Astro 121: Using the Spiral Arms as Cosmic Laboratory to Understand Hydrogen in the Universe

Neil Pichay
Partner: Elma Chuang Mia Birkelund
Professor: Aaron Parsons
University of California Berkeley

(Dated: July 13, 2024)

We present the data acquisition and data analysis techniques of observing the structure of spiral arms in the Milky Way Galaxy. This lab applies a variety of skills and knowledge from fitting software, coordinate transformations, and using horns for capturing radio frequencies. This lab report strongly emphasizes data analysis techniques for understanding and characterising the HI 21cm line. Fitting plays an important role in measuring physical quantities like brightness temperature, Doppler shift, and Doppler Width. We also attempt to characterize the transmission lines in our instruments. This lab provided us with a deep understanding of how to probe hydrogen in our galaxy.

I. INTRODUCTION

The goal of this lab to understand astronomical timekeeping, coordinate transformations, transmission lines, and observe HI (hydrogen-one, H-one) emission in the Cassiopeia constellation using a variety of data analysis techniques. This lab report carefully details data acquisition methods for observing Cassiopeia. The figures presented in this lab applies several fitting and error propagation techniques that allow us to understand the kinematics of several hydrogen (H) clouds.

II. BACKGROUND

A. 21cm Line

Neutral hydrogen (H) describes a proton and an electron orbiting in the lowest ground state. The internal structure dictates energy transitions in the electronic ground state due to hyper-fine transitions caused by electron spin flips. A neutral H atom transitioning from spin state $\uparrow\uparrow$ to $\uparrow\downarrow$ emits about 5.9x10^{-6}, which represents emission of 1420 MHz/21 cm radiation. This spin state transitions occurs on the order of 10^6 years. While this may seem like a long time, the abundance of hydrogen in our universe - especially in our own galaxy - allows us to observe this transition.

B. Time

Timekeeping in astronomy is important for understanding observations. The location of an astronomical object crucially depends on both location and time. It is not as easy as keeping track of the date and time like how we do in our daily lives. Astronomers use a variety of timekeeping systems (Local Sidereal Time, Julian Date, Unix time, etc.) and translate them from one to the other. Our data is timekept with `time.time()` such that all of our observation are in Unix time, which is the time kept from 00:00:00 UTC on January 1, 1970. Depending on the program being used, this Unix time is translated to other units to specify azimuthal coordinates, barycentric radial velocity, etc..

C. Astronomical Coordinate Transformations

To keep track of the positions of astronomical objects, astronomers use a various coordinate systems. We have galactic coordinate systems, systems based off an observer's current time and location, and some centered on Greenwich, England (boo). To switch to and from coordinate systems is a matter of matrix transformations that determine specific coordinates for the same object. Understanding coordinate transformations is important for this lab because we need to know where Cassiopeia is located from the roof of New Campbell Hall at any given time.

D. Transmission Lines

The characterize the transmission lines in our instruments, we first measured the speed of light with a 26ft long cable. Using square wave reflectometry seen in Figure 1 and measuring the time delay of the reflection we calculated of cable speed of light to be 0.65c ± 0.5. We applied the same reflectometry to measure the cable length to be about 77ft; we immediately recognize that this measurement is likely far away from the true measurement and we suspect that our inability to use a high RF signal affected our measurements.

Our group also made an attempt at properly characterizing the amplifiers, filters, and cable loss but weren't able to make any substantial conclusions. However, it was interesting to note how the length of the transmis-

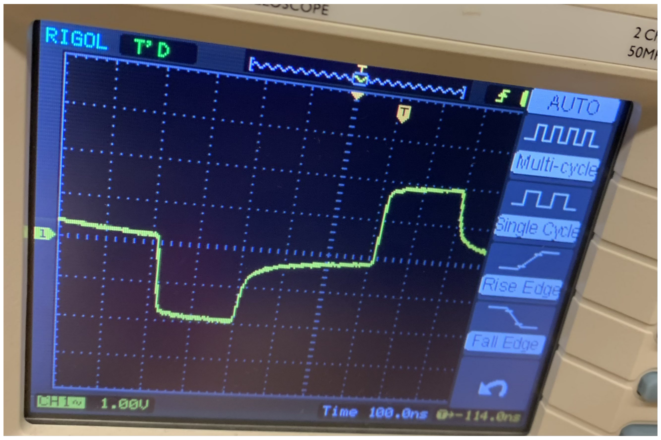

FIG. 1. Square wave inputted to an oscilloscope. Without proper termination, the square wave is reflected with a time decay corresponding to the speed of light in the cable and the length of the cable

sion line had a strong affect on the power spectra such that a drop of 20% power when comparing a direct RF signal into the SDR and the signal being sent through the signal chain. We were also unsure how to properly draw the signal chain, so we referred to the drawing of the signal chain in the Github.

