

## HiFAST: An H I data calibration and imaging pipeline for FAST

Yingjie Jing<sup>1\*†</sup>, Jie Wang<sup>1,2,3</sup>, Chen Xu<sup>1,3</sup>, Ziming Liu<sup>1,3</sup>, Qingze Chen<sup>1,3</sup>, Tiantian Liang<sup>1,3</sup>,  
Jinlong Xu<sup>1,6</sup>, Yixian Cao<sup>4</sup>, Jing Wang<sup>5</sup>, Huijie Hu<sup>1,3</sup>, Chuan-Peng Zhang<sup>1,6</sup>, Qi Guo<sup>1,2,3</sup>,  
Liang Gao<sup>1,2,3,7</sup>, Mei Ai<sup>1,6,8</sup>, Hengqian Gan<sup>1,8</sup>, Xuyang Gao<sup>1,8</sup>, Jinlin Han<sup>1,8</sup>, Ligang Hou<sup>1</sup>,  
Zhipeng Hou<sup>1,3</sup>, Peng Jiang<sup>1,6</sup>, Xu Kong<sup>9,10</sup>, Fujia Li<sup>9,10</sup>, Zerui Liu<sup>1,3</sup>, Li Shao<sup>1</sup>, Hengxing Pan<sup>11</sup>,  
Jun Pan<sup>1,12</sup>, Lei Qian<sup>1,8</sup>, Jinghai Sun<sup>1</sup>, Ningyu Tang<sup>13</sup>, Qingliang Yang<sup>1</sup>, Bo Zhang<sup>1,8</sup>,  
Zhiyu Zhang<sup>14,15</sup>, and Ming Zhu<sup>1,6,8</sup>

<sup>1</sup>National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China;

<sup>2</sup>Institute for Frontiers in Astronomy and Astrophysics, Beijing Normal University, Beijing 102206, China;

<sup>3</sup>School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing 100049, China;

<sup>4</sup>Max Planck Institute for Extraterrestrial Physics, Garching 85748, Germany;

<sup>5</sup>Kavli Institute for Astronomy and Astrophysics, Peking University, Beijing 100871, China;

<sup>6</sup>Guizhou Radio Astronomical Observatory, Guizhou University, Guiyang 550000, China;

<sup>7</sup>Institute of Computational Cosmology, Department of Physics, University of Durham, Durham DH1 3LE, UK;

<sup>8</sup>CAS Key Laboratory of FAST, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China;

<sup>9</sup>Deep Space Exploration Laboratory/Department of Astronomy, University of Science and Technology of China, Hefei 230026, China;

<sup>10</sup>School of Astronomy and Space Science, University of Science and Technology of China, Hefei 230026, China;

<sup>11</sup>Astrophysics, Department of Physics, University of Oxford, Oxford OX1 3RH, UK;

<sup>12</sup>College of Earth Sciences, Guilin University of Technology, Guilin 541004, China;

<sup>13</sup>Department of Physics, Anhui Normal University, Wuhu 241002, China;

<sup>14</sup>School of Astronomy and Space Science, Nanjing University, Nanjing 210023, China;

<sup>15</sup>Key Laboratory of Modern Astronomy and Astrophysics, Nanjing University, Nanjing 210023, China

Received October 25, 2023; accepted January 29, 2024; published online April 12, 2024

The Five-hundred-meter Aperture Spherical radio Telescope (FAST) has the largest aperture and a 19-beam L-band receiver, making it powerful for investigating the neutral hydrogen atomic gas (H I) in the universe. We present HiFAST (<https://hifast.readthedocs.io>), a dedicated, modular, and self-contained calibration and imaging pipeline for processing the H I data of FAST. The pipeline consists of frequency-dependent noise diode calibration, baseline fitting, standing wave removal using an FFT-based method, flux density calibration, stray radiation correction, and gridding to produce data cubes. These modules can be combined as needed to process the data from most FAST observation modes: tracking, drift scanning, On-The-Fly mapping, and most of their variants. With HiFAST, the root-mean-square (RMS) noises of the calibrated spectra from all 19 beams were only slightly (~5%) higher than the theoretical expectation. The results for the extended source M33 and the point sources are consistent with the results from Arecibo. The moment maps (0, 1 and 2) of M33 agree well with the results from the Arecibo Galaxy Environment Survey (AGES) with a fractional difference of less than 10%. For a common sample of 221 sources with signal-to-noise ratio  $S/N > 10$  from the Arecibo Legacy Fast ALFA (ALFALFA) survey, the mean value of fractional difference in the integrated flux density,  $S_{\text{int}}$ , between the two datasets is approximately 0.005%, with a dispersion of 15.4%. Further checks on the integrated flux density of 23 sources with seven observations indicate that the variance in the flux density of the source with luminous objects ( $S_{\text{int}} > 2.5 \text{ Jy km s}^{-1}$ ) is less than 5%. Our tests suggest that the FAST telescope, with the efficient, precise, and user-friendly pipeline HiFAST, will yield numerous significant scientific findings in the investigation of the H I in the universe.

\*Corresponding author (email: [jjy@nao.cas.cn](mailto:jjy@nao.cas.cn))

†FAST fellow

**radio telescopes and instrumentation, H I regions and 21-cm lines, observation and data reduction techniques, algorithms and implementation****PACS number(s):** 95.55.Jz, 98.38.Gt, 95.75.-z, 07.05.Kf

---

**Citation:** Y. Jing, J. Wang, C. Xu, Z. Liu, Q. Chen, T. Liang, J. Xu, Y. Cao, J. Wang, H. Hu, C.-P. Zhang, Q. Guo, L. Gao, M. Ai, H. Gan, X. Gao, J. Han, L. Hou, Z. Hou, P. Jiang, X. Kong, F. Li, Z. Liu, L. Shao, H. Pan, J. Pan, L. Qian, J. Sun, N. Tang, Q. Yang, B. Zhang, Z. Zhang, and M. Zhu, HiFAST: An H I data calibration and imaging pipeline for FAST, *Sci. China-Phys. Mech. Astron.* **67**, 259514 (2024), <https://doi.org/10.1007/s11433-023-2333-8>

---

## 1 Introduction

The Five-hundred-meter Aperture Spherical radio Telescope (FAST) is the largest single-dish and most sensitive radio telescope in the world [1]. It has an illuminated aperture of 300 m and a 19-beam L-band receiver that covers frequencies from 1050 to 1450 MHz [2-5], making it a powerful instrument for observing the neutral hydrogen atomic gas (H I lines). The raw spectral data must be calibrated in terms of amplitude and transformed by frequency before any scientific analysis or comparison with other telescopes can be conducted. In this paper, we introduce HiFAST<sup>1</sup>, a modular and flexible H I data calibration and imaging pipeline for FAST.

The paper is organised as follows. We describe the pipeline workflow in sect. 2. The calibration procedures, including temperature and flux calibration, are presented in sect. 3. The Radio Frequency Interference (RFI) flagging and Doppler correction are described in sects. 4 and 5. The imaging procedures are described in sect. 6. A pipeline verification test on FAST data is presented in sect. 7. The code implementations are described in sect. 8, and a concluding summary is provided in sect. 9. More details and tests on the module flux calibration, standing wave removal, and stray radiation correction are extensively described in separate papers: Liu et al., Xu et al., and Chen et al. (in preparation, hereafter referred to as papers II, III, and IV, respectively).

## 2 Scheme of HiFAST

HiFAST supports the processing of the data observed in various modes, including tracking, drifting scan, and On-The-Fly (OTF) mapping, as well as most of their variants<sup>2</sup> (see more details on FAST observation modes in ref. [3]).

In Figure 1, we present a typical calibration and imaging workflow in HiFAST. The raw spectral data recorded by FAST are stored in the Flexible Image Transport System (FITS) files, mainly including sky frequency and uncalibrated intensity (power) in each spectral channel recorded as

single-precision floating point data. Additionally, the trajectory of the phase centre of the feed is stored in a corresponding Office Open XML Workbook (.xlsx) file. HiFAST loads the raw spectral data from the FITS files and calibrates them to the temperature using a noise diode as reference power input. To obtain the antenna temperature of the H I line, the standing wave is first removed, followed by the off-source subtraction or the removal of the baseline to eliminate the system temperature. The temperature is then calibrated to the flux density. Meanwhile, HiFAST calculates the right ascension (RA) and the declination (DEC) of each spectrum using the three-dimensional coordinates of the feed and the measurement time. RFI flagging and Doppler correction procedures are also included. For mapping observation modes, such as drifting scans or OTF, the imaging process consists of an iterative stray radiation correction and gridding. The final cube data is saved in a standard FITS file. These modules can be combined in different ways depending on user needs. Detailed descriptions of these modules are presented in the following sections.

## 3 Calibration procedures

The raw observation data record the detected power  $P$  of the receiver output, comprising various contributions [3, 6, 7]. This can be expressed as:

$$P = gT \quad (1)$$

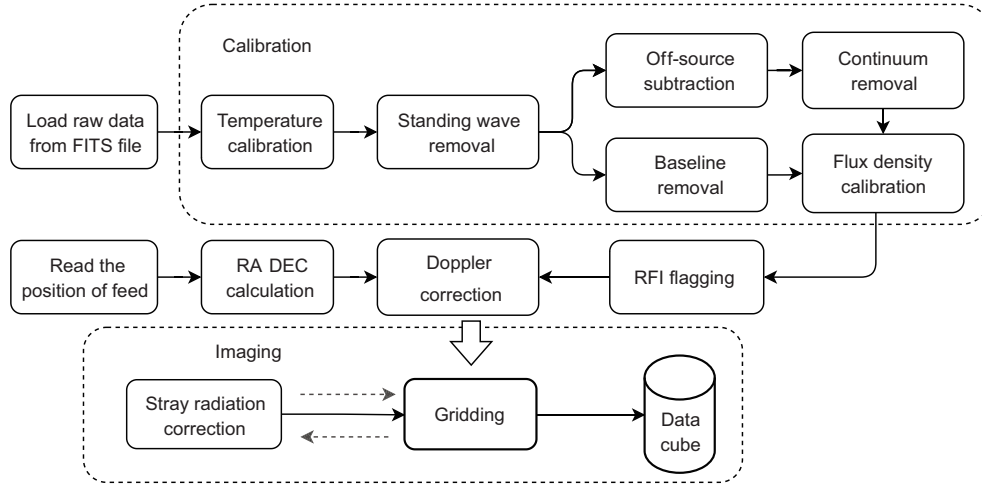
$$= g(T_{\text{src}} + T_{\text{sys}}) \quad (2)$$

$$= g(T_{\text{src}} + T_{\text{rec}} + T_{\text{sw}} + T_{\text{gr}} + T_{\text{atm}} + T_{\text{bg}} [+T_{\text{cal}}]), \quad (3)$$

where  $g$  represents the gain of the receiving system.  $T_{\text{src}}$  is the antenna temperature induced by the flux density of the astronomical source, including line profiles  $T_{\text{src}}^{\text{line}}$  (emission or absorption) and a continuum contribution  $T_{\text{src}}^{\text{cont}}$ .  $T_{\text{sys}}$  is the system temperature that accounts for all other contributions. Specifically,  $T_{\text{rec}}$  represents the receiver's noise temperature,  $T_{\text{sw}}$  is the contribution from standing waves,  $T_{\text{gr}}$  arises from

1) <https://hifast.readthedocs.io>.

2) <https://fast.bao.ac.cn/cms/article/24/>.



**Figure 1** Flowchart of the typical HiFAST calibration and imaging process.

ground radiation,  $T_{\text{atm}}$  is the atmospheric contribution,  $T_{\text{bg}}$  includes the cosmic microwave background and the Galactic emissions, and  $T_{\text{cal}}$  is the calibration noise injected by a noise diode for calibration purposes. The term  $T_{\text{cal}}$  only contributes when the noise diode is activated. Previous studies detail the signal path in the telescope [2, 8, 9].

For scientific analysis, the raw data are calibrated to transform power  $P$  to temperature  $T$  using a noise diode as reference power input (see sect. 3.1). Subsequently, the line temperature  $T_{\text{src}}^{\text{line}}$  is derived (refer to sect. 3.2). A standard source with known flux density is used to convert the temperature to the flux density (see sect. 3.3). This calibration is applied to each spectrum, obtained under different beams and polarizations ( $XX$  and  $YY$ ). Figure 2 illustrates this calibration for 2200 spectra in polarization mode  $XX$ . The following subsections describe each calibration stage in detail.

