University of Virginia

Department of Astronomy
ASTR 4998 - Undergraduate Senior Thesis

Reduction and Calibration of GBT Radio Data of the Andromeda Galaxy

Presented by:
Jacob Halenda

Supervisor:
Professor Bradley Johnson
Co-Supervisor:
Liam Walters

Spring 2024 – Graduation Date: 05/18/2024

This thesis is submitted in partial completion of the requirements of the BA Astronomy Major on 05/09/2024.

Abstract

The primary objective of this thesis is to express the importance of data analysis through the reduction and calibration techniques throughout observational radio astronomy. I use a Python data pipeline written by my co-supervisor to explore data filtration techniques and data visualization methods. I do not explore every aspect of the pipeline but attempt to understand and convey all aspects I do cover.

Contents

Chapter 1

Introduction

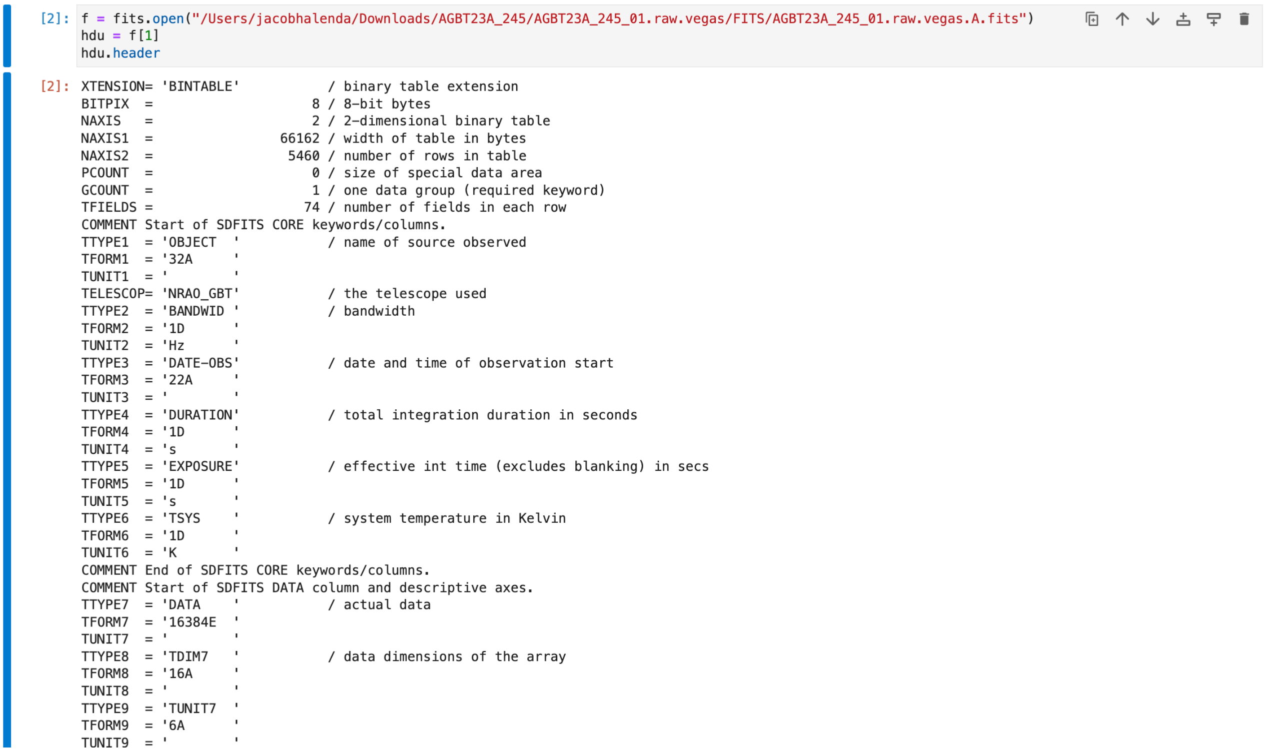
My motivation and goal is to find evidence of dark matter through axion spectral line emission. I am observing neutron stars because they are a great environment for axion signals in the form of axion-photon conversions. These signals come from the strong magnetic field emitted around the neutron star (NS). The electron plasma surrounding the neutron star enhances the probability of this axion-photon conversion which can be detected as a microwave signal. These emerging photons can be recognized as a spectral line in the GHZ frequency range and measured with radio detectors (1). Andromeda galaxy (M31) has many neutron stars, is the closest galaxy to the Milky Way (2.537 million light years (ly)), and easy to observe. The observations of M31 were done by my co-supervisor from the Green Bank Telescope (GBT) in July 3-5 and October 19, 2023, and the data was compiled into a single-dish fits (SDFITS) file. I use the data entries from these four observations of the C-band frequency range (split up into sections (banks) of the 4-8 GHz range), and plot the frequency values of the reduced signals with their respective spectral flux density. I use a data pipeline written by my co-supervisor with examples of the reduction and calibration functions, then I use these functions to reduce, calibrate, and filter my data. Plotting is an important tool in data analysis in order to visualize what my data is doing. Learning the details and intricacies of the data pipeline is an important goal of mine to further my understanding of not only data calibration techniques but also the processes of radio astronomy as a whole.

