

Mapping diffuse neutral hydrogen with the GBT

Spencer A. Wolfe^{1*}, D.J. Pisano^{1, 2}, Felix J. Lockman³

April 21, 2015

Abstract

Observations of the gas that resides between galaxies is essential to understanding how large spirals like our Milky Way acquire material for the formation of new stars. The Robert C. Byrd Green Bank Radio Telescope (GBT) has the capability to map the neutral component of this gas directly in emission. This work details the process we have developed to observe extremely faint 21 cm radiation, in which we reach spectral noise limits ≈ 3.6 mK with $\Delta\nu \approx 5$ km s⁻¹, where $\Delta\nu$ is the spectral channel width. We provide an explicit discussion of our observing techniques and reduction process, such as the stability of calibration temperatures T_{CAL} over several months, as well as a blanking issue inherent to the GBT Spectrometer.

¹Department of Physics and Astronomy, West Virginia University, P.O. Box 6315, Morgantown, WV 26506

²Adjunct Astronomer at the National Radio Astronomy Observatory

³National Radio Astronomy Observatory, P.O. Box 2, Rt. 28/92, Green Bank, WV 24944

*Corresponding author: swolfe4@mix.wvu.edu

Contents

1 Mapping HI in the “Cosmic Web”	1
1.1 Background	1
1.2 GBT Mapping Sensitivity	1
2 Observing Method	2
2.1 Concept	2
2.2 Front End and Back End Configuration	2
2.3 Observing the Calibrator 3C48	2
2.4 Mapping the HI	3
3 Data Reduction	3
3.1 Concept	3
3.2 T_{CAL} from 3C48	3
3.3 HI Mapping Reduction	4
4 Blanking Issues with the Spectrometer	5
5 Gridding the Spectral Line Data	5
Acknowledgments	6
References	6

1. Mapping HI in the “Cosmic Web”

1.1 Background

To further our understanding of how galaxies acquire fresh gas, we must investigate the diffuse material that resides in filamentary structures between galaxies. The main issue with observing such gas is that it is mostly ionized [1], with perhaps 1% of the gas in the neutral state [2], appearing at HI column densities of $N_{\text{HI}} \sim 10^{17}$ cm⁻². Thus, it takes significant sensitivity for a radio telescope to observe such emission. One must generally resort to UV absorption line measurements instead, but these require a bright background source for such observations to be possible.

In 2004, extremely faint HI emission in the Local Group between our neighboring spiral galaxies M31 (Andromeda) and M33 (Triangulum) was detected [3]. In that work, a column density sensitivity of $N_{\text{HI}} \sim 10^{17}$ cm⁻² was reached by using every dish in the Westerbork Synthesis Radio Telescope (WSRT) as a single receiver. After thousands of hours and extensive smoothing of the data to 49' resolution and 18 km s⁻¹ velocity resolution, a very faint, extended HI feature between M31 and M33 was discovered (Figure 1). Follow-up observations [4] of the region with the GBT have confirmed the existence of this gas, but those observations didn't reach the same sensitivity as the WSRT data.

1.2 GBT Mapping Sensitivity

Reaching 21 cm sensitivities of $\sim 10^{17}$ cm⁻² with no spatial smoothing using the GBT is feasible. The radiometer equation tells us:

$$\sigma_{\text{T}_B} = \frac{K1 \cdot T_{\text{sys}}}{\sqrt{K2 \cdot N_{\text{pol}} \cdot \Delta\nu \cdot t_{\text{eff}}}} \quad (1)$$

where σ_{T_B} is the predicted root-mean-square (RMS) of the brightness temperature of a typical spectrum over one GBT beam. T_{sys} is the system temperature, around 18-20 K. N_{pol} is the number of polarizations, usually 2. $\Delta\nu$ is the bandwidth per unit spectral frequency channel and t_{eff} is the effective integration time of the observation. $K1$ and $K2$ are correction factors specific to the GBT Spectrometer; specifically, $K1 = 1.032$ and $K2 = 1.21$. So, the radiometer equation becomes, using $T_{\text{sys}} = 18$ K and $\Delta\nu = 1.52$ kHz :