III. CHARACTERIZING THE HI LINE

This section details all methods for calibrating, acquiring, and analyzing our dataset to characterize the observed HI line. The reproducibility of this science strongly relies on carefully following these methods.

A. Calibration

The first step in understanding how a signal is processed by the horn is observing how the signal chain reacts to an inputted signal. We can characterize the signal chain by injecting a signal of known frequency and properly setting our LO to mix our signal into a frequency we can capture. Lab 1 taught us that when an observed RF signal is mixed with a SSB mixer of set frequency (LO), the output signal is the difference between the RF and LO. For example, an LO of 1420 MHz set in the SDR would capture an HI signal at 0.405 MHz.

After proper signal injection, the next step is to take power spectra used for temperature calibration. In order to calibrate the temperature of the observed object, we have to compare how much power comes out of other objects of known temperatures.

Figure 2 shows the power spectra for S_{cal} (calibration) and S_{cold}. S_{cal} is the power spectra of a known, luminous blackbody signal; for this lab the blackbody signal was ourselves. To acquire S_{cal}, a few of our group members + friends stood in front of the horn and covered at least 95% of the aperture. The S_{cal} power spectra in green is mostly smooth except for unwanted resonances at various frequencies possibly due to electronic and station RF interference. S_{cold} is the power spectra with the horn pointed up at the cold sky. We initially had trouble with capturing S_{cold} without the presence of H, so the S_{cold} observation used in our was sure to avoid the galaxy at zenith. S_{cold} is important for lab because it provides us with a baseline power spectra of the sky so that HI observations can be properly isolated from external RF-emitting sources. It is evident that S_{cal} is more powerful (i.e. luminous) than S_{cold} because hotter blackbodies emit more radiation at all frequencies.

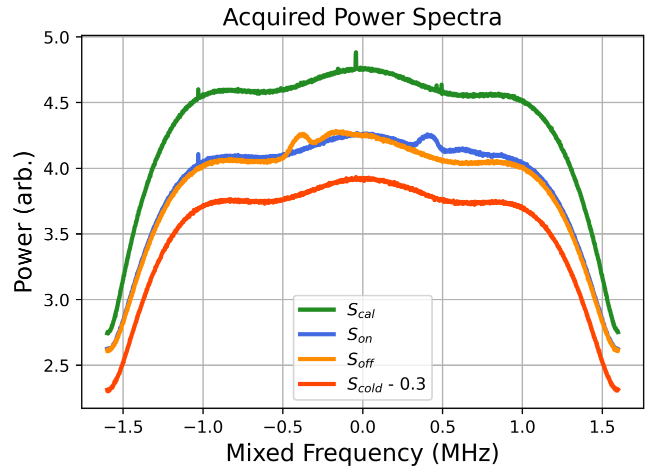

FIG. 2. The acquired power spectra of S_{cal} (LO=1420 MHz), S_{on} (LO=1420 MHz), S_{off} (LO=1420.8116 MHz), and S_{cold} (LO=1420 MHz). The entire data set was taken with 10000 blocks and 2048 samples per block.

B. Observing HI Regions

Once calibration data taking is finished, we can finally observe the HI line. Cassiopeia is located at galactic coordinates l = 120° and b = 0°. To convert from galactic to azimuthal coordinates, we used time and location-dependent `astropy.coordinates` to provide us with a set of coordinates relative to cardinal directions. We also referenced the very useful planetarium software programs `Sky Guide` and `Stellarium` to precisely point the horn.