Chapter 2

Methods

2.1 Observations

For this thesis, in order for me to make plots and visualize this data, I needed to understand the formatting of the FITS file. The file is in a binary table extension format and the header contains information on each of the rows (2). Some information on the observation's date, temperature, elevation, frequency resolution, bandwidth, integration time, etc. are found in figure 2.1. I worked with the DATA column in order to graph my intensity vs. frequency plots. I used the reduction function through the pipeline which cuts off 10% of the channels from either side of the data bank which avoids the drop in sensitivity at the edges. I plotted four of the pairs of calibration on/off source with diode on/off and four pairs of on/off source of M31 with diode on/off in Figure 2.2. This observing method is called position switching where the telescope is moved at an offset position close to an on-source position by looking through the same airmass (3). This is accomplished by choosing an offset position not too different in declination, since looking at a lower declination would introduce more noise in your signal by looking through more atmosphere. The on-source coordinates of M31 are RA/Dec (00:42:44.330, +41:16:07) with the off-source position being RA/Dec (00:45:00.000, +41:02:00). We use this particular coordinate for the off-source reference pointing because it is a region of non-emission, in order to measure the contributing signal from this background source. This is required when looking at a complex emission region like the Andromeda galaxy. It is useful to have these ratios of on/off scans to differentiate between emission and non-emission regions.



```
[2]: f = fits.open("/Users/jacobhalenda/Downloads/AGBT23A_245/AGBT23A_245_01.raw.vegas/FITS/AGBT23A_245_01.raw.vegas.A.fits")
     hdu = f[1]
     hdu.header
```

```
[2]: XTENSION= 'BINTABLE'           / binary table extension
     BITPIX  =                    8 / 8-bit bytes
     NAXIS   =                    2 / 2-dimensional binary table
     NAXIS1  =                66162 / width of table in bytes
     NAXIS2  =                 5460 / number of rows in table
     PCOUNT  =                    0 / size of special data area
     GCOUNT  =                    1 / one data group (required keyword)
     TFIELDS =                   74 / number of fields in each row
     COMMENT Start of SDFITS CORE keywords/columns.
     TTYPE1  = 'OBJECT  '           / name of source observed
     TFORM1  = '32A     '
     TUNIT1  = '        '
     TELESCOP= 'NRAO_GBT'           / the telescope used
     TTYPE2  = 'BANDWID '           / bandwidth
     TFORM2  = '1D      '
     TUNIT2  = 'Hz      '
     TTYPE3  = 'DATE-OBS'           / date and time of observation start
     TFORM3  = '22A     '
     TUNIT3  = '        '
     TTYPE4  = 'DURATION'           / total integration duration in seconds
     TFORM4  = '1D      '
     TUNIT4  = 's       '
     TTYPE5  = 'EXPOSURE'           / effective int time (excludes blanking) in secs
     TFORM5  = '1D      '
     TUNIT5  = 's       '
     TTYPE6  = 'TSYS    '           / system temperature in Kelvin
     TFORM6  = '1D      '
     TUNIT6  = 'K       '
     COMMENT End of SDFITS CORE keywords/columns.
     COMMENT Start of SDFITS DATA column and descriptive axes.
     TTYPE7  = 'DATA    '           / actual data
     TFORM7  = '16384E  '
     TUNIT7  = '        '
     TTYPE8  = 'TDIM7   '           / data dimensions of the array
     TFORM8  = '16A     '
     TUNIT8  = '        '
     TTYPE9  = 'TUNIT7  '
     TFORM9  = '6A      '
     TUNIT9  = '        '
```

