Measuring the 21-CM Line and Waveguides

Matthew Cardozo
matthewcardozo012@berkeley.edu

March 12, 2019

Abstract

Constituting nearly 3/4 of all matter, hydrogen is the most abundant element in our universe. This makes the emission of hydrogen at the 21-cm line one of the easiest for radio astronomers to measure. With this in mind, the objective of this paper was to lead an investigation into the methods and techniques required to measure the 21-cm line. We learned that to resolve such a signal, proper analysis must account for factors like noise and Doppler shifts both of which are inherent to our data. Our research suggested that data collected from a pointing at galactic coordinates $(120^\circ, 0^\circ)$ returns a higher calibrated line intensity centered at 0.355 MHz than the non-pointed data centered about 0.405 MHz. We concluded that despite the non-pointed data being closer to the expected 21-cm value of 0.4 MHz, the signal is actually much more heavily distorted with noise causing its peak amplitude to differ with other peaks by only 15 kelvin whereas our pointed data returned a signal whose peak was roughly 40 kelvin higher than the rest of the line.

The latter portion of this lab demonstrated that the use of waveguides in calculating wave velocities can return a measurement of the width of the waveguide with an error of only 3.64%. Both calculations in the report exercise the importance of statistical methods, outlining how techniques such as polynomial fitting help solve for relationships amongst variables in collected data.

1 Introduction

This report is split into two portions: the first serves to build on the measurement of the 21-cm line, while the second investigates the use of waveguides as a method of measuring various wave velocities. The report is thus organized so as to give proper consideration of both.

Section 2 of this report concerns itself with the electronics of the horn and the digitalization of our signal. Section 3 reviews our pointed and non-pointed data, and covers how the noise was reduced in effort to calibrate a final signal intensity. Section 4 dives into the analysis of our signal and the development of a signal intensity as a function of frequency. The final section of this report covers the waveguide portion of the lab and compares the velocities calculated in both the C-Band and X-Band devices.

2 Telescope Electronics

The Big Horn atop Campbell Hall served as our telescope for the duration of this lab. We used the horn to measure signals from pointings at the zenith, above Campbell Hall, and at galactic coordinates $(120^\circ, 0^\circ)$.

Because the signals from space fluctuate at relatively high frequencies, we used a series of mixers (DSB/SSB) and band-pass filters (BPF) to reduce the frequencies down to a range between -2 MHz and 2 MHz, which were in turn further amplified (AMP) to distinguish from other frequencies. We worked within this range to focus on measuring the Hydrogen I 21-cm line (HI) which we know emits at about 0.4 MHz once reduced and properly digitalized.

2.1 Signal Path

Despite the mechanics of the horn, the data acquisition process actually began prior to the measurement of the signal. This goes to consider that we were able to set our own input parameters for how to acquire the signal. Our initial parameters included a volt range of 50 Mv, a sampling frequency of 12.5 MHz, and a set of 99 blocks of data.

The first two amplifiers were within the horn and served the purpose of amplifying our signals about a 15 MHz bandwidth (BW) at a frequency of 1420.4 MHz, that of the HI line. This helped focus on a range of frequencies that captured our desired measurement. Following these two pre-amplifiers, the signal reached the mixer board located in the lab and was further reduced by a BPF with a BW of 320 MHz. It then interacted with an additional two amplifiers before reaching the first DSB Mixer at an amplified signal around 1420.4 MHz. The DSB mixer met our frequency with an LO of 1230 MHz, which after taking the sum and difference of the two left us with signals at 190.4 MHz and 2650.4 MHz. We then used our second BPF focused at 190.4 MHz and with a BW of 20 MHz to isolate this signal from the larger 2650.4 MHz signal. The signal then met our fifth and final AMP before entering the SSB mixer, which with an LO of 190.0 MHz gave us a signal centered at 0.4 MHz. To account for Doppler shifts in our frequencies we imposed a low pass filter of 2 MHz to include frequencies ranging from -2 MHz to 2 MHz, still while expecting to see the HI line at about the 0.4 MHz frequency mark.