$$\sigma_{\text{T}_B} = \frac{306 \text{ mK}}{\sqrt{t_{\text{eff}}}} \quad (2)$$

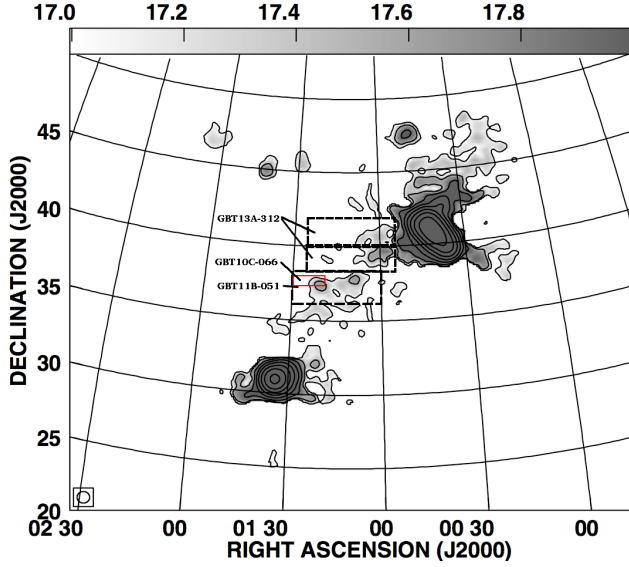


Figure 1. An HI column density map of the original WSRT data [3, Figure 9]. The dashed boxes are areas we have mapped with the GBT (GBT11B-051, GBT13A-312). The small red outline is the area we mapped as an exploratory project (GBT10C-066).

HI column density is determined from the brightness temperature integrated over the spectral line in frequency (or velocity). Assuming that the HI is optically thin:

$$N_{\text{HI}} = 1.822 \times 10^{18} \int_{\text{line}} \left(\frac{T_{\text{B}}(v)}{\text{K}} \right) \left(\frac{dv}{\text{km s}^{-1}} \right) \text{ cm}^{-2} \quad (3)$$

If the spectral noise is Gaussian in nature, we can write the error in column density as:

$$\sigma_{N_{\text{HI}}} = 1.822 \times 10^{18} \cdot \sigma_{T_{\text{B}}} \cdot \sqrt{W \cdot \Delta v} \text{ cm}^{-2} \quad (4)$$

where W is the linewidth of the HI spectrum and Δv is the width of a spectral channel. Both are in units of km s^{-1} . A 1.52 kHz channel width converts to 0.32 km s^{-1} . Assuming a typical HI linewidth of 25 km s^{-1} , a 3σ HI column density detection limit would require, inverting Equation 4, $\sigma_{T_{\text{B}}} = 6.5 \text{ mK}$. Inserting this result into Equation 2 and solving gives an effective integration time of 37.2 min per GBT beam.

Given that the main GBT beam is about $9.1'$ across, there are roughly 44 beams (43.5) per square degree. Thus, to map 21 cm radiation down to 10^{17} cm^{-2} requires an effective time of about 27 hours per square degree. While this may seem like a significant amount of time, with careful planning of the observations, one can still map rather sizeable regions. Furthermore, the unblocked aperture design of the the telescope (Figure 2) significantly reduces the amount of contamination by stray radiation, with the side-lobes of the beam being roughly 30dB below that of the main

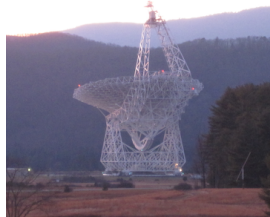


Figure 2. The GBT
Image taken by the author.

beam. This, plus the stability of the receivers and low system temperature also ensure that the telescope can **consistently** map faint HI emission.

