Visualizing the Molecular Structure of Protoplanetary Disks

Astra Võ

May 2025

Instructor & Advisor: Dr. Ilsedore Cleeves

Department of Astronomy

University of Virginia

This thesis is submitted in partial completion of the requirements of the BS Astronomy-Physics major.

Abstract

Protoplanetary disks, the birthplace of planets, exhibit fascinating and diverse molecular structures. However, visualizing these structures remains difficult due to the conventional plotting of singular, isolated molecules. In this thesis, I present a Python-based visualization script that is designed to interpolate and overlay molecular data from ALMA observations, specifically with applications to the files from ALMA's DECO program. This script interpolates both the spatial and spectral grids of given FITS files, and allows the user to customize how they would like to plot the interpolated files atop one another. Demonstration of this script is through the application to HP Tau, a young protostar with an interesting disk structure in Taurus. The script reveals interesting molecular streams in carbon-13 based carbon monoxide (^{13}CO), large structures in abundant carbon-12 based carbon monoxide (^{12}CO), and distinct structures with hydrogen cyanide (HCN) and carbon monosulfide (CS). These structures are consistent with light scattering as seen in SPHERE images. The Python script used is publicly available via a public GitHub repository. This work lies the foundation for future customization and enhancement, including more complex molecular overlays and structure visualization of multiple and more complex disks.

Contents

1 Introduction

Much about protoplanetary disks is still an enigma in the field of astronomy. They are the beginnings of planetary systems, as the name suggests, but also contain young stars shortly after their birth before the dust clears. The reason that knowledge is limited in this field is that the dust surrounding these new stars is extremely opaque in optical wavelengths—optical telescopes like the Hubble Space Telescope are unable to penetrate into the dust clouds to see the structure of these young stars and forming planets. A telescope that operates in the infrared or radio spectrum, however, would indeed be able to penetrate through the dark, dusty clouds that surround these protostars.

1.1 ALMA

The Atacama Large Millimeter/sub-millimeter Array (ALMA) is one of the largest and the most powerful radio telescope on Earth. It has been an invaluable instrument to the protoplanetary disk and planet formation fields. ALMA consists of an *array* of telescopes, as the name suggests. Aligning the array to a single object in the sky (especially an object as small as a disk) is an extremely difficult task and requires a high degree of precision. The large number of antennae acting together allow the array to take data on the sky distribution of light from an astrophysical source, which can be reconstructed into high resolution images. By combining the signal from all antennae, interferometers can provide high spatial resolution information as if the telescopes were a single much larger dish.

In addition to high spatial resolution, ALMA also provides high spectral resolution data. Spectral resolution is important, since different molecules emit at unique wavelengths of light. For example, carbon monoxide (CO) emits at around 230 GHz, while carbon monosulfide (CS) emits at around 290 GHz. The emission from molecules is furthermore Doppler shifted based on the motion of the gas, which can include thermal motions, turbulent motions, and orbital motions, for example. Thus, ALMA is very well suited to both imaging and spectrally resolving molecular emission on disk scales.

Disks are heterogeneous with extremely diverse chemical structures. As seen in Law et al. (2021), the structure of disks can vary wildly. One disk (AS 209) might have very vivid HCN emission, while in another (IM Lup), HCN would be considered nondetection. Law et al. (2021) also mentions that, in the Molecules at Planet-forming Scales (MAPS) survey, IM Lup was the only disk to have spiral structure in its continuum, but not in its molecular emissions. AS 209 is more compact than the others, while GM Aur is considered a transitional disk. Disks can have exceptionally and radically different structures from one another, despite being from the same general environment.

ALMA's Disk-Exoplanet C/Onnection (DECO) program is a survey of over 80 disks orbiting low-mass stars. Channel maps (position-velocity intensity graphs) show the variety of structures that these disks have. A channel map is an image of the object at a specific velocity "channel", showing the emission of a single molecule at these specific velocity slices (L. Ricci 2022). In these channel maps, disks may have *butterfly* shapes due to Doppler shift in conjunction with the Keplerian rotation of the rotating disks. An example of this from simulations from L. Ricci (2022) can be found in Figure 1. Since the array can only visualize and image one frequency at a time, it is difficult to be able to visualize exactly how the molecules trace regions and processes in 3D. Conventionally, molecules are plotted either separately on their own or reduced to a two dimensional projection like integrated intensity. In the present thesis, I present a process to overlay channel maps with interpolation over the sky coordinates and frequency coordinates. This routine will enable those working with multi-line data to better visualize their observations. This thesis focuses on the application of this code to the HP Tau star and disk system, described further in Section 1.2. We describe our methods to develop the code in Section 2. We summarize our our findings and future applications in Section 3.

1.2 HP Tau

HP Tau is a young T Tauri star located in Taurus, about 171.2 parsecs away from our solar system. As stated in Garufi et al. (2025), a survey of planet-forming disks in Taurus, it is a large star with a mass of about 1.7 $M_\odot$, nearly as massive as two Suns. The star hosts a compact disk with a radius of approximately 21 AU. Despite being the most massive star referenced in Garufi et al. (2025), it has the lowest mass accretion rate of about 10^{-11} $M_\odot$ per year. It also has the highest free-free emission (electromagnetic radiation of decelerating