2.2 Relevant Programming

Below are two of the functions we defined to help capture the data used for this lab. This data can be found under DIR Lab 2, in the SUBDIR titled 'Final Lab 2'. What follows are the doctrings listing the input parameters and variables that were returned as a result of their implementation.

```
def scale(data, volt_range):
Purpose - To rescale sampled data.

Parameters:
data (array): data to be rescaled
volt_range (double) : volt range input parameter

Returns:
output (array): rescaled data


def load(filename, volt_range, nblocks, nsamples = 16000):
Purpose - To load a file with sampled data.

Parameters:
filename (string): name of the file
```

```
volt_range (double): volt range input parameter
nblocks (int): number of data blocks

Returns:
complex (array): rescaled data, split by block
real (array): real component
imag (array): imaginary component
```

The purpose of the scale function was to account for our desired voltage amplitude of 50 Mv and reduce the data to a symmetric data set of fewer cycles per second, making the data agree with the parameters of our Pico Sampler and with smaller, more reasonable numbers to work with. What stands out from our load function, other than serving to load in our sampled data, is the parameter that requests the number of blocks which for this lab was analogous to the 99 individually collected data sets. The load function uses 'nblocks' to splice our real and imaginary components of the signals into 99 arrays each with 16,000 samples.

3 Observing

Since this lab was concerned with the measurement of the HI line, it was important to compare the results taken from two different pointings of the telescope - one being pointed directly up at the zenith, and the other being pointed at galactic coordinates $(120°, 0°)$ where we presume to receive the highest level of HI emission.

3.1 Not Pointed vs. Pointed

Our non pointed measurement refers to our observation of the signals received from the zenith. The coordinates in RA and DEC were $3.59°$ and $37.87°$ respectively. The galactic coordinate of the pointed data translated to an RA and DEC of $-6.58°$ and $-104.81°$.

Despite measuring a peak for our non-pointed at 0.4 MHz, this data differed from our pointed measurement in that it was characterized by high amounts of noise. We noticed that there were a series of peaks in the calibrated line intensity that challenged the strength of our HI measurement. It is evident from Figure 1 how these various peaks were absorbed in the calculation of the signal's on-line velocity. We realized that the non-pointed velocity not only deviated further from its Doppler velocity than our pointed measurement, but it also seemed to undergo a red-shift in contrast to the blue-shift experienced by that of our pointed signals. This data suggests that the source of HI emission we measured from the zenith is actually moving away from Earth so as to induce this relative red-shift.

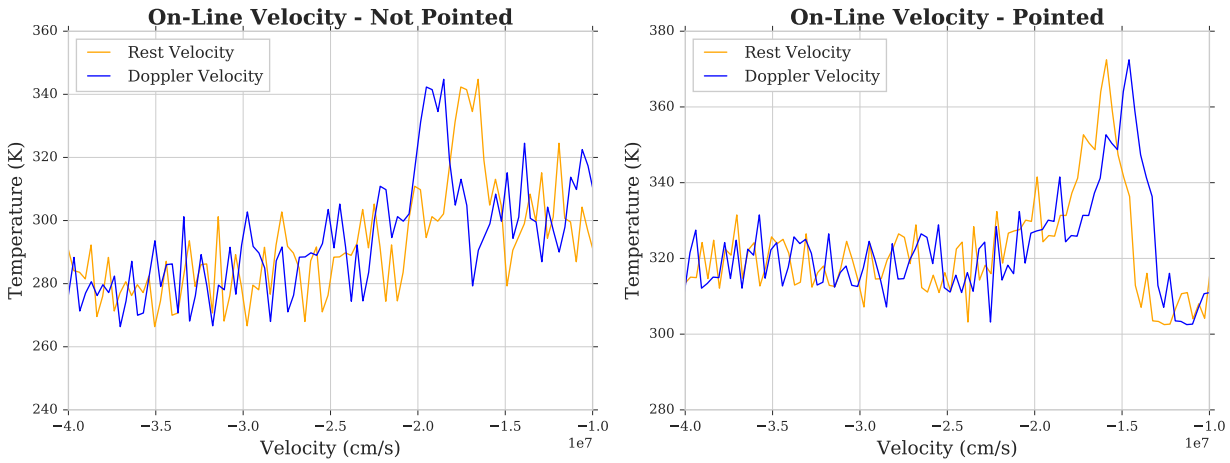

Figure 1: This graph compares the rest velocity with the Doppler velocity of our on-line signal in both the non-pointed and pointed data. The non-pointed data represents a red shift as the velocity decreases, whereas the pointed data represent a blue shift with an increasing Doppler velocity relative to the rest.