Figure 2 shows the mean power spectra for S_{on} and S_{off} plotted on top of one another. To reduce the error in all of our power spectra, we took the mean of 10000 blocks (2048 samples per block). S_{on} is taken at LO = 1420 MHz and S_{off} is taken at LO = 1420.8116 MHz. S_{on} and S_{off} compared to S_{cal} and S_{cold} show a similar overall shape but reveal one distinguishable feature: this is HI! Figure 3 plots S_{on} and S_{off} with the corrected frequencies to take a closer look at the HI line of both power spectra. Immediately we notice the appearance of multi-

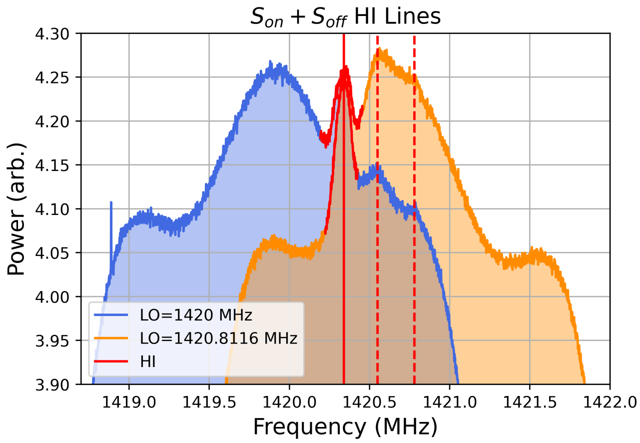

FIG. 3. The S_{on} and S_{off} power spectra corrected with their appropriate frequencies. The HI lines on both of the power spectra do not look the same because frequency switching changes the overall filter response in the range where HI is observed.

ple HI lines; our data analysis will uncover the nature of these lines. Additionally, the amount of power that each of the HI lines emits isn't the same: if we are observing the same HI region, shouldn't these emission lines be the same as well? The next section will explain the purpose of taking S_{on}+S_{off} and the differences between them.

C. Data Analysis

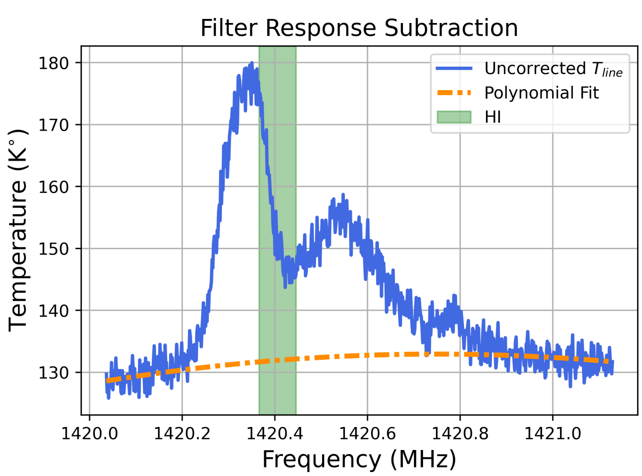

FIG. 4. Temperature profile of HI line uncorrected from instrumental baseline filter response. This filter response is modelled by a 2nd order polynomial model (in orange) that is subtracted from the temperature profile.

To calibrate the temperature of the signal, we first need to choose a region to calibrate from. Throughout my analysis I used the second half of S_{on} and frequency switch to S_{off} to compute S_{line}, where:

$$S_{line} = \frac{S_{on}}{S_{off}} \tag{1}$$

S_{line} is a normalization of the emitting HI region in S_{on} to a non-emitting HI region in S_{off}. The normalization of S_{on} can then be converted into temperature space with the help of G:

$$G = \frac{T_{sys,cal} - T_{sys,cold}}{\sum(S_{cal} - S_{cold})} \sum S_{cold} \tag{2}$$

where $T_{sys,cal} = 300K^{\circ}$ and $T_{sys,cal} = 10K^{\circ}$. This G-factor compares the temperature and power output of two blackbodies to determine the temperature (T_{line})of a normalized emitting HI region in frequency domain.

$$T_{line} = GS_{line} \tag{3}$$

The T_{line} of our observation is shown in Figure 4. Although T_{line} more clearly distinguishes the HI lines, the goal is to eliminate sources of thermal contamination such that we only observe emission from the Cassiopeia. In the previous section we saw that the emission lines from S_{on} and S_{off} had different amounts of power. This is because of the instrument's filter response in the signal chain that hosts a frequency-dependent sensitivity. We account for this baseline filter response by treating the baseline of T_{line} as second-order polynomial model (P) and subtracting it from T_{line}.

$$P(\nu, A, B, C, I, J) = A(\nu - i)^2 + B(\nu - j) + C \tag{4}$$

Figure 4 shows the baseline polynomial model in orange and the resulting corrected T_{line} in Figure 5. With the baseline filter response at around $0K^{\circ}$, the fun can begin!