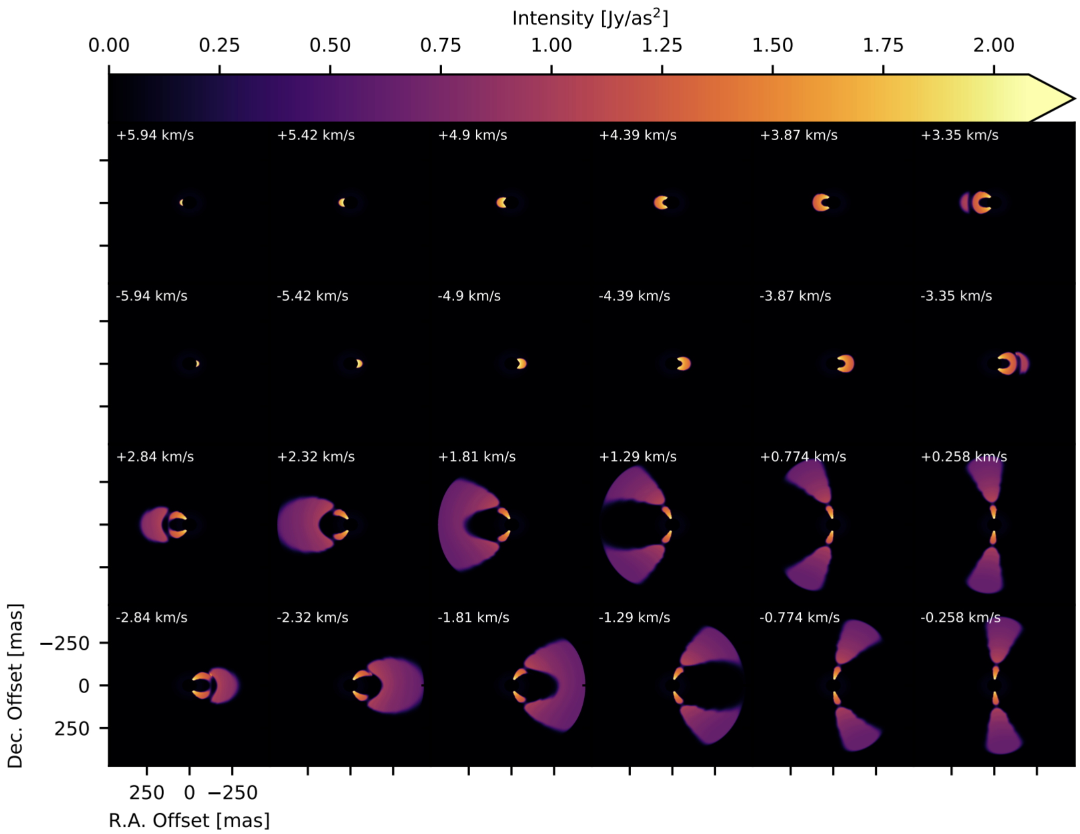

Figure 1: Model synthetic channel maps for the ^{12}CO ($J = 1 - 0$) line emission for the disk+planet system considered in this study. The central channel velocity relative to the rest frame velocity of the system is shown on the top side of each panel. Image and caption credit: L. Ricci (2022).

Figure 2: Hubble Space Telescope image of HP Tau and surrounding area. Image Credit: NASA, ESA, G. Duchene (Universite de Grenoble I); Image Processing: Gladys Kober (NASA/Catholic University of America)

charged particles) of ∼10 mJy while the other stars in this survey have a rate of <1 mJy.

HP Tau is the most anomalous case of the disks surveyed in this paper. This specific disk was selected due to its interesting structure and the surrounding reflection nebula, as seen in Figure 2, and we were curious about the molecular structure.

ALMA's DECO data is condensed into Flexible Image Transport System (FITS) cubes, 3D cubes of data, with RA and Dec axes and frequency as the third axis. The goal of this thesis project is to create a functioning, streamlined Python script that allows for customization of FITS cubes (in the case that the user needs to normalize dimensions or to compress files), overlaying molecules on a channel per channel basis. This process is described in more depth in Section 2.

Overlaying the molecules can allow a better understanding of the chemical structure of disks with application to HP Tau. The scripts will be able to interpolate all given FITS files to the same dimensions (if the files are not already normalized) and then overlay the interpolated files atop each other according to the user's preferences. The user can customize the overlay by changing the color maps that each individual molecule would be plotted with, the intensity (continuum minimum and maximum) that each molecule is plotted under, the order to plot the files, the number of channels to plot, the velocity to center the plots on, and the arcsecond window of the plots.

2 Method

The full DECO files contain images for a select number of molecules. Each molecule's channel maps were analyzed by eye for any structure within the noise. Many channel maps were empty, and thus those molecules were ignored. The files, 3D FITS cubes of emissions of carbon dioxide (both the most abundant isotopologue $^{12}C^{16}O$ and carbon-13 based $^{13}C^{16}O$), carbon monosulfide (CS), and two hyperfine lines of hydrogen cyanide (HCN), were downloaded from the DECO database. These files and images were cleaned beforehand by Dr. Charles Law using CASA's `tclean` algorithm. These specific molecules were chosen based off visible emissions from the channel maps. Though there were three hyperfine HCN lines, only the first and third lines were used as the second line emissions were too faint and not positively identified.

The files, though cleaned, did not have the same dimensional values. Both the frequency step and pixel size values were not normalized and had to be interpolated in order to uniformly overlay the images. The hyperfine images specifically had finer resolution. The frequency values were interpolated using `SciPy`'s 1D linear interpolation package, while the pixel size values were interpolated using `SciPy`'s `RegularGridInterpolator` function. These specific interpolation types were chosen because they were the fastest. Other interpolation modes resulted in poor `RAM` usage (slowing down my laptop significantly to the point where it would crash and restart) and long processing times. This decision was further motivated by issues encountered in using a remote desktop on one of the group's local servers, but provides the added benefit that anyone will be able to combine these fits cubes on their personal computer, rather than requiring a large machine to read these large datasets into memory. Instead of choosing arbitrary values for the dimensions, the script reads an inputted file and sets the pixel size, interpolating all other files. There was no reasonable way to have the reference file be pre-selected before running without having two separate script files, so the inputted reference file name is the assumed name of the file after it has been velocity/frequency interpolated. This is nominally the file with the best resolution, though the user can choose any file they wish to interpolate on, as long as they know what the file will be named after the initial interpolation.