3.2 Time and Coordinate Systems

The time logs used throughout this lab were provided by the *ugradio.timing* package. The logging of the Julian Day was done at the time of measurement, whereas the the Local-Sidereal Time was calculated with the *lst* function using our Julian Day as the relevant input parameter. These times were used to determine the RA and DEC of our pointings, though our pointed measurement, while not simply pointed at the zenith, required additional application of rotation matrices. We started by defining our rotation matrix $R_{eq,gal}$ and galactic coordinates vector $\vec{\boldsymbol{x}}$ as:

$$R_{eq,gal} = \begin{bmatrix} -0.054876 & -0.873437 & -0.48383 \\ 0.494109 & -0.444830 & 0.746982 \\ -0.867666 & -0.198076 & 0.455984 \end{bmatrix}, \vec{\boldsymbol{x}} = \begin{bmatrix} 120 \\ 0 \\ 0 \end{bmatrix} \tag{1}$$

As labeled, the rotation matrix was used to convert from equatorial to galactic coordinates. By inverting $R_{eq,gal}$ and taking its dot product with $\vec{\boldsymbol{x}}$, we were able to calculate a vector of the following format which gave us the RA and DEC with respect to Campbell Hall.

$$\vec{\boldsymbol{r}} = \begin{bmatrix} -6.58 \\ -104 \\ -58.1 \end{bmatrix} = \begin{bmatrix} RA \\ DEC \\ \gamma \end{bmatrix} \tag{2}$$

With this method we were able to calculate a projected Doppler velocity of 12.9 km/s for our pointed signal, leaving our signal's adjusted velocity just under the speed of light as graphed by Figure 1.

3.3 Noise Reduction

Noise served a significant role in the capturing of our data. Figure 2 demonstrates the smoothing techniques we applied to remove noise from our signals. These techniques included calculating average and median

signals amongst the 99 captured data sets. Therefore, we were able to combine said data to construct a new signal that essentially displayed constructive and destructive interference. As displayed below, the averaged power spectra produced data with a much higher signal to noise ratio, although the smoothed median power spectra was the only plot able to remove our peak at $\nu = 0$, thus leading us to use these upper and lower bands in constructing our on-line and off-line signals.