### 3.1 Temperature calibration

FAST utilises a noise diode with a known temperature (approximately 1 or 10 K) to measure the gain ( $g$ ) and calibrate the recorded power to temperature. The stability of the gain  $g$  can be maintained for several minutes, with fluctuations on the order of a few percent for FAST [3]. Consequently, the noise diode needs to be periodically injected, such as for 1 s every 5 min. Within each injection period, the measured gain, denoted as  $g_m$ , can be determined using the equation:

$$g_m(\nu) = \frac{[P_{\text{cal.on}} - P_{\text{cal.off}}^{\text{near}}]^{\text{smo}}(\nu)}{T_{\text{cal}}^{\text{smo}}(\nu)}, \quad (4)$$

where  $T_{\text{cal}}$  represents the pre-measured temperature of the noise diode<sup>3)</sup>,  $P_{\text{cal.on}}$  corresponds to the power measured with

the noise diode activated, and  $P_{\text{cal.off}}^{\text{near}}$  represents the power of the closest spectrum obtained with the noise diode deactivated, close in time to  $P_{\text{cal.on}}$ . The term “smo” denotes the process of smoothing the frequency-dependent power and temperature of the noise diode along the frequency  $\nu$  (i.e., channel) using a Gaussian filter or moving average. This smoothing reduces the variance induced by random noise. The typical smoothing width is 2 MHz, which covers 263 channels in wide-band data<sup>4)</sup> with a frequency resolution of 7.6 kHz.

The calibration of power to temperature for the spectra obtained when the noise diode is deactivated is performed using the equation:

$$T_{\text{cal.off}}(\nu) = \frac{P_{\text{cal.off}}}{g_m(\nu)}. \quad (5)$$

For the spectra obtained when the noise diode is activated (with the contribution from the noise diode subtracted), the calibration of power to temperature is given by

$$T_{\text{cal.on}}(\nu) = \frac{P_{\text{cal.on}}}{g_m(\nu)} - T_{\text{cal}}^{\text{smo}}(\nu), \quad (6)$$

where  $g_m$  represents the nearest measured gain in time.

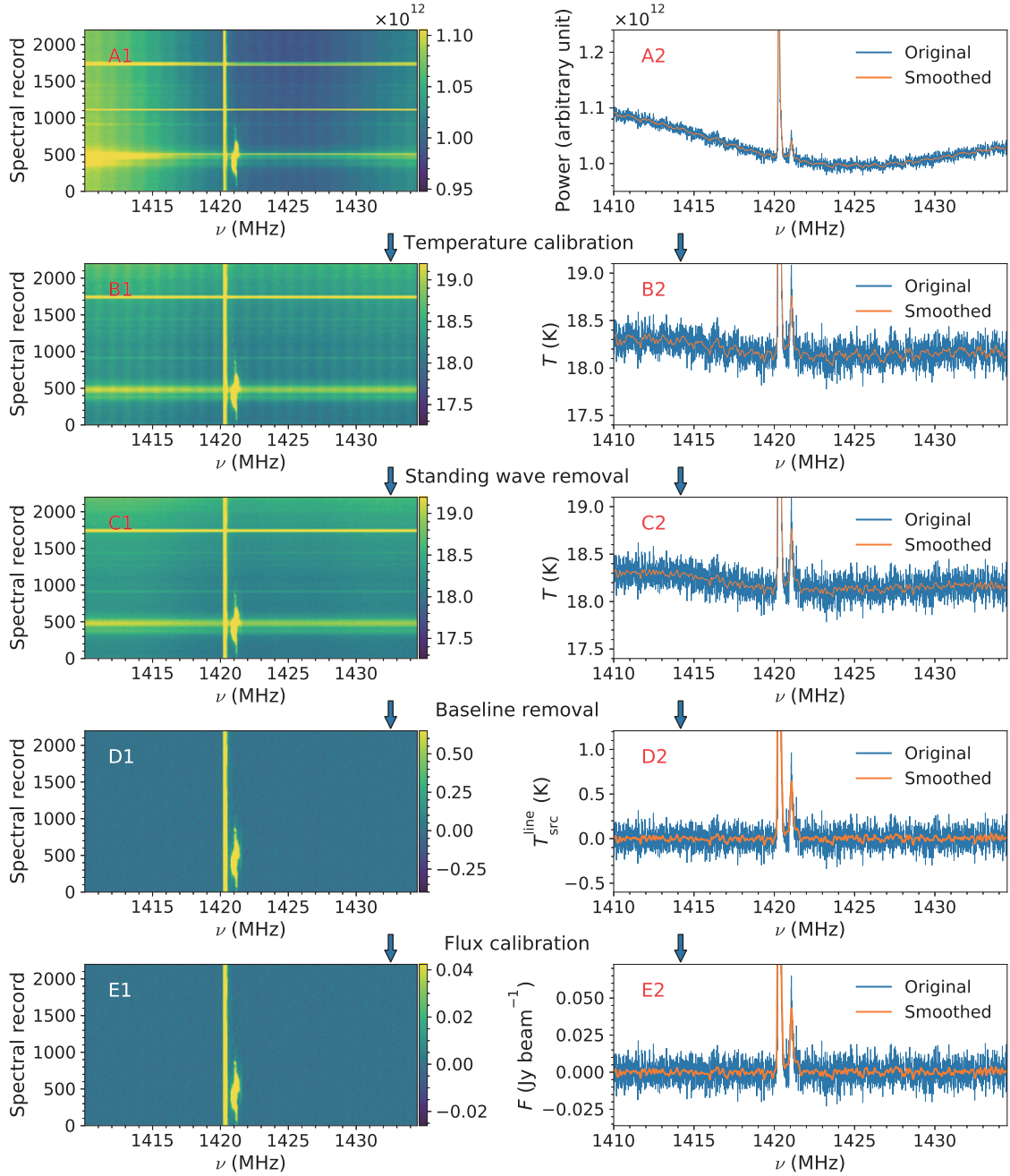
An example of the temperature calibration process is shown in the top two panels of Figure 2. The power of the spectrum is shown in the first rows, A1 and A2, while the calibrated temperature of the spectra is presented in the subsequent rows, B1 and B2.

### 3.2 Source temperature

The temperature  $T$  obtained by gain ( $g$ ) calibration is the composition of the target astronomical source temperature

3) [https://fast.bao.ac.cn/cms/category/telescope\\_performance\\_en/noise\\_diode\\_calibration\\_report\\_en/](https://fast.bao.ac.cn/cms/category/telescope_performance_en/noise_diode_calibration_report_en/).

4) <https://fast.bao.ac.cn/cms/article/26/>.



**Figure 2** (Color online) An example of typical calibration process for one polarization of spectra from one beam taken in a mapping scan. Left panels: waterfall plots, i.e., power (A), temperature (B, C, and D), and flux density (E) as a function of frequency  $\nu$  and spectral record number, for 2200 continuous spectral records. Each spectrum is recorded with a 0.5-s exposure time. Right panels: the average of 10 consecutive records (corresponding to 5 s of integration time, depicted in blue lines) selected from the left panels. The smoothed spectra are shown in orange lines for enhanced clarity. The arrows between the panels indicate the sequence of procedures: A→B (refer to sect. 3.1) involves calibrating the raw power data to the temperature of the spectra using the noise diode; B→C entails the removal of standing waves via the FFT-filter method; C→D involves baseline subtraction; and D→E includes calibrating the temperature to the flux density using a stable standard flux calibrator.

$T_{\text{src}}$  and the system temperature  $T_{\text{sys}}$ .  $T_{\text{src}}$  potentially includes line profiles  $T_{\text{src}}^{\text{line}}$  (emission or absorption) and a continuum contribution  $T_{\text{src}}^{\text{cont}}$ . In studies of H I the primary objective is to determine the line profile contribution  $T_{\text{src}}^{\text{line}}$ .

To obtain  $T_{\text{src}}^{\text{line}}$ , there are two approaches that can be

used. The first method involves measuring  $T_{\text{sys}}$  by observing a blank sky region as an off-source measurement [6]. This measurement is subtracted from the total temperature  $T$ . Subsequently, a polynomial baseline fitting can be applied to eliminate the continuum contribution (including  $T_{\text{src}}^{\text{cont}}$  and

potential changes in  $T_{\text{sys}}$ ) to extract line profiles. The second method involves fitting a baseline directly onto source-on spectra across frequencies and then subtracting it, effectively removing both  $T_{\text{sys}}$  and the continuum contribution.

However, both techniques have limitations. It can be difficult or even impossible to select off-source observations in certain cases, such as extended sources like M31 or the ubiquitous H I lines of the Milky Way. Additionally, fitting the baseline can be a challenge when  $T_{\text{sys}}$  has a complex frequency dependence. To address these issues, HiFAST has modules designed to handle these two methods, depending on the specific situation.

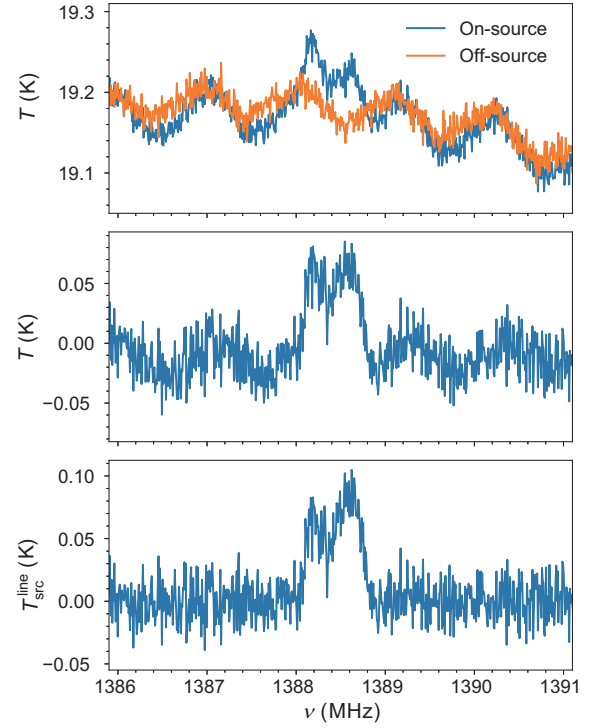
In particular,  $T_{\text{sw}}$ , which is involved in these methods, varies slightly over time (see details in Figure 3 of sect. 3.2.1 and Paper III), leading to different  $T_{\text{sys}}$  values in the off-source and on-source observations. The primary standing wave in FAST exhibits a period of about 1 MHz, which is comparable to or smaller than the width of numerous source signals. Consequently, it cannot be eliminated during baseline fitting either. Therefore, a separate method is needed to eliminate it (sect. 3.2.3).

In the following subsections, we describe the required modules separately in detail.

### 3.2.1 Off-source subtraction

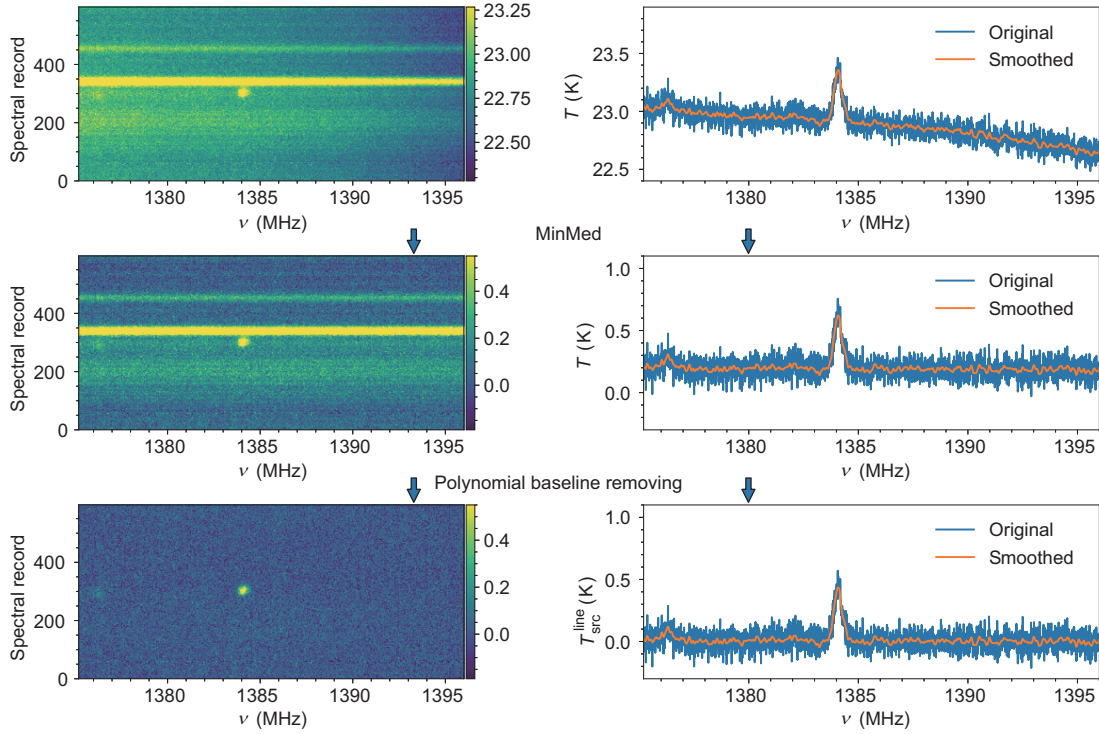
The estimation of the off-source spectrum is different in the tracking and mapping modes. We will explain these two approaches below.