The overlay script takes inspiration from UVA alumna Claire Thilenius' CASA guide script for plotting (Thilenius 2024). Though the plotting function is heavily modified, the function that reads in the values of the FITS cubes (e.g. `readingrainbowCASA` and `RADEC`), is still largely the same. The files had varying minimum and maximum values for the contour. Through trial and error (by plotting ∼5 or so graphs per output image per molecule), minimum and maximum values that reduced the most background noise and didn't compress the values into indiscernible blobs were chosen. Since these values were chosen by eye, they may not be the most accurate, and some structure may have been lost if the values were too close to the background noise.

3 Results & Discussion

3.1 HP Tau's Molecular Environment

With the molecules overlapped and continuum plotted on top (as seen in Figure 3), we can see that though there is some emission and structure near the center (in mainly CS and HCN), the ^{12}CO emission is much larger and has more structure. ^{13}CO dominated the more outer regions in what appears to be some high-speed stream from 8.21 km/s to 9.24 km/s. We can see that CO is mostly bright around the 8.1 km/s which the channels are centered on, while the other molecules emit more at higher velocities. Plots of ^{12}CO with ^{13}CO can be found in Figure 5. Separately, they can be found in Figure 6 and Figure 7 respectively. Plots of CS with HCN and on its own can be found in Figure 8 and Figure 9 respectively.

As seen in the figures, the emissions at these velocities are concentrated to the lower left of the continuum. We can see with the overlay of SPHERE images (Garufi et al. 2024) in Figure 10 that these emissions are consistent with the scattered light.

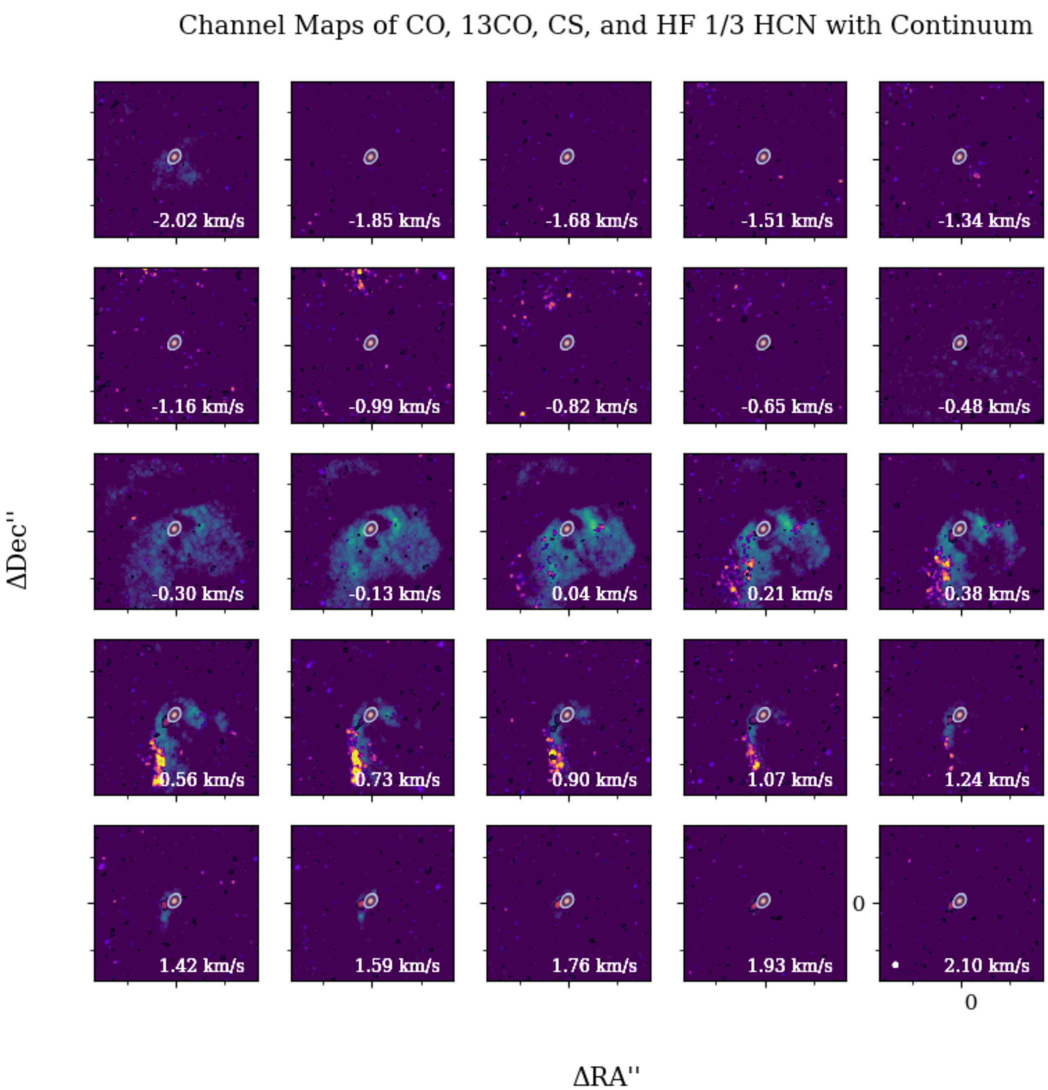

Figure 3: Channel maps of HP Tau centered at 8.1 km/s (all listed velocities given relative to 8.1 km/s) in a 10×10 arcsecond window. The emissions in the `viridis` color palette are ^{12}CO emissions, while the bright `plasma` emissions are ^{13}CO, CS, and HCN. HCN and CS emissions are around the center, while ^{13}CO is the streamer emission. HP Tau's continuum is plotted overtop as contour lines.