2. Observing Method

2.1 Concept

A major issue with observing extremely faint emission is ensuring that there is little to no baseline structure in the spectral data. Frequency switching can produce baseline artifacts that could be potentially confused with real emission features. Position switching can eliminate many of these structures by providing more stable baselines, but the telescope must generally slew to a different part of the sky and spend time integrating to obtain off-source spectra for reference. For a project requiring hundreds of hours of observations, the accumulated time required to obtain reference spectra can greatly increase the amount of observing time needed.

For our projects (GBT10C-066, GBT11B-051, GBT13A-312), we decided to choose a region of our proposed mapping area to use as our reference positions. Using the WSRT map, we determined that the eastern edge of our observing areas would be suitable to use for reference positions (Figure 1). This gave us the advantages of having a reference for each scan, allowing us to choose the total number of integrations that can be used as reference and allowed us to devote the majority of our observing time to actually mapping the region without needing to slew the telescope off source. See Figure 3 for a simple illustration. One must take careful note, however, that the area of our scans chosen for the offs must contain no emission (at least down to the desired column density sensitivity).

2.2 Front End and Back End Configuration

The data were collected in total power mode, which simply collects photons as the telescope is scanned, since the actual reference selection and subtraction would be done during the reduction. Our observing setup used the L-band receiver (1.10 - 1.75 GHz) and GBT Spectrometer with a 12.5 MHz bandwidth using 8192 channels centered near the 1420 MHz line, giving $\Delta v = 1.52 \text{ kHz}$. We also tried using a 50 MHz bandwidth and four different frequency windows (IFs), with the other three centered on the OH lines (1612, 1665, 1667, 1720 MHz). Strong radio frequency interference (RFI) in the 1612 MHz window, however, caused large swings in the power levels of the other IFs, so we focused solely on the 1420 MHz window and shrunk the bandwidth to 12.5 MHz.

2.3 Observing the Calibrator 3C48

To ensure the spectra were properly calibrated, that is, converted from raw counts to antenna temperature, the nearby flux calibrator 3C48 was observed before the HI mapping. We first conducted pointing scans, an example of which is shown in Figure 4, and then conducted an “OffOn” position switching with each position lasting 60 seconds. An example of the raw spectra is shown in Figure 5.

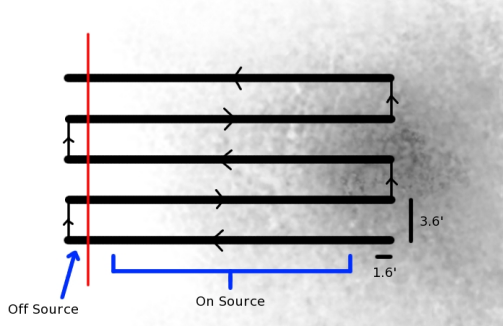


Figure 3. An illustration of the observing method. We scan in right ascension, recording data at 1.6' intervals, and then jump in declination to the next row by 3.6'. Our map is extended to the east to include areas believed to be free of emission.

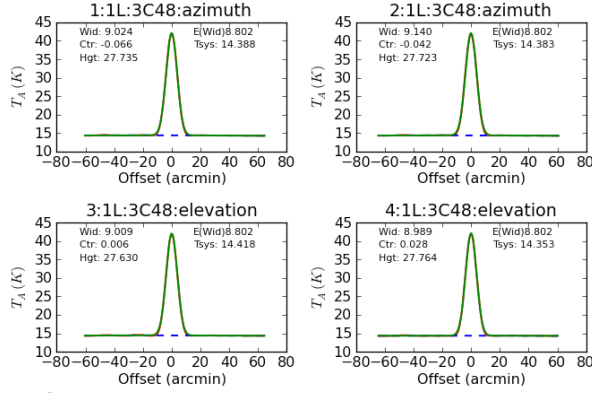


Figure 4. An example of a GBT pointing on 3C48, as seen in ASTRID, to examine the main beam response of the telescope. Here, the beamwidth is correct ($\sim 9'$) and the gaussian height corresponds to the known brightness of 3C48 (in K).