In tracking mode, position switching is commonly used to acquire blank sky information. This involves the telescope observing the target object for a certain amount of time and then moving to a blank sky region for an equal duration as an off-source observation. An example of this data processing is illustrated in Figure 3. The top panel displays the temperature-calibrated spectrum of the on-source (blue) and off-source (orange) observations, with a  $T_{\text{sys}}$  of approximately 19.06 K, including a sine standing wave. Additionally, an H I emission line is present at a frequency of approximately 1388.5 MHz. The middle panel shows the results of subtracting the off-source spectrum from the on-source spectrum. The temperature of the baseline is now close to zero, as most of the  $T_{\text{sys}}$  has been removed. However, a minor residual sine wave remains, caused by the phase shift of the standing wave with time. This can be eliminated using the sine fitting method described in sect. 3.2.3, as seen in the bottom panel. An alternative approach would be to remove the standing wave before performing the off-source subtraction, as the standing wave in the off-source observations differs from that in the on-source observations.



**Figure 3** (Color online) An example of off-source subtraction in position-switching tracking mode. The top panel shows the calibrated temperature of on-source (blue) and off-source (orange) observation, which is a blank sky region close to the source. The middle panel shows the target spectrum obtained by subtracting the off-source from the on-source spectrum. The bottom panel illustrates the result after removing the residual standing wave.

In drifting scans and OTF mapping observations for point or small extended sources, off-source observation should be carefully selected [6, 10] in general. The “MinMed” method is usually used in this process. This method was introduced by refs. [11, 12] and has been used in a number of investigations [13, 14]. The process involves dividing the scanned spectra into segments over time for each channel, calculating the median value for each segment, and then selecting the minimum of these medians as the off-source spectra. The median value is unaffected by the on-source spectra data when the angular size of the sources is sufficiently small compared with the area covered by each segment. Figure 4 shows an example of data processing in this method, where the waterfall plot and the average of 10 consecutive records near the target source are shown in the left and right panels, respectively. Initially, we applied the “MinMed” method to estimate and subtract the off-source from temperature-calibrated spectra. The resulting waterfall plot for a portion of the scanning data is displayed in the left panels. Notably, a source was present at a frequency of approximately 1384 MHz, located near the 325th spectral record. In addition, a few spectra with elevated continuum levels are present around the 329th and 340th spectral records. To further refine the data, we removed



**Figure 4** (Color online) Similar to Figure 2, but for the “MinMed” method used in off-source subtraction. Top panels: spectra with the standing wave removed; middle panels: spectra applied with the MinMed method; bottom panels: spectra applied with a lower-order polynomial baseline fitting to remove continuum contribution in the source and other possible origins.

a lower-order polynomial baseline from each spectrum using the method described in sect. 3.2.2. The refined result can be found in the bottom panels of Figure 4.

### 3.2.2 Baseline fitting

The baseline fitting process involves isolating the line profile  $T_{\text{src}}^{\text{line}}$  by treating the continuum contribution  $T_{\text{src}}^{\text{cont}}$  and the system temperature component  $T_{\text{sys}}$  as the baseline. To obtain an accurate line profile, the baseline must be fitted in a signal-free region and then removed from the target spectrum. In drifting scan and OTF modes, manually selecting a signal-free region for baseline fitting can be difficult and time-consuming, especially when dealing with numerous spectra. To address this issue, the arPLS (asymmetrically reweighted penalized least squares smoothing) [15] method is employed for automatic baseline fitting without having to select a signal-free region beforehand. The arPLS method is a type of penalized least squares method [16, 17] and is effective even in the presence of small noise in the data.

In the arPLS method, the baseline  $\mathbf{z}$  of a spectrum with data points  $\mathbf{y}$  is obtained using the following equation:

$$\mathbf{z} = (\mathbf{W} + \lambda \mathbf{D}^T \mathbf{D})^{-1} \mathbf{W} \mathbf{y}, \quad (7)$$

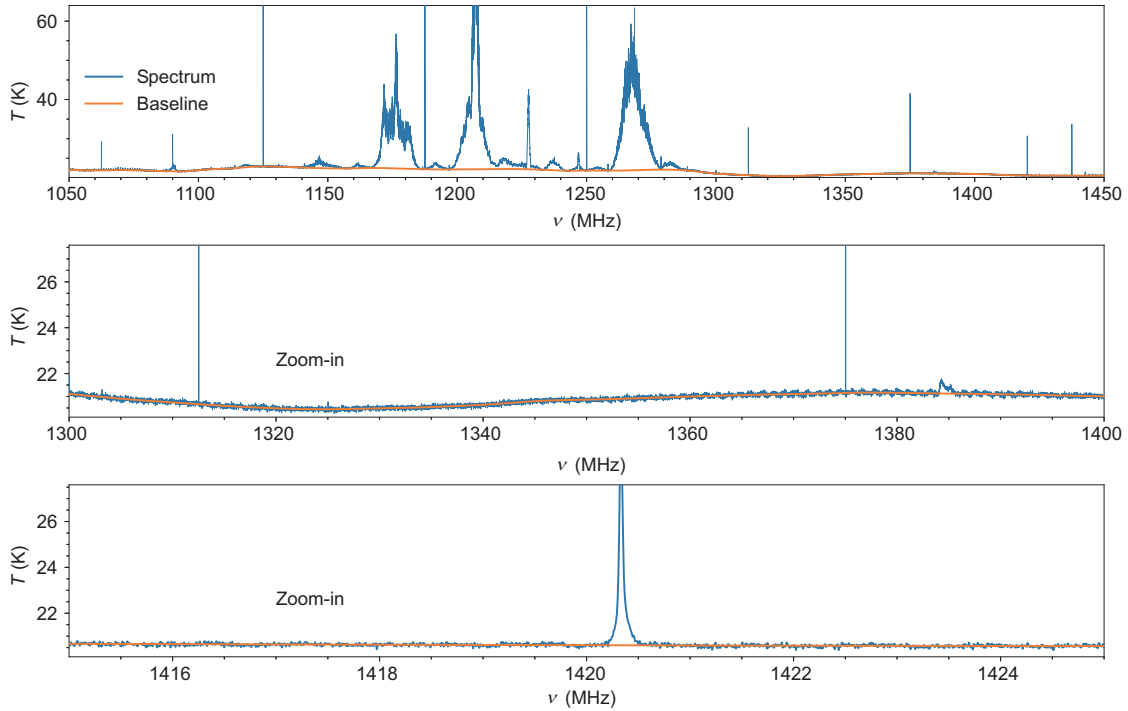
where  $\mathbf{D}$  is the second-order difference matrix,  $\lambda$  is a param-

eter that adjusts the balance between  $\mathbf{y}$  and  $\mathbf{z}$ , and  $\mathbf{W}$  is a diagonal matrix with weight vector  $\mathbf{w}$  of  $\mathbf{y}$  on its diagonal. The arPLS method iteratively adjusts and applies weights to the data points, refining the baseline until the weight changes become negligible or fall below a predetermined threshold.

In Figure 5, we illustrate an example of the baseline fitting using the arPLS method. The top panel shows the complete spectrum, while the middle and bottom panels show the zoom-in of the upper panel for a better view. We find that the arPLS method performs well in the Milky Way region ( $\sim 1420.4$  MHz in the bottom panel), the signal region ( $\sim 1385$  MHz in the middle panel), and the RFI-contaminated region (1140-1300 MHz).

However, it is important to note that the iteration of the weight adjustment and application process in the arPLS method is time-intensive. For a rapid and efficient fitting of the baseline, an alternative method, arPoly, is provided, in which eq. (7) is replaced by a polynomial fitting. This method is significantly faster than arPLS, but is only suitable for simple baseline situations, such as fitting  $T_{\text{src}}^{\text{cont}}$  after off-source subtraction.

One way to improve the quality of the baseline fitting is by stacking the spectra. Typically, the integration duration for recording a single spectrum does not exceed 1 s. In the case of wide-band data with a 7.6 kHz frequency resolution, noise



**Figure 5** (Color online) Baseline fitting result using arPLS. The blue and orange lines represent the spectrum and the fitted baseline, respectively. The upper panel displays the complete spectrum, while the middle and bottom panels provide a zoomed-in view.

levels during this short integration time are typically between 0.1 and 0.3 K. These elevated noise levels can negatively impact the effectiveness of baseline removal algorithms, such as arPLS. To mitigate this issue, spectra recorded over several minutes are combined through stacking. This method presupposes the stability of the baseline over the duration, allowing the fitting of a common baseline to the combined spectra. Once this common baseline is determined and subtracted, a lower-order arPoly fit can be employed on each spectrum within the stack to adjust for any remaining baseline variations. Figure 2 illustrates this process, where a comparison between panels C1 (C2) and D1 (D2) reveals the effective removal of the baseline.

### 3.2.3 Standing wave fitting

The standing wave  $T_{sw}$  is induced by signals entering the receiving system with different paths and is widely observed in radio telescopes [18–20]. For FAST, the primary standing wave arises from the reflection between the reflector and the receiver cabin [3], where the path length difference of the signal is approximately 276 m (approximately twice the distance between the reflector and the receiver cabin). This difference results in a delay of approximately 0.92 ms and gives rise to a standing wave with a period of approximately 1.09 MHz across the spectrum.

HiFAST provides two methods for removing standing

waves: sine-fitting and FFT-filter. Both approaches require a preliminary baseline removal. After isolating the standing wave, it is possible to subtract it from the spectra that include the baseline. This allows for the selective removal of only the standing wave, enabling off-source subtraction of the processed data.

The sine-fitting method utilizes the least-squares algorithm to fit the spectrum in the signal-free region using a sine function as follows:

$$T_{sw}(v) = a_0 + a_1 \sin(2\pi f v + \phi), \quad (8)$$

where  $a_0$ ,  $a_1$ ,  $f$ , and  $\phi$  are the standing wave's displacement, amplitude, frequency, and phase, respectively. The initial guess value of  $f$  is set to  $\sim 0.92$ . With the baseline already subtracted, the signal-free regions are identifiable by selecting temperatures below a threshold, typically several times the root-mean-square (RMS) noise level of the spectra.

In the FFT-filter approach, the Fast Fourier Transform (FFT) is initially applied to the spectra, followed by the identification of modes associated with standing waves in Fourier space. The standing wave is then reconstructed through an inverse Fourier transformation of these modes. Illustrations of the FFT-filter technique's efficacy in removing standing waves are presented in panels C1 and C2 of Figure 2, which depict a waterfall plot and a 5-s integrated spectrum, respectively. Comprehensive details of the FFT-filter method, including criteria for selecting standing wave components in

Fourier space and its comparison with alternative methods like running median, running mean, and sine-fitting, are elaborated in Paper III.

### 3.2.4 RMS noise level

In this subsection, we evaluate the calibration performance by comparing the measured RMS noise level with the theoretical expectations. Considering that the temperature and the RMS vary with the frequency and time, we performed the test on 300 spectra with an integration time of 0.5 s and a frequency ranging from 1415 to 1417 MHz.

For each spectrum, we estimate the theoretical RMS using the following equation:

$$\text{RMS}_{\text{theoretical}} = \frac{T_{\text{input}}}{\sqrt{\Delta\nu\Delta t}}, \quad (9)$$

where  $\Delta\nu$  represents the width of the spectral channel (7.6 kHz), and  $\Delta t$  denotes the integration time (0.5 s). We selected  $T_{\text{input}}$  as the temperature  $T$  obtained from the temperature calibration as described in sect. 3.1. To calculate  $T_{\text{input}}$ , we use the mean value of  $T$  in the frequency range, which is typically approximately 20 K. It is important to note that the actual RMS was considered by measuring the spectra after removing the baseline and standing waves. Approximately 262 channels in the frequency range were used for the calculations.