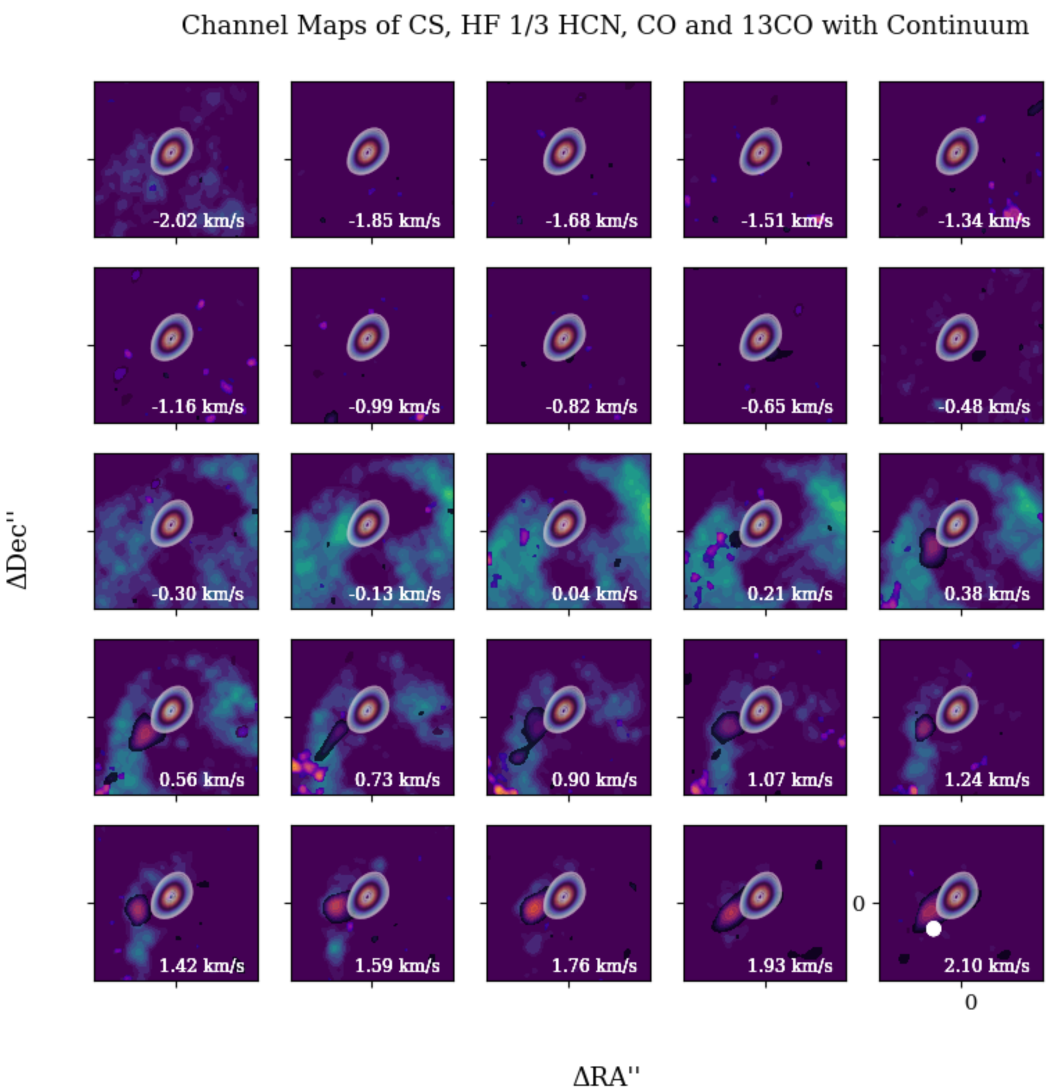

Figure 4: Channel maps of HP Tau centered at 8.1 km/s (all listed velocities given relative to 8.1 km/s), zoomed to a 3 × 3 arcsecond window. The emissions in the `viridis` color palette are ^{12}CO emissions. HCN and CS emissions are around the center (these molecules can be found plotted in together in Figure 8 and with CS isolated in Figure 9). HP Tau's continuum is plotted overtop as contour lines.