Figure 2.1: Header of the Binary Bin Table for a FITS File

2.2 Data Analysis

2.2.1 Data Structure

The structure of the data is in the form of a binary header table with many different entries and rows that specify which data is contained within. The VErsatile GBT Astronomical Spectrometer (VEGAS) contains 8 spectrometers which are called banks, and the data from each bank is written to a separate FITS file with a letter A-H that consist of a certain frequency bandwidth. As shown in Figure 2.1, the PORT binary table is shown with two rows and groups of three headers in 74 fields of columns. TTYPE specifies the name of the data field, TFORM specifies the format (arrays, one-dimensional, ASCII characters) of the data, and TUNIT shows which units (seconds, Hertz, Kelvins) that the data is in (2). Each data bank is further divided into frequency channels, which in this case add up to 16,384 total channels. I take the mean values of the intensity source associated with each frequency channel. The dynamic range output of the data is limited by the Analog to Digital Converter which is 14 bits ($2^{14} = 16384$). This converts the signal to Analog-to-Digital Units (ADU's) that are readout as distinct values. Unresolved background starlight, moonlight, scattered sunlight, auroral emission from the atmosphere, and light pollution contribute to this single pixel readout. All of this contributing background noise affects the data I am trying to analyze, so I proceed with reducing and calibrating the data.

2.2.2 Reduction and Calibration

There are many different ways to reduce and calibrate data which vary depending on the type of astronomy one does. I simply reduce the data from on/off source pointings into temperature calibrated values. It is interesting to point out that the National Radio Astronomy Observatory (NRAO) have scale calibrators that predict a source of flux density with functions of known

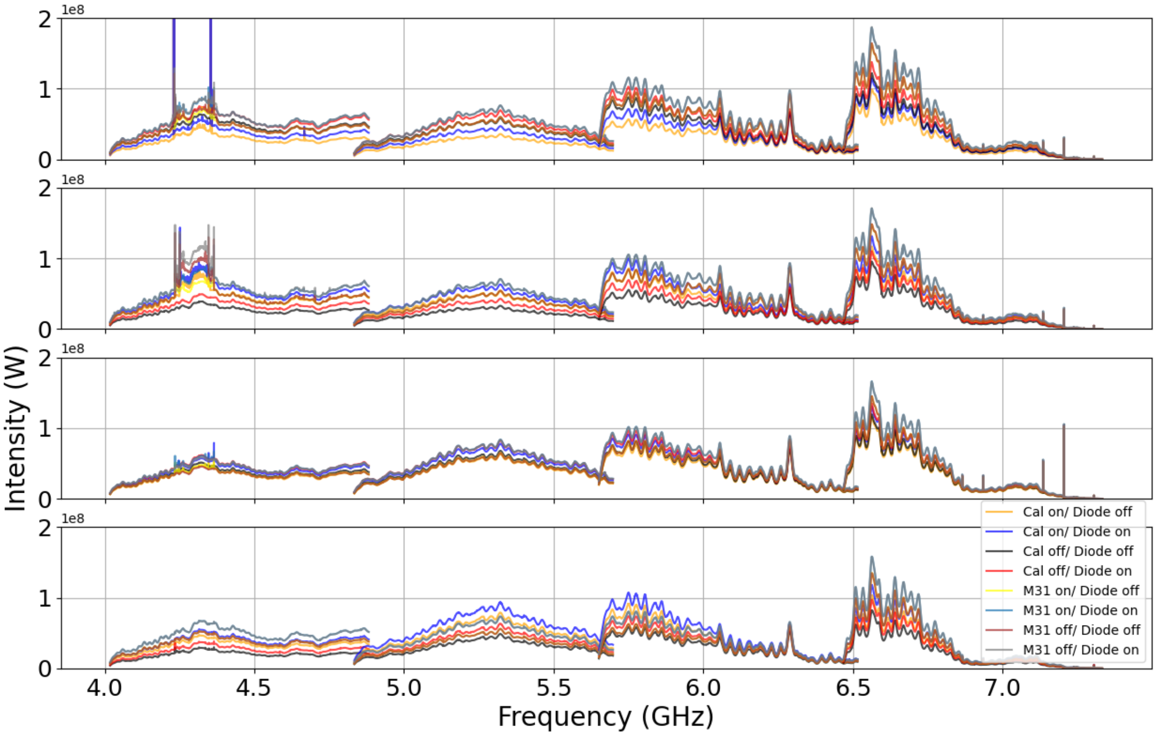


Figure 2.2: 4 Observations of All Reduced Data with the Intensity vs. Frequency Plotted