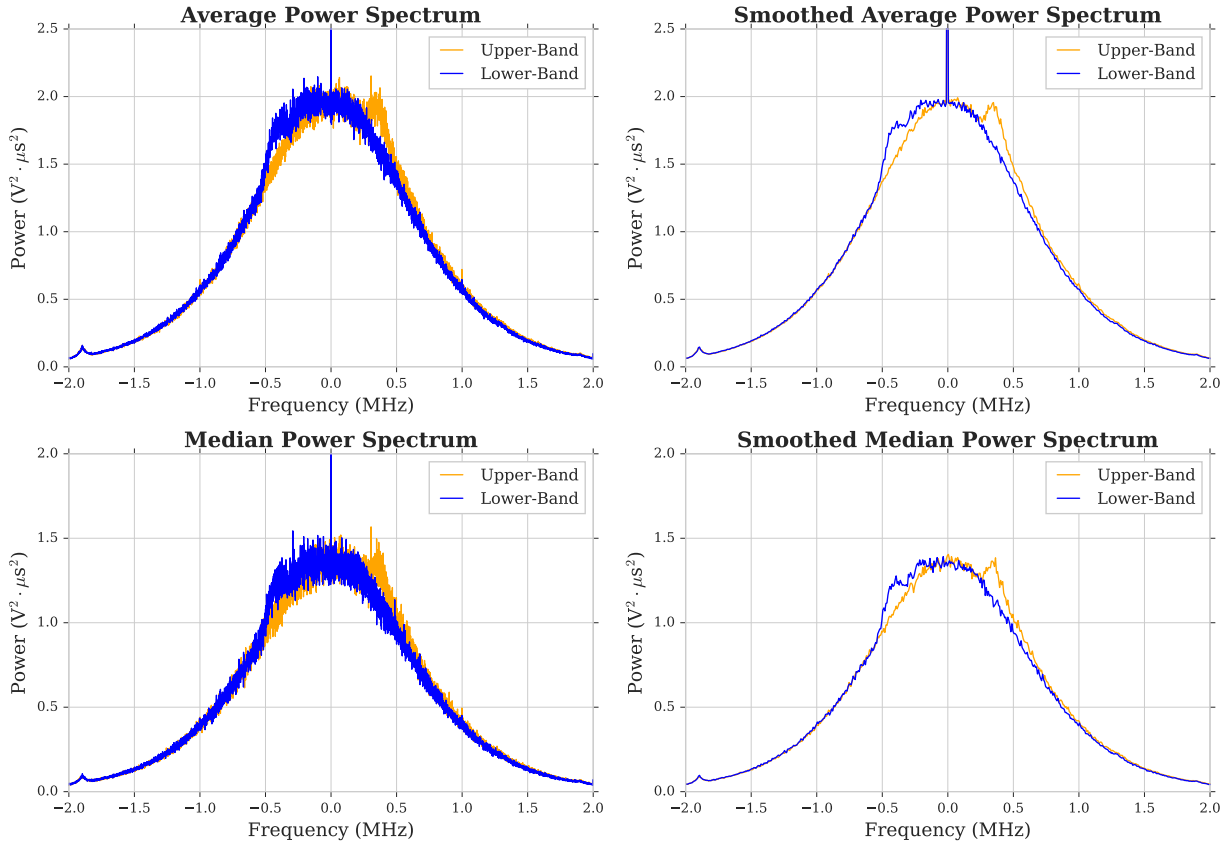

Figure 2: This graph compares the average and median power spectra with the smoothed average and median power spectra for our pointed data. The peaks in the upper and lower band represent the implementation of 1 MHz low pass filters, accounting for higher intensity at the 21-cm line - -0.4 MHz and 0.4 MHz.

4 Data Analysis

4.1 Online and Offline

The online and offline measurements were reminders of our initial 1 MHz low pass filter preset in the SSB mixer. As expected, the upper sideband of our power spectrum corresponded to a spike in the positive frequency range at approximately 0.35 MHz, whereas the lower sideband corresponded to a spike at -0.35 MHz - refer to Figure 2. To construct an on-line signal, we used *np.concatenate* to join the left-half of the lower band with the right-half of the upper band, and vice-versa for the off-line signal. It was made clear

from our graphs in Figure 3 that our pointed data would contain much less noise than our non-pointed data - both equally smoothed by a bin division of 10. The difference in signal to noise ratio between the two spectra also helps explain why the pointed line shape and intensity in Figure 1 fluctuates less than that of the non-pointed line.