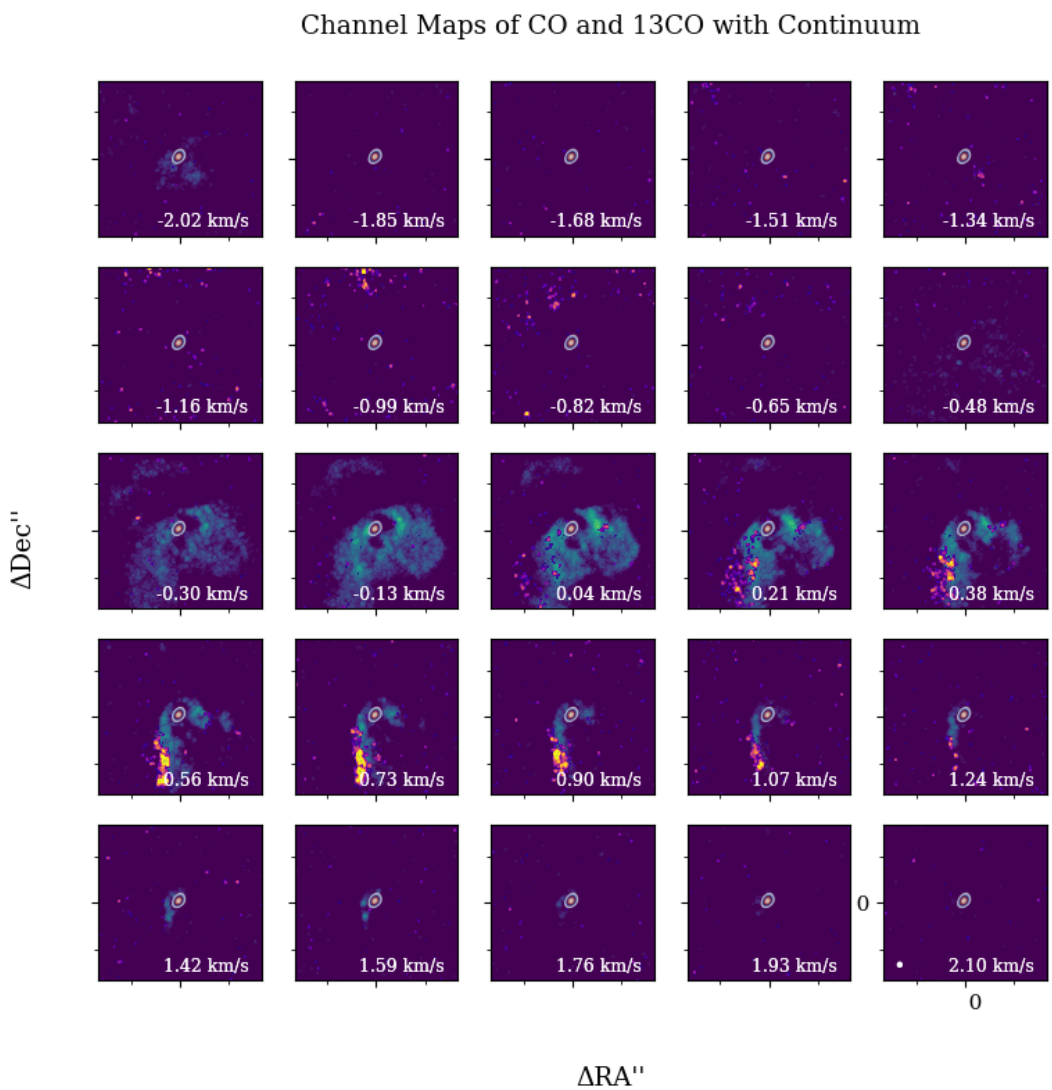

Figure 5: Channel maps of HP Tau centered at 8.1 km/s (all listed velocities given relative to 8.1 km/s) in a 10 $\times$ 10 arcsecond window. The emissions in the `viridis` color palette are ^{12}CO emissions, while the bright `plasma` emissions are ^{13}CO. CS and HCN have been removed from this graph. HP Tau's continuum is plotted overtop as contour lines.