Figure 6 presents a comparison between the theoretical RMS (orange) and the actual measured RMS (blue) for the XX polarization of each beam. Our analysis revealed that the measured RMS values were slightly larger than the theoretical expectations. On average, the fractional difference, calculated by  $(\text{RMS}_{\text{measured}} - \text{RMS}_{\text{theoretical}}) / \text{RMS}_{\text{theoretical}}$ , is approximately 5%, with a 1-sigma dispersion of approximately 4.5% for each beam.

### 3.3 Flux density calibration

Once the temperature of the line  $T_{\text{src}}^{\text{line}}$  was determined, it was necessary to convert it into flux density by dividing it with the gain  $G$  of FAST. The gain  $G$  can vary depending on the telescope's condition during observation. Therefore, it is essential to observe a standard source with a precisely known flux as a calibrator observation before or after the target source observation. The gain  $G$  is determined by calculating the ratio of the antenna temperature to the flux density of the calibrator. For detailed procedures on processing calibrator observation data, please refer to Paper II.

However, observing a calibrator every time may not always be feasible due to time constraints and the availability of calibrators. Alternatively, we can utilise a pre-measured

gain value obtained from previous observations. A significant concern is the accuracy of these pre-measured gains. In Paper II, we examine the stability of gain  $G$  over an extended period of observation. We found that using the pre-measured gain  $G$  introduces an uncertainty of less than 3% in general, although it can escalate to approximately 8% in certain instances. We suggest that studies that need an accuracy better than 3% in the flux density should incorporate the observation of a calibrator together with their target sources.

HiFAST provides two options for pre-determined gain values that are suitable for observations without a calibrator. The first option, as outlined in Paper II, has a high-frequency resolution of 5 MHz. The second option, as reported by ref. [3], is based on observations of the stable calibrator 3C286 using 19 FAST beams at varying zenith angles during August and December 2019. The frequency resolution in Paper II is finer than the 50 MHz resolution described in ref. [3].

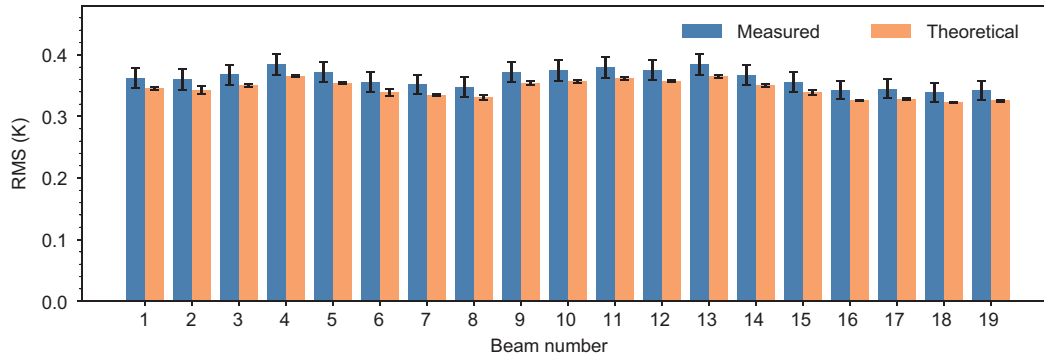
## 4 RFI flagging

The signal of interest is inevitably contaminated by artificial radio emission, which is normally called RFI. This noise is inevitable for large single-dish radio telescopes such as FAST. The RFI in FAST can be roughly divided into three types: harmonic RFI, wideband RFI, and narrowband RFI.

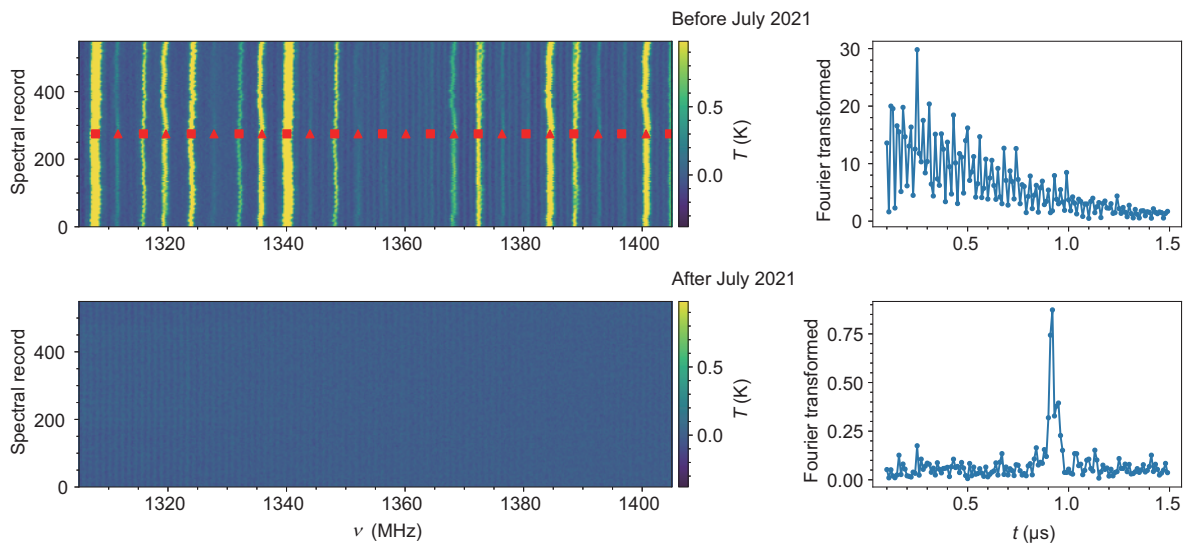
The harmonic RFI manifests as a sequence of bump signals with a coherent period in the frequency domain. Consequently, this kind of RFI disrupts the entire frequency range and effectively obstructs the channels of the target signal. We have detected several instances of such harmonic RFI. In this work, we provide a brief overview of an example with a period of approximately 8.1 MHz, which previously posed significant challenges to early H I observations of FAST. Further information regarding other periods associated with this type of RFI can be found in Paper III.

The 8.1 MHz harmonic RFI has a width of about 1 MHz and a period of approximately 8.1 MHz. In fact, there are two or three sets of harmonic RFI with an 8.1 MHz period in the spectra. Therefore, they can be easily identified as non-harmonic RFI [3, 21]. These sets of 8.1 MHz harmonic RFI spread over the entire bandwidth of the 19-beam L-band receiver, significantly impacting the observation of faint H I sources. In the top panel of Figure 7, these RFIs are clearly visible in the waterfall plot (indicated by yellow and green vertical strips) and in the Fourier space of the 10-s integrated example spectrum (as high spikes). Specifically, the waterfall plot shows two groups of 8 MHz harmonic RFI, with the positions of the bumps marked by red square and triangle markers. Fortunately, the source of this RFI has been identified in the three sets of compressors in the cabin and was shielded





**Figure 6** (Color online) Comparison of RMS noise measured from spectra (blue) and the theoretical expectation (orange) for the  $XX$  polarization of each beam. The vertical bar represents the mean value of the 300 spectra for the  $XX$  polarization of each beam. The error bars represent the 1-sigma dispersion of the values.



**Figure 7** (Color online) The 8.1 MHz harmonic RFI in the comparison of sample data observed before July 2021 (top panels) and after that (bottom panels). The left panels show the waterfall plot, i.e., temperature (coded by colour) as a function of frequency and the spectral record number for 500 spectra. The red square and triangle markers in the top left panel denote the position of two groups of 8.1 MHz RFIs, which disappear in the bottom left panel. The right panels show the Fourier space of 10-s integrated spectra selected from the corresponding left panel.

in July 2021. The bottom panels show the results observed after July 2021. The interference disappeared in the waterfall plot, and only the 1 MHz standing wave component ( $0.92 \mu\text{s}$ ) remained.

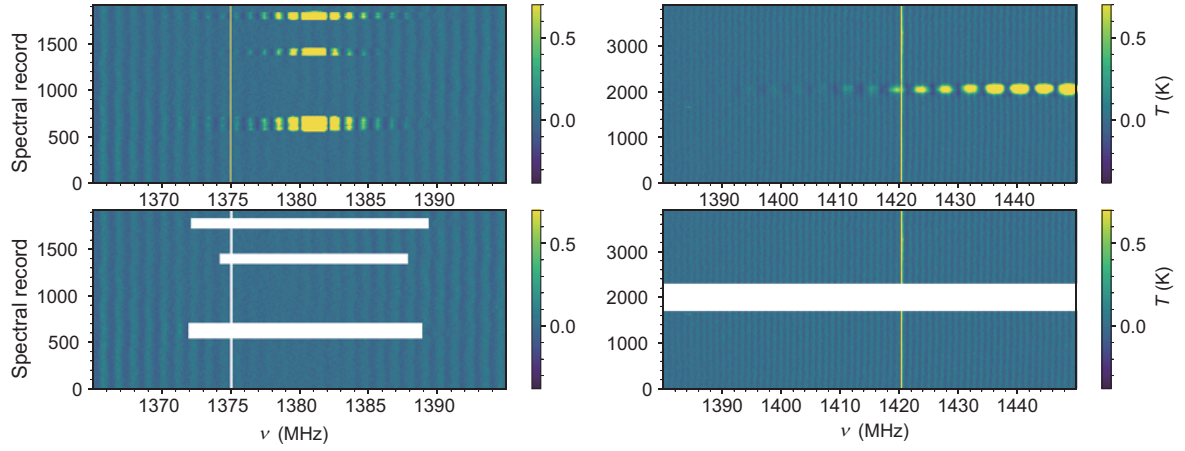
The wideband RFI is caused by satellites or civil aviation in the sky, and primarily affects the spectra in the frequency range between 1155 and 1295 MHz. The probability of occurrence can reach 70%, and even 99% based on statistical analysis [21] (based on  $\sim 300$  h of FAST observation data). Observations in this frequency range should generally be avoided when using FAST. If necessary, a thorough check should be conducted before the observation by referring to ref. [21].

The narrowband RFI typically affects only one or two channels in spectra stored in the wide-band data with a frequency resolution of 7.6 kHz. An example can be found in

the left panels of Figure 8, the region near 1375 MHz. The narrowband RFI was significantly reduced after April 2019 with the installation of electromagnetic shielding [3], and there are only a few narrowband RFI left across the entire bandwidth (Figure 24 of ref. [3]).

HiFAST provides several automatic processes to mask the RFI considering that the RFI is different from the  $H\text{I}$  signal in the polarization, time domain, and the possibility of accruing 19 beams. Usually,  $H\text{I}$  emissions are not polarized and tend to last for a short time, not across multibeams.

As an example, Figure 8 displays a receipt to mask the RFI at 1381 MHz related to GPS L3 emissions [22, 23] in the left panels, and a wideband RFI larger than 1300 MHz possibly caused by geostationary satellites in the right panels. The upper panels illustrate the spectra temperature (coded by colour) in the frequency-time plane, and the bottom panels



**Figure 8** (Color online) RFI flagging procedure for two different spectral frequency ranges. Each panel shows the spectra temperature (coded by colour) as a function of frequency and time. The upper panels show the original spectra, and the bottom panels show the corresponding RFI candidates are flagged as blank regions. Note that the regions near 1420 MHz are the Milky Way emissions, which are not flagged as desired.

show the corresponding flagged RFI regions for RFI candidates as blank areas. The three regions centred at a frequency of 1380 MHz and located near the 550th, 1460th, and 1500th positions are examples of wideband RFI due to GPS L3 emissions. We identified and flagged them mainly based on their abrupt changes in time at the beginning and end of their occurrence. The right panels demonstrate a type of RFI that appeared sporadically over a wide frequency range larger than 50 MHz. The regions around 1420 MHz were kept because they were H I emissions from the Milky Way.

Certainly, automatic processes cannot identify all RFI. HiFAST offers a manual approach as a supplementary method to mask RFI. This involves selecting box regions to cover the RFI in the waterfall plots viewed in CARTA [24].

## 5 Doppler correction

In FAST, observations occur in a topocentric mode, where the reference frame aligns with the telescope's geographical position. This frame differs from the standard in astrophysical research because it does not account for Earth's rotation, its orbital motion around the Sun, or the solar system's trajectory within the Milky Way. To rectify this and correct for the Doppler effect in the spectrum, a transformation to an alternate reference frame is necessary. HiFAST offers two such frames: the Heliocentric frame, considering Earth's rotation and orbit around the Sun, and the Local Standard of Rest (LSR), which additionally incorporates the solar system's motion relative to nearby stars.

