

Cloud fragmentation and chemical evolution of the high-mass star-forming region G327.3-0.6

PhD Thesis



Author: Tianwei Zhang

Supervisor: Prof. Dr. Peter Schilke

Second reviewer: Dr. Volker Ossenkopf-Okada

Institute of Physics I
Faculty of Mathematics and Natural Sciences
University of Cologne

September 5, 2023

Abstract

In the struggle to understand how stars form in a cluster, it is important to study the morphology, kinematics and chemistry of the star-forming clouds. This thesis focuses on the high-mass star-forming region G327.3-0.6, which is a 3 pc filament at a distance of 3.3 kpc, hosting one hot molecular core and a set of cold dense cores. It was observed with the Atacama Large Millimetre/Sub-millimetre Array (ALMA) in four spectral windows around 1.3 mm, and a $4'' \times 4''$ region was mapped with high resolution ($2''$).

First, we begin with a continuum analysis. The ALMA 12m and 7m array data were self-calibrated to improve SNR by a factor of 2. The high-resolution continuum map ($2''$ at 1.3 mm) was then combined with the ACA ALMA data and the single dish data to correct the interferometer-missing flux. The combined image was analyzed by various algorithms to identify the cores. The dendrogram algorithm, together with the background subtraction, performed the best and determined a list of 66 compact cores. We used minimum spanning tree to build a network of the core distribution, and derived a Q-value of 0.49, which favored sub-structures in our source. The two-point correlation function did support double peaks of the nearest-neighbor separations. For each distinct core, the prestellar mass was derived from the continuum flux. We fitted the core mass function with an index of -0.83, which is a hint of high-mass star-forming regions. The mass-radius relation follows Larson's power law with an index of 2.26, indicating more high-mass cores in our samples. The mass segregation ratios are calculated to be close to unity, leading to the conclusion of a uniform mass cluster. The study of fragmentation in the filament and core scale is based on the theory of cylindrical fragmentation and (non)thermal Jeans fragmentation. Both scales show a dominant role of turbulence support rather than thermal pressure, while the filament/cores are not preferred to continue fragmenting.

For spectral analysis, all average spectra for the core area were extracted. First, we fitted the hot core spectrum using toolbox XCLASS. 26 molecules and 39 isotopes were fitted together, and an error estimation using MCMC method was done towards CH_3OH , proving the reliability of its high temperature (270 K) and small source size ($0.485''$ 1500 au). We identified the new vibrational state ($v_{13}+v_{21}=1$) lines of ethyl cyanide with the most recent laboratory data, and the fitting is improved with two components of cold (53 K) and hot (218 K) layer. Fitting the entire hot core region gives a gradient of the line parameters. For the other cores, formaldehyde was first fitted under the LTE assumption. We tested to fit with band 7 low resolution data, to fit two components, and to fit under non-LTE, all of which tell us that the three para- H_2CO lines are not enough to limit the temperature error. Despite that, the derived H_2CO abundance ratio has a peak distribution at the literature value (10^{-10} – 10^{-9}), although it has a much wider range (10^{-12} – 10^{-7}) because of high sensitivity and the hot environment. Methanol was also fitted for each core, whose temperatures are not the same as those of formaldehyde.

Then we used statistical methods to examine the evolutionary status of the cores. Principal component analysis (PCA) and clustering algorithm were applied to 65 cold cores. Eight principal components were adopted, and the cores were assigned to four groups according to their characteristic lines. Different groups possibly trace different evolutionary stages of early star formation: Group 1 shows weak lines of extended molecules, located far from the main filament; Group 2 and 3 have similar and richer lines, while Group 3 is closer to the hot core and its line strength is slightly stronger than Group 2; Group 4 traces the most significant cores on the filament and presents the richest chemistry among the four groups. The cores in Group 4 have a temperature at around 100 K and usually show intermediate mass, which tends to

enter a hot core phase. Hence, PCA has been shown to be profitable in classifying cores.

For the whole filament, we derived the moment-zero maps of 42 transitions of 24 molecules. We used two methods, PCA and histogram of gradient (HOG), to examine the correlation between the molecules. We then investigate the eigenvectors of PCA, the mean resultant vector (MRV), the projected Rayleigh statistic (PRS), and alignment measurement (AM) of HOG. We found that the transitions with similar energy levels tend to show highly correlated moment zero maps, as well as the molecules with the same special atom (e.g., sulfur and deuterium). The most distinct molecules are CCD and H_2CO . N_2D^+ is also different from all the other images. Compared to the continuum image, we conclude that DCN, H_2CO , CH_3OH , HC_3N , OCS, and ^{13}CS are the most relevant molecules in descending order, while $\text{c-C}_3\text{H}_2$, CCD, $\text{C}_2\text{H}_5\text{CN}$, and HNC are the least correlated. In addition, the kinematics of the entire filament was explored using the DCN central velocity map (moment 1 map), which shows hints of global collapse towards the hot core.

As for the infrared counterparts, we looked into Spitzer RGB image and found 4 new extended green objects (EGOs) associated with the sub-mm cores. We also found 16 GLIMPSE high-reliability point sources appear at 13 sub-mm cores out of 66. Comparing the infrared image with the moment-zero map of a photon-dissociation region (PDR) tracer $\text{c-C}_3\text{H}_2$, we probably find a new PDR towards the west of the $\text{c-C}_3\text{H}_2$ arc. For the hot core, it is very possible to be a face-on disk and to fragment into multiple systems.

Contents

1	Introduction	1
1.1	Star formation process	1
1.1.1	Low-mass star formation	1
1.1.2	High-mass star formation	2
1.2	Morphology of star-forming regions	4
1.3	Chemistry in star-forming process	6
1.3.1	Cold core chemistry	8
1.3.2	Hot molecular core	8
1.4	A massive star-forming region G327.3-0.6	9
1.5	Motivation of this work	11
2	Observation and data reduction	13
2.1	Interferometry basics	13
2.2	ALMA observation	16
2.3	Calibration	18
2.4	Imaging	19
2.4.1	CASA version	20
2.4.2	Number of iterations	21
2.4.3	Weighting	21
2.4.4	Masking	23
2.4.5	Deconvolution algorithms	24
2.5	Self-calibration	25
2.5.1	Phase calibration	27
2.5.2	Amplitude calibration	28
2.6	Data combination	29
2.6.1	Continuum of ALMA 12m and 7m-array	31
2.6.2	Continuum of interferometer and single dish	33

2.6.3	Spectral combination	39
3	Continuum study	42
3.1	Core identification	42
3.1.1	Background subtraction: Findback	43
3.1.2	Core extraction: Dendrogram	43
3.2	Spatial distribution of the sources	51
3.2.1	Minimum spanning tree	51
3.2.2	Clustering	54
3.3	Mass estimation	56
3.3.1	Core mass function (CMF)	58
3.3.2	Larson's law	59
3.3.3	Mass segregation	60
3.4	Fragmentation	61
3.4.1	Cylindrical fragmentation	62
3.4.2	Core fragmentation	64
4	Molecular spectral analysis	65
4.1	Line fitting basics	65
4.1.1	Radiative transfer	65
4.1.2	Non-LTE fitting	66
4.1.3	Line profiles	68
4.1.4	Software: XCLASS for spectral analysis	69
4.2	Hot core spectra	70
4.2.1	Vibrational excited lines	73
4.2.2	Error estimation	75
4.2.3	Source size	76
4.2.4	Gradient with radius	77
4.3	Temperature derivation of cold to warm cores	78

4.3.1	Formaldehyde, H ₂ CO	78
4.3.2	LTE temperature	82
4.3.3	Non-LTE modeling with RADEX	89
4.3.4	Two components fitting	91
4.3.5	Other molecular tracers	92
4.4	Core classification	95
4.4.1	Principle component analysis (PCA) basics	95
4.4.2	Preparations before PCA	97
4.4.3	Applying PCA to the spectra of cores	98
4.4.4	Clustering and classification	98
4.4.5	Common spectra of groups	101
4.4.6	Spatial distribution of groups	106
4.4.7	Limitations and discussion	106
5	Large scale structures	108
5.1	Deriving moment maps	108
5.1.1	Correlation of different transitions	111
5.1.2	Applying PCA to moment 0 maps	113
5.2	Histogram of gradient (HOG) analysis of moment 0 maps	114
5.2.1	Computation of the HOG	116
5.2.2	The mean resultant vector	118
5.2.3	The projected Rayleigh statistic	120
5.2.4	The alignment measurement	122
5.2.5	Correlation between different molecules	123
5.3	Velocity distribution	125
6	Star formation in the infrared	127
6.1	Extended green objects (EGO)	131
6.1.1	Possible PDR	131

6.2	Hot core and a possible face-on disk	132
6.2.1	VLT PAH continuum	133
6.2.2	Possible disk fragmentation	134
7	Summary and outlook	138
	References	140
	Appendix	153
	Acknowledgement	155

1 Introduction

Stars are fundamental objects in astronomy. By transforming gas into stars, the star formation process determines the structure, luminosity, and evolution of the galaxies. Star formation is inextricably tied up with planet formation and creates most of the elements that make up us and the world around us. Hence, star formation constitutes one of the basic topics in astrophysics. Multiple theories and observations have been proposed to understand the morphology, dynamics, chemistry, and process of star formation, which are introduced in this section.

1.1 Star formation process

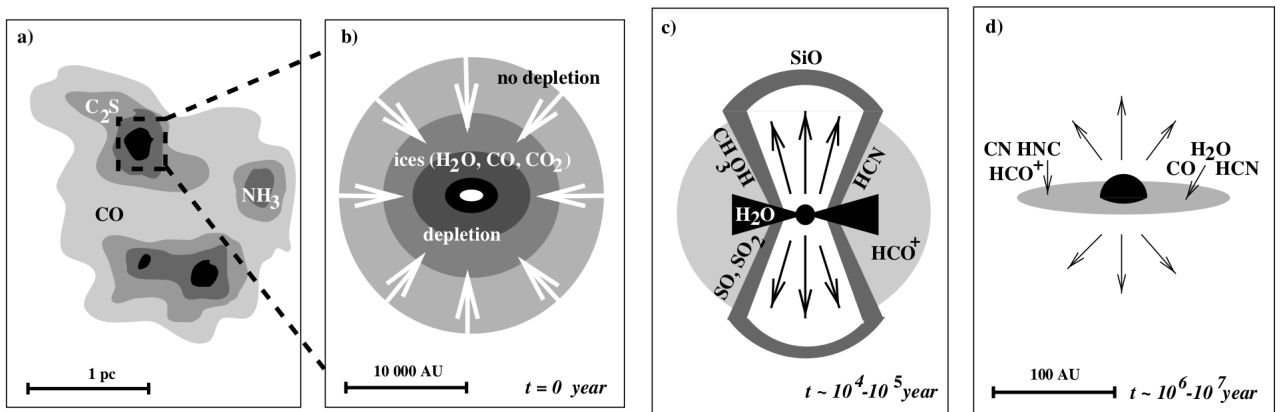


Figure 1.1: A schematic view of low-mass star formation. (a) Dark cloud cores, roughly 1 pc in size, gradually contract until (b) magnetic support is overcome and inside-out collapse begins at t equals zero. (c) For $\sim 10^4 - 10^5$ years, a phase of both high accretion and supersonic outflow occurs in deeply embedded protostars (young stellar objects or YSOs). (d) Gradual clearing by the outflow leaves only the young T Tauri star and a residual protoplanetary accretion disk, which, on time scales of $10^6 - 10^7$ years, leads to the formation of a mature planetary system. (Hogerheijde, 1998)

1.1.1 Low-mass star formation

The theory of low-mass star formation was well developed (Draine, 2011), shown in Figure 1.1. The first step is the formation of a self-gravitating dense core, which is half or several solar masses (M_{\odot}) and likely to form a single star or a close multibody system. When the core becomes gravitationally unstable, it will begin to undergo isothermal collapse. When the dust becomes optically thick and cannot cool the gas efficiently, adiabatic collapse starts. The hydrostatic equilibrium will soon be reached, with a mass of $10^{-2} M_{\odot}$ and temperature ~ 170 K, indicating the birth of the first core. As the temperature increases, H_2 starts to dissociate and later H starts to ionize, leading to the isothermal collapse again. After all hydrogen is ionized, the gravitational energy turns into heating and a second core, also named *protostar*, is formed when the temperature reaches 10000 K and volume density $\sim 10^{-2} \text{ g cm}^{-3}$. At the same time, the infalling gas from the envelope will generally have nonzero angular momentum, and the material will be flattened to form a rotational disk around the protostar. To slow down the rotation of the protostar, bipolar outflows appear to transport excess angular momentum

outward. At the end of this process, the outflows turn weak, planets form, and the temperature of the central star becomes high enough to trigger nuclear fusion, at which moment the real main-sequence star forms.

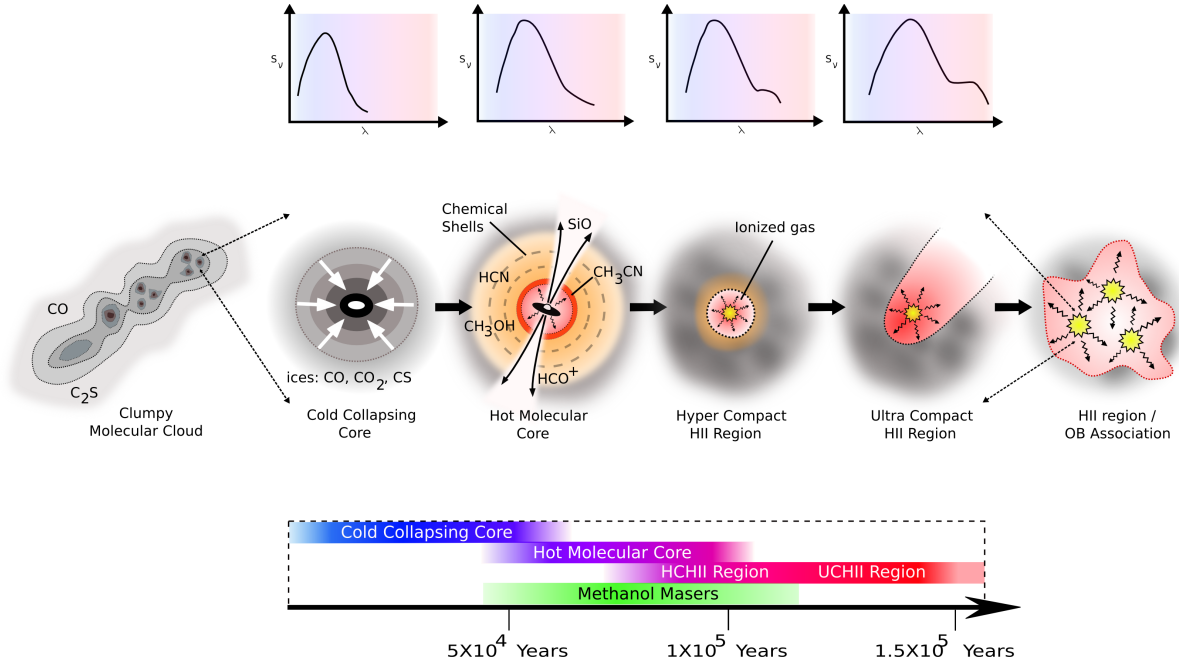


Figure 1.2: Middle panel: Process of high-mass star formation. Top panel: Spectral energy distributions (SEDs) for each stage. Lower panel: Corresponding timescale.

However, the star formation processes of low-mass stars and high-mass stars involve separate mechanisms.

1.1.2 High-mass star formation

The general scenario of high-mass star formation emerging from the recent observations is (Purcell, see Figure 1.2): Clumpy molecular clouds \rightarrow Cold collapsing core \rightarrow Hot molecular cores (HMCs) \rightarrow hyper/ultra-compact HII regions \rightarrow HII region surrounding the ionizing high-mass stars (OB association). The first two steps are the same as those of low-mass stars, whereas the deviations are larger and larger in the later stages. The difference originates from the accretion. The time of accretion onto protostar is:

$$t_{acc} = M_*/(dM_{acc}/dt),$$

which increases with mass. The gravitational contraction time of protostar to the main sequence is the Kelvin-Helmholz time (KH time):

$$t_{KH} = GM_*^2/R_*L_* \propto M_*^{-1.8},$$

which decreases with mass. When a low-mass or intermediate-mass star is born, the KH time is longer than the accretion time, which means accretion stops before the gravitational collapse (left part in Figure 1.3). But for a massive star, it seems to be still accreting when already formed, which is quite unreasonable since the radiation pressure (maybe together with wind and ionizing flux) would halt accretion. This means that high-mass stars ($> 10M_\odot$) cannot

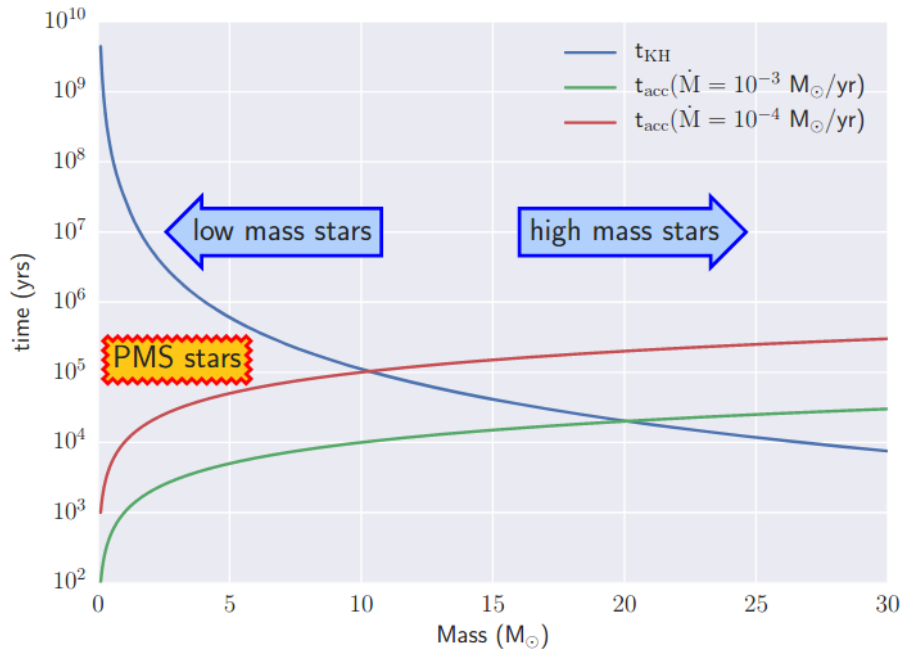


Figure 1.3: Kelvin-Helmholtz time scale vs. accretion time scale of high mass stars vs. low mass stars, for two accretion rates. Low mass stars have a pre-main sequence phase after they stop accreting but still gain energy from gravitational contraction. (Schilke, 2015)

be formed simply by gravitational collapsing and subsequent gas accretion similar to low-mass stars. Hence several alternative theories arose.

One of the two prominent models is known as the *monolithic collapse*, or by another name the *turbulent core model*. For a spherical accretion, usually only thermal support is considered, leading to a typical accretion rate of $10^{-5} M_{\odot} \text{ yr}^{-1}$ (Shu, 1977), which is not sufficient to form high-mass stars. McKee and Tan (2003) therefore suggested the turbulent support, allowing for much higher accretion rates. However, it requires a strong assumption on the modeling conditions: the starting density distribution is highly peaked ($n \propto r^{-1.5}$); the final stellar mass is preassembled, isolated from the rest of the cloud, which is not the truth in space.

An alternate theory is the *competitive accretion* model, where high mass star formation takes place in a clustered environment (Bonnell et al., 2001; Bonnell and Bate, 2006). All of the star-forming cores are created with equal mass, which then accrete mass through Bondi-Hoyle accretion. The accretion rates depend on the location of the core in the cluster potential, which means that the central core near the gravitational well can accumulate more than the ones that are located outward. In other words, the cores must compete to gain mass, hence the name competitive accretion. The accretion will terminate until gravitational interactions kick the star out of the rich feeding zones, or all the material falls onto these cores. The basic assumptions of this model involve the isothermal and uniform gas density at the beginning.

These two theories of high-mass star formation are almost orthogonal approaches and use very specific initial conditions, neglecting the effect of feedback process. Therefore, the debate changes from only one of them being right, to a unified high-mass star-forming theory, where more physical processes are considered for, e.g., the presence of magnetic fields and disks. Recent studies towards several star-forming clusters (Rivilla et al., 2013; Schilke, 2015; Yuan et al., 2016) support a coexistence of two theories by showing strong turbulence inside the cores and mass flows along the filaments toward the central cores. Meanwhile, the level of fragmentation has a profound influence on the final results.

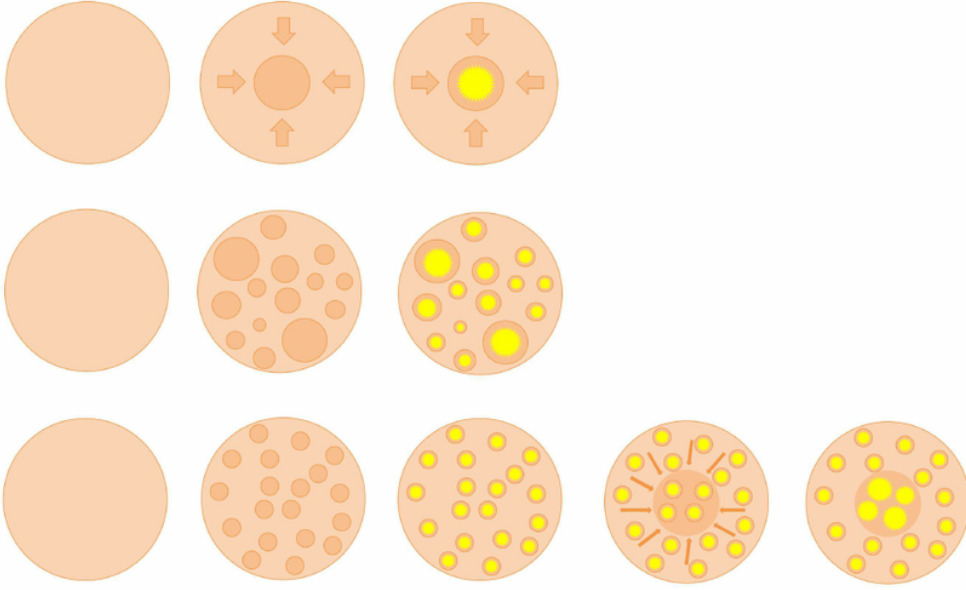


Figure 1.4: Scenarios for massive star formation. Upper panels: low fragmentation and monolithic core accretion. Middle panel: intermediate fragmentation and subsequent subcore accretion. Lower panel: high fragmentation, subcore accretion forming stellar seeds, and competitive accretion. The dominant role of each theory is mainly determined by the level of fragmentation in the core (Rivilla et al., 2013)

Few alternative theories like the low-mass protostar merging into high-mass ones, and cloud-cloud collision enhanced star formation are also put forward. However, in summary, there is still no concrete theory of high-mass star formation. Observational studies of high-mass star formation are highly challenging in many aspects. Firstly, massive embryos are deeply embedded in the thick envelope, which is only visible in (sub-)mm bands. As seen in the spectral energy distribution (SED) in Figure 1.2, the dust emission from the envelope dominates the SED, while the hot core emission is only a small bump, which changes slightly over evolution and is hard to recognize. In addition, the short lifetime of high-mass protostars (10^5 yrs vs. 10^7 yrs for low-mass ones) and large distances from us make it even more difficult to find samples and to determine stages of development. Hence, any sample of high-mass star-forming regions is of great importance to explore the early evolution of massive stars.

1.2 Morphology of star-forming regions

Since fragmentation is very likely to play an important role in the high-mass star formation and the formation of clusters, some basics about the morphology of star-forming regions should be mentioned. First, several widely referred terms are summarized.

- Giant molecular clouds (GMCs) refer to large and dense molecular clouds that exceed the mass of $10^5 M_{\odot}$, which is enough to form hundreds of thousands of stars. They are very cold, having temperatures down to 10 kelvin, which can last for 10 to 100 million years. They usually appear as dark lanes in the optical wavebands, while far-infrared observations can penetrate the thick molecular cloud, revealing the star formation inside.
- Cloud is the similar and simplified version of GMC, which describes the large structure of our interest on scales of ≥ 10 pc. A cloud fragments into clumps.

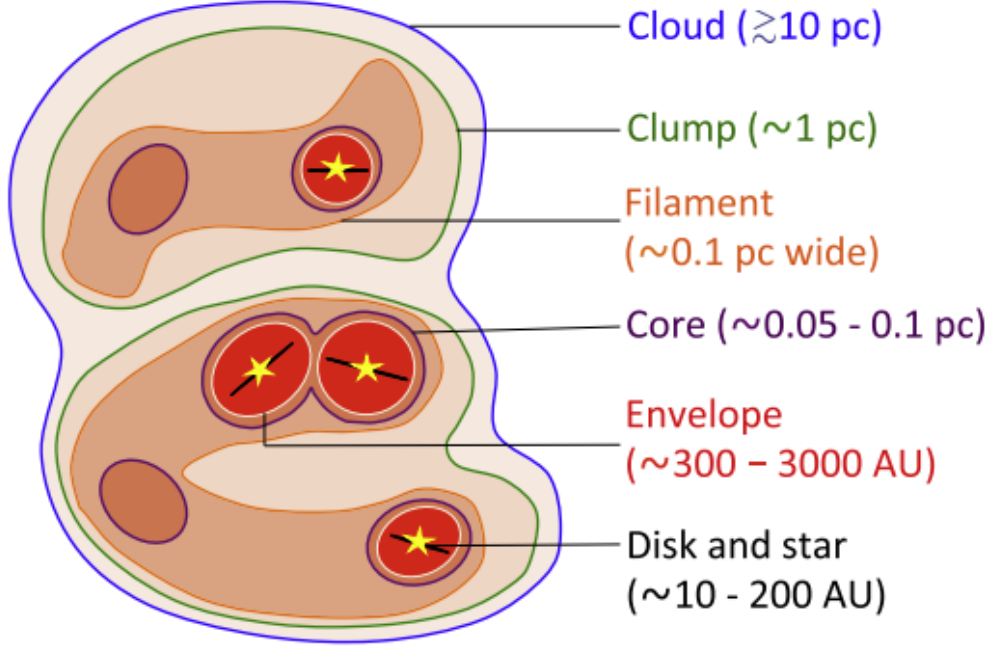


Figure 1.5: Hierarchical fragmentation schematic. The figure shows the cloud, clumps, filaments, cores, envelopes, and protostellar systems. The image is not drawn to scale.

- Clump is used to describe the self-gravitating dense regions with masses $\sim 10^3 M_\odot$ and ~ 1 pc in size. A clump may fragment into multiple cores, leading to the formation of stars and star clusters.
- Core is the specific site where new stars (usually single or binary) are formed, at a scale of ~ 0.1 pc.
- Envelope is a dense condensation of size scale ~ 300 - 3000 au. Inside the envelope, there are usually an inner envelope, a protostellar disk, and a central young star.
- Disk is the rotating circumstellar disk of dense gas and dust surrounding the young stellar objects, providing materials from the inner edge to the surface of the star. It is usually several hundreds of au wide (up to 1000 au), ending in a planetary system.
- Filament is the elongated gaseous structure that inhabits multiple cores. They were observed in different sizes, for example, a small scale of 0.1 pc, a large scale of 10 pc. The giant molecular filaments can spread several hundred pc (Smith et al., 2020). The average width of the filament is around 0.1 pc.

During the process of cloud collapsing into stars, the morphology changes over time through hierarchical fragmentation. A cartoon in Figure 1.5 shows hierarchical structures in star formation: cloud \rightarrow clump, filament \rightarrow core \rightarrow envelope, protostellar system. The fragmentation is due to the Jeans instability.

For a non-rotating and unmagnetized spherical gas cloud, when the gravitational force exceeds the thermal pressure, it would be unstable and begin to collapse. As shown in Figure 1.6, the pressure force is

$$F_p = pA \sim p \frac{V}{r}$$

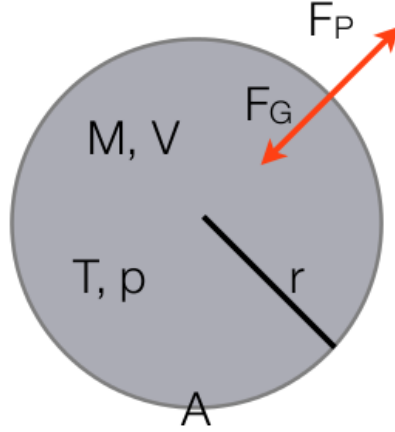


Figure 1.6: A spherical cloud balanced by thermal pressure and gravitation.

The gravitational force is

$$F_g = G \frac{M^2}{r^2} \sim G \rho^2 V r$$

If two forces are in balance, a typical length scale named Jeans length is derived:

$$r = \sqrt{\frac{p}{G \rho^2}} \propto T^{1/2} \rho^{-1/2} \quad (1.1)$$

Jeans mass is thus acquired, $M_J \propto T^{3/2} \rho^{-1/2}$.

A cloud will start to fragment into small pieces if it exceeds the Jeans length. With low temperature and high density, a cloud tends to be less stable and hence more likely to form stars. For a molecular cloud with a temperature of 10 K and a number density of 100 cm^{-3} , the Jeans length is around 1.7 pc, which is one magnitude smaller than the normal scale of the cloud. Therefore, a cloud could fragment into pieces, the so-called *clumps*. When the clumps continue the isothermal collapse and when the density becomes higher, the Jeans length is reduced, and another level of fragmentation from clump to cores is triggered. This cascade of fragmentation is named *hierarchical fragmentation*. The fragmentation terminates when a central star begins to form, since the condensation becomes optically thick and the radiation heats up the gas, thus leading to an increase in Jeans length which will prevent further fragmentation.

By observing and analyzing the distribution of clumps and cores, the fragmentation status can be figured out. By measuring the line width and the kinetic temperature, the turbulence inside condensations and their stabilities can be derived, giving hints to the initial conditions of star formation.

1.3 Chemistry in star-forming process

Astrochemistry can provide important clues about the physical conditions of molecular clouds. With the improvement of observational tools and increasingly sophisticated theory, chemical evolution in different stages of star formation is broadly explored. The scenario for the formation of a single, isolated low-mass star was discussed in Section 1.1, and now the chemistry is roughly introduced.

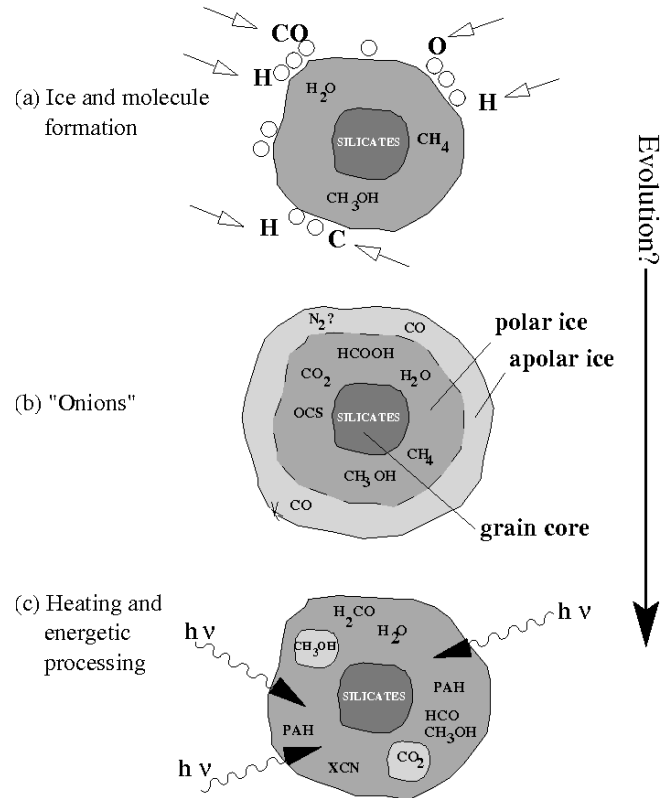


Figure 1.7: A grain and the evolution of its icy mantles (Boogert and Ehrenfreund, 2004). (a) H_2O rich ice mantels and molecules are rapidly formed with a high atomic abundance. (b) "Polar ice" refers to hydrogenated species like H_2O , CH_3OH and NH_3 that have large dipole moments, which evaporates at ~ 90 K. Apolar ice may form under low temperature and form onion-like grain mantels. (c) Surface diffusion brings adsorbed species into close contact. And with the energy provided by cosmic rays and UV photons, molecules and radicals react to form complex organic molecules. Later they are released to gas phase by heating process.

1.3.1 Cold core chemistry

Cold molecular clouds are dominated by chemical reactions that are gas-phase, ion-neutral, and neutral-neutral reactions, leading to the formation of small radicals and unsaturated molecules, e.g., long carbon chains and sulfur chemistry. When a molecular cloud begins to collapse, the density increases while the temperature remains low; therefore, molecules accrete onto the grains to form icy mantles (see Figure 1.7). The hydrogenation of solid CO leads to species like H_2CO and CH_3OH , which are essential ingredients of more complex polymerization reactions. The grain surface then acts as a catalyst for the chemical reactions, enhanced with the help of ultraviolet photons, X-rays, and cosmic rays. After the formation of a new protostar, its radiation heats the surrounding gas and dust to release some of the molecules back into the gas phase. Meanwhile, the outflows from the young stellar objects penetrate the envelope by high-temperature shocks, in which the molecules within the icy mantles can be returned to gas or even dissociated. Due to the strong shock, some refractory material containing silicon also appears in gas form; thus, SiO is adopted as an outflow tracer (see Figure 1.1 (c)). Also, freshly evaporated molecules can drive more complex chemical reactions, which is a period $\sim 10^5$ years named *hot core phase* (van Dishoeck and Blake, 1998). Finally, for low-mass stars, the envelope is dispersed by the stellar winds, whereas for massive stars, the ultraviolet photons will dissipate through the region to form photon-dominated regions (PDRs).

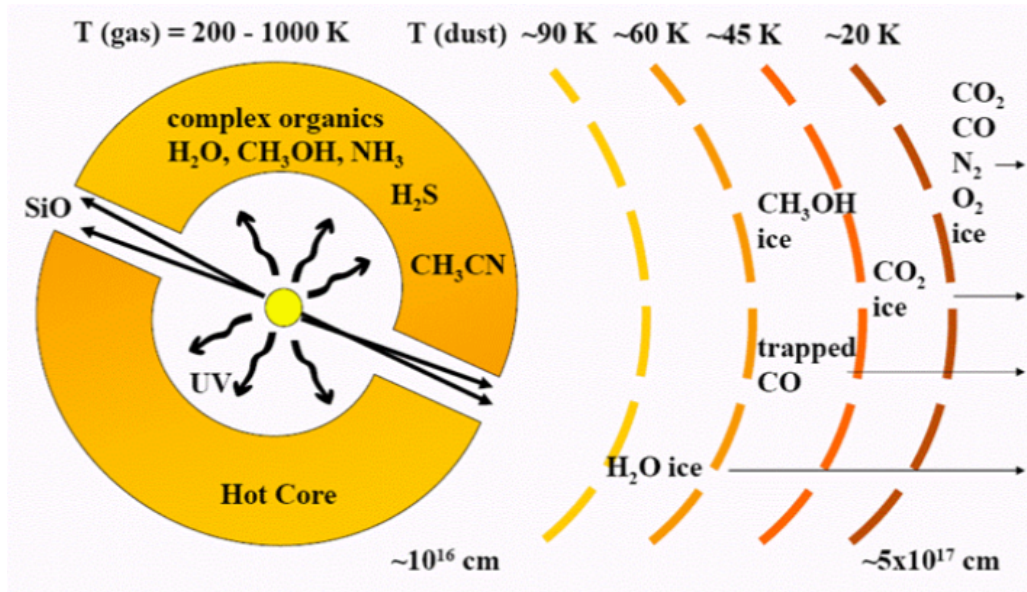


Figure 1.8: A sketch of the hot core and chemistry within. In the envelope, there are abundant of complex organic molecules and their ingredients. Radial distribution of temperature is also shown.

1.3.2 Hot molecular core

Among all the evolutionary stages, the hot core phase is emphasized by a series of researches due to its complexity and its role in the transition from a dense core to a real star. The term "hot core" or "hot molecular core (HMC)" is named after the original source in Orion, which are hot ($\sim 100\text{--}300$ K) and dense ($\sim 10^7$ cm^{-3}) condensations of molecular gas and dust within the innermost several thousand au. In most cases, they are found inside massive star-forming regions and have the richest gas-grain phase chemistry in the interstellar medium (van

Dishoeck and Blake, 1998). A sketch can be found in Figure 1.8. An equivalent hot core phase for low-mass stars is known as the *hot corino*.

Numerous (sub-)millimeter wavelength observations toward the hot cores reveal a rich chemistry in nitrogen-, sulfur- and oxygen-bearing molecules. In addition, complex organic molecules (COMs), which consist of six or more atoms such as ethanol ($\text{C}_2\text{H}_5\text{OH}$), dimethyl ether (CH_3OCH_3) and methyl formate (CH_3OCHO), contribute a large number of emission lines at (sub-)mm wavebands, detected in a handful of individual hot cores (e.g., Sewilo et al. (2018); Csengeri et al. (2019); Gieser et al. (2019)). The line intensities also vary between different sources and evolutionary stages. To model the chemical composition and development, chemical networks and kinetic approaches should be applied appropriately. For example, various 3D models for HMCs were developed (Choudhury et al., 2015) when gas-grain chemistry was coupled with the radiative transfer model. By varying the temperature, molecular abundances and spatial distribution, the models can produce synthetic spectrum along the radius through evolutionary time to compare with the observations.

1.4 A massive star-forming region G327.3-0.6

G327.3-0.6 (hereafter G327) is a widely searched massive star-forming region at a relatively close distance of 3.3 kpc (kinematic distance derived by HI absorption) near the galactic plane (Urquhart et al., 2012).

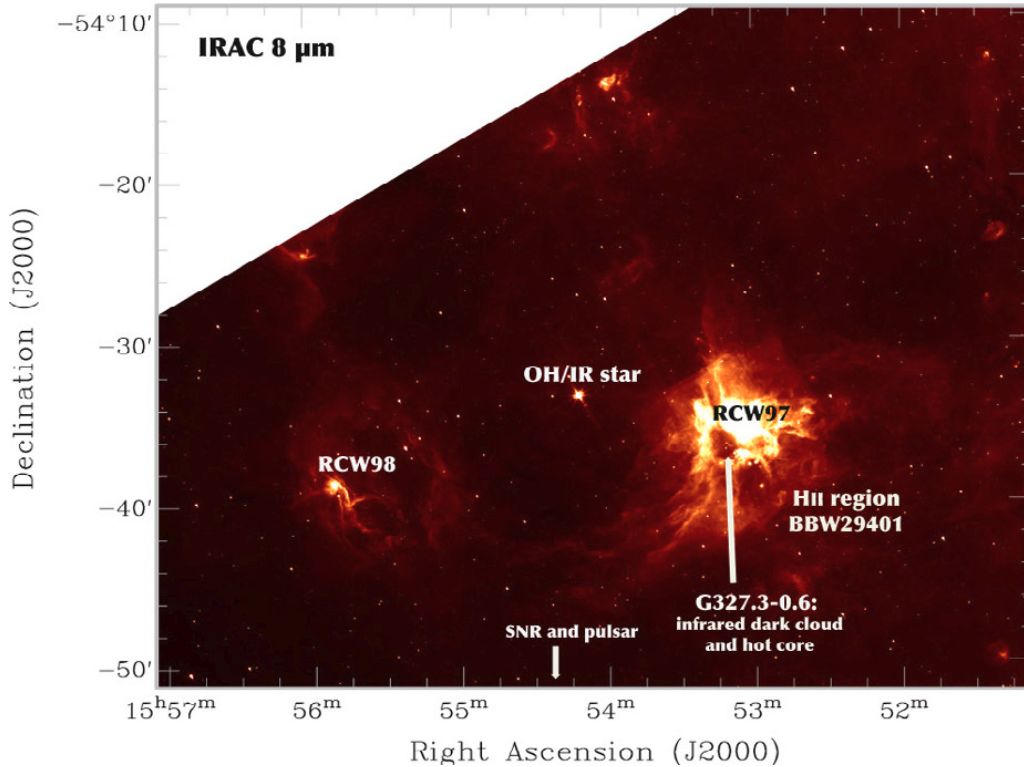


Figure 1.9: Overview of G327.3-0.6 environment. *Spitzer*/IRAC image at 8 μm shows RCW97 and HII region BBW29401 at the border of large scale infrared bubble with G327 embedded. (Minier et al., 2009)

It first attracted attention because of a nearby $\text{H}\alpha$ emission region RCW97 (Rodgers et al., 1960) and a bright expanding HII region BBW29401 (Goss and Shaver (1970), see Figure 1.9).

Later in 1971, OH absorption spectra at 1665, 1667, 1720 MHz and OH emission at 1612 MHz were first detected towards this source near the velocity -49 km s^{-1} (Robinson et al., 1971). At the same period, prominent H_2O emission attracted attention, whose intensity changed over years (Johnston et al., 1972). According to Caswell et al. (1974) and Batchelor et al. (1980), it remained one of the strongest water maser sources near the galactic plane.

To better understand the maser-related regions, several single-pointing spectral line surveys were conducted (Whiteoak and Gardner, 1978; Gardner and Whiteoak, 1978; Batchelor et al., 1981; Brand, 1983), which covered the frequency range of HCN (1-0), CS (1-0), HCO^+ (1-0), and CO (2-1). The radial velocity of these lines are well-consistent at around -47.0 km s^{-1} , while the narrow line widths and simple Gaussian profiles indicate that G327 has relatively simple physical structure, which is excellent for the follow-up study.

The observations towards G327 have made great progress since several telescopes at the southern hemisphere, especially Chile, were established. Multiple lines of CH_3CN and CH_3CCH were detected by the Swedish-ESO Submillimeter Telescope (SEST) (Bergman, 1992), from which two adjacent dense cores were found: one cold core (kinetic temperature $T=30 \text{ K}$) and one hot core ($T=100\text{-}200 \text{ K}$). For the latter position, a compact core of 0.06 pc , $100\text{-}200 \text{ K}$, $10^6 - 10^8 \text{ cm}^{-3}$ and a cooler and less dense envelop of 0.3 pc , $40\text{-}60 \text{ K}$, $10^5 - 10^6 \text{ cm}^{-3}$ were identified (see left panel of Figure 1.10).

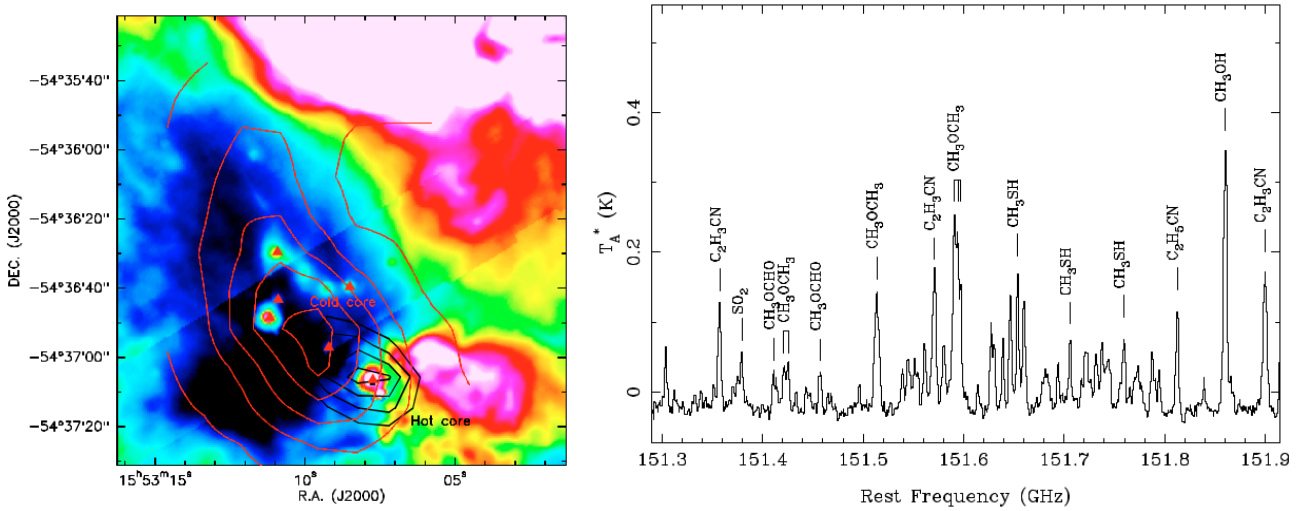


Figure 1.10: Left: APEX image of N_2H^+ (3-2) in red contours and CH_3OH (7-6) in black contours trace cold core and hot core separately. Background is $8 \mu\text{m}$ GLIMPSE emission, showing massive YSOs by triangles (Wyrowski et al., 2006). Right: Sample spectrum around 151.6 GHz. (Gibb et al., 2000)

Soon, the distribution of warm gas and large-scale environment were explored in detail. CO (3-2) mapping (Wyrowski et al., 2006) and mid-J CO survey (Leurini et al., 2013) were done by the Atacama Pathfinder Experiment (APEX), showing excitation temperatures at $20\text{-}80 \text{ K}$ towards the gas around the HII region. H_2 column density ranges from a few 10^{21} cm^{-2} in the inter-clump gas to $3 \times 10^{22} \text{ cm}^{-2}$ towards the hot core. The presence of high-density inter-clump gas indicates that it is produced by relatively dense components with a low beam filling factor. The continuum data gives a mass of $420 M_\odot$ for the hot core. Minier et al. (2009) examined two infrared bubbles and HII regions surrounding the infrared dark cloud where the hot core is located, suggesting that the bubbles expand towards the hot core region, hence triggering star formation. Outflows from massive young stellar objects are also found (Leurini et al., 2017).

With the development of the hot core theory, the chemistry of the G327 hot core was investigated in more detail. Ethylene oxide ($c\text{-C}_2\text{H}_4\text{O}$) and its isomer acetaldehyde (CH_3CHO) were first detected (Nummelin et al., 1998) with an excitation temperature of 20-40 K, indicating these species are excited either in the cool envelope or subthermally. Gibb et al. (2000) presented 1-3 mm spectrum with high abundance of nitrile species (CH_3CN , $\text{C}_2\text{H}_3\text{CN}$, $\text{C}_2\text{H}_5\text{CN}$, and NH_2CHO) and saturated organics (CH_3OH , $\text{C}_2\text{H}_5\text{OH}$, CH_3OCH_3) towards the hot core at the center $\sim 2''$, while extended emission (OCS , H_2CS , CH_3CCH , and HCOOH) could be fitted in the zone of the cool envelope (see right panel of Figure 1.10). G327 also gains the first detection of CH_3SH outside the Galactic center, indicating the evaporation of icy grain mantle material instead of a simple gas-phase chemical model. Schilke et al. (2006) first attempted to perform a submillimeter line survey toward G327 other than Orion-KL. 245 lines were found, of which 238 were identified, providing evidence of a very compact and hot source.

1.5 Motivation of this work

The massive protostar is deeply embedded in molecular clouds, making it difficult to explore. Luckily, the intermediate phases between the IRDCs and the compact HII regions, known as the hot cores, provide clues of central protostars, since they heat their surrounding gas and dust to high temperatures (> 100 K) without ionizing them (still no UV photons are created). The evaporation of water and organic-rich ice mantles leads to the detection of a wide variety of COMs, which also trace the evolutionary stages in detail. Therefore, the study of physics and chemistry of hot molecular cores can reveal intricate details of early phases in high-mass star formation, especially when the initial evolution conditions are still not well known.

G327.3-0.6 provides a superior sample for examining the hot core as well as the surrounding cold cores for the following reasons. The mass and luminosity of its hot core is of $420 M_\odot$ and $10^5 L_\odot$, and it is comparable to classical luminous hot cores such as G10.47+0.03 and G31.41+0.31 (Hatchell et al., 2000), but the advantage of G327.3-0.6 compared to those regions is its relative nearby distance (factor of 2 smaller). Another outstanding feature is that the different evolutionary phases of star formation coexist in a small (~ 3 pc) region, hence it is very efficient for studying the cluster formation and the mass accretion from the filament to the hot core. Furthermore, the G327 hot core is remarkable in that it does not display a HII region feature, proving that it is an excellent candidate for a young high-mass protostar. It is also good for modeling, since no strong feedback like UV radiation comes into effect.

The complex chemical and physical processes occurring in G327.3-0.6 were not fully understood due to the lack of high-resolution spectral and mapping observations. Recently we obtained some new ALMA data towards G327, which is significant to unveil the complex processes in this active star forming region.

To start with, the high-resolution continuum map was analyzed. Multiple algorithms were adopted to identify the cores accurately, to get a glimpse of fragmentation. For these cores, their prestellar masses could be derived to study their stability, and their spectra were extracted for further fitting and comparison. The kinematics and evolutionary status of the filament were explored. Since we barely know anything about the derived cores, some machine learning techniques were also used to statistically extract crucial features and to gauge the effects of noise. Subsequently, the cores can be assigned into multiple groups according to their characteristic lines. The moment 0 maps of each molecule are explored with machine learning methods, in order to characterize the flux distribution and find correlations between the molecules. Our

long-term goal is to use certain molecules, whose abundance can be temporarily enhanced or weakened, as signposts of the evolutionary state of an object. This will be very practical for ALMAGAL, which is a 1000+ sources super-high-resolution ($< 0.5''$) ALMA project to study massive cluster formation.

In Section 2, the interferometric observation and data reduction are introduced, including the method of self-calibration and data combination. Section 3 presents the results of continuum cores, including the core mass function and fragmentation. Section 4 discusses the spectral line fitting process and the classification of the core evolutionary stages. In Section 5, the large-scale structures in the moment maps are examined, to understand the similarity and difference between molecules. Section 6 shows the related infrared observations. The last Section 7 gives a summary and a plan for future work.

2 Observation and data reduction

2.1 Interferometry basics

The angular resolution of single-dish radio telescopes is insufficient to obtain (sub-)arcsecond features to study the internal structure of high-mass star formation regions. For example, a 100 m single dish observing at 3 mm has a resolution of $\theta = 1.22 \frac{\lambda}{D} = 7.5''$. However, because of the technical limit of making single dishes with larger size and the lack of resources at the high altitudes required for (sub-)millimeter wavelength to minimize atmospheric effect, it would be impossible to improve the angular resolution with single dishes.

Interferometry then creates an opportunity to improve the resolution by combining multiple dishes as a large single telescope. The highest resolution depends on the maximum separation, that is, the baseline, between the dishes, and the sensitivity of interferometry is proportional to the total combined area of all the individual dishes (Wilson et al., 2013).

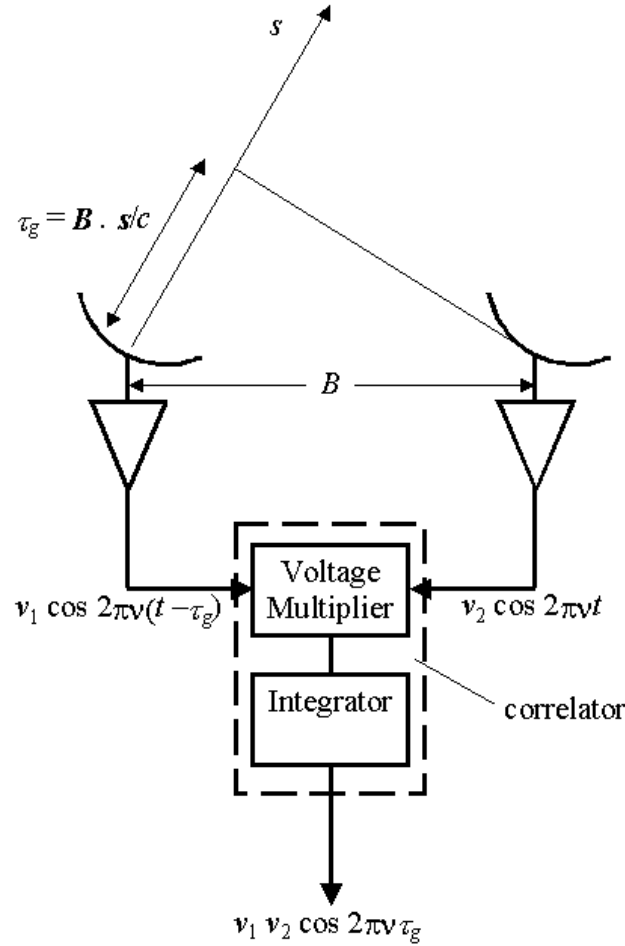


Figure 2.1: The sketch of two-element interferometer, with the relationship between baseline B and observing direction S .

Figure 2.1 shows a sketch of the simplest two-element interferometer¹, which consists of two antennas. Assuming the incoming wavefront is planar over the baseline for the cosmic

¹<https://web.njit.edu/~gary/728/Lecture7.html>

sources, the geometrical time delay of the signal arriving at the two antennas is

$$\tau_g = \frac{\mathbf{B} \cdot \mathbf{S}}{c} \quad (2.1)$$

The received signals pass amplifiers and give voltage outputs from each antenna at time t as

$$V_1 = v_1 \cos 2\pi\nu(t - \tau_g), \quad V_2 = v_2 \cos 2\pi\nu t \quad (2.2)$$

where ν is the frequency of the observation and v is the amplitude of the signal. As the voltage signals are fed into the correlator which works as voltage multiplier, the response is

$$V_1 \cdot V_2 = \frac{v_1 v_2}{2} \cos 2\pi\nu(2t - \tau_g) + \frac{v_1 v_2}{2} \cos(2\pi\nu\tau_g), \quad (2.3)$$

in which the high frequency term can be removed by low-pass filter. Usually, we adopt the antenna response as identical for each antenna, i.e., $v_1 = v_2$. And then the signal is averaged over a short period of time:

$$\langle V_1 \cdot V_2 \rangle = \frac{v_1 v_2}{2} \cos(2\pi\nu\tau_g) \quad (2.4)$$

which is simply depending on the time delay τ . Since the geometric delay changes due to the rotation of the earth, the cosine term oscillates slowly with time ($\tau_g = \frac{b}{\lambda} \sin \omega_{\text{earth}} t$), which means that the output of the correlator describes a fringe pattern. If we observe the source for sufficient time to see a full oscillation, we can infer the flux density of the source from the fringe amplitude.

Considering we observe a point source that is offset from the central field of view with a small angle $\Delta\theta$, the response of the interferometer V after solving for amplitude and phase can be written in terms of flux density F_ν ,

$$V = F_\nu \cos(2\pi \frac{b}{\lambda} \Delta\theta). \quad (2.5)$$

This equation only depends on the baseline and position at a certain wavelength. Hence, by observing at different baselines, we can determine the amplitude variation, which gives information of the source size.

The response pattern is in the form of a cosine function, which can be in sine form when a delay is added to one of the antenna pairs. Recalling $e^{ix} = \cos x + i \sin x$, Eq. 2.5 can be extended to the complex domain with the combination of the cosine and sine parts as:

$$V(b/\lambda) = F_\nu e^{i2\pi \frac{b}{\lambda} \Delta\theta}. \quad (2.6)$$

Now we analyze an extended source using exponential form. The visibility function will be the sum of the visibilities from all the infinitesimal points that make up the extended source. If we integrate $T(x,y)$, the sky brightness distribution at the angular coordinates (x,y) in the sky, the complex visibility becomes:

$$V(b/\lambda) = \int \int T(x, y) e^{-i2\pi(b_x/\lambda)x} e^{-i2\pi(b_y/\lambda)y} dx dy. \quad (2.7)$$

$V(b/\lambda)$ can be written as a 2D visibility function $V(b_x/\lambda, b_y/\lambda)$. Defining $u=b_x/\lambda$ and $v=b_y/\lambda$, we get

$$V(u, v) = \int \int T(x, y) e^{i2\pi(ux+vy)} dx dy. \quad (2.8)$$

Here, (u,v) are the East and North spatial frequencies in the unit of wavelength (i.e., uv plane), and (x,y) are the coordinates on the tangent plane of the sky in the unit of radians (i.e., image plane). The geometric relationship of the coordinates (x,y) and (u,v) is illustrated in Fig. 2.2, in which we ignore the w axis by assuming $w=0$. From the equation we see mathematically that the complex visibility function $V(u,v)$ is the 2D Fourier transform of the sky brightness distribution $T(x,y)$, under the condition of a small field of view, a far field, and an incoherent source (Burnard F. Burke, 2002). In turn, the source brightness can be recovered with the 2D Fourier inverse transform of the detected visibility:

$$T(x, y) = \int \int V(u, v) e^{-i2\pi(ux+vy)} du dv \quad (2.9)$$

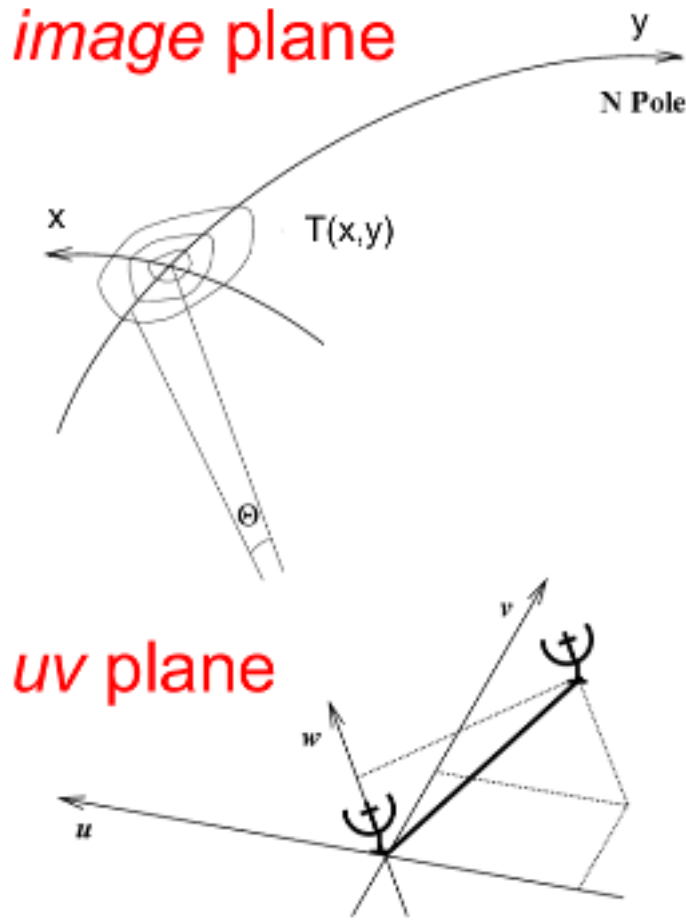


Figure 2.2: Geometric relationship between (u,v,w) and (x,y) . (x,y) are the coordinates on the sky and (u,v) represent the baseline vectors (Thompson et al., 2017).

The next step is to consider an N -element interferometer. There are $N(N-1)/2$ antenna pairs, and each of them forms a fringe. The observed visibility is a complex and discrete measurement, since the antennas are discrete and cannot continuously cover the entire uv plane. In other words, the interferometer observation is a sampling of the Fourier transform of $T(x,y)$ at discrete positions (u,v) . According to the Fourier transform, the longest baseline of the interferometer determines the smallest scale that it is sensitive to, while the shortest baseline is related to the large-scale structure. We note the observed visibility data set as \tilde{V}_{ij}

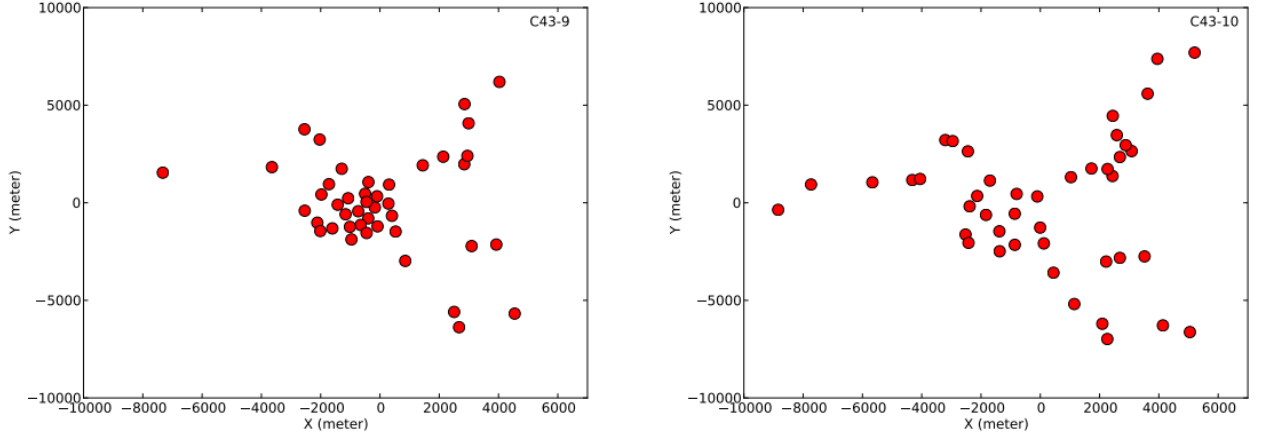


Figure 2.3: Two examples of ALMA interferometer configurations in the ground coordinates. Left: C-9 configuration with a minimum baseline of 368m and a maximum of 14 km. Right: The most extended configuration C-10 with a maximum distance of two telescopes as 16 km.

and the real visibility as V_{ij} . The calibration equation is

$$\tilde{V}_{ij}(t) = G_{ij}(t)V_{ij}(t), \quad (2.10)$$

where $G_{ij}(t)$ is the baseline-based gain at a time t . It can be approximated by the product of the two associated antenna-based complex gains $g_i(t)$ and $g_j(t)$,

$$G_{ij}(t) = g_i(t)g_j^*(t) = a_i(t)a_j(t)e^{i(\phi_i(t)-\phi_j(t))}, \quad (2.11)$$

where $a_i(t)$ is the antenna-based amplitude correction and $\phi_i(t)$ is the antenna-based phase correction. To determine the complex gain, we need calibrator observations, which require the calibrator source to be a point source with known flux (i.e. F_ν Jy) and position (i.e. phase=0 degree). The complex gain is then derived by

$$G_{ij}(t) = \tilde{V}_{ij}(t)/F_\nu, \quad (2.12)$$

which can be applied in two-step calibration, amplitude and phase calibrations (see more discussion in Section 2.3 and Section 2.5).