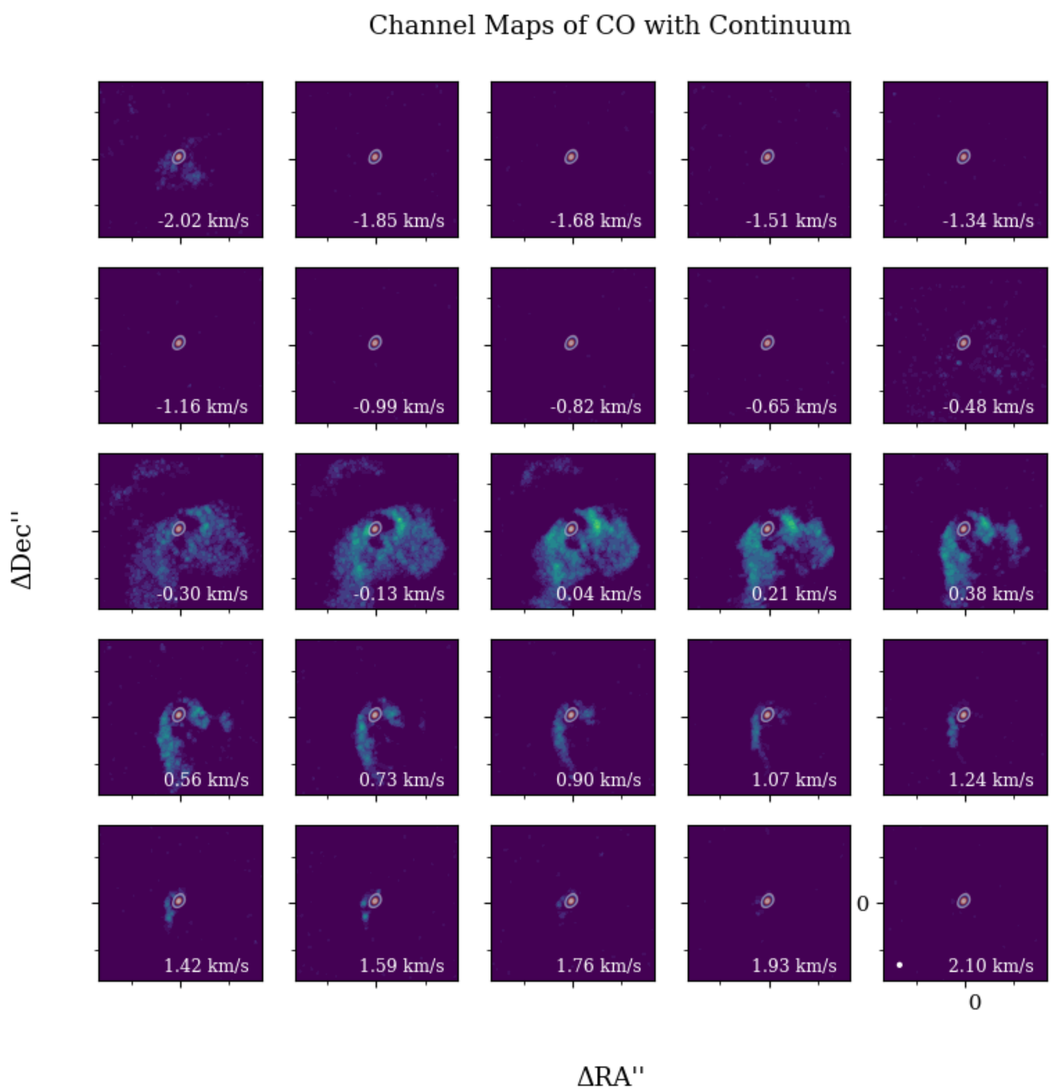

Figure 6: Channel maps of HP Tau centered at 8.1 km/s (all listed velocities given relative to 8.1 km/s) in a 10 × 10 arcsecond window. The emissions in the `viridis` color palette are ^{12}CO emissions. HP Tau's continuum is plotted overtop as contour lines.

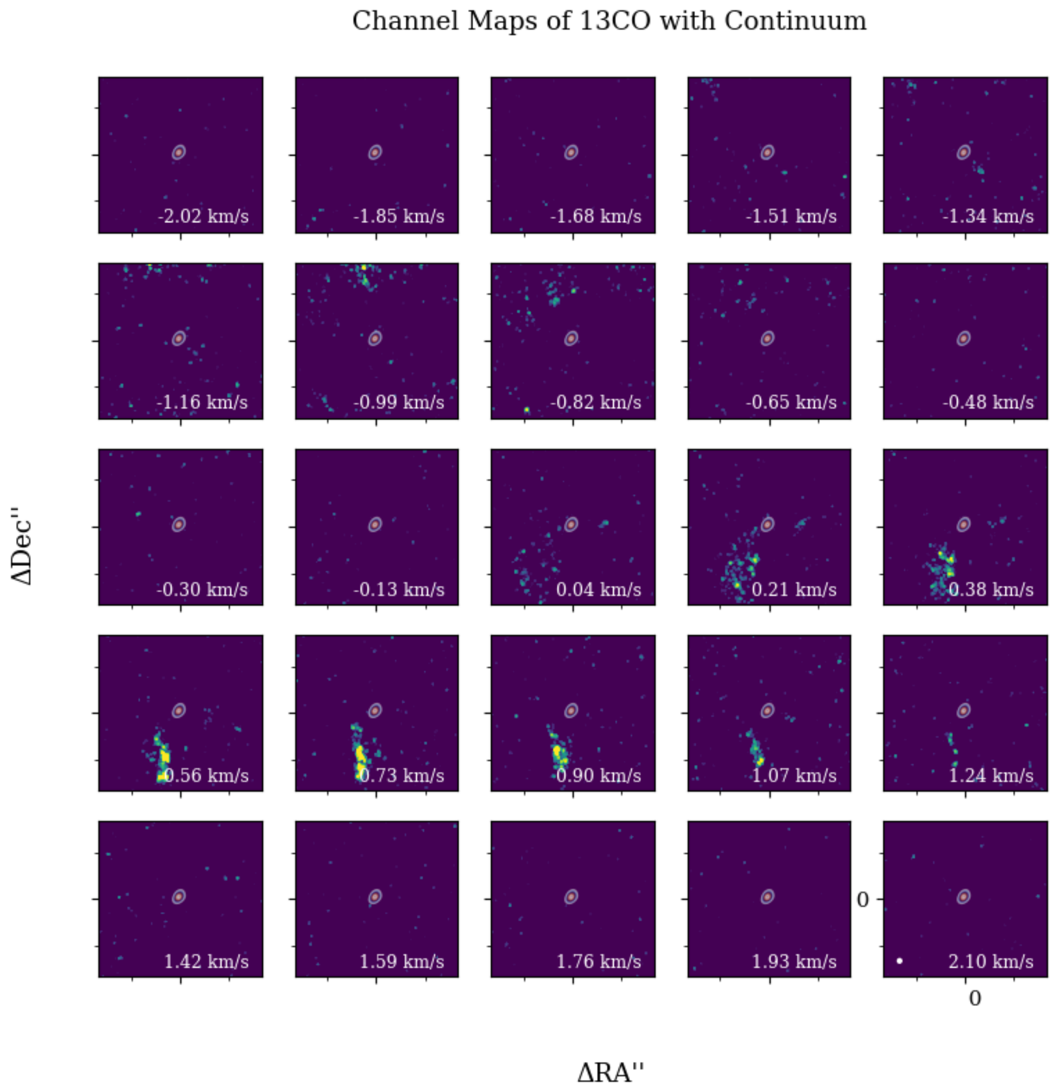

Figure 7: Channel maps of HP Tau centered at 8.1 km/s (all listed velocities given relative to 8.1 km/s) in a 10×10 arcsecond window. The emissions in the `viridis` color palette are ^{13}CO emissions. HP Tau's continuum is plotted overtop as contour lines.