2.4 Mapping the HI

Our maps are $6^\circ \times 2^\circ$ in size, in which we chose to perform a raster scan in right ascension, sampling at three-second intervals every 1.6'. The telescope then jumped in declination to the next scan row by 3.6'. This ensured that our area was Nyquist sampled so that no spatial information in our maps was lost. There are 34 rows in all, with each row taking 11.25 minutes to complete. This area was observed repeatedly for a total of 400 hours. An example spectrum from one three-second integration is shown in Figure 6.

3. Data Reduction

3.1 Concept

To remove the bandpass structure in spectral data, one must subtract the reference spectra from the on source spectra so that only the spectra of the desired source remains. In terms of temperature:

$$T_{\text{src}} = T_{\text{CAL}} \cdot \frac{P_{\text{sig}} - \langle P_{\text{ref}} \rangle}{\langle P_{\text{CALon}} - P_{\text{CALoff}} \rangle_{\text{ref}}} \quad (5)$$

T_{CAL} can be determined from observations of a flux calibrator as described in Section 3.2. P_{sig} and P_{ref} are the input power levels, in counts, and P_{CALon} and P_{CALoff} are the power levels when the front end noise diode is on and off respectively. Since we use the eastern most integrations of our map as the

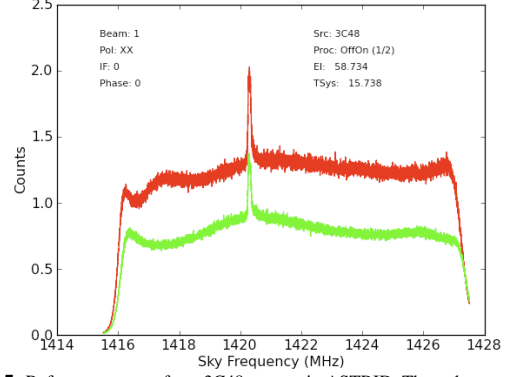


Figure 5. Reference spectra from 3C48 as seen in ASTRID. The red spectrum is for the X polarization and green is for the Y polarization.

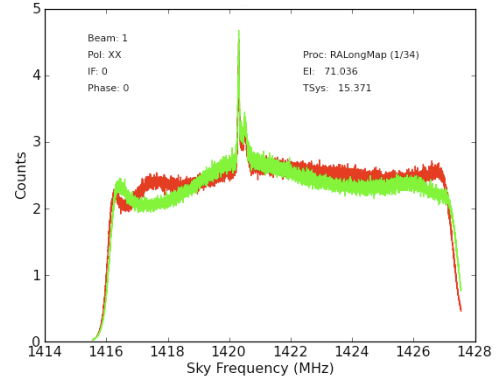


Figure 6. The spectra, X and Y polarizations, from one three second integration of the HI mapping scans.

reference, we average them for Equation 5. The reduction of the raw data was done in GBTIDL¹.

3.2 T_{CAL} from 3C48

The “OffOn” observations of the flux calibrator 3C48 were used to determine the T_{CAL} values for a given observation. The data from a particular observing session were first loaded into a working directory by using the `sdfits` command.

We then took the 3C48 scans and calculated the T_{CAL} values using `scal.pro`², contributed by Ron Maddalena. The inputs of relevance here are for the opacity and aperture efficiency, namely:

$$\text{tau}=[0.01], \text{ ap_eff}=[0.657543]$$

The T_{CAL} values obtained from our observations were remarkably stable, even over the period of several months, with values of $1.56 \pm 0.05(1\sigma)$ for both polarizations. A histogram of the T_{CAL} values over 103 observing sessions spanning 11 months is shown in Figure 7.