frequency like the quasar 3C48 (4). Measuring M31's flux density relatively by comparing it to 3C48's known intensity would be an effective calibration technique. However, in this case, I am not using 3C48 to calibrate. I use the reduction function to distinguish the calibration source of M31 from 3C48 in order to confirm my on-source values. The different pairings in the plot indicate the subtle differences of the on/off source 3C48 scans and the M31 scans when the diode is on or off. The diode on the detector has a low constant temperature but still contributes to this signal with thermal induced noise. This is the reason why I am showing the subtle effects of having it on/off. Having this plot in Figure 2.2 serves as a visual aid to see certain data points before I calibrate further. Since I only want to see the M31 data, I plotted another graph of just the M31 on/off source scans in Figure 2.3. After acquiring the M31 on/off source data, I can calibrate accordingly. Each scan constitutes a pairing of 1 off-source pointing and 1 on-source pointing. I take the median values of the off-source and on-source integration pointings since the pairings are not always one-to-one. The on-source pointing is integrated for 30 seconds, and the off-source pointing is integrated for 15 seconds. The resulting plot would have a value of both on-source and off-source points which can be folded to improve the signal-to-noise ratio. This method effectively shows the on-source observations only. Attached to the detector is the diode which is at a specific, constant temperature of 2.7 K which is used to calculate a certain, constant wavelength for calibration purposes. Relying on this constant temperature calibration keeps the signal intensity values in check. It also corrects for the contribution of the diode's constant temperature since this thermal emission produces unwanted noise. The calibration details of the signal are discussed further in this journal: (5). The resulting plot is a less noisy distribution signal in Janskies (Jy)[1] in Figure 2.4. The calibrated plot also has a lot less signal variation because I have taken the median value of all the integrations for each channel in the frequency range.

---

[1] **Jy** = $(10^{-26} W/m^2/Hz)$ which is non-SI unit of spectral flux density commonly used in radio astronomy.