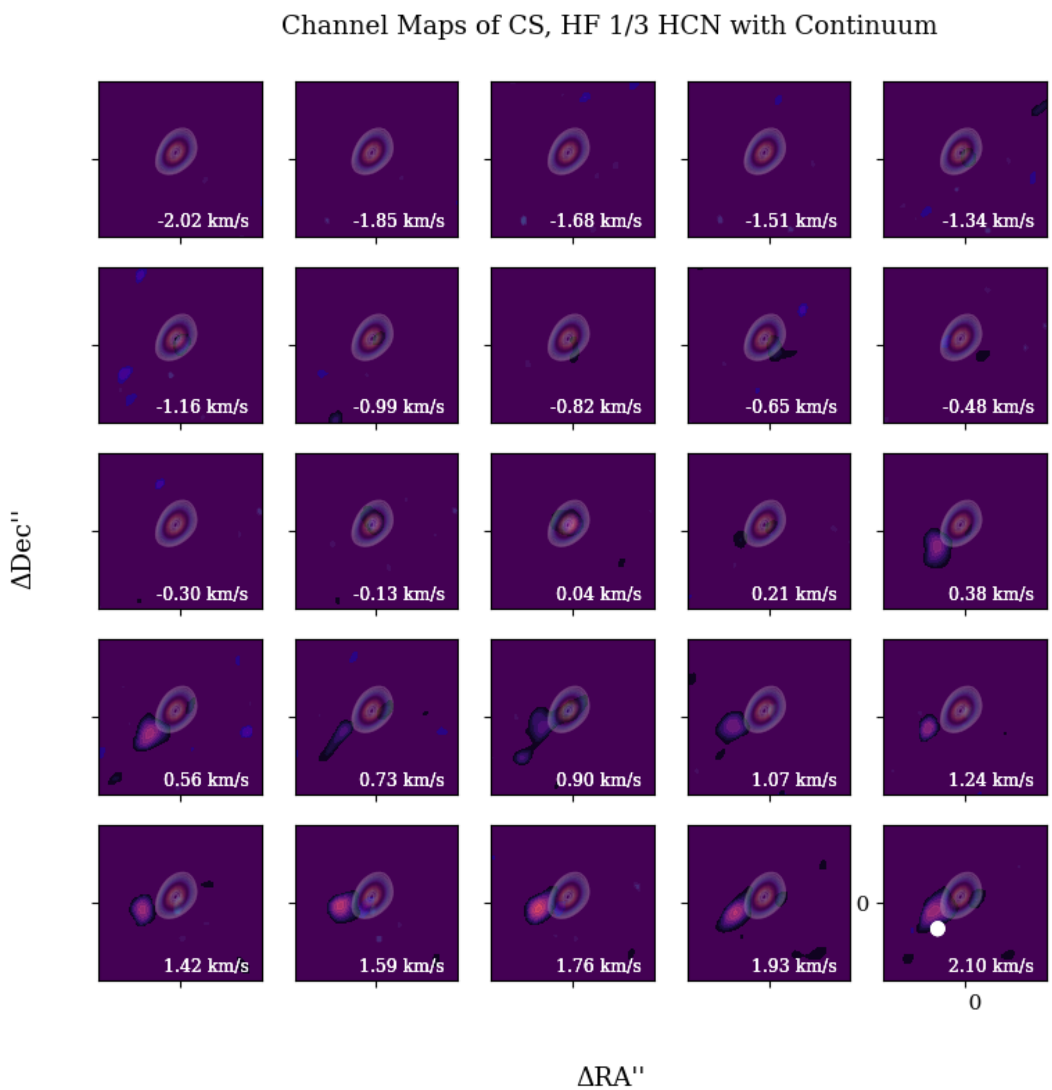

Figure 8: Channel maps of HP Tau centered at 8.1 km/s (all listed velocities given relative to 8.1 km/s) in a 3 × 3 arcsecond window. HCN is the bright emission. HP Tau's continuum is plotted overtop as transparent contour lines.