The T_{CAL} values for each polarization were then applied to the reduced spectra (Equation 5) for that particular observing session. The stability of T_{CAL} over such an extended

¹<http://gbtidl.nrao.edu/>

²<http://www.gb.nrao.edu/GBT/DA/gbtidl/release/contrib/contrib/scal.html>

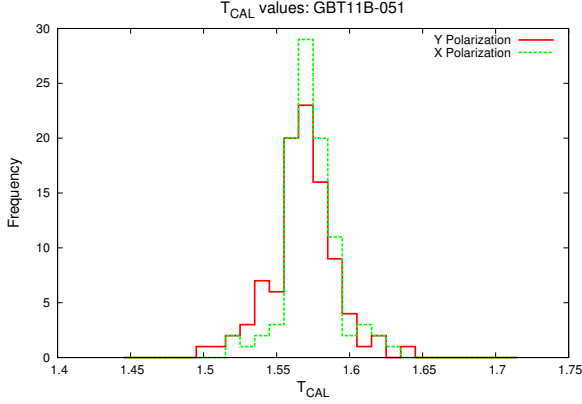


Figure 7. A histogram of calculated T_{CAL} values, for both polarizations, over a period of 11 months.

period is quite significant. For instance, if the observation of a flux calibrator is not possible for a particular observing session, one could use the inferred value of T_{CAL} from recent observations with a likely error of only a few percent.

3.3 HI Mapping Reduction

Each telescope scan in right ascension is defined by a sequence number, in our case, ranging from 1 (the first row) to the 34 (the last row). For our observations, odd sequences scanned from west to east and even numbers east to west. Thus, to select the eastern most integrations per scan, we first called a scan with `gettp` and used a modulo 2 argument to determine whether to use the first few or last few integrations to average for the reference.

Once we determined which integrations to use for the reference, we then extracted the CALon and CALoff phases separately to take the difference. This was accomplished with the `get` command. The difference between the on and off phase was then averaged over all 10 OFF integrations to get $\langle P_{\text{CALon}} - P_{\text{CALoff}} \rangle_{\text{ref}}$. We then collected the reference integrations with `select` and averaged to get $\langle P_{\text{ref}} \rangle$. Then, using `gettp`, each signal integration (P_{sig}) was collected to compute the ratio in Equation 5 and then multiplied by T_{CAL} .

To correct for atmospheric effects, one must multiply the result by a factor that is dependent upon the telescope's elevation at the time the spectra were taken. One can retrieve the telescope elevation, in degrees, with the `!g` structure, scaling the reduced spectra by $e^{\frac{\tau}{\sin(\text{el})}}$ where τ is the opacity of the atmosphere at zenith and el is the elevation of the integration.

A second order baseline was then fit to the spectra to ensure any remaining large scale baseline structure was removed. The region chosen for fitting did not include frequency ranges believed to contain any HI emission.

After the baseline was subtracted, the RMS of the spectrum over an emission free region was measured. If the RMS was less than three times the expected value, we continued, otherwise, the spectrum was discarded. This ensured that most spectra with baseline structure, mainly due to RFI and instrumental effects, were removed. Only a small fraction of the reduced spectra ($\sim 4\%$) had significant enough baseline

structure to be discarded. The remaining spectra were then truncated to the desired frequency range using `dcextract`. Finally, the data were smoothed from 0.3 km s^{-1} to 5 km s^{-1} . We did this using a 16 channel boxcar smoothing and removed the redundant channels. The spectrum was then saved to an output file. This process was carried out for each polarization. An example of the reduced spectra is shown in Figure 8.