2.2 ALMA observation

The Atacama Large Millimeter/Sub-millimeter Array (ALMA) is an interferometer that works in wavelengths of 0.32 to 3.6 millimeter. The ALMA consists of 66 antennas, of which 50 have a diameter of 12 meters, while 12 are 7-meter dishes. The 66 antennas can be arranged in different configurations, i.e. the position pattern of the dishes can change from the most extended configuration (maximum distance 16 km, named C-10) and the most compact one (150 m, named C-1), seen two examples in Fig. 2.3. ALMA is the most powerful interferometer until now, except for very long-baseline interferometers (VLBI).

If a source is extended and cannot be covered by a single pointing, an observation with multiple pointings should be conducted. This observing mode is named *mosaic*, which is also adopted in this work. Since the primary beam is different for 12m-array and 7m-array, we used different pointing schemes. 109 12m-array pointings were performed (see the yellow circles in

Fig. 2.4), as well as 38 pointings of 7m-array, in order to cover the entire $4' \times 4'$ region of G327.

Regarding the frequency coverage and the spectral setup, ALMA can obtain the continuum with high sensitivity and at the same time observe the spectral lines. Polarization can also be observed with requirement. In this study, the high-resolution observation was performed in band 6, consisting of 4 spectral windows, as seen in Table 2.1. Traditionally, spectral line observations require a higher spectral resolution, which means a narrower channel width and fewer frequency coverage. In our setup, the spectral window 2 (spw2) in band 6 has a channel width as narrow as 7.812 kHz, corresponding to a velocity resolution of $\sim 0.1 \text{ km s}^{-1}$, which is sufficient to resolve the targeted spectral line like N_2D^+ .

G327.3-0.6 was observed with ALMA (et. al, 2015) during its cycle 4 on 28 March 2017, under project number 2016.1.00168.S. We used the 12m array consisting of 40 antennas, the Atacama Compact Array (ACA) with 10 7m-antennas, and the total power (TP) array. For mapping the whole region, the 12m and 7m mosaic observing mode was adopted. Each mosaic covered a region of $198'' \times 71''$, spacing by $12.9''$ and $22.0''$ for 12m and ACA, respectively (Fig. 2.4). 109 of 12m- and 38 of 7m-pointing mosaics were obtained, and in total a data cube of size $4' \times 4'$ was generated. The center of the mosaics was set to $\alpha(\text{J2000})=15\text{h}53\text{m}08\text{s}$, $\delta(\text{J2000})=-54^\circ36'54''$ for all bands. The ALMA correlator was configured to cover specific frequency ranges within the ALMA band. In order to be sensitive to the continuum emission and spectral lines, four spectral windows (Table 2.1) were tuned at frequencies of specific molecular transitions of dense gas and shock tracers. A total observation time of 32 hours was counted.

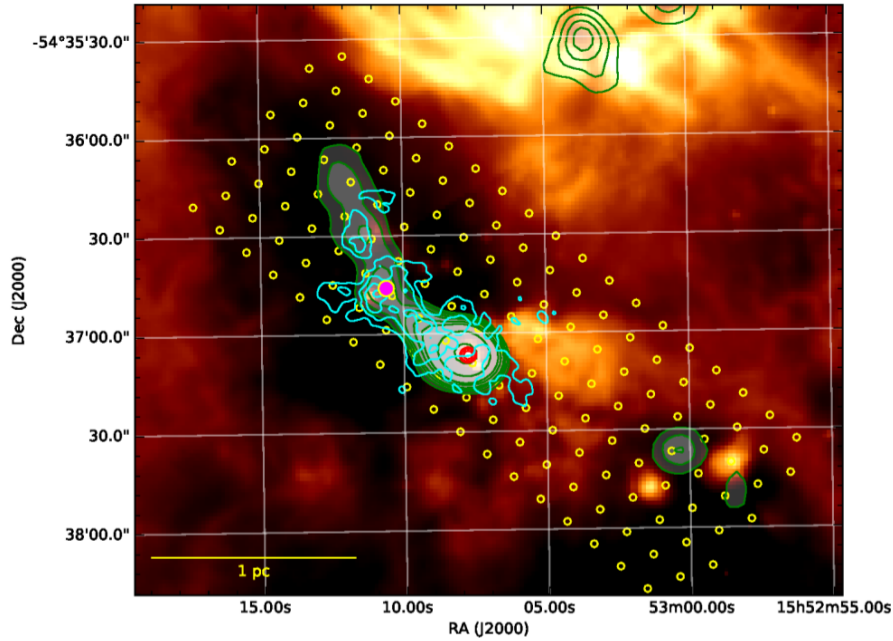


Figure 2.4: The ALMA Mosaic observing regions shown in yellow circles. Background: $8 \mu\text{m}$ GLIMPSE map. Grey scale with green contours: $450 \mu\text{m}$ APEX map. Red contours: ATCA 3 mm continuum map, depicting the hot core. Light blue contours: N_2H^+ (1-0) map, whose peak is marked by a magenta dot showing the cold core.

Table 2.1: Spectral windows setting in ALMA band 6

Spw	Center Freq MHz	Channels	Bandwidth MHz	Resolution MHz	Vel. Res. km/s
1	216418	3500	1709	0.48832	0.677
2	218420	1740	1700	0.97664	1.341
3	231293	2840	173	0.06104	0.079
4	230538	3500	1709	0.48832	0.635

The ALMA data were calibrated using the ALMA calibration pipeline available in CASA version 5.4.0. The calibration procedures solve for flux and gain solutions as well as for the bandpass response. The images were produced from the calibrated visibilities using the task TCLEAN, with the multi-scale synthesis option. The 12m-array-only images have a final restored beam of $2.2'' \times 1.8''$ (original beam size), equivalent to a linear resolution of about 6500 au. For each data cube, we created a continuum image file and a line-only data cube using the STATCONT method (Sánchez-Monge et al., 2018). In order to recover spatial scales that are filtered out in the 12m data set, the continuum images were combined with the ACA 7m data and with single-dish observations from the ATLASGAL survey (hereafter 12M+7M+TP continuum image). The three data sets were merged using the FEATHER task within CASA, after taking into account that the flux of ATLASGAL at $870 \mu\text{m}$ has to be rescaled to the frequency of the ALMA observations at 230 GHz (or 1.3 mm). For the line-only data cubes, only the 12m and 7m data are available, and they were merged together using the task FEATHER. The combined data cubes are later used to study the spectral lines and to carry out our analysis.

2.3 Calibration

Prior to creating an image product by applying an inverse Fourier transform to visibility data, some basic calibration should be done to remove unwanted effects, e.g., instrumental response, terrestrial interference, atmospheric fluctuations, and poor quality data. There are several steps in calibration.

The first calibration applied to the data comes from the long-term variations of the telescope system, i.e. the positions of antennas and the instrumental settings in the period of weeks or longer. In most cases, this is not required to be done by the user since it is calibrated prior to the output of observation.

The second set of calibration are to account for effects that are short-term, e.g., the atmospheric variations during one observation. This stage of calibration will correct the system temperature change with elevation and weather, since the telescope beam is pointed through a varying amount of atmosphere by time. Hence, calibration sources are used to measure the interferometer response during an observation, which should satisfy four "selection rules". To get an efficient high signal-to-noise response, they are supposed to be bright over the required bandwidth of the receiver. To reduce the time of adjusting the telescope pointing direction and obtain a correlated observing effect, they should be close to the target science source, especially in elevation (Thompson et al., 2017). The third requirement is to be an unresolved point source, to avoid precise visibility measurement and facilitate the reconstruction of their signal. Additionally, these sources are required to display stable positions and fluxes within an observation track, rather than quick variation during a certain period.

The interferometer response to the calibrator is treated as a point spread function (PSF),

which is applied to the raw data from the correlator before creating the final intensity map. Since the interferometer does not fully sample the uv-plane, the final beam deviates from a simple Gaussian and can be very complex. This real beam is essential to reconstruct the real visibilities both in amplitude and phase.

As described previously, the complex visibility is calibrated in four stages in real practice (Burnard F. Burke, 2002).

- Data inspection. The first step is to identify and remove bad data. If such data are not flagged, it will be used to incorrectly clean the image, resulting in significant artifacts. Therefore, we carried out some initial inspections by plotting: (1) the amplitude and phase against time, as there should not be sudden jumps but rather smoothly distributed; (2) the system temperature against elevation and time, and again jumps indicate bad data with excess noise; (3) amplitude against frequency, in which spikes from terrestrial interference should be removed. In general, clear discontinuity should be flagged throughout the calibration process.
- Bandpass calibration. The response of the receiver system is often dependent on frequency, which stems from the unequal voltage measurement in different bandpass channels. We used the calibrator with a bright and flat continuum spectrum over the bandpass, to correct the receiver variation, especially the reduced response at edge frequencies.
- Flux calibration. Since the flux standards are usually planets, which are seldom unresolved or located close to the target, the moderate-strong nearby flux calibrators are first calibrated against the accurately known flux standards and then applied to the target.
- Phase calibration. Phase calibrators are usually unresolved, so that their visibility can be considered as unity. They are not necessarily to be exactly flux stable, but must be exceptionally stable in position. Phase calibrations are done frequently in an observation, and therefore it is important that the calibrators are as close to the target as reasonably possible. Calibration is performed by flattening the phase response of the interferometer to the calibrator sources.

If a source is extended and cannot be covered by a single pointing, an observation with multiple pointings should be conducted. The technique used to combine these pointings is named *mosaicing*, which can naturally consider the variable primary beam during each pointing and describe the spatial sensitivity across the entire source.

2.4 Imaging

The process from calibrated visibilities to the final image is named *imaging*. It can often be greatly improved through the application of deconvolution algorithms. The most common method is the CLEAN algorithm, which convolves the interferometer beam with the point source responses that are deconstructed from an initial image to produce the final image. The detailed process is as follows according to Thompson et al. (2017):

- Create the initial image through the Fourier transform of total calibrated visibilities. This initial image is named *dirty image* as all the irregularities of the interferometer beam are

included. At the same time, a *dirty beam* image is created from the point response of the interferometer to show its beam pattern.

- Identify the highest intensity point in the dirty image, which is assumed to be a point source with an amplitude of a fraction of the total intensity at this point. By convolving the dirty beam with the point amplitude, the interferometer response towards this theoretical point is calculated, and then the dirty image is subtracted. The position and amplitude of the point source are recorded as a δ function in the *model image*.
- Iterate the previous step to identify all intensity peaks until no significant structure remains on the leftover intensity map, which is defined as *residual image*.
- Convolve the model image with the *clean beam* to create an ideal noise-free map, which represents a full (u,v) sampling by the telescope. The clean beam of the interferometer is usually a Gaussian beam with the FWHM equal to the central response of the dirty beam.
- Add the residual image to the convolved model image to obtain a final map of the source as seen by the interferometer. This allows flux of over-subtraction or residual excess to be included.
- Divide the final map by the antenna primary beam to correct the flux for the edge pixels that have lower sensitivity. This step is optional, but important for the source at the border of the field of view.

There are various modified algorithms towards the original CLEAN method, which have been developed to be suitable for extended source emission. For example, the multi-scale CLEAN models emission as a combination of several circular Gaussians around a δ function (Cornwell, 2008). The newly implemented CASA deconvolution algorithm "Adaptive Scale Pixel (ASP)" (Bhatnagar and Cornwell, 2004) parameterizes the sky brightness distribution into a collection of Gaussians and does a formal constrained optimization on their parameters. The number of flux components is also adjusted as the iterations progress.

2.4.1 CASA version

The basic calibration was performed using CASA version 5.3.0, and then we downloaded the calibrated visibility for further cleaning. From that time on, CASA has been greatly improved to version starting with 6. The major improvements from version 5.x.x to 6.x.x are: (1) A new task `sdintimaging` is available for joint deconvolution of single dish and interferometer data. (2) A powerful new CARTA tool is produced to visualize data. (3) Some minor errors in the `concat`, `gaincal`, `tclean` and its parallel running are corrected. So we expect some deviation of the output when using different CASA version. A simple check is to run a `tclean` with CASA version 5.7.0 and 6.3.0 using the same script. The result is shown in Fig. 2.5, which does not present apparent difference on the first glimpse. Then we check them by subtracting the 5.7.0 result from the 6.3.0 image. The largest difference appears in the hot core (0.03 Jy/beam), while the medium difference is only around 0.1 mJy. The difference is also located in noisy regions (the green pixels in the right panel of Fig. 2.5) rather than along the filament (the white part), indicating the consistency between CASA versions, especially in the cleaning task. To get a good comparison of images in this study, we used the most recent version 6.3.0 for self-calibration and cleaning.

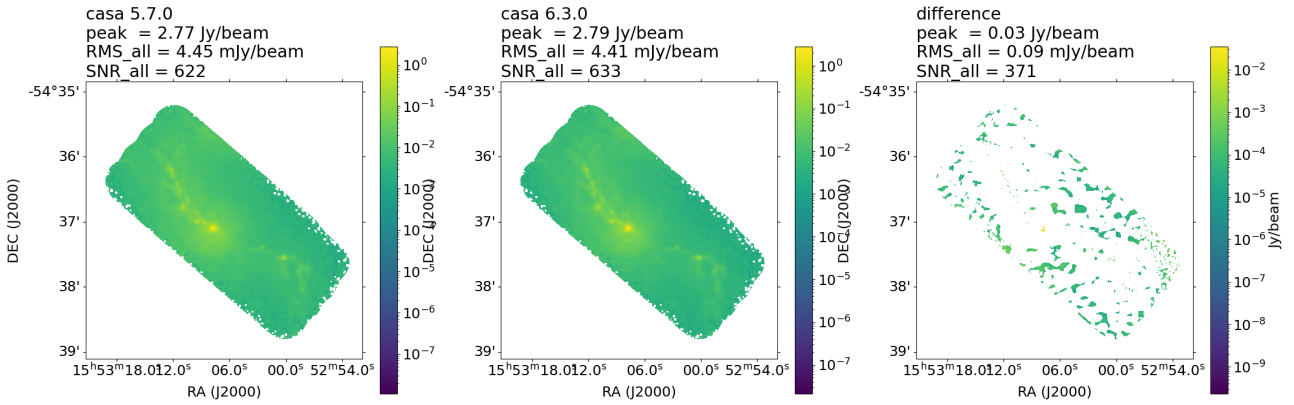


Figure 2.5: Comparing the results of CASA version 5.7.0 and 6.3.0 when using the same parameters and commands. The difference between the two resulting images are not visible by eye, also proved by the subtracted difference image (differ mostly with 0.1 mJy/beam).

2.4.2 Number of iterations

The number of cleaning iterations is one of the main factors determining the reliability of the image. It can be controlled by setting an upper limit or a stopping criterion, such as the noise level of the residual image. Usually more iterations lead to overall fewer artifacts and a higher dynamic range (e.g. comparing (a) and (b) in Fig. 2.6), so it is set to a large number in order to reach a certain sensitivity. However, it can be over-cleaned when too many iterations are conducted, during which some artifacts are taken into the model image, leading to stripes or ring-like structures (e.g. (c) in Fig. 2.6 and the right panel in Fig. 2.7). This commonly happens in high-dynamic-range images, since the PSF of bright spots shows positive and negative parts, leading to a huge error in the subtraction of these bright parts from the residual image. In this case, it introduces artificial positive spots into the model image and then into the final image.

Another factor that functions similar to the number of iterations is the so-called *loop gain*. The loop gain controls the portion of the peak flux to be transported from the residual image to the model image. In consequence, a larger loop gain makes it easier to reach the desired sensitivity as well as the over-cleaned image. We selected a loop gain of 0.1, the same as the default value.

2.4.3 Weighting

When gridding the visibility into cells, a weighting factor w , is multiplied in the data to maintain a balance between sensitivity and resolution at different uv distances. There are various weighting schemes in the CASA `tclean` task. Three important ones are named *natural*, *uniform*, and *briggs* separately.

The natural weightings are calculated to be the inverse noise variance of the specific visibility, which are only the data weights. The resulting sensitivity is the highest among the three weighting methods. However, the inner uv plane is usually much better sampled than the outer part of the uv plane; hence the lower spatial frequencies have higher signal-to-noise ratios, indicating a good-quality low-resolution image.

The uniform weighting gives equal weight to each measured visibility regardless of the sample density. The resulting PSF has the narrowest main lobe and suppressed sidelobes,

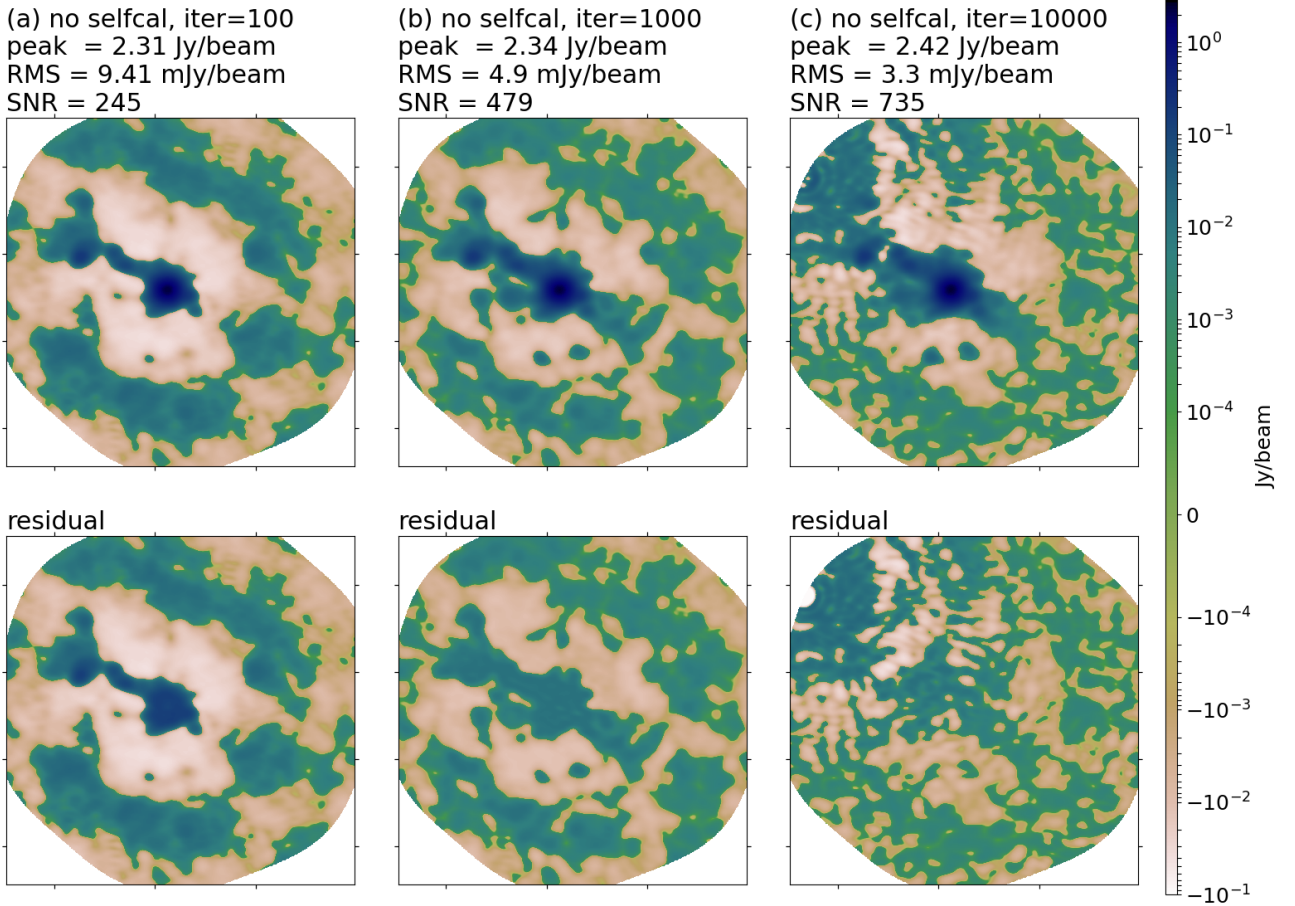


Figure 2.6: The cleaning results of one observed field with increasing number of iterations while other parameters are settled. (a) With 100 iterations, the peak intensity is almost recovered with slight deficiency, while the rms is extremely high and hence SNR is low. The residual map clearly shows the core feature, indicating the number of iterations is insufficient. (b) With 1000 iterations, we found back more peak intensity and less noise than that of (a). The residual map shows filament feature, which still requires more iterations. (c) With 10000 iterations, the peak intensity and SNR are the highest of three tests. However, the residual map already shows ring-like structure at the top-left corner, indicating over-cleaned. In conclusion, the best number of iterations is located between 1000 and 10000, to find which we can set a peak-residual threshold to 4 mJy to have an automatic stop.



Figure 2.7: The cleaning results of all fields with increasing number of iterations while other parameters are settled. We set a very low threshold (1.0 mJy) to allow the cleaning running through all of the given iterations. The most right panel shows significant artifacts at a large number of iterations.

which is best suited for the sources with high SNR to minimize sidelobe contamination between sources. In this case, the data in the densely sampled region are weighted down; therefore, the sensitivity is apparently worse than the natural weighting. One should pay attention to isolated visibility, which can give false weights and result in artifacts in the PSF.

The Briggs weighting (Briggs, 1995), also named Robust weighting, will compromise between natural and uniform weighting according to their SNR and a parameter that defines a threshold. High SNR samples are weighted by sampling density to optimize the shape of PSF, and low SNR visibilities are used for sensitivity optimization. In CASA briggs weighting, there is an essential sub-parameter *robust* which ranges between -2.0 (close to uniform weighting) to 2.0 (close to natural weighting).

In this study, we adopted Briggs weighting and the robust parameter to be 0.5 (same as default), to get a good resolution image as well as large-scale faint emission.

2.4.4 Masking

The masks to be used for deconvolution can be set by the user. Otherwise, there are two masking methods named *pb* and *auto-multithresh*. The method **pb** constructs a mask at a certain gain level of the primary beam. For example, `pbmask = 0.2` means creating a mask at 0.2 pb and excluding the low-gain noisy part from being selected as the model in the tclean process. The method **auto-multithresh** is an auto-masking algorithm with multiple thresholds for deconvolution, which mimics what an experienced user would do when manually masking the images. This algorithm first identifies regions that are either above the given noise level or sidelobe level. The mask regions should be larger than some fraction of the beam size; otherwise, they are removed to avoid spurious peaks within the masks. It will include the adjacent extended emission regions around the initial mask using a process called binary dilation, to count in the flux of low-SNR regions. In the end, the mask is expanded by convolving with a Gaussian and then only regions that have fluxes above a certain value will be retained as the final mask. In this study, we used *auto-multithresh* for the tclean process with its default values².

²https://casaguides.nrao.edu/index.php/Automasking_Guide

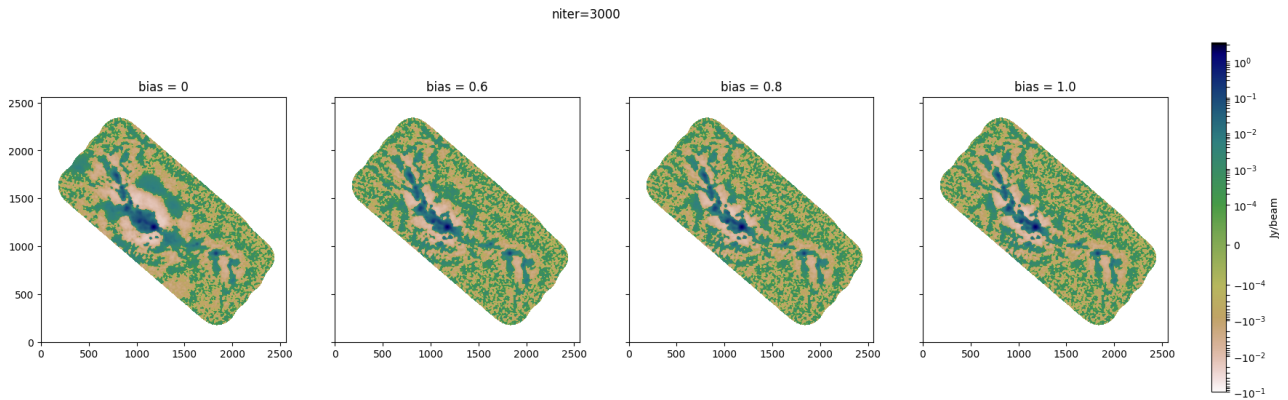


Figure 2.8: The cleaning results of multiscale method with increasing values of `smallscalebias` from left to right, while the other parameters are set to the same (`scales`=[0,20,60], `niter`=3000).

2.4.5 Deconvolution algorithms

The deconvolution refers to the process of Fourier transform, to be more specific, reconstructing a model from the sky brightness distribution when providing a residual image and the PSF structure. There are various deconvolution methods to select from the CASA imaging task, in which the most widely used and introduced here are *hogbom*, *multiscale*, and *mtmfs*. We also briefly introduce the newly released and under-experiment method *ASP*.

The **hogbom** method is a point source model of the sky brightness distribution, which requires the entire PSF to be used in the iterative deconvolution steps (Högbom, 1974). It is adopted as the default method, since it can treat well towards low-dynamical-range images and does not require adjustment. But it produces artifacts when the imaging weights are chosen inappropriately, especially when the PSF has strong side lobes and the image has a large dynamical range.

Unlike the hogbom method that assumes the sky to be an ensemble of point sources that may not be the case for imaging extend sources, the **multiscale** deconvolver regards the emission from the targets as a linear combination of the emission on different scale sizes. Two sub-parameters *scales* and *smallscalebias* can be set. The one or more integers in the scale parameter are multiplied by the beam size to be used as the scales in the imaging process. A reasonable starting point for setting the scales is [0, beam, 3*beam], where 0 stands for the unresolved sources, and the largest scale size should be similar to the large-scale features. The parameter *smallscalebias* determines how much weight is given to small scales, ranging from 1 (large scale weighted to zero) to -1 (small scale weighted to zero), while the default value is 0 (weights all scales equally). If the residual image shows negative regions, it means over-cleaning in the large scale, and hence a larger value of *smallscalebias* is suggested. For our 12m high resolution data, we tested different values for *smallscalebias* parameter (see Fig. 2.8). When the *smallscalebias* is set to zero, there are significant missing fluxes (negative bowl) surrounding the filament, and over-cleaning (positive bowl) as well. This phenomenon is weakened as the value of bias increases. In another word, a large bias value helps to strengthen the small-scale structures (which are what we need in the study) and suppress large-scale artifacts. Therefore, we adopted *smallscalebias*=1.0 when using multiscale deconvolver in the following study.

When there is a wide frequency range of line-free visibilities, the shape of the continuum emission needs to be taken into account. Only in this case, the method **mtmfs** (multi-term multi-frequency synthesis) is adopted, in which the selected visibility data are averaged with

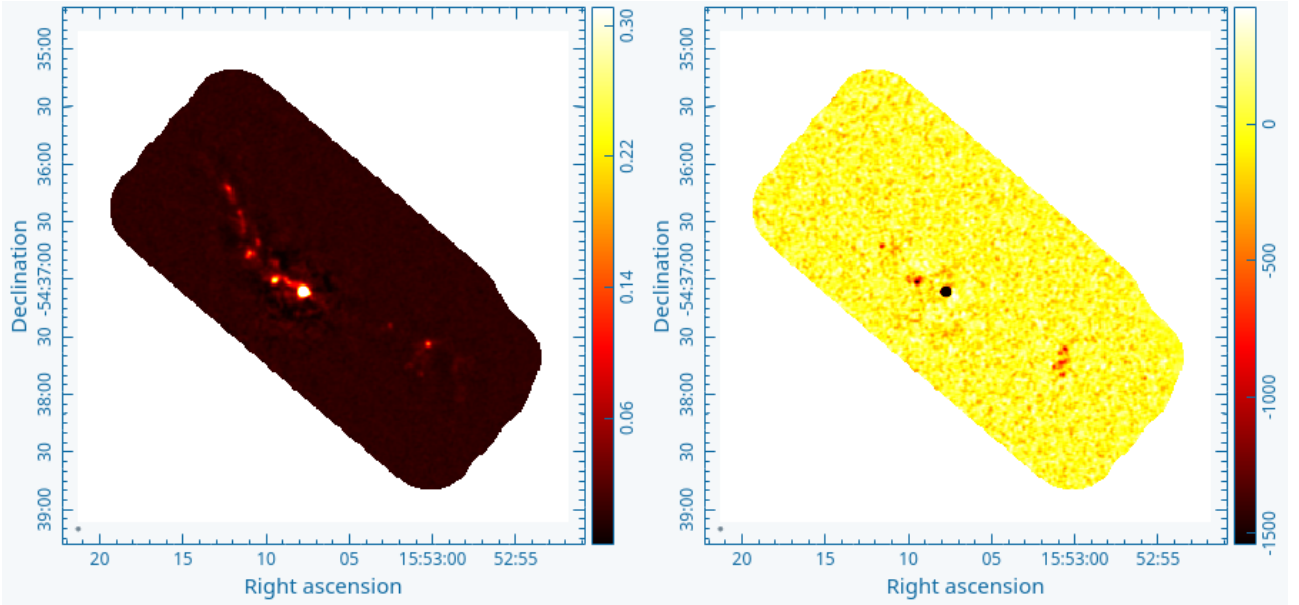


Figure 2.9: The cleaning results of mtmfs method. Left panel: $tt0$ image, showing the flux at the central frequency of the band, similar as the normal `tclean` result with parameter `spec-mode='mfs'`. Right panel: $tt1$ image, showing the spectral index, which is suspicious due to high deviation from zero.

polynomial weightings during uv gridding. The sub-parameter *nterms* defines the order of polynomial, which is considered to be a Taylor expansion of the real SED at a certain frequency. A common practice is a power-law SED, corresponding to *nterm*=2. The resulting images are the zero-order image $tt0$ which is the flux at the central frequency of the band, and the first-order image $tt1$ representing the spectral index. Examples of the $tt0$ and $tt1$ images are shown in Fig. 2.9. However, our observation covers a frequency range of almost 5 GHz, and the uv coverages change with frequency. Thus, the visibilities are actually sensitive to different spatial scales in different frequency, leading to a complex spectral index map rather than that in the $tt1$ map (see right panel in Fig. 2.9). This method works better when having a broadband of spectra and is useful in generating a good continuum map.

The recent implemented algorithm **adaptive scale pixel** (ASP) (Bhatnagar and Cornwell, 2004) in CASA shows better performance than the previous methods. It parameterizes the sky brightness distribution into a series of Gaussian functions and conducts an optimization according to the location and extension of the peak residual at each iteration. The final residual image usually looks more homogeneous and hence the cleaned image contains more flux as well as faint features (see Fig. 2.10 for an example). This method can be used for both the continuum and the spectrum. However, the produced continuum images do not have a significant difference by eye, and the processing time of ASP is twice that of the multiscale method. Considering the fact that computationally expensive and its performance, it is not recommended to use this method for a large interferometer observational data set.

2.5 Self-calibration

The sensitivity of interferometric images is often limited by residual calibration errors. Here, we recall the standard calibration process mentioned in Sec. 2.3. The usual phase calibration targets on a reference point source for a fixed time interval, which has the disadvantage of losing

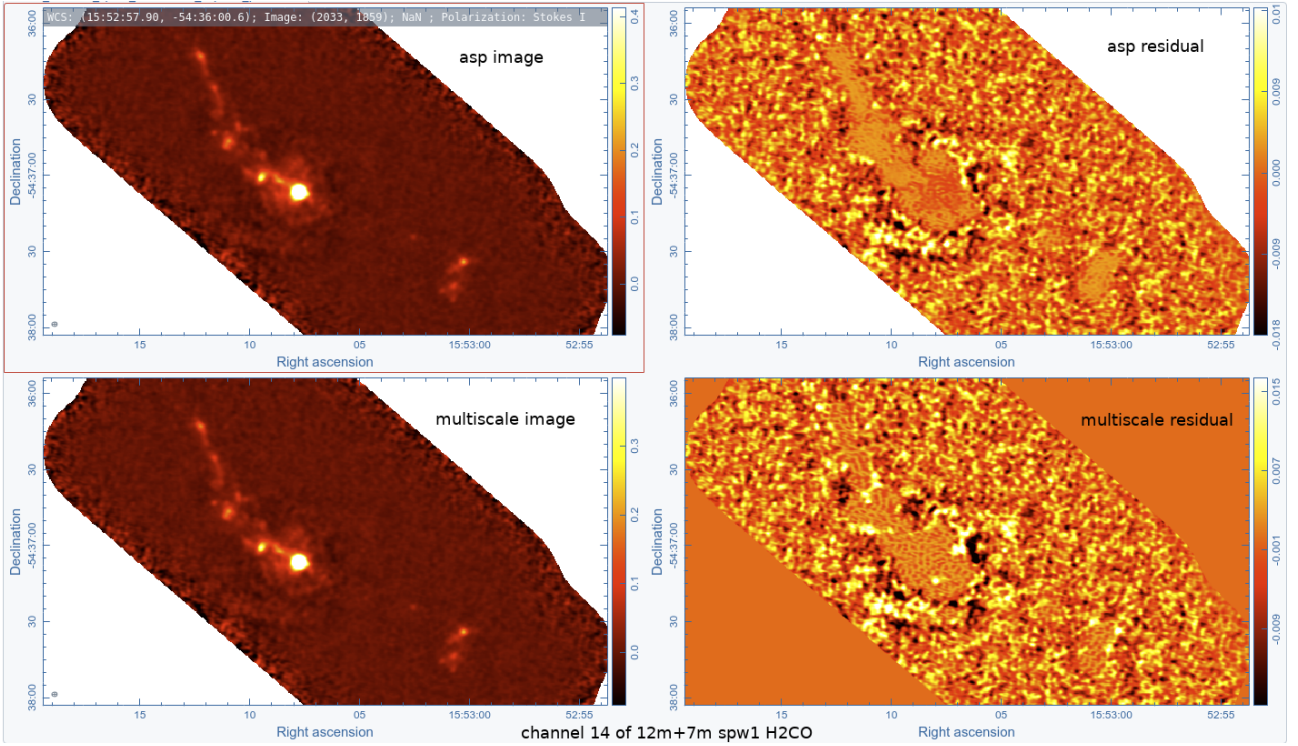


Figure 2.10: Comparison of the cleaned images and corresponding residual images with two different algorithms. Upper panels: ASP method shows slightly better result since the residual image is more homogeneous and contains less emission from the filament. Lower panels: results of MTMFS method. The residual still shows some emission at the central core, which is not preferred.

observing time and failing to track the rapidly changing phases between the time interval. In addition, the usual calibration cannot deal with artifacts (i.e. negative bowls in continuum, fake-absorption lines in spectra) that come from bright, dominating sources. This problem can be partly solved by a method named **self-calibration** (hereafter "self-cal", (Pearson and Readhead, 1984)). To better calibrate the complex gains (G_{ij} in Equation 2.11, including phases and amplitudes) and utilize the on-source time, self-cal uses the science target itself as a model to calibrate the data. The precondition of applying phase self-cal is that SNR should be larger than 3 for a single antenna and solution time interval, and at least 10 for amplitude self-cal. Or, as a rule of thumb, self-cal is worth a try if SNR is larger than 20 for ALMA data.

As we notice, G327 presents a prominent hot core with a peak flux density of 2.34 Jy/beam at the 12m-array continuum image (shown in panel (a) in Fig. 2.14). Compared to the residual noise clipped with sigma ($\sigma_{image} = 4.90$ mJy/beam) of the cut-out region of the seven mosaic fields that covered the hot core (see the left panel of Fig. 2.11), the resulting high signal-to-noise ratio (479) suggests a unique opportunity to apply self-cal, since the hot core itself can be used as a model to calibrate the data and hence recover a more reliable continuum.

To examine the validation of self-cal in detail, we converted the noise (σ_{image}) to a single antenna and solution (σ_{self}) with the following equation (Liu et al., 2010):

$$\sigma_{self} = \sigma_{image} \times \sqrt{n-3} \times \sqrt{\frac{t_{on-source}}{t_{solint}}}, \quad (2.13)$$

where n is the number of antenna, t is the on source time and single interval time. Knowing that 40 plates of 12m array are used, the on-source and solution interval times for each field

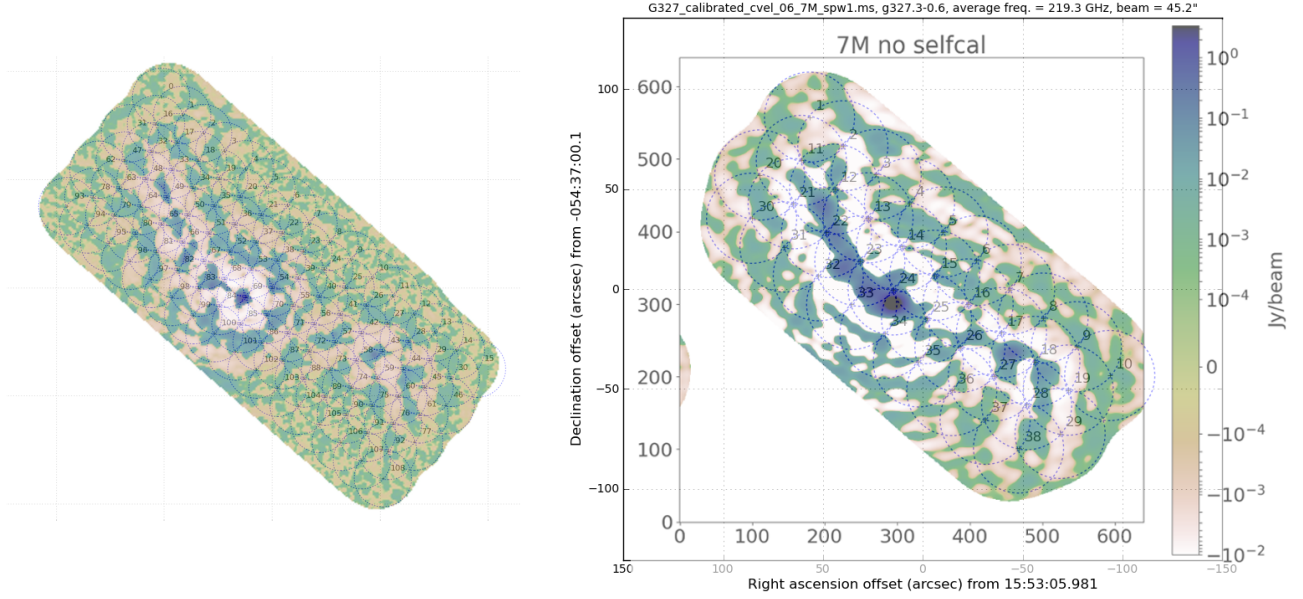


Figure 2.11: **Left:** The 12m continuum dirty image overlaid by the 109 observing fields. The 7 pointing fields that cover the hot core are No. 53, 54, 68, 69, 70, 84, 85, which are used for self-cal. **Right:** The 7m continuum overlaid by the 38 fields, in which field No. 14, 15, 23, 24, 25, 33, 34 are used for self-cal.

were 24.08 s and 6.02 s, a SNR_{self} ($=\frac{I_{peak}}{\sigma_{self}}$) of around 39 was derived, which far exceeded the minimum requirement of 3 for phase self-cal and 10 for amplitude self-cal.

2.5.1 Phase calibration

To begin with, we tested phase calibration on the continuum imaging process. We selected 386 line-free channels (278 MHz) in the four observed spectral windows (see Fig. 2.12) to derive the continuum. And we separated the visibility data of the seven central fields that covered the hot core into a new file.

The first step of self-cal was to make a shallow clean with only 100 iterations and save the model (panel (b) in Fig. 2.14). The second step was to run an initial phase self-cal with a long solution time interval, i.e. infinite. We found that the phase change over time was non-trivial, though not enormous, with a range of a few 10s of degrees (upper panel in Fig. 2.13). Then we applied this phase self-cal to the line-flagged measurement set and ran cleaning with 1000 iterations, which improved SNR from 245 to 788 (comparing panel (b) and (c) in Fig. 2.14). The next step was with the shortest solution time interval, i.e. per integration (solint=int in CASA gaincal task). We applied the int phase self-cal to the first inf phase self-cal modified measurement set and ran cleaning with 1000 iterations, resulting in a marginally improved SNR from 788 to 802 (panel (d) in Fig. 2.14). At the same time, the corresponding phase distribution was almost perfectly centered around zero over time (lower panel in Fig. 2.13), therefore we do not expect further phase self-cal to improve the results.

With the calibration table produced in each phase-cal step, we applied the gain calibration towards the original visibility files with all the fields. In detail, similar things have been done

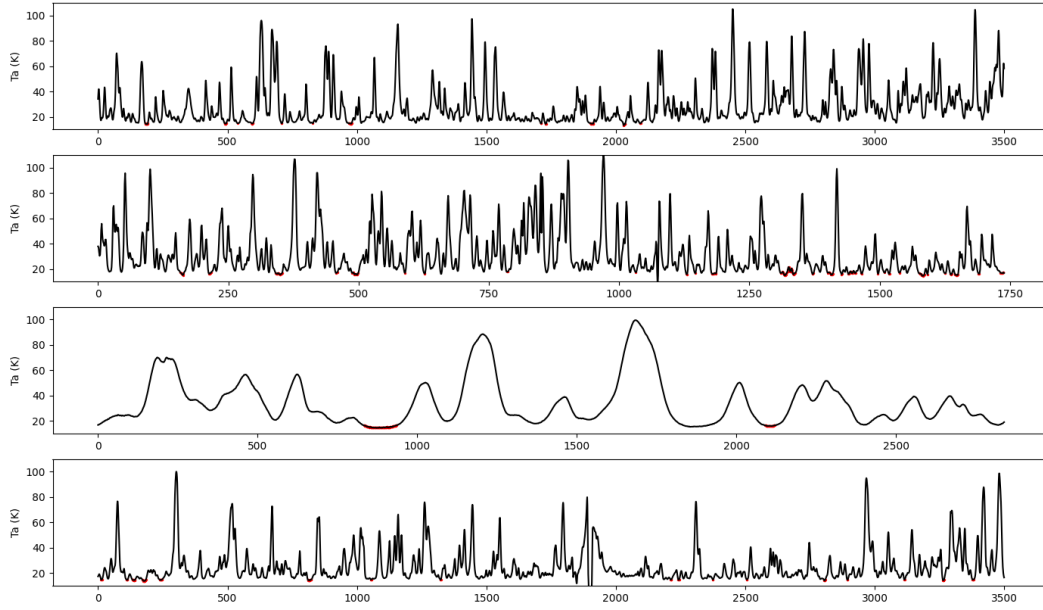


Figure 2.12: The line-free channels that used for continuum imaging marked as red dots on top of the hot core spectra. In total, we adopted 386 channels (278 MHz) in the four observed spectral windows.

- first apply the infinite-time phase-self-cal table and ran 1000 iterations, then apply the per-integration-time phase-self-cal table to run a cleaning task. The results are shown in Fig. 2.16. To evaluate the effect of self-cal within the whole field of view, we drew a random box region to derive the RMS and SNR. The RMS and SNR of the entire field were also calculated for comparison. We have seen that the RMS decreases after applying self-cal, and decreases further when applying deeper and shorter time interval self-cal. We note that the "decreasing" behavior of RMS of the box region agrees well with that of the whole field, though the RMS of the box region is always higher, which is due to fewer pixels in the box region and the uncovered low-noise hot core region.

The primary beam-corrected images based on the image at the same column are also shown in Fig. 2.16. The SNR is worse than the non-pb-corrected one because the noise at the border is enlarged. In general, the SNR always increases after the phase self-cal, no matter how to count the SNR and whether to apply pb correction.

2.5.2 Amplitude calibration

The next step of self-cal is amplitude self-cal. We used the longest time interval "inf" to measure the amplitude. The most worrying situation, the change in source characteristics, did not appear or at least cannot be recognized by the eye. With the confirmation of reliability, the SNR was again improved to 889 (see panel (e) in Fig. 2.14). One more try of "int" amplitude self-cal was tested, but the resulting SNR decreased, hence we stopped with two phase and one amplitude self-cal. In conclusion, complex gain self-cal improved SNR by a factor of around two (compare the SNR in panel (a) and (e) in Fig. 2.14).

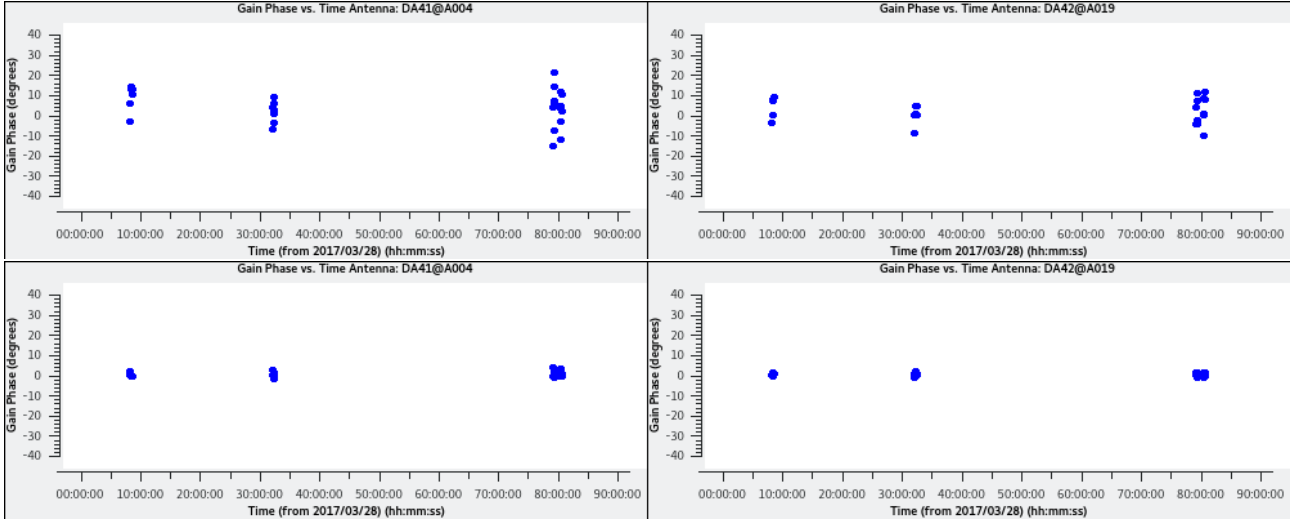


Figure 2.13: Phase distribution over time of two example antennas DA41 and DA42. Upper panel: Applying a phase self-cal with infinite solution time (solint=inf). Lower panel: Applying a second phase self-cal with the shortest solution time (solint=int). The phase was significantly improved to be centered around zero after phase self-cal.

In the end, we performed cleaning on the self-cal modified data until reaching the peak residual stopping threshold of 4.0 mJy, which leads to a final SNR of 1369 (see panel (f) in Fig. 2.14)).

For 7m-array data, seven central fields are valid for self-cal (see the pointing and beam distribution in the left panel of Fig. 2.11). Then two iterations of phase self-cal and one amplitude self-cal were made, and the resulting SNR improvement is around 2 times as well (see 2.15). The final best SNR of the 7m data is 1134, comparable with the 12m data.

The improvement factor of two is also found in Ginsburg et al. (2022), while some other works (Beuther et al., 2020; Ginsburg et al., 2022) also derived a factor of three. There are some reasons that our datasets cannot be further improved by self-cal. First, G327 was observed on three separate nights and in total four scans were taken for each mosaic field. Considering different complex gains throughout the nights, we cannot combine different scans of a certain field when generating self-cal tables. Naturally, we noticed that adapting a solution time interval of per integration led to a better SNR correction than using the infinite time, which is expected from the theory. We also point out that the remaining mosaic fields which did not include bright sources such as the hot core were not applied with self-cal.

2.6 Data combination

The interferometer has the advantage of a high resolution and smooth baseline. One configuration of the interferometer is sensitive to a specific range of spatial scales according to the dish sizes, numbers, and observed frequency. Therefore, using various dishes with different sizes or configurations to do interferometry is natural to collect all the emission of extended sources Wilson et al. (2013). In other words, one configuration usually misses part of the uv coverage. A second configuration of the same array in a more compact/sparse configuration or a combination of another array is then a necessary supplement to obtain the total emission of desired spatial scales. A single dish can provide the zero-spacing information and therefore the largest spatial scale. The detailed steps of data combination are explained as follows.

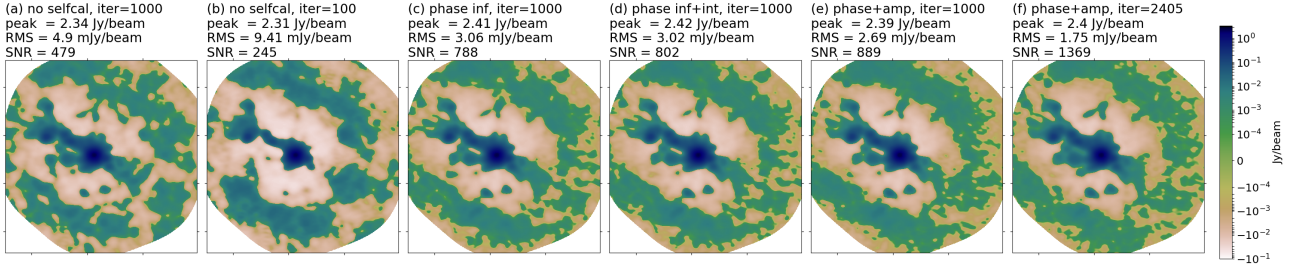


Figure 2.14: The central 7 fields of 12m-array cleaned images before and after self-cal. The RMS was generated from the residual map. (a) The initial continuum produced by 1000 cleaning iterations with line-flagged measurement set. (b) Same as (a) but with 100 cleaning iterations. (c) The continuum produced by one phase self-cal with infinite solution time and 1000 cleaning iterations. (d) The continuum produced by applying solint=int phase self-cal on (c) and 1000 cleaning iterations. (e) The continuum produced by applying amplitude self-cal on (d) and 1000 cleaning iterations. (f) The self-calibrated continuum when reaching cleaning threshold of 4.0 mJy.

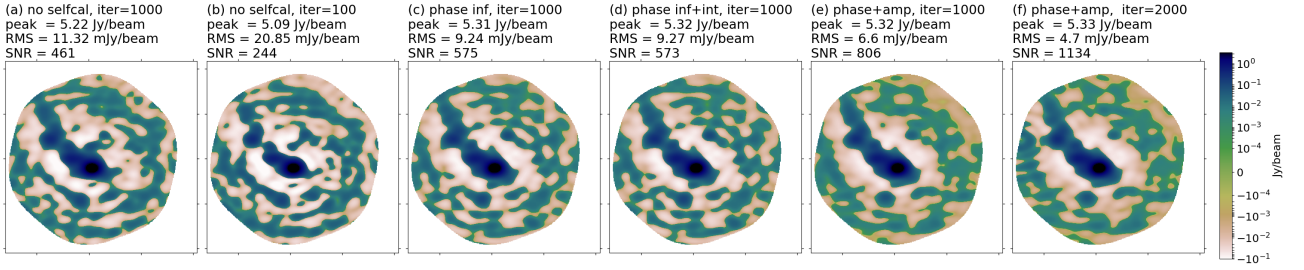


Figure 2.15: The 7m cleaned images before and after self-cal. The RMS was generated from the residual map.

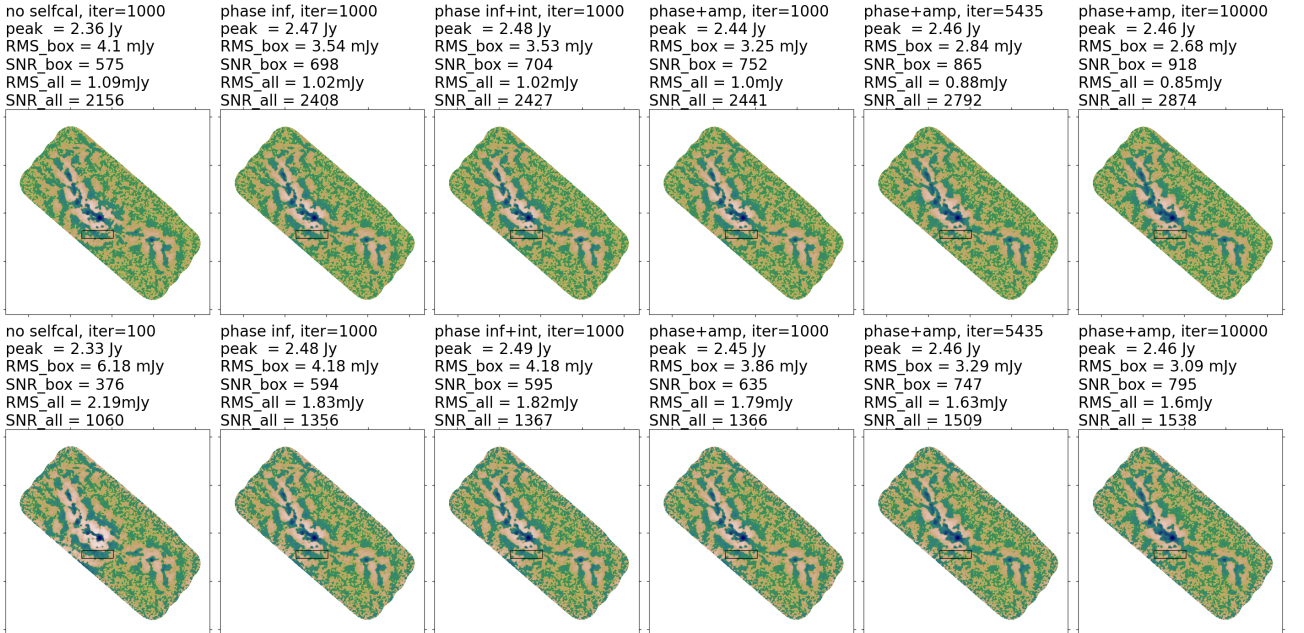


Figure 2.16: Upper panels: the cleaned continuum images with all the mosaic fields using different calibration methods and iterations. The peak intensity, RMS, SNR within a box region and the whole image are listed above the the corresponding image. Lower panels: the primary beam corrected images based on the image at the same column. The SNR is worse than the non-pb-corrected one because the noise at the border is enlarged.

There are mainly two aspects to combine the emission from different arrays, either before deconvolution in the uv-plane or after deconvolution in the image-plane. Some other methods (e.g., MACF, SDINT, FSSC) have been tested recently as well (Plunkett et al., 2023). A quick conclusion is that the recovered flux is always closer to the real flux when using data combination than without any total power data. The least computationally expensive method is to combine the images after separate CLEAN. A technique named **feather** is implemented in the CASA toolbox to achieve this. The feather algorithm is implemented as follows:

- Regrid the low-resolution images to match the coordinate system, pixel size, and image shape of the provided highest-resolution image.
- Fourier transform each image onto uniformly gridded spatial frequency axes (i.e. the uv plane).
- Scale the low-resolution images in Fourier space by the ratio of the volumes of the clean beam, which is defined as a relative weight wt towards the highest resolution data. Scale the Fourier transformed high-resolution image by a factor of $(1-wt)$.
- Add all the images in Fourier space and transform back to the image plane.

Here, the joint image and its PSF are created by adding the separate images and PSFs together, hence it requires well-defined beam shapes for all the images, specifically in the fits header. It is also supposed to have the same flux density normalization scale in Jy/beam rather than Kelvin. Meanwhile, different observations should have a comparable noise level and no artifacts in any of them. Otherwise, the scaling factor between different arrays is not understood straightforward and acts as a trade-off between the flux correction and sensitivity (i.e., low noise maintenance). In the combination of continuum, it does not calculate through multiple channels to derive the continuum and already starts with the good-quality continuum images. Therefore, it is quick, easy to control the quality, and can avoid significant artifacts.

Another approach is to first combine the data in the uv-plane to create a single visibility file of all the arrays and subsequently apply **Joint deconvolution** to derive a joint image. For single dish, pseudo-visibility can be synthesized from the image, which is then applied a constant weighting factor to be continuous within the Fourier space with the interferometric data. For more detail, the factor is the ratio of the low-resolution over high-resolution image at the overlap region in the spatial frequency (Fourier) domain. This step is currently limited to human inspections to define the weight. And a drawback is the artifacts such as sharp discontinuity beyond the edge of the image due to unsampled scales of the source structure and hence inaccurate interpolation.

These two methods are examined and compared in the following sections.

2.6.1 Continuum of ALMA 12m and 7m-array

Before combining the 12m and 7m visibility data, we first examine the 12m continuum generated in different observed frequency windows, in order to verify the line-free channels and the reliability of the final results. From Fig. 2.17, the cleaning results of four spectral windows show a slight difference in the peak flux intensity. The peak fluxes are known to increase with frequency as $I \propto f^\alpha$. We fit the power-law index α to be 1.76, close to 2 from the black body. To

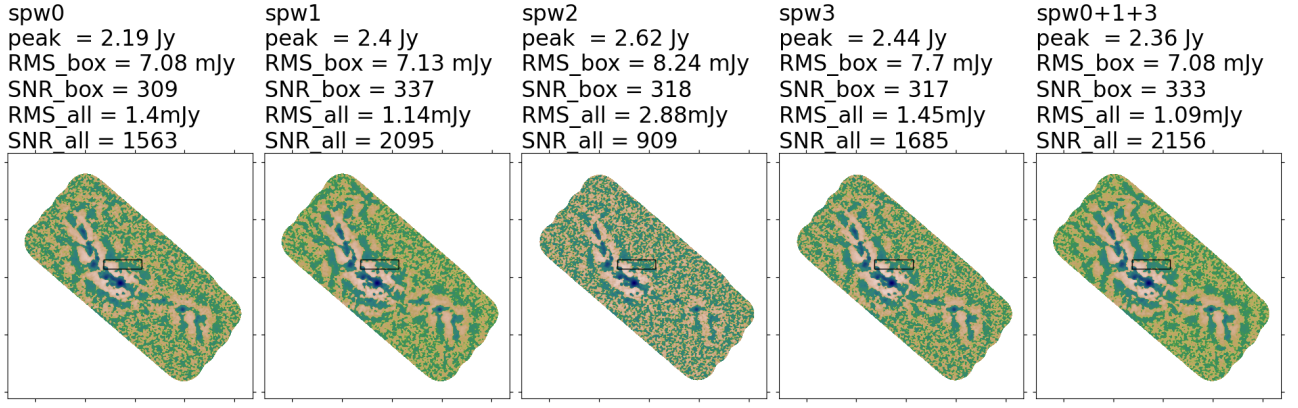


Figure 2.17: Cleaning result without selfcal. Two rms were derived: RMS_box counted the standard deviation (std) within the selected box, RMS_all was the 3-sigma-clipped std.

gain insight into the numbers, we can perform a simple calculation. Considering spw0 centered at 216.5 GHz and spw3 at 230.5 GHz and a black body function in the Rayleigh-Jeans approximation, $\frac{I_0}{I_3} = (\frac{f_0}{f_3})^2$. Hence $I_3 = 2.19 * (230.5/216.5)^2 = 2.48$, very similar to the observed spw3 peak intensity 2.44. But for the peak fluxes of spw1 and spw2, they are below the black-body expectation, indicating a modified black-body with a lower spectral index. Then we calculate the RMS according to $T_{rms} = \frac{T_{sys}}{\sqrt{\Delta t \Delta \nu}} \propto (N_{channel} * \Delta \nu)^{-0.5}$. Consider RMS_all for spw0 and spw2, $RMS_2 = RMS_1 * \sqrt{\frac{N_1 * \Delta \nu_1}{N_2 * \Delta \nu_2}} = 1.4 * \sqrt{\frac{71 * 0.48832}{134 * 0.0610395}} = 2.88$, which is the same as the result produced. However, since the spw2 is a narrow band and is 8 times higher in spectral resolution than the other spectral windows, we only adopted the other three bands for a continuum. Finally, the spw0+1+3 continuum gave a reasonable peak flux and rms.

First, we test joint deconvolution. We used the *concat* task in CASA to combine the self-calibrated visibility files of the 12m and 7m array. Then the same line-free channels indicated in the previous paragraph were selected to produce a combined image. The result after primary beam correction is shown in the left panel of Fig. 2.18. The resulting continuum image shows the nice filament and no obvious artifacts. We derived the 3-sigma-clipped standard deviation throughout the image as the RMS, to evaluate the quality of data combination.

To test the Feather method, we first ran cleaning on the same line-free channels of the 12m and 7m self-calibrated visibility files separately. Then we combined the primary beam corrected 12m and 7m array continuum images together. Since the 12m and 7m data were observed at the same frequency, it is not necessary to apply a scaling factor (*sdf* parameter in the feather task) for one of the data, and the weight was determined automatically by the spatial frequency response of each image. The right panel of Fig. 2.18 shows the Feather result, which looks good and has a higher SNR than the joint deconvolution method. This is because the noise is much lower in the Feather case, though the peak intensity is also slightly lower.

Despite the fact that joint deconvolution is more reliable in the sense of mathematics, the feather task is quicker and produces a higher SNR. We adopted the Feather result for further analysis.