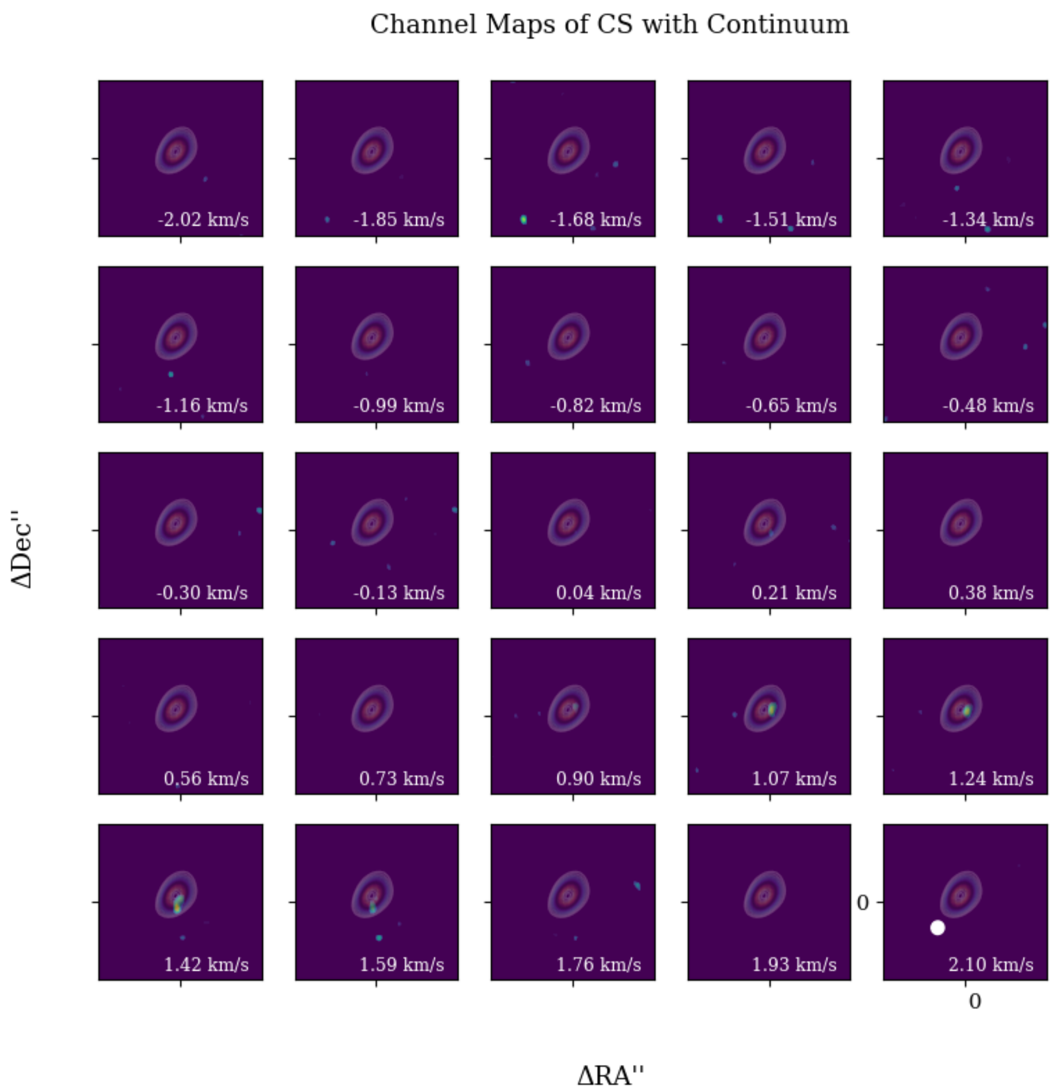

Figure 9: Channel maps of HP Tau centered at 8.1 km/s (all listed velocities given relative to 8.1 km/s) in a 3 × 3 arcsecond window. CS has very faint emissions near the center from about 1–1.6 km/s. HP Tau's continuum is plotted overtop as transparent contour lines.

Channel Maps Overlayed with SPHERE Images

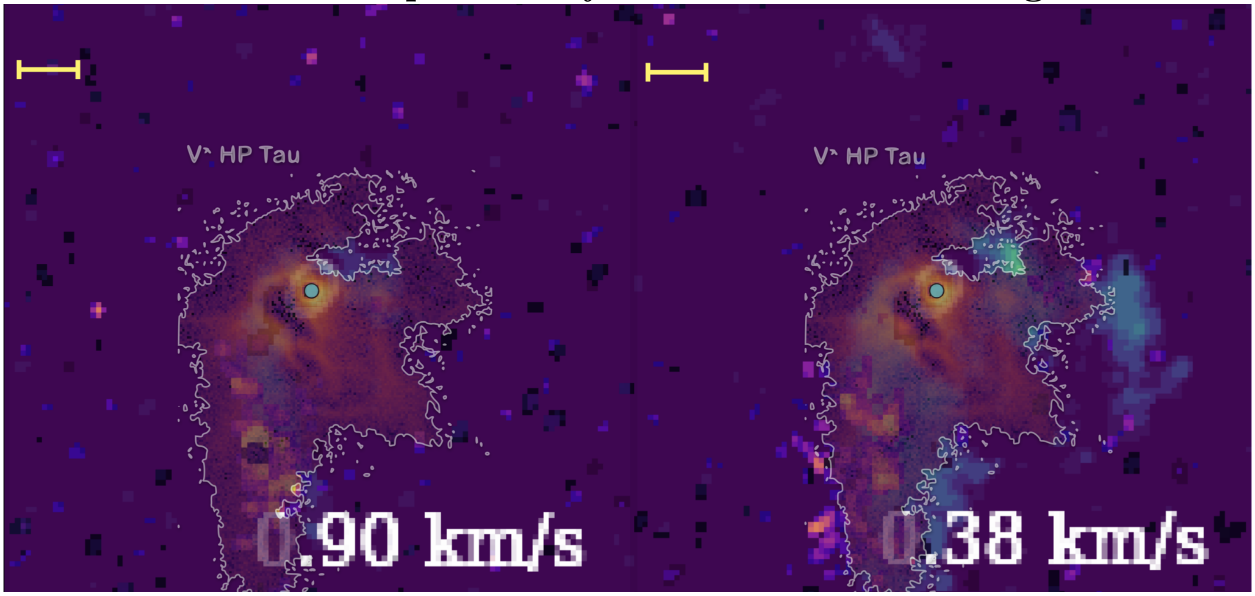

Figure 10: SPHERE images overlayed atop channel maps of ^{12}CO, ^{13}CO, CS, and HCN at 0.90 km/s and 0.38 km/s relative to the source velocity of 8.1 km/s. at 55% opaque. Yellow line indicates the length of 1 arcsecond. Streams of ALMA data are consistent with structure of SPHERE data.

3.2 Python Scripts

The final product contains two python scripts: `data_interpolate.py` and `data_overlay.py`. The first script (interpolate) interpolates both the frequency steps and pixel size of given FITS files in the given directory. The second script (overlay) allows the user to enter which files they wish to overlay and which file will be the reference file, which will also be the bottom-most image of the overlay. The script also overlays the continuum file overtop all channels. The scripts can be found in this public GitHub repository: `https://github.com/nyf5gb/pythondatamanipscripts`. The scripts are available publicly for access at any time.

The hope for this project is that future radio astronomers have an easy way to visualize their disks. The scripts are meant to be easily modified and flexible. They can be used to interpolate multiple complex disks and overlay multiple molecules.

References

Garufi, A., Ginski, C., van Holstein, R. G., et al. 2024, A&A, 685, A53

Garufi, A., Carrasco-González, C., Macías, E., et al. 2025, A&A, 694, A290

L. Ricci, B. Burrill, I. R. Z. Z. 2022, The ngVLA Memo Series

Law, C. J., Loomis, R. A., Teague, R., et al. 2021, ApJS, 257, 3

Thilenius, C. 2024, Discovering CASA: An Introduction to Imaging Radio Telescope Data