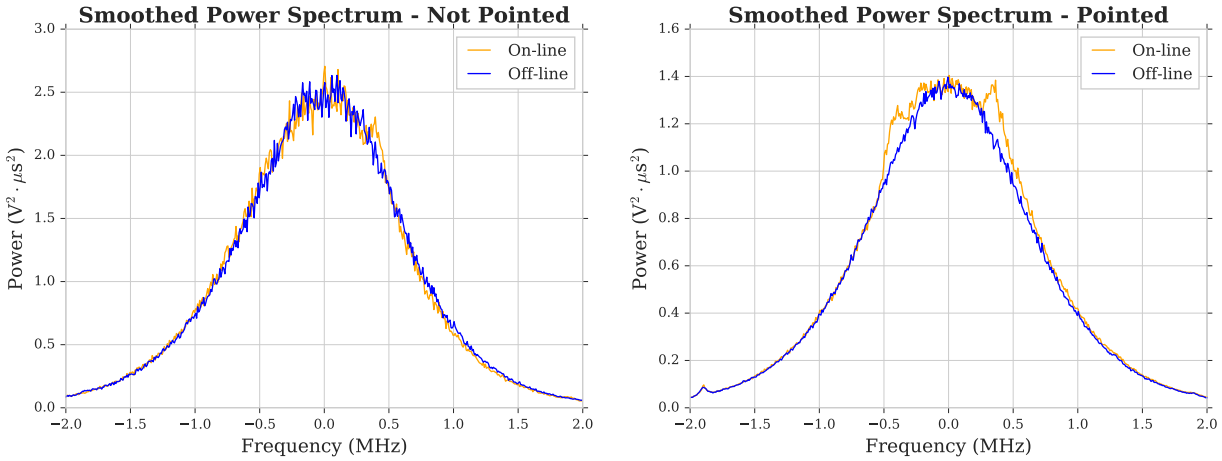

Figure 3: This figure compares the smoothed power spectrum of our non-pointed data with the smoothed power spectrum of our pointed data. The on-line signals for both contain the peaks that we obtained from upper and lower bands. It is clear from these plots that our non-pointed was noisier than our pointed data.

4.2 Calibrating Line Intensity

As mentioned earlier in the section, the gain factor was important in developing a relationship between the intensity of the signal and its range of frequencies. Using Equation 3, we were able to determine a gain factor of 304.2. We got to this number by taking the product between the difference in our system temperature (300 Kelvin) and the sum of our average cold powers, and dividing that number by the sum of the differences between our average calibrated powers and average cold powers. Using this gain factor, we were then able to scale the ratio of our on-line to off-line signals to achieve a line intensity in units of Kelvin as demonstrated by Equation 4.

$$G = \frac{T_{sys,cal} - T_{sys,cold}}{\sum(s_{cal} - s_{cold})} \sum s_{cold} \tag{3}$$

$$T_{line} = \frac{s_{on-line}}{s_{off-line}} \times G \tag{4}$$

The following plot shows the intensity as a function of frequency for both our pointed and non-pointed signals.

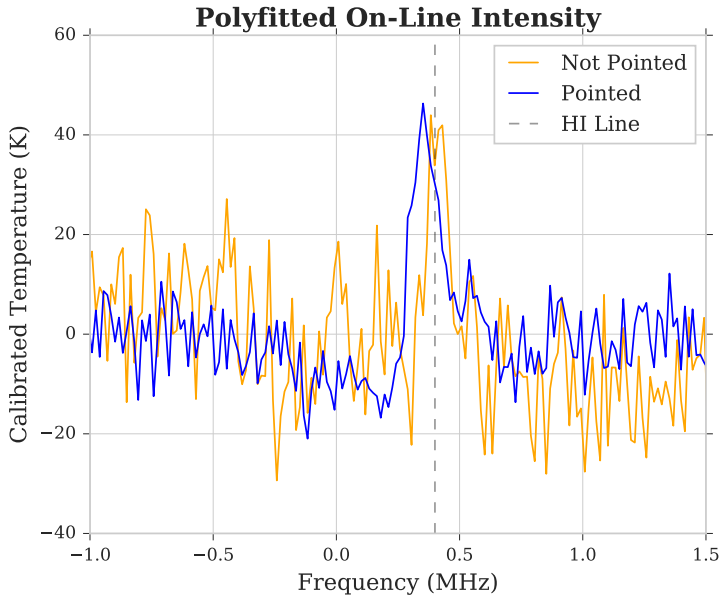

Figure 4: This graph compares the polyfitted on-line intensity of our non-pointed data with the pointed data. The dashed line represents were we expected to find the peaks about the 21-cm line. It is evident that the non-pointed data had a peak frequency who's amplitude was more closely proportional to the line than our pointed data. Therefore, the pointed data produced a more a accurate measurement of the 21-cm line.