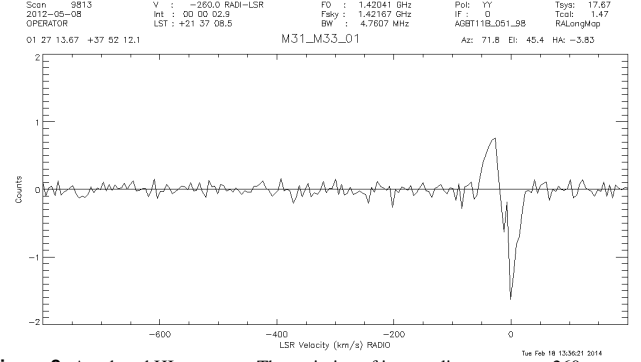


Figure 8. A reduced HI spectrum. The emission of interest lies at or near -260 km s^{-1} , but one integration is not enough for any diffuse emission to be visible. The large swings near 0 km s^{-1} are due to incomplete subtraction of the Milky Way HI.

The RMS noise of the reduced spectrum, from -800 to -260 km s^{-1} , was also measured and recorded. We also computed the expected RMS noise, using the radiometer equation. The effective integration time, using 10 integrations for the reference and 2.9s for the signal, is:

$$t_{\text{effective}} = \frac{t_{\text{sig}} \cdot t_{\text{ref}}}{t_{\text{sig}} + t_{\text{ref}}} = \frac{10}{11} \cdot t_{\text{sig}} = \frac{10}{11} \cdot 2.9 \text{ s} = 2.6 \text{ s} \quad (6)$$

Smoothing the original resolution, 12.5 MHz over 8192 channels by a factor of 16 gives:

$$\Delta\nu = \frac{12.5 \text{ MHz}}{8192} \cdot 16 = 24.4 \text{ kHz} \quad (7)$$

which is equivalent to 5.2 km s^{-1} . T_{sys} was also measured over several scans using

$$T_{\text{sys}} = T_{\text{CAL}} \cdot \frac{\langle P_{\text{ref}} \rangle}{\langle P_{\text{CALon}} - P_{\text{CALoff}} \rangle_{\text{ref}}} \approx 20 \text{ K}. \quad (8)$$

Using this for T_{sys} , along with $K1 = 1.032$ and $K2 = 1.21$ gives:

$$\sigma_T = \frac{1.032 \cdot 20 \text{ K}}{\sqrt{1.21 \left(\frac{12.5 \text{ MHz}}{8192} \cdot 16 \right) \left(2.9 \text{ s} \cdot \frac{10}{11} \right)}} = 74 \text{ mK} \quad (9)$$

From the RMS noise measurements, $\sigma_T = (75 \pm 10) \text{ mK}$, in very good agreement with the predicted value. See Figure 9. The tail toward higher RMS values is due to integrations at low elevation and only includes a small fraction of the data.

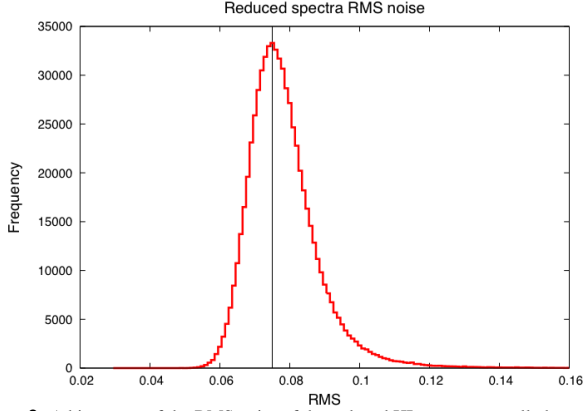


Figure 9. A histogram of the RMS noise of the reduced HI spectra over all observing sessions. The vertical line represents the expected RMS noise value of 74mK from Equation 9.

4. Blanking Issues with the Spectrometer

Upon inspection of our exploratory data (GBT10C-066), we discovered that the system temperatures were increasing as we scanned across our map. This was always an increase, regardless of the scan direction, indicating that this was not an effect due to the sky. The drifts were severe enough, around 2 K, to make the initial data unusable.

