise diode; a low-frequency rise for the system equivalent flux density, and the expected inverse-frequency dependence for the beam size.

4. Internal RFI was found to be problematic with early data. We were able to suppress this considerably by subtracting median spectra calculated in 19 beams over short periods of time. Further RFI flagging was developed to remove data contaminated by RFI, and ensure that strong H\textsc{i} sources were not affected.

5. For our initial target field data in FUDS0 and FUDS1, we employ the similar algorithms for bandpass calibration and imaging as for HIPASS and AUDS. The final cube for each FUDS field is 1 deg × 1 deg, with pixel size of 1 arcmin × 1 arcmin and frequency resolution of 22.9 kHz. Sample spectra are presented, and a comparison with data from the Arecibo Ultra-Deep survey (AUDS) is made.

Acknowledgements. We would like to thank Jiguang Lu for his suggestions on calibration of 19 beams. Shijie Huang’s kindly provided an RFI list for the FAST site. Pei Zuo helped improve the paper.

This work made use of the data from FAST (Five-hundred-metre Aperture Spherical radio Telescope). FAST is a Chinese national mega-science facility, operated by the National Astronomical Observatories, Chinese Academy of Sciences. The work is supported by the National Key R&D Program of China under grant number 2018YFA0404703, and the Open Project Program of the CAS Key Laboratory of FAST, NAOC, Chinese Academy of Sciences. Parts of this research were supported by the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), through project number CE170100013. BP would like to acknowledge the CAS-MPG Legacy funding ‘Low-Frequency Gravitational Wave Astronomy and Gravitational Physics in Space’.

References

Blyth, S., et al. 2016, in MeerKAT Science: On the Pathway to the SKA, 4
Calabretta, M. R., Staveley-Smith, L., & Barnes, D. G. 2014, PASA, 31, e007
Chang, T.-C., Pen, U.-L., Bandura, K., & Peterson, J. B. 2010, Natur, 466, 463
Hickish, J., et al. 2016, JAI, 5, 1641001
Jarvis, M., et al. 2016, in MeerKAT Science: On the Pathway to the SKA, 6
Johnson, P., et al. 2020, RAA, 20, 064
Lake, S. E., & Wright, E. L. 2016, OAJA, 1, 3
Nan, R., et al. 2011, IJMPD, 20, 989
A. Appendix

A.1 Beam model

<table>
<thead>
<tr>
<th>Index</th>
<th>XX</th>
<th>YY</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A (m)</td>
<td>B (m)</td>
<td>C (m)</td>
</tr>
<tr>
<td>01</td>
<td>297.76 (0.01)</td>
<td>61.90 (0.14)</td>
<td>−40.78 (0.38)</td>
</tr>
<tr>
<td>02</td>
<td>288.59 (0.01)</td>
<td>94.45 (0.23)</td>
<td>−61.65 (0.60)</td>
</tr>
<tr>
<td>03</td>
<td>291.40 (0.01)</td>
<td>76.97 (0.18)</td>
<td>−37.90 (0.47)</td>
</tr>
<tr>
<td>04</td>
<td>293.63 (0.01)</td>
<td>61.24 (0.19)</td>
<td>−19.72 (0.51)</td>
</tr>
<tr>
<td>05</td>
<td>295.88 (0.01)</td>
<td>81.11 (0.18)</td>
<td>−51.87 (0.48)</td>
</tr>
<tr>
<td>06</td>
<td>303.30 (0.01)</td>
<td>77.14 (0.25)</td>
<td>−63.12 (0.47)</td>
</tr>
<tr>
<td>07</td>
<td>307.93 (0.01)</td>
<td>56.37 (0.14)</td>
<td>−32.23 (0.37)</td>
</tr>
<tr>
<td>08</td>
<td>301.42 (0.01)</td>
<td>88.80 (0.27)</td>
<td>−71.80 (0.70)</td>
</tr>
<tr>
<td>09</td>
<td>291.24 (0.01)</td>
<td>105.87 (0.24)</td>
<td>−84.27 (0.62)</td>
</tr>
<tr>
<td>10</td>
<td>291.20 (0.02)</td>
<td>140.28 (0.33)</td>
<td>−164.66 (0.85)</td>
</tr>
<tr>
<td>11</td>
<td>289.79 (0.03)</td>
<td>137.75 (0.48)</td>
<td>−119.92 (1.25)</td>
</tr>
<tr>
<td>12</td>
<td>295.03 (0.02)</td>
<td>115.27 (0.31)</td>
<td>−134.72 (0.79)</td>
</tr>
<tr>
<td>13</td>
<td>302.00 (0.01)</td>
<td>78.51 (0.23)</td>
<td>−57.53 (0.61)</td>
</tr>
<tr>
<td>14</td>
<td>307.93 (0.01)</td>
<td>56.37 (0.14)</td>
<td>−32.23 (0.37)</td>
</tr>
<tr>
<td>15</td>
<td>301.42 (0.01)</td>
<td>88.80 (0.27)</td>
<td>−71.80 (0.70)</td>
</tr>
<tr>
<td>16</td>
<td>291.82 (0.02)</td>
<td>123.85 (0.32)</td>
<td>−142.91 (0.81)</td>
</tr>
</tbody>
</table>