4.3 Relevant Programming

After obtaining our on-line signal we encountered the issue of dealing with two peaks, one in the positive frequency range and another in the negative frequency range. In effort to resolve the signal down to a single peak, we designed a filter to trim our data and remove both the frequencies and their associated powers outside of our -2 MHz to 2 MHz frequency range. We used *np.argmax* to return the indices of our two maximum values and indexed our arrays to have two smaller arrays about the two maximum values. By overlaying one peak with the other, we were able to determine a final calibrated peak by averaging over the sum of both arrays. The resulting array produced one peak at 0.35 MHz.

```
filter = np.ones(len(smooth_med), dtype = bool)
for i in range(len(med_freqs)):
    if med_freqs[i] <= -2 or med_freqs[i]>= 2:
        filter[i] = False

smoothed_arr = np.array(smooth_med)
filtered_smoothed_temp = smoothed_arr[filter]
filtered_med_freqs = np.array(med_freqs)[filter]
max_inrange = np.argmax(np.fft.fftshift(filtered_smoothed_temp))

total = np.fft.fftshift(filtered_smoothed_temp)[80:130] +
        np.fft.fftshift(filtered_smoothed_temp)[130:180]
```

```
mean = total / 2

new_y = np.concatenate((np.fft.fftshift(filtered_smoothed_temp)[0:80], mean,
        np.fft.fftshift(filtered_smoothed_temp)[180:]))
```

4.3.1 Function Fitting

Function fitting was used throughout the lab for various reasons. The Gaussian fitting approach allowed us to exercise a brute-force method in determining elements of our data such as the amplitude and standard deviation from the mean. It helped in visualizing the contribution that frequencies about 0.35 MHz had on the final line's intensity. The Polynomial fitting technique solved for a least-squares solution to a polynomial which shared similar input parameters with our measured signal. As shown below and in the plot of Figure 4, the resulting Polyfit model helped us reshape our data in a way that centered the signal about $T_{cal} = 0$.

```
Polyfit
new_x = -np.fft.fftshift(filtered_med_freqs)[:-50]
new_y = np.concatenate((np.fft.fftshift(filtered_smoothed_temp)[0:80], mean,
        np.fft.fftshift(filtered_smoothed_temp)[180:]))

fitter = modeling.fitting.LevMarLSQFitter()
model = modeling.models.Gaussian1D()

adjusted_y = new_y - fitted_model(new_x)
```

5 Waveguides

The C-Band measurements corresponded to a wave whose frequency was at 3 GHz, whereas the X-Band measurements corresponded to waves of frequencies 7.5 GHz, 8 GHz, and 8.5 GHz. The table below shows the velocity of each of our measurements with respect to their frequency.

Band	Open/Closed	ν(GHz)	Velocity (10^9 cm/s)
C	Open	3.0	2.994
C	Closed	3.0	-0.843
X	Open	7.5	0.105
X	Closed	7.5	0.105
X	Open	8.0	0.068
X	Closed	8.0	0.068
X	Open	8.5	0.069
X	Closed	8.5	0.075

5.1 C-Band

The velocities of the waves in the C-Band were quite different than those measured from the X-Band. As Table 1 notes, then open velocity in the C-Band was approximately that of the speed of light, whereas the

closed waveguide measured a negative velocity a fraction of that of the speed light. Our initial interpretation was that human error had faulted our second measurement. After careful consideration, we instead interpreted this calculation as being a reflected wave. Provided the initial open velocity had read very near that of the speed of light, we found it reasonable to assume that the closed waveguide, with an impedance of 60 Ohms, was actually an insufficient amount of impedance so as to cause our wave to be reflected thereby leading us to measure a velocity in the opposite direction.

5.2 X-Band

The velocities of the waves in the X-Band were nearly the same in both cases - the exception was our measurement of the wave whose frequency measured 8.5 GHz. This was unexpected considering the 8.0 GHz wave had an open velocity 2/1000 smaller than that of the 8.5 GHz wave. We therefore attributed this difference to human error, suspecting that our 0.25 cm margin of error in the null measurements led to a difference in the calculated velocities.

The Voltage Standing Wave Ratios we measured from the 8 GHz X-Band were 2.0 for the open waveguide and 15.0 for the closed. This indicated that the magnitude of our minimum voltage was lower than the magnitude of our maximum voltage. With an impedance of 1 corresponding to a VSWR that is also 1, this data led us to conclude that the impedance of our X-Band waveguide in both the open and closed scenarios was able to match the frequency of our waves. This supported why our measured X-Band velocities, as shown in Figure 5, were nearly the same in both cases.