Doppler correction varies depending on the time of observation, the pointing position in the sky, and the telescope's location. Although this variation might be minor in tracking mode, it is crucial in scanning mode, especially when

spanning a wide sky area or conducting observations on different days. Therefore, in HiFAST, the Doppler correction is applied individually to each spectrum. Subsequently, all spectra are interpolated onto a uniform frequency sampling grid.

## 6 Imaging procedures

### 6.1 Stray radiation correction

The stray radiation arises from H I emission entering the telescope receiver through the side lobes, rather than the main beam. In order to accurately determine the flux of the target sources, it is essential to take into account the effect of stray radiation. The measured temperature  $T_a$  consists of two parts: one from the main beam (MB) and the other from the stray pattern (SP). Our algorithm does not attempt to resolve the temperature distribution in the main beam, but rather derives a beam-averaged brightness temperature  $T_b$  by solving the following equation:

$$T_b(x, y) = \frac{T_a(x, y)}{\eta_{MB}} - \frac{1}{\eta_{MB}} \int_{SP} P(x-x', y-y') T_b(x', y') dx' dy', \quad (10)$$

where  $P(x, y)$  denotes the beam pattern and  $\eta_{MB}$  represents the main beam efficiency, defined as:

$$\eta_{MB} = \int_{MB} P(x, y) dx dy. \quad (11)$$

This procedure is described in more detail in several publications, such as refs. [25-28]. In HiFAST, the correction is confined to near-field stray radiation and utilises the

beam pattern,  $P(x, y)$ , from the observational results of Radio Galaxy PKS 0531+19 as described in Paper IV. We implement iterative processes, beginning with the observed  $T_a(x, y)$  as an initial approximation of  $T_b(x, y)$  on the right side of eq. (10), and then substituting the outcome into the subsequent iteration. This correction process is applied separately to each spectrum and each beam, but  $T_b(x, y)$  is updated simultaneously in the entire target range before the next iterative round.

This comprehensive procedure in HiFAST can be found in Paper IV. In Paper IV, we also provide the beam pattern for each beam and a thorough examination of the stray radiation correction on the extended source M33 and point sources. After the stray radiation correction, the flux of point sources increases from 3% to 6% depending on the beam used, whereas the flux of some part of the extended source could be affected by more than  $\pm 10\%$ .

## 6.2 Gridding map

To obtain different moment plots of the target source or region, the observed spectra should be properly assembled into a three-dimensional cube. The different moment plots can then be obtained by integrating the cube. The Nyquist sampling criteria for the number of spectra should be satisfied as the first principle. For FAST, the On-The-Fly (OTF) mapping and drift scanning observation modes have been meticulously designed to meet the Nyquist sampling criteria. This is achieved by rotating the 19-beam receiver at a specific angle, as detailed in sect. 4.1 of ref. [3].

To assemble the individual spectra into cube data, normally one needs to assign the proper spectra into a regular grid in the Right Ascension (RA)-Declination (DEC) plane. This process is similar to creating a map of the sky, in which each grid point contains a single assembled spectrum. The initial step involves generating a World Coordinate System (WCS) header [29]. The WCS is a standard framework in astronomy that associates a physical coordinate (here, RA and DEC) with each pixel in an image. This header corresponds to a regular grid based on the RA and DEC ranges of the input spectra.

For each pixel in the grid, we consider the spectra located within a distance  $r$  from the pixel, where  $r$  is less than a predefined cutoff distance  $r_{\text{cut}}$ . These spectra contribute to the assembled spectrum in the pixel using a proper weighting method. This process ensures that only the relevant data are included for each grid point. The corrupted spectra are discarded to maintain the integrity of the data.

A weight  $w_k$  is assigned to each remaining spectrum using a Gaussian or Bessel\*Gaussian kernel, following the method in ref. [30]. This weighting process takes into account the

distance  $r$  of each spectrum from the grid point and the beam size to ensure the appropriate scaling of the data.

The spectrum  $S$  on the grid is then calculated using the following formula:

$$S = \frac{1}{\sum w_k(r)} \sum_{k=1}^n w_k(r) S_k, \quad (12)$$

where  $S_k$  represents the remaining spectra. The resulting 3D RA-DEC-velocity or RA-DEC-frequency cube data are stored in a standard FITS file. For the integration of the cube at different moments along different dimensions, one could obtain the H I density map, position-velocity map, velocity map, and velocity dispersion map, which are widely used to investigate the physical processes occurring in the target sources.

## 7 Test results

This section presents the results of scientific tests that demonstrate the calibration and imaging performance of HiFAST. First, we provide a brief introduction to the observation data used. Then, we discuss the results for the extended source (M33 galaxy) and point sources separately.

The observed sky covers the regions with  $-5.2^\circ < \text{RA} < 27^\circ$  and  $29.0^\circ < \text{DEC} < 36.5^\circ$ . Within this region, M33, one of the galaxies closest to us, is prominently situated. Its position and the extensive observations previously conducted by the Arecibo telescope make it an exemplary target for our test on extended sources. At the same time, this specific celestial region overlaps significantly with the Arecibo Legacy Fast ALFA (ALFALFA) [23, 31], carried out using the Arecibo telescope, making it equally suitable for the examination of point sources.

Drift scanning was utilised to observe these target regions. The 19-beam receiver array was rotated by  $23.4^\circ$  during scanning to satisfy the Nyquist sampling criterion. To calibrate the temperature, a high noise diode ( $\sim 10$  K) was injected for a duration of 2 s every 5 min. The quasar 3C48, which is close to M33, was observed on the same day as the flux calibrator. Each beam recorded the spectra at an exposure time of 0.5 s. The frequency resolution was approximately 7.63 kHz, corresponding to a velocity resolution of approximately  $1.61 \text{ km s}^{-1}$  at 1420 MHz.

### 7.1 H I map of M33 region

#### 7.1.1 Data reduction

The raw data were processed using the HiFAST modules described in this paper (see also Figures 1 and 2). Specifically,

the system temperature was removed by combining the baseline and standing wave fitting. The spectral channels were shifted to the heliocentric frame. The pixel size was set to  $1'$  during gridding, and spectra within a radius of  $4.14'$  were considered with a weight of a Gaussian kernel with a sigma of  $1.38'$ . Eventually, we obtained a data cube with a velocity range of  $-1500$  to  $500 \text{ km s}^{-1}$ .

### 7.1.2 Comparison with AGES

We compared our data cube with that of the Arecibo Galaxy Environment Survey (AGES) [32]. To maintain consistency, we used the same pixel size (i.e.,  $1'$ ) and grids in the RA-DEC plane as the data cube<sup>5)</sup> from AGES. Since these two observations have different beam sizes, we convolved our data cube to match the beam size of the AGES. During this process, we assumed the beam shape to be a two-dimensional Gaussian function with a full width at half maximum (FWHM) equal to the beam size. Consequently, this assumption and the measurement of the beam size affected the comparisons related to the flux density.

We binned the spectral channels by a factor of 3, resulting in a change in the velocity resolution from  $1.61$  to  $4.83 \text{ km s}^{-1}$ . This adjustment brings our resolution closer to the  $5.2 \text{ km s}^{-1}$  resolution of AGES. Subsequently, we aligned our data cube with AGES by interpolating it onto the same frequency sampling grid. We subtracted a linear baseline along the RA direction for each channel in the cube data to adhere the treatment on H I of Milky Way in AGES. To further minimize interference from the Milky Way, we computed the Moments in the velocity range of  $-350$  to  $-75 \text{ km s}^{-1}$ . Additionally, during the calculation, we used a mask thresholding at  $5\sigma$  of the noise of the spectra during the calculation process to reduce potential data artifacts. Our analysis revealed a systematic velocity offset of approximately  $4.75 \text{ km s}^{-1}$  between the FAST and AGES data, which falls within the velocity resolution of AGES. This offset, which likely arises from differences in velocity resolution, is discussed in detail in Appendix. We have adjusted for this systematic offset in our comparative analysis to ensure accuracy.

In Figure 9, we present a comparison of the integrated flux density (Moment-0) and the intensity-weighted velocities (Moment-1) in the left and right panels, respectively. The top and middle panels illustrate the H I maps acquired from the FAST and Arecibo data, respectively. The third row displays a map showing the fractional difference between these observations, while a scatter plot is used for pixel-by-pixel comparisons.

The Moment-0 maps clearly demonstrate the structural similarities between the inner and outer regions of M33 in both observations. To quantify this agreement, a measure of the fractional difference is used, which is typically less than 10%. This scatter is likely attributable to systematic differences between the two telescopes, encompassing factors such as facility stability, receiver efficiency, and variations in post-processing of the data, among others. One notable source of this scatter is the dissimilar beam shapes of the two telescopes. FAST is equipped with 19 beam receivers, and apart from the central beam (beam 0), the shapes of the other beams are irregular, with slight differences in the beam size (refer to ref. [3] and our Paper IV for more details). We note that the relatively high fractional difference in certain pixels far away from M33 disk is due to the interference from the Milky Way and the velocity cut-off at  $-75 \text{ km s}^{-1}$ . This will introduce a non-negligible error in the determination of the Moment-0 and subsequent Moments calculations.

The Moment-1 maps also exhibit a significant level of similarity in their structure, and most of the pixels in the fractional difference map deviate by less than 5%. However, it is worth mentioning that the median fractional difference is close to zero and does not show a strong correlation with the velocity values.

Furthermore, we compared the weighted velocity dispersion (Moment-2, left panels) and position-velocity maps (right panels) of both datasets, as depicted in Figure 10. For Moment-2, the maps displayed a satisfactory level of agreement on structure. Pixel-by-pixel analysis shows that the median fractional difference is approximately 4%. Furthermore, the absolute difference is minor, approximately  $(0.6 \text{ km s}^{-1})^2$ . Analogous to the Moment-0 maps, the typical discrepancies in the flux density for each positive-velocity pixel were also below 10%.

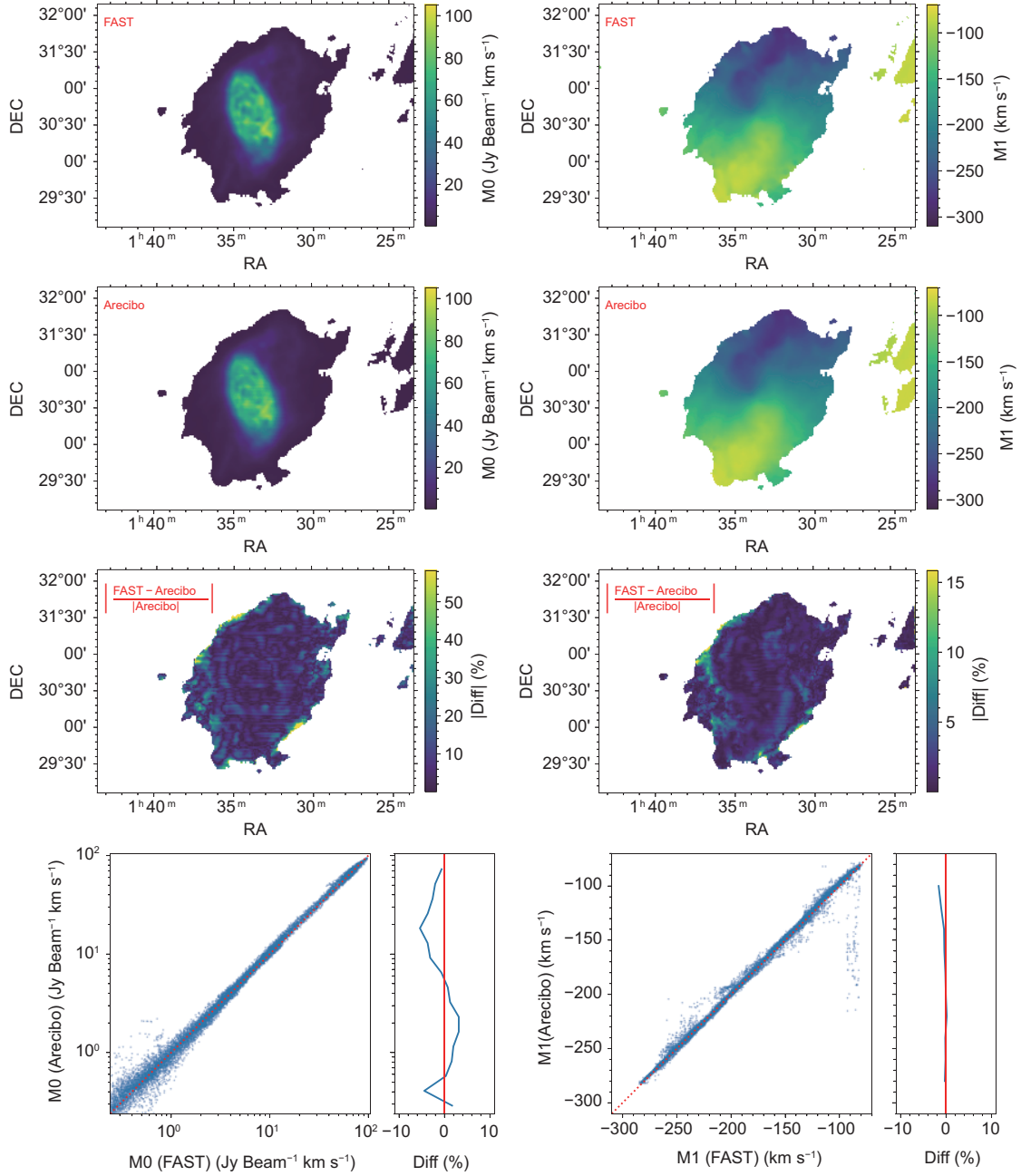
In summary, considering the systematic differences between the two datasets, the agreement on Moments 0, 1, 2, and the position-velocity map is quite noteworthy.