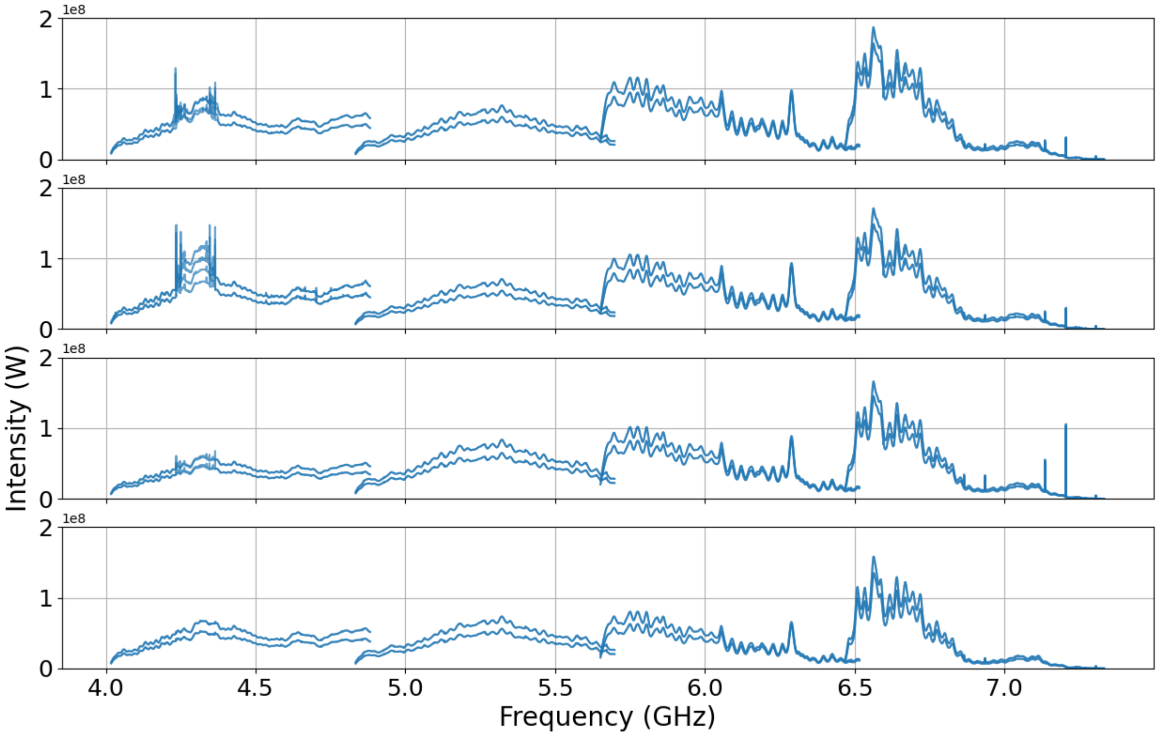


Figure 2.3: Lots of RFI and Noise in this Non-Calibrated Data Plot of only M31 Scans with Intensity of the Source vs. Frequency of Signal

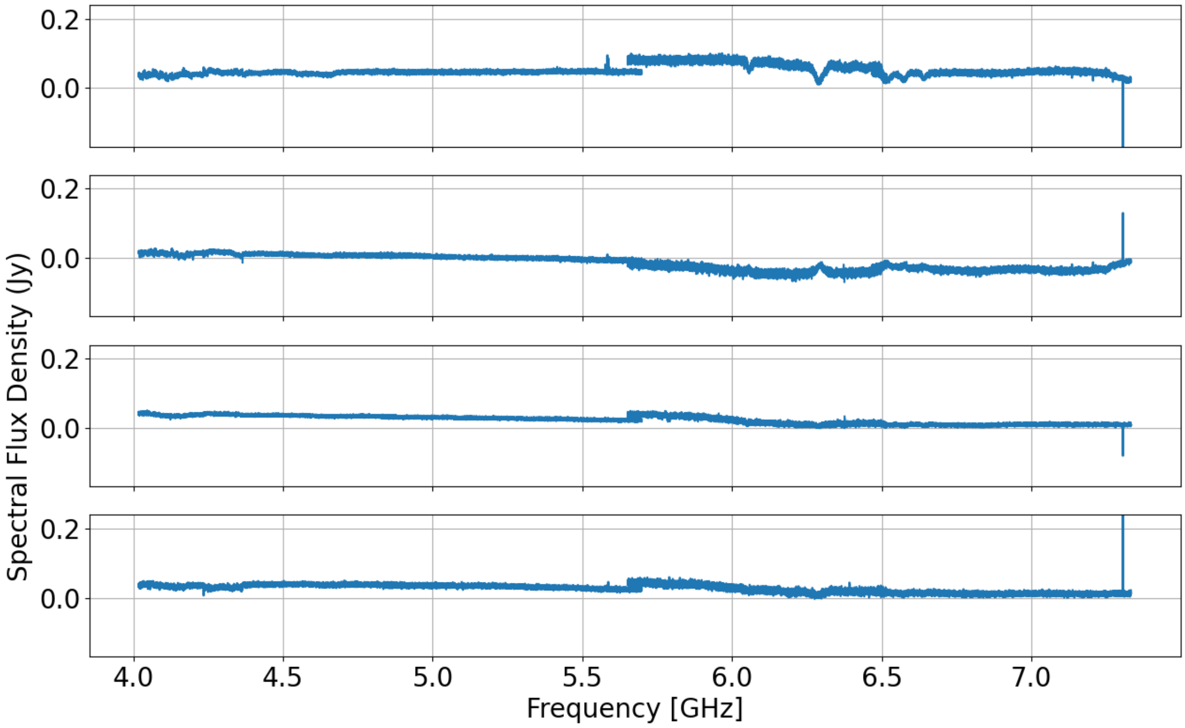


Figure 2.4: Less RFI and Noise in this Calibrated Data Plot of Spectral Flux Density (Jy) vs. Frequency from the Median Calibrated Values of Each Channel