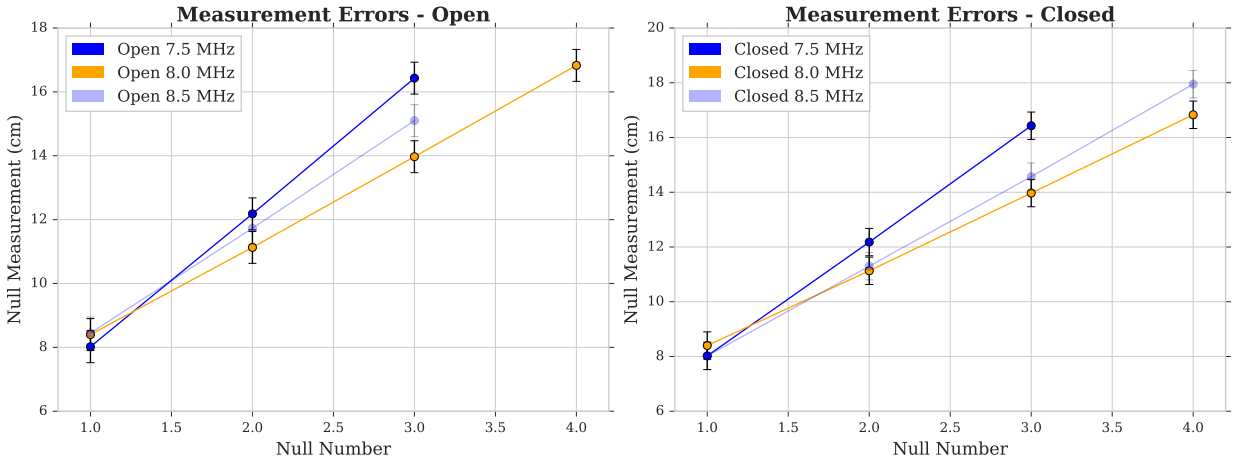

Figure 5: This figure compares the errors associated with the measurements taken from the open and closed X-Band waveguide. The plots demonstrate that the measurements were mostly the same in most cases with an exception to the 8.5 MHz signal. The error associated with these measurements was 0.25cm.

6 Conclusion

Although it was unclear to us exactly what was the source of HI emission from the zenith, we were able to support our theory of Doppler shifting by taking our pointed data with respect to the spiral arms of the

Milky Way. Understanding that these coordinates aimed at a point on a spiral arm which in fact was moving towards Earth, was a huge clue in helping us validate our pointed data. As expected, Figure 1 highlights how the Doppler velocity of our pointed on-line signal was greater than its rest velocity, thereby indicating the presence of a blue-shift. It was also clear to us which of the two on-line signals had a greater intensity. With the amplitude of our pointed data over 20 kelvin larger than our non-pointed data, we were able to confirm our expectation that the pointed source was a stronger emitter of HI than the non-pointed source.

The majority of the analysis done in this lab was concerned with developing our time-based frequency signal into a plot of intensity as a function of both frequency and velocity. As we've encountered before in the course, our initial signal represented a wave shape as a function of time. We used the Fast Fourier Transform to display our signal as a function of frequency, then allowing us to calculate the power of the signal as a function of frequency. To extrapolate an intensity from this power, we determined a gain factor 304.2 and used it to scale our power spectrum in units of temperature. The result was a calibrated line intensity with a peak just about the 21-CM line for both the pointed and non-pointed data.

The second portion of this lab was concerned with measuring and comparing the velocity of waves in a waveguide whose end was either open (without stop) or closed (with stop). Using this data and an application of the polynomial fitting, we were able to determine a waveguide width of approximately 2.12 cm, short of our expectation of 2.2 cm by only 3.64%.

6.1 Contributions

The functions used to capture data for this lab were written and shared primarily by Romain - these are the functions reviewed in the Relevant Programming section 2.2. The filter program covered in section 4.3 was designed by both Oscar and myself. Outside of the *ugradio* package, *astropy.modeling* was used to construct the polynomial fitting reviewed in section 4.3.1.