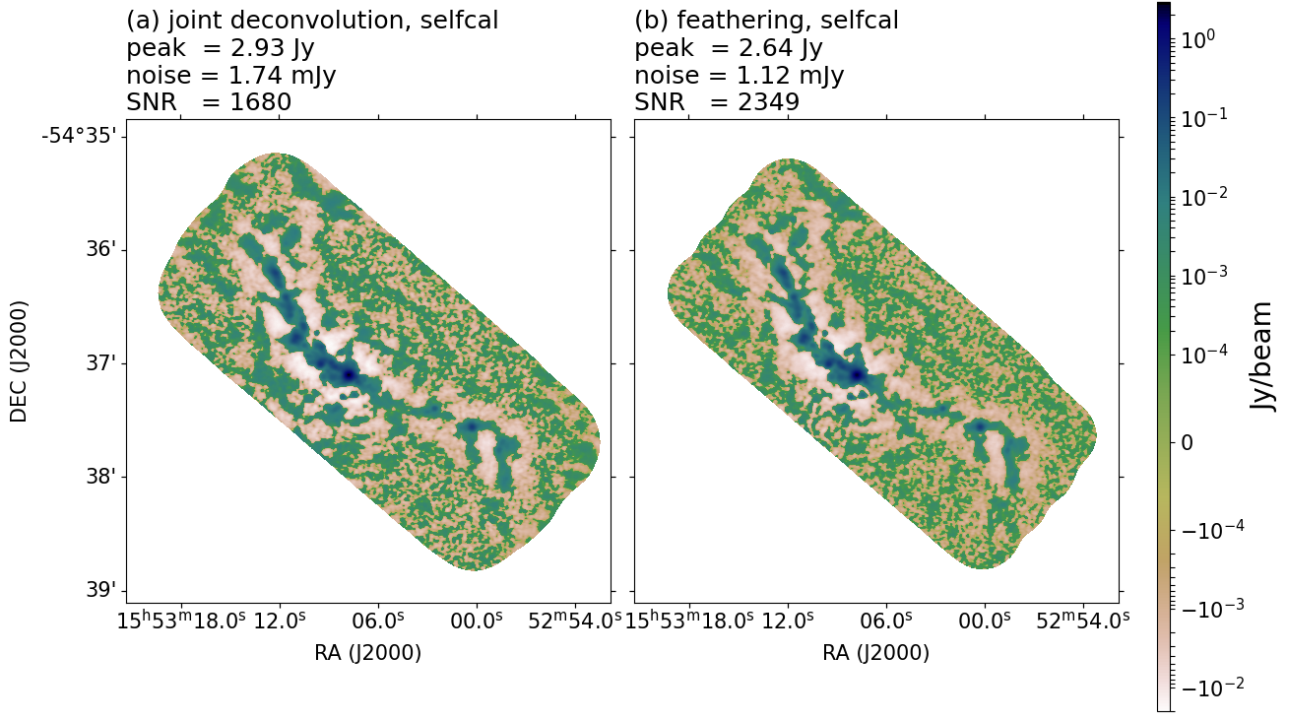


Figure 2.18: Comparing the two data combination methods. Left: Applied joint deconvolution after combining the 12m and 7m visibility files. Right: Applied Feather using the 12m and 7m cleaned images. Both are self-calibrated and primary beam corrected.

2.6.2 Continuum of interferometer and single dish

In order to generate a well-representing continuum image that includes fluxes of both extended emission and compact sources, the data combination of a single-dish image and an interferometer image was applied. For the low-resolution data, ALMA total power observation was not conducted; instead, we collected the 870-micrometer dust emission obtained from the combination of Planck/HFI and APEX/LABOCA data of the ATLASGAL survey (Schuller et al., 2009). Again, two methods CASA Feather and Joint deconvolution are utilized.

To convert the low-resolution intensity to that corresponding to the high-resolution intensity, a single-dish scaling factor (sdf) should be taken to confine the same flux scale. Considering ALMA data was centered on 1.3 mm while single-dish data was at 870 μm , a scaling factor of 0.3 should be applied to the single-dish image, which was derived from the Rayleigh-Jeans approximation of a single-grey-body function, $I \propto f^3$ ($\beta = 1$ following Elia et al. (2017)).

The casafeather visual interface provides the tools to inspect slices of the $u-v$ plane of both the original and weighted deconvolved input images. This is especially useful while setting the sdfactor parameter, which applies a flux scaling factor to the low-resolution image. For the conservation of flux, one would expect the fluxes of the weighted deconvolved high- and low-resolution images to be roughly equal. One can achieve this by altering the sdfactor until the casafeather plots demonstrate this equality. Given the uncertainty in the flux calibration of the SD, we expected that a factor would be required, and we find a satisfactory result with a sdfactor of 0.3. This demonstrates that the calibration factor applied to the APEX data in the first instance was sound and appropriate.

Figure 2.21 shows a comparison of the APEX, ALMA and combined integrated intensity

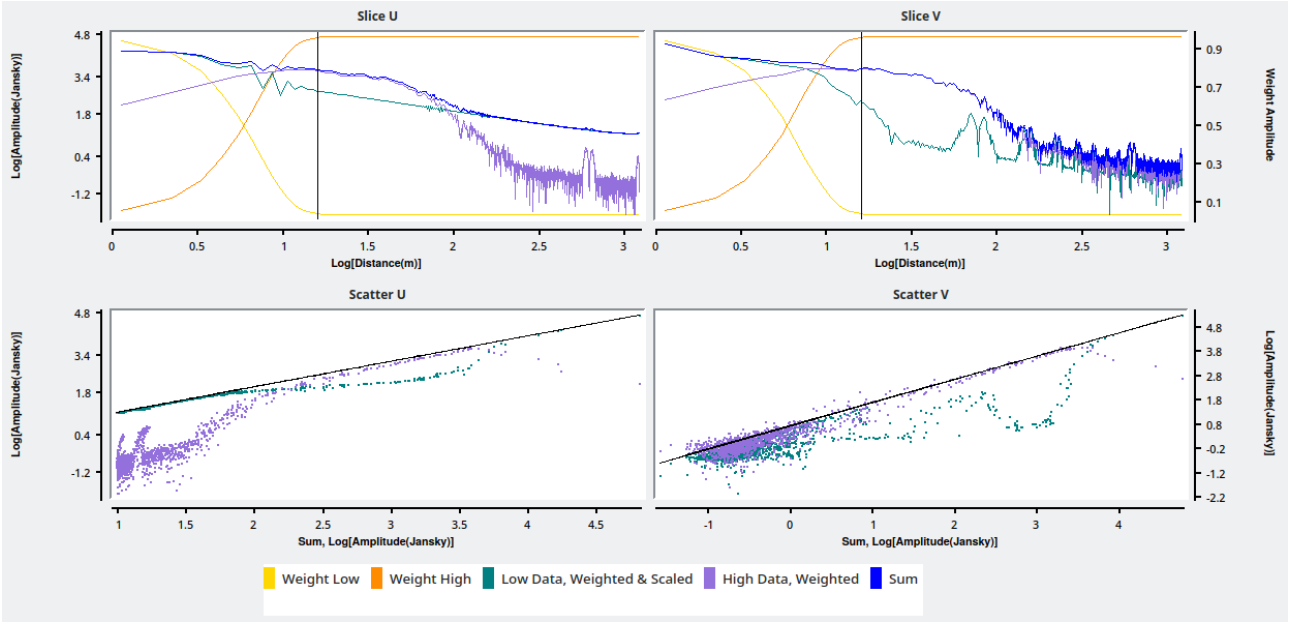


Figure 2.19: The UV distribution when combining data with $\text{sdf}=0.448$. Upper panels: show the data slice which are cuts through the u and v directions of the Fourier-transformed input images. The weight functions are shown in yellow for the low resolution data and orange for the high resolution data. Green is the weighted low resolution data convolved with the high resolution beam and purple the weighted high resolution data. The combined, feathered outputs are shown in blue. The vertical line shows the location of the effective dish diameter, taken from the single dish beam size. Lower panels: The scatter plot of the high data versus the weighted and scaled low data. The feather result is better when the scatters locate closer to the black equality line.

maps of continuum. It is clear that the combined data recover more extended flux than the ALMA-only, resulting in a three times of the ALMA flux on the whole, whilst retaining the small-scale detail and RMS noise of the interferometer. We can clearly see that the negative bowl features, a consequence of missing flux, that surround the cloud structure in the ALMA-only map are mostly eradicated in the combined map and filled by the recovered extended emission.

This factor was examined to be good with the scatter plot (lower panels of 2.19) and the rms count (2.20). The scatter plot, especially in the V direction, shows the tendency of centering around the black equality line and a high overlapping between the two resolution fluxes, indicating a good combination. In addition, the interferometer image contains the actual flux measurements made by the telescope. Therefore, if the single dish scaling is correct, the flux in the high-resolution image convolved with the low-resolution beam and appropriately weighted should be the same as the flux of the low-resolution data convolved with the high-resolution beam once weighted and scaled. When counting peak fluxes for different feather factors, the best match is exactly located at $\text{sdf} = 0.2$ (seen in 2.20). Seen by eye, the image at $\text{sdf}=0.3$ also best presents both the compact sources and the extended emission.

The final combined continuum images have an RMS noise level of $1.28 \text{ mJy beam}^{-1}$ in sigma-clipped signal-free regions and the peak brightness of G327 hot core is $2.73 \text{ Jy beam}^{-1}$, leading to a dynamic range of 2120.

Another way of combining data is to first use the TP2VIS task to transform the single dish continuum image to the visibility file, then combine the visibility data, and then run CASA

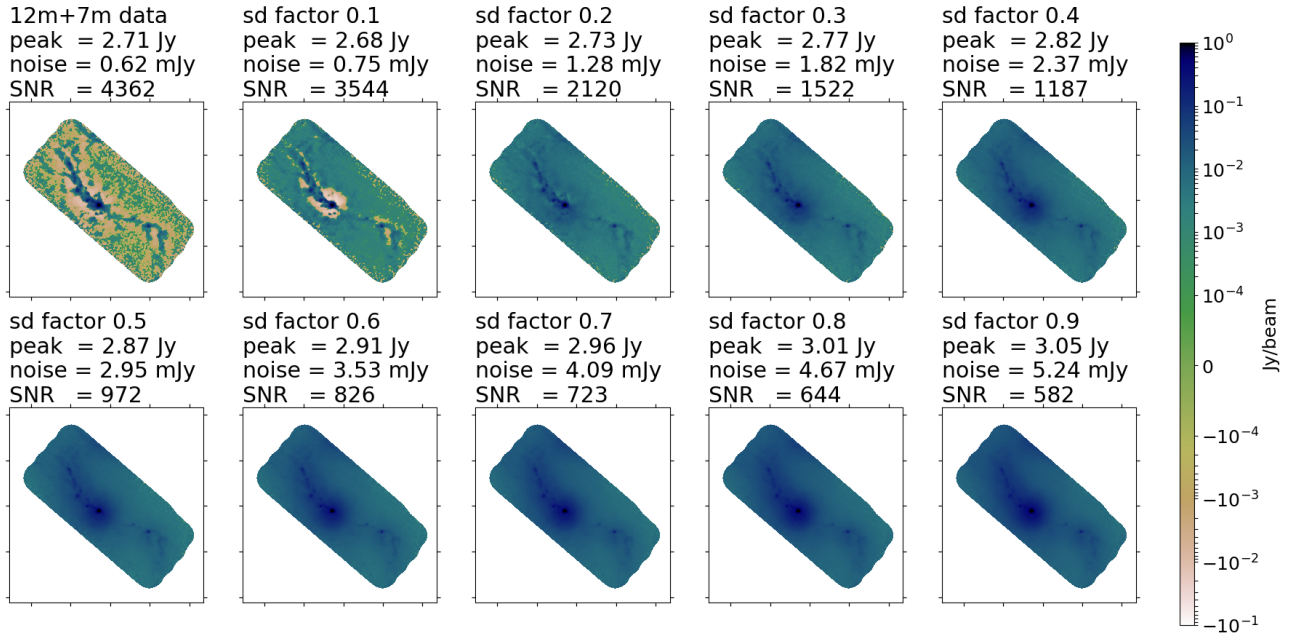


Figure 2.20: Combined image of interferometer and single dish data with different scaling factors. The first image is the combination of 12m+7m data, without total power.

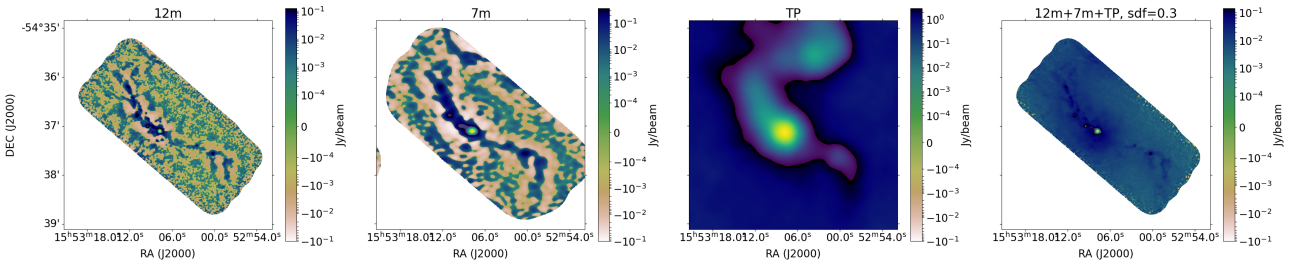


Figure 2.21: A comparison of the 12m, 7m, single-dish image, and the combined image with sdf=0.3.

TCLEAN. TP2VIS is a github script using the method of [Koda et al. \(2011\)](#) and [Koda et al. \(2019\)](#). The steps are listed as follows:

- Make a pointing file out of one of the interferometer visibility files. This is to determine the coordinates and the grid used to create the visibility file.
- Find a reasonable RMS for the single dish image, which is the 3-sigma noise. If one has a spectral cube, the RMS can be estimated from the line-free channels of the cube.
- Apply the TP2VIS function to the single dish data file together with pointing file and noise information. The output is the corresponding visibility file.
- Plot the weight statistics and evaluate the consistency of the weights from different arrays. If the weights of single-dish are several magnitudes lower/higher than that of the interferometer, one should manipulate the weights, e.g., scaling the weights of single-dish to be continuous to the 7m.
- Combine the visibility files of the interferometer and single dish using the CASA concat task. To avoid the inconsistency in previous calibration CASA versions, here the parameter copypointing should be set to False.
- The final step is to do the joint deconvolution using CASA tclean.

In Fig. 2.22, we evaluated whether we can apply the joint deconvolution. In the upper left image, the uv distribution shows a vacancy around the single dish data points (ring-like gap around the red color). The 7m uv coverage (in green color) partly overlaps with the single dish and partly with the 12m-array (in blue). This gap is not symmetric, larger in the v direction than in the u direction, possibly leading to artifacts like negative bowls or stripes. The amplitudes of the three observations agree well, in which the single dish has the highest amplitude, since it traces the strongest component in the short spacing. The weight density distribution through the uv distance is shown in the bottom panels, which is modified to match each other. The original weight density of the single dish data is 10^8 times lower than that of the 7m and 12m data, so we applied this factor to correct the single dish data.

Despite everything looks perfect in the previous steps, the final joint-deconvolved image shows significant artifacts. The most important reason might be the lack and uneven distribution of uv coverage between the 7m and single dish data. [Plunkett et al. \(2023\)](#) suggests using a larger single dish if the data exists, e.g., IRAM 30m rather than APEX 12m, to cover the whole missing shorter spacing of the interferometer. They also encourage the society to build a 50-meter telescope for this specific purpose, and claim the data combination should be a standard process in the data reduction pipeline in order to measure the total flux as accurate as possible.

We did several tests to find other possible reasons. One problem stems from transforming the single dish to virtual visibility, since it generates a fake uv coverage for a single dish data. We reproduced the single dish image with the generated visibility file, which already shows clear ring structures surrounding the center of view (see Fig. 2.23). When the quality of single dish visibility file is not good, the combined visibility will also be heavily affected, same for the cleaned image.

The scaling factor might be the second origin of the problem. We tested different scaling factors from 10^9 to 10^2 to correct the weight distribution. However, the artificial structure

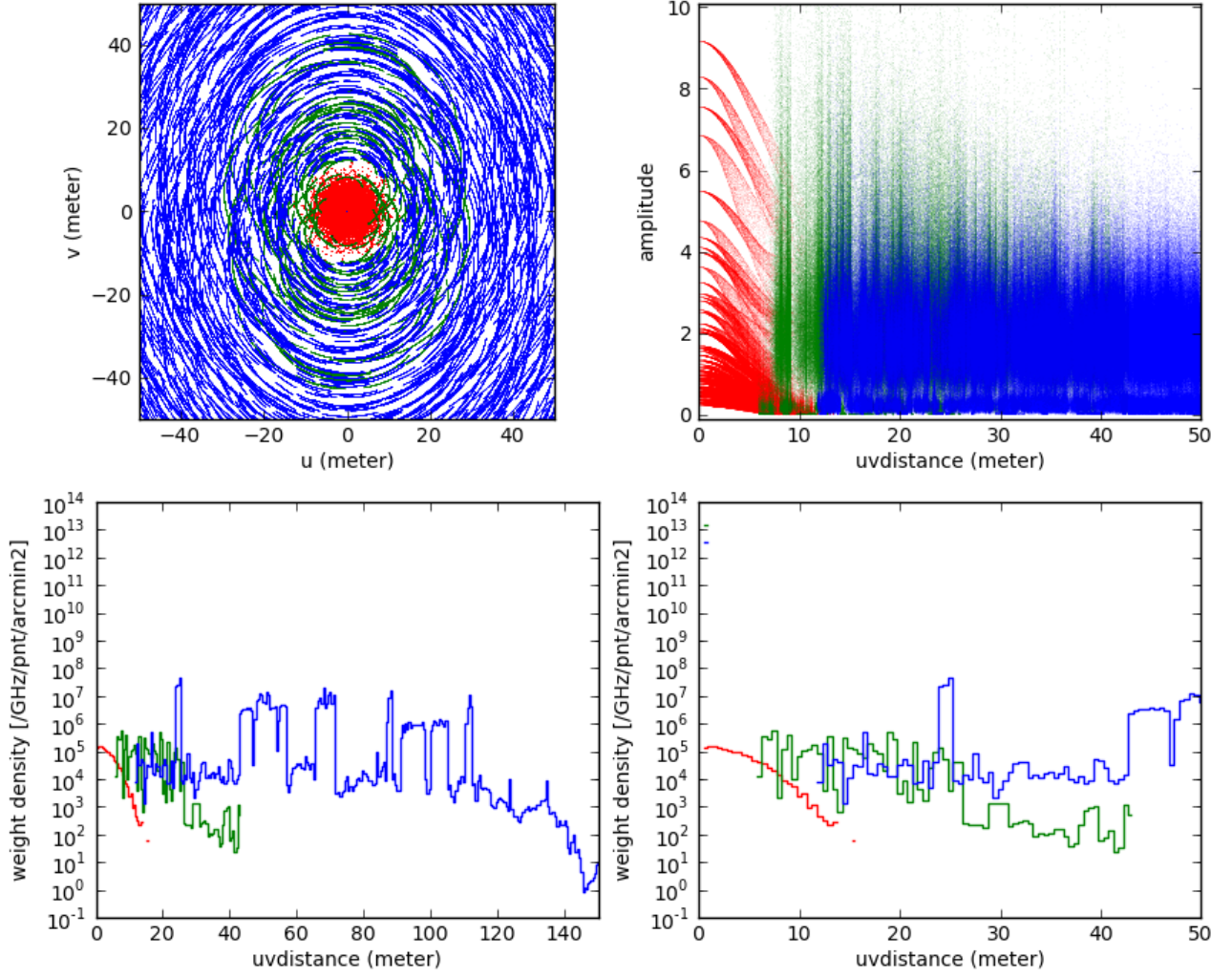


Figure 2.22: The TP2VIS evaluation images. Color blue: 12m-array; green: 7m-array; red: single dish. Upper left: The uv distribution of the three observations. Upper right: The weight-corrected amplitude distribution versus the uv distance. Bottom left: The corrected weight density of the three observations. Bottom right: The same as bottom left with a narrower uv distance range, in order to clearly show the agreement of the weight density of different observations.

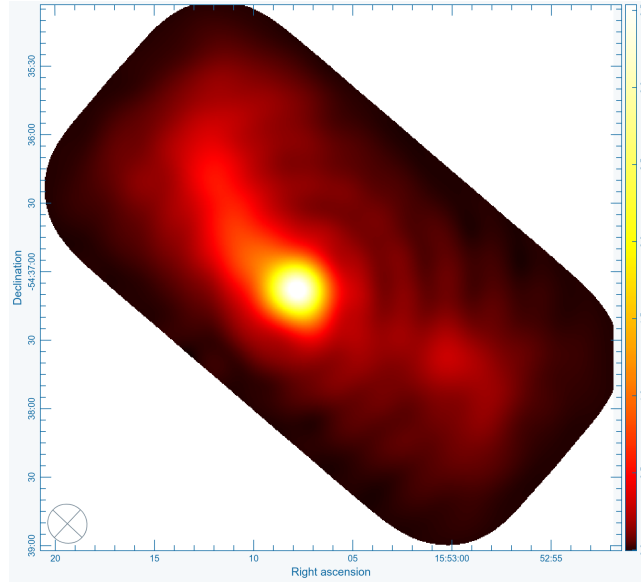


Figure 2.23: The reproduced single dish image with the generated visibility file showing artifacts, indicating a poor TP2VIS quality.

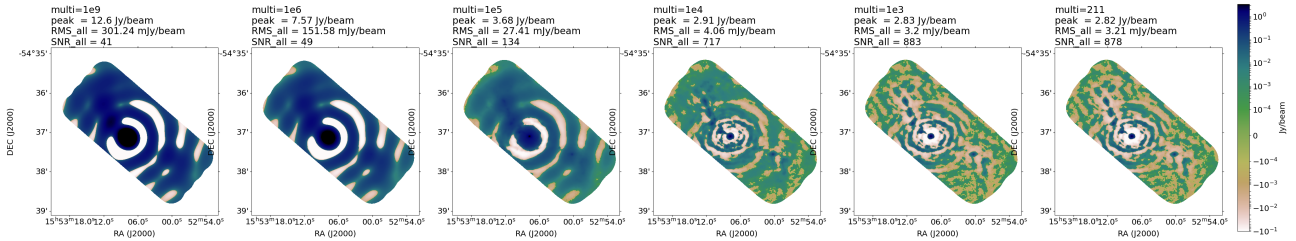


Figure 2.24: The cleaning results when applying different scaling factor, noted in the image as multi=1e9 for example. From left to right, the multiply factor becomes smaller, leading to a fainter artificial ring-like structure.

always appears in cleaned images, as seen in Fig. 2.24, while a smaller multiplier leads to a less significant ring-like structure. From this test we would say that the factor 10^8 read from the weight difference between 7m/12m and single dish data needs to be reconsidered. A factor of around 10^4 would be enough to keep the high-resolution data visible and weaken the artifacts at the same time.

There are a couple of parameters to play with in the TP2VIS method. It doesn't read the original beam size from the fits file to infer the single dish telescope size, instead take the size given by the user. Since the single dish data is at a different frequency (353 GHz) than the 12m and 7m data (220 GHz) and has a higher resolution, some modification should be done before converting to visibility. There are two ways to get the equivalent single dish data. One way is to scale the TP flux by a factor of 0.3 (same as the sdf in feather) and smooth the continuum by the resolution at 220 GHz (33.8"). In this case we will decrease the resolution, hence less preferred. Another way is to create a fake telescope with a larger dish size to increase the resolution. The real APEX telescope is 12m in diameter, corresponding to a resolution of 21.0" at 353 GHz and 33.8" at 220 GHz. Now we want a resolution of 21.0" at 220 GHz, so the virtual telescope size can be calculated as $12\text{m} \times 353\text{GHz} / 220\text{GHz} = 16.9\text{m}$. This will be defined in the TP2VIS task as a dictionary of virtual telescope: telescope diameter, beam size, maximum uv distance of TP visibility distribution (maxuv), number of antenna (nant), and visibility group (nvgrp). We tested different numbers of visibility group, i.e., the nvgrp, ranging

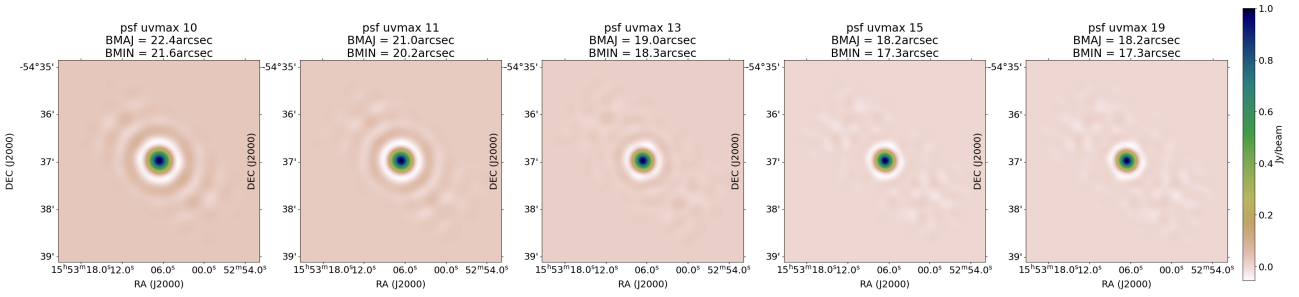


Figure 2.25: The point spread function of the virtual single dish data with different uvmax value. The selected uvmax values and the resulting beam sizes are noted above each image.

from 4 to 40 by a factor of 10. The results are similarly bad, since the most severe problem of lacking sensitivity information at a certain uv distance is not solved in this case. Then we modified the maximum uv distance, whose default value for a 12-meter single dish is until 10m. But considering we set a larger telescope, the parameter should be enlarged to better recover the uv coverage. This can be seen in the change in PSF when maxuv is set to 10, 11, 13, 15, and 19 2.25. The low uvmax leads to water-ripple artifacts in the PSF, and the lower uvmax produces stronger artifacts due to less uv coverage. When uvmax equals 19, the PSF is round and symmetric, corresponding to a beam size of $\sim 18''$ at 220 GHz, therefore we adopted this value.

In addition, there exists a mismatch between dirty and clean/restore beam areas. To overcome this problem, we used the *tp2vistweak* function to compare the two beam areas to rescale the flux density in the residual map and recalculate the cleaned map. This correction is important to give accurate flux; see the comparison of tweak-corrected and non-corrected image in Fig. 2.26. It also decreases the noise especially at the edge, the opposite of the pb-correction, which usually increases the noise at the edge.

From the previous discussion, we conclude that TP2VIS results heavily rely on the collective effect of many parameters, including the beam size of the single dish data, the scaling factor of single dish to interferometer, the maximum counted uv distance, and the beam correction method. We did not derive a satisfying combined continuum image for our sample mainly due to the lack of uv-coverage of single dish data, while it is worth a try for a future data set.

We also intended to use the newly implemented CASA **dintimaging** method to combine the single dish and interferometer data. However, we encountered the same problem of combining data at different frequencies. In the future, we will use the vpmanager tool to specify virtual voltage patterns and primary beams, in order to combine our data correctly.

2.6.3 Spectral combination

In terms of spectrum, we also applied the data combination algorithm to recover the total flux in each frequency channel. We used the image-plane data combination method to avoid the problem of converting single dish data to a virtual visibility file. Similarly as continuum, we first merged the 12m and 7m spectra using Feather function, which were then combined with the scaled APEX single dish spectra ($sdf=0.3$). The original spectra and the data combination results of a representing molecular line, H_2CO , are shown in Fig. 2.27. To compare the flux covered by 12m, 7m, APEX, and the combined data, we drew a circular region that is the same size of the APEX beam and then derived the spectra in units of Jy within the region. Generally,

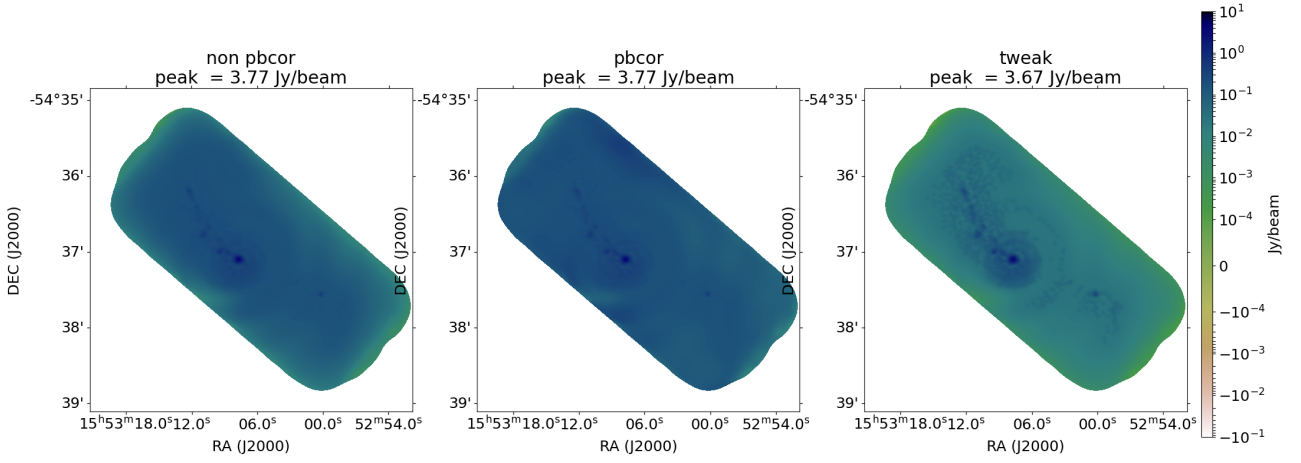


Figure 2.26: A comparison of different scaling methods. Left: The original image. Middle: The primary beam corrected image. Right: The image generated by the `tp2vistweak` function, which compares the single dish/interferometer beam areas to rescale the flux density in the residual map and recalculate the cleaned map. This correction is important to give the accurate flux and decrease the noise.

the spectra are consistent with each other in line width, velocity offset, and flux density. If we look carefully, the 12m spectrum (in blue color) shows double peaks, while the others do not. The 12m+7m+APEX combined spectrum (purple) is the strongest, and the single dish (red), 12m (blue), and 7m (green) spectra decrease in order. This can be explained by the missing flux in the interferometer and the lack of high-resolution structure in the single dish. When we plot the channel map at the peak frequency of this H_2CO line, the final combined image (bottom right) is basically similar to the 12m highest-resolution image (upper left), but shows clearly more extended emission and prominent cores.

Hence we draw the conclusion that our spectral combination results are reasonable and reliable, which can be applied for further line fitting and classification.

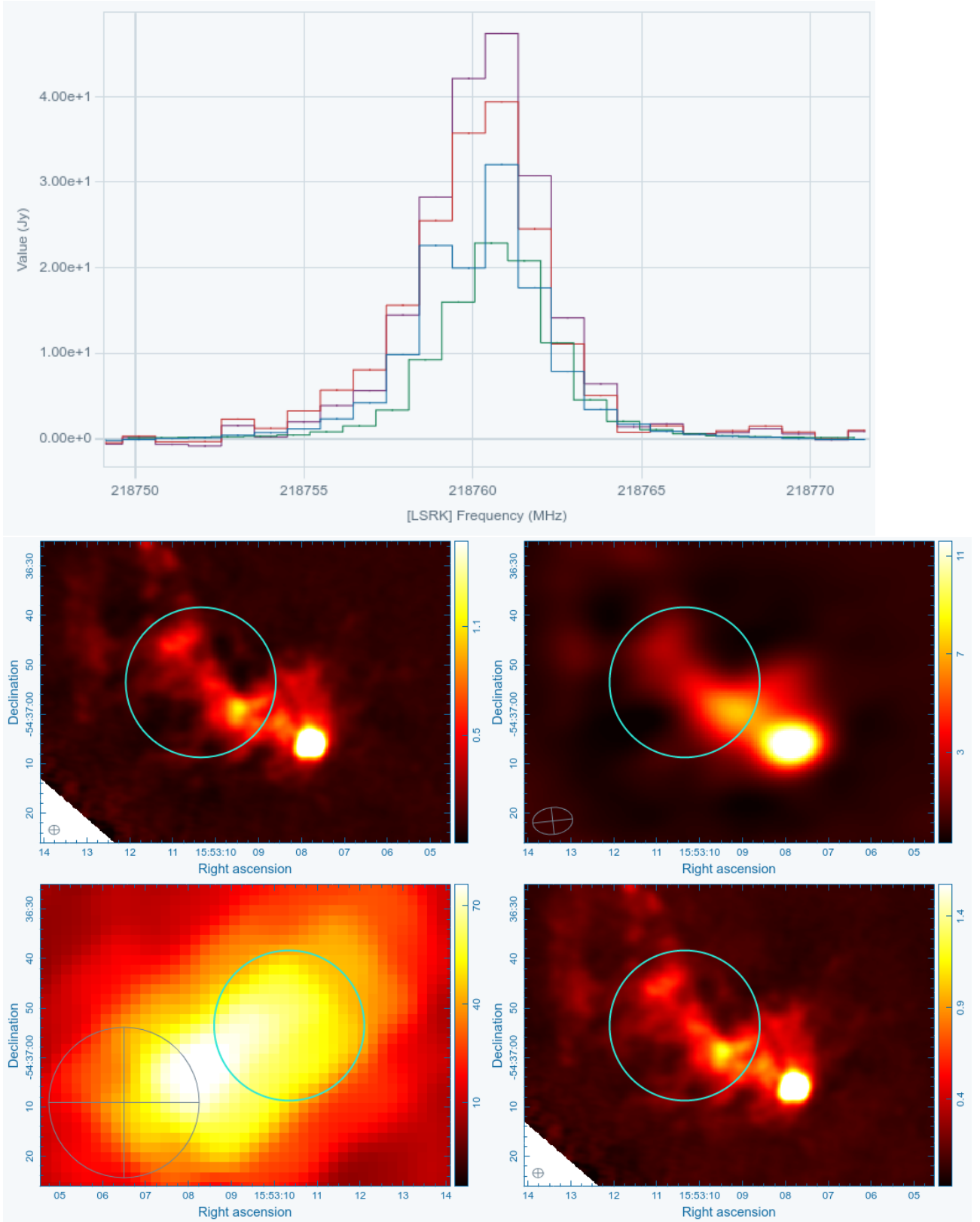


Figure 2.27: **Upper panel:** Comparing the H_2CO spectra of different telescopes and data combination, calculated within the single dish beam size. Purple: 12m+7m+APEX combined spectrum; Blue: 12m spectrum; Green: 7m spectrum; Red: APEX single dish spectrum. **Lower panel:** Comparing one of the channel maps of different telescopes and combined results. The beam size are shown in the bottom left corner of each image. The blue circle is the region used to derive the spectra shown in the upper panel. Upper left: 12m; Upper right: 7m; Bottom left: APEX; Bottom right: 12m+7m+APEX. The final combined image shows clearly more extended emission and prominent cores.

3 Continuum study

The large scale ($4' \times 4'$) 12M+7M+TP continuum image shows a beautiful filament at 1.3 mm for the first time (Figure 3.1). Along the filament, there are distinct compact sources harboring, among which the central source is especially intensive, with a peak flux density of 3.24 Jy/beam. In the extended emission, there are also a couple of sources scattered. These sources are defined as cores whose physical parameters are important for understanding cloud fragmentation and evolutionary stages. Their spectra are fitted in the next section.

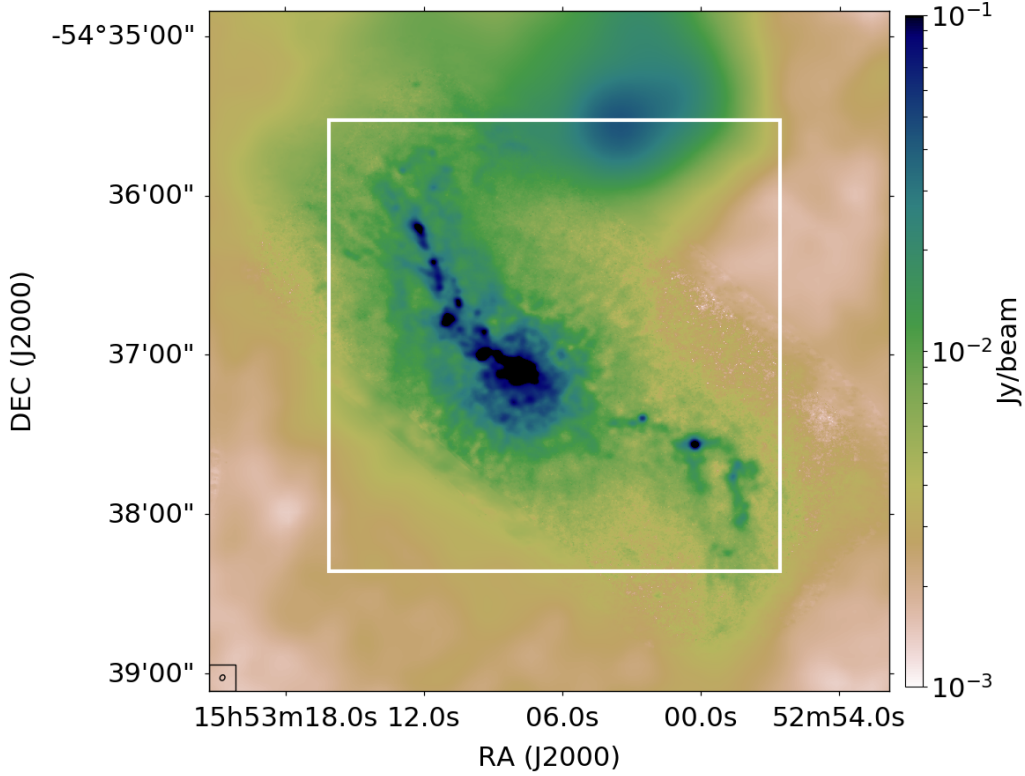


Figure 3.1: The $4' \times 4'$ 12M+7M+TP continuum images with a white box showing the region of interest. Due to the high dynamical range ($-1.36 \times 10^{-5} \sim 3.24$ Jy/beam) and the unusually strong central emission, the plot is shown in log scale and the limit is set to 0.001-0.1 Jy/beam.

The extended roundish feature at the north-western corner belongs to another region RCW97 (Rodgers et al., 1960), which is a $H\alpha$ emission region only observed by single dish (APEX) but not covered by the interferometer (ALMA). We do not see an obvious interaction between this region with our target filament. Most of the emission is located within the white box of Figure 3.1, hence the regions outside are excluded for further analysis.

3.1 Core identification

Determining the cores in G327 is of great importance for the following study, as cores refer to condensed areas in molecule clouds, which are precursors of protostars. The subsequent determination of hot molecular cores (HMCs) that have temperatures above 100 kelvin and cold cores whose temperature is closer to those of the ISM also provides the first insight to early star formation. Mass segregation and cloud fragmentation also rely on the core distribution. In this section, we introduced the pre-processing, core identifying algorithms, and some direct statistics.

3.1.1 Background subtraction: Findback

Findback ³ is an algorithm to determine the background (large scale structure) of an image, which in turn highlights compact sources if we subtract the background from the image. The embedded logic of this algorithm is to remove the low-density component of the source on a specified scale size. In detail, firstly we define a desired scale size in the shape of a box, which should be an odd number in the unit of pixel; otherwise the algorithm takes the next higher odd value. Then we replace each pixel value with the minimum value inside the box centered on the pixel, which produces a minimum value map. Again, we replace each pixel value in the minimum value map by the maximum inside the box, to generate the lower envelop of data. With some smoothing and correction, the background can finally be determined.

The background subtraction algorithm included in the STARLINK findback function works in the following way: First, we define a box size, which is the size of the desired smallest features to remain in the background subtracted image. Then each pixel value is replaced by the minimum value within the given box size. It is continued with a second replacement by the maximum value within the box of the minimum image. The third replacement is to calculate the mean value within the box of the maximum image. Finally, some correction is applied to noisy regions. Note, box size is the essential parameter in the findback algorithm.

To observe the effect of box size, the same 12M+7M+TP continuum image was processed using different box sizes in units of beam size, marked in Fig. 3.2. Since the beam size is an even number (20 pixels) but the findback function requires an odd one, the applied number of pixels is always added by 1 pixel, e.g., box size = 5 Beam = $5 \times 20 + 1 = 101$ pixels. From left to right of Fig. 3.2, the box size increases, and the background-subtracted images do show less extended emission and more prominent sources than the original image. When varying the box sizes from 1 beam size to 9 beam sizes of the 12m-array, the background images look more diffuse (upper panels in Fig. 3.2) while the background-subtracted images show brighter cores (central panels in Fig. 3.2). In order to obtain the most distinguishing cores in the core extraction step, an appropriate background should be subtracted. However, it is hard to claim which of these box sizes is better for finding cores. Therefore, some statistics are necessary for quantitative comparison, shown in the next section.

3.1.2 Core extraction: Dendrogram

The dendrogram is a well-established method for identifying clumpy structures (Goodman et al., 2009). A dendrogram is made up of two types of structures (see Figure 3.3): branches, which are structures splitting into multiple substructures, and leaves, which are structures that do not have substructure. Branches can split into branches and leaves, which allows hierarchical structures to be adequately represented. The term trunk is used to refer to a structure that does not have a parent structure. The smallest structure, leaf, corresponds to the core in the molecular cloud.

The way the algorithm works is to construct the tree starting from the brightest pixels in the data set, and progressively adding fainter and fainter pixels. If a local maximum is reached, it creates a new structure, named leaf. When a new pixel is adjacent to more than one existing structure, the structures are merged into a branch, and a chunk in the end. Otherwise, the pixels are joined to an existing structure. After getting a leaf, the corresponding major/minor

³<http://starlink.eao.hawaii.edu/docs/sun255.htx/sun255ss5.html>

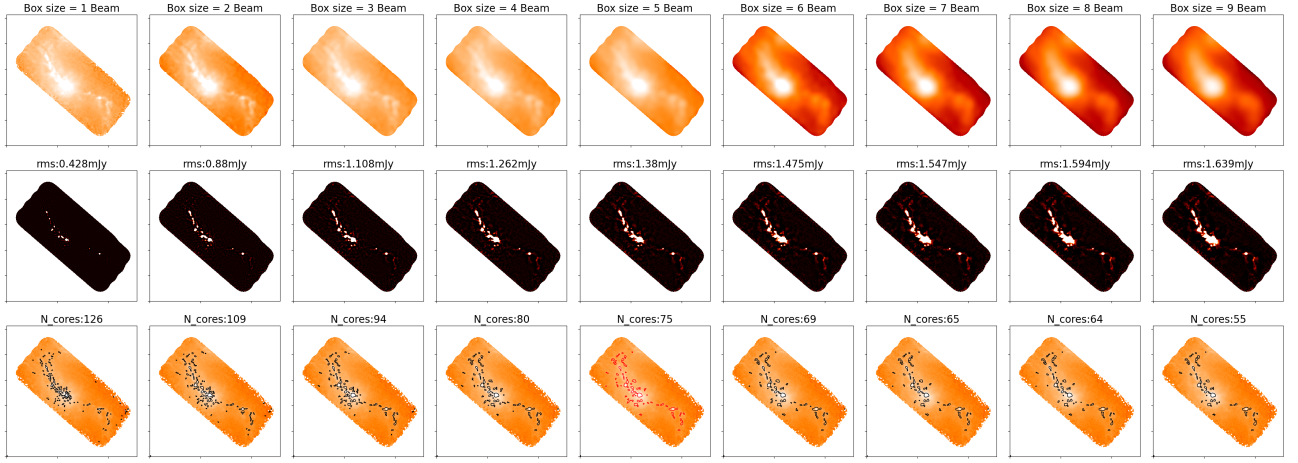


Figure 3.2: Upper panels: the background of 12M+7M+TP continuum image derived using STARLINK findback function, with varying parameter *box size* from 1 beam size to 9 beam sizes; Central panels: background subtracted images, in which cores are more significant; Lower panels: core identification result towards the background subtracted images using dendrogram algorithm with the same parameters.

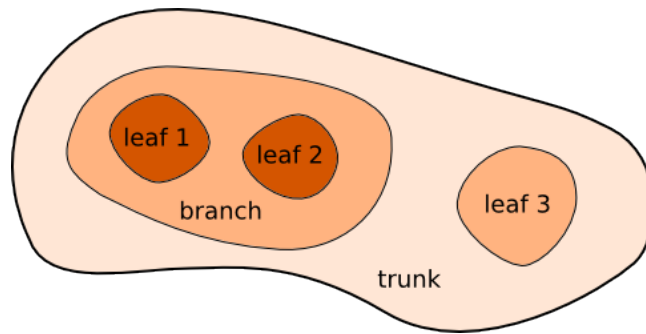


Figure 3.3: The dendrogram of a two-dimensional map of a hierarchical structure. Trunks, branches and leaves are shown.

axes will be fitted as the full-width at half-maximum (FWHM) in the direction of the greatest elongation in the position-position (PP) plane. This can lead to a much smaller size than the leaf size, especially for sources such as hot cores that present a huge gradient of intensity. Therefore, we use the original leaf as the core area to sum up the fluxes and average the spectra, while the fitted major/minor axes are treated as the densest scale of the core, and compared with the size of leaf.

To deal with the noise in real data sets, the user can set a minimum value that should be taken into account. To avoid noise spikes above the minimum value, we can define the minimum height required for a structure to be retained. The minimum number of pixels that a structure should contain in order to remain an independent structure can also be specified by the user. With these three key parameters, the structures determined are significant and reliable.

Now we applied the dendrogram algorithm to identify the G327 continuum cores. In the G327 data set, the RMS value of the background-subtracted image is 0.57 mJy/beam. Three key parameters are defined as follows. The minimum value of a core peak intensity is limited to three times the RMS, and the border intensities of two adjacent cores are required to differ beyond the RMS. At the same time, the minimum number of pixels is set to half of the beam area (156 pixels). We accept a source less than the beam size because the definition of major/minor axes in the dendrogram is FWHM rather than fitting the border to an ellipse. Nevertheless, sources should exceed half the beam size to qualify as a core.

To have a statistical comparison, we applied the dendrogram on the background-subtracted images with the same settings (Value= 3σ , Delta= 1σ , Npix=156), while the σ is different for each image, considered the standard deviation derived by sigma-clipping. In general, the dendrogram successfully identified the most significant cores in all images, as seen in the bottom panel of Fig. 3.2. We discuss the results in the next paragraph.

When the box size equals one beam size, the determined cores tend to surround the central hot core, possibly because of some noise spikes. When box size grows, the cores become more widely distributed, while the number of cores decreases, which can be explained by the increasing noise level and hence raising the detection limit. Meanwhile, we check the source sizes and fluxes within the cores to determine the best box size. For the source size, there are two choices: one is the encoded major/minor sigma provided by the dendrogram, and another is estimated from the leaf area ($A=\pi r^2$). The kernel density estimation (KDE) plots of both sizes are shown in the upper panel of Fig. 3.4, in which the legend shows the peak position of the kde distribution. The cores mainly assemble at small sizes, which presents a Gaussian peak on the small-size side. Large cores are not dominating. Nevertheless, there are more large cores when increasing the size of the box. Naturally, as previously discussed, the leaf size is larger than the FWHM size by a factor of two. Due to the setting of a minimum number of pixels, around 30% of the leaf cores are smaller than the beam size, most of which remain reliable since their peak intensities exceed 4 RMS. Since the leaf size is mostly larger than the beam size and more reliable, we adopted the leaf size for further analysis, e.g., to count the flux within the core, and to get the average spectra.

After defining the region of cores, we obtained the fluxes of cores on both the background-subtracted images and the original image. The fluxes derived on the 12M+7M+TP image (marked with BG) include the flux of a leaf and the corresponding branches, which are naturally larger than the background-subtracted fluxes (noted as noBG) that represent the leaf flux only. We plotted the KDE distribution of fluxes in both cases in the bottom panel of Fig. 3.4. The first impression is that different background subtractions are highly consistent, indicating that

the dendrogram is a robust method to extract cores. To watch closer, when subtracting too much background (e.g., BG1), the source fluxes tend to be small; while subtracting a large-scale, weak background (e.g., BG9) leads to large source fluxes. In addition, the flux at the peak of distribution gradually increases from 1 to 9 beam sizes, no matter whether the background emission is included or not. The peak flux is stable at around 5 beam size.

In conclusion, choosing an optimal box size is to balance the core size/flux and the core number. A smaller box size leads to more cores with smaller sizes, which can be affected by artifacts. A larger box size results in fewer and larger cores, while it may miss small cores. A middle box size helps to find cores in noisy regions but not spikes. In this case, we use the 5 beam as the best box size for background subtraction. This is also consistent with previous studies (Guzmán et al., 2015; Yuan et al., 2018; Zhang et al., 2018), which usually set the box sizes to around five times the beam size, although determined by visual inspection.

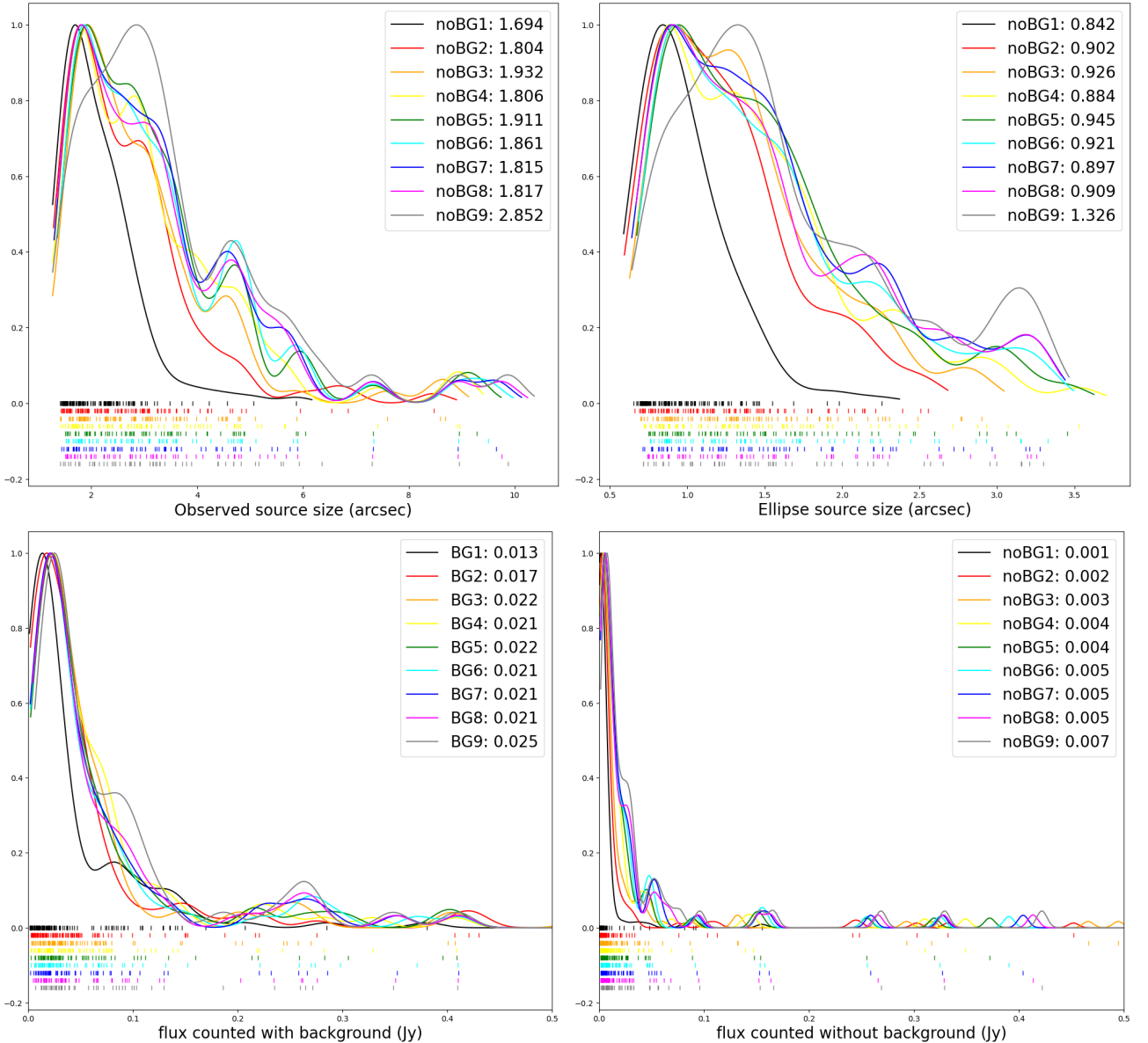


Figure 3.4: Upper left: kde plot of source size distribution estimated from the border of dendrogram polygons; Upper right: kde plot of dendrogram default source sizes that are FWHM of the polygons; Bottom left: kde plot of flux counted on the original images; Bottom right: kde plot of flux counted on background subtracted images. Different colors correspond to different box sizes of background subtraction.

As we adopted box size = 5 Beam, the dendrogram algorithm found 75 cores in total. However, when we look into more detail, there are nine cores not satisfying and removed from further analysis, in which one locates at the noisy edge of the observing field, four are very weak and show less than 3 spectral lines, four are surrounding the hot core and show artifacts. Therefore, we detected 66 reliable cores in the G327 filament, shown in Fig. 3.5. The polygon of leaves, the corresponding ellipse sizes and the core numbers are shown there.

In Table 3.1, we summarize the derived center positions, the major/minor axis from the dendrogram, the estimated source radius from the identified leaves, the rotating angle and the fluxes of the ellipses. Note that the core number starts from zero, the same as in the following discussion. We also plotted the KDE distribution of the peak intensity, total flux, and estimated source size in the upper panel of Fig. 3.6, which are all Gaussian-shaped at the low value end, and show a long tail toward the high value. Considering that the fluxes are plotted in logarithm scale, the small fluxes might follow a logarithm-normal distribution, while the large fluxes might be affected by turbulence and deviate from the distribution. The average major axis of the sources is $2.3''$ and the average minor axis is $1.4''$, which is equivalent to a round cloud with a diameter of $1.8''$ (corresponding to 0.03 pc). The estimated source size is derived from the area of leaves, which has a median value of $2.80''$ (~ 0.045 pc), 56% larger than the approximation of major/minor axes. Nevertheless, both small source sizes confirm that we resolve the star forming region in the core scale, while not yet in the sub-core scale due to the resolution limit.

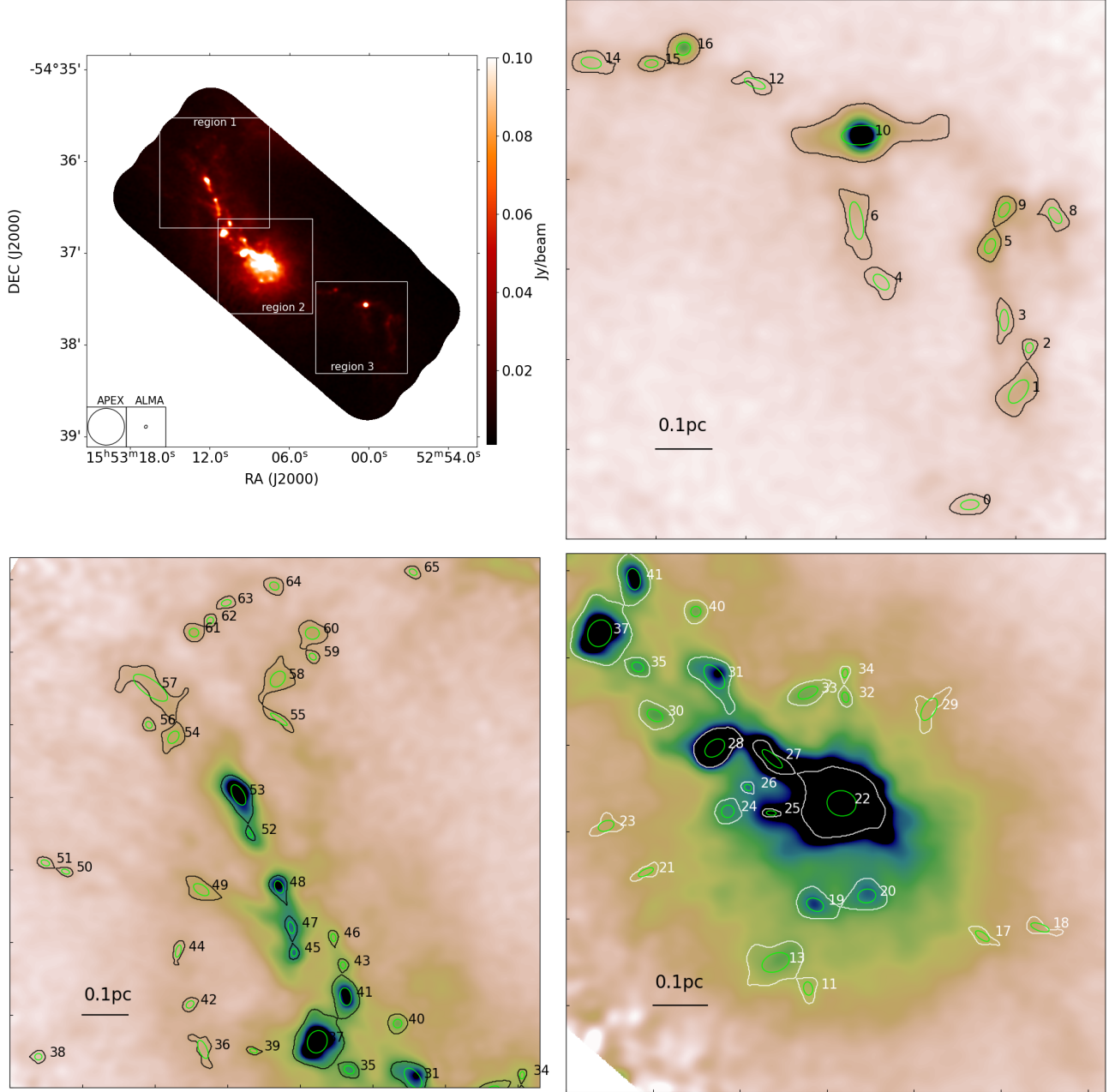


Figure 3.5: The 12m+7m+APEX continuum image showing the three zoom-in regions, marked with identified cores with core numbers: region 1 in bottom left, region 2 in bottom right, region 3 in upper right.

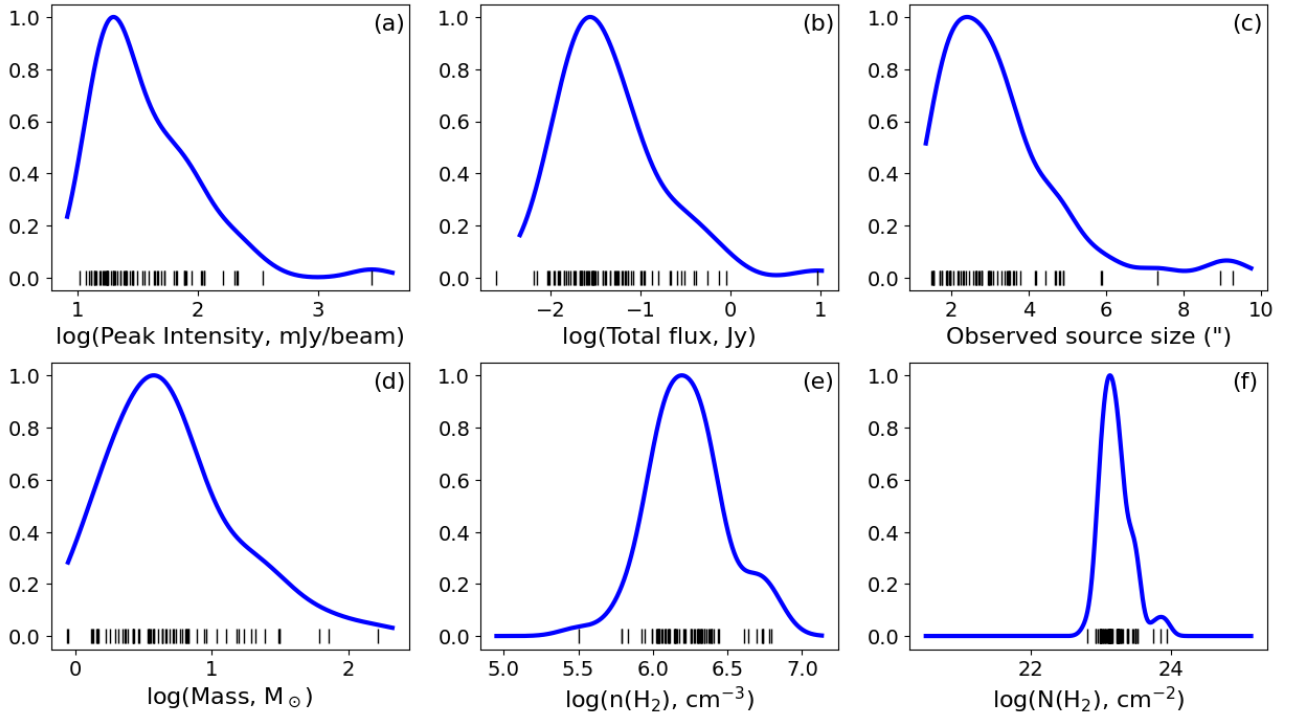


Figure 3.6: KDE distribution of the identified 66 cores towards (a) peak intensity, (b) total flux with background emission, (c) source size estimated from dendrogram leaves, (d) continuum mass including dust and gas, (e) volume density (f) column density in respect to the leaf size. Note, some KDE are plotted in logarithm scale.