## 7.2 H I point sources

### 7.2.1 Data reduction

In this section, we analyse the FAST detected point sources within the frequency spectrum of 1300 to 1415 MHz. This particular frequency range has been selected due to its near absence of RFI and its overlap with ALFALFA [23, 31]. As the angular size of most galaxies in this frequency range is smaller than the FAST's beam size, we use the "MinMed" method to fit the off-source observation in system tempera-

5) [http://www.naic.edu/~ages/public\\_data.html](http://www.naic.edu/~ages/public_data.html).



**Figure 9** (Color online) Comparisons of Moment-0 (M0, left panels) and Moment-1 (M1, right panels) of M33 H I from FAST (this work) and Arecibo (AGES). The sequence includes FAST data maps (top row), Arecibo data maps (second row), and the absolute value of fractional difference between the two datasets (third row), calculated using the formula  $(\text{FAST} - \text{Arecibo})/|\text{Arecibo}| \times 100\%$ . The fourth row contains scatter plots on the left, comparing pixel values from FAST and Arecibo, and plots on the right, displaying the median value of fractional difference as a function of the Arecibo values; the red lines in these plots denote the one-to-one correspondence between the two datasets.

ture removal (see sect. 3.2.1). The remaining processes were the same as those used to process the M33 H I data cube. The sources were identified using SOFIA 2 software [33, 34].

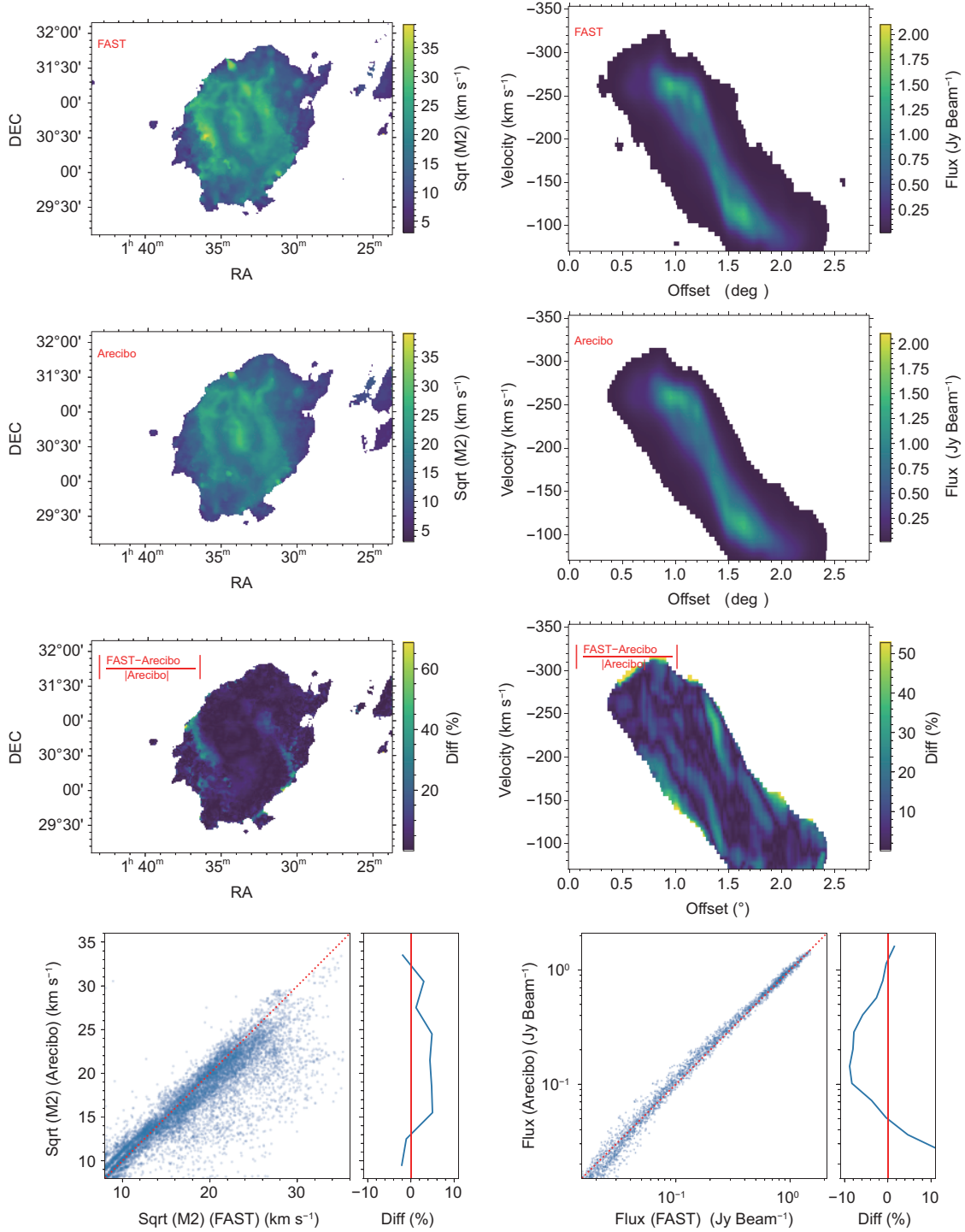
### 7.2.2 Comparison with ALFALFA

In this subsection, we compare the properties of the detected sources with the ALFALFA catalogue. The flux density  $S_\nu$

from each source is measured by spatially integrating the spectra  $s_\nu$  in the pixels as following [35], given by

$$S_\nu = \frac{\sum_{x_0} \sum_{y_0} s_\nu(\Delta x, \Delta y)}{\sum_{x_0} \sum_{y_0} B(\Delta x, \Delta y)} [\text{Jy}], \quad (13)$$

where the numerator is the sum of  $s_\nu$  over a selected solid angle covered by the region within the chosen isophotal fit [23] and the denominator is expressed as the sum of a centred,



**Figure 10** (Color online) Similar to Figure 9, but for Moment-2 (left panels) and the position-velocity map (right panels). The position-velocity maps are extracted along the line of RA = 01h33m51s.

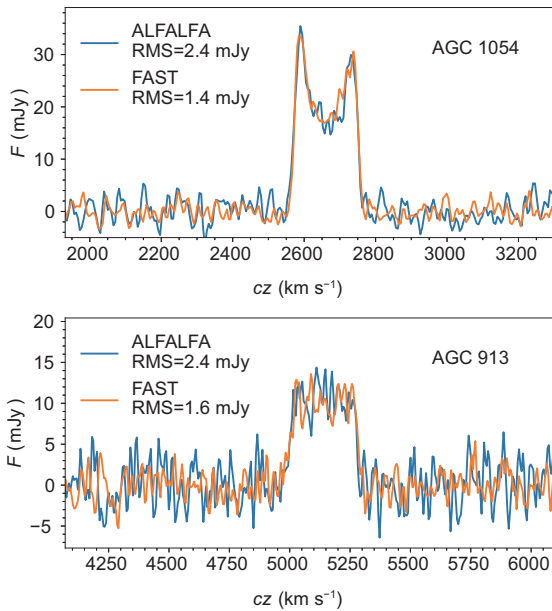
grid-sampled beam in the same region. The total integrated H I line flux density  $S_{\text{int}}$  in  $\text{Jy km s}^{-1}$  is then summed over all velocity ranges covered by the signal. Meanwhile, the signal-to-noise ratio  $S/N$  and the error of  $S_{\text{int}}$  are calculated using the same method used in ref. [23].

We cross-matched our sources with the ALFALFA based on their positions in the sky and the centre velocity. As ref. [23] pointed out, the pipeline algorithm used for the ALFALFA catalogue may miss flux from very extended or highly asymmetric sources [23]. For a fair test, we limit

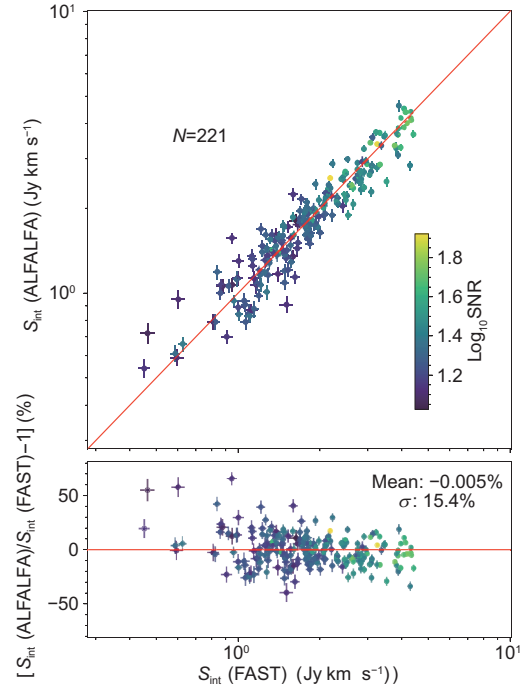
the comparison to point sources with a characteristic size of the H I disk smaller than  $2'$  (similar to ref. [36]), which is less than the beam size of two telescopes. The characteristic size is estimated using the H I size-mass relation in ref. [37]. We also limited the sample to  $S/N > 10$ . This resulted in 221 sources matched from both datasets.

First, two examples of spectra comparison are presented in Figure 11. The spectra's RMS noise are indicated on the labels. We notice that the spectra from FAST had lower RMS values with mean values of approximately 1.47 and 2.24 mJy for FAST and ALFALFA in 221 sources, respectively. Both observations were performed using the drifting scan mode. While the sampling spacing in DEC is around  $1.05'$  [38] in ALFALFA's two-pass strategy, which is smaller than the spacing of approximately  $1.14'$  in our FAST one-pass observation. This indicates a slightly longer integration observation time in ALFALFA. Given the shorter integration time and lower RMS noise, FAST demonstrated excellent performance in observing H I.

In Figure 12, we compare the integrated H I line flux density  $S_{\text{int}}$  for 221 sources. The top panel displays the scatter plot of  $S_{\text{int}}(\text{FAST})$  versus  $S_{\text{int}}(\text{ALFALFA})$ , while the bottom panel shows the fractional difference  $S_{\text{int}}(\text{ALFALFA})/S_{\text{int}}(\text{FAST}) - 1$  as a function of  $S_{\text{int}}(\text{FAST})$ . The data points are in close agreement with the equal value line (red line), with a mean fractional difference of approximately  $-0.005\%$  and a dispersion of  $15.4\%$ . The scatter of the differences showed a slight dependence on



**Figure 11** (Color online) The spectra of two example galaxies detected in ALFALFA (blue lines) and FAST (orange lines). The  $x$ -axis represents the optical velocity, corrected for the Heliocentric rest frame. Labels also include RMS noise of the spectra. The ALFALFA AGC number for each source is indicated in the top-right corner of the plot.