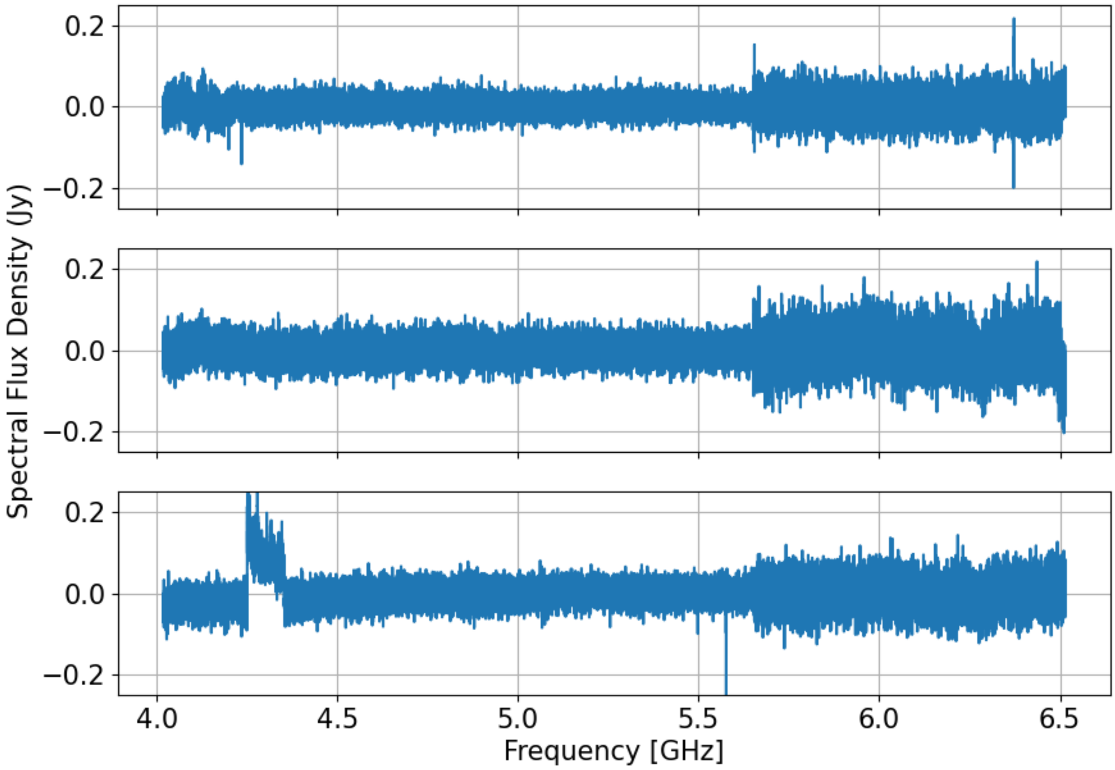


Figure 2.5: Thresh Cut and Linear Fit Calibration of 3 Observations of the C Band Data of Spectral Flux Density (Jy) vs. Frequency with an Error Preventing a Complete Graph

2.2.3 Data Filtration

After the C Band data is calibrated to the necessary reference sources and the right system temperatures of the diode, there are still necessary filtration techniques to be done. I proceed to filter the data with a threshold cut and linear fit function. The threshold function cut out any outliers that would skew the distribution of measurements. After the threshold cut function, the shape of the graph cut out the signal outliers by establishing a threshold of maximum 10 Jy. Keeping the range of data constrained allows for more accurate median values used for the analysis. The linear fit function subtracts unnecessary frames of data, so I can only see the signals coming from the source of interest, in this case, the M31 galaxy. This function takes an integration and fits a line to all the channels of data in that specific integration to subtract a zero point baseline which takes out the sloping in Figure 2.4. The resulting filtration technique reduces some of the noise and RFI and has a constant zeroed baseline, which is shown in Figure 2.5. Overall, these functions reduce the statistical outliers that are unnecessary in my analysis, and more filtration techniques that I did not get to are discussed in Section 3.2.

Chapter 3

Conclusion

3.1 Results

The resulting plot is shown in figure 2.5 with three observations in the A, B, C Bank. The resulting filtration technique reduced some of the noise but mainly reduced RFI. The shape of the continuum in the plot does not have much sloping variation and is very linear across the banks. Keeping the continuum constrained to a certain range helps focus on the actual data that is readout correctly. The linear fit function took out some of the RFI which can be seen as giant spikes in the plot, and subtracted out the off-source, light polluted frames. It also improved the linearity of the plot to make it quite constant which can then be fitted to a linear function to figure out flux density values based on a specific frequency. This result is a good indicator of the methods used throughout this project. I took radio signal data from a GBT FITS file of the Andromeda galaxy; reduced it to only on/off source data, calibrated the scans based off median integration of the channels and constant diode temperature; and used filtering data techniques to get rid of the outliers, such as noise and RFI; and shown the importance of calibrating and filtering data.