Table 3.1: Core list found by Dendrogram

No	RA,DEC (J2000) degree,degree	a,b ", "	Radius "	Angle °	Flux mJy	Mass M _☉	N_{H_2} cm ⁻²	n_{H_2} cm ⁻³	T K
0	238.24524836,-54.63743686	1.0*0.5	3.2	6	21.2	4.4	1.16e+23	1.10e+06	30
1	238.24264802,-54.63393670	1.5*0.8	4.7	53	61.3	11.0	1.36e+23	8.85e+05	34
2	238.24206897,-54.63258431	0.6*0.4	2.0	77	9.3	1.8	1.28e+23	1.99e+06	32
3	238.24341486,-54.63173136	1.2*0.5	3.0	90	22.5	4.7	1.42e+23	1.44e+06	30
4	238.24998158,-54.63055608	1.1*0.6	3.3	136	23.5	4.4	1.08e+23	9.87e+05	33
5	238.24417017,-54.62944046	0.9*0.6	3.0	68	37.8	4.4	1.36e+23	1.40e+06	50
6	238.25129672,-54.62865442	2.1*0.7	4.9	99	64.2	8.8	9.99e+22	6.22e+05	43
7	238.28131100,-54.62908410	1.2*0.7	3.4	163	26.3	5.4	1.30e+23	1.17e+06	30
8	238.24068861,-54.62849529	1.0*0.6	3.0	125	23.5	3.5	1.04e+23	1.05e+06	40
9	238.24341195,-54.62832136	0.9*0.5	2.8	60	32.1	3.7	1.28e+23	1.39e+06	50
10	238.25105994,-54.62601354	2.2*1.1	9.3	5	561.2	30.7	9.69e+22	3.17e+05	100
11	238.28416396,-54.62433925	0.8*0.5	2.6	101	40.3	2.9	1.21e+23	1.43e+06	76
12	238.25673280,-54.62441936	1.2*0.5	2.7	158	15.1	2.7	9.83e+22	1.09e+06	34
13	238.28595587,-54.62350581	1.6*1.0	5.9	21	258.5	31.2	2.47e+23	1.28e+06	48
14	238.26546983,-54.62378192	1.1*0.6	3.5	171	30.5	5.9	1.31e+23	1.14e+06	32
15	238.26225275,-54.62380573	0.7*0.4	2.3	3	17.9	3.4	1.71e+23	2.23e+06	32
16	238.26053957,-54.62334467	0.8*0.7	3.6	29	67.1	6.1	1.27e+23	1.07e+06	62
17	238.27455256,-54.62266219	0.9*0.4	2.4	146	29.8	4.1	1.94e+23	2.47e+06	43
18	238.27137304,-54.62238680	1.1*0.4	2.5	161	21.7	3.8	1.67e+23	2.06e+06	35
19	238.28378017,-54.62166178	1.0*0.7	4.2	155	213.4	20.9	3.26e+23	2.37e+06	58
20	238.28092547,-54.62138161	1.0*0.8	4.2	9	219.1	19.6	3.03e+23	2.20e+06	63
21	238.29311239,-54.62062259	0.9*0.4	2.3	25	24.7	3.4	1.78e+23	2.39e+06	43
22	238.28231934,-54.61842379	1.7*1.5	9.0	170	9269.8	163.6	5.55e+23	1.88e+06	299
23	238.29530714,-54.61914367	0.9*0.6	2.7	21	29.6	2.6	9.74e+22	1.09e+06	63
24	238.28859487,-54.61867865	0.7*0.7	3.0	36	107.3	6.7	1.98e+23	1.99e+06	88
25	238.28619857,-54.61873329	0.6*0.3	1.5	176	45.8	2.1	2.45e+23	4.92e+06	120
26	238.28748168,-54.61791930	0.4*0.3	1.5	147	33.2	2.4	2.76e+23	5.43e+06	76
27	238.28614255,-54.61700218	1.5*0.5	3.5	138	395.5	31.1	7.07e+23	6.20e+06	71
28	238.28930957,-54.61665747	1.3*0.9	4.8	38	747.5	71.9	8.54e+23	5.42e+06	59
29	238.27748980,-54.61541944	1.4*0.7	3.5	59	55.8	7.8	1.74e+23	1.51e+06	42
30	238.29261731,-54.61561460	1.0*0.6	3.5	153	103.9	5.2	1.14e+23	9.80e+05	109
31	238.28933672,-54.61438052	1.6*0.9	4.9	128	305.0	15.7	1.79e+23	1.11e+06	106
32	238.28211074,-54.61503546	0.6*0.4	2.0	106	29.5	1.7	1.17e+23	1.80e+06	96
33	238.28417485,-54.61488842	1.2*0.6	3.6	24	100.7	4.9	1.00e+23	8.38e+05	112
34	238.28212935,-54.61427435	0.5*0.3	1.6	83	15.7	0.9	9.77e+22	1.90e+06	98
35	238.29356076,-54.61406503	0.7*0.5	2.6	157	69.7	3.6	1.40e+23	1.62e+06	106
36	238.30321685,-54.61328475	1.3*0.6	3.1	113	30.3	2.2	6.31e+22	6.20e+05	76
37	238.29566130,-54.61299168	1.6*1.3	7.3	66	906.3	61.1	3.10e+23	1.28e+06	82
38	238.31408900,-54.61356139	0.5*0.4	1.9	3	6.6	1.4	1.01e+23	1.60e+06	30
39	238.29987880,-54.61333875	0.5*0.3	1.5	166	9.8	0.9	1.07e+23	2.16e+06	62
40	238.29035540,-54.61229736	0.6*0.6	2.8	50	54.1	6.7	2.33e+23	2.54e+06	47
41	238.29378844,-54.61126583	1.2*0.8	4.7	100	282.8	15.3	1.89e+23	1.22e+06	101
42	238.30406619,-54.61156484	0.6*0.4	2.1	40	12.9	1.3	8.51e+22	1.25e+06	55
43	238.29395775,-54.61006442	0.5*0.4	1.7	99	22.4	1.5	1.37e+23	2.45e+06	85
44	238.30484425,-54.60952328	0.9*0.3	2.2	75	14.3	1.5	8.55e+22	1.20e+06	55
45	238.29720323,-54.60957949	0.5*0.4	1.8	92	41.4	3.5	3.11e+23	5.38e+06	66
46	238.29459954,-54.60900378	0.5*0.3	1.8	100	21.0	2.7	2.34e+23	4.06e+06	46
47	238.29738407,-54.60862668	0.8*0.4	2.3	102	71.8	6.5	3.33e+23	4.37e+06	62
48	238.29819096,-54.60707016	0.9*0.6	3.0	112	133.6	6.8	2.08e+23	2.12e+06	107
49	238.30332519,-54.60717250	1.2*0.6	3.7	148	57.5	6.5	1.30e+23	1.07e+06	51
50	238.31228843,-54.60647942	0.6*0.3	1.7	163	7.1	1.5	1.38e+23	2.46e+06	30
51	238.31362142,-54.60614514	0.6*0.4	2.0	157	9.5	2.0	1.37e+23	2.11e+06	30
52	238.30009241,-54.60503648	0.6*0.3	1.5	116	28.3	2.7	3.05e+23	6.01e+06	60
53	238.30086560,-54.60354967	1.5*0.8	4.8	119	409.0	24.7	2.89e+23	1.83e+06	91
54	238.30515858,-54.60134237	0.9*0.7	3.4	52	45.8	6.4	1.48e+23	1.31e+06	42
55	238.29819679,-54.60065488	1.4*0.3	2.7	145	25.5	3.8	1.45e+23	1.66e+06	40

Continued in the next page

No	RA,DEC (J2000) degree,degree	a,b ", "	Radius "	Angle °	Flux mJy	Mass M _☉	N_{H_2} cm ⁻²	n_{H_2} cm ⁻³	T K
56	238.30677716,-54.60086727	0.5*0.4	1.9	128	11.0	1.4	1.11e+23	1.79e+06	45
57	238.30667016,-54.59944591	2.8*1.1	5.9	144	112.0	17.1	1.33e+23	6.85e+05	39
58	238.29826649,-54.59911764	1.2*0.9	4.4	49	79.1	12.8	1.77e+23	1.21e+06	37
59	238.29595629,-54.59825858	0.5*0.4	1.9	138	12.6	2.3	1.72e+23	2.73e+06	33
60	238.29598609,-54.59736690	1.0*0.8	3.8	178	51.3	9.2	1.75e+23	1.40e+06	34
61	238.30380177,-54.59733879	0.7*0.6	2.8	168	27.3	5.2	1.83e+23	2.00e+06	32
62	238.30272263,-54.59687818	0.5*0.4	1.9	55	11.1	2.3	1.74e+23	2.78e+06	30
63	238.30166784,-54.59620257	0.7*0.4	2.2	17	14.0	2.9	1.64e+23	2.27e+06	30
64	238.29848905,-54.59555844	0.7*0.5	2.6	152	22.1	4.6	1.78e+23	2.04e+06	30
65	238.28932864,-54.59502858	0.6*0.4	2.0	151	12.1	1.3	9.12e+22	1.40e+06	53

3.2 Spatial distribution of the sources

The spatial distribution of stars in young prestellar clusters has received a substantial amount of attention, both in observations and in simulations, since they reflect the interaction between the star and the environment, which is important to understand the star formation.

3.2.1 Minimum spanning tree

To characterize the spatial distribution of the cores down to scales of 6600 AU and examine the preferential fragmentation scales, we computed the basic statistics of the projected nearest neighbour separations. Here, we used the method of minimum spanning tree (MST), which is an algorithm searching for the minimum total edge weight that can connect all the vertices together in a graph. We constructed the MST using the Python package *NetworkX*⁴ which implements the MST according to Kruskal’s algorithm (Kruskal, 1956)

In Fig. 3.7 we present the connecting segments of the minimum spanning tree for the whole filament. To prove the statistical significance of the MST result, the Kolmogorov-Smirnov test (KS test) (Frank J. Massey, 1951) was applied, assuming that the null hypothesis is no correlation between core locations, i.e. random distribution. The returned p-value is 0.09, higher than 0.05, thus accepting the null hypothesis and indicating a trend of random distribution.

Distance distribution Next, we use the MST results to evaluate the distances between the nearest neighbors, which is shown in Fig. 3.8. Separations for each determined pair of ALMA sources have mean and median values of 0.167 and 0.147 pc. The KDE plot also shows four peaks, located at 0.088, 0.148, 0.341, and 0.414 pc, separately. The average and peak separations are a few times of the typical core size (~ 0.1 pc), indicating a relatively crowded core environment, although the mean separation is larger than Orion A (0.037 pc) and NGC 6334 (0.045 pc) (Sadaghiani et al., 2020).

Angle distribution In previous simulations and observations, many filaments were found to be either parallel or perpendicular towards the main filament. To examine whether this phenomenon exists in our field, the angle distribution throughout the filament is plotted in Fig. 3.9.

⁴Available at <https://networkx.github.io/>

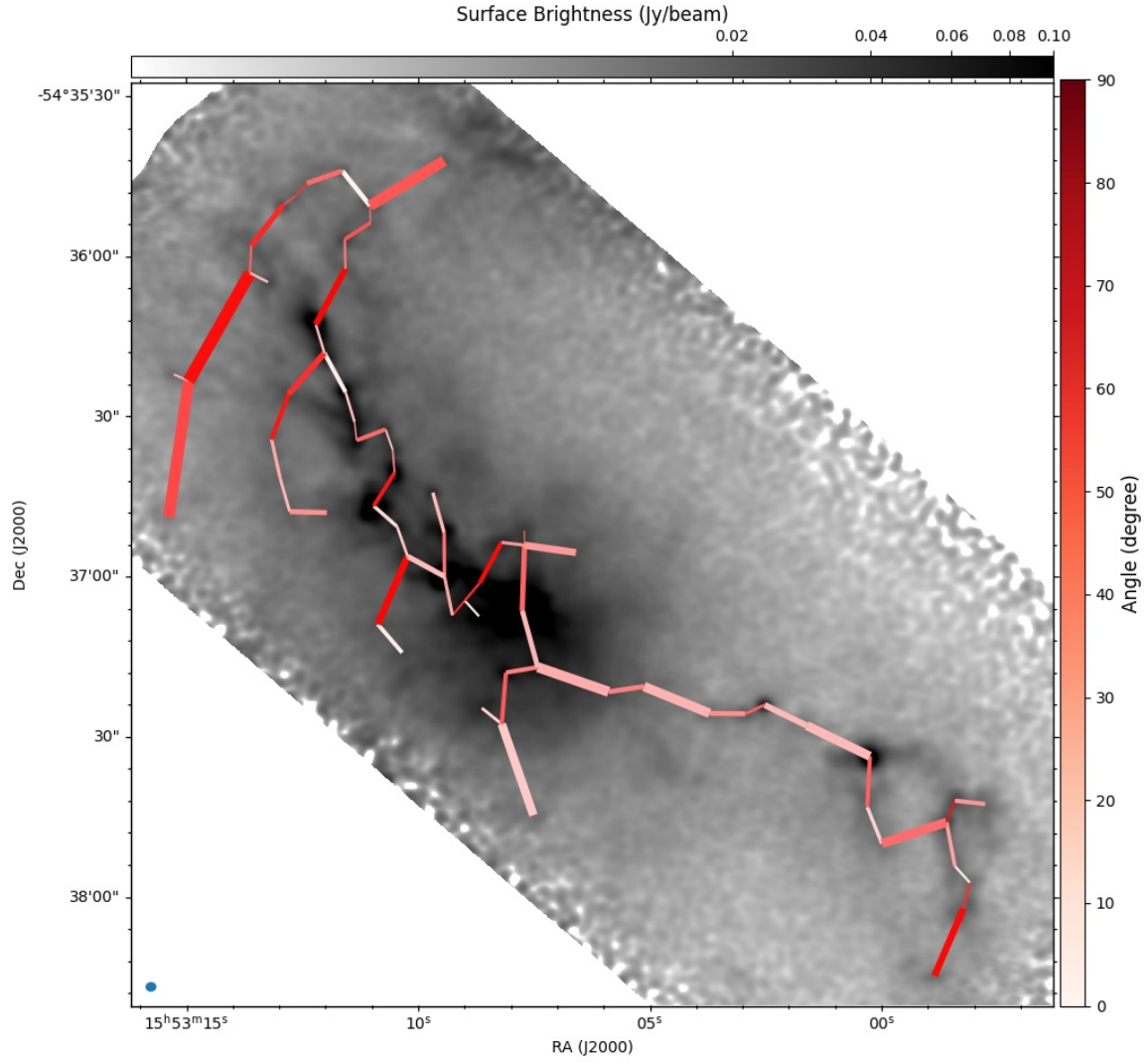


Figure 3.7: Minimum spanning tree (MST) for the ALMA continuum sources. The thickness of the connecting lines reflects the 2D projected distance between each pair of sources. The color indicates the relative angles towards the main filament.

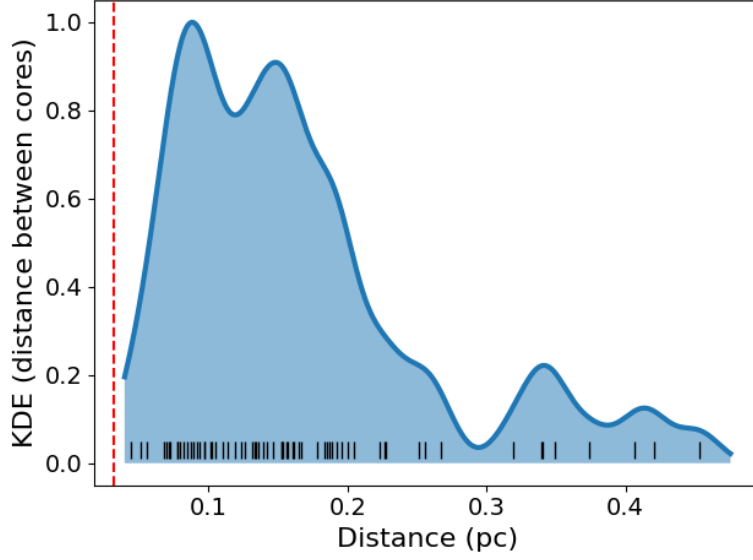


Figure 3.8: The distribution of the 2D projected separation between nearest neighbours for ALMA sources, showing multiple peaks and a median separation of 0.15 pc. The red vertical dashed line represents the ALMA resolution limit $2''$ in this study.

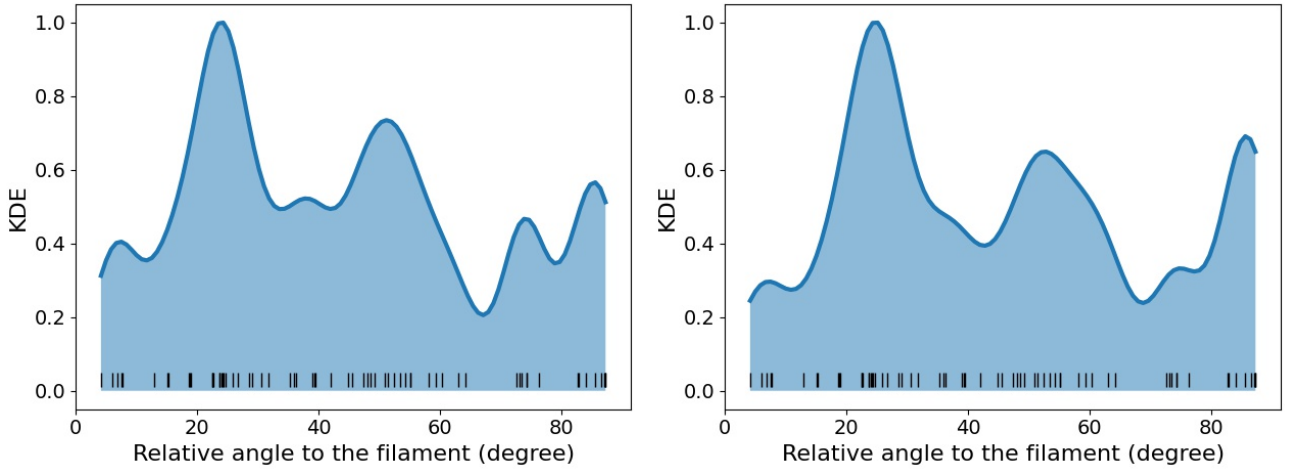


Figure 3.9: Relative angle distribution (left panel) and weighted by the corresponding distances (right panel).

Firstly, the angle is corrected to be the relative angle towards the main filament, which is separated into two sections: the north-eastern part ($RA > 15:53:04.8535$) is 40 degrees clockwise to the x axis, and the south-west part ($RA < 15:53:04.8535$) is 13 degrees clockwise to the x axis. After correcting the angles, we show the distribution of angle distribution, which is obviously not uniform. There are two significant peaks at 24 and 51 degrees, together with four minor peaks at 8, 38, 74, and 86 degrees. After using the corresponding separation as weight to plot the KDE, there are three peaks at 24, 51, and 86 degrees, and two very weak peaks. Although there are significant peaks that deviates from 90 degree, we do have a 86 degree preference in many cases, proving the theory of perpendicular sub-filament.

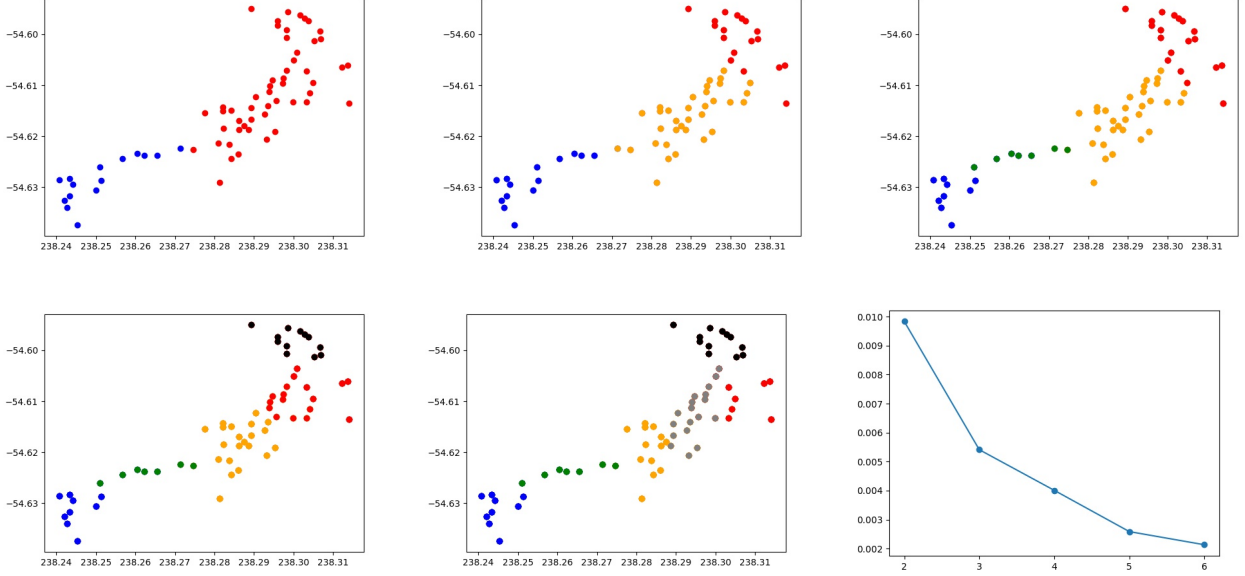


Figure 3.10: From upper left to bottom middle: the clustering results using a Gaussian mixture model with a given number of clusters (from 2 to 6). Bottom right: the within-cluster sum of squares (WSS) versus different number of clusters. The elbow point of 3 refers to the optimal number of clusters.

3.2.2 Clustering

Since the filament shows a turning point in the hot core and is not continuous in gas distribution, we try to determine different sections of the filament by clustering algorithm. Since our source distribution is on a non-straight filament, we need some algorithms that can well identify clusters in the elongated sample. Gaussian mixture models⁵ would be a good choice in this case, by assigning multivariant normal distribution components to the data points. With the help of the Scikit-learn 1.2.2 package, we run the Gaussian mixture clustering function for a given number of clusters, ranging from 2 to 6, as shown in Fig. 3.10. For each run, we calculate the within-cluster sum of squares (WSS). The WSS indicates the compactness of the clusters and tends to decrease as we increase the number of clusters, as seen in the bottom right panel of Fig. 3.10. The value at which the rate of decrease changes sharply is taken as the number of clusters (Malika Charrad, 2014). From this point on, adding another cluster does not improve the total WSS. This point is the so-called elbow point, which is 3 in our case. Hence, three clusters throughout the filament is the optimal case.

Gaussian mixture models are useful for clustering, whereas another algorithm KMeans⁶ can be seen as a special case of a Gaussian mixture model with equal covariance per component. We run the K-Means algorithm for a range of K values that change from 2 to 6, and the corresponding WSSs are calculated, shown in Fig. 3.11. The clustering results of KMeans are highly consistent with the Gaussian mixture results, both in the elbow point determination and the cluster distribution.

With all kinds of clustering methods, the hot core is unexpectedly always sitting between two clusters, rather than at the center of a cluster. Thus, we suspect that a hot core is formed

⁵<https://scikit-learn.org/stable/modules/mixture.html>

⁶<https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>

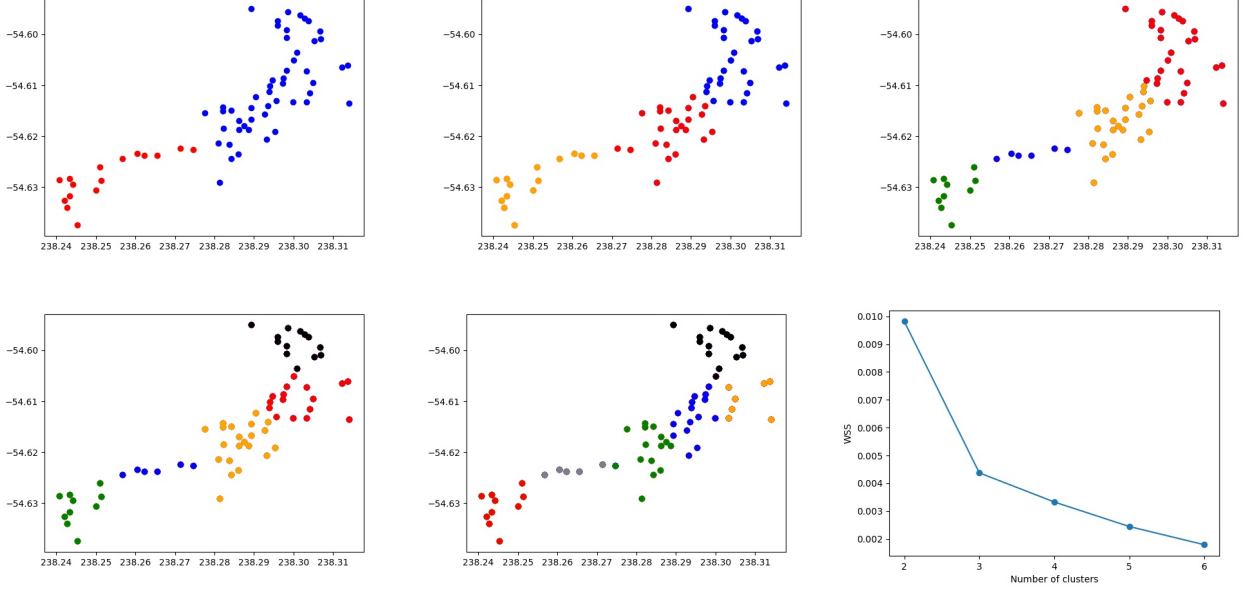


Figure 3.11: From upper left to bottom middle: the clustering results using Kmeans algorithm with a given number of clusters (from 2 to 6). Bottom right: the within-cluster sum of squares (WSS) versus different number of clusters, showing an elbow at 3 clusters.

by a cloud-cloud collision of the north-eastern and the south-western fragments.

The Q parameter The existence of sub-structures can also be examined statistically. [Cartwright and Whitworth \(2004\)](#) defined a parameter Q to characterize the spatial distribution. Q is defined as the ratio of the normalized mean length of the segments of a minimum spanning tree (MST), \bar{l}_{MST} , to the correlation length, \bar{s} , i.e., the mean projected separation between the data points normalized by the cluster radius ([Hunter et al., 2014](#)). The radius of the cluster R is defined as the distance from the mean position of all members of the cluster to the farthest core. For \bar{l}_{MST} , the normalization term is given by $\sqrt{A/N}/(N-1)$ [Parker \(2018\)](#), where A is the area of the circle with radius R encompassing all the cores, and N is the number of cores. Therefore, Q is derived with the equation:

$$Q = \frac{\bar{l}_{MST}}{\bar{s}} = \frac{\frac{\sum \frac{L}{N-1}}{\frac{\sqrt{N\pi R^2}}{N-1}}}{\frac{\sum \frac{S}{N(N-1)/2}}{R}} = \frac{\sum L / \sqrt{N\pi}}{\sum \frac{S}{N(N-1)/2}} = \frac{0.47}{0.97} = 0.49, \quad (3.1)$$

where we calculate mean length of MST and correlation length to be 10.43'' (~ 0.17 pc) and 96.82'' (~ 1.55 pc), respectively. Taking into account a radius of the cluster of 100.00 (~ 1.60 pc), we obtain $Q = 0.49 \pm 0.1$.

The Q parameter is a dimensionless measure of the smooth over densities and fractal structures in a cluster of data points. A value of $Q < 0.8$ indicates the presence of fractal sub-structures in the data set, while $Q > 0.8$ corresponds to a centrally concentrated structure with a large-scale radial density gradient ([Cartwright and Whitworth, 2004](#)). The low Q value we derived confirms the presence of fractal structures or clusters within the cores identified in the G327 filament.

Other star-forming regions show varies Q values. For the Pipe nebula cloud, [Alfaro and Román-Zúñiga \(2018\)](#) found similar small Q values of 0.4, while it is a region with very low star formation activity. [Parker \(2018\)](#) analyzed the three sub-regions in Orion B, and obtained

$Q = 0.72, 0.65$, and 0.71 for L1622, NGC 2068/2071, and NGC 20223/2024, respectively. Nevertheless, these two studies are broadly consistent with our findings and makes it reliable.

Two-point correlation function (TPCF) In the next approach, we explored the clustering behavior of the cores using the two-point correlation function, $\xi(r)$, which describes the excess probability per unit area of an object located at a separation r from another object, compared to the expectation of the random distribution. We adopted the TPCF defined by [Landy and Szalay \(1993\)](#):

$$\xi(r) = \frac{DD(r) - 2DR(r) + RR(r)}{RR(r)}, \quad (3.2)$$

where $DD(r)$ is the observed distribution of pair separations, $DR(r)$ is the distribution of separations between an observed sample and a random one, $RR(r)$ is the simulated random pair distribution. $\xi(r)=0$ can be interpreted as a random distribution at separation of r , while positive or negative amplitudes correspond to excess or deficit.

With the Python/C library [FragMent](#) introduced by [Clarke⁷](#), we first calculated the pair separation distribution ($DD(r)$) of 66 ALMA cores after straightening the filament. We then used Monte Carlo simulation to create random separation distributions ($DR(r)$ and $RR(r)$) by placing 10,000 random points within the ALMA mapping area. All the separation distribution points were convolved with a Gaussian kernel and finally led to adequate statistics. The uncertainty of TPCF follows the relation [Sawangwit et al. \(2011\)](#)

$$\sigma_\xi = \sqrt{\frac{1 + \xi(r)}{DD(r)}}. \quad (3.3)$$

The TPCF is shown in [Fig. 3.12](#), together with a zoom-in image for distances within 0.5 pc. It has a significant peak at 0.080 pc, which is consistent with the peaks derived by MST, 0.088 pc. Two other peaks exist at longer distances, 0.235 pc and 0.370 pc, indicating possible three level fragmentation. When the separation exceeds 0.5 pc, the TPCF and its errors are partly below the threshold zero, hence not considered for further peaks. The high peaks at the tail can be neglected since they are at a scale of the whole mapping field, and the large error prevents convincing explanation.

3.3 Mass estimation

Dust grains produce continuum emission, following the rule of blackbody radiation. Therefore, the measured continuum flux depends on $B_\nu(T_d)$, the Planck function for a blackbody of dust temperature T_d , distance to the source D , and the dust emissivity. Assuming the emitting source is optically thin, the flux follows:

$$S_\nu = \frac{4\pi a^3}{3} \cdot \rho N \kappa_\nu B_\nu(T_d) / D^2, \quad (3.4)$$

where a is the grain radius ($0.1 \mu\text{m}$), ρ is the grain density (3 g cm^{-3}), N is the total number of grains, which forms the total mass of dust. The dust mass opacity coefficient κ_ν is frequency-dependent:

$$\kappa_\nu = \kappa_0 \left(\frac{\nu}{\nu_0} \right)^\beta, \quad (3.5)$$

⁷<https://github.com/SeamusClarke/FragMent>

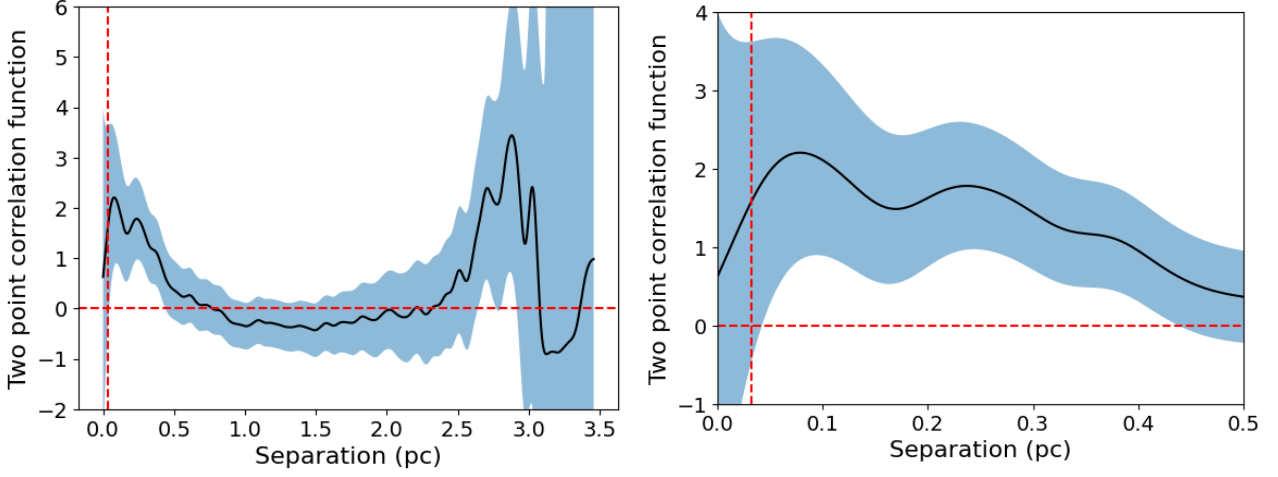


Figure 3.12: Left: Two-point correlation function of the continuum source distribution in G327. The blue areas show the 1σ confidence intervals. The dashed vertical line corresponds to the resolution limit, and the dashed horizontal line indicates a uniform distribution. Right: A zoom-in image for distances within 0.5 pc.

where $\kappa_0=1 \text{ cm}^2/\text{g}$ at $\nu_0=250 \text{ GHz}$ (Ossenkopf and Henning, 1994), and $\beta=2$, as in Beuther et al. (2002). Then the gas mass can be derived with a dust-to-gas ratio $R_d=0.01$:

$$M = \frac{S_\nu D^2}{\kappa_\nu B_\nu(T_d) R_d}. \quad (3.6)$$

In practice, we sum up the continuum emission around 230 GHz (1.3 mm) within each core to calculate the flux, listed in the seventh column of Table 3.1. However, the original integrated intensity is usually expressed in units of Jy/beam, which needs to be converted to Jy by the equation: total flux = flux (Jy/beam) summed over a number of pixels / number of pixels in a beam, where the number of pixels in a beam = beam area (arcsecond²) / (pixel length)², and beam area = $\pi \times B_{maj} \times B_{min} / 4 \ln 2$.

According to the above discussion, the total flux value S_ν is computed with a pixel length of $0.1''$, B_{maj} of $2.2''$, B_{min} of $1.8''$, on the background-subtracted image and the original image. The dust temperature is assumed to be the same as the kinetic temperature, which is approximately the same as the excitation temperature under local thermodynamic equilibrium (LTE). Here we adopted the excitation temperature of H_2CO (detail explained in Section 4). Together with Equation 3.6, the masses of the dust continuum are estimated and are listed in the last column of Table 3.1 and are plotted in the bottom left panel of Figure 3.6. The median mass of our cores is $4.4 M_\odot$, with a maximum of $163.6 M_\odot$, and a minimum of $0.9 M_\odot$. The minimum core mass is ten times larger than the possible minimum mass ($0.08 M_\odot$), which is the theoretical lower limit of stellar mass to trigger hydrogen burning. The maximum mass is for the hot core, which is lower than the previous results due to the better confined core size ($R_{major}=1.44''$, $R_{minor}=1.18''$). The model by Bergman (1992) gave a mass of $2000 M_\odot$ since it included the envelope mass as well. Using $450 \mu\text{m}$ APEX continuum observations, Minier et al. (2009) estimated the mass of the hot core as $3769 M_\odot$, which could be attributed to the lower resolution of the APEX image ($10''$) and, therefore, the larger source size and dust mass. Another observation from APEX $850 \mu\text{m}$ (Wyrowski et al., 2006) presented a mass of $420 M_\odot$ (only twice of ours), when using a dust emissivity of 1.8 and a temperature of 100 K. The difference in temperature (299 K in our case) explains the deviation of the mass. Although the calculation of mass can deviate within one magnitude due to different fluxes, dust emissivities,

distances and temperatures, the mass of the hot core is definitely much larger than $10 M_{\odot}$, highly supporting its future as a high-mass star or even star cluster. In addition, there are in total 8 cores with a mass larger than $20 M_{\odot}$, which have the potential to become future high-mass star-forming cores, although currently some of them are warm and quiescent, devoid of rich lines. There are 23 cores less than $3 M_{\odot}$. If we consider a star formation efficiency of 30%, it is very possible to end up with low-mass stars. The other 35 cores are possible precursors of intermediate-mass stars.

With the core sizes derived in the previous section, we also calculated the column density and the number density, shown in Figure 3.6. The median column density is $1.44 \times 10^{23} \text{ cm}^{-2}$, and the median volume density is $1.64 \times 10^6 \text{ cm}^{-3}$.

3.3.1 Core mass function (CMF)

The core mass function (CMF) usually follows a power-law spectrum such as $dN/dM \propto M^{-\alpha}$ in differential form. This was the form originally adopted by Salpeter (1955) to describe the stellar initial mass function (IMF), a $\alpha = -2.35$ was obtained by fitting the observational data. However, this function diverges as it approaches zero, so it is clear that there must be a break or turn-over in the IMF at low masses. Kroupa (2001) derived the IMF as a piecewise-defined function:

$$\alpha = \begin{cases} -0.3, & (m < 0.08) \\ -1.3, & (0.08 \leq m < 0.5) \\ -2.3, & (m \geq 0.5) \end{cases} \quad (3.7)$$

The form of the IMF at the low mass end is still relatively uncertain and subject to ongoing debate. The philosophical difference between one power-law or multiple sections is that whether one believes that star formation is a continuous process or whether there are distinct physical processes that dominate in certain mass regimes, e.g., low-mass and high-mass stars.

To explore the power index α more intuitively, people usually plot the CMF in logarithmic scale on both axes, i.e., $\log(M)$ as the x axis and $\log(dN/d(\log M))$ as the y axis. The slope of such a plot on logarithmic scale is exactly the power index. We can also study the cumulative form of CMF, that is, mass as the x axis and $\int N dM$ as the y axis.

With the masses of identified 66 cores, we derive the CMF in both differential form ($\log(dN/d(\log M))$ - $\log(M)$) and cumulative form ($\int N dM$ - M) shown in Fig. 3.13. The completeness limit is derived according to three times the noise level. Thus, the 3σ flux within a beam size at 50 K corresponds to a mass of $0.9 M_{\odot}$, which is marked in the CMF plot. And all of our cores are more massive than this limit. We fit a single power law at the mass range beyond the peak in the differential CMF, and got a power index of -0.83 with a small error of 0.01. This value is far from the slope of -2.35 (Salpeter, 1955), but is getting closer to the number 1.35 derived from the high mass end of the IMF (Offner et al., 2014). As a good comparison, we also plot the shifted IMF from Kroupa (2001) on top of our cumulative CMF (see the right panel of Fig. 3.13), which is much steeper ($\alpha=-1.3$) than our CMF ($\alpha=-0.90$).

Our results challenge the idea that the shape of the IMF is directly inherited from the shape of the CMF and that there is a self-similar mapping from the CMF to the IMF (Goodwin et al., 2008; Hopkins, 2013; Cheng et al., 2018). A flatter CMF shape in the whole section of our CMF

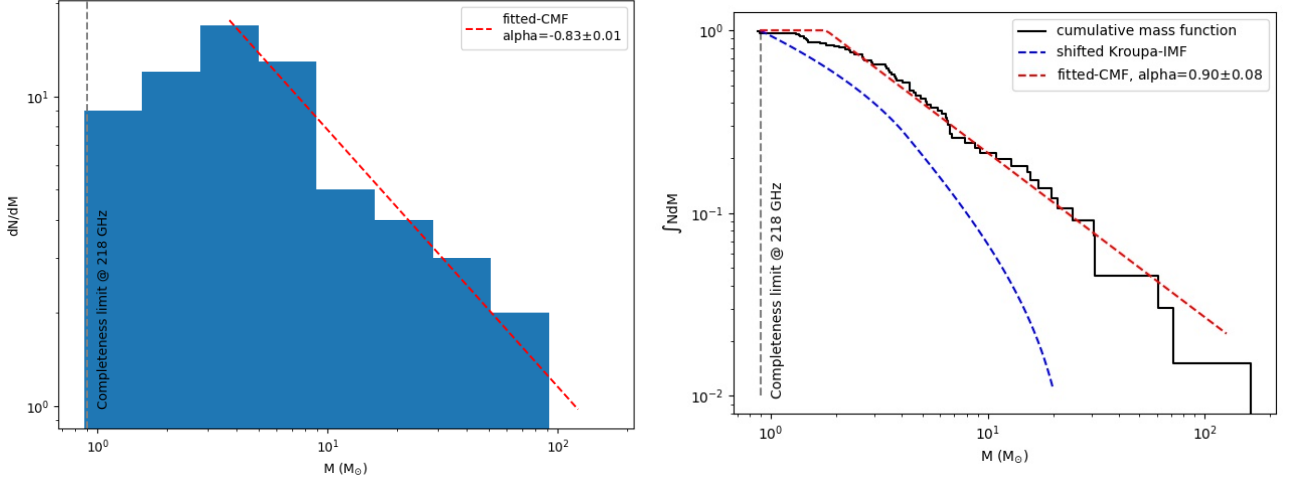


Figure 3.13: Left: Differential core mass function (CMF) of the identified 66 cores, fitted with a single power law index of -0.83. Right: Cumulative CMF. The fitting results shown in red dashed line corresponds to a power index α of -0.9, which is flatter than the shifted Kroupa’s IMF with α of -1.3.

may indicate a feature of HMSFRs according to [Li et al. \(2007\)](#). It is also possible that the index of CMF relies on different stages of star formation and does not have a universal form.

3.3.2 Larson’s law

With the known mass, source size, and volume density, we plotted them against each other to check the relationship. The plot on the left of Fig. 3.14 shows the core masses against the equivalent radius, that is, the mass-size relation. Plotted on logarithmic scale, we see a good correlation between these two quantities, and we fit a power law to the mass-size map, such as $M \propto r^{2.26}$. As for a constant column density, the theoretical mass-size relation would be $M \propto r^2$, with a smaller power law index than us. Therefore, we find a steeper relation, which means we have more high-mass/column-density cores in our samples. We also want to mention that $M \propto r^3$ holds for a constant volume density, which is not the case in our study. Usually the volume density decreases with the source size for star-forming cores.

Thus, we plotted the volume density versus source size in the right panel of Fig. 3.14. The fitted relation is $n \propto r^{-0.74}$. We note that in the classical 3rd Larson relation, [Larson \(1981\)](#) also found that for his sample of CO molecular clouds, the mean density is inversely related to the size ($n \propto r^{-1.1}$). Again, this indicates that we have larger volume densities at larger size than Larson’s low-mass samples. However, the mechanism of creating such high density and large size sources is not fully understood.

In the previous study, mass-size relations with various exponents can be found. While some studies report relations close to the classical Larson’s 3rd law ($\alpha=2$, e.g., [Heyer et al. \(2009\)](#); [Lombardi et al. \(2010\)](#)), steeper relations have also been found. For example, [Kainulainen et al. \(2011\)](#) found a mass-size relation with exponent 2.7 when sampling only the highest column density parts of their studied molecular clouds. It is also interesting to note that $M \propto r^3$ was found in the central molecular zone of our Milky Way ([Krieger et al., 2020](#)).

[Ballesteros-Paredes et al. \(2020\)](#) gave a possible reason for various mass-size relations with power-law indices from below 2 to 3. They claim that cores are usually derived from constant

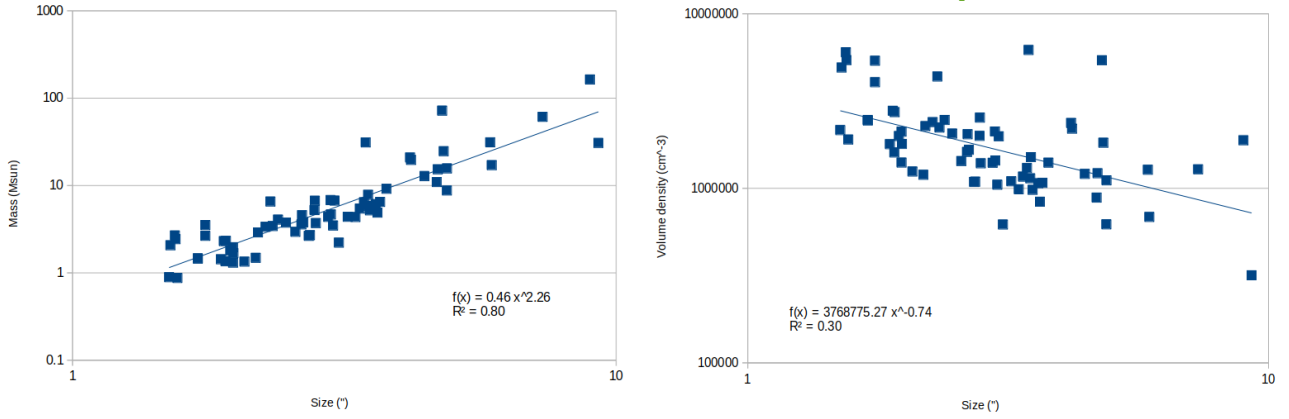


Figure 3.14: Left: The mass-size plot fitted with a power law index of 2.26. Right: The volume density-size plot in logarithmic scale, fitted with a power law of index -0.74.

column density thresholds (e.g., 5σ) if the filling factor of the densest portions of the cores is small, the mass-size relations should naturally become $M \propto r^2$. However, if the small filling-factor hypothesis failed for the densest gas in the continuum data, the cores may exhibit sharp boundaries and the lower column density material would not contribute to the mean column density. Such observational conditions would imply a steeper slope.

3.3.3 Mass segregation

Previous observations (Plunkett et al., 2018; Busquet et al., 2019) find that in young clusters containing high-mass stars, the most massive members tend to concentrate in the center, while low-mass members are distributed throughout the whole cluster. This effect is called *mass segregation*, and its study provides constraints for numerical simulations that describe the formation of stellar clusters.

Following Allison et al. (2009), the degree of mass segregation can be quantified by the mass segregation ratio (Λ_{MSR}):

$$\Lambda_{MSR} = \frac{\langle l_{random} \rangle}{l_{massive}} \pm \frac{\sigma_{random}}{\sigma_{massive}}, \quad (3.8)$$

where $l_{massive}$ is the mean MST length of the N most massive objects, with $\sigma_{massive}$ the standard deviation. $\langle l_{random} \rangle$ is the average of the mean MST lengths of randomly generated groups with N members uniformly distributed with a given standard deviation of σ_{random} . A value of $\Lambda_{MSR} \sim 1$ implies that massive objects are distributed in the same way as a uniform cluster, while $\Lambda_{MSR} > 1$ corresponds to mass segregation in the cluster, i.e. the MST length of the N most massive objects in the cluster is shorter, and they are more closely distributed compared to a uniform distribution. On the contrary, $\Lambda_{MSR} < 1$ suggests inverse mass segregation, with the N massive stars more spread outward than the rest.

We divided the cores of each cluster into low-mass ($M < 20 M_{\odot}$) and high-mass ($M > 20 M_{\odot}$) cores, and built individual MSTs for the high-mass populations (8 cores). Assuming a star formation efficiency of 30%, a compact source with a mass of $20 M_{\odot}$ ends up with forming a B-type star, thus allowing to separate between potential low-mass and high-mass stars. Therefore, we derived an $\langle l_{massive} \rangle = 35.9''$ and $\sigma_{massive} = 27.8''$.

For a good determination of $\langle l_{massive} \rangle$, we generated 500 random clusters with 8 sources

in the same area of the observation field, and built individual MSTs for each cluster. The final l_{random} is derived as $31.1''$, and σ_{random} $17.7''$. Thus, we obtained a Λ_{MSR} of 0.87 ± 0.64 , which is slightly below 1. This may suggest that the G327 cores are not uniformly distributed, even slightly spreading in mass. Other studies have found larger Λ_{MSR} , e.g., NGC 2023/2024 presents a $\Lambda_{MSR}=28$ for the 4 most massive cores, and NGC 2068/2071 shows mild levels of mass segregation ($\Lambda_{MSR}=2$) (Kirk et al., 2016). Compared to these values, our Λ_{MSR} is therefore too small to convince a mass segregation.

The Λ_{MSR} method has been modified by Olczak et al. (2011) using the geometric mean rather than the arithmetic mean to minimize the influence of outliers. This method works similarly to Λ_{MSR} by constructing the MST for the N most massive stars and determining the mean edge length $\gamma^{massive}$. The difference is that here we construct the MST of the same number of randomly selected stars from the entire sample and determine the mean edge length γ^{random} . Following Olczak et al. (2011), the mass segregation ratio, Γ_{MSR} , is then defined as:

$$\Gamma_{MSR} = \frac{\langle \gamma^{random} \rangle}{\gamma^{massive}} \pm \frac{\sigma_{random}}{\sigma_{massive}}. \quad (3.9)$$

Again, we applied 500 iterations to draw random clusters out of the 66 cores. The resulting γ^{random} is $37.4''$, with a σ_{random} of $33.4''$. Surprisingly, we obtained a Γ_{MSR} of 1.04 ± 1.20 , which is very close to 1 and shows a large uncertainty. Together with the results of Λ_{MSR} , the values close to 1 in both cases imply that the most massive and the randomly selected sources are distributed in a similar manner, or in other words, uniformly distributed.

Li et al. (2021) found a mass separation at $10 M_{\odot}$, and claimed that the masses of low-mass cores ($M_{core} < 10 M_{\odot}$) are determined by fragmentation, whereas massive cores ($M_{core} > 10 M_{\odot}$) grow mostly through accretion. We also evaluate this mass separation and find 14 cores with mass larger than $10 M_{\odot}$. We derived $\langle l_{massive} \rangle = 26.91 \pm 25.28$, $\langle l_{random} \rangle = 22.55 \pm 12.37$, and $\gamma^{random} = 25.94 \pm 24.27$. Therefore, we got $\Lambda_{MSR} = 0.84 \pm 0.49$, $\Gamma_{MSR} = 0.96 \pm 0.96$, which are both close to 1 and with large error range. This again proves our sources are uniformly distributed.

3.4 Fragmentation

Fragmentation is a universal phenomenon in molecular clouds, discovered, e.g., SgrB2 (Sánchez-Monge et al., 2017), NGC6334 (Sadaghiani et al., 2020), and G28 (Wang et al., 2011). A clump fragments into cores with its initial physical conditions including density, temperature, turbulence, etc. Some cores continue to fragment into condensations when density and temperature increase. Thus, we discuss the fragmentation of G327 in two different scales, filament and core, respectively.

Above all, we estimate the observational mass density averaged over the filament for fragmentation analysis. The continuum map (Fig. 3.1) contains $725 M_{\odot}$ in the identified leaf structures, leading to a linear mass density of $453 M_{\odot} \text{pc}^{-1}$ with a filament length of 1.6 pc ($100''$). From the MST analysis, the projected median separations of the cores are measured as 0.15 pc ($9.5''$) from the nearest neighbor algorithm, thus a segment mass of $68 M_{\odot}$ is derived. At the core level, we computed the core masses in Section 3, $4.4 M_{\odot}$ as median value. The corresponding median source size is 0.03 pc from the dendrogram ellipses and 0.045 pc from the approximation of identified leaves.

3.4.1 Cylindrical fragmentation

In this section, we compare the observational results with the theoretical predictions of cylindrical fragmentation. The continuum cores appear to be regularly spaced along a 'cylinder', although it presents a turning point between the north-eastern and south-western segments. This fragmentation in cylinder was first explained by 'sausage instability' when assuming an isothermal gas cylinder (Chandrasekhar and Fermi, 1953), in which self-gravity overcomes internal pressure (represented by σ) once the mass per unit length along the cylinder, i.e. linear mass density, exceeds a critical value of

$$(M/l)_{crit} = 2\sigma^2/G = 465(\frac{\sigma}{1 \text{ km s}^{-1}})^2 M_{\odot} \text{pc}^{-1}. \quad (3.10)$$

Here, if the cylinder is supported by thermal pressure, σ is the sound speed c_s ; if, on the other hand, it is mainly supported by turbulent pressure, then σ is replaced by the velocity dispersion. When the observed linear mass density is far beyond the critical one, the cylinder becomes gravitationally unstable and fragments into a chain of equally spaced fragments with a typical spacing of

$$\lambda_{cl} = 22v\sqrt{4\pi G\rho_c} = 1.24 \text{pc} (\frac{\sigma}{1 \text{ km s}^{-1}})(\frac{n_c}{10^5 \text{ cm}^{-3}})^{-1/2}. \quad (3.11)$$

In the above two equations, ρ_c (n_c) is the gas (volume) density in the center of the cylinder. The fragment mass is therefore

$$M_{cl} = \lambda(M/l)_{crit} = 575.3 M_{\odot} (\frac{\sigma}{1 \text{ km s}^{-1}})^3 (\frac{n_c}{10^5 \text{ cm}^{-3}})^{-1/2}. \quad (3.12)$$

These numbers in the simplified equation are adopted from Wang et al. (2014)

Assuming the cylinder temperature is the median excitation temperature (50.0 K), the thermal velocity is $\sigma_{thermal} = c_s = \sqrt{\frac{k \cdot Tex}{m_H \cdot \mu}} = 0.38 \text{ km s}^{-1}$. Together with the volume density in the central hot core, $5.43 \times 10^6 \text{ cm}^{-3}$, it yields a linear mass density of $67 M_{\odot} \text{pc}^{-1}$, fragment spacing of 0.06 pc and segmentation mass of $\sim 4.3 M_{\odot}$.

For the turbulent support, the related non-thermal motion velocity is derived with $\sigma_{nontherm} = \sqrt{(\frac{dv}{2\sqrt{2}ln2})^2 - \frac{k \cdot Tex}{m_{H2CO} \cdot \mu}}$, in which dv is the average line width through the filament. With the line width distribution map in Section 5.3, we obtain $dv \sim 3.1 \text{ km s}^{-1}$ and $\sigma_{nontherm} \sim 1.31 \text{ km s}^{-1}$, hence the corresponding linear mass density, fragment spacing and mass are $798 M_{\odot} \text{pc}^{-1}$, 0.22 pc and $176 M_{\odot}$, respectively.

Interestingly, the observed linear mass ($453 M_{\odot} \text{pc}^{-1}$) far exceeds the thermal expectations ($67 M_{\odot} \text{pc}^{-1}$), indicating a large-scale on-going fragmentation. At the same time, the observed linear fragment spacing (0.15 pc) locates between the value derived by thermal and non-thermal, indicating a dominating role of non-thermal motion. However, it is worth to mention that these projected separations are measured as 1D, which can be underestimated partly due to projection effect and partly because of high resolution. If we consider a correction factor of $\sqrt{3}$ to obtain the 3D distance from 1D, the observed fragment spacing becomes 0.26 pc, which is larger than the non-thermal fragmenting length. In any case, the filament definitely exceeds the thermal Jeans length, while we cannot tell whether the filament is unstable and tends to fragment into smaller pieces or not. It requires further observation with higher resolution.

In an ALMA study of very young high-mass star-forming regions, Svoboda et al. (2019) estimated the cylindrical fragmentation scale to be roughly a factor of 3.5 larger than the typical

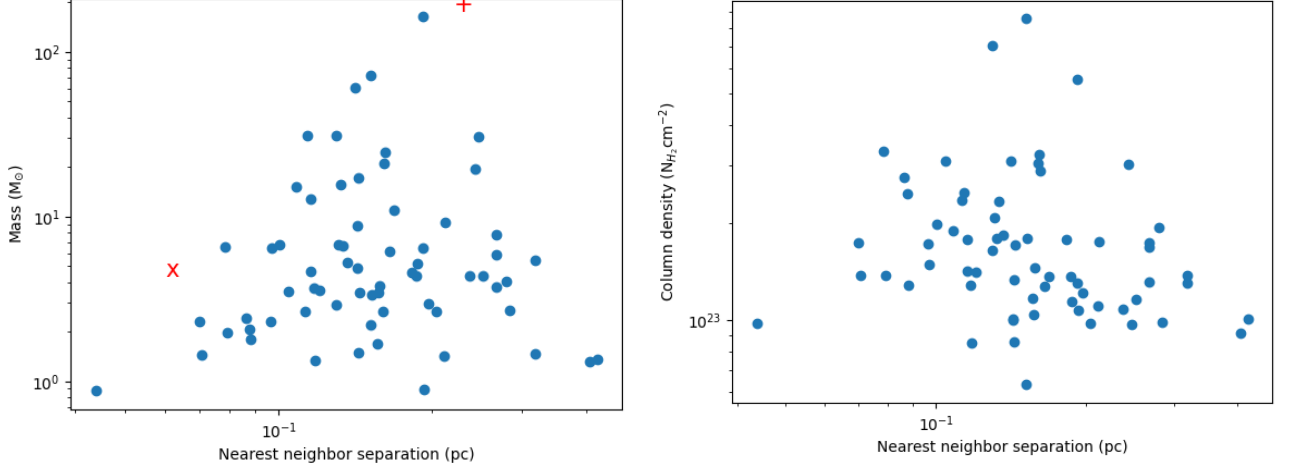


Figure 3.15: Left: Core masses against nearest-neighbor separations from the minimum spanning tree analysis, plotted in logarithmic scale. The thermal and non-thermal cylindrical Jeans parameters are shown in red cross and plus sign respectively. Right: Column density of each core versus separations.

thermal Jeans length. This is not far away from what we found. There are also reports in the literature supporting the claim of more important turbulent pressure than thermal pressure (Wang et al., 2011; Zhang et al., 2015; Sadaghiani et al., 2020).

Mass-separation distribution The nearest neighbor separation is defined as the average length of the MST edge for a certain core. If the core is only connected with one other core, the edge length is the nearest neighbor separation. Thus, we plotted the left panel of Figure 3.15 to present the core masses versus the nearest neighbor separations for the individual cores, which seems like scattered in a uniform distribution. In order to correlate the image with the cylindrical Jeans analysis, we also plotted the Jeans lengths and Jeans masses for both the thermal and non-thermal case. The thermal data point is marked with a red cross in the left corner of the distribution. Since most of the cores show a larger separation than the thermal one, they tend to fragment when only thermal pressure provides the support. The non-thermal data point shown in red plus sign is on the other hand far away at the top right corner, which means most of the cores are not reaching the non-thermal Jeans length or mass. This phenomenon tells us that our cores are not exceeding the critical mass or length to fragment further due to the non-thermal pressure such as turbulence. Such an analysis has been done in Beuther et al. (2021), showing a similar global trend for individual cores in two young high-mass star-forming regions, ISOSS22478 and ISOSS23053.

We also plotted the column density distribution with respect to the nearest-neighbor separations in the right panel of Fig. 3.15, which shows a random distribution by eye. Since the column density is proportional to the surface density, what we expect is a power law relation, as predicted by Li et al. (2021), $\Sigma_{\text{edge}} \propto l_{\text{core}}^{-0.28}$. The authors also claimed that fragmentation is controlled by a scale-dependent turbulent pressure. Maybe we have a homogeneous turbulence through the cores, or we overestimated the surface density due to a high column density within the cores rather than averaging the filament.

3.4.2 Core fragmentation

The previous discussion is on filament scale, and for a core, the Jeans length is defined differently as:

$$\lambda_J = c_{eff} \left(\frac{\pi}{G\rho} \right)^{1/2} \left(\frac{n_c}{10^5 \text{ cm}^{-3}} \right)^{-1/2}, \quad (3.13)$$

where c_{eff} is the effective sound speed defined above; ρ is the density at a hydrogen number (volume) density n_{H_2} of the sphere, that is, the average molecular mass 2.72, times the atomic hydrogen mass 1.6735×10^{-24} g. The Jeans mass is the mass of the gas in a sphere with a radius of $\lambda_J/2$:

$$M_J = \frac{\pi^{5/2}}{6G^{3/2}} c_{eff}^3 \rho^{-1/2}. \quad (3.14)$$

We assume that the thermal pressure is dominant in the fragmentation process, c_{eff} corresponds to the sound speed of the gas and thermal Jeans mass becomes:

$$M_J^{thermal} = 0.6285 M_\odot \left(\frac{T}{10 \text{ K}} \right)^{3/2} \left(\frac{n_c}{10^5 \text{ cm}^{-3}} \right)^{-1/2}, \quad (3.15)$$

in which T is the kinetic temperature. When turbulence dominates the pressure, the non-thermal fragment mass is therefore

$$M_J^{non-thermal} = 0.8255 M_\odot \left(\frac{\sigma}{0.188 \text{ km s}^{-1}} \right)^3 \left(\frac{n_c}{10^5 \text{ cm}^{-3}} \right)^{-1/2}, \quad (3.16)$$

with a non-thermal velocity dispersion σ . The values in the equations are adopted from [Palau et al. \(2015\)](#).

Under the condition of a median temperature 50 K, a sound speed 0.38 km/s, a non-thermal velocity dispersion 1.31 km/s, and a median volume density $1.64 \times 10^6 \text{ cm}^{-3}$, the thermal Jeans length and mass are estimated to be 0.03 pc and $1.74 M_\odot$, while the non-thermal Jeans length and mass are 0.11 pc and $68.97 M_\odot$. The thermal Jeans length is comparable with the real core size of 0.03–0.045 pc, whereas the observed median mass of $4.4 M_\odot$ is 2–3 times larger than the expected thermal Jeans mass and 16 times lower than the non-thermal Jeans mass. This again suggests that turbulence is the dominating factor in core fragmentation, which stops the cores from forming subcore features. Similar results are found in the massive star formation region G28.34-P1 ([Wang et al., 2011](#)). For low-mass star-forming regions, it usually suggests thermal Jeans fragmentation, e.g., in NGC6334-I(NW) [Sadaghiani et al. \(2020\)](#).

Apparently, thermal Jeans fragmentation can not fully explain our data, and the observations favor the turbulent scenario against the thermal one in the fragmentation from the filament (~ 1 pc) to the core scale (< 0.1 pc), which is the same case in [Liu et al. \(2018\)](#). However, the G327 filament may be far from equilibrium and can fragment into smaller segments, which requires further observation with higher resolution.