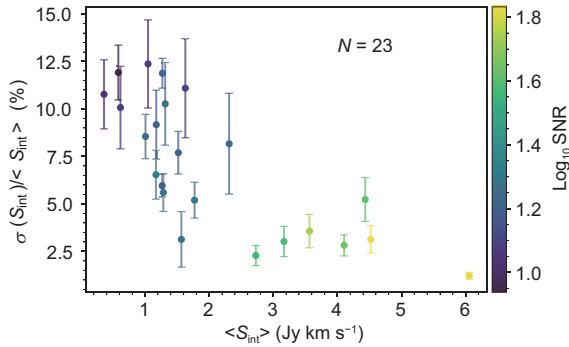


**Figure 12** (Color online) Comparison of integrated H I line flux between FAST (this work) and the ALFALFA catalogue. The colour codes the signal-to-noise ratio of the source in FAST. Top panel: scatter plot of  $S_{\text{int}}(\text{FAST})$  versus  $S_{\text{int}}(\text{ALFALFA})$ ; bottom panel: fractional difference as a function of  $S_{\text{int}}(\text{FAST})$ . The red line in the top panel is a one-to-one relation and in the bottom panel the fractional difference equals 0.

flux:  $16.1\%$  for  $S_{\text{int}}(\text{FAST}) < 2.5 \text{ Jy km s}^{-1}$  and  $11.7\%$  for  $S_{\text{int}}(\text{FAST}) > 2.5 \text{ Jy km s}^{-1}$ .

Given that the data from ALFALFA and FAST were obtained from different facilities and processed through distinct pipelines, the scatter observed in the comparison between the two datasets could potentially be induced from various sources. These include differences in baseline complexity, uncertainties in beam shape, the multibeam synthesis (7 beams in ALFALFA and 19 beams in FAST), and uncertainties in source parameter estimation from the data cube. To estimate the inherent systematic error in FAST with HiFAST, we conducted a meticulous examination of multiple observations of the same sources. A 10 square-degree region of the sky was scanned seven times over a few months, leading to the identification of 23 sources.

In Figure 13, we illustrate the 1-sigma dispersion of integrated H I line flux density  $S_{\text{int}}$  measured in seven observations as a function of the mean value. The error for the dispersion is estimated using the bootstrap resampling method. For sources with  $S_{\text{int}} > 2.5 \text{ Jy km s}^{-1}$ , the dispersion is less than  $5\%$ . However, for  $S_{\text{int}} < 2.5 \text{ Jy km s}^{-1}$ , the dispersion increases to approximately  $10\%$  as the uncertainty in the calibration procedure (particularly in system temperature removal operations) becomes more pronounced for fainter sources.



**Figure 13** (Color online) Multiple observation results for 23 sources. The x-axis and y-axis denote the mean and 1-sigma dispersion of integrated H I line flux density measured in 7 times observations, respectively. The error bars are estimated by the bootstrap resampling method. The colour codes the mean signal-to-noise ratio.

## 8 Code implementation and parallelization

HiFAST is written in Python programming language [39], using the Python standard library and several third-party libraries: NumPy [40] and SciPy [41] for data structure and calculating; Astropy [42] for the calculation of RA-DEC and Doppler correction; Matplotlib [43] for diagnostic plotting; and h5py [44] for intermediate data storage (a comprehensive list is available on our website). HiFAST is an ongoing project, and its dependencies may change in the future.

Parallel calibration is a critical feature in HiFAST because of the significant amount of data recorded by the 19-beam receiver of FAST. The data from the 19 beams were stored in separate files, and the majority of the procedures described previously were performed on each beam. This allowed HiFAST to facilitate a streamlined simultaneous calibration. For certain time-consuming calibration processes, e.g., baseline fitting, stray radiation correction, and gridding, we employed the multiprocessing package to parallelize the calibration on each spectrum or the computation of the spectrum at each grid point.

## 9 Discussion and summary

We have presented HiFAST, a pipeline for processing H I data obtained from FAST observations. This pipeline is compatible with tracking, drift scanning, and OTF mapping modes, as well as most of their variants. HiFAST provides modular tasks to calibrate the raw data and create a data cube from the spectra. These tasks include temperature calibration with a noise diode as reference power input; system temperature removal using the “MinMed” method, baseline and standing wave fitting; flux density calibration by using a calibrator observed on the same day with the target or applying a predetermined gain function with zenith angle; an automated

process to mask common RFI; Doppler correction of spectra from topocentric to Heliocentric or LSR reference frame; an iterative process to correct stray radiation; and interpolation of spectra to a regular grid of a data cube.

To assess the performance of HiFAST, we used H I data from the extended source M33 and point sources within the 1300-1415 MHz frequency range. After processing this raw data with the pipeline, we produced an H I image of M33 and a catalogue of point sources. Our comparison of the M33 data with the AGES results showed similar structures in the moment maps (0, 1, and 2), with a median fractional difference of less than 10%. Additionally, we cross-matched our point sources with the ALFALFA survey catalog. In a common sample of 221 sources with a S/N ratio greater than 10 from both datasets, the mean fractional difference in integrated flux density  $S_{\text{int}}$  was approximately 0.005%, with a dispersion of 15.4%. This dispersion varied with the value of  $S_{\text{int}}$ : 16.1% at  $S_{\text{int}}(\text{FAST}) < 2.5 \text{ Jy km s}^{-1}$  and 11.7% at  $S_{\text{int}}(\text{FAST}) > 2.5 \text{ Jy km s}^{-1}$ . Using multiple observations with FAST, we estimated the inherent systematic uncertainty of HiFAST to be less than 5% for sources with  $S_{\text{int}} > 2.5 \text{ Jy km s}^{-1}$  and about 10% for those with  $S_{\text{int}} < 2.5 \text{ Jy km s}^{-1}$ .

HiFAST has been successfully employed to process the FAST data in a number of studies, such as the FAST All Sky H I survey (FASHI) [45], M106 [46], M51 [47], DDO168 [48], M94 [49], and NGC4490/85 [50]. While HiFAST focuses mainly on processing the H I data, its capabilities can be readily extended to handle OH line (Zhang et al. submitted) and ultra-wide band (UWB) data from FAST [51] in the future.

*This work was supported by the National Key R&D Program of China (Grant No. 2022YFA1602901), the National Natural Science Foundation of China (Grant Nos. 11988101, 11873051, 12125302, and 12373011), the CAS Project for Young Scientists in Basic Research Grant (Grant No. YSBR-062), the K. C. Wong Education Foundation, and the science research grants from the China Manned Space Project. Yingjie Jing acknowledges support from the Cultivation Project for FAST Scientific Payoff and Research Achievement of CAMS-CAS. We would like to express our gratitude to Prof. Rendong Nan and his team for their groundbreaking efforts in establishing the FAST telescope. Additionally, we would like to acknowledge the FAST operation team for their invaluable assistance and support in the development of this pipeline. We also thank Martha P. Haynes for generously sharing ALFALFA spectra data and insightful discussions on source flux calculations, which significantly enhanced our study. This work made use of the data from FAST (Five-hundred-meter Aperture Spherical radio Telescope). FAST is a Chinese national mega-science facility, operated by National Astronomical Observatories, Chinese Academy of Sciences.*