3.2 Future Work

If I had more time in the semester or took more than just my spring semester to do my senior thesis, I would have done more data filtering throughout the pipeline. There are a couple bugs I did run into that did not get sorted out in the essence of time. An example is my last plot in Figure 2.5 which did not have all four observations or the complete range of data. When attempting to plot, the first observation and D bank were not accounted for in the linear fit function, which I would have fixed if I had the time. One of the data filtration techniques that I would have run my data through is the Kurtosis flagging algorithm. This function looks at excess Kurtosis[1] and flags any distribution with an absolute Kurtosis of above 3, which is deemed not Gaussian or "normal." Seeking a normal distribution among the thousands of channels in this dataset is important to gather intel on statistical anomalies. Another filtration technique is the smoothing function which is a modified boxcar smoothing algorithm that takes a median value of ten channels and subtracts that value from each integration to provide a smoother baseline of your values (6). Finally, there is a Gaussian width test that fits a Gaussian to the spectrum at a particular channel. The time domain takes the width of that channel and compares its width to a spectrum of all the remaining channel widths to acquire a cohesive

---

[1] **Kurtosis** is the statistical measurement of how skewed or asymmetric a data distribution is.

spectrum along all of the data. Doing these filtration techniques would have smoothed my plots more effectively and, ultimately, would improve the data visualization results.

3.3 What I learned & Why I Want to Learn More

I have learned a lot about research and what it takes to understand our world through the astronomical lens. Throughout this senior spring semester of my undergraduate career, I have gained a whole new level of understanding with data analysis in Python. I extracted data through a fits file by seeing the multiple extensions with different headers, and chose which columns of headers to look at and analyze. I started by making a simple scatter plot and histogram of the altitude and azimuth of the observations. I then used the reduction function to plot the actual data in the X-band (8-12 GHz) to get practice before doing it with the C-band (4-8 GHz) as seen in figure 2.2. I took the time-domain median of the on/off-source scans and calibrated them to where I integrated the on-source scans only, so the plots would show the source light intensity with nothing around it contributing. I learned how different filtration functions affected the data and the spectrum shapes of the plots. There are also so many aspects to radio astronomy that I have never known until now. The techniques like on-off pointing scans to diode temperature calibrations were very intriguing. Just by doing this project, I discovered the intricacies of radio astronomy, and how different it is than optical astronomy. The most important takeaway I have from this semester is not only understanding the science of radio astronomy, but actually doing it.

I want to learn more about data analysis and methods in observational astronomy because I plan to continue these studies in graduate school. I think it is very important to comprehend the data that is acquired in a telescope observing session. Getting the practice now with data visualization and calibration methods is the key to making better observations in order to produce better results and have a more purposeful, influential research conclusion. I want to know more about how the universe works, and this project aided in that request. Eventually, I aim to go above and beyond to make real scientific discoveries. Finding evidence of dark matter through axion-photon conversions started the motivation for this project. One day, I want to reach that goal of making a scientific breakthrough by sharing it with the scientific community and the world.

Bibliography

[1] A. J. Millar, S. Baum, M. Lawson, and M. C. D. Marsh, “Axion-photon conversion in strongly magnetised plasmas.” https://arxiv.org/abs/2107.07399, Nov 2021.

[2] R. Garwood and Brandt, “Gbt vegas fits file specification.” https://www.gb.nrao.edu/GBT/MC/doc/dataproc/gbtVEGASFits/gbtVEGASFits.pdf, Dec 2017.

[3] J. Magnum, “Observing modes used in radio astronomy.” https://safe.nrao.edu/wiki/pub/Main/RadioTutorial/radio-obs-modes.pdf, Apr 2006.

[4] “Calibrating the flux density scale.” https://science.nrao.edu/facilities/vla/docs/manuals/oss/performance/fdscale, Jan 2019.

[5] J. Braatz, “Calibration of gbt spectral line data in gbtidl v2.1.” https://www.gb.nrao.edu/GBT/DA/gbtidl/gbtidl_calibration.pdf, Oct 2009.

[6] L. Walters, J. Shroyer, M. Edenton, P. Agrawal, T. Edwards, B. Johnson, B. J. Kavanagh, D. J. E. Marsh, S. Ransom, and L. Visinelli, ”Searching for Axions in the Andromeda Galaxy with the Green Bank Telescope.” PRL, in preparation.