Figure 5 shows the corrected T_{line} fitted with triple Gaussian model (G) represented by the function:

$$G(\nu, A, \mu, \sigma) = \sum_{i=1}^{3} A_i \exp\left(\frac{-(\nu - \mu_i)^2}{2\sigma_i^2}\right) \tag{5}$$

The motivation for fitting a triple Gaussian model was due to the appearance of two emission lines and one small line around 1420.7 MHz that was pointed out during presentations. A Gaussian is a good model because its parameters (A, μ, σ) allow us to measure physical characteristics of the H clouds in the later section. Now we need to convert to velocity space.

One important concept is that the earth orbits around the sun, and on top of that the sun moves relative to Cassiopeia. This motion Doppler shifts the frequencies of incoming light and makes kinematic measurements of the HI clouds less accurate. This is further visualized in Figure 6, where V_{earth} is the tangential velocity of the earth, V_{sun} is the velocity of

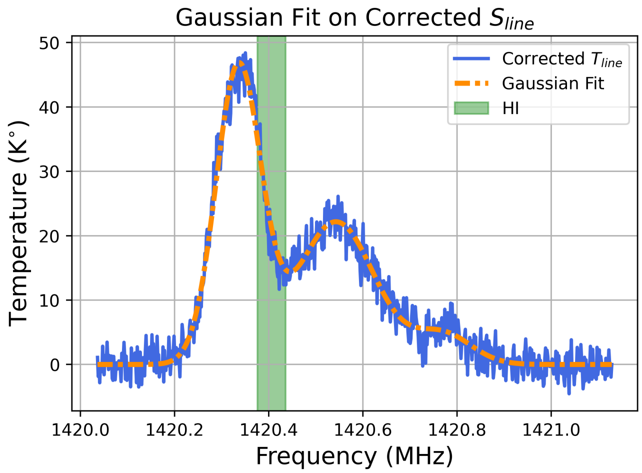

FIG. 5. Corrected temperature profile fitted with a triple Gaussian function. The green region corresponds to LSR frequency HI; we observe three different emission amplitudes.

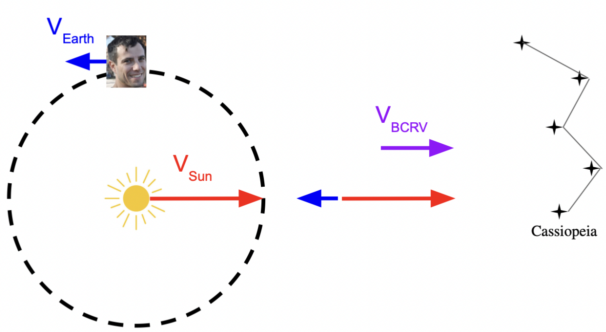

FIG. 6. Diagram representing how to calculate the barycentric radial velocity (BCRV) relative to Cassiopeia. The observer's Doppler motion is then corrected in frequency space.

the sun, and V_{BCRV} is the sum of both velocities in the direction of Cassiopeia called barycentric radial velocity. To account for this motion, we calculated the earth's Doppler motion relative to Cassiopeia by using `ugradio.doppler.get_projected_velocity` which outputs the V_{BCRV}. The frequency axis is shifted by the following:

$$f_0 = f(1 - \frac{V_{BCRV}}{c})^{-1} \tag{6}$$

where f_0 is the frequency in the local standard of rest (LSR), f is the observed frequency, and c is the speed of light.