This effect was primarily due to Spectrometer blanking. There are two types of blanking. When the noise diode switches phases, from CALoff to CALon and vice versa, the data are blanked during the transition. This is “internal blanking”. “External blanking” occurs when the reference frequency is shifted to account for Doppler tracking across the sky and occurs at indefinite times. In our case, the external blanking was occurring during every CALoff phase and was slowly drifting into the CALon phase. As a result, the exposure times, the total integration duration minus blanking, were changing across the scan and thus causing the system temperatures to appear to drift. See Figure 10 for an example of this.

If we define t_{duration} as the total time of an integration, and t_{internal} and t_{external} as the amount of time the data were blanked by internal and external blanking respectively, then the Spectrometer should, in principle, correct blanking by scaling the power levels by:

$$P_{\text{scaled}} = P \cdot \frac{1}{t_{\text{duration}} - (t_{\text{internal}} \cup t_{\text{external}})}. \quad (10)$$

The data however, are actually scaled by:

$$P_{\text{scaled}} = P \cdot \frac{1}{t_{\text{duration}} - \max(t_{\text{internal}} \text{ OR } t_{\text{external}})}. \quad (11)$$

depending upon whether t_{internal} or t_{external} is larger.

This caused the apparent drift in system temperature. It should be noted that this drift is not an issue with frequency switched data, given that blanking occurs automatically before each integration, preventing the drift. Tracking a source for

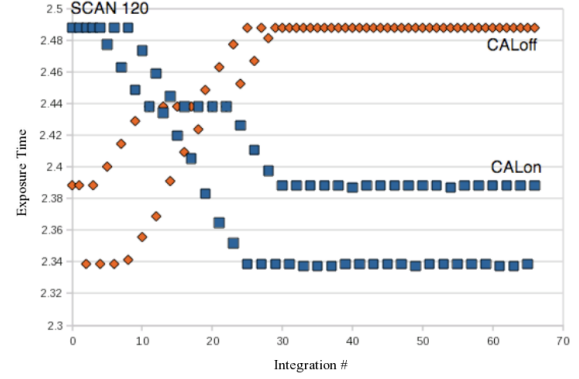


Figure 10. The effect of Doppler blanking on the exposure times of both the CALon and CALoff phases. The rapid oscillations are because the blanking happens every other CALoff phase and then drifts to the CALon phase across the scan.

position switching also reduces the effect since one is not moving across the sky which would require regular updates to the Doppler correction. Furthermore, the closer one is to the vernal equinox point on the sky, the more corrections that are required, enhancing the effect. We have confirmed that the effect is present in other data sets, as shown in Figure 11.

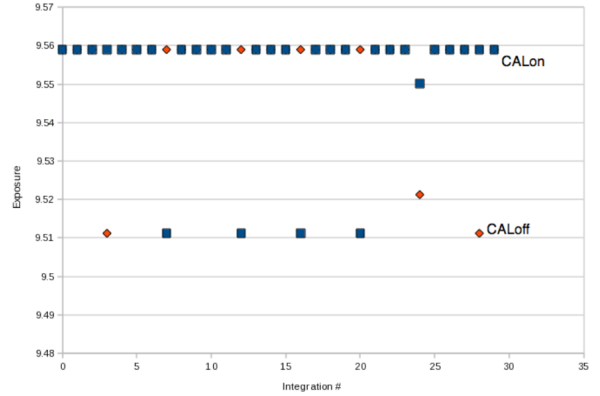


Figure 11. Exposure times taken from a tracking scan of GBT10A-043. The effect of drift is present, although not as severe. Data retrieved from the NRAO archive.³

To prevent this issue from occurring, the following was added to our observing scripts, after configuration and before the telescope slew:

```
SetValues("L01", {"tolerance": 1000000})
SetValues("L01", {"state": "prepare"})
```

This increases the amount of allowed, yet still small, drift in frequency before Doppler tracking occurs and corrected the problem of system temperature drift in our scans.