**Conflict of interest** The authors declare that they have no conflict of interest.



- 2 P. Jiang, Y. L. Yue, H. Q. Gan, R. Yao, H. Li, G. F. Pan, J. H. Sun, D. J. Yu, H. F. Liu, N. Y. Tang, L. Qian, J. G. Lu, J. Yan, B. Peng, S. X. Zhang, Q. M. Wang, Q. Li, and D. Li, *Sci. China-Phys. Mech. Astron.* **62**, 959502 (2019), arXiv: 1903.06324.
- 3 P. Jiang, N. Y. Tang, L. G. Hou, M. T. Liu, M. Krčo, L. Qian, J. H. Sun, T. C. Ching, B. Liu, Y. Duan, Y. L. Yue, H. Q. Gan, R. Yao, H. Li, G. F. Pan, D. J. Yu, H. F. Liu, D. Li, B. Peng, and J. Yan, *Res. Astron. Astrophys.* **20**, 064 (2020), arXiv: 2002.01786.
- 4 L. Qian, R. Yao, J. Sun, J. Xu, Z. Pan, and P. Jiang, *Innovation* **1**, 100053 (2020), arXiv: 2011.13542.
- 5 J. Yin, P. Jiang, and R. Yao, *Sci. China-Phys. Mech. Astron.* **66**, 239513 (2023).
- 6 K. O'Neil, in *Single-dish calibration techniques at radio wavelengths: Single-Dish Radio Astronomy: Techniques and Application, ASP Conference Series, Vol. 278*, edited by S. Stanimirović, D. R. Altschuler, P. F. Goldsmith, and C. J. Salter (ASP, 2022), pp. 293-311.
- 7 B. Winkel, A. Kraus, and U. Bach, *Astron. Astrophys.* **540**, A140 (2012), arXiv: 1203.0741.
- 8 A. Dunning, M. Bowen, S. Castillo, Y. S. Chung, P. Doherty, D. George, D. B. Hayman, K. Jeganathan, H. Kanoniuk, S. Mackay, L. Reilly, P. Roush, S. K. W. Smart, R. D. Shaw, S. L. Smith, T. Tzioumis, and V.-C. J. Venables, in *Design and laboratory testing of the five hundred meter aperture spherical telescope (FAST) 19 beam L-band receiver: 2017 XXXIInd General Assembly and Scientific Symposium of the International Union of Radio Science (URSI GASS)* (Montreal, 2017), pp. 1-4.
- 9 K. Jeganathan, A. Dunning, J. Chengjin, Z. Haiyan, Y. Chung, S. Mackay, P. Doherty, H. Kanoniuk, L. Reilly, P. Roush, R. D. Shaw, S. Severs, M. A. Bowen, and D. B. Hayman, in *The FAST multi-beam receiver design with RF over fiber link: 2017 IEEE Asia Pacific Microwave Conference (APMC)* (Kuala Lumpur, 2017), pp. 244-247.
- 10 S. A. Wolfe, D. J. Pisano, and F. J. Lockman, *Mapping Diffuse Neutral Hydrogen with the GBT*, GBT Memo 289 (2015).
- 11 M. E. Putman, V. de Heij, L. Staveley-Smith, R. Braun, K. C. Freeman, B. K. Gibson, W. B. Burton, D. G. Barnes, G. D. Banks, R. Bhathal, W. J. G. de Blok, P. J. Boyce, M. J. Disney, M. J. Drinkwater, R. D. Ekers, P. A. Henning, H. Jerjen, V. A. Kilborn, P. M. Knezek, B. Koribalski, D. F. Malin, M. Marquarding, R. F. Minchin, J. R. Mould, T. Oosterloo, R. M. Price, S. D. Ryder, E. M. Sadler, I. Stewart, F. Stootman, R. L. Webster, and A. E. Wright, *Astron. J.* **123**, 873 (2002), arXiv: astro-ph/0110416.
- 12 M. E. Putman, L. Staveley-Smith, K. C. Freeman, B. K. Gibson, and D. G. Barnes, *Astrophys. J.* **586**, 170 (2003), arXiv: astro-ph/0209127.
- 13 J. I. Davies, R. Auld, L. Burns, R. Minchin, E. Momjian, S. Schneider, M. Smith, R. Taylor, and W. van Driel, *Mon. Not. R. Astron. Soc.* **415**, 1883 (2011).
- 14 R. Taylor, R. F. Minchin, H. Herbst, J. I. Davies, R. Rodriguez, and C. Vazquez, *Mon. Not. R. Astron. Soc.* **443**, 2634 (2014), arXiv: 1407.0016.
- 15 S. J. Baek, A. Park, Y. J. Ahn, and J. Choo, *Analyst* **140**, 250 (2015).
- 16 P. H. C. Eilers, *Anal. Chem.* **75**, 3631 (2003).
- 17 Z. M. Zhang, S. Chen, and Y. Z. Liang, *Analyst* **135**, 1138 (2010).
- 18 R. Padman, *Publ. Astron. Soc. Aust.* **3**, 111 (1977).
- 19 F. H. Briggs, E. Sorar, R. C. Kraan-Korteweg, and W. van Driel, *Publ. Astron. Soc. Aust.* **14**, 37 (1997), arXiv: astro-ph/9701182.
- 20 J. E. G. Peek, C. Heiles, K. A. Douglas, M. Y. Lee, J. Grcevich, S. Stanimirović, M. E. Putman, E. J. Korpela, S. J. Gibson, A. Begum, D. Saul, T. Robishaw, and M. Krčo, *Astrophys. J. Suppl. Ser.* **194**, 20 (2011), arXiv: 1101.1879.
- 21 C. P. Zhang, J. L. Xu, J. Wang, Y. Jing, Z. Liu, M. Zhu, and P. Jiang, *Res. Astron. Astrophys.* **22**, 025015 (2022), arXiv: 2111.11018.
- 22 M. J. Meyer, M. A. Zwaan, R. L. Webster, L. Staveley-Smith, E. Ryan-Weber, M. J. Drinkwater, D. G. Barnes, M. Howlett, V. A. Kilborn, J. Stevens, M. Waugh, M. J. Pierce, R. Bhathal, W. J. G. de Blok, M. J. Disney, R. D. Ekers, K. C. Freeman, D. A. Garcia, B. K. Gibson, J. Harnett, P. A. Henning, H. Jerjen, M. J. Kesteven, P. M. Knezek, B. S. Koribalski, S. Mader, M. Marquarding, R. F. Minchin, J. O'Brien, T. Oosterloo, R. M. Price, M. E. Putman, S. D. Ryder, E. M. Sadler, I. M. Stewart, F. Stootman, and A. E. Wright, *Mon. Not. R. Astron. Soc.* **350**, 1195 (2004), arXiv: astro-ph/0406384.
- 23 M. P. Haynes, R. Giovanelli, B. R. Kent, E. A. K. Adams, T. J. Balonek, D. W. Craig, D. Fertig, R. Finn, C. Giovanardi, G. Hallenbeck, K. M. Hess, G. L. Hoffman, S. Huang, M. G. Jones, R. A. Koopmann, D. A. Kornreich, L. Leisman, J. Miller, C. Moorman, J. O'Connor, A. O'Donoghue, E. Papastergis, P. Troischt, D. Stark, and L. Xiao, *Astrophys. J.* **861**, 49 (2018), arXiv: 1805.11499.
- 24 A. Comrie, K.-S. Wang, S.-C. Hsu, A. Moraghan, P. Harris, Q. Pang, A. Pińska, C.-C. Chiang, T.-H. Chang, Y.-H. Hwang, H. Jan, M.-Y. Lin, and R. Simmonds, *CARTA: The Cube Analysis and Rendering Tool for Astronomy (2.0.0)* (Zenodo, 2021).
- 25 P. M. W. Kalberla, U. Mebold, and W. Reich, *Astron. Astrophys.* **82**, 275 (1980).
- 26 P. M. W. Kalberla, W. B. Burton, D. Hartmann, E. M. Arnal, E. Bajaja, R. Morras, and W. G. L. Pöppel, *Astron. Astrophys.* **440**, 775 (2005).
- 27 A. I. Boothroyd, K. Blagrove, F. J. Lockman, P. G. Martin, D. Pinheiro Gonçalves, and S. Srikanth, *Astron. Astrophys.* **536**, A81 (2011), arXiv: 1110.1765.
- 28 B. Winkel, J. Kerp, L. Flöer, P. M. W. Kalberla, N. Ben Bekhti, R. Keller, and D. Lenz, *Astron. Astrophys.* **585**, A41 (2016), arXiv: 1512.05348.
- 29 E. W. Greisen, and M. R. Calabretta, *Astron. Astrophys.* **395**, 1061 (2002).
- 30 J. G. Mangum, D. T. Emerson, and E. W. Greisen, *Astron. Astrophys.* **474**, 679 (2007).
- 31 M. P. Haynes, R. Giovanelli, A. M. Martin, K. M. Hess, A. Saintonge, E. A. K. Adams, G. Hallenbeck, G. L. Hoffman, S. Huang, B. R. Kent, R. A. Koopmann, E. Papastergis, S. Stierwalt, T. J. Balonek, D. W. Craig, S. J. U. Higdon, D. A. Kornreich, J. R. Miller, A. A. O'Donoghue, R. P. Olowin, J. L. Rosenberg, K. Spekkens, P. Troischt, and E. M. Wilcots, *Astron. J.* **142**, 170 (2011), arXiv: 1109.0027.
- 32 O. C. Keenan, J. I. Davies, R. Taylor, and R. F. Minchin, *Mon. Not. R. Astron. Soc.* **456**, 951 (2016), arXiv: 1511.02710.
- 33 P. Serra, T. Westmeier, N. Giese, R. Jurek, L. Flöer, A. Popping, B. Winkel, T. van der Hulst, M. Meyer, B. S. Koribalski, L. Staveley-Smith, and H. Courtois, *Mon. Not. R. Astron. Soc.* **448**, 1922 (2015), arXiv: 1501.03906.
- 34 T. Westmeier, S. Kitaeff, D. Pallot, P. Serra, J. M. van der Hulst, R. J. Jurek, A. Elagali, B. Q. For, D. Kleiner, B. S. Koribalski, K. Lee-Waddell, J. R. Mould, T. N. Reynolds, J. Rhee, and L. Staveley-Smith, *Mon. Not. R. Astron. Soc.* **506**, 3962 (2021), arXiv: 2106.15789.
- 35 G. S. Shostak, and R. J. Allen, *Astron. Astrophys.* **81**, 167 (1980).
- 36 P. Zuo, D. Yang, J. Wang, L. Staveley-Smith, X. Lin, B. Q. For, T. Westmeier, J. Wang, K. Spekkens, V. Kilborn, O. Ivy Wong, D. Li, K. Lee-Waddell, B. Catinella, L. C. Ho, B. Koribalski, B. Lee, and M. Zhu, *Res. Astron. Astrophys.* **22**, 095016 (2022).
- 37 J. Wang, B. S. Koribalski, P. Serra, T. van der Hulst, S. Roychowdhury, P. Kamphuis, and J. N. Chengalur, *Mon. Not. R. Astron. Soc.* **460**, 2143 (2016), arXiv: 1605.01489.
- 38 R. Giovanelli, M. P. Haynes, B. R. Kent, P. Perillat, B. Catinella, G. L. Hoffman, E. Momjian, J. L. Rosenberg, A. Saintonge, K. Spekkens, S. Stierwalt, N. Brosch, K. L. Masters, C. M. Springob, I. D. Karachentsev, V. E. Karachentseva, R. A. Koopmann, E. Muller, W. van Driel, and L. van Zee, *Astron. J.* **130**, 2613 (2005), arXiv: astro-ph/0508300.
- 39 G. Van Rossum, and F. L. Drake, *Python 3 Reference Manual* (CreateSpace, Scotts Valley, CA, 2009).
- 40 C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, *Nature* **585**, 357 (2020), arXiv: 2006.10256.
- 41 P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J.

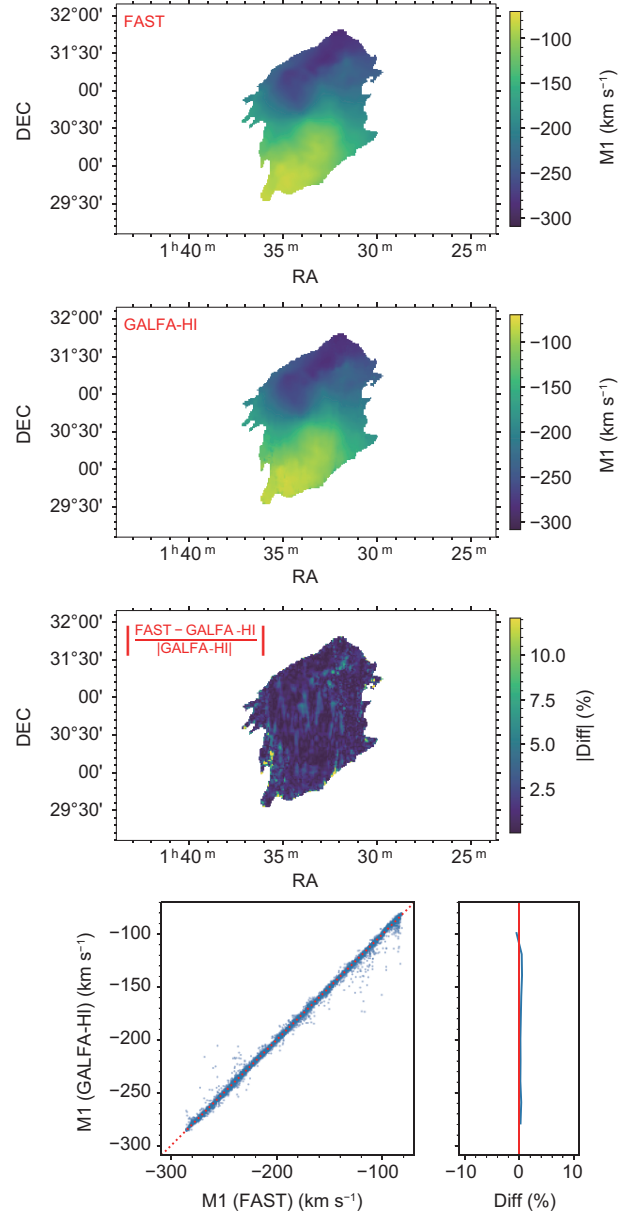
van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Í. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, *Nat. Methods* **17**, 261 (2020), arXiv: 1907.10121.

- 42 The Astropy Collaboration, A. M. Price-Whelan, P. L. Lim, N. Earl, N. Starkman, L. Bradley, D. L. Shupe, A. A. Patil, L. Corrales, C. E. Brasseur, M. Nöthe, A. Donath, E. Tollerud, B. M. Morris, A. Ginsburg, E. Vaher, B. A. Weaver, J. Tocknell, W. Jamieson, M. H. van Kerkwijk, T. P. Robitaille, B. Merry, M. Bachetti, H. M. Günther, and Astropy Project Contributors, *Astrophys. J.* **935**, 167 (2022), arXiv: 2206.14220.
- 43 J. D. Hunter, *Comput. Sci. Eng.* **9**, 90 (2007).
- 44 A. Collette, *Python and HDF5* (O'Reilly, 2013).
- 45 C. P. Zhang, M. Zhu, P. Jiang, C. Cheng, J. Wang, J. Wang, J. L. Xu, X. L. Liu, N. P. Yu, L. Qian, H. Yu, M. Ai, Y. Jing, C. Xu, Z. Liu, X. Guan, C. Sun, Q. Yang, M. Huang, and Q. Hao, *Sci. China-Phys. Mech. Astron.* **67**, 219511 (2024), arXiv: 2312.06097.
- 46 M. Zhu, H. Yu, J. Wang, J. L. Xu, W. Du, L. Yuan, J. Wang, Y. Jing, M. Ai, and P. Jiang, *ApJL* **922**, L21 (2021).
- 47 H. Yu, M. Zhu, J. L. Xu, M. Ai, P. Jiang, and Y. Yang, *Mon. Not. R. Astron. Soc.* **521**, 2719 (2023), arXiv: 2302.03270.
- 48 N. P. Yu, M. Zhu, J. L. Xu, C. P. Zhang, X. L. Liu, P. Jiang, and J. J. Wang, *Mon. Not. R. Astron. Soc.* **521**, 737 (2023).
- 49 R. Zhou, M. Zhu, Y. Yang, H. Yu, L. Yuan, P. Jiang, and W. Xi, *Astrophys. J.* **952**, 130 (2023), arXiv: 2306.05080.
- 50 Y. Liu, M. Zhu, H. Yu, M. Ai, P. Jiang, S. Liu, and L. Yuan, *Mon. Not. R. Astron. Soc.* **523**, 3905 (2023), arXiv: 2304.13964.
- 51 C. P. Zhang, P. Jiang, M. Zhu, J. Pan, C. Cheng, H. F. Liu, Y. Zhu, C. Sun, and F. Collaboration, *Res. Astron. Astrophys.* **23**, 075016 (2023), arXiv: 2304.11895.
- 52 J. E. G. Peek, B. L. Babler, Y. Zheng, S. E. Clark, K. A. Douglas, E. J. Korpela, M. E. Putman, S. Stanimirović, S. J. Gibson, and C. Heiles, *Astrophys. J. Suppl. Ser.* **234**, 2 (2018).

## Appendix Velocity comparison with GALFA-H I

In sect. 7.1.2, we found that there is a systematic offset of about  $-4.75 \text{ km s}^{-1}$  between the FAST and AGES data. After accounting for this offset, the Moment-1 comparison displayed consistency, as shown in Figure 9. In this section, we further expand our comparison by comparing the data from FAST to the data from the Galactic Arecibo L-band Feed Array H I Survey (GALFA-H I) [52]. GALFA-H I also utilizes the Arecibo telescope but has a higher velocity resolution of approximately  $0.7 \text{ km s}^{-1}$ . Figure a1 illustrates the Moment-1 comparison between FAST and GALFA-H I. The median difference is approximately  $0.52 \text{ km s}^{-1}$ , which is smaller

than the difference between FAST and AGES. This suggests that the velocity resolution may contribute to this systematic velocity offset.



**Figure a1** (Color online) Similar to the right panels of Figure 9, but for comparison of data from FAST (this work) and Arecibo (GALFA-H I) without velocity adjustment.