4 Molecular spectral analysis

In this section, we adopted molecular spectral lines to analyze the physical environment and the chemical evolutionary status using radiative transfer theory and an automatic line fitting tool XCLASS (Möller et al., 2017). In the subsections, we introduce the approximation and condition used in the spectral line analysis (e.g. LTE and non-LTE radiative transfer), the algorithms to promote the fitting speed, and the error estimation for the results. In the end of the section, we test an unsupervised machine learning technique, Principal Component Analysis (PCA), to characterize the spectra and classify them into groups that are in different evolutionary stages.

4.1 Line fitting basics

To fit a spectral line, we need to know the relationship between the strength of the line and the corresponding physical conditions, including temperature, optical depth, and density. Thus, we first introduce the theory of radiative transfer and then give the solutions under (non-)local thermodynamic equilibrium in detail.

4.1.1 Radiative transfer

The propagation of radiation through a medium is affected by absorption, emission, and scattering processes. The theory of radiative transfer is developed to describe these interactions mathematically, which is valid in a wide variety of stellar objects and frequency ranges. The analytic solutions to the equation of radiative transfer are derived for simple cases, which helps to evaluate the temperature and density through the strength of spectral lines. For a more complex scenario, numerical methods are required to solve the equation. We follow Draine (2011) to discuss the theory of radiative transfer.

Thinking of the energy observed by a telescope along the line of sight, it is a function of the on-sky projection within a solid angle and of the frequency. Now when we consider the emission with intensity I passing through an interstellar medium with an absorption coefficient κ and emission coefficient j_ν at frequency ν , the change in intensity along the distance s is given by

$$\frac{dI_\nu}{ds} = -\kappa_\nu I_\nu + j_\nu. \quad (4.1)$$

The emission of a certain line transition consists of both spontaneous and stimulated emission from an upper energy level to a lower one. The absorption coefficient determines the percentage of photons reabsorbed to excite the same line, but inversely forms lower to upper energy levels. Then the optical depth τ is defined as the integrated absorption along the path length s of photon before absorption:

$$\tau_\nu = \int_0^s \kappa_\nu(s) ds. \quad (4.2)$$

We define the ratio $\frac{j_\nu}{\kappa_\nu}$ as the source function S_ν , representing the ISM emission. Together

with the definition of τ , equation 4.1 then becomes:

$$\frac{dI_\nu}{d\tau} = S_\nu - I_\nu \quad (4.3)$$

Integrating the equation on both sides and assuming S_ν is constant, we derive:

$$I_\nu = I_\nu^0 \exp^{-\tau_\nu} + S_\nu(1 - \exp^{-\tau_\nu}), \quad (4.4)$$

in which the first term explains how a background intensity I_ν^0 drops down after passing through the ISM, and the second term reflects the re-emission of the ISM. Equation 4.4 is called the **formal radiative transition equation** for a simple situation with a known fixed source function. If the source function is not homogeneous and changes with the distance throughout the ISM, this equation does not hold, and one should turn to a numerical solution.

The typical on-off observation mode provides two close-by observations, one targeting on the target cloud and one on the background. By subtracting the background emission from the total emission, we derive the intensity of the target source as $I_{source} = S(T)(1 - \exp^{-\tau})$.

In ISM, it is essential to distinguish the optically thin and thick cases according to the τ value. For optical depths $\ll 1$, the emission is considered to be optically thin, and hence the emission from the entire line of sight is recovered. In this case, $\exp^{-\tau}$ is approximately equivalent to $1 - \tau$. Thus, the resulting intensity can be simplified as $S(T)\tau$. Optically thick ($\tau \gg 1$) on the other hand infers that only emission from a finite depth into the ISM is observed, while all emission from further depth is fully absorbed and cannot be recovered. Since $\exp^{-\tau}$ approaches zero when $\tau \gg 1$, the final intensity will not change with optical depth and only depends on $S(T)$.

In the ISM where gas and dust are thermalized, the particle occupation in different energy states follows the thermal distribution, i.e. they can be described by the Boltzmann distribution at a certain kinetic temperature T_k . This connects temperature and emission using a blackbody distribution $I_\nu = B_\nu(T_k)$. This is known as **local thermal equilibrium** (LTE) assumption. In the optically thick case, the source function S is equal to a thermal radiation field B at temperature T .

4.1.2 Non-LTE fitting

The local thermodynamic equilibrium (LTE) condition does not hold in many cases, e.g., hot stars and super-giants (Gieser et al., 2019). In these situations, the density is very low and temperature is high; hence the radiative transitions are more frequent than collisional transitions, which requires a non-LTE description. Now we discuss this with equations.

The collisions between atoms or molecules can produce transitions from any excitational state i to another state j . Assuming that for every state i , the rate of being excited out of this state is equal to the rate of being re-populated from the other levels j , the equation becomes:

$$n_i \sum_{j \neq i}^N P_{ij} = \sum_{j \neq i}^N n_j P_{ji}, \quad (4.5)$$

where P_{ji} is the formation rate of state i , and P_{ij} is the corresponding destruction rate. They include the contribution of spontaneous emission A, the absorption B, and the collision C. Then P is written as:

$$P_{ij} = \begin{cases} A_{ij} + B_{ij}u_{ij} + C_{ij}, & (i > j) \\ B_{ij}u_{ij} + C_{ij}, & (i < j) \end{cases} \quad (4.6)$$

We use $B_{ij}u_{ij}$ as the probability of induced emission, and $B_{ji}u_{ij}$ as the probability of absorption, where u is the local radiative energy density. We define $u_{ij} = \int_0^{\text{inf}} \phi_{i,j}(\nu) I_\nu d\nu$, in which I_ν is the radiative field and $\phi_{ij}(\nu)$ represents the line profile for the transition from state i to j . Equation 4.5 becomes:

$$\sum_{j>i} [n_j A_{ji} + (n_j B_{ji} - n_i B_{ij}) u_{ji}] - \sum_{j<i} [n_i A_{ij} + (n_i B_{ij} - n_j B_{ji}) u_{ij}] + \sum_{j \neq i} [n_j C_{ji} - n_i C_{ij}] = 0, \quad (4.7)$$

where C_{ij} represents the collision rate per second per particle of interest. And it is related to the density of the collision partner n_{col} , hence written as $C_{ij} = K_{ij} n_{col}$. The collision partner can be H2, H, ortho-H2, para-H2, etc. In most cases, they are H2. The collision rate coefficient K_{ij} is the velocity-integrated collision cross section in units of $\text{cm}^3 \text{s}^{-1}$. When collisions dominate, i.e. more collisions than radiations ($n_j C_{ij} \gg A_{ij} + B_{ij} u_{ij}$), the collision rates are defined through detailed balance:

$$K_{ji} = K_{ij} \frac{g_i}{g_j} e^{-h\nu/k_B T_{kin}}. \quad (4.8)$$

Hence the level populations are described mainly by collision. In Equation 4.7, we retain the collision term while omitting the radiations, then we obtain the level population ratio as

$$\frac{n_j}{n_i} = \frac{C_{ij}}{C_{ji}} = \frac{g_j}{g_i} e^{-\Delta E_{ji}/k_B T_{kin}}, \quad (4.9)$$

where $\Delta E_{ji} = E_j - E_i$ represents the energy difference between the two levels, and T_{kin} is the kinetic temperature.

Conversely, when radiation dominates and $i > j$, the level occupations can be expressed as:

$$\frac{n_j}{n_i} = \frac{P_{ij}}{P_{ji}} = \frac{A_{ij} + B_{ij}u_{ij}}{B_{ji}u_{ji}} := \frac{g_j}{g_i} e^{-\Delta E_{ji}/k_B T_{rad}}, \quad (4.10)$$

Then people define the critical number density ($N_{crit} = \frac{A_{ij}}{C_{ij}}$) as the density above which collisions are frequent enough to use LTE, while below which it deviates from LTE.

In most cases, $n_j C_{ij} \approx A_{ij} + B_{ij} u_{ij}$, at which we defines an excitation temperature through:

$$\frac{n_j}{n_i} = \frac{g_j}{g_i} e^{-\Delta E_{ji}/k_B T_{ex}}. \quad (4.11)$$

Note that the excitation temperatures are different for different lines/transitions. In general, the radiation temperature is not described by the Boltzmann distribution, hence, not excitation temperature. We can compare the derived excitation temperature and the radiation temperature to determine the LTE state. LTE approximation is valid when the temperatures are similar to each other for all transitions, while under non-LTE, the excitation temperature is higher than any of the radiation temperatures. In some extreme cases, such as the maser source, the radiation temperature can even be negative.

4.1.3 Line profiles

There are some other important features of the spectral line that are not directly shown in the radiative transfer and may affect the standard fitting process. Therefore, we discuss the line profiles in this section.

In addition to basic radiative transfer theory, we understand how spectral lines form and when they appear as emission or absorption. Suppose the radiation $I_\nu(0) = B_\nu(T_0)$ goes through a layer of ISM with $S_\nu = B_\nu(T_1)$, so when $T_1 > T_0$ this produces an emission line, and if $T_1 < T_0$ it produces an absorption line.

Line width is an important parameter for describing a spectral line. In an ideal case, the line will appear as a delta function at a well-defined frequency corresponding to the excitation energy. However, the line has a width as a result of the uncertainty principle, called natural broadening effect. And in most astrophysical situations, the thermal distribution of particles cause velocity Doppler shifts to build a line profile as a Gaussian shape. In the cases of non-thermal dispersion, e.g., collision and turbulence, the line width will be greater than the thermal one (Ward-Thompson and Whitworth, 2011).

In a real physical case of a deeply embedded star-forming core, many components, including hot core, outflow, and envelop, may exist along the line of sight and cause an extremely complicated line profile and large line width. The superposition and interaction of components may present some special signatures of particular star-forming activity through the distinct profile.

The self-absorption feature can be translated as an infalling gas. When we have a warm collapsing core surrounded by a cool envelope, the outer cool component strongly absorbs the emission and produces a double-peaked profile. This effect is known as self-absorption. It happens only in the optically thick case, for example, in CO as it is the second most abundant molecule. The low J-level CO transitions are often optically thick, causing saturation of line intensities, i.e. many transitions show the same brightness and the line center is flattened. Self-absorption can easily occur here.

The asymmetry in the line profile has various explanations. Considering a spherical collapse with higher temperature and infall speed towards the inner region (also known as *inside-out collapse*, see a sketch at Fig. 4.1), more attenuation will occur at the red-shifted components since the hot photons created at the central fast-falling region tend to be absorbed by the cooler outer region. For the blue-shifted components, absorption is more difficult since the cold slow photons cannot catch up with the hot quick ones. Thus, a double-peaked profile with a higher blue-shifted peak is the hint of infall, named blue-asymmetric line profile. For similar reasons, a red-asymmetric line profile represents the expanding envelop.

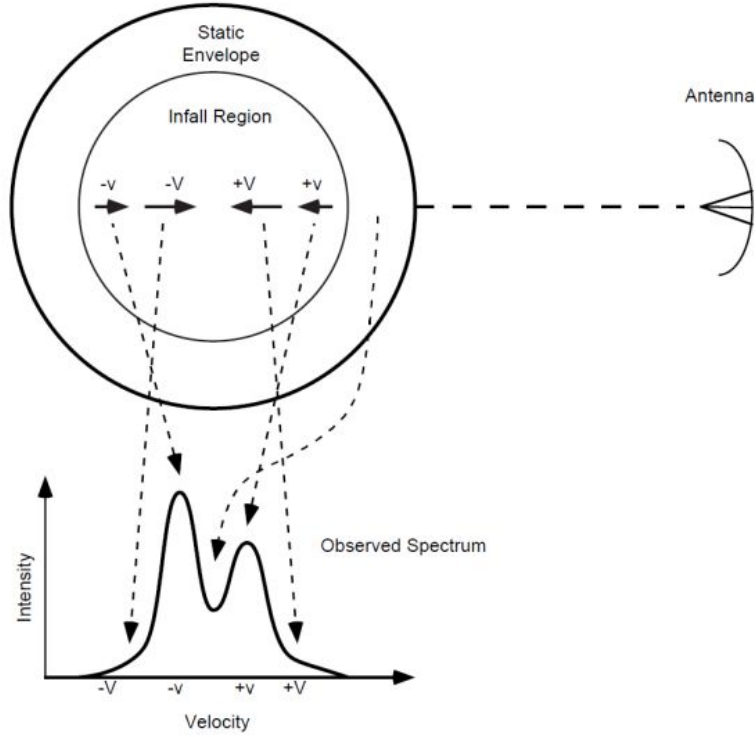


Figure 4.1: The origin of various parts of the line profile for a cloud undergoing inside-out collapse. The static envelope produces the central self-absorption dip, the blue-shifted peak comes from the back of the cloud, and the red-shifted peak from the front of the cloud. The faster collapse near the center produces line wings, which are usually mixed with outflow wings.

Line profiles do not necessarily have to be Gaussian. One example is line wings, which is the flux excess towards a Gaussian fit at the border of the line. They appear as high-velocity tails on both the red and blue sides when accreting protostars drive fast (a few to a thousand km/s) and bipolar jets. The profile can skew toward either red or blue side if the outflow is not symmetric. In these situations, the line widths are much larger than the typical $\sim 2\text{--}5$ km/s for quiescent gas (Bally, 2016).

4.1.4 Software: XCLASS for spectral analysis

The eXtended CASA Line Analysis Software Suite (XCLASS) is a toolbox adopted for spectral analysis, especially for modeling interferometric and single dish spectra (Möller et al., 2017).

The toolbox provides a function `myXCLASS` to model the data by solving the radiative transfer equation for an isothermal object in 1D, whereas the finite source size and dust attenuation are considered as well. XCLASS requires molecular data from the Cologne Database for Molecular Spectroscopy (CDMS)⁸ and the Virtual Atomic and Molecular Data Centre (VAMDC) (Dubernet et al., 2016).

Additionally, the toolbox contains an interface for the model optimizer package MAGIX (Modeling and Analysis Generic Interface for external numerical codes, Möller et al. (2017)), which helps to find the best description of observational spectra using a certain model, i.e., find the parameter set that reproduces the data most accurately.

⁸<https://cdms.astro.uni-koeln.de/cdms/portal/>

Furthermore, the XCLASS toolbox provides several functions (myXCLASS, myXCLASSFit, and myXCLASSMapFit) to simplify the interface for MAGIX using the myXCLASS program. The myXCLASS function is useful for producing synthetic spectra together with the quantum numbers of transitions. The myXCLASSFit function fits single spectra, whereas the myXCLASSMapFit function fits one or more complete FITS data cubes.

The basic idea of the fitting is explained below. Considering m molecules and c core components, the solution of the radiative transfer equation is:

$$T^{core}(\nu) = \sum_m \sum_c \left[\eta(\theta^{m,c}) [S^{m,c}(\nu)(1 - e^{-\tau^{m,c}(\nu)}) + T_{bg}^{core}(\nu)(e^{-\tau^{m,c}(\nu)} - 1)] \right] + (T_{bg}^{core}(\nu) - T_{CMB}), \quad (4.12)$$

Where T is the antenna temperature, η is the beam filling (dilution) factor, θ is the angular size, S represents the source function in form of temperature, τ is the optical depth of all lines, T_{bg} is the beam-averaged continuum background temperature, T_{CMB} is the cosmic background temperature.

Hence, for a given set of source size, excitation temperature, column density, velocity offset, and line width, the intensity at a certain frequency can be derived. The goodness of a model is described by the χ^2 distribution, which is a function of the relative quadratic differences between the observational and model values. A widely used fast-converging algorithm to get the smallest χ^2 is the Levenberg-Marquardt algorithm, although it can get stuck in local minima.

The myXCLASSFit function uses an algorithm or a combination of several of them in MAGIX to optimize the input parameters, when iteration times, path, and name of observation data files are specified. Isotopologues are also taken into account to reduce the number of input parameters. Finally, it returns the optimized parameters and corresponding modeled spectra from the best fit.

With the 66 cores selected using the Dendrogram, detailed analysis could be applied to obtain the chemical and physical properties. Toward these cores, we extracted the spectra from 12m-only cube for four spectral windows and then averaged them, see examples in Figure 4.2. All cores show strong CO emission lines together with absorption feature due to the interferometer effect, while sometimes H₂CO and CH₃OH also show absorption for the same reason, which should be carefully examined during the line fitting and classification process. Most of the cores show several strong lines that trace extended gas and large-scale motion, including H₂S, OCS, ¹³CS, and DCN. Some cores also present more core-related species like HNCO and HC₃N, while some show complex molecules, including e.g., CH₃OCHO and C₂H₅CN. Therefore, a detailed line study is necessary to get a better understanding of the spectra and the condition of cores.

4.2 Hot core spectra

Toward the hot core, the spectra of the central pixel (RA 15:53:07.7549, DEC -54:37:06.271) were selected for a detailed fitting, see Figure 4.3. Each line was selected by the GetTransitions task in XCLASS to be assigned to a possible molecular transition. As mentioned before, the CO line is excluded from the XCLASS fitting, since it is highly affected by the interferometer and shows significant absorption. In total, there are 26 molecules and 39 isotopes taken into account; see the first column of Table 7.1 and Table 7.2 in the appendix. Next, each molecule was examined separately inside XCLASS-GUI, which provides an interface to produce synthetic

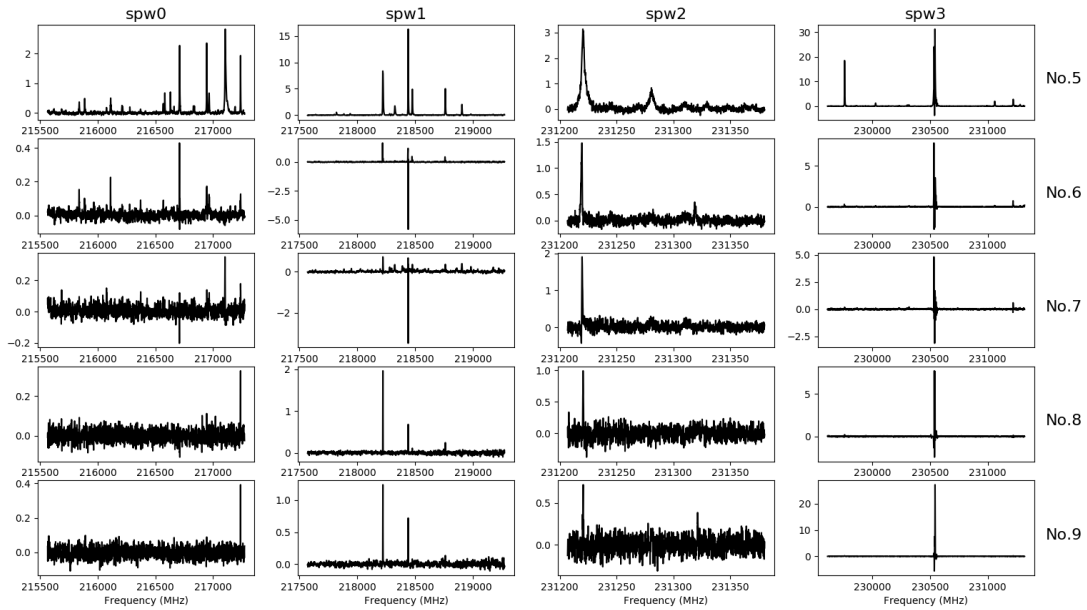


Figure 4.2: Examples of average spectra of derived cores in the four frequency ranges. Each row is one core and each column is one spectral window. See the Appendix for all of the average spectra.

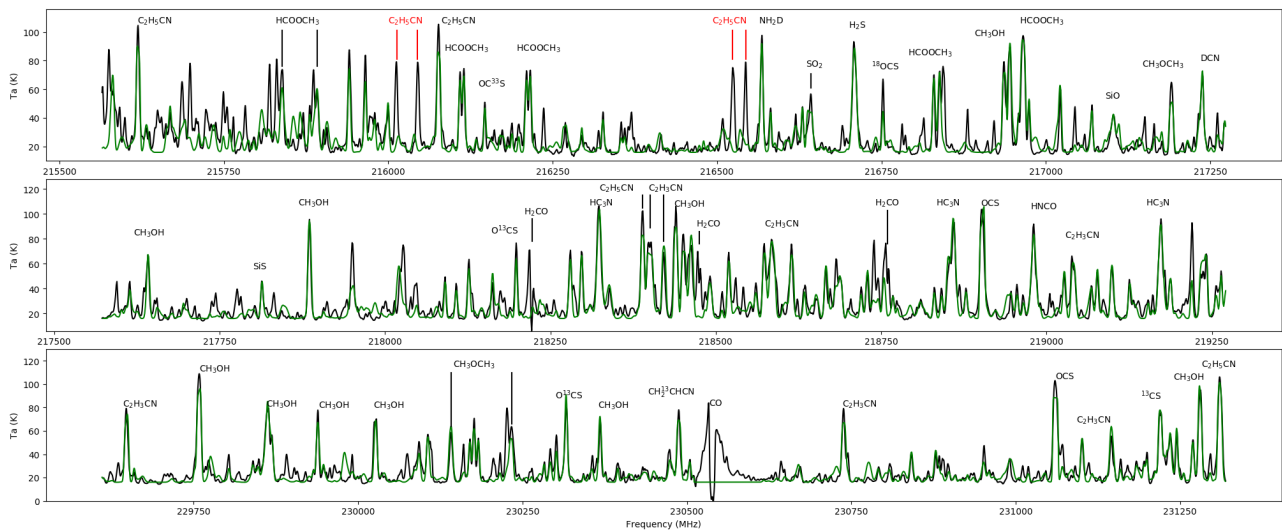


Figure 4.3: The best fit (green line) of hot core spectra for spw0 (top), spw1 (middle), spw3 (bottom), including 26 molecules and 39 isotopes, overlaid on the original data (black line).

spectrum at real time when the parameters were manually adjusted. The parameters, including source size, excitation temperature, column density, velocity dispersion, and central velocity, were set properly to obtain a synthetic spectrum similar to the observed one. If there is only one transition visible in our observing range, the excitation temperature will be fixed to 200 K. The single-molecule fitting is then done by the Levenberg-Marquardt algorithm in XCLASS-GUI. This algorithm calculates the gradient in the χ^2 space, thus it can quickly find a minimum χ^2 (possibly local) even if it starts very far away. It is also an algorithm that strongly depends on the starting values of the parameters, so we should set the initial guess carefully. Otherwise, the algorithm can easily become stuck in a local minimum of the global solution. The parameters fitted by GUI are used as the starting point for fitting all the molecules together with the myXCLASSFit function, to obtain a more reliable and satisfied result.

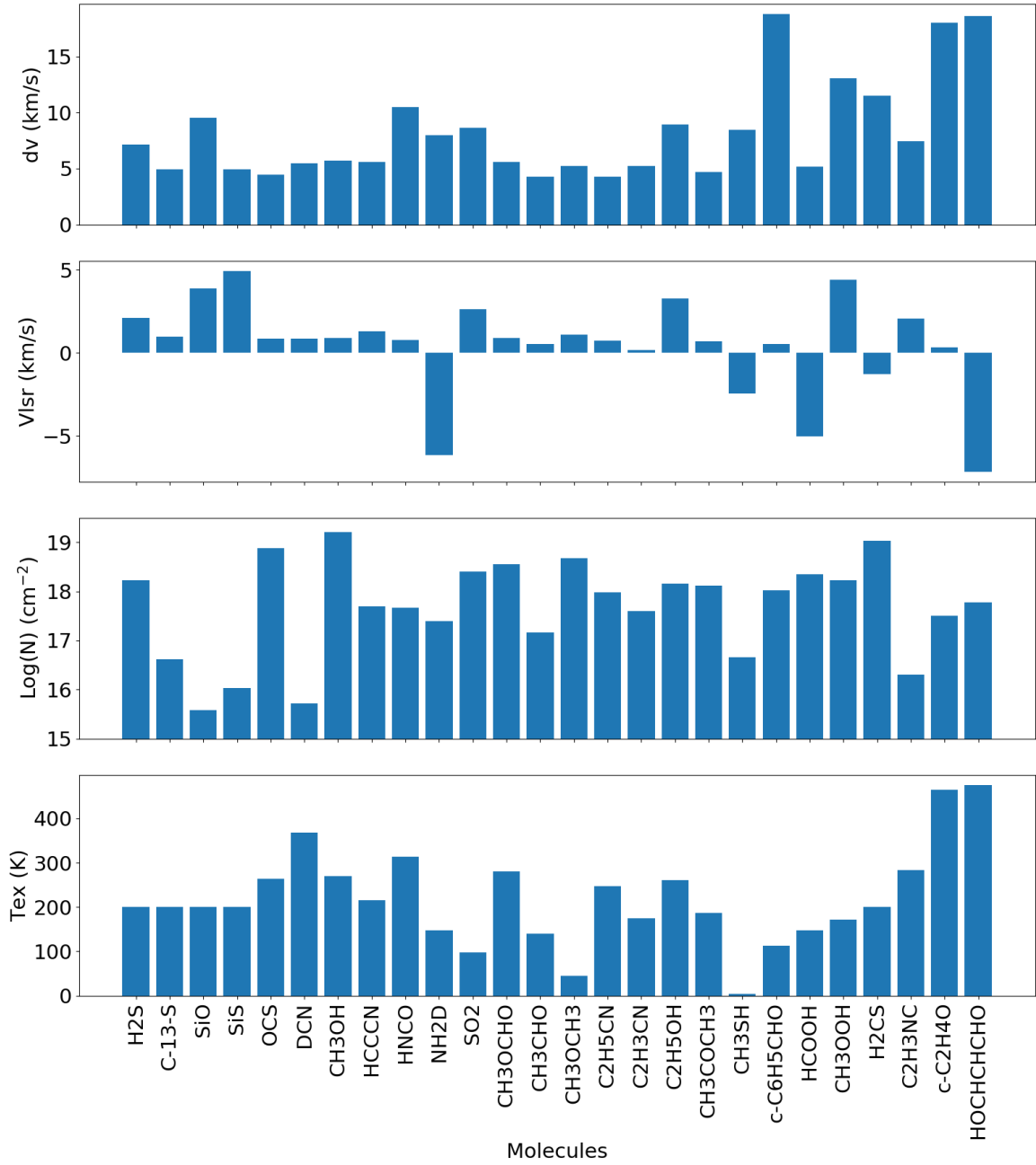


Figure 4.4: The visualization of myXCLASSFit results (Table 7.1) for the hot core.

Now, the initial guess of the parameters for all the molecules was acquired with high accuracy, and we kept the source size fixed and the same for all during the fitting, which

accelerates the fitting itself. The number of components for each molecule is set to one since only one clear velocity component is shown in G327. The fitting process took 7 to 8 hours for 10 iterations or until χ^2 reached 10^{-6} . To make it easier to read, the fitted parameters are plotted in Figure 4.4 for all molecules. The detailed number (XCLASS molfit file) is also listed in Table 7.1 in the appendix. The synthetic spectrum of the final result can be seen in Figure 4.3.

In general, the fitting result represented most of the lines with a small chi-square value, and the derived parameters are given within a reasonable range. In Figure 4.4, the molecules on the left side are generally more widely distributed and show stronger lines than on the right side. In addition to molecules that are set to 200 K, most of the excitation temperatures are also around 200 Kelvin, again proving the character as a hot core. Some extreme excitation temperatures like 465 K for $\text{c-C}_2\text{H}_4\text{O}$ and 34 K for CH_3SH may be fake, possibly due to their weak and blended lines. The column densities are one to two magnitudes higher than usual, indicating a well-evolved hot core at the position. The line widths are all within 20 km s^{-1} and much larger than the thermal line width ($\sim 0.3 \text{ km s}^{-1}$ at 200 K), implying strong turbulence inside the hot core. The central velocities are centered at zero with some deviations. For example, NH_2D has negative vlsr, but still shows good fit.

4.2.1 Vibrational excited lines

There are still some lines not fitted, e.g., $\text{C}_2\text{H}_5\text{CN}$ lines at 216 GHz and 216.5 GHz (shown by red lines in Figure 4.3), which corresponds to vibrational states $v_{13} + v_{21} = 1$ (Belloche et al., 2013) that were not included in the molecular database before Endres et al. (2021). These two modes of the same vibrational states are fairly close in energy and interacting strongly, hence not easy to measure in the laboratory of molecular physics. With the most recent experimental data, we derived a $\text{C}_2\text{H}_5\text{CN}$ column density of $9.5 \times 10^{17} \text{ cm}^{-2}$ with a temperature around 250 K, which are comparable to the values derived for other species, e.g., CH_3OCHO ($3.4 \times 10^{18} \text{ cm}^{-2}$, 270 K), HC_3N ($5.0 \times 10^{17} \text{ cm}^{-2}$, 235 K), and HNCO ($3.9 \times 10^{17} \text{ cm}^{-2}$, 290 K).

The fitted results are shown in Fig. 4.5. Furthermore, considering the bright detection of $v_{13} + v_{21} = 1$, we expect the same brightness of other higher-vibrational states, e.g., $v_{13} + v_{21} = 2$. Other unidentified lines may also belong to the vibrational states of complex molecules and their isotopes. Further laboratory measurements involving these excited states will be crucial to better understanding and characterizing the structure of astronomical sources.

Comparison between $\text{C}_2\text{H}_5\text{CN}$ ground and vibrational states can be used to obtain information on the temperature structure of the hot core. Seen in Table 4.1, when fitting only $\text{C}_2\text{H}_5\text{CN}$ $v=0$ lines, the excitation temperature is 109 K, while only fitting the $v=1$ lines gives a twice higher temperature of 223 K. This means that the $v=0$ and $v=1$ transitions are tracing different environment and gas components. The column density of $v=0$ (7.98×10^{17}) is twice that of $v=1$ (3.89×10^{17}). However, when we fit the two vibrational states as one component together with a free parameter of the isotope ratio, the resulting temperature is only 104 K with an isoratio of 0.41, which is 5 times lower than the ratio derived from fitting $v=0$ and $v=1$ separately. Then we use two components to fit the $v=0$ & 1 together. When the iso ratio is left open, the resulting temperatures of two components are lower than expected, and the fitted spectra are not satisfying. Thus, we fixed the isotope ratio to the derived value of 2 and reran the fitting of two components. This time we got a satisfying result with two core components of 53 K and 218 K separately, which provides a sketch of a two-temperature-layer core.

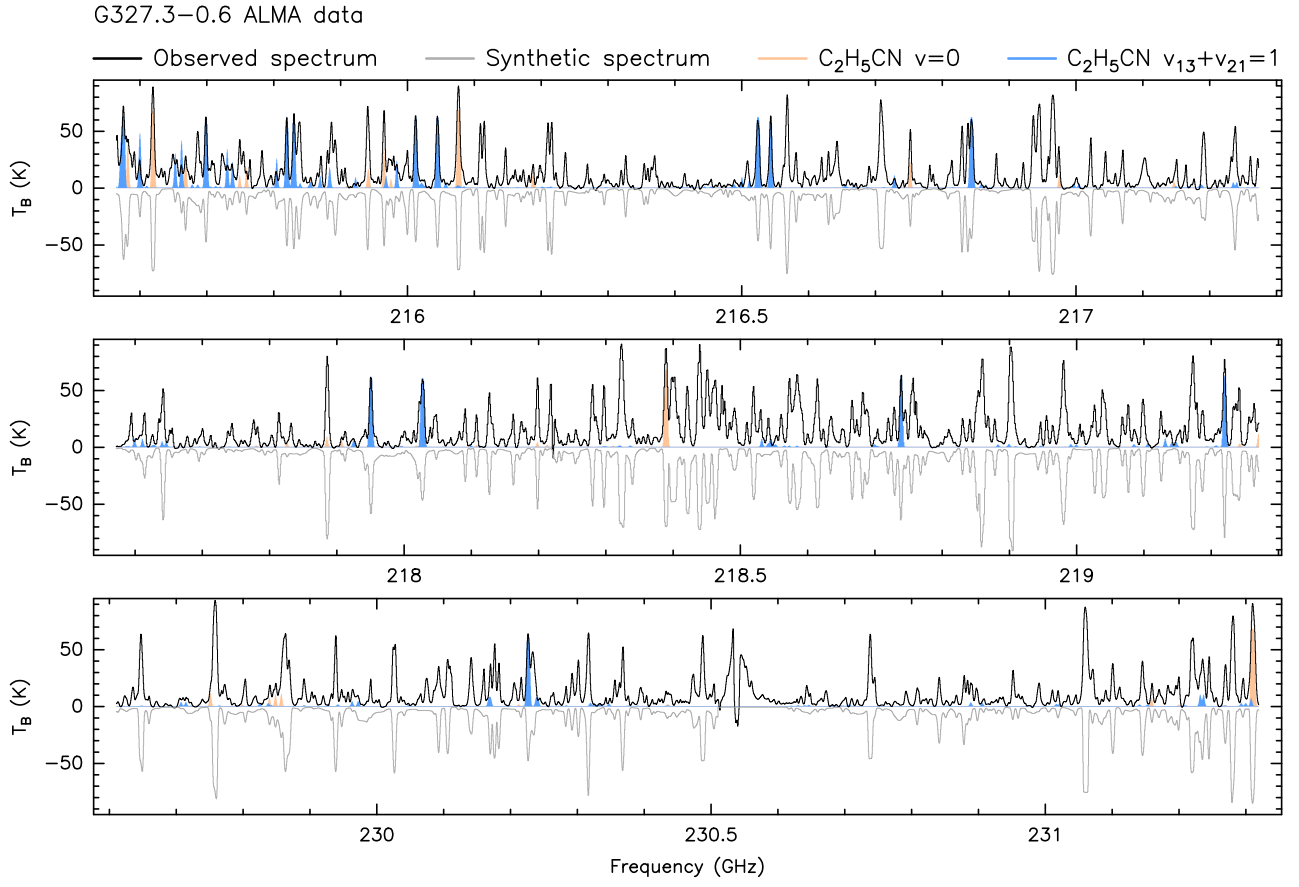


Figure 4.5: Hot core spectrum (black) with the synthetic spectrum including all identified molecular species (grey), which is multiplied by a factor of -1. The contribution from $\text{C}_2\text{H}_5\text{CN } \nu=0$ and $\nu_{13} + \nu_{21} = 1$ transitions are highlighted in orange and blue, respectively.

We also realize that the lines with high lower/upper energy levels, e.g., the vibrational excited states and the high-K transitions, help to confine the temperature, especially in the case of hot core. It will require laboratory progress and the high-sensitivity spectra showing the high-energy transitions that are possibly weak.

Table 4.1: The C₂H₅CN line fitting

Fitted states	components	T _{ex} K	N cm ⁻²	dv km/s	vlsr km/s	size "	isotope ratio
v=0	1	109	7.98×10 ¹⁷	3.85	0.74	0.13	-
v=1	1	223	3.89×10 ¹⁷	6.10	0.73	0.12	-
v=0&1	1	104	7.67×10 ¹⁷	4.05	0.75	0.10	0.41 (free)
v=0&1	2-1	35	9.76×10 ¹⁶	7.07	0.72	0.14	0.44 (free)
	2-2	110	6.51×10 ¹⁷	4.40	0.64	0.14	0.44 (free)
v=0&1	2-1	53	3.48×10 ¹⁷	6.32	0.76	0.14	2 (fixed)
	2-2	218	5.96×10 ¹⁷	4.82	0.75	0.13	2 (fixed)

4.2.2 Error estimation

When applying the Levenberg-Marquardt algorithm in a single spectral fitting, there is a problem of missing an error estimation of parameters. Therefore, the Markov Chain Monte Carlo (MCMC) method ([Don van Ravenzwaaij, 2018](#)) together with the myXCLASSFit function were applied to the line fitting process for the first time. The principle of MCMC in error estimation is adopting a number of steps for Markov-chain, and then each step randomly draws a set of samples of parameters from the likelihood function in a small ball around the priority preferred position. After the algorithm is finished, the probability distribution and the corresponding highest posterior density interval are calculated. The initial steps of sampling always deviates largely from the best results, which should not be included in the final statistics. Therefore, more steps are added as "burn-in" phases before real counting. The whole MCMC process is quite computationally expansive, requiring reducing free parameters. Since the LM fitting already gave good results of the line width and central velocity, these two parameters are fixed in our error estimation. Then we have three free parameters: source size, temperature, and column density.

The parameter evolution map shows how the parameter behaves through steps in the MCMC phase only. When we set the burn-in steps to 60 and the MCMC steps to 100, the upper left panel of Figure 4.6 shows that it is still converging in the first 80 steps. The lower left panel shows the corner plot, which describes the projected 2D histograms of two corresponding parameters. The shape of the corner plot is not regular due to the small sampling size. Then in the next try, the burn-in phase is set to 200 with MCMC steps of 1000. The results are exciting. The well-converged evolution map of source size is shown in the upper right panel of Figure 4.6. The corner plots in the lower right panel show 2D Gaussian shapes, as expected, while the plot between temperature and column density shows a different diagonal direction from the others. This means that when the temperature is higher, the column density is also higher, which is possible in physical condition. At higher temperatures, more levels are populated, and many lines will get weaker, which has to be compensated by higher compensated by higher column density. The best-fitting results are clearly shown by the histogram peak: source size of

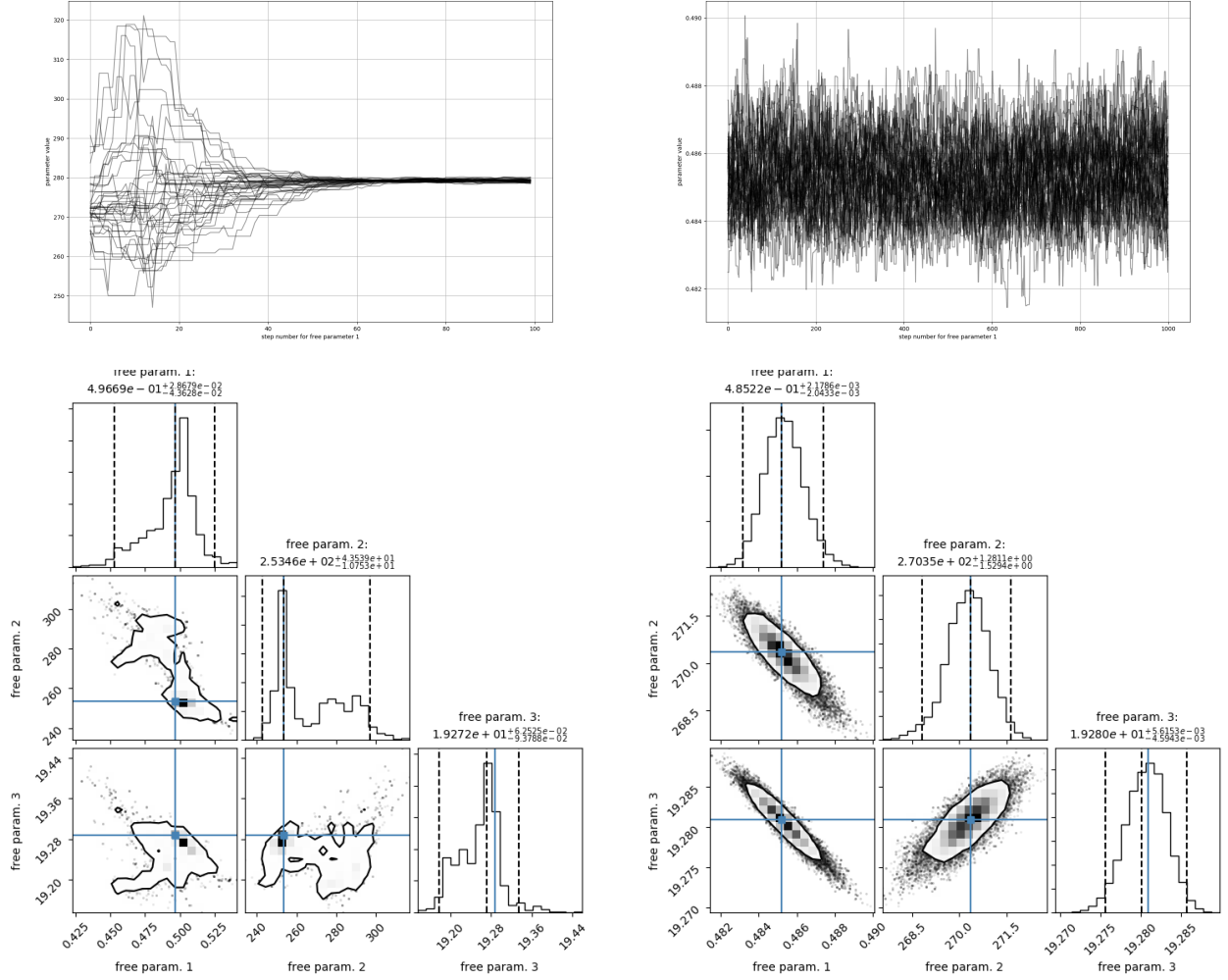


Figure 4.6: Error estimator results. Left side: 60 burn-in steps and 100 MCMC steps. Right side: 200 burn-in steps and 1000 MCMC steps. Upper panels: the parameter evolution map, in order to check converging or not. Lower panels: corner plots between free parameters. Free parameter 1, 2, 3 are source size, excitation temperature, and column density, respectively.

0.485", T_{ex} of 270 K and logarithmic column density of 19.280 cm^{-2} . The two sigma confidence intervals are marked by dashed lines in the histogram, which are two magnitudes smaller than the fitted value, convincing us of well-confined parameters within a narrow range. With the error estimation, we again prove that it is a quite evolved hot core and that the fitted parameters are reliable.

4.2.3 Source size

When checking the spectrum for the methanol lines, a problem appeared that it could not be well fitted when leaving the source size fixed and larger than the telescope resolution (0.7"). The best fit is seen from the blue lines in Figure 4.7, showing that the first fitted peak is higher than the real data, while the other three are lower. Simply adjusting the temperature would not give a better fit since temperature and intensity are in positive correlation. The source size may affect the fitting, although it is normal to assume an extend source that fills the beam size.

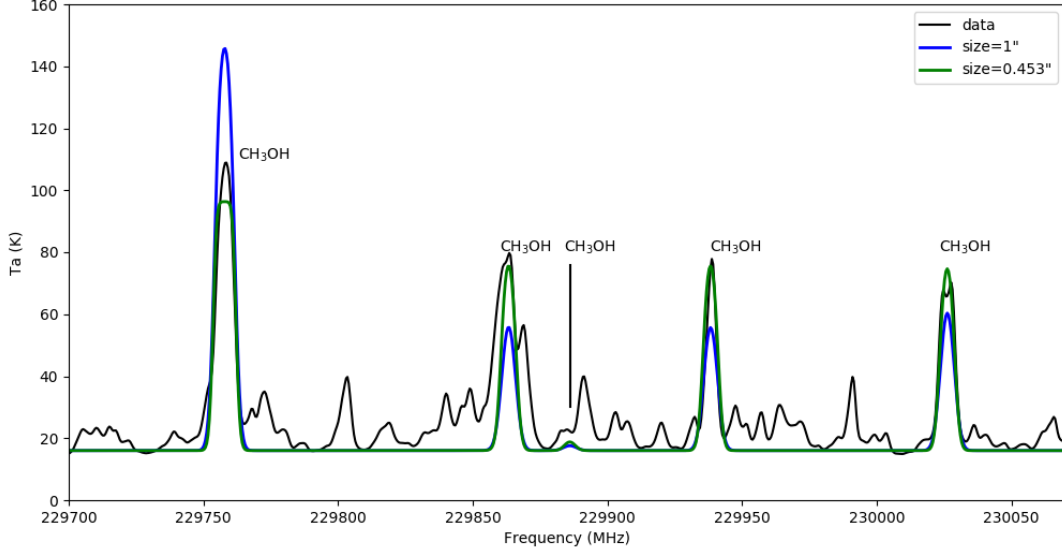


Figure 4.7: An example frequency of methanol line analysis. Black line shows the original data set; Blue line is the best fit with a fixed source size of 1'' (larger than telescope size 0.7''); Green line is fitted when leaving source size open, giving a result of 0.453''.

To understand the physics behind it, we need to refer to radiative transfer again. The observed intensity is proportional to $\eta T_{ex}(1 - e^{-\tau})$, in which η is the beam filling factor related to the source size, T_{ex} is the excitation temperature, τ is the optical depth. If a certain transition of methanol is optically thin, the intensity is simplified to $\eta T_{ex}\tau$; If it is optically thick, the intensity is only defined by ηT_{ex} . η equals 1 when the source (θ_s) already passes the beam size (θ_b), otherwise it equals $\theta_s^2/(\theta_b^2 + \theta_s^2)$. When T_{ex} is known, η can be derived from an optically thick line, hence τ from an optically thin line. Under the LTE assumption, T_{ex} is the same for all the transitions and η is identical for a certain molecule, while the optical depths are different. So when there are enough optically thin and optically thick lines, we can fit multiple transitions to get all the parameters. With only optically thin or optically thick lines, η and T_{ex} could not be decoupled, reminding us to check the optical depth after fitting.

We tried to leave the source size adjustable within a range of 0.01 to 10, and other parameters are also open. The range of other parameters are 3-1000 K for T_{ex} , $10^{12} - 10^{20} \text{ cm}^{-2}$ for column density, 1-10 km/s for line width, and -10 to 10 km/s for vlsr. The fitted source size is 0.453'', resulting in a beam filling factor of 5%, which produces a better methanol spectrum; see the green line of Figure 4.7. In total, we identify at least 22 significant lines of methanol, its vibrational state $v_{12}=1$, and its isotopologue $^{13}\text{CH}_3\text{OH}$, $\text{CH}_3^{18}\text{OH}$, whose optical depths range from 8.2×10^{-6} to 6.2. This fulfills the requirement of getting a reliable fitted source size.

4.2.4 Gradient with radius

The gradients of temperature, density, vlsr, and line width in hot core are not well studied so far, due to the lack of computational power and high-quality data. With the data and software we have, it is possible to derive the line parameters pixel by pixel. Firstly, we chose a 40*40-pixel region centered on the well-fitted peak pixel. The myXCLASSMapFit function provided by the XCLASS software was applied to the region with the same initial parameters as the peak

pixel. Because it is very computationally expensive to fit all the molecules in a large number of pixels, a limited number of molecules were tested.

Complex organic molecules are good tracers of the hot core and usually do not appear outside of the hot core region (Kobayashi, K., 2011). Therefore, we fitted $\text{C}_2\text{H}_5\text{CN}$ lines within the core region to obtain the gradient of the line parameters, as seen in Fig. 4.8. Interestingly, the fitted temperature results show a sharp gradient (from 150 K to 70 K within four pixels) at the edge of the hot core, indicating that there is a physical boundary condition to form $\text{C}_2\text{H}_5\text{CN}$ efficiently. There are also clear temperature gradients from the center of the core towards the edge, while it is asymmetric, especially in south-north direction. Some pixels show a temperature drop at the center region, which might stem from a possible space gap between the disk/envelop/outflow. For the column density distribution, we clearly see a dense core with a diameter of $2''$, which is exactly the size of a beam. The velocity offset shows a small but obvious gradient of 2 km/s from southeast to northwest ($7''$ angular distance), indicating a possible rotating envelope. The line width also has a gradient of 2 km/s from the center of the core to the edge through a radius of $4''$.

4.3 Temperature derivation of cold to warm cores

Determining temperatures throughout the filament in an early stage of evolution is crucial to understanding the initial conditions of star formation, e.g., thermal motion and Jeans stability. A high gas temperature will affect the Jeans masses of dense cores and may affect the initial mass function (IMF) of star formation, resulting in a top-heavy IMF.

However, the usual temperature tracers, such as CO, are easily affected by the interferometer effect, acting like self-absorption (van Dishoeck and Blake, 1998). Other linear molecules such as HCN and HCO^+ are usually optically thick, therefore, suffer from the coupling of temperature and density, which means the line strengths fit with a high density at a low temperature or a low density at a high temperature (Snow and McCall, 2006). The metastable inversion lines of NH_3 is extremely affected by high UV flux and forms a variety of layers at different temperatures (Ho and Townes, 1983). Symmetric top molecules such as CH_3CN are not widespread and are usually very faint (Ilee et al., 2021). Therefore, we need some other universal (symmetric or slightly asymmetric top) molecule to derive the kinetic temperature of the entire cloud. We introduce some examples in the following section.

In addition to the hot core, there are 65 other cores to explore. However, each spectrum has its own unique features, with wide range of molecular species and intensities, making it time consuming to analyze one by one in detail, although we already took the average spectra of cores. Therefore, in the present case, we only adopted two molecules that are the most representative and widely existing molecules, CH_3OH and H_2CO , for direct XCLASS fitting of the averaged core spectra. This helps us to know the basic physical conditions of extended emission and more evolved core region. We first discuss formaldehyde.

4.3.1 Formaldehyde, H_2CO

Formaldehyde (H_2CO) is proved to be ubiquitous and the variations of H_2CO abundance are usually within one order of magnitude (Johnstone et al., 2003). Formaldehyde line ratios involving different K_a ladders of ortho/para H_2CO are established as good tracers of the kinetic

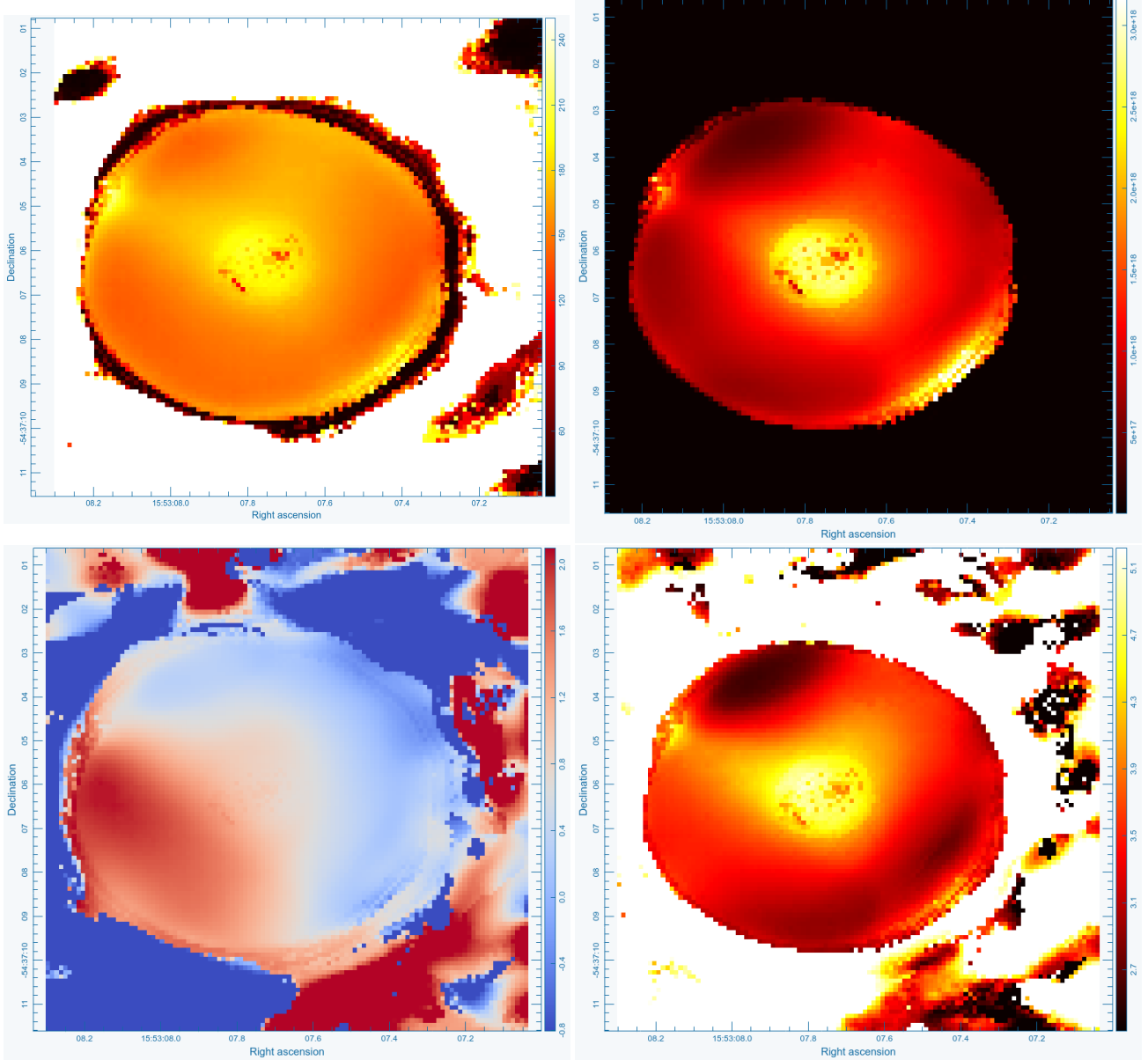


Figure 4.8: Upper left: The excitation temperature distribution of the hot core using $\text{C}_2\text{H}_5\text{CN}$ as a tracer, produced by XCLASS myXCLASSMapFit function. The temperature shows a gradient from the center of the core towards the edge, while the edge is very sharp. Upper right: The column density distribution, showing a dense core with $r=1''$, comparable with the beam size $2''$. Bottom left: The velocity offset shows a gradient of 2 km/s from southeast to northwest, indicating a potential rotating envelope. Bottom right: The line width shows a gradient of 2 km/s from the center of the core to the edge.

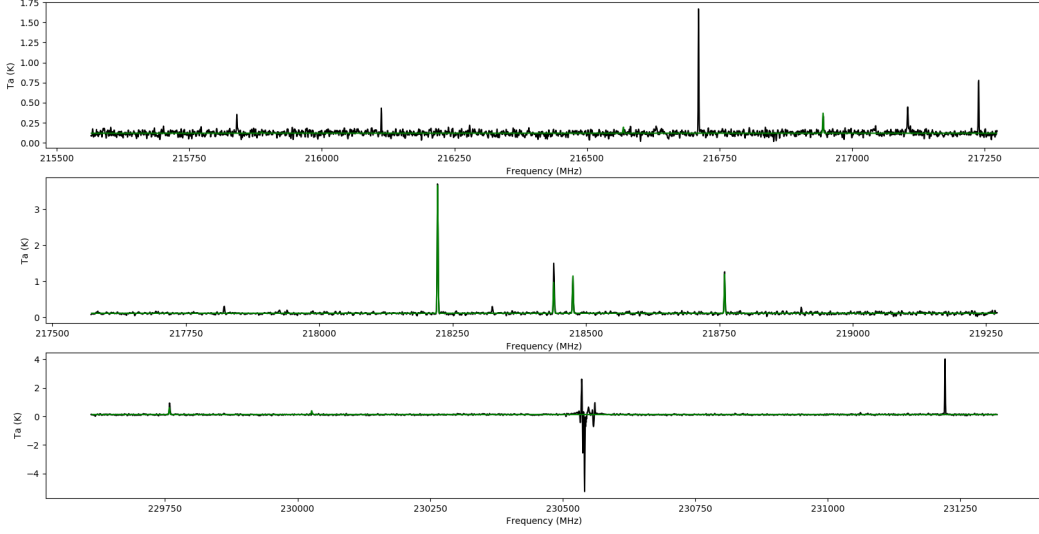


Figure 4.9: An example of the CH_3OH and H_2CO fitting results of averaged core spectra.

temperature (Mangum and Wootten, 1993; Mühle et al., 2007), particularly in gas below 100K (Mangum et al., 1993). The relative populations of the K_a ladders of H_2CO are shown in Fig. 4.10 as an energy level diagram. Since the lines cover a wide range of energy levels, from ten to several hundred kelvin, we can use a rotational diagram to derive the temperature with a relatively small uncertainty under the assumption of Local Thermal Equilibrium (LTE). In practice, it is found that H_2CO lines are dominated by non-LTE processes, i.e., collision, rather than LTE (Tang et al., 2017b). Analysis of non-LTE conditions will not only provide us with a more reliable kinetic temperature, but also the information on hydrogen density, which allows a determination of H_2CO abundance ratio.

To study the H_2CO properties, we first list all the H_2CO transitions included in our observations, as seen in Table 4.2. The line parameters are archived from the CDMS database.

In ALMA band 6, the brightest transition of H_2CO is the $J=3(0,3)-2(0,2)$ at 218.222 GHz, with an upper energy level temperature of 21.0 K and a lower energy level of 10.5 K. The remaining two transitions have very similar upper-level excitation temperatures at 68.1 K and a lower energy level of 57.6 K, and appear with very similar morphology and brightness. These two lines therefore provide near-degenerate measurements by constraining almost identical conditions. The line ratio of these two lines with energy states of 10.5K and 57.6K is sensitive to gas kinetic temperatures below 100 K in the case of optically thin H_2CO emission. At higher temperatures, these lines are less ideal since small changes in the ratios yield significant changes in the kinetic temperature. Thus, with only two independent data points, constraining the temperature condition faces a large uncertainty at the same time.

There is one more transition $J=9(1,8)-9(1,9)$ included in spw 0 of the band 6 observation, which has a high $E_{low}=164$ K. Due to its high excitation energy, many weak cores cannot excite such a transition; hence, there is usually no detection of this line, and we can only fit the three lines mentioned above. Despite that, it helps to determine an upper limit of the excitation temperature; that is the reason we always include this frequency in the fitting. For the hot cores, such a transition with higher J number becomes a better tracer of temperature and helps to narrow down the error in line fitting. It already surpasses previous studies such as Tang

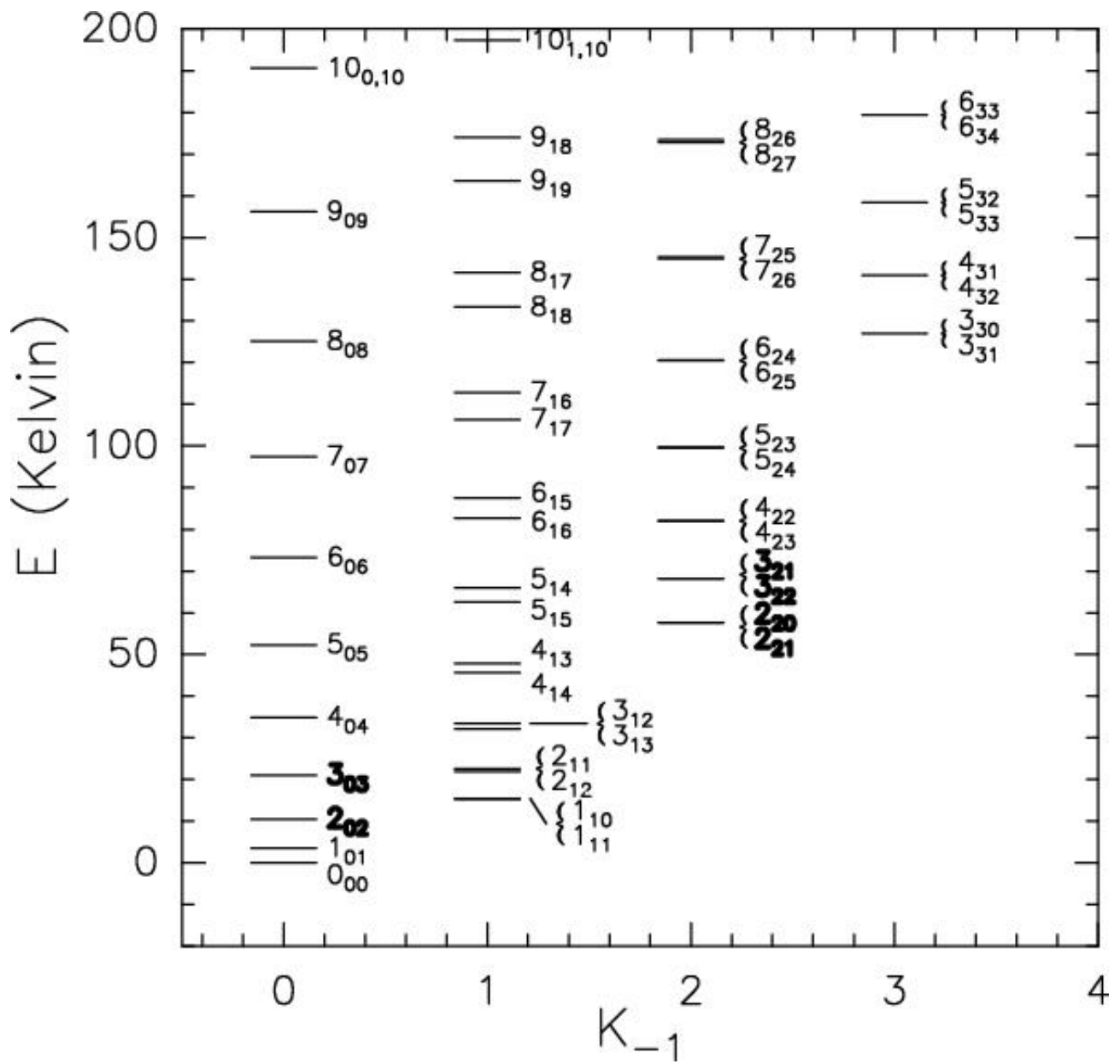


Figure 4.10: Para-H₂CO energy level diagram up to 200K. The three strongest transitions at band 6 218 GHz are shown in bold. (Mangum and Wootten, 1993)

et al. (2017b,a) when this line is added to our analysis.

It is necessary to mention that the 2 hydrogen atom in the H_2CO molecule can spin in parallel or in the opposite direction, which is known as the para/ortho state. These two states have different ground energy levels and consequently need to be considered separately, i.e., as different molecules in the fitting. In a molecular database such as CDMS, people usually calculate the ortho energy state relative to the para-ground state, rather than the ortho-ground state itself. For example, the ortho H_2CO at 216.568 GHz should have a $E_{\text{low}}=148.432$ K, while it is noted to be 163.595 K when searching for H_2CO in the CDMS database. The difference of 15 K do affect the fitting results. The practical way of solving this problem is to fit them as isotopes with a ratio, at the same time sharing the other physical parameters. In summary, we should always consider the para and ortho H_2CO separately in our fitting.

Recalling that we have ALMA ACA band 7 data, it is also useful to mention the covered H_2CO transitions. There are 6 H_2CO lines in the frequency windows of the band 7 with different energy levels ranging from 21 K to 127 K (see Table 4.2). In the 6 H_2CO lines, 3 ortho lines are in 2 energy states, and the other 3 para lines are also in 2 energy states.

In total, there are 10 H_2CO lines, in which 4 are ortho- H_2CO and 6 are para- H_2CO (marked on a real spectrum shown in Fig. 4.11). They belong to 7 energy levels ranging from 10K-164K, which provides a strong constraint for the rotational diagram to determine a temperature with high accuracy.

Table 4.2: The H_2CO line list in this paper

Rest frequency GHz	Resolved QNs	lower energy state K	ortho/para
216.56865100	9(1, 8)- 9(1, 9)	163.5952	ortho
218.22219200	3(0, 3)- 2(0, 2)	10.4834	para
218.47563200	3(2, 2)- 2(2, 1)	57.6086	para
218.76006600	3(2, 1)- 2(2, 0)	57.6120	para
281.52692900	4(1, 4)- 3(1, 3)	32.0589	ortho
290.62340500	4(0, 4)- 3(0, 3)	20.9564	para
291.23776640	4(2, 3)- 3(2, 2)	68.0938	para
291.38044230	4(3, 2)- 3(3, 1)	126.9544	ortho
291.38436150	4(3, 1)- 3(3, 0)	126.9545	ortho
291.94806740	4(2, 2)- 3(2, 1)	68.1109	para

4.3.2 LTE temperature

Usually we suppose that the stellar objects are in a stable state, and we start with deriving the temperature under the assumption of Local Thermal Equilibrium (LTE).

Core temperature We first fitted the average spectra of the cores to get the temperature. Since the cores are identified with the ALMA band 6 continuum at a high resolution of $2''$, we extracted the core spectra from the band 6 12m+7m data cubes. As discussed above, usually

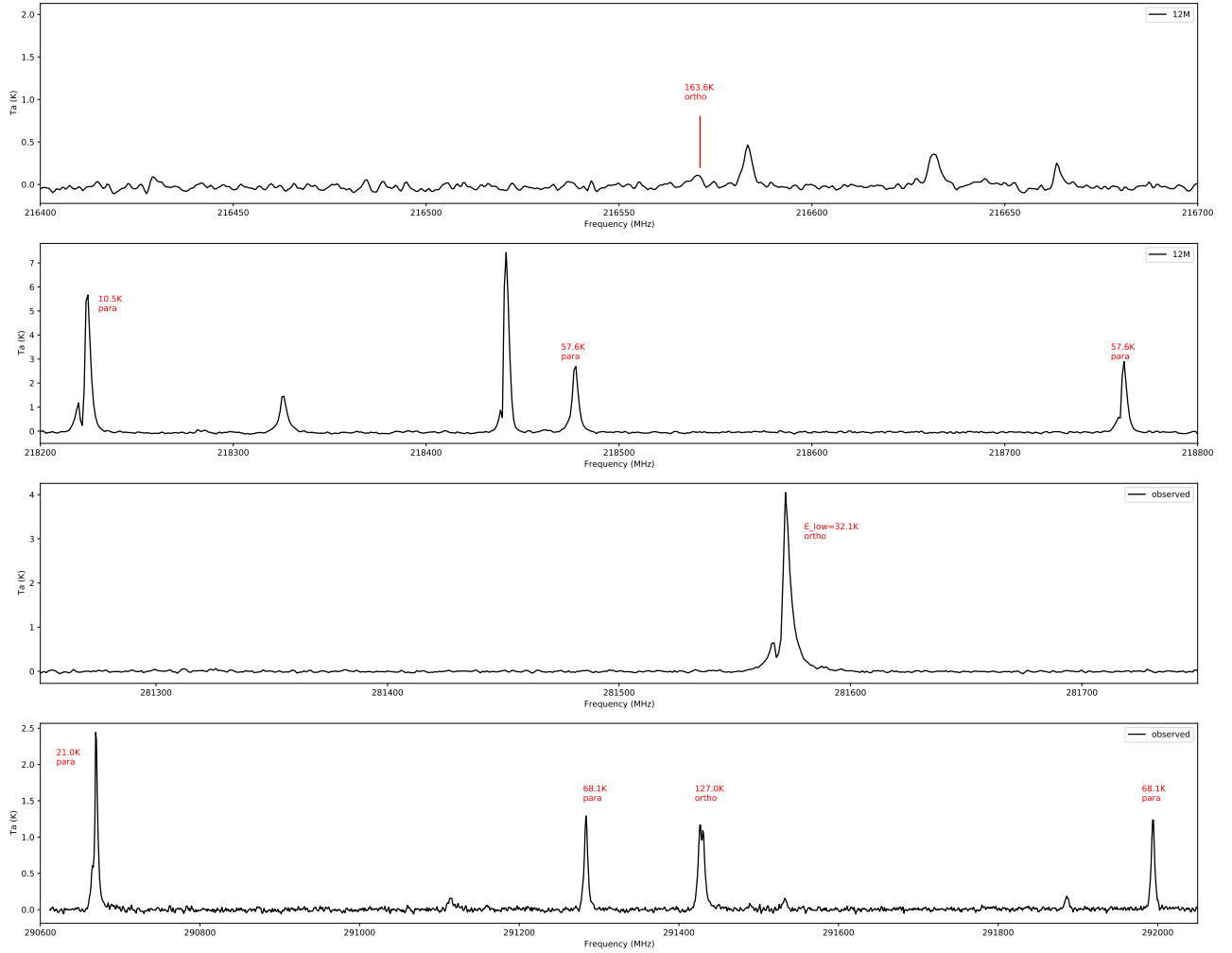


Figure 4.11: Example of the covered H₂CO lines.

only three para H₂CO lines appear at the cores, which requires well-confined initial values for LTE fitting. We set up the fitting pipeline as follows.

- Fit a single Gaussian towards the strongest H₂CO line at 218.222 GHz and obtain the line parameters that include intensity, width, and central velocity. The latter two are adopted as initial values for the XCLASS fitting.
- Set the initial temperature according to the peak intensity derived. For a peak intensity less than 0.2 K, we set the initial value as 30 K. For a peak above 3 K, the initial temperature is set to 120 K. For the sources in between, the initial guess follows an equation as $T_{ex}=32T_{peak}+24$.
- Set a large and fixed source size as 100'', a free column density as $1.0\times 10^{14}\text{cm}^{-2}$, and an isoratio between para/ortho as 1. Save the individual molfit file for each core.
- Set the fitting algorithm: Levenberg-Marquardt algorithm with 50 iterations and $10^{-4}\%$ variation in each step. The stopping limit is set to 10^{-6} for a calculated χ^2 .
- Run the myXCLASSFit function to get a fit and plot.

The fitting results are listed in Table 4.3. We go through all of the fitted spectra by eye and find that it is, in general, satisfactory. We get temperatures in range of 30–299 K with a median of 50 K, H₂CO column density of 2.96×10^{13} – $2.98\times 10^{16}\text{cm}^{-2}$ with a median of $2.13\times 10^{14}\text{cm}^{-2}$. The velocity dispersion ranges from 2.03 to 14.92 km/s, while the median is 4.38 km/s. The central velocity indicates the relative motion of the cores, which is from -4.09 to 2.51 km/s. The total velocity difference is only 6.6 km/s, inferring a relatively stable cloud without significant collisions or large-scale infalling features. The distribution of these parameters is shown in a violin plot in Fig. 4.12. The extreme value in the violin plot comes from the central hot core, since we fit only one Gaussian towards H₂CO while it is actually double peaked.

Compared with previous research, Tang et al. (2017b) fitted temperatures from 30–61 K (average 46 K) for 30 Galactic massive clumps at various stages of high-mass star formation, similar to our source. It is similar to the lower limit and the median value of our temperatures, while we have a higher upper limit due to the additional line $J = 9(1,8)$ – $9(1,9)$. For the hot central molecular zone (CMZ), Ao et al. (2013) found a temperature ranging from 50 K to above 100 K, with an average of 65 K, which indicates a hotter environment than our source. For the Large Magellanic cloud (LMC) with lower metallicity, Tang et al. (2017a) obtained the H₂CO kinetic temperatures as the 35–63 K (average 47 K), suggesting our source may also have low metallicity. The comparison of H₂CO column density is discussed in detail in terms of the abundance ratio.

Table 4.3: XCLASS automatic fitting results of 66 cores of H₂CO and CH₃OH

Number	CH ₃ OH				H ₂ CO			
	N cm ⁻²	Tex K	dv km/s	vlsr km/s	N cm ⁻²	Tex K	dv km/s	vlsr km/s
0	4.27e+13	30	2.03	-0.17	5.34e+13	78	1.01	-0.04
1	6.31e+13	34	2.42	0.39	3.22e+13	37	1.46	1.34
2	5.32e+13	32	2.21	0.10	5.34e+13	46	2.64	0.42
3	7.14e+13	30	2.62	0.06	1.04e+13	10	1.70	-0.89
4	1.02e+14	33	4.30	-1.10	8.30e+13	14	2.64	-0.37
5	2.01e+14	50	2.41	-0.16	4.70e+14	52	1.99	-0.30

Continued in the next page

Number	CH ₃ OH				H ₂ CO			
	N cm ⁻²	Tex K	dv km/s	vlsr km/s	N cm ⁻²	Tex K	dv km/s	vlsr km/s
6	3.37e+14	43	7.97	-4.09	8.38e+14	32	11.39	-4.39
7	3.55e+13	30	3.93	0.64	1.50e+14	184	1.03	-2.08
8	1.03e+14	40	2.75	-0.52	2.13e+14	35	1.00	-0.75
9	1.46e+14	50	2.35	-0.47	1.89e+14	46	2.53	-0.51
10	1.22e+15	100	5.12	0.06	4.79e+15	46	3.87	0.22
11	5.62e+14	76	4.55	2.05	1.59e+15	53	2.66	2.49
12	5.64e+13	34	2.20	0.39	7.31e+13	55	1.01	-0.12
13	2.18e+14	48	5.32	2.45	3.51e+14	28	3.22	3.60
14	7.95e+13	32	2.66	-0.03	1.36e+14	34	1.17	-0.04
15	6.28e+13	32	2.43	0.06	1.73e+14	82	2.74	0.21
16	3.60e+14	62	3.60	0.68	1.60e+15	62	4.70	0.01
17	1.17e+14	43	3.66	1.32	5.21e+15	5	3.60	0.03
18	8.68e+13	35	5.68	-0.19	4.89e+13	38	2.69	2.42
19	3.85e+14	58	4.76	2.51	6.99e+14	28	3.68	3.30
20	3.71e+14	63	4.70	1.65	1.80e+15	17	9.64	0.00
21	2.08e+14	43	3.59	-0.03	1.65e+14	15	1.26	-0.86
22	2.98e+16	299	14.92	-0.59	1.72e+17	37	4.80	-0.30
23	5.02e+14	63	5.53	-0.10	1.03e+15	39	2.88	-0.45
24	1.22e+15	88	5.10	-0.25	6.53e+15	33	7.56	1.53
25	1.84e+15	120	5.87	1.81	1.10e+15	223	1.76	-0.05
26	1.41e+15	76	4.96	-1.19	3.12e+15	183	2.29	-0.97
27	1.51e+15	71	4.71	-1.30	3.06e+15	138	3.04	-0.69
28	1.36e+15	59	5.21	-2.71	5.21e+15	202	4.46	-2.11
29	1.56e+14	42	4.38	-1.21	3.60e+09	10	2.87	-0.99
30	1.20e+15	109	4.37	-0.99	2.74e+15	48	2.37	-0.76
31	1.12e+15	106	6.38	1.27	2.02e+15	46	3.38	2.24
32	9.46e+14	96	4.74	-1.27	3.48e+15	36	4.28	-1.64
33	1.27e+15	112	4.60	-0.33	5.94e+15	39	3.22	-0.37
34	9.58e+14	98	4.92	-0.47	3.41e+15	37	2.69	-1.00
35	1.03e+15	106	4.45	-0.91	2.84e+15	52	2.55	-0.81
36	5.54e+14	76	4.32	-0.05	1.63e+15	43	2.08	-0.71
37	1.15e+15	82	6.03	-1.76	7.95e+15	60	3.72	-1.72
38	2.96e+13	30	3.57	-0.32	4.10e+09	28	1.50	-0.51
39	7.22e+14	62	10.67	2.21	3.34e+15	31	8.55	0.21
40	3.49e+14	47	6.32	1.01	5.22e+08	28	2.24	-1.49
41	1.69e+15	101	8.27	0.35	5.79e+15	48	7.79	-0.01
42	2.58e+14	55	4.97	-0.88	1.08e+15	35	3.84	-1.15
43	6.16e+14	85	7.71	-0.60	1.38e+15	62	4.31	-0.77
44	3.04e+14	55	5.01	-1.39	4.68e+14	29	3.05	-0.95
45	4.70e+14	66	5.31	-1.61	1.29e+15	95	5.23	-2.38
46	1.14e+14	46	4.40	0.28	4.82e+14	25	3.23	2.41
47	3.57e+14	62	4.17	-1.35	4.48e+14	38	2.62	-1.25
48	1.20e+15	107	4.97	-3.04	8.13e+15	67	4.12	-2.51
49	1.84e+14	51	4.91	-0.82	3.19e+14	34	1.61	-0.61
50	4.56e+13	30	2.82	-0.26	1.42e+14	137	2.63	1.31
51	4.64e+13	30	3.07	-0.36	4.73e+13	119	1.13	-0.48
52	3.25e+14	60	6.55	-1.01	1.01e+15	42	2.36	-0.81
53	7.72e+14	91	7.03	-2.22	4.72e+15	65	5.30	-0.31
54	9.58e+13	42	2.11	-0.21	6.86e+13	40	1.33	-0.30
55	9.86e+13	40	3.18	-0.47	2.95e+14	361	1.09	-1.46
56	1.32e+14	45	2.63	-0.07	2.03e+15	6	1.90	-0.53
57	1.09e+14	39	2.52	-0.48	1.04e+14	26	1.79	-0.68
58	8.26e+13	37	3.34	-0.39	5.20e+13	33	1.88	-0.17
59	1.01e+14	33	4.35	-1.07	9.87e+13	116	1.84	-1.60
60	1.00e+14	34	4.02	-0.87	7.98e+13	20	1.64	-0.69
61	7.30e+13	32	2.84	-0.67	2.66e+14	10	2.57	-0.70
62	4.73e+13	30	2.92	-0.35	5.23e+14	421	5.19	0.41

Continued in the next page

Number	CH ₃ OH				H ₂ CO			
	N cm ⁻²	Tex K	dv km/s	vlsr km/s	N cm ⁻²	Tex K	dv km/s	vlsr km/s
63	4.50e+13	30	2.58	-0.44	3.53e+13	29	1.11	-0.75
64	3.95e+13	30	2.99	-0.50	1.15e+14	49	3.25	-0.30
65	1.71e+14	53	2.68	-0.81	2.11e+14	39	2.01	-1.34

Filament temperature We fitted the whole G327 observation field with the myXCLASSMap-Fit function for pixels showing H₂CO lines above 5σ . The results are shown in Fig. 4.13. The H₂CO excitation temperature shows a gradient (10–150 K) throughout the filament, which peaks at the central hot core and decreases toward the edge of the filament. The majority of the filament is hot with a temperature above 100 K. The column density map is peaked at the significant cores and also presents a high value on the filament. The cold and diffuse gas shows a low column density of several 10^{13} cm⁻², which is 4 magnitudes lower than that of the hot gas. The line width of H₂CO is varied, ranging from 2 to 10 km/s. Interestingly, the cores are not always associated with the high line width; instead, the pixels around the cores tend to show strong velocity dispersion. This may be explained by the accompanying outflows, which efficiently broaden the line widths. Finally, the central velocity appears to be homogeneous on the filament, with some spikes at the noisy edge region. This may be not true due to the multiple components of H₂CO, while we fit only a single Gaussian towards each spectrum.

Abundance ratio By dividing the H₂CO column density and the N_{H_2} derived from the dust continuum, we obtained the H₂CO abundance ratio in the whole filament. We plot the distribution between abundance and excitation temperature for all pixels in Fig. 4.14. From the density plot, we see most of the pixels gather at an abundance of 10^{-10} – 10^{-9} , while the total range is 10^{-12} – 10^{-7} . There is another group of data located at the high-temperature high-abundance end, which corresponds to the hot cores and surrounding areas. In general, the abundance increases with temperature, indicating a higher H₂CO forming efficiency at the higher temperature. At the same time, we overplotted the 66 identified cores (orange dots) in the distribution, which shows a similar tendency as discussed: the hotter cores have higher abundance (except one outlier at 48 K but with high abundance of 2×10^{-7}).

Previous observations gave a very narrow fractional abundance of para-H₂CO, usually between 10^{-10} – 10^{-9} and within one magnitude variance (Johnstone et al., 2003; Ao et al., 2013; Tang et al., 2017b,a), which agrees well with our main result. We have a much larger diversity in the abundance ratios, extending in both directions of the low limit and the high limit, since our observation has a higher resolution (2'') than the previous ones (e.g. APEX 30'') to detect weak sources and resolve hot sources.

The error in temperature For the most case of the 12m band 6 spectra of cores, we only have three lines and two independent data points, thus leading to a large uncertainty in the fitting results. It would be good to include more lines, e.g., the ones at band 7, to improve the temperature fitting. However, we do not have a band 7 observation (5.7'') at similar resolution as the band 6 12m data (2''). Only the ALMA ACA data for band 7 are available, which is not enough to recover the core-to-subcore scale temperature. In this case, we select a region of 6'', extract the average spectra of the ACA data of both bands 6 and 7, and fit the temperature. We intend to check the difference of fitting only band 6 lines, only band 7 lines, and all together.

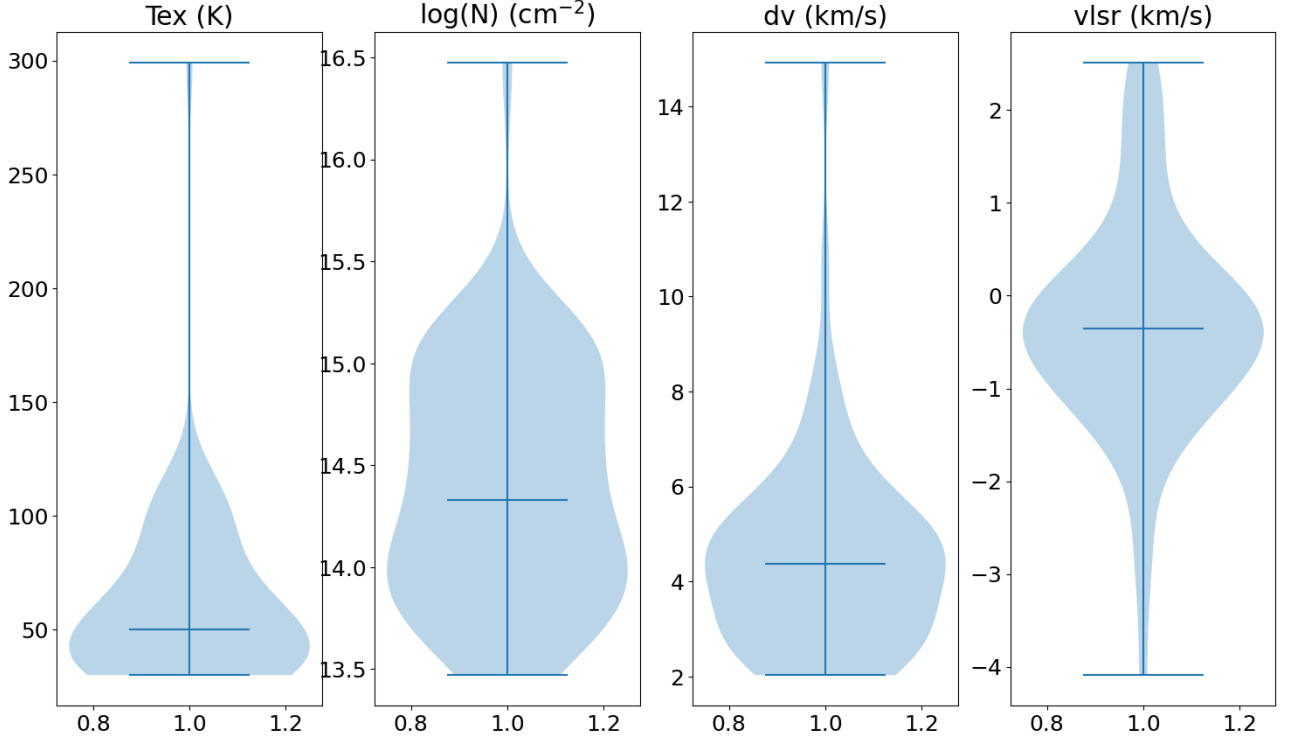


Figure 4.12: The violin distributions of the four derived line parameters of H_2CO . The median values of each parameter are marked by a horizontal line.

Here are the MCMC fitting results for the ACA band 6 lines, band 7 lines, and all together (see Fig. 4.15). We have set the same starting value (from the best fit of the LM algorithm for band 6+7 data) and MCMC parameters for all the fitting; see the table 4.4. Since we fit the ortho and para molecules separately and the ratios between are important, we set the isoratio = ortho/para in the molfit file. The band 6 only data do not give a good constraint of the temperature, neither for the line width nor the isoratio, which is predictable since there are basically only two energy states in the fitting. The band 7 only data gives a better limit of the parameters, while the error is still large. The band 6+7 results are the best, since it gives the minimum error ranges for column density, line width, vlsr, and isoratio. For the temperature, it peaks at around 166 K, similar as the band 6/7 only result, which shows the importance of the 164 K line covered by band 6 to limit the upper boundary of the temperature. However, all of the temperature density plots in the MCMC results are elongated banana-shaped. This is a sign of the degeneracy between the temperature and the column density, which is the case in the optically thin regime. In other words, if all the lines are optically thin, it is not possible to derive a reliable temperature or density.

Table 4.4: H_2CO input and output fitting parameters with errors

type	size	temperature	column density	line width	vlsr	iso-ratio
	"	K	10^{14}cm^{-2}	km/s	km/s	
input	100	112	0.73	4.55	-3.33	1.0
band 6 output	100	169^{+131}_{-131}	$1.49^{+2.81}_{-1.47}$	$8.45^{+1.55}_{-0.62}$	$-2.63^{+8.03}_{-7.34}$	$6.73^{+3.26}_{-6.06}$
band 7 output	100	156^{+143}_{-69}	$1.02^{+1.96}_{-0.53}$	$5.14^{+3.99}_{-1.68}$	$-3.39^{+1.29}_{-1.11}$	$0.53^{+2.01}_{-0.31}$
band 6+7 output	100	166^{+133}_{-76}	$1.24^{+2.00}_{-0.69}$	$5.39^{+3.20}_{-1.95}$	$-3.32^{+0.81}_{-0.83}$	$0.62^{+0.33}_{-0.27}$

In conclusion, the lack of lines and energy states will not determine a good temperature. We require supplement observations of the G327 filament, either higher resolution 12m band

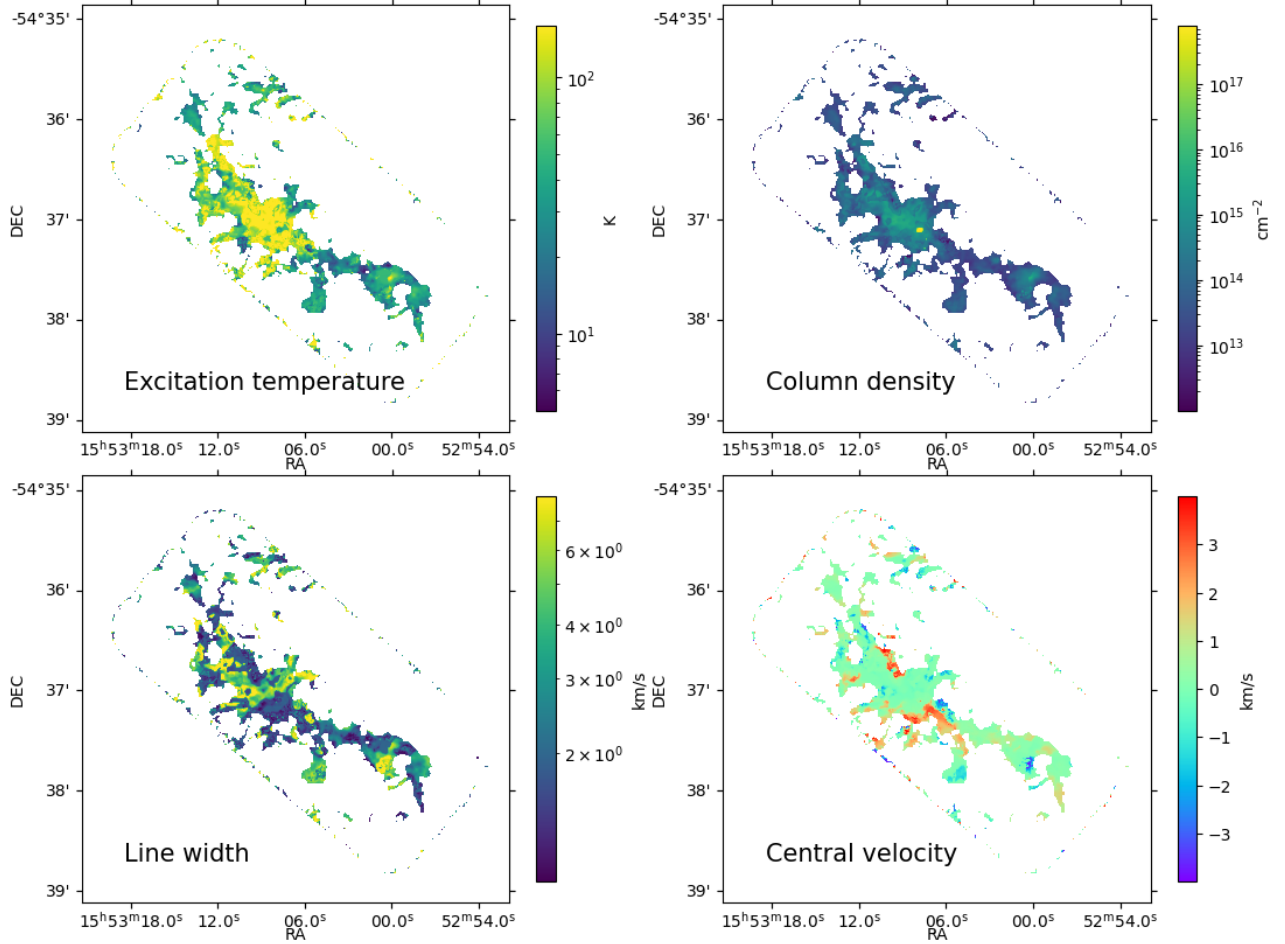


Figure 4.13: The H₂CO parameters fitted by XCLASS myXCLASSMapFit function with a mask of 5σ . Upper left: The excitation temperature distribution of the whole filament using H₂CO as a tracer. Upper right: The column density distribution peaks at the cores. Bottom left: The line width shows a gradient from the center of the core (10 km/s) to the edge (2 km/s). Bottom right: The velocity offset may be not reliable due to single Gaussian fit at the case of double peaks.

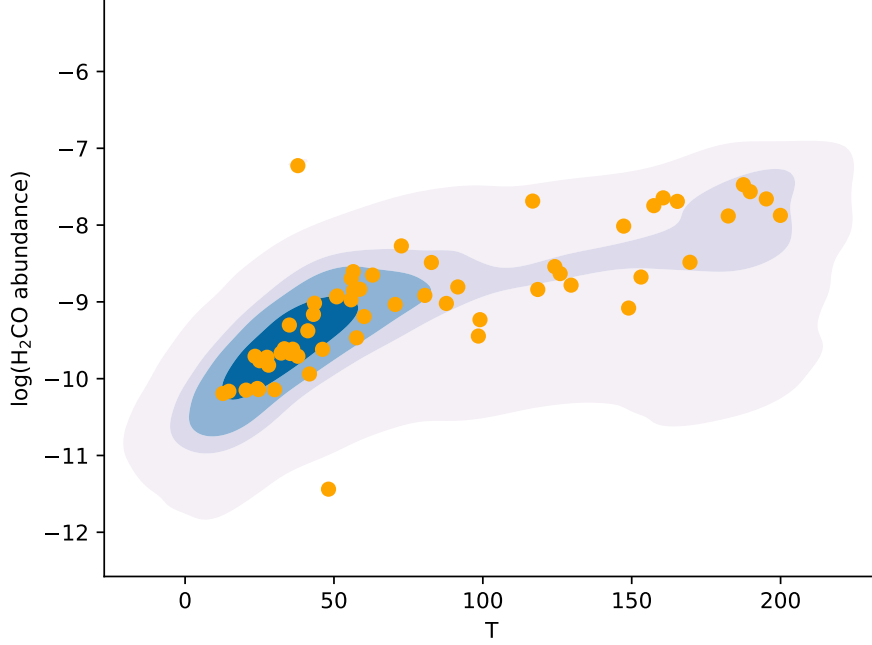


Figure 4.14: The H_2CO abundance ratio (in logarithmic scale) versus excitation temperature in density contours, showing a positive correlation. The 66 identified cores are plotted as orange dots.

7 data, or other frequency bands with more H_2CO lines, to better study the temperature of cores, which is a crucial step in understanding the evolution of star formation.

4.3.3 Non-LTE modeling with RADEX

[Ao et al. \(2013\)](#) has previously studied the same 218 GHz H_2CO triplet, though observed by APEX rather than ALMA, and the object is the clump scale of the Galactic center, rather than the core scale of a closer star-forming region. Their spectral and spatial resolutions are 2 km/s and $30''$, separately. The authors used LVG models to constrain the kinetic temperatures through all Galactic center clouds (in average 65 K), while the central part is the hottest ($>100\text{K}$). They claimed that the high temperature part may be caused by turbulent heating and cosmic-ray heating rather than photon heating, which needs further study.

[Tang et al. \(2017b\)](#) also observed the same lines with the APEX telescope, at a spectral resolution of 0.33km/s, three times higher than ours. They compared the gas temperature obtained from CO with the APEX and the dust temperature from the Hi-GAL data and the new kinetic temperature from para- H_2CO lines, and the result shows a good agreement of all of them. The authors believe that this suggests that the kinetic temperatures traced by para- H_2CO are correlated with the ongoing massive star formation. Considering NH_3 observed with Effelsberg telescope, its temperature is only half of H_2CO and the abundance is very low, indicating NH_3 is irradiated by UV photons while H_2CO is less affected by photo-dissociation of the young massive stars and hence more widespread.

They created model grids using RADEX codes to generate the line brightness for temperatures between 10 and 110 K, $n(\text{H}_2)$ between 10^3 and 10^8 cm^{-3} , and $N(\text{p-}\text{H}_2\text{CO})$ from 10^{12} to 10^{16} cm^{-2} . The line width is fixed to the observed value (5 km/s) and the background radiation temperature is set to the CMB value of 2.73 K. The initial fit is to determine the beam averaged

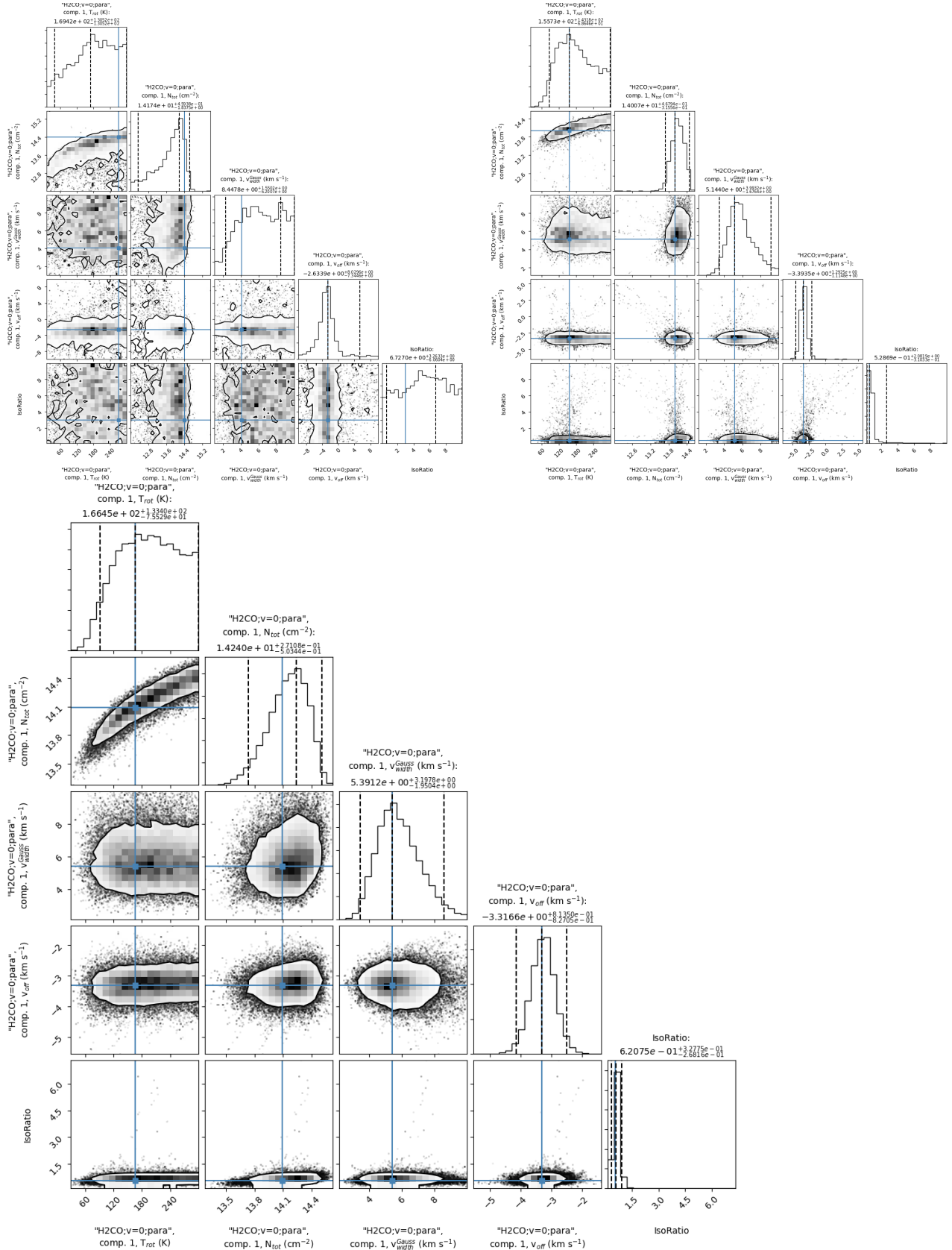


Figure 4.15: **Left:** MCMC results when fitting H_2CO observed by ACA at band 6. **Right:** MCMC when fitting H_2CO of band 7 ACA data. This shows a better limit of the physical parameters, including temperature, column density, and para/ortho ratio. **Bottom:** MCMC results when fitting H_2CO observed at band 6 and 7 ACA together, which narrow down the error of parameters while still do not give non-degenerated physical parameters.

H₂CO column density using the brightness temperatures of the two energy states. It is a minimization of χ^2 with one degree of freedom, with a fixed temperature of 40 K. At low volume densities below 10^6 cm^{-3} , the line brightness depends on H₂ density, H₂CO column density and the temperature. At higher densities, χ^2 only slowly reduces with changes in $n(\text{H}_2)$. They found a lower limit of $N(\text{H}_2\text{CO})$ at 10^{14} cm^{-2} and a preference for densities at 10^5 cm^{-3} , which is consistent with ATLASGAL clump values. Then the authors fix the molecular hydrogen density and leave the temperature and H₂CO column density open. In the second following paper, Tang et al. (2017a) fix the H₂CO column density to fit the temperature, which works on the clump scale.

Although it works with the RADEX model of H₂CO fitting at the clump scale, there is still no common sense at the interferometer scale. In this case, we tested RADEX on the high resolution band 6 data (observed by the 12m array of ALMA). Three parameters are set to variables, including temperatures between 70 and 190 K, n_{H_2} between 10^5 and 10^7 cm^{-3} , and $N(\text{p-H}_2\text{CO})$ from 10^{14} to 10^{16} cm^{-2} , each with 51 grids. The brightness temperatures of the four lines in band 6 were derived, using $\chi^2 = \Sigma(\text{model-observation})^2$ to evaluate the deviation from the model to the real data. Here, χ^2 are calculated in two ways, one by the brightness temperature and the other by the integrated intensity, since the two lines with the same energy state show a slightly different line strength. I plotted the resulting grided χ^2 onto a 3D map in the format of a vtk file, which can be easily read and shown by a software PARAVIEW⁹. By rotating the 3D grids and setting a reasonable color contour in the software, it does not present a clear minimum χ^2 throughout the 3D space, while the minimum behaves as an elongated banana in both cases of brightness and line integration. When plotted in 2D, as shown in Fig. 4.16 left panel, the minimum position marked by a white cross shifts dramatically when the density of the H₂ numbers is varied. It is similar when plotting the line intensities in 2D, as shown in the right panel of Fig. 4.16. Different temperatures produce different line intensity integrals, and our observation data is marked by the green dashed lines, indicating a high temperature of the core, while it is difficult to confirm the temperature due to the sparse part of the plot. Hence we draw the conclusion that the H₂CO in band 6 cannot give a best fit of the temperature, given the reason that the temperature and column density are degenerate due to the lack of tracers of different energy states.

4.3.4 Two components fitting

The single dish can detect the missing flux of the interferometer, hence we archived the APEX data at the same frequency ranges to explore the real emission from our cores. So we compared the spectra observed by single dish, ALMA ACA, and 12m-array, see an example in Fig. 4.17. For the H₂CO line at 218.222 GHz, the ACA spectra show double peaks, while the single dish spectra only show one peak, which corresponds to the right peak of the ACA. The 12m spectra shows one peak but a broad blue wing, in which the peak agrees with the left peak of the ACA data and the blue wing seems like another weaker component. In this scenario, the most possible case is the so-called "red profile", where the red-shifted peak of a double-peaked line is stronger for optically thick lines Liu et al. (2011). This is usually considered as a hint of outflow. Since we did not observe two components by the single dish (resolution 30"), and the 7m/12m arrays are more sensitive to the small scale structures ($\leq 7''$), we may detect a very compact outflow. This outflow is also confirmed by the SiO emission; see the bottom right panel of Fig. 4.17. And the velocity separation of the two peaks of the SiO line agrees well with the H₂CO lines (4 MHz), though the channel width of the SiO line is twice of the H₂CO

⁹<https://www.paraview.org/>

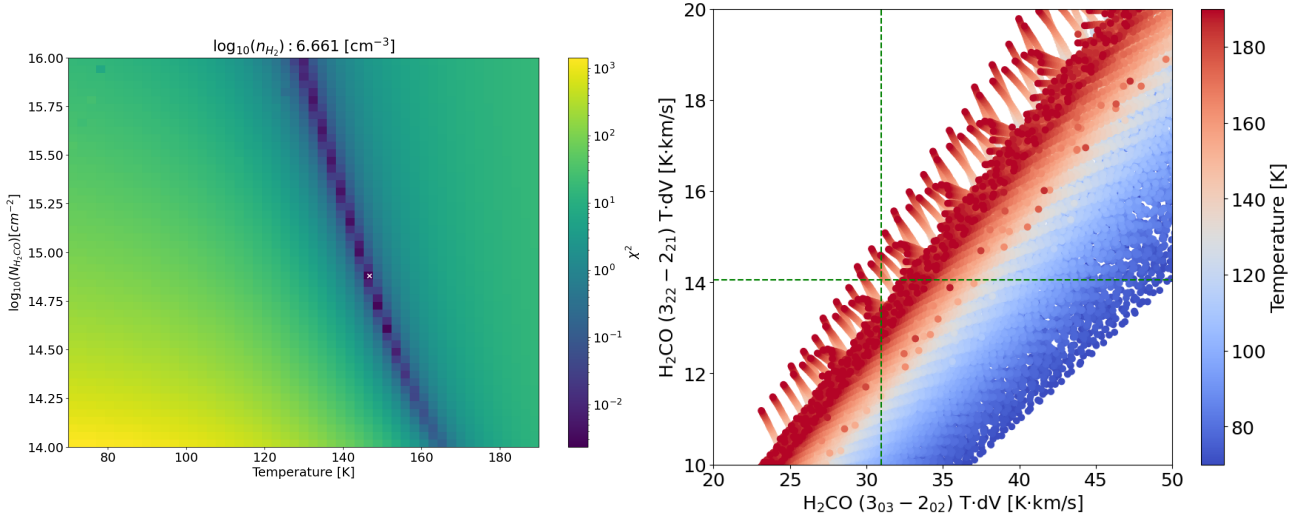


Figure 4.16: **Left:** The 2D χ^2 image at a fixed nH_2 of $10^{6.661} \text{ cm}^{-3}$. Two variables are temperature and H_2CO column density. The minimum χ^2 position is marked by a white cross, which changes at different nH_2 value. **Right:** The line intensity integral comparison of two H_2CO lines under different parameters, with real observation data marked by the green dashed lines, indicating a high temperature of the core. It is difficult to confirm the temperature due to the sparsity around the desired values.

lines. If we consider that the right peak is the core component and the left peak is the outflow, the outflow speed relative to the core is around 5.50 km/s.

The cores with outflows show double-peak H_2CO spectra. Fitting with only one component leads to a large discrepancy between the real data and the fitted result (see the upper panel of Fig. 4.18), which is basically due to the outflow peak in addition to the main peak of the core component. Therefore, naturally, we tried to fit the spectra with two components with different velocity offsets. The source sizes of both core and outflow are fixed to a large number, i.e. $100''$, which means the beam filling factor is close to unity. Leaving the other eight parameters free, we derived a disappointing fitting result, deviating from the observation, despite the double peaks being shown.

We realized that eight free parameters might be too much for only three formaldehyde lines. Therefore, we fixed the velocity offset of the two components to 4.5 km/s for the core and 0.5 km/s for the outflow component (see Table 4.5). With six free parameters, we derived a better fitting result (see the bottom panel of Fig. 4.18) with a lower fitted χ^2 than a one-component fitting. The core component is fitted with a temperature twice higher and also the column density twice higher than the one-component fitting, indicating a possible hot core ($>100 \text{ K}$). For the outflow component, the excitation temperature is around 34 K, lower than the core since the outflow is more extended. A low column density was found for the outflow. We also persuade ourselves that there exists a real outflow at the core by looking into the moment 1 map of H_2CO (see Sect. 5.3 for more details) and the 3D distribution of SiO (Kandpal Master thesis, 2019).

4.3.5 Other molecular tracers

Since H_2CO has its disadvantage in determining the temperature, we require other tracers to reliably constrain the conditions of the cores. Methanol is another abundant molecule,

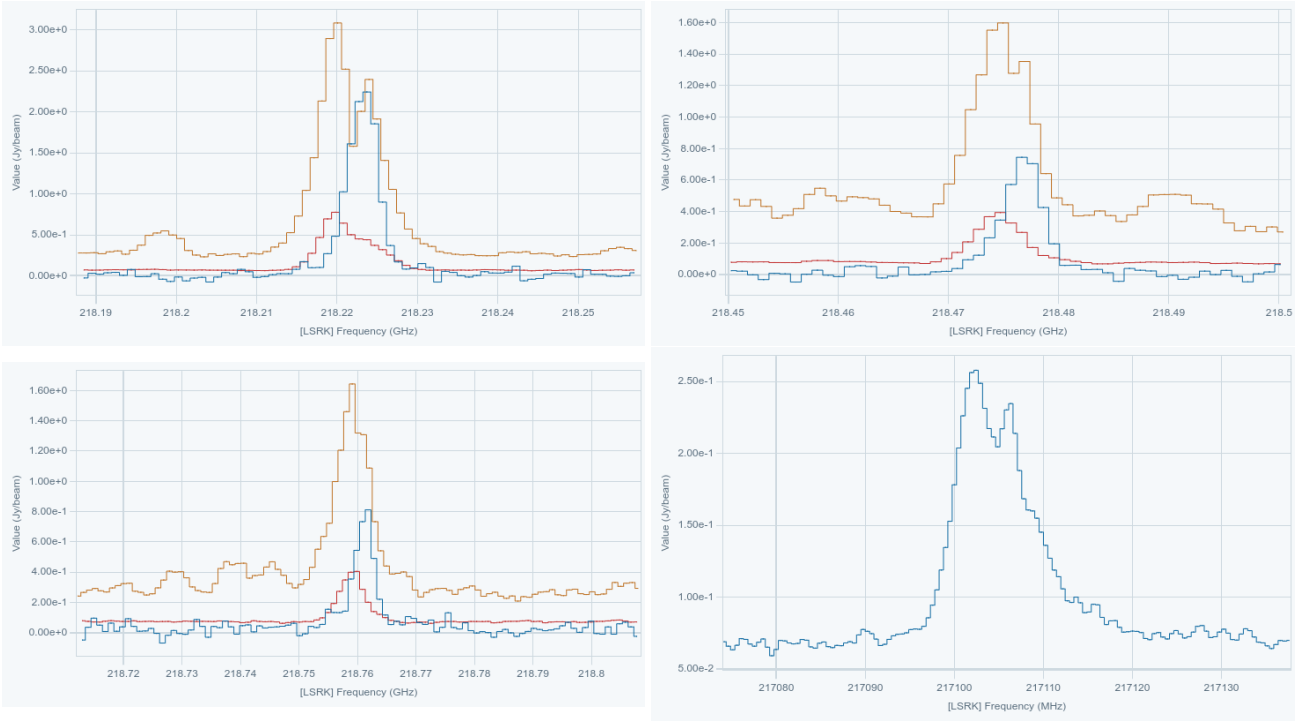


Figure 4.17: The average H_2CO spectra of region 42, which represents 218.222, 218.476, 218.760 GHz transition, see the axis label. Blue: APEX single dish data; Orange: ALMA ACA data; Red: ALMA 12m-array data. **Bottom right:** The average SiO spectra of region 42, whose two peaks are apart for 4 MHz, similar as the separation of the two peaks in ALMA ACA data.

Table 4.5: H_2CO two-component fitting results

type	size "	temperature K	column density cm^{-2}	line width km/s	vlsr km/s	fit vlsr or not
1 comp	1.0e+02	55.6	8.64e+14	9.31	4.31	yes
2 comp - core	1.0e+02	21.2	1.33e+15	6.09	3.88	yes
2 comp - outflow	1.0e+02	13.1	1.72e+15	2.10	0.00	yes
2 comp - core	1.0e+02	118.6	1.62e+15	8.64	4.50	no
2 comp - outflow	1.0e+02	33.9	3.05e+13	1.14	-0.5	no

formed by additional hydrogenation relative to formaldehyde on the surface of dust grains (van Dishoeck and Blake, 1998). It is widely distributed in high-density environments, including many evolutionary stages from infrared dark cloud to hot protostellar cores (Menten et al., 1988). As an asymmetric rotor, it is well established as a tracer of a warm molecular gas and a temperature higher than H_2CO . Other molecules are either too diffuse (OCS) or too condensed (COMs), not to mention most of them only show one line in the whole spectra (HC_3N , DCN , H_2S), and some of them are highly affected by outflow (SiO) and therefore are not ideal to trace our cores.

In our frequency ranges in band 6, we have 14 methanol lines that have $E_{\text{low}} < 1000$ K, in which 7 of them cover the range of 29–166 K. We used the average spectra of each core to fit these methanol lines. The fitting results are listed in Table 4.3. We get temperatures in range of 5–421 K with a median of 39 K, which span over larger ranges than the H_2CO temperatures. We plot the relationship between two temperatures in Fig. 4.19, in which the red diagonal line represents the same temperature for both. Most cores tend to have higher excitation temperatures in H_2CO than CH_3OH , while the distinct dots in the upper left corner are not

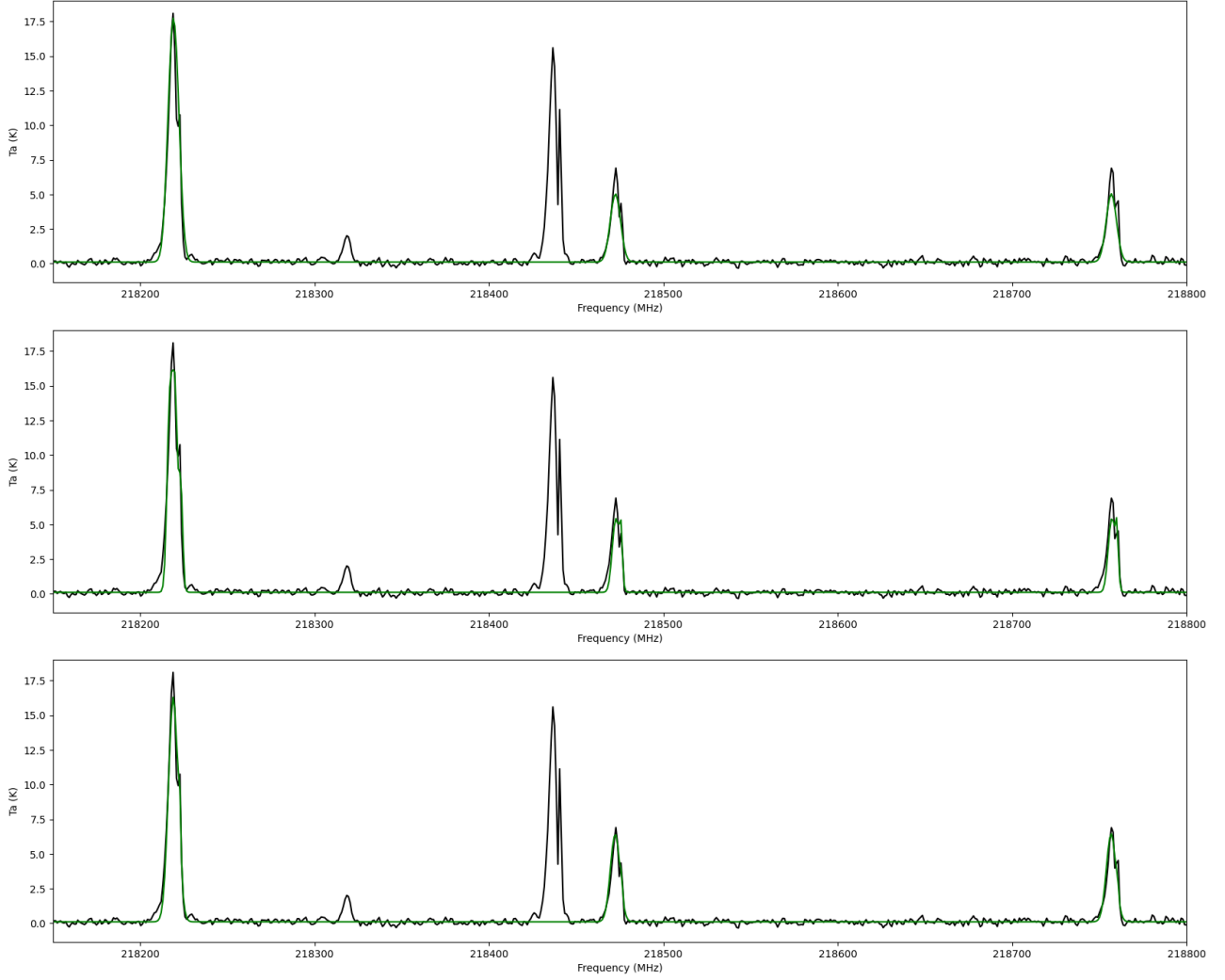


Figure 4.18: The H_2CO fitting results of average spectra of region 45. **Upper panel:** Fitting with only one component, showing a large discrepancy between the real data and the fitted result. **Central panel:** Fitting with 2 components and 8 free parameters, in which the double peaks were fitted separately. **Bottom panel:** Fitting with 2 components and 6 free parameters (fixing v_{lsr}), indicating a hot core component and a cooler outflow shell.

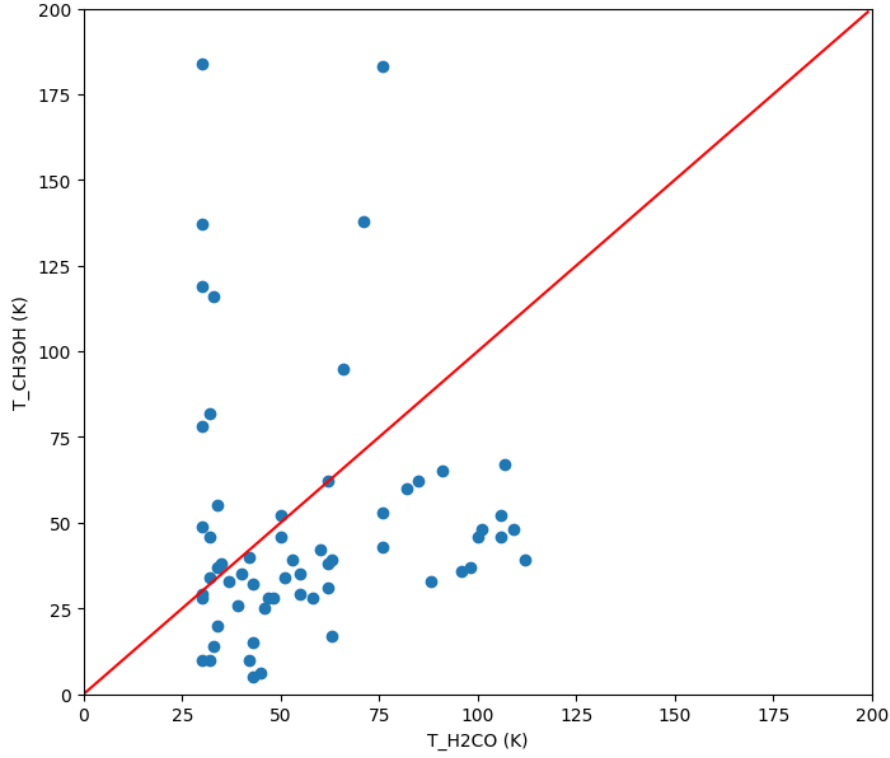


Figure 4.19: The excitation temperature comparison between CH_3OH and H_2CO . The red diagonal line represents the same temperature for both molecules.

reliable, as they show CH_3OH temperatures four times higher than the H_2CO temperatures. In addition, CH_3OH column densities vary from 5.22×10^8 to $1.72 \times 10^{17} \text{cm}^{-2}$ with a median value of $4.76 \times 10^{14} \text{cm}^{-2}$. The low densities originate from the weakest cores, which naturally have little methanol. The velocity dispersion ranges from 1.0 to 11.39 km/s, which is smaller than the formaldehyde value. The median line width of 2.64 km/s is again 40% less than that of H_2CO (4.38 km/s), indicating H_2CO is more affected by non-thermal motion assuming that the kinetic temperature is similar for the two molecules. The central velocity indicates the relative motion of the cores, which is from -4.39 to 3.60 km/s, centered on -0.51 km/s. The distribution of these parameters is shown in a violin plot in Fig. 4.20. The extreme value in the violin plot comes from either the hot core or the weak cores, since some cores do not show enough methanol lines to obtain a good fitting.

4.4 Core classification

4.4.1 Principle component analysis (PCA) basics

Traditionally, explicit programs are necessary to perform a specific task. However, when the data set becomes larger and larger, the characteristics and relationships of the variables remain unknown, leading to problems in detailed programming. An idea arose as teaching computers to implement human learning behaviors to acquire new skills and reorganize existing knowledge structures. Thus, machine learning (ML) is born as a rapidly developing core subject in the field of artificial intelligence (AI). ML studies algorithms and statistical models that computer systems can use to perform tasks without specific instructions, only given a set of sample data to train the machine itself.

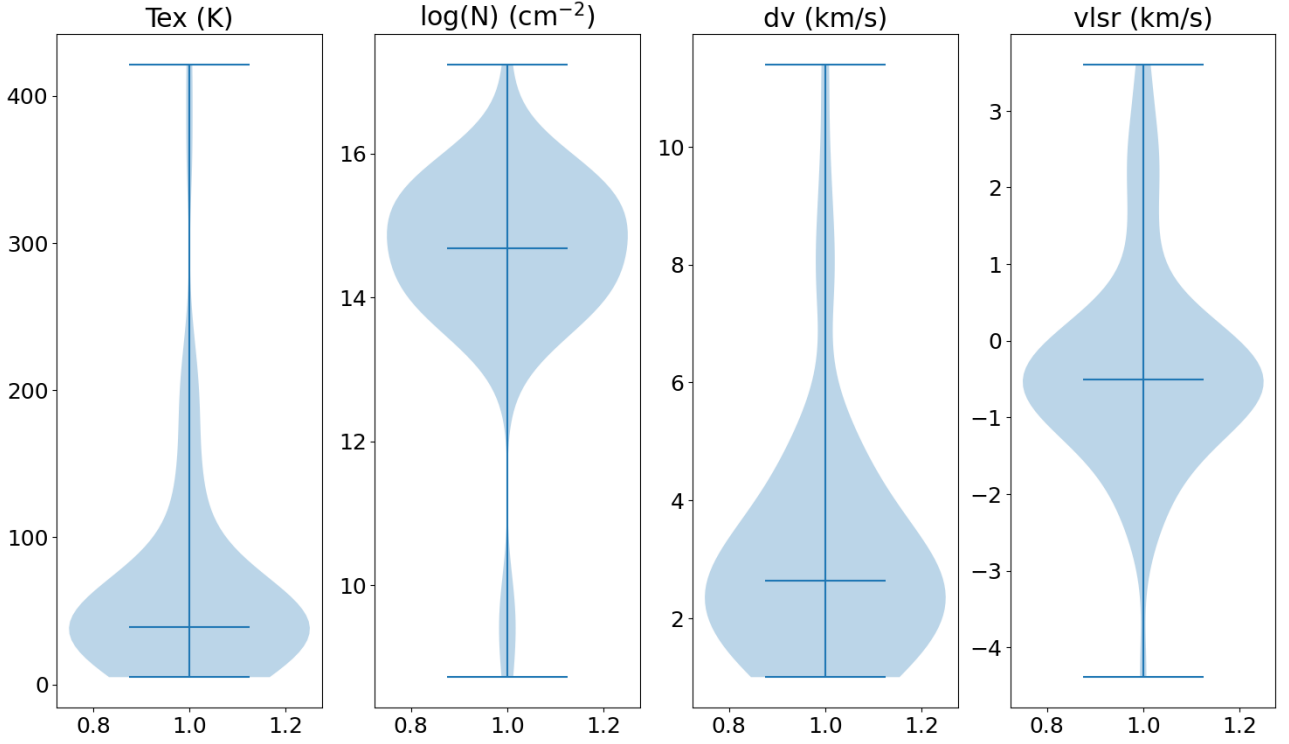


Figure 4.20: The violin distributions of the four derived line parameters of CH₃OH. The median values of each parameter are marked by a horizontal line.

The principal component analysis (PCA) has significant applications in machine learning. It is a statistical method that helps to find the correlation between original variables and reduce the dimension to focus on the important ones. The mathematical concept behind is to use orthogonal transformation to convert observed variables to a set of linearly uncorrelated new variables, which are called principal components (PCs, see Figure 4.21 for a 2D example). Thus, the first principle component includes the largest possible variance, and the succeeding principal component shows the next highest variance under the constraint that it is orthogonal to the preceding one. By selecting the top several principle components, the computational complexity is reduced, and machine learning algorithms can run faster.

When implementing PCA to data set, here are five main steps to follow:

- Normalize the data.
- Calculate the covariance matrix.
- Compute the eigenvalues and eigenvectors.
- Choose the top components and form a feature vector.
- Form principal components.

PCA was applied to examine a large-scale spectral line image by [Heyer and Peter Schloerb \(1997\)](#), and it successfully determined three major emission components within the Sh 155 (Cep OB3) cloud complex and spatial distributions, proving it as a powerful tool.

PCA is a widely applied technique in machine learning, to find the correlation between original variables, reduce the dimension, and extract principal components. If a large data

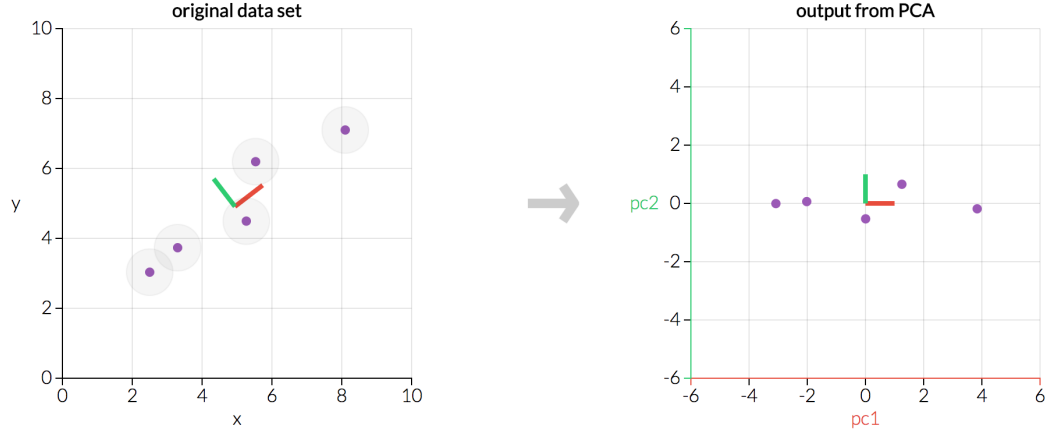


Figure 4.21: A 2D example for PCA. The two charts show the same data while the right one reflects the original data transformed onto two orthogonal principle components.

set needs to be analyzed without knowing much characteristics in advance, PCA can help to decrease the noise and irrelevant information, to focus on important variables.

The bandwidth of our spectra is ~ 15 GHz, consisting of a plethora of molecules with various line features, while the relationship between them remains unknown. Hence, PCA is believed to shed light on the classification of star-forming core spectra and their evolutionary characteristics.

4.4.2 Preparations before PCA

To apply PCA to our spectra, the first step is to align the data, since the cores are spread along the filament and have different velocities. According to the Doppler effect, the frequency will be larger if the core is moving towards us, or smaller when flying apart. To get the frequency difference, the correlation function is considered, which measures the similarity of two sequences of signals as a function of the displacement of one relative to the other. When the shift leads to a larger overlap, the function reaches a higher value. The frequency difference is hence the displacement at the maximum of the correlation function, counted by channel number.

In our case, the spectra towards the window spw1 (218.5 GHz) usually display significant H_2CO 3(0,3)-2(0,2) line, which is adopted to examine the frequency shift. H_2CO 3(2,2)-2(2,1) and 3(2,1)-2(2,0) are not favored because they are too weak in several cores to be taken into account. Other molecules such as DCN and SiO are even weaker than H_2CO , which is also not reliable. Therefore, the spectrum in the range 218.183-218.253 GHz is selected out, the continuum subtracted and the absolute value is normalized by the sum. The frequency offset is then derived by the correlation function with respect to the spectrum of region 5 (shown in 4.2), because it clearly shows a single line peak and without any absorption. Later, all four spectral windows are modified by this offset.

The second consideration before PCA is to remove the lines that are blended or with absorption. CO is one of the most extended molecules in the interstellar medium, and CO (2-1) tends to show an artificial absorption feature because of the interferometer observation. Other lines including CH_3OH (4-3) and H_2CO 3(0,3)-2(0,2) also present strong absorption due to bad channels in the data cube, originating from the imaging process. Therefore, the spectra

between 230400-230800 MHz (CO 2-1), 218417-218457 MHz (CH₃OH 4-3) and 218183-218253 MHz (H₂CO 3(0,3)-2(0,2)) are excluded from further PCA. For spw2, it has a narrow frequency range (174 MHz) and mostly overlaps with spw3. The high frequency end of spw3 covers the most significant lines (¹³CS, CH₃OH, and C₂H₅CN) in spw2, so spw2 is also discarded in PCA.

The third and most widely discussed issue in PCA is the scaling of the original variables. The aim of PCA is to find a combination of variables that maximize the variance. If one variable varies more than another one, e.g., weight in kilograms vs height in meters, the derived principal components will be more closely correspond to the one with higher variance (here weight) if these features are not normalized. Hence, scaling is of great importance in PCA. Usually, the data are normalized to a standard normal distribution, with a mean of zero and a standard deviation of one. However, if the data is far from the normal distribution or measured under the same scale, it is recommended to normalize by the sum of each data set.¹⁰ For our data, the variables are the spectral intensity in different channels. To preclude the effect of the absolute intensity of the cores and strengthen the relative intensity of different lines, we standardize each spectrum by the sum of the area under its profile. Before scaling, the spectra in three windows are already smoothed to the same frequency resolution (0.97664 MHz per channel), continuum subtracted, aligned using the velocity shift, removed "bad" lines, and then connected in the sequence of frequency. After scaling, the central hot core shows a spectrum very similar to white noise or to an extremely weak core. This hot core is also excluded from PCA; otherwise, it will be sorted to the very cold core category.

4.4.3 Applying PCA to the spectra of cores

Now PCA is ready to start. The `decomposition.PCA` function in the `sklearn`¹¹ Python package is adopted to analyze normalized data (66 cores and 4706 channels). The percentage of eigenvalues in descending order are shown in Figure 4.22. The first eigenvalue counts the highest percentage and almost half of the features (49.7%), while the eighth eigenvalue is close to 1%. The continuing eigenvalues are not significant enough to be considered. In total, the eight eigenvalues count for 78.4%, which corresponds to eight principal components for further analysis. The first six corresponding eigenvectors are plotted in Figure 4.23, from which we can see the decisive lines in each eigenvector. The first eigenvector is very similar to the spectrum of region 5, which shares some common features of all the spectra. The other eigenvectors show both emission and absorption, which means that some lines are only significant for certain cores. For a specific core, its spectrum could be mostly reproduced by a linear combination of eight eigenvectors. The linear coefficients define the core location in an eigenvector space, or, in other words, a principal component space. More intuitively, the original 66 cores are projected onto a pair of principal components and shown on the PC axis (Figure 4.24).

4.4.4 Clustering and classification

With the new position of the data in the principal component space, the `sklearn.cluster.KMeans` function is utilized to classify the cores. KMeans is a clustering algorithm, which separates data into groups with equal variance and maximum internal coherence¹², which requires a user-defined number of clusters.

¹⁰https://scikit-learn.org/stable/auto_examples/preprocessing/plot_scaling_importance.html

¹¹Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.

¹²<https://scikit-learn.org/stable/modules/clustering.html#k-means>

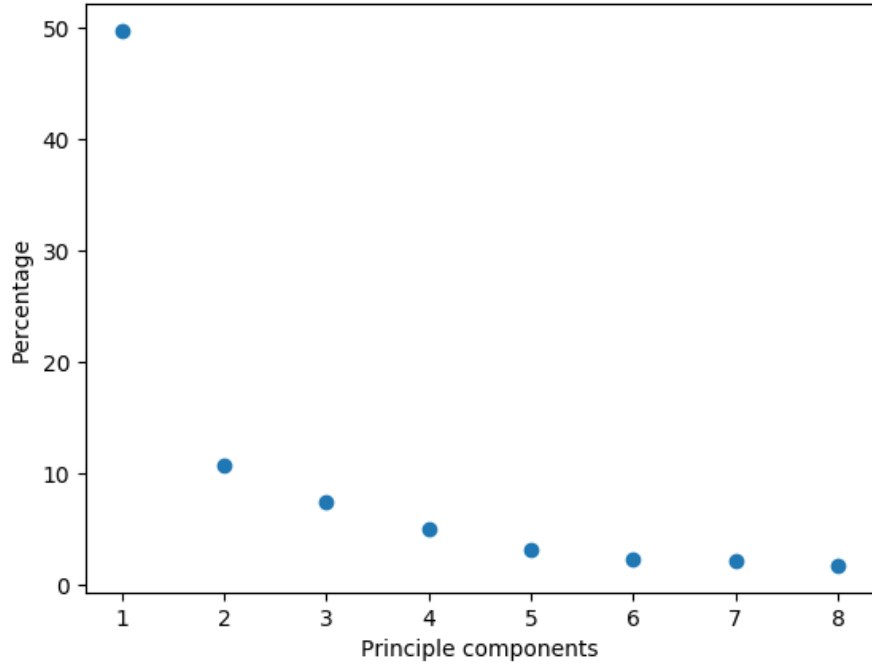


Figure 4.22: The percentage of first eight eigenvalues in descending order.

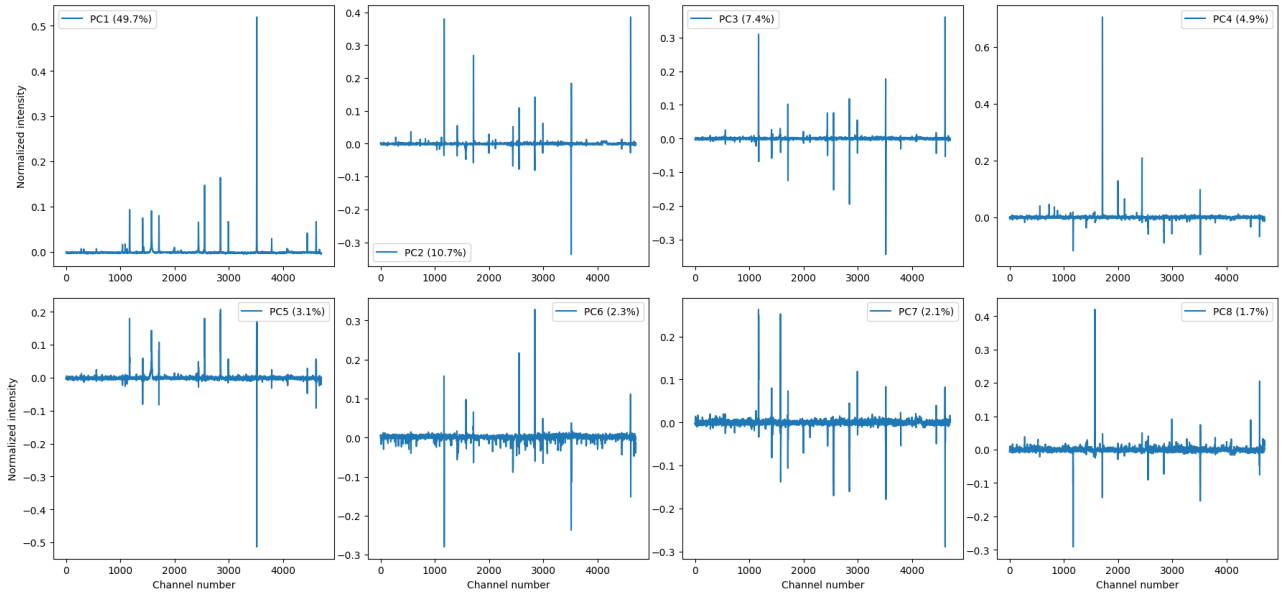


Figure 4.23: Eight eigenvectors showing the spectral feature. The first eigenvector counts the highest percentage and nearly half of the features (49.7%), which includes the most common and extended lines. The other eigenvectors are plotted in descending order. In total, the eight eigenvectors represent 78.4% of line features.

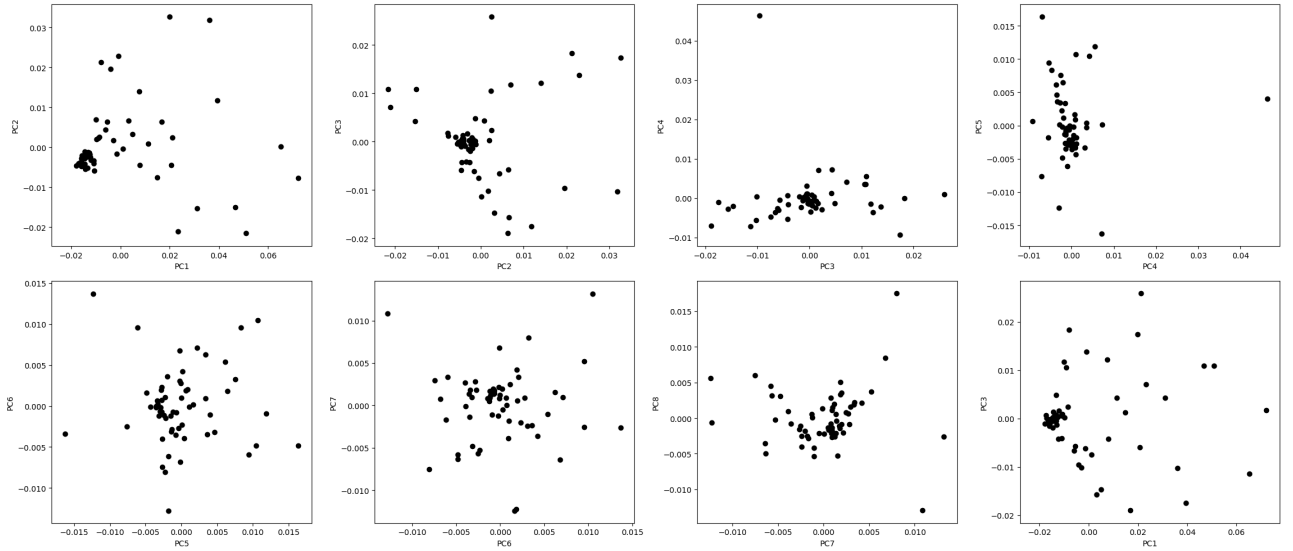


Figure 4.24: Core distribution in different principle component space. Some cores tends to cluster together while some are more widely scattered.

To obtain the best number of clusters, we applied several tests. In the first approach, introduced as elbow-point analysis ([Malika Charrad, 2014](#)) in Section 3.2.2, we run the K-Means algorithm for a range of K values changing from 2 to 9. For each run, we calculate the within-cluster sum of squares (WSS). As seen in the left panel of Fig. 4.25, the elbow point is at $K = 3$, where the decrease rate changes the most.

The second Calinski-Harabasz score (also known as the Variance Ratio Criterion, [Calinski and Harabasz \(1974\)](#)) is defined as the ratio of the sum of between-cluster dispersion and of within-cluster dispersion. The between clusters sum of squares (BSS), measures the variability between the clusters accounting for the center of the data. The WSS measures the variability between the data points assigned to cluster K and the corresponding centroid. Because the total variation in the data, the total sum of squares (TSS), is fixed, minimization of WSS corresponds to the maximization of BSS. The Calinski-Harabasz score is then given by:

$$CH = \frac{\frac{BSS}{K-1}}{\frac{WSS}{N-K}}. \quad (4.13)$$

A large value of this index for a given value of K is an indication of a clustering solution with low within variability and large between cluster variability. As seen in the middle panel of Fig. 4.25, the CH score quickly drops from $K = 2$ to $K = 5$ and remains equally low for $K > 5$. Then the best K is 2.

The third approach is to calculate the Silhouette value, which is associated with each data point in a cluster, ranging from -1 to 1 ([Kaufman, 1990](#)). A high Silhouette value indicates dense and well-confined clusters, verified in previous studies, e.g., ISM segmentation ([Colombo et al., 2015](#)). Then we run the K-Means algorithm similarly as the first method and compare the average Silhouette values for different numbers of clusters (2–9). The configuration with the highest value is the optimal one, which is $K=2$ according to the right panel of Fig. 4.25. The Silhouette value also has a local peak at $K=6$, which is still smaller than the value at $K=2$ and $K=3$.

In conclusion, statistics prefer a small number of clusters, the best of which should be 2 or 3. It is not suggested to try clustering cores into more than 4 groups. This may infer the

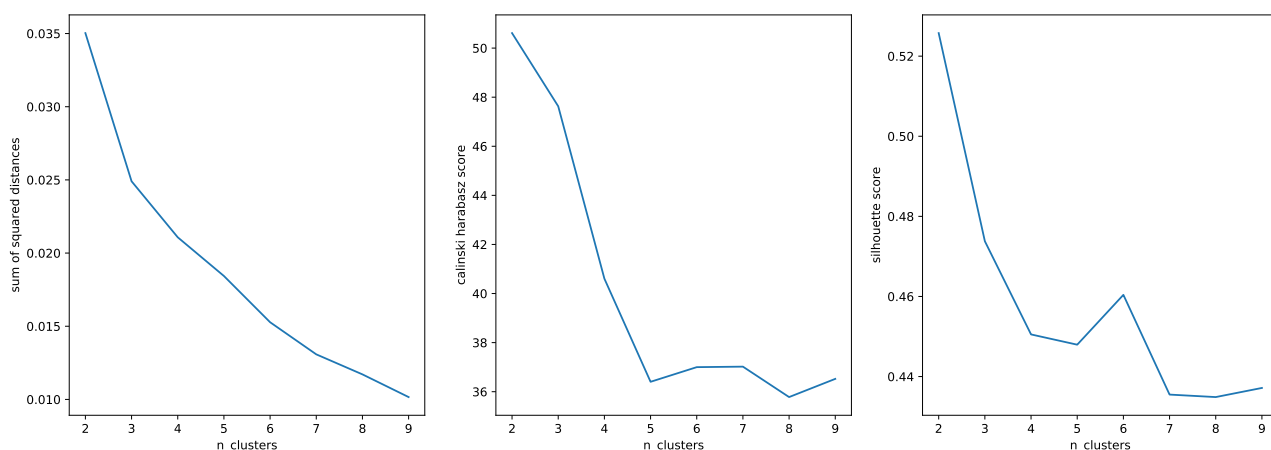


Figure 4.25: The statistical tests for determining the optimal number of clusters. Left: The sum of squared distances (WSS) changing by the number of clusters. The elbow point at 3 clusters is the best. Middle: The Calinski-Harabasz score is larger for a better clustering, showing the best number of cluster is 2. Right: The silhouette score peaks at 2 clusters, which is the best point.

continuity of core properties: if the core data are naturally strip-shaped or non-clustered, they are difficult to be classified into different clusters. In this case, the star-forming cores may not show distinct evolutionary stages between the cold and hot cores.

When we set the desired number of clusters to three, clear clustering features are visible in the upper panel of Figure 4.26, plotted in different colors for different clusters. The three groups contain 6, 15, 39 cores, respectively. We also tried to classify the cores into four groups, and the resulting groups each consist of 5, 6, 10, and 39 cores (lower panel of Figure 4.26). A larger number of clusters is not plausible since our sample size is not very large, while fewer clusters result in coupling two distinct groups together.

From three clusters to four clusters, only one core in the red group goes to the purple group, and only one black core is shifted to the blue group. The original blue group is separated into two subgroups. Therefore, the red group and the black group are believed to be well classified, and the other cores may belong to one or several intermediate groups between the red and black group. The cores in the same group are considered to stay in a similar star-forming phase, hence providing some hints of classifying evolutionary stages. We will discuss the reasonable classification results by analyzing the average spectra of each group in Section 4.4.5.

4.4.5 Common spectra of groups

To examine whether a group shares some evolutionary hints or not, the spectra are analyzed in detail. Take, for example, the three groups (Figure 4.27), each normalized spectrum is marked by a color corresponding to one of the three classified groups and reordered. The black spectra show a relatively high level of noise and several weak lines. The blue and red spectra show smooth baselines and prominent molecular lines. To know more about the internal chemistry, it is helpful to check the average spectra of the originals in each group (upper panel of Figure 4.28). From bottom to top, the groups are named from group 1 to group 3, showing an increasing number of lines, as well as the line intensity. For group 1, the lines are barely seen, except the lines of simple molecules that trace extended emission, including H_2CO , H_2S , ^{13}CS , DCN ,

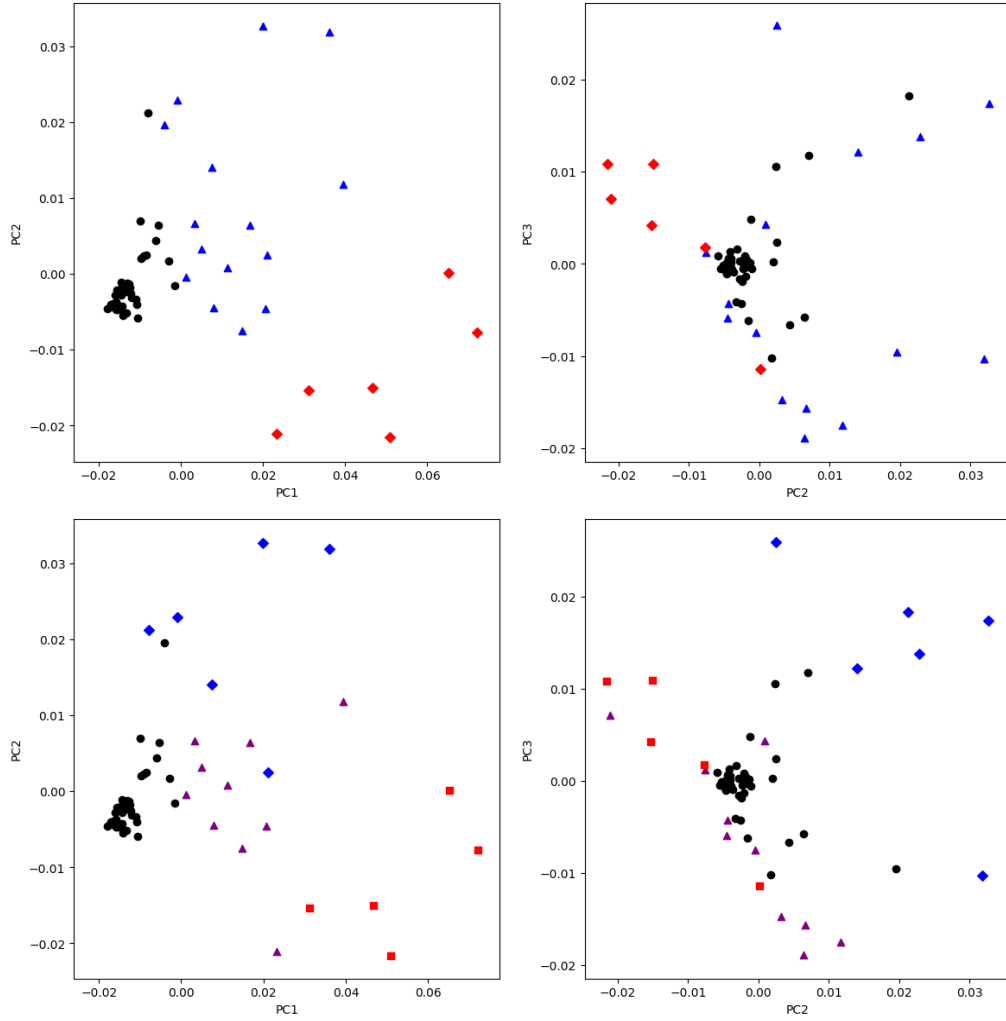


Figure 4.26: Core clustering results in the first three principle component space. Upper panel: use three clusters, containing 6, 15, 39 cores, respectively. Lower panel: use four clusters, each consists of 5, 6, 10, and 39 cores. The members in red group and black group are almost the same in both case.

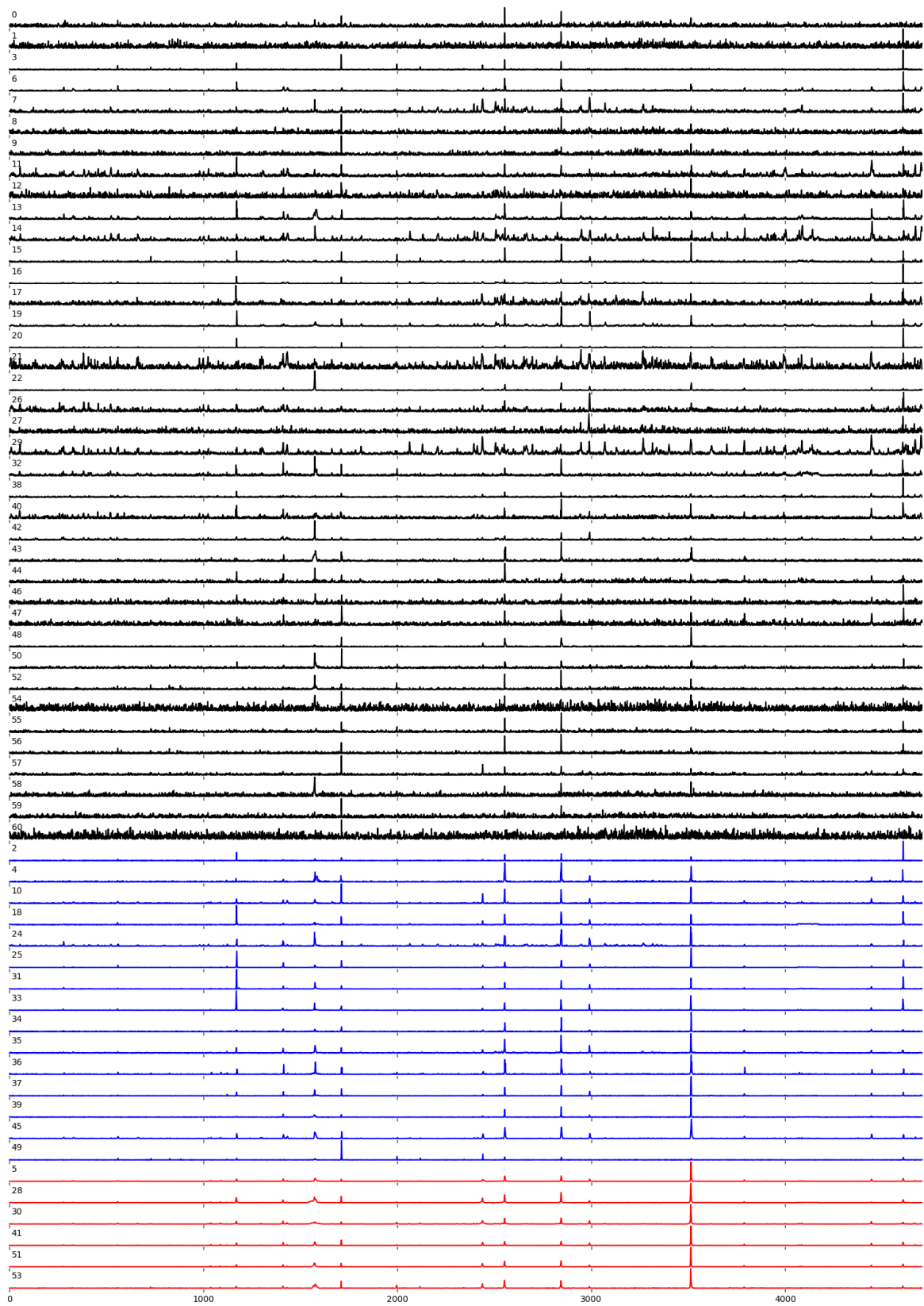


Figure 4.27: The aligned and classified spectrum of 60 cores for three groups.

and OCS. Weak CH_3OH and SiO also exist in group 1 spectra, implying nearby star-forming activity. Group 2 is more evolved, showing peaks of several complex molecules, like CH_3OCHO and CH_3CHO , though relatively weak. Group 3 starts to have more molecules related to the hot core, e.g., clear lines of CH_3OCH_3 and CH_3COCH_3 , indicating it is at a later evolutionary stage than the other groups.

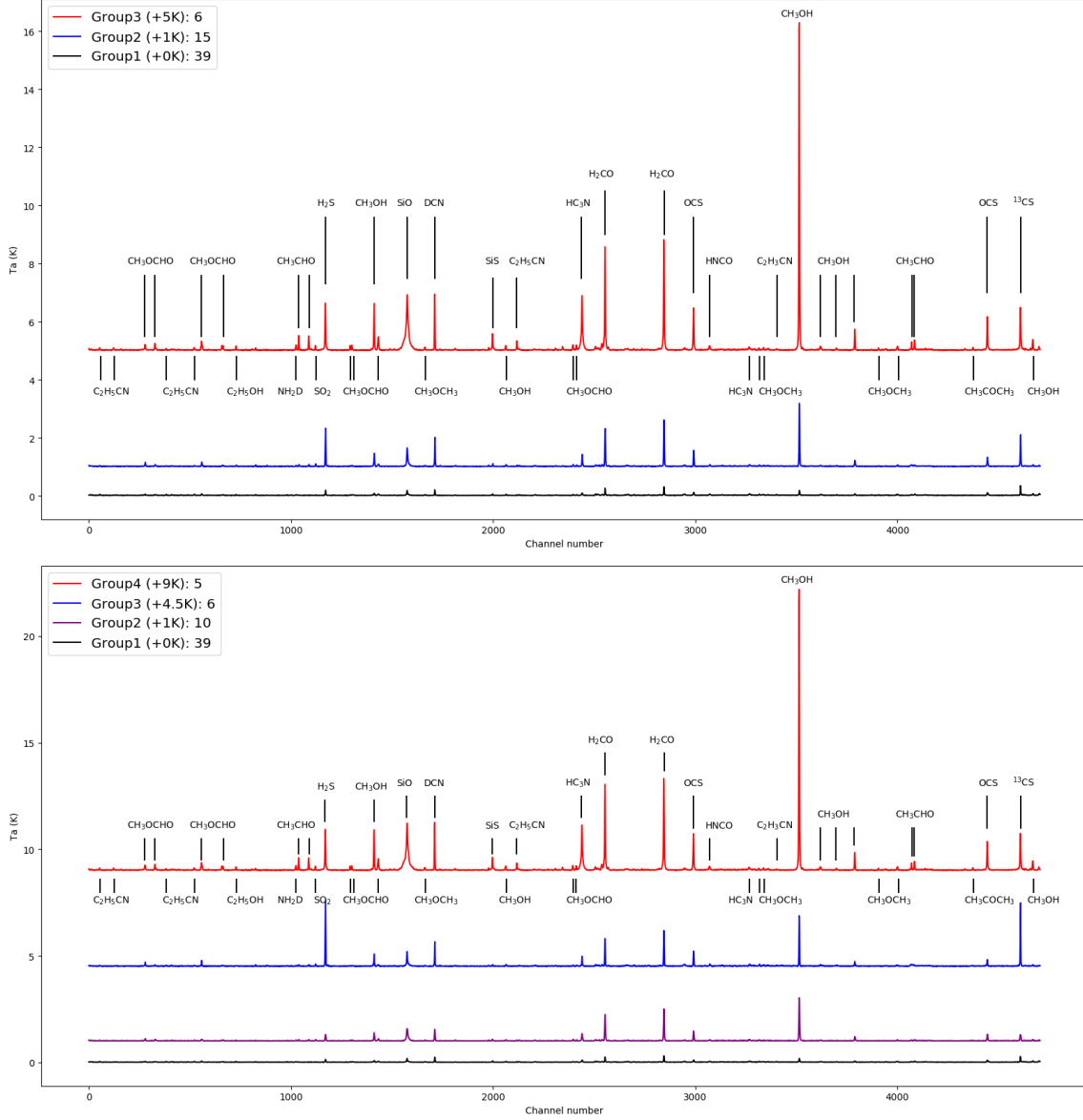


Figure 4.28: The average spectra of different groups. Upper panel: Classified into three groups. Lower panel: four groups.

When we classify them into four groups, the first and fourth groups are similar as before. The previous intermediate group is divided into two subgroups. In the new classification, group 3 resembles group 2, except the H_2S and ^{13}CS lines. These two lines in group 3 are even stronger than those in group 4, which indicates that sulfur-bearing species may trace different stages of star formation. However, OCS, SO_2 and SiS remain the same for these two groups, which is a question of the rule that sulfur plays. The different trend of these sulfur species is possibly due to the limit of oxygen and silicon abundances. When carefully looking at the spectra, more complex molecular lines could be found in group 3, and sometimes stronger than in group 2, e.g., CH_3OCHO .

Table 4.6: Molecules detected in different groups

Group 1	Group 2	Group 3	Group 4
weak H ₂ CO	H ₂ CO	H ₂ CO	strong H ₂ CO
very weak CH ₃ OH	weak CH ₃ OH	CH ₃ OH	strong CH ₃ OH
very weak H ₂ S	weak H ₂ S	strong H ₂ S	H ₂ S
weak ¹³ CS	weak ¹³ CS	strong ¹³ CS	¹³ CS
very weak OCS	OCS	OCS	strong OCS
very weak SiS	very weak SiS	very weak SiS	SiS
very weak SiO	SiO	SiO	strong SiO
very weak DCN	weak DCN	DCN	strong DCN
very weak HCCCN	weak HCCCN	HCCCN	strong HCCCN
	very weak SO ₂	very weak SO ₂	SO ₂
	very weak NH ₂ D	very weak NH ₂ D	NH ₂ D
	very weak HNCO	very weak HNCO	weak HNCO
	very weak CH ₃ OCHO	weak CH ₃ OCHO	CH ₃ OCHO
	very weak CH ₃ CHO	very weak CH ₃ CHO	CH ₃ CHO
		very weak CH ₃ OCH ₃	weak CH ₃ OCH ₃
		very weak C ₂ H ₅ CN	weak C ₂ H ₅ CN
			weak C ₂ H ₃ CN
			weak C ₂ H ₅ OH
			weak CH ₃ COCH ₃

Table 4.7: Physical parameters of different groups

Parameter	Group 1	Group 2	Group 3	Group 4
Methanol T _{ex} (K)	50.2	82.0	86.7	105.0
Methanol N (cm ⁻²)	2.11×10 ¹²	3.05×10 ¹⁵	2.18×10 ¹⁵	6.17×10 ¹⁶
Median Mass (M _⊙)	0.76	1.15	0.94	5.91

To be more accurate, XCLASS is applied to facilitate the identification of molecules and quantification of their physical properties. The molecules detected in each group are shown in Table 4.6, from which a distinct chemical ladder can be seen through the groups. The unsaturated carbon chain molecule HCCCN exists in all of the groups, implying the early condition of our cores. It is also interesting to see CH₃OCH₃ and C₂H₅CN appear in group 3 but not in group 2, indicating some real chemical difference between these two groups. The fourth group contains lines not only stronger than the same one in the other groups, but also lines that never appeared previously, including C₂H₃CN, weak C₂H₅OH, and CH₃COCH₃.

We also fitted methanol lines to obtain an average temperature for each group. And not surprisingly, the excitation temperatures also increase from group 1 to group 4, as noted in the first line of Table 4.7. Group 1 is actually warmer and richer than expected, possibly conceiving low-mass protostars, since weak SiO is detected. With T>100 K and high column density of methanol, this group of cores is believed to be possible candidates for hot cores. However, group 2 and group 3 share similar temperatures and column densities, as well as molecules.

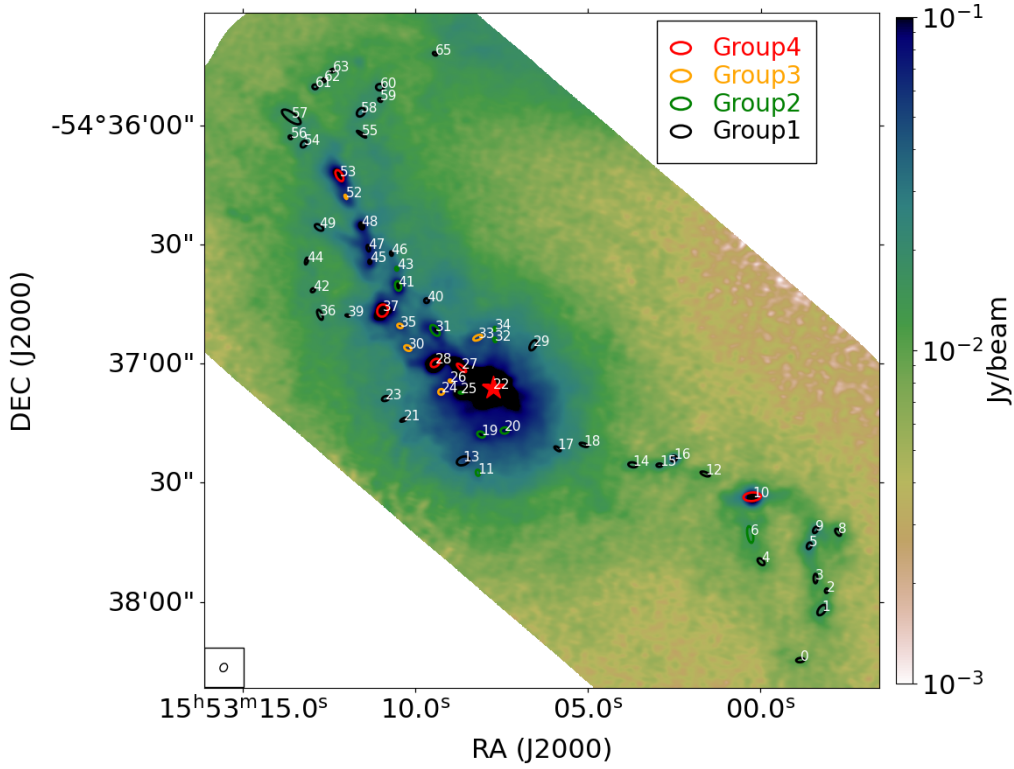


Figure 4.29: The distribution of groups through the filament, classified into four groups.

4.4.6 Spatial distribution of groups

It is worth studying the spatial distribution of groups to understand the morphology. Checking the results of the four groups shown in the lower panel of Figure 4.29, group 1 is usually far from the main filament, in the dispersed region. The second group is located along the filament in a more enhanced region. The third group mainly surrounds the hot core, except one that is at the tail of the filament (region 2). The region around the hot core is deep into the potential well and accumulates a lot of mass, which leads to the outstanding feature of groups 3. It should also be mentioned that this region is affected by the second lobe of the interferometer mapping and may duplicate the hot core spectrum as artificial features, which might be the reason that PCA classified these cores as one group. Group 4 traces the most significant cores on the filament and shows rich chemistry. They are not as strong as the central hot core, but already have some features related to a hot core, and therefore they are very likely to form at least a low-mass star.

4.4.7 Limitations and discussion

G327.3-0.6 region is a large filament that hosts a hot molecular core, several high-mass cold cores, as well as other lower mass cores. PCA could provide some hints of the complex star-forming activity in the filament.

The direct hints are pointing towards chemical evolution in cluster formation. In the observation and model of [van der Tak et al. \(2000\)](#), it is suggested that the temperature of CH_3OH can be used as an evolutionary indicator during the embedded phase of massive star formation, independent of source optical depth or orientation. We also found a gradient of CH_3OH temperature and density in different groups. If we assume a methanol abundance of

10^{-8} , the H_2 density of group 4 can reach 10^{24} cm^{-2} , which is very dense for a core.

In the five cores of group 4, four of them have intermediate mass, ranging from $4.0\text{--}6.9 M_\odot$. These cores have higher potential to form intermediate-mass stars or even high-mass systems. Group 3 and group 2 seldom include intermediate-mass cores, whereas low-mass cores are very common. The median mass of group 1 is $0.76 M_\odot$, which rather tends to disperse, not even forming a low-mass star. There are two high-mass cores in group 1 since their core sizes are much larger than the average. In general, PCA can partly classify high- and low-mass cores by simply analyzing the spectra, which could be very useful for filtering large samples. Alternatively, a massive core classified into the first group could be a good example of high-mass star formation in an extremely early stage.

Another interesting thing is that the numbers of cores in each group decrease as the lines become richer, reminding us of the star formation efficiency. Current star formation efficiencies from a molecular cloud to a star range from 3% to 6% ([Evans et al., 2009](#)). For our samples, the first group is in the earliest stage and may not form stars at the end, which counts for 65% of the cores. Only five cores are chemically rich, though still far away from being a real star. These five cores share 8.3% of all cores, which is very close to the star formation efficiency. This indicates that the evolution rate from early stage to latter stage may also follow a similar efficiency. At least the tendency is clear: the more evolved, the fewer cores.

5 Large scale structures

The molecular spatial distribution is found to be various from the coldest dense cores to the warm environment and hot regions (e.g. [Qin et al. \(2022\)](#)): dense gas tracers like DCN are visible across the filament and peak on the cores; extended emission tracers like H₂CO are luminous through the entire filament and even strong at the shocked regions; temperature tracers like CH₃OH are stronger towards the hotter regions; hot core tracers, basically complex organic molecules (COMs), are condensed on the limited hot core regions rather than spreading through the entire star-forming cloud. Hence, it is worth systematically assessing the distribution association between different molecules and dust continuum.

5.1 Deriving moment maps

To characterize the spatial distribution, the first step we adopted is to derive the corresponding moment maps for each line and each molecule, neglecting the complex spectral profiles if they deviate from Gaussian. We use the CASA function *immoment* to derive the moment maps. We considered all pixels with positive values in the corresponding channel ranges of a certain line. The moment-zero map is the integrated intensity map at a certain velocity range ($\int I_v dv$), which roughly gives the amount of gas at each pixel. The moment-one map shows the velocity fields (vlsr), in other words, the intensity-weighted coordinate ($\int v I_v dv$). The moment-two map can be derived by ($\int v^2 I_v dv$), which represents the velocity dispersion (or line width). To make it easy to remember and discuss, people name the moment maps by the order of frequency ν in the derivation equation.

The moment 0 maps of 24 molecules and 42 transitions are derived, shown in Fig. 5.1. It is impressive that some molecules are very extended through the whole field of view, while some are relatively compact and weak. For the same molecule, the moment 0 maps of different transitions are also varied. Usually, the higher the transition energy, the less extension of the molecular emission. Inspired by the search for similar spatial distributions of different molecules/transitions, we adopt a statistical approach in the following section.

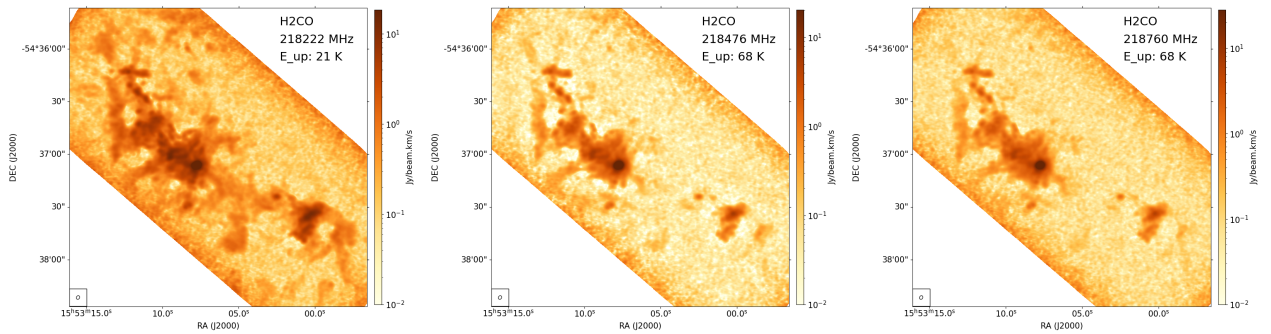
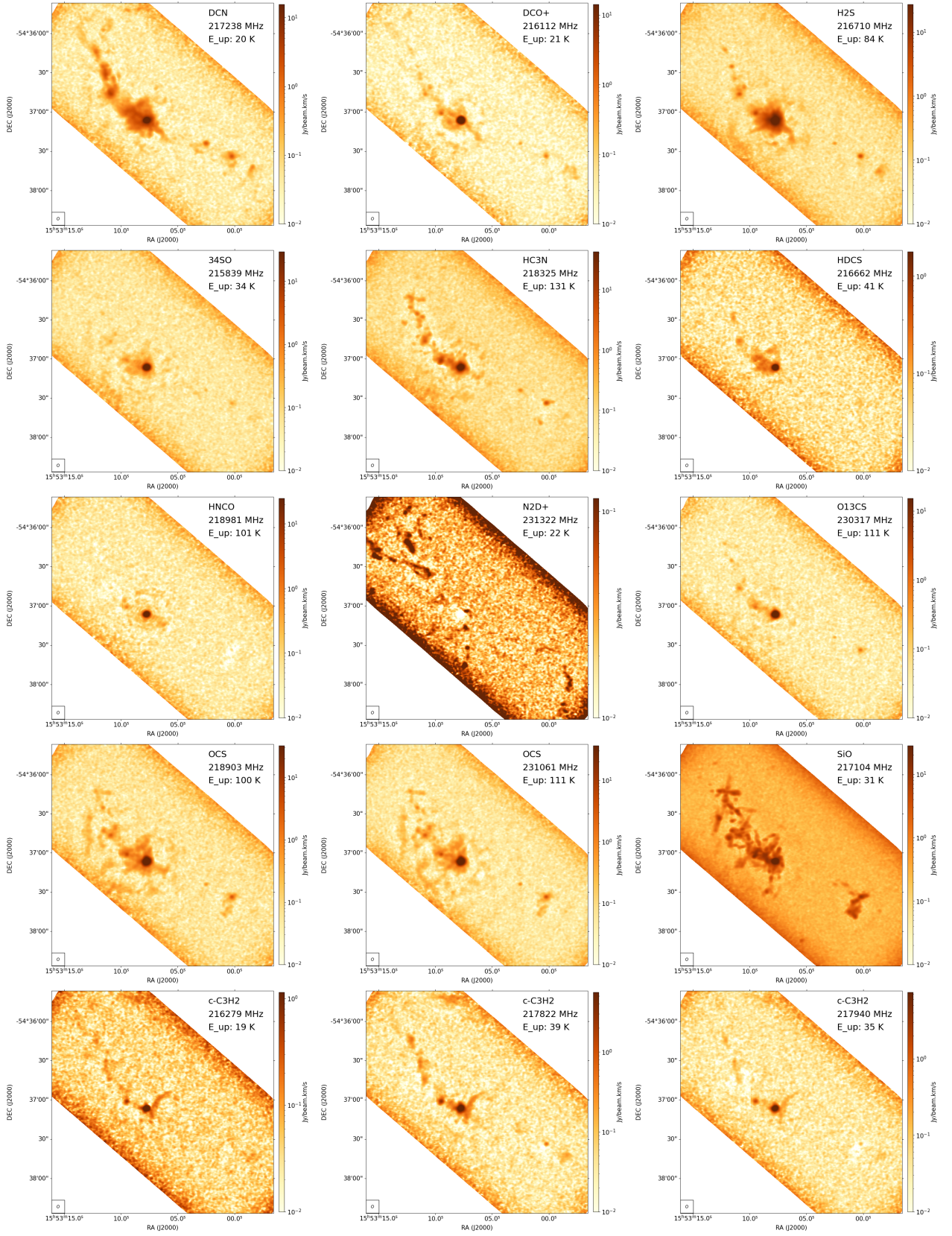
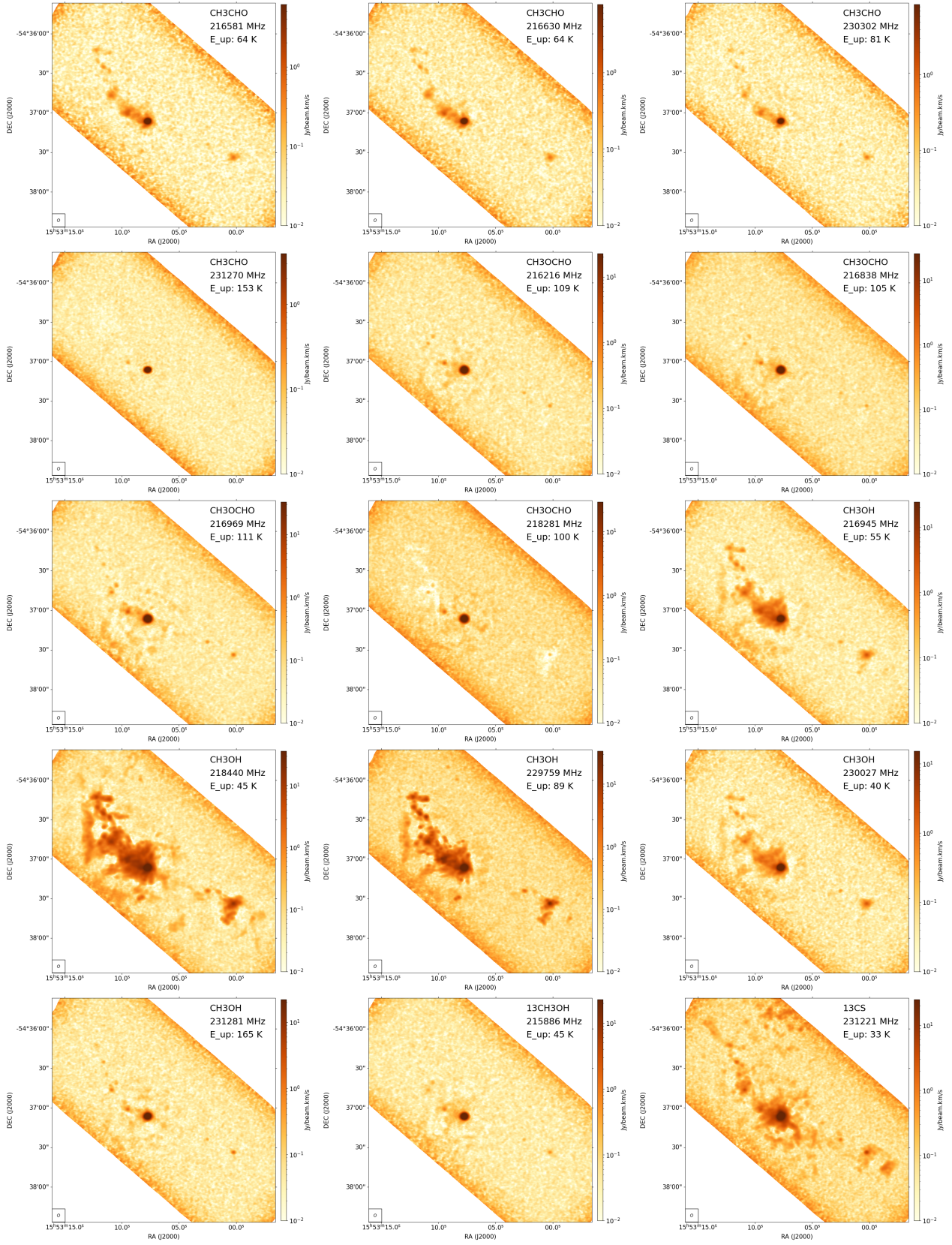
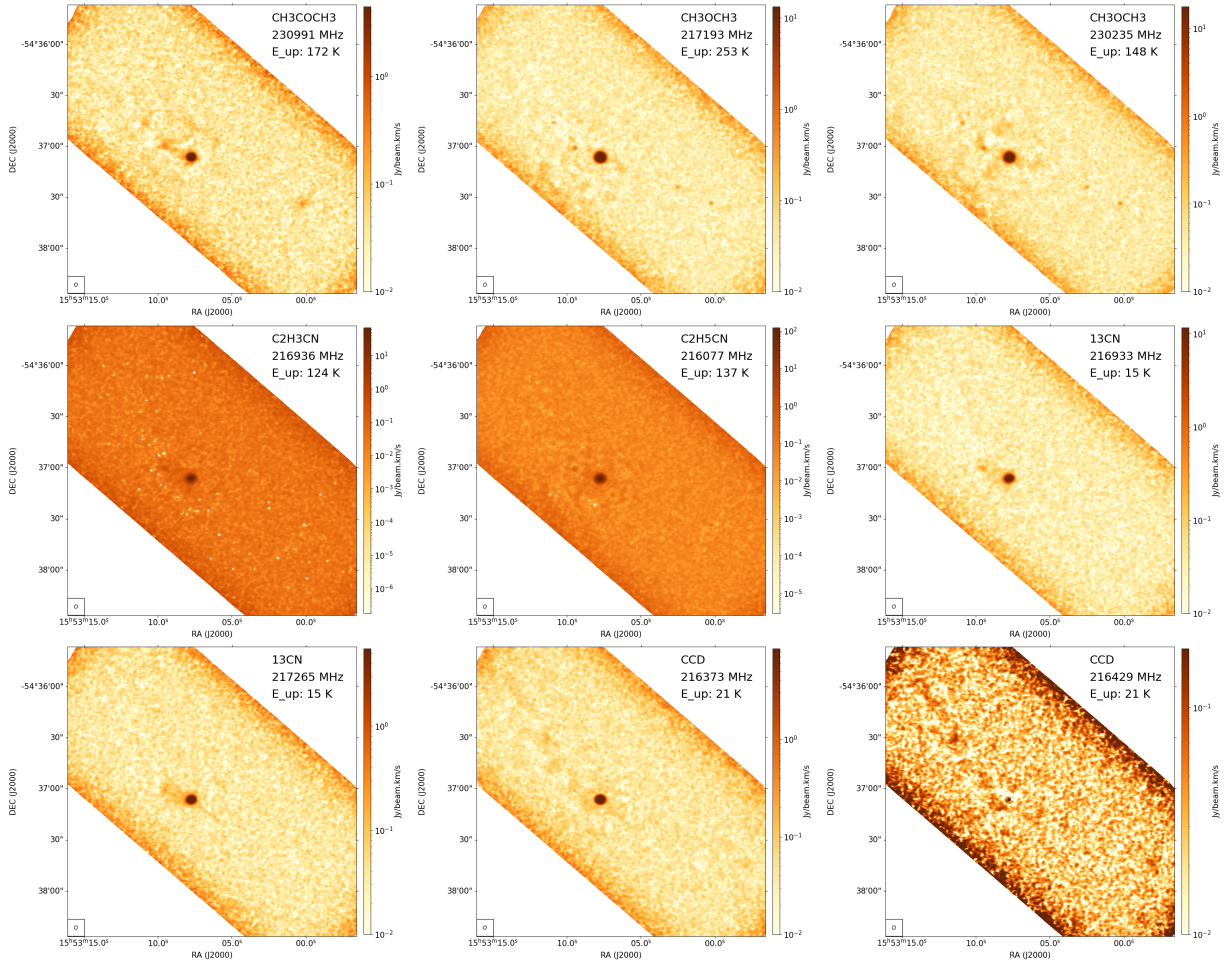


Figure 5.1: The moment 0 maps of the 24 molecules and 42 transitions that were analyzed in this study, shown in a logarithm color scale. The molecular name, the corresponding rest frequency and upper energy level of each transition are shown in the upper right corner of each plot. The beam size of the observation is shown in the bottom left corner. Continued in the following pages.







5.1.1 Correlation of different transitions

After deriving the moment maps, we are interested in comparing them for the different transitions of the same molecule. In our sample, the molecule CH_3OH shows the largest number of transitions that are extended. Therefore, we adopted the 5 lines of CH_3OH to plot a correlation map. For one transition, we got the intensity at a certain pixel of the moment 0 map, and another intensity at the same pixel of another transition. Then using the intensity of two transitions as an axis, we plotted the intensity distribution of all pixels. To compare every two lines, we generated 10 correlation maps (see Fig. 5.2). From the plot, we could tell there are different components in different transitions. For the transition at 216945 MHz (55 K) and 218440 MHz (45 K), the intensity distribution is almost diagonal and can be fitted with a straight line, which means they are similar to each other, despite a ratio of 4 in the intensity for the lower energy state line towards the higher state one. However, for the line 2312281 MHz with $E_{\text{up}}=165$ K, the intensity distribution is not simply correlated with a lower energy line, instead showing multiple components. For instance, the 230027 MHz (40 K) versus 2312281 MHz map (bottom left panel of Fig. 5.2) shows at least five components, each can be fitted by a line with different slopes. The component that has the highest intensity in both transitions may come from the hot core with a slope of 1. The other components also show a similar slope around 1, while the intercepts are different, representing different core components throughout the field of view. Since it is more difficult to excite the high-energy lines, the cold and diffuse cores are not visible at the higher-energy moment maps, while the hot and dense cores stand out in the correlation map. However, we cannot quantitatively measure the correlation with these intensity distribution plots. Hence, we try some recent and popular methods in the next sections.

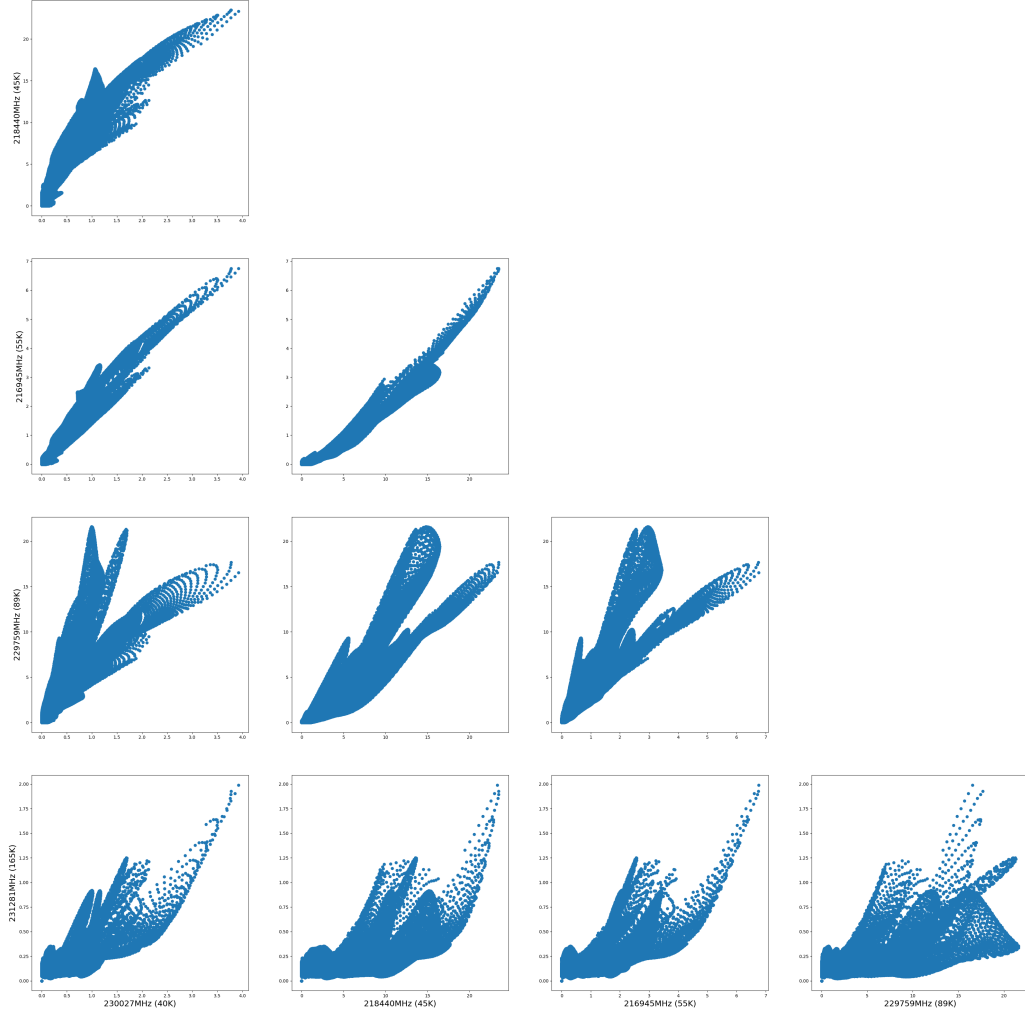


Figure 5.2: The correlation between the moment 0 maps of five different transitions of CH_3OH . For each subplot, the x and y axis corresponds to the intensity of moment 0 maps of two transitions, including all pixels within a high-SNR area.

5.1.2 Applying PCA to moment 0 maps

Since PCA is useful in the classification of the cores by their spectra, it is expected to characterize the moment 0 maps of different molecules, to explore the similarity of the spatial distribution between the transitions. A similar test has been performed in Liu et al. (2020).

Since all of the moment 0 maps have the same resolution and pixel size, we do not need any smoothing or re-gridding steps. Then we flatten each image (2560*2560 pixels) into an array (1*6553600) and normalized it by the corresponding maximum value. We also made a mask to remove the noisy edge part, nan values, and the hot core region. In total, the data matrix used for PCA is of size 42*553600. The resulting PCA eigenvalues (in percentage) for the principal components are shown in Fig. 5.3, in which the first 8 PCs already count for more than 80% of the total eigenvalues. Interestingly, the first PC only describes 39.1% of the variance in the data. This is much lower than the values in other studies, e.g., 73.6% in Liu et al. (2020), and 80% in the spectral analysis in this thesis. The dropping slope of eigenvalues is also flatter than other works, which infers the complexity of the moment 0 maps and more PCs are needed to better explain the variance.

The eigenvectors in the PCA analysis describe the correlations between the molecules and the PCs, which are listed in Table 5.1. A negative/positive eigenvector indicates the molecule is anti-/correlated with the PC. The larger the value, the stronger the correlation holds. After reshaping the eigenvectors back to 2D and constructing the principal components as images in Fig. 5.4, we clearly see that many PCs are dominated by certain molecules. Therefore, we explore the representative molecule for each PC by reading the eigenvector table and the eigenvector image. For example, PC1 peaks at H₂CO with a value of 73, reading from Table 5.1), which means that PC1 is strongly related with H₂CO that traces the large scale extended structures as well as dense cores. We marked the most important molecules in the upper right corner of each plot in Fig. 5.4. The other PCs are analyzed similarly. PC2, PC3 are mostly correlated with CCD and HDCS separately, which trace more towards the core rather than extended emission. PC4 and PC5 were surprisingly found to be highly correlated with N₂D⁺, which shows emission in cold cores rather than hot components. This confirms the importance of N₂D⁺ as an early star-forming tracer, which is expected to find more cold cores than the continuum.

Table 5.1: The eigenvectors of the principle components

Component	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
¹³ CN-2	-9	-5	11	-3	-1	-8	-5	6
DCN	27	9	13	-16	-4	10	3	-8
CH ₃ OCH ₃ -2	-25	-4	-3	-1	-2	-2	-0	4
SiO	33	0	-14	0	1	-12	-5	5
CH ₃ OCHO-4	-14	-6	-1	3	-1	-2	-1	-0
CH ₃ OH-3	13	-2	5	-1	-1	3	-3	1
¹³ CN-1	-17	-4	2	1	-0	-4	-2	4
CH ₃ OCHO-2	-18	-5	-1	0	-3	-2	-1	2
H ₂ S	-7	-4	6	-10	-12	12	-6	1
HDCS	36	7	50	-12	20	-12	2	10
CH ₃ CHO-1	14	-5	3	12	7	-2	-3	-9
CH ₃ CHO-2	25	-4	4	14	8	1	-5	-14
CCD-2	-31	82	1	15	-4	1	-5	1
CCD-1	-21	7	5	-10	-3	-2	36	-8
c-C ₃ H ₂ -1	-19	-2	-7	-0	1	-14	6	-11
CH ₃ OCHO-3	-19	-5	-1	0	-2	-2	-1	3

Continued in the next page

Component	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
DCO ⁺	1	1	5	-3	-4	1	4	-4
¹³ CH ₃ OH	-13	-4	1	0	-2	-0	-1	3
³⁴ SO	-6	-3	5	-6	-7	5	-3	3
HNCO	-13	-5	-0	-2	-9	1	-0	-5
OCS-1	-8	-4	-1	-2	-5	4	-2	3
H ₂ CO-3	48	5	-7	-4	-4	2	1	-0
H ₂ CO-2	45	4	-8	-1	-2	2	0	-1
CH ₃ OH-2	41	3	2	-4	-2	3	-2	2
HC ₃ N	-5	-5	-1	6	1	-6	-3	-4
CH ₃ OCHO-1	-22	-5	-2	1	-2	-3	-1	2
H ₂ CO-1	73	11	-30	-5	-9	-7	9	10
c-C ₃ H ₂ -3	-18	-4	-3	2	1	-9	1	-4
c-C ₃ H ₂ -2	-14	-4	-4	2	1	-8	1	-5
¹³ CS	-6	-2	3	-11	-12	11	-3	0
CH ₃ CHO-4	-23	-6	-6	6	0	-15	1	17
CH ₃ OH-5	-16	-6	-1	2	-3	3	-2	1
N ₂ D ⁺	-27	8	-26	-30	43	9	-4	-1
CH ₃ OH-4	29	-2	-1	10	6	-2	-3	-5
CH ₃ OH-1	6	-2	1	-1	-3	5	-3	2
CH ₃ OCH ₃ -1	-23	-5	-3	0	-2	-1	0	2
CH ₃ CHO-3	13	-6	2	17	7	1	-5	-13
O ¹³ CS	-5	-5	-0	8	2	1	-2	-4
CH ₃ COCH ₃	4	-9	1	33	17	23	15	15
OCS-2	-8	-4	-1	-1	-4	3	-2	3
C ₂ H ₅ CN	-6	-2	3	-10	-11	11	-3	0
C ₂ H ₃ CN	-15	-6	-2	3	-2	1	-2	0

Fig. 5.5 shows the plot of the eigenvectors of each transition in the PC space. In these maps, we mark the upper energy levels in different colors to check the tendency of the temperature distribution. In the PC1 vs. PC2 plot, the different energy levels are not well separated. The most distinct moment maps belong to CCD and H₂CO, which occupy different corners in the PC1/PC2 plot. When looking a step closer, we find the high temperature transitions are located at a very narrow and negative zone of both PC1 and PC2, so they are anti-correlated with the transitions located at positive PC zones. Oppositely, the low temperature transitions usually spread more widely through the parameter spaces, not necessarily to be anti-correlated with the high energy ones. The same conclusion holds for other PCs as well. Therefore, it is not trivial to classify the transitions into different groups. They are all correlated with each other more or less.

5.2 Histogram of gradient (HOG) analysis of moment 0 maps

The histogram of oriented gradients (HOG) is a feature descriptor developed for machine vision to systematically characterize the spectral line observation of atomic and molecular gas (McConnell, R K, 1986; Soler et al., 2019). A feature descriptor is a representation of an image by extracting one or multiple characteristics. Here in the case of HOG, the representation is based on the distribution of local intensity gradients or edge directions. The edge direction is defined to be perpendicular to the direction of the gradient. The HOG method is widely applied in the recognition of objects (e.g. Freeman (1994)), evaluating the morphological changes across velocity channels, while in this work we adopted HOG to compare two extended spectral line images and provides an estimate of their correlation. Now the HOG is a representation of the relative orientations between local gradient orientations in the two images, thus it is a

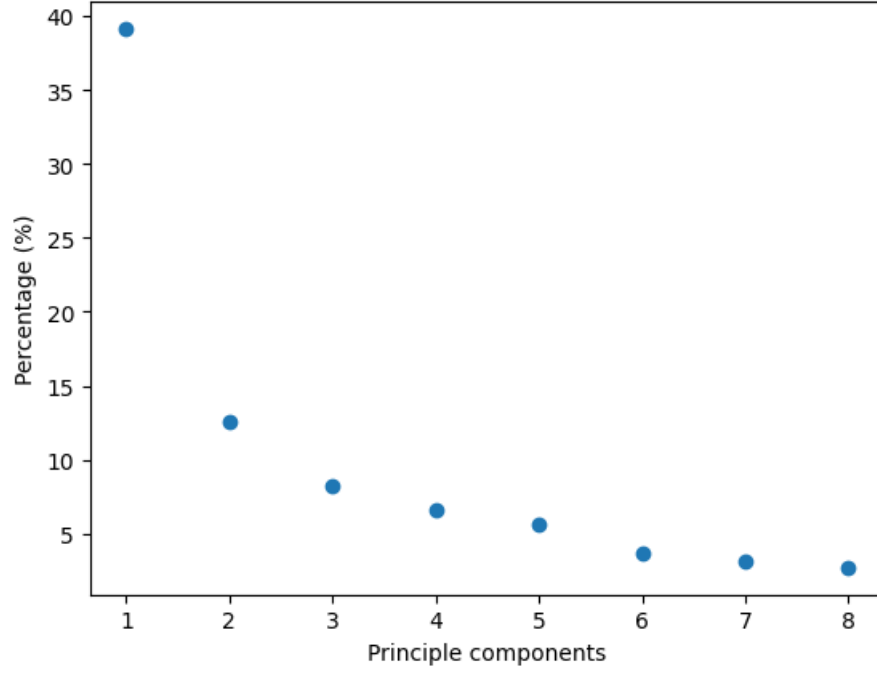


Figure 5.3: The eigenvalue percentage against the component number of the principle components (PCs) in descending order, in which we take the first 8 PCs for further analysis.

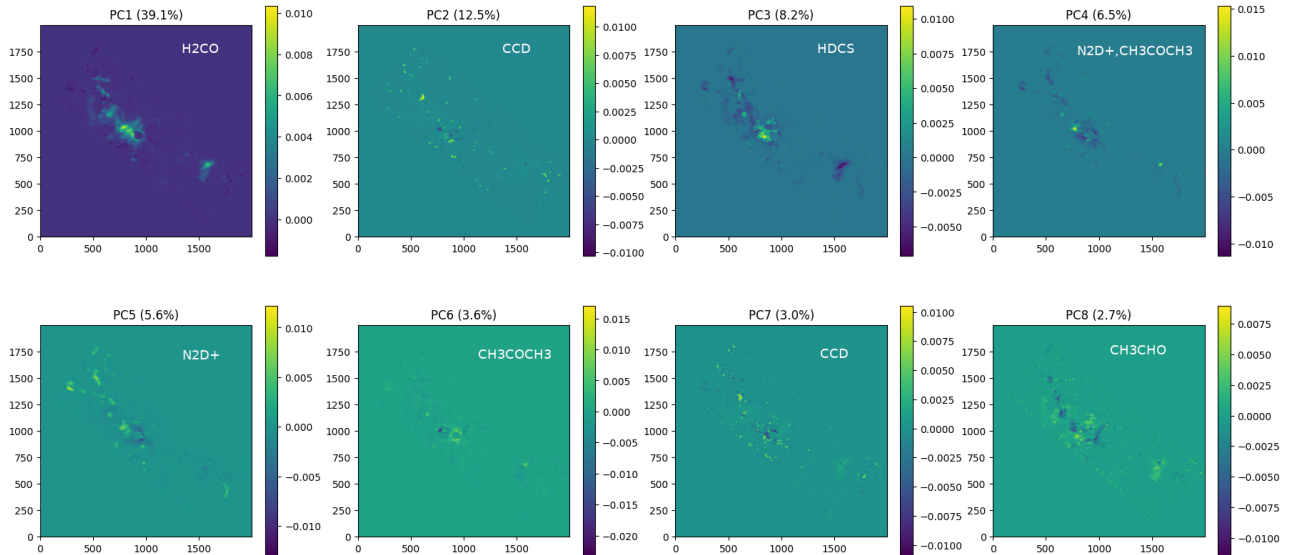


Figure 5.4: Constructed images out of eigenvectors of the first 8 PCs. The images are not normalized to their maximum values.

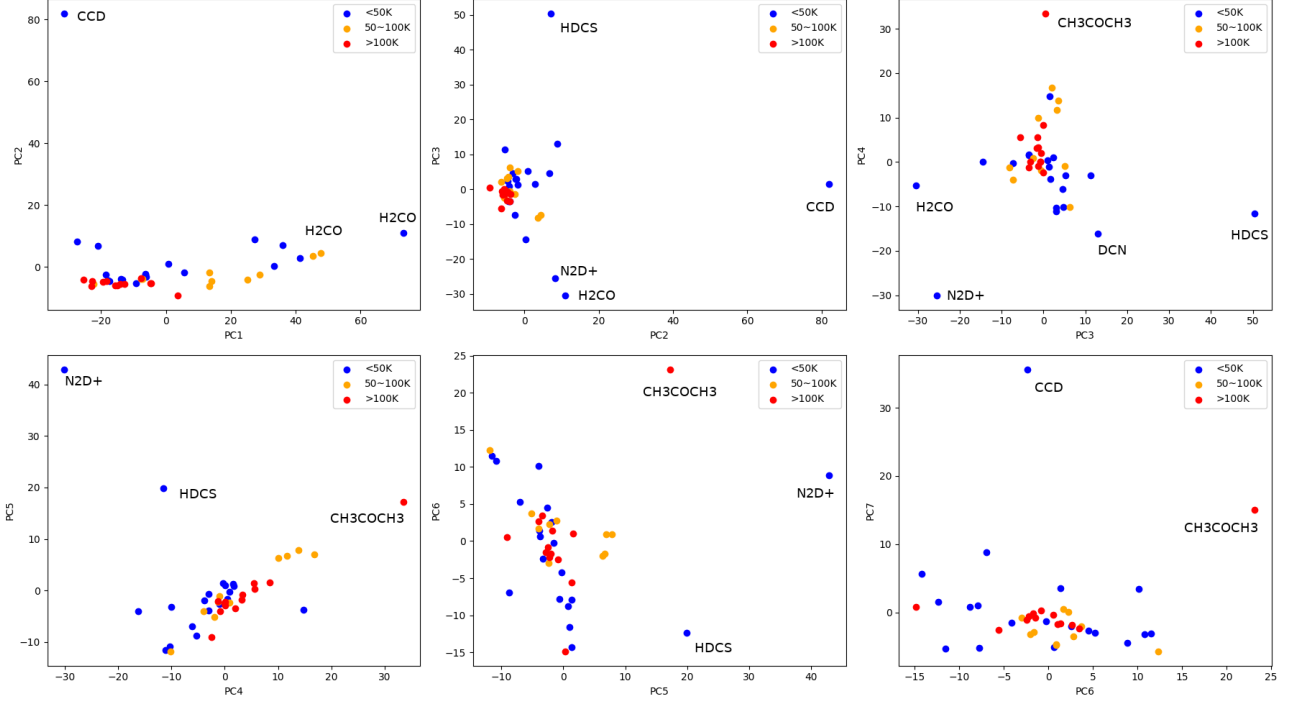


Figure 5.5: Plots of eigenvectors of a certain pair of PCs. Each of the eigenvectors represents the component of that molecule in the relevant PC.

measurement of how the edges in the two images match each other.

5.2.1 Computation of the HOG

We characterized HOG using moment-zero maps of 22 different molecules (including isotopes) as well as the dust continuum emission. The magnitude of the intensity gradient for a pair of adjacent i,j pixels is defined as $|\nabla I_{i,j}|$. $\nabla I_{i,j}$ is calculated by:

$$\nabla I_{i,j} = I_{i,j} \otimes \nabla \left(A \exp \left[-\frac{x^2}{2\sigma_g^2} - \frac{y^2}{2\sigma_g^2} \right] \right), \quad (5.1)$$

where \otimes is the convolution operator, A is a normalization factor to adjust the gaussian integral to unity, σ_g is the gaussian kernel size.

We compute the gradients using Gaussian derivatives, explicitly, by applying the multidimensional Gaussian filter routines in the *gaussian_filter* function of the *Scipy* package. Gaussian derivatives are the result of the convolution of the image with the spatial derivative of a two-dimensional Gaussian function. The width of the Gaussian determines the area of the vicinity over which the gradient is calculated, which is set to be one beam size (20 pixels) to reduce the effect of noise in the pixels as well as retain the signal.

The orientation angle is defined as the gradient direction, written as $\arctan \left(\frac{\nabla I_{i,j} \cdot \hat{z}}{|\nabla I_{i,j}|} \right)$, where \hat{z} is the north direction (90°) and the angle is counterclockwise. Due to the definition of an arc-tangent function, the orientation angle has a range $[-\pi/2, \pi/2]$. However, the gradient should have a direction in the range $[-\pi, \pi]$. Similarly to common practice, we map each angle twice itself, to test for uniformity in a later evaluation stage.

In Fig. 5.6, we present two examples of the moment 0 images, together with the orientation angle map, and the corresponding magnitude of intensity gradients. For areas with weak emission, the orientation angles are less confidential due to the possible noise spikes (left panels). When applied a threshold of 3σ to exclude noisy regions, the angles better trace the intensity gradients, which is clearly visible by the eye (right panels) and can be considered as a large improvement. As for the magnitude of intensity gradients, the central hot core shows the strongest gradients, and the other cores located on the filament are partially visible.

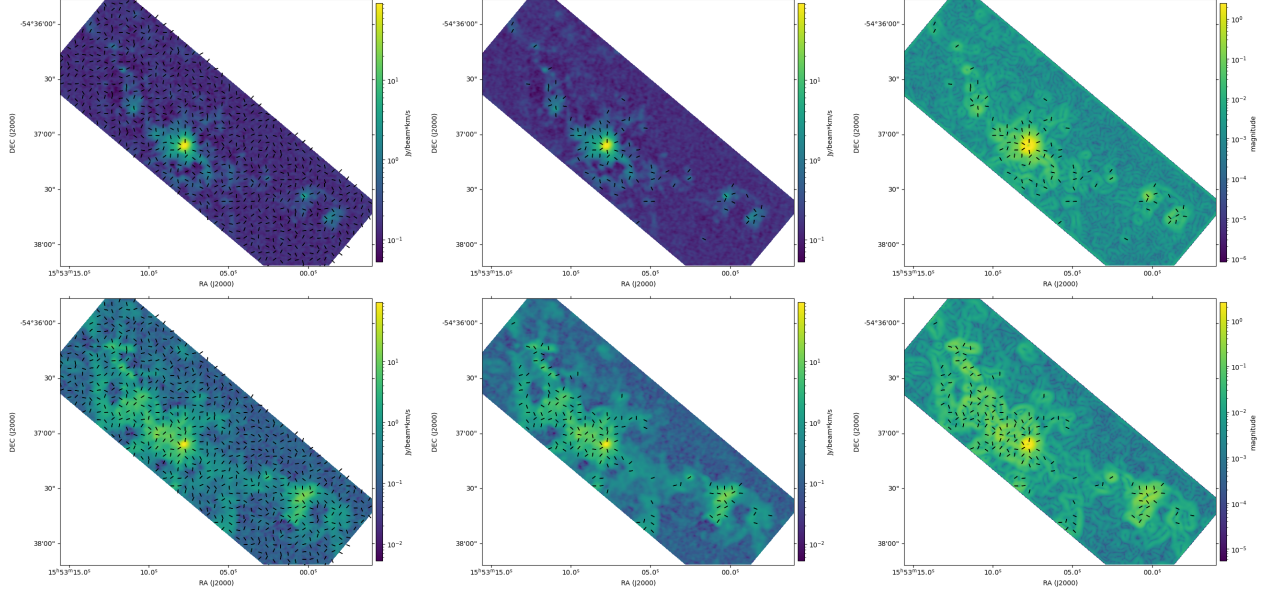


Figure 5.6: Examples of applying HOG method for ^{13}CS (upper panels) and H_2CO (lower panels). Left: the original relative orientation angle map in the step of every 40 pixels, and the corresponding moment 0 image as the background. Middle: the relative orientation angles when applied a threshold of 3σ to exclude the noisy regions. Right: the corresponding magnitude of intensity gradients, dominated by the strong cores on the filament.

Since the H_2CO is more extended than ^{13}CS seen from the moment 0 maps, the relative orientation angles are credible in more areas for H_2CO than for ^{13}CS . Taking into account the common valid regions of two images, we can compare the difference of orientation (defined as **relative orientation angle**) pixel by pixel by subtracting the angle of one image from the other, to examine the similarity of the intensity gradient. Seen from the left panel of Fig. 5.7), the relative orientations (short black lines) are mostly horizontal, i.e. close to zero, especially at the central hot-core region. We plotted a histogram of relative orientation angles to check the distribution, as shown in the right panel of Fig. 5.7. The histogram is not a simple Gaussian, while the highest peak is centered at zero degree. This implies some similarities between the two moment maps and some differences with a preferred direction. Therefore, we need some statistical tools to evaluate the difference of gradients and preferred orientations, presented in the following sections.

Explicitly, we calculated the correlation between two maps (relative orientation angles) by the following steps presenting as mathematical equations.

The two moment 0 maps $I_{i,j}^A$ and $I_{i,j}^B$ are already aligned to the same grid in the imaging process. Indexes i and j correspond to the spatial coordinates, Galactic longitude and latitude. We calculate the relative orientation angle ϕ between intensity gradients following Equation

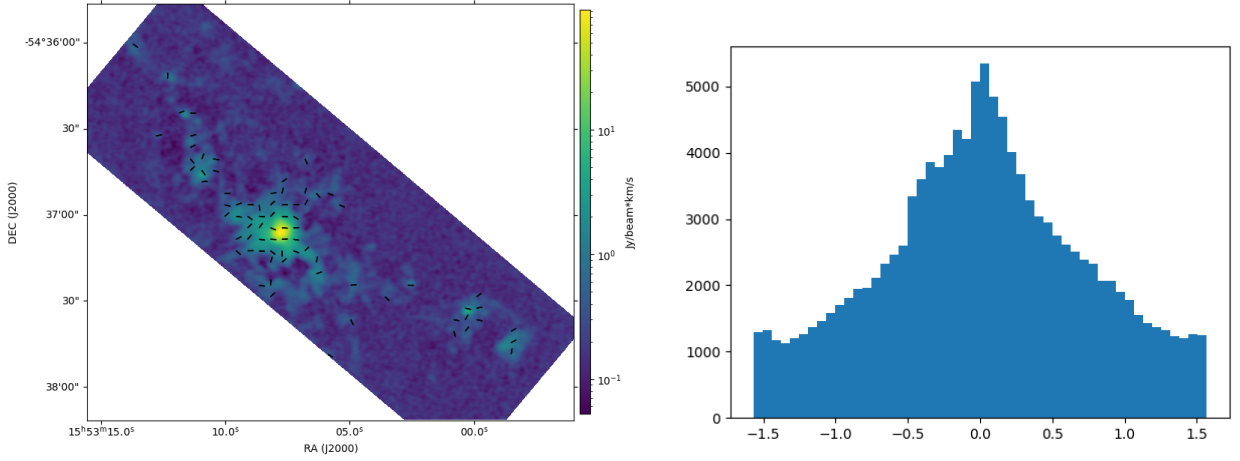


Figure 5.7: Left: the difference of gradient orientation angles of H₂CO and ¹³CS moment 0 maps, under which is the ¹³CS moment 0 map for a comparison. Right: the histogram of the the angle difference is peaked at zero degree but asymmetric, indicating some similarities between the distribution of two molecules.

(1) in Soler et al. (2019) in a two-dimension case:

$$\phi_{i,j} = \arctan \left(\frac{(\nabla I_{i,j}^A \times \nabla I_{i,j}^B) \cdot \hat{z}}{\nabla I_{i,j}^A \cdot \nabla I_{i,j}^B} \right), \quad (5.2)$$

where the differential operator $(\nabla I_{i,j}^A \times \nabla I_{i,j}^B) \cdot \hat{z} = |\nabla I_{i,j}^A| |\nabla I_{i,j}^B| \sin \phi_{i,j}$ corresponds to the gradient. The term is the projection on the z -axis of the cross product. The term $\nabla I_{i,j}^A \cdot \nabla I_{i,j}^B = |\nabla I_{i,j}^A| |\nabla I_{i,j}^B| \cos \phi_{i,j}$ is the scalar product of the vectors. In practice, the use of the Gaussian derivative allows us to vary the width of the Gaussian kernel to reduce the noise in the pixels (Soler et al., 2013). Equation 5.2 implies that the relative orientation angles are in the range $(-\pi/2, \pi/2)$, thus accounting for the orientation of the gradients, not just their direction. The value of ϕ only makes sense when both $\nabla I_{i,j}^A$ and $\nabla I_{i,j}^B$ are significant, for which we used a threshold of three RMS of each corresponding image.

Once we obtain the relative orientation angles $\phi_{i,j}$ for a pair of maps, first we double the angle $\phi_{i,j}$ to $2\phi_{i,j}$. This is a common technique to convert the axial data that carry information of orientation but not direction to circular data that are valid for circular statistics.

Then we summarize the spatial correlation contained in these angles by marginalizing over all the spatial coordinates. Inspired from circular statistics (Jow et al., 2018), we use three tools to evaluate the correlation: **the mean resultant vector** (MRV) and **the projected Rayleigh statistic** (PRS), and alignment measurement (AM).

5.2.2 The mean resultant vector

In this work, we use the definition of the mean resultant vector according to Batchelor et al. (1981):

$$MRV = \sqrt{([\sum_{i,j} \omega_{i,j} \cos(2\phi_{i,j})]^2 + [\sum_{i,j} \omega_{i,j} \sin(2\phi_{i,j})]^2) / \sum_{i,j} \omega_{i,j}}, \quad (5.3)$$

where the indexes i and j run over the pixels in two spatial dimensions and $\omega_{i,j}$ is the statistical weight of each angle $\phi_{i,j}$. The weight is defined as $(\delta x / \lambda)^2$, where δx is the pixel size and λ is

the diameter of the derivative kernel that we used to calculate the gradients. For pixels that are negligible, we set the weight to zero.

The mean resultant vector MRV is a descriptive quantity that can be interpreted as the percentage of vectors pointing in a preferred direction. If all the angles are identical, the MRV equals one. If the MRV is close to zero, then there is no single preferred direction. This may be explained by a phenomenon where all directions are equally distributed, or in other cases of multimodal distribution, e.g., two opposite directions are equally probable. In all the cases, the MRV is between zero and one, which roughly corresponds to the percentage of angles that represent a preferential orientation. We derived an MRV value per pair of moment maps and made the heat map shown in Fig. 5.8. When the MRV is close to zero, it means a uniform distribution of the relative orientation angles. In other words, a whiter color on the heat map means less consistency for the two transitions. Conversely, an MRV close to unity (redder color) means that the relative orientation angles of two transitions tend to be parallel throughout the imaging area. The parallel relative angles are not necessary to be zero degree, therefore we cannot easily determine the similarity for the corresponding moment 0 maps.

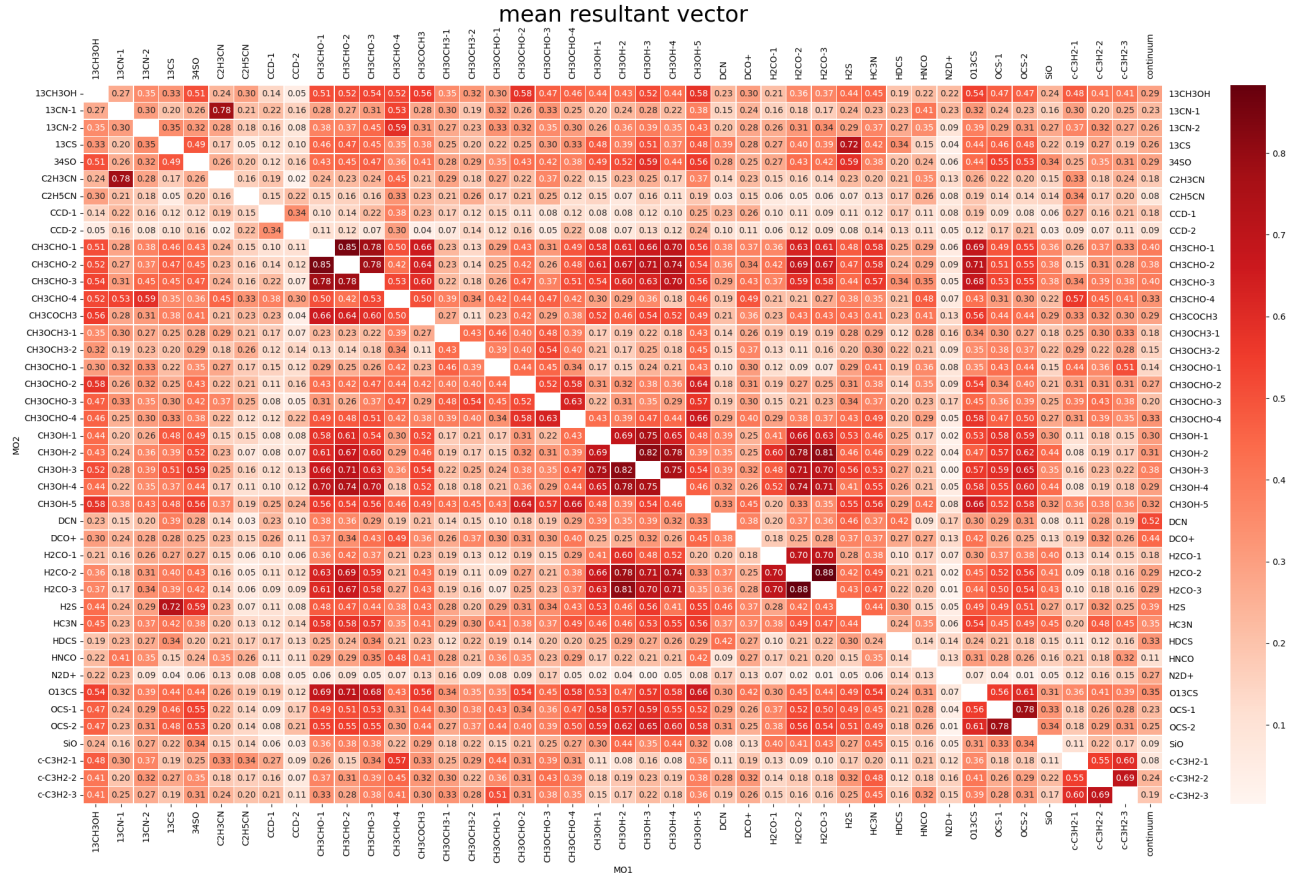


Figure 5.8: The heat map of HOG could not provide any information on the shape of the angle distribution, therefore we consider the **Rayleigh Z test** (h. Zar, 1974). It assumes that the sample is generated from a von Mises distribution, which is a continuous wrapped normal distribution

However, MRV could not provide any information on the shape of the angle distribution, therefore we consider the **Rayleigh Z test** (h. Zar, 1974). It assumes that the sample is generated from a von Mises distribution, which is a continuous wrapped normal distribution

on the circle, in other words, a circular analogue of the normal distribution. The Rayleigh test value Z is defined as the magnitude of the mean resultant length in the most common instance of unspecified mean direction:

$$Z = [MRV^2 \Sigma_k^N \omega_k]^{1/2}. \quad (5.4)$$

The null hypothesis H_0 of the Rayleigh test is that the population is uniformly distributed around the circle. Therefore, the alternative hypothesis H_A is not uniformly distributed. The null hypothesis H_0 is rejected when Z is large.

The value Z can be explained more intuitively by the random walk model. We consider a random walk with N steps, each with a characteristic length ω_k . If the walking angles θ_k are uniformly distributed, the net displacement of this random walk is close to zero, which is the Z value defined above. If the angles are not uniformly distributed, the net displacement is larger than zero. The values of Z (shown in Fig. 5.9) clearly show that many of the image pairs pass the non-uniformity test in redder color, and the corresponding MRV values are also significant. Most of these large Z values are located along the diagonal of the plot, representing the different transitions of the same molecule, consequently proving that our method of calculation is correct. Z values are useful to evaluate if there is a preferential orientation in a distribution of angles, but given that this angle can be different than zero degree, it is not adequately informative to draw a conclusion.

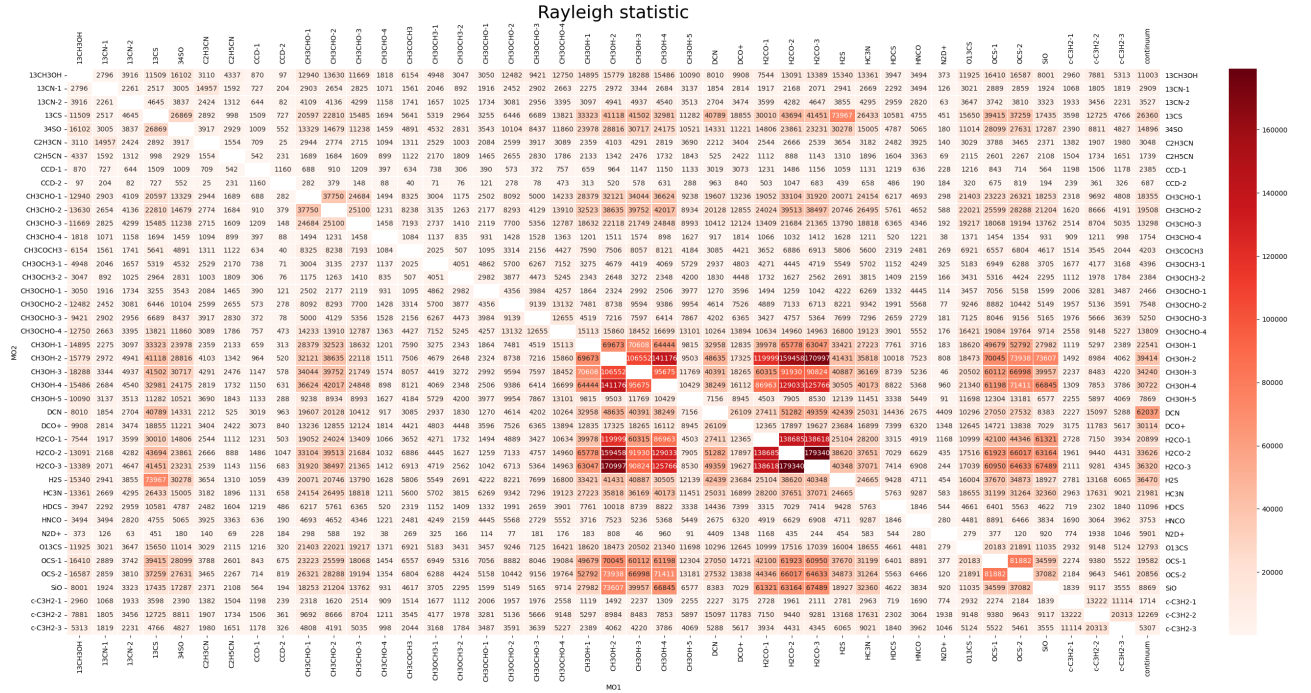


Figure 5.9: The heat map of Rayleigh test Z values between the relative orientation angles of every two transitions. Z value close to zero (whiter color) means random distribution of the relative orientation angle, hence less consistency for the two transitions. A higher Z value (redder color) means a non-uniform distribution and more similarity for the two transitions.

5.2.3 The projected Rayleigh statistic

In the HOG application, we want to know not only the uniformity, but also the preferential mean orientation. The optimal statistic to evaluate the distribution of angles is non-uniform and

peaked at zero degree is the projected Rayleigh statistic (PRS, [Jow et al. \(2018\)](#)), also known as the V statistic:

$$V = \frac{\sum_{i,j} \omega_{i,j} \cos(2\phi_{i,j})}{\sqrt{\sum_{i,j} [(\omega_{i,j})^2/2]}}, \quad (5.5)$$

which follows the same conventions mentioned previously.

Each value V represents the likelihood of a von Mises distribution, or in other words, the likelihood that two emission maps have a similar difference in their gradients in various pixels. The null hypothesis is a random distribution; therefore, the samples are independent and in uniform circular distribution. The alternative hypothesis is assumed to have a main direction, e.g, parallel (0°) in our case.

Now we can use the V values from the projected Rayleigh statistic to test whether the relative orientation is preferentially parallel, which corresponds to $\phi=0$, and what is the statistical strength of that trend. We derived one V value per pair of moment maps and made a heat map shown in Fig. 5.10. The resulting $V>0$ indicates a mostly parallel relative orientation between the gradient vectors, while $V\leq 0$ corresponds to mainly perpendicular relative orientations, i.e., no particular correlation when comparing two images. A larger V indicates greater similarity for two images. This test has been proved to be practical when testing the correlation between column density structures and the magnetic field ([Soler et al., 2017](#)).

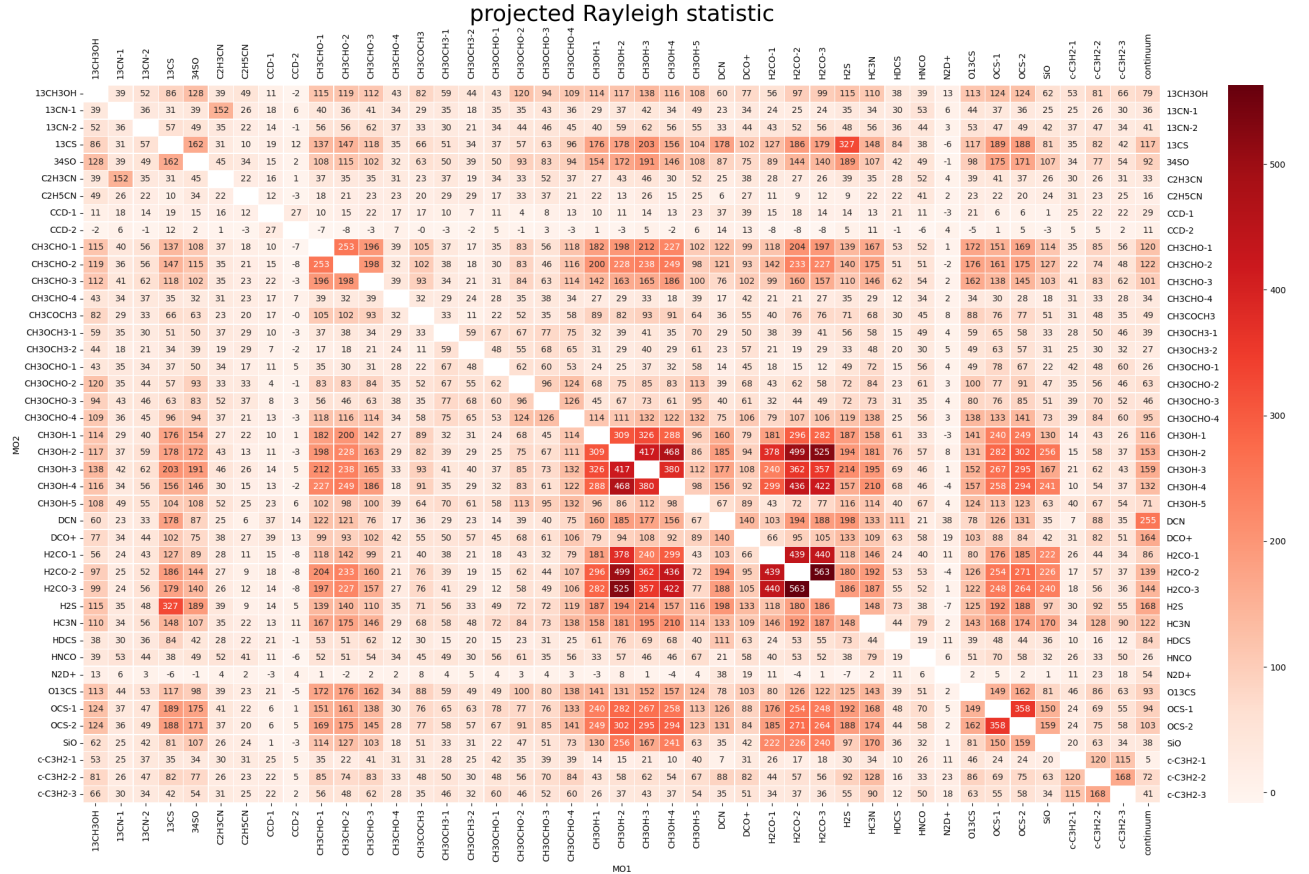


Figure 5.10: The heat map of projected Rayleigh test V values between the relative orientation angles of every two transitions. Small or even negative V value (whiter color) means perpendicular relative orientation angle or random distribution of the relative orientation angle. A higher V value (redder color) refers to preferentially parallel and a peak at zero degree, hence more similarity for the two transitions.

In the large N limit of the V distribution, the asymptotic limit is the standard normal distribution. Then we define the variance in V simply as the variance of each $\cos(2\phi)$:

$$\sigma_V = \frac{2\sum_k^N (\cos 2\phi_k)^2 - V^2}{N}. \quad (5.6)$$

We derived the variance for each pair of transitions, shown in Fig. 5.11. The lower variances (whiter color) indicate a less extended distribution of the V values, which corresponds to the peak at zero degree.

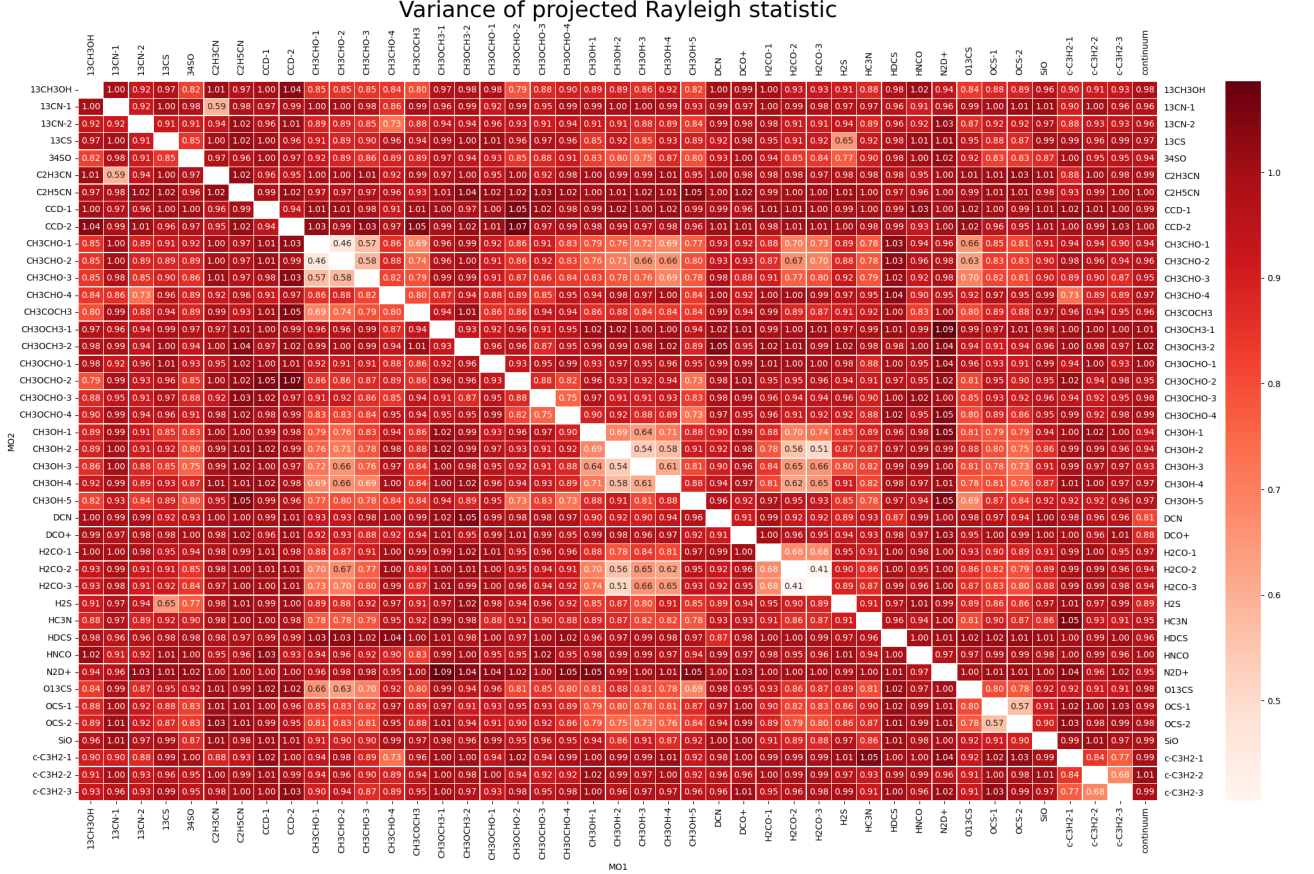


Figure 5.11: The heat map of the variance of projected Rayleigh test V values between the relative orientation angles of every two transitions. The whiter color refers to better confined V values, i.e., peak at zero degree.

5.2.4 The alignment measurement

We also calculated the alignment measurement (AM), an alternative method to estimate the degree of alignment between vectors. This quantity is defined as:

$$AM = \langle 2\cos\phi_{i,j} - 1 \rangle_{i,j}, \quad (5.7)$$

where $\langle \dots \rangle_{i,j}$ represents the average over the relative orientation angles. We derived the AM values for each pair of transitions, shown in Fig. 5.12. The values of AM should range from -1 to 1 (can be narrower in a real test), corresponding to perpendicular to parallel accordingly. We can immediately see that AM has a lower dynamic range than V, so the contrast between regions with low- and high-correlation is much lower as well. Alignment measurement is widely used in the study of dust grain alignment with magnetic fields (e.g. Lazarian and Hoang (2007)).

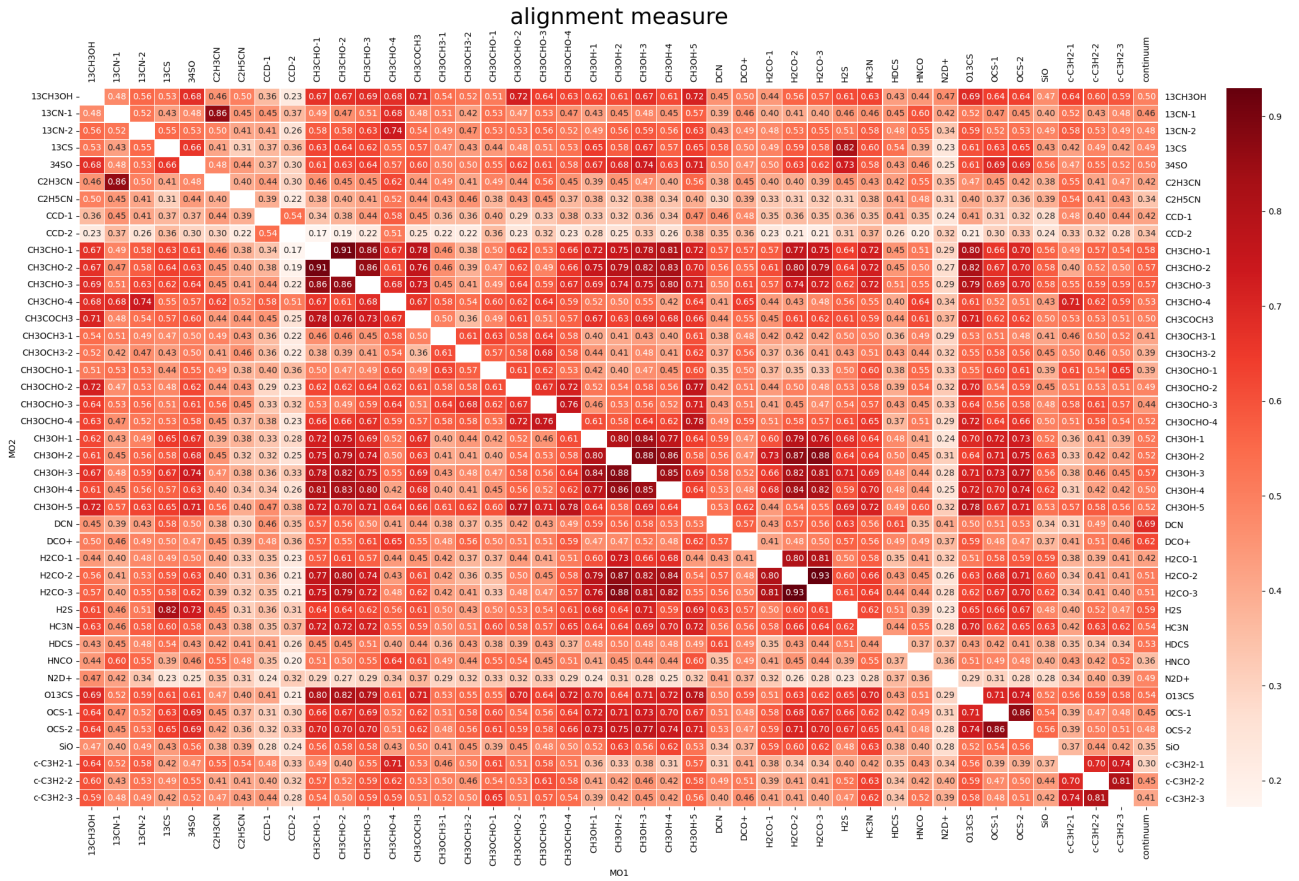


Figure 5.12: The heat map of the alignment measurement values between the relative orientation angles of every two transitions. The AM values close to 1 (redder color) is corresponding to a zero-degree peaked parallel distribution. A smaller AM value tends to be the hint of a random distribution.

5.2.5 Correlation between different molecules

In the previous sections, we present the results of the HOG analysis in terms of MRV, Z, V, σ , and AM. The values of MRV, Z, and AM are only meaningful for our purposes when they are validated by V values, since large values of the mean resultant vector and the Rayleigh statistic only indicate a preferential orientation, not necessarily $\phi=0^\circ$. Therefore, we should assess the relation based on the combined information of V and MRV, rather than drawing conclusions from the values of MRV alone.

In this work, the interpretation of astroHOG is first examined in terms of the projected Rayleigh statistic V, which is sensitive to the number of gradients that are parallel, thus indicating that the gradients of two images match. Taken the last column in Fig. 5.10 showing the V values between the continuum and different transitions, we can see that DCN has the highest V value of 255, followed by H₂S of 168, DCO⁺ of 164. Not surprisingly, the lower energy-state CH₃OH also presents high V values, while the highest energy state is below 100 due to its compact emission compared to the continuum. Similar for CH₃CHO. For H₂CO lines, an interesting thing is that the V values increase with the excitation energy. This can be explained by too much extended emission (which can be affected by outflow) in the H₂CO line with lower excitation energy, hence less correlated with the continuum image. The remaining molecules showing V values greater than 100 are HC₃N, OCS, and ¹³CS. Many of the mentioned molecules are well-known dense core tracers, which certifies our result. Regarding the molecules with the

lowest continuum correlated, they are $c\text{-C}_3\text{H}_2$, CCD, $\text{C}_2\text{H}_5\text{CN}$, and HNCO with V values below 30, which will not be considered in the further analysis of the MRV and AM statistics.

When we look into the V statistics between different transitions, the standing out values in the heat map are those of the same molecule and similar energy states, e.g., H_2CO lines at 218476 and 218760 MHz with the same $E_{\text{up}}=68$ K, OCS lines at 218903 MHz of 100 K and 231061 MHz of 111 K, since the close energy states can be interpreted by similar formation environment and condition. For the same reason, different molecules having similar energy states also show high V values, especially between the CH_3OH line ($E_{\text{up}}=45, 55, 89\text{K}$) and H_2CO lines ($E_{\text{up}}=68$ K) with V values above 400, indicating a strong correlation between these molecules. The other large V values above 200 happen between CH_3OH and ^{13}CS , CH_3OH and CH_3CHO , CH_3OH and H_2CO , CH_3OH and H_2S , CH_3OH and HC_3N , CH_3OH and OCS , CH_3OH and SiO , H_2CO and CH_3CHO , H_2CO and SiO , H_2CO and OCS , H_2S and ^{13}CS . This tells us some intrinsic correlations between the most popular molecules in the cores, e.g., H_2CO and CH_3OH might be highly affected by the outflow, since it traces components similar to SiO , while the high-energy-state CH_3OH traces the dense gas cores as HC_3N does. We consider the V values above 100 to be the significant limit for further analysis. Furthermore, there are some V values close to zero or even below zero between N_2D^+ , CCD and other molecules, indicating a completely different physical condition to produce these two molecules. It is also worth to mention that N_2D^+ and CCD are in better, though still weak, correlation with other deuterated molecules, including DCN and DCO^+ .

For the variance of projected Rayleigh statistic (Fig. 5.11), the lower values are written in black color (<0.67). We notice some low variance, and hence similar orientation angles, besides the diagonal pixels. They are the molecular pairs of $\text{C}_2\text{H}_3\text{CN}$ and ^{13}CN , H_2S and ^{13}CS , CH_3CHO and O^{13}CS , CH_3CHO and H_2CO , CH_3CHO and CH_3OH , CH_3OH and H_2CO , four of which are also selected in V values above 200. The high correlation between $\text{C}_2\text{H}_3\text{CN}$ and ^{13}CN , H_2S and ^{13}CS may originate from the same element in each pair, here Nitrogen and Sulfur, respectively.

Then we analyze the MRV heat map. Taking into account the continuum, it is best correlated with DCN ($\text{MRV} = 0.52$), the same as the conclusion from the V statistics. The other highly correlated molecules ($\text{MRV}>0.37$) are DCO^+ , CH_3CHO , H_2S , and CH_3OH , which is also consistent with the result of the V statistics. Since the MRV values have a small range of 0.08–0.52, it is hard to draw a limit to distinguish the uniform angle case from the parallel one. Hence we just tested if the MRV values show a similar tendency as the V values.

Regarding the relationship between different transitions, the peak MRV values (>0.69) are located between the different transitions of the same molecule, including CH_3CHO , CH_3OH , H_2CO , OCS , $c\text{-C}_3\text{H}_2$. There is no significant clue to correlate N_2D^+ with any other molecules mentioned. Interestingly, the MRV values between the molecules are usually larger than those between the continuum and a certain molecule, which emphasizes the importance of continuum study as well as line research, since each has unique properties.

The pattern of Z values from the Rayleigh statistic is extremely similar to the V values, except that the Z values are around 100 times larger than the V values. In general, the MRV and Z should be high when V is high, but the opposite is not always the case, which is the same as what we found from the heat map results.

The alignment measure values are unexpectedly all above zero, whose mathematical range is $[-1,1]$. This can be interpreted as that there exists some similarity between each pair of

transitions rather than being completely inconsistent. The AM values for the continuum also peak at DCN, followed by DCO⁺. And similar conclusions hold as for the V statistics.

Now we discuss the limits and benefits of the HOG method. There are three main advantages. First, we do not need to match the scales of the images or assume a priori towards the shape of the sources that we observed. Second, it does not require a sharp edge of the molecular emission image, but can systematically characterize and correlate the intensity contours including filaments and clumps without visual inspection. Third, we do not assume any physical parameters or models to examine the correlations. The limitations are obvious: it requires a suitable Gaussian scale size to estimate the orientations, while it can be difficult to explain the local relative orientations due to resolution and fluctuation. In addition, the statistical values are hard to interpret quantitatively, e.g., defining a threshold of statistical parameters to distinguish a high-correlation case from a low one. To draw a conclusion one should be extremely careful.

5.3 Velocity distribution

In order to understand the kinematics of G327, we have used molecular line velocity maps. DCN usually traces extended emission from the molecular cloud as well as the kinematic motion. It shows a clear peak through the whole region; hence a Gaussian profile is fitted to the spectrum of each pixel in the observed region. The central velocity of the Gaussian fitting is adopted as the radial velocity of the cloud at a certain pixel. When plotting the distribution of the central velocity in Figure 5.13, a large-scale velocity gradient is identified: the region towards the north-east is relatively blue shifted, whereas the central and south-west regions of the filament are relatively red shifted. The region close to the hot core (marked by a star) possesses a central velocity around zero. The whole scenario indicates a possible global infall from the filament to the central hot core, which is also shown by the high-J HCN lines observed by Rolffs et al. (2011).

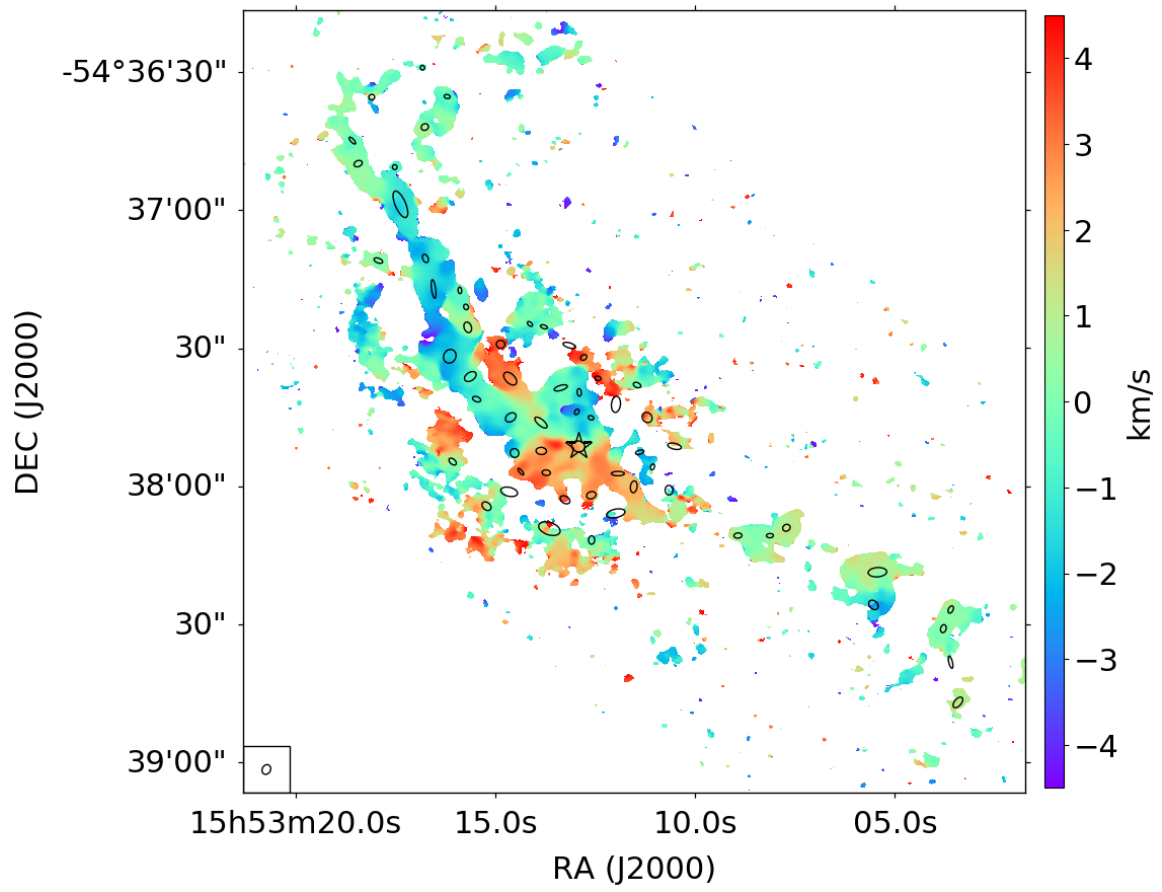


Figure 5.13: Central velocity distribution of DCN. Cores found by dendrogram are over plotted to trace the filament. The hot core is marked by a star.

6 Star formation in the infrared

Infrared dark clouds (IRDCs) are characterized by extinction features seen against the bright background, which are dense ($n > 10^5 \text{ cm}^{-3}$), cold ($T < 20 \text{ K}$), and bright at submillimeter wavelengths (Carey et al., 2000). These extreme properties make them the progenitors of massive stars and clusters (Chambers et al., 2009). Possible clues of the earliest evolutionary stages of star formation can be found within the IRDC, including young stellar objects (YSOs), ultracompact HII regions, and outflows. Some IRDCs are lacking any star formation activity, which serve as good targets for the sub-millimeter to study the earliest phases prior to collapse (Busquet et al., 2016). Therefore, the infrared observations can help us to categorize the star formation into different evolutionary states, in supplement to the ALMA data we have.

We collected infrared data from the Spitzer IRAC four bands, the GLIMPSE I Spring 07 Catalogue (highly reliable), the Wide-Field Infrared Survey Explorer (WISE) point source catalogue, the Herschel/PACS point source catalogue, and the VVV survey catalogue.

Firstly, we plotted the Spitzer three-color image of source G327, shown in Fig. 6.1. As a common practice, the IRAC 1, 2, 4 bands (corresponding to 3.6, 4.5 and 8 μm emissions) are coded in blue, green, and red, respectively. The identified continuum cores by ALMA 1.3 mm observations are overlaid on top to check the correspondence. The north-eastern sub-mm filament is located in a relatively dark cloud, which shows some scattered point sources corresponding to the continuum cores. The south-western section of the filament, on the other hand, is located on a fairly bright background and accompanied by some infrared emission blobs. Hence we understand the reason for having more cores in the north-eastern part compared to the south-western side: There is more extinction as infrared dark at the north-east, thus more material and mass to form the cores.

To check the star formation activity throughout the filament, we plotted all the point sources found in the WISE catalogue, shown as black circles in Fig. 6.1. We found that only three cores have corresponding WISE point sources, whereas the WISE positions are offset from the peak of Spitzer emission. This phenomenon is common between the WISE point sources and the Spitzer image in our case, possibly because of different resolutions and wavelengths of the two telescopes. Therefore, we also checked the point sources from the GLIMPSE I Spring 07 Catalogue (highly reliable), plotted as red circles in Fig. 6.1. We found that there are 16 GLIMPSE sources associated with 13 cores, and some cores even include two GLIMPSE sources. Only one of the three WISE sources is covered by the GLIMPSE catalogue at the same time, though the other two WISE sources have corresponding GLIMPSE points within a radius of 1". Since a young star is more visible at a longer wavelength, they are usually red or green in Spitzer images. Red sources usually show MIR spectra dominated by PAH emission and sometimes forbidden lines of a massive YSO close to the main sequence (Simpson et al., 2014). The point sources that are close to blue are usually foreground stars or highly evolving stars, which are excluded from the further analysis. Yellow sources combined by green and red are considered to be YSOs with icy envelopes, since the central protostar has not evolved to a temperature high enough to evaporate the ice or excite PAHs (Simpson et al., 2014). There are no correlated point sources from the Herschel catalogue or VVV survey.

Note that the RGB images in Fig. 6.2 and 6.1 are from the same data set, which are produced by different tools and stretch of color, which may affect visual inspection. The SAOImage-ds9 tool provides a best high-contrast image (lower panel of 6.1), while it requires hand adjustment rather than a script and is less efficient in adding marks and labels.

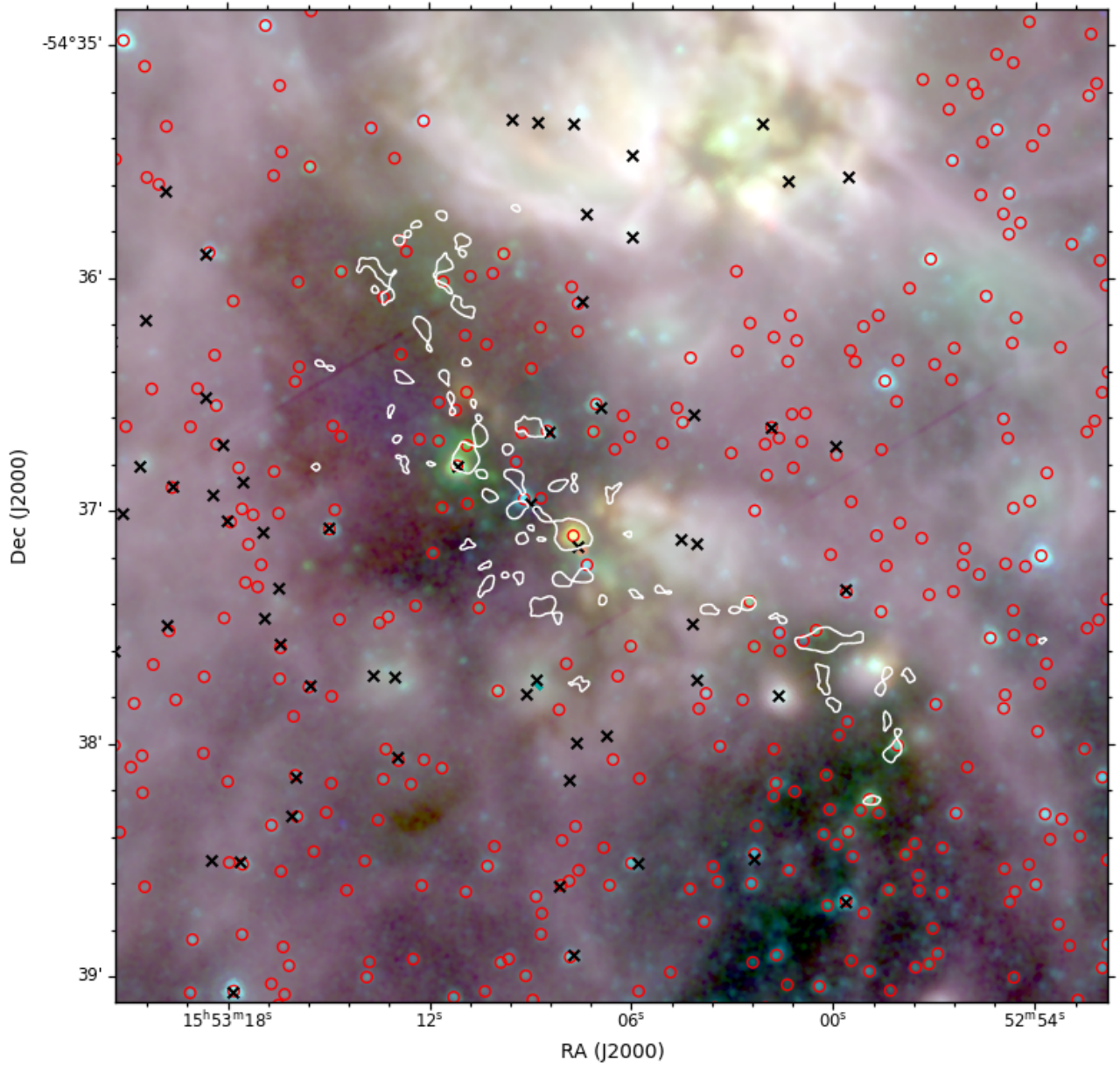


Figure 6.1: Spitzer three-color composite image with the 3.6, 4.5 and 8 μm emissions coded in blue, green and red, respectively, which is generated by python package APLpy `make_rgb_cube` function in log scale. The white contours are the identified continuum cores at 1.3 mm. The red circles are the potential stars found in the GLIMPSE I Spring 07 highly reliable Catalogue, and the black crosses are from the WISE point sources catalogue.