5. Gridding the Spectral Line Data

To make maps of the emission, one must grid the spectral line data into a data cube. We used the Astronomical Image Processing System, or AIPS⁴. First, the data were converted from their IDL format to an sdfits format that AIPS

³<https://archive.nrao.edu/archive/advquery.jsp>

⁴<http://www.aips.nrao.edu/index.shtml>

could utilize. This was accomplished by using the command `idlToSdfits`, written by Glen Langston. It is important to note that data blocks with repeating scan numbers can not be used in `idlToSdfits`, or else the data with repeated scan numbers will not be converted. The repeating scans must be loaded separately. It should also be noted that `idlToSdfits` can not convert 100,000 or more scans at one time.

After the data were converted into `sdfits`, each block of data was then loaded into AIPS using the task `UVLOD`. Each reduced data block had to be concatenated together for the gridding by using the task `DBCON`.

Once the data were concatenated together, we used the task `SDGRD` for the gridding. Here, one can use inputs that are typical for HI mapping with the GBT [5], such as:

```
CELLSIZE 210
BCHAN 0
ECHAN 0
XTYPE -16
YTYPE -16
XPARM 550 275 462 2
YPARM 550 275 462 2
```

which convolves the data with a exponentially truncated, circular Bessel function. Other types of functions can be used via the `XTYPE`, `YTYPE` inputs and the corresponding `XPARM`, `YPARM` parameters, depending upon the users preference. One can review the AIPS `SDGRD` help page at aips.nrao.edu for further information.

Once the data are gridded into a cube, the standard analyses can be performed, such as creating moment maps with the `TRANS` and `MOMNT` tasks. It is also advisable that one subtract out any remaining baseline structure using the task `XBASL`. An example moment-zero map of our current survey data (GBT11B-051, GBT13A-312), using a `CELLSIZE` of 105, is shown in Figure 12.

Acknowledgments

We would like to thank the Green Bank staff for their support, especially in resolving the blanking issues with the GBT Spectrometer and utilizing `idlToSdfits`. We also thank the telescope operators for their support during the observations and conducting some of the observations when the projects were declared operator-run.

References

- [1] Maloney, P. 1993, *ApJ*, 414, 41
- [2] Dove, J. B., & Shull, J. M. 1994, *ApJ*, 423, 196
- [3] Braun, R. & Thilker, D.A. 2004. *A&A*, 417, 421
- [4] Lockman, F.J., Free, N.L., & Shields, J.C. 2012, *AJ*, 144, 52
- [5] Mangum, J. G., Emerson, D. T., & Greisen, E. W. 2007, *A&A*, 474, 679

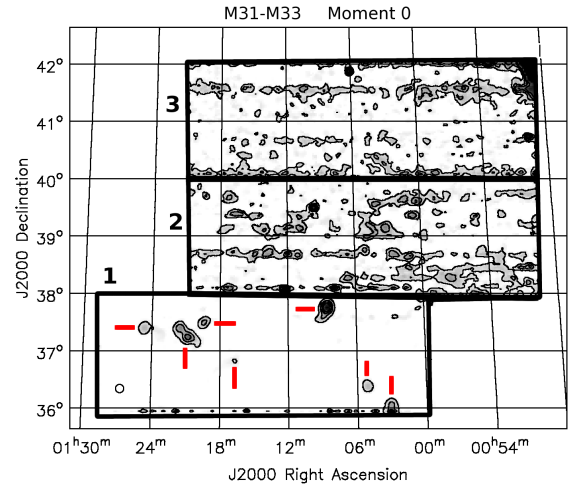


Figure 12. A moment zero map of our HI survey data. Region 1 is for 400 hours of observations (GBT11B-051), while regions 2 and 3 are for ~ 70 hours and ~ 30 hours respectively (GBT13A-312). In region 1, we reach a 3σ sensitivity of $N_{\text{HI}} \sim 10^{17.3} \text{ cm}^{-2}$. We reach sensitivities of $\sim 10^{17.6} \text{ cm}^{-2}$ and $\sim 10^{17.8} \text{ cm}^{-2}$ in regions 2 and 3 respectively. The red lines are currently detected 21 cm sources.