To convert from LSR frequency to velocity I used the equation:

TABLE I. Physical Characters of Three H Clouds

Cloud	#1	#2	#3
Temp. (K°):	45.92 ± 0.38	22.16 ± 0.30	4.77 ± 0.32
Dop. shift ($\frac{\text{km}}{\text{s}}$):	14.79 ± 0.12	-28.90 ± 0.35	-76.61 ± 1.51
Dop. width ($\frac{\text{km}}{\text{s}}$):	23.81 ± 0.28	40.98 ± 1.32	30.52 ± 3.31

$$\frac{v}{c} = -\frac{\Delta f}{f_0} = \frac{f_0 - f}{f_0} = 1 - \frac{f}{f_0} \tag{7}$$

$$v = c(1 - \frac{f}{f_0}) = c(1 - \frac{f_{LSR}}{f_{HI}}) \tag{8}$$

$$\tag{9}$$

where f_{LSR} is the LSR frequency and f_{HI} is 1420.4085 MHz. Figure 7 is the resulting velocity profile fitted with another triple Gaussian model. Another reason why a triple Gaussian model was chosen was because the residual for a double Gaussian model had feature that stood out around -70 $\frac{km}{s}$. Since the residual for the triple Gaussian model looks noise-like, any statistical variation in the model can be accounted as noise in the velocity profile.

The error analysis in computing the residual is also accompanied by a calculation of the reduced chi-squared (χ_r^2). The reduced chi-squared is a tool for measuring the uncertainty of a model and is represented by the formula:

$$\chi_r^2 = \frac{1}{N - M} \sum_i \frac{(f(v_i) - G_i)^2}{\sigma_i^2} \tag{10}$$

where N is the number of data points, M is the number of open parameters in our models (14), $f(v_i$ is the data, G_i is the model, and σ_i is the noise. For our calculation the standard deviation of the residual was used as noise. To maximize the likelihood of our fit matching with our data, χ_r^2 needs to be minimized to equal about 1. The χ_r^2 measured with our models was ≈1.02.

Since our model matches quite well with our data, it's time to measure physical parameters! Table I lists the physical properties of the three observed H clouds: brightness temperature (amplitude), Doppler shift (μ), and Doppler width (≈2.355σ) caused by the Doppler motion of H from within the cloud.

D. Discussion of Velocity Profile

The velocity profile in Figure 7 is fitted with three independent Gaussian that show evidence of the spiral arms. Being said, I don't quite understand how there can be redshifted and blueshifted H clouds.

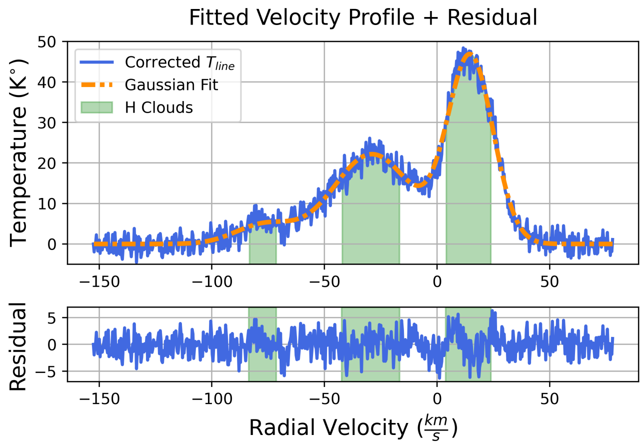

FIG. 7. The temperature-velocity profile in the Local Standard of Rest (LSR) fitted with a triple Gaussian function. The three independent Gaussian curves represent three hydrogen clouds observed with different kinematics and temperatures. The noise-like behavior of the residual indicates the triple Gaussian is a good fit for the profile.

An interpretation of this is that our group observed three spiral arms. However if we pointed the horn at only one spot in the sky, how can these spiral arms be moving both to and away from us? Are there spiral arms that move faster and slower with respect to one another? Is this some kind of galactic retrograde effect? This is why I refer to each of the three Gaussian profiles as an "H cloud" and not "spiral arm" because I am unsure about how they should be properly addressed. Overall the velocity profile is quite fascinating to me and I wish I knew the answer to these questions.

IV. CONCLUSIONS

My partners and I learned a lot taking on-sky radio data with completing this lab. I thought it was very interesting that we could understand the structure of our own galaxy with a horn on a roof. We spent a lot of time analyzing our data and got a lot more comfortable with error analysis. Despite getting soaked in rainwater to take data, I definitely had a lot of fun with this lab and I'm excited to further our understanding of the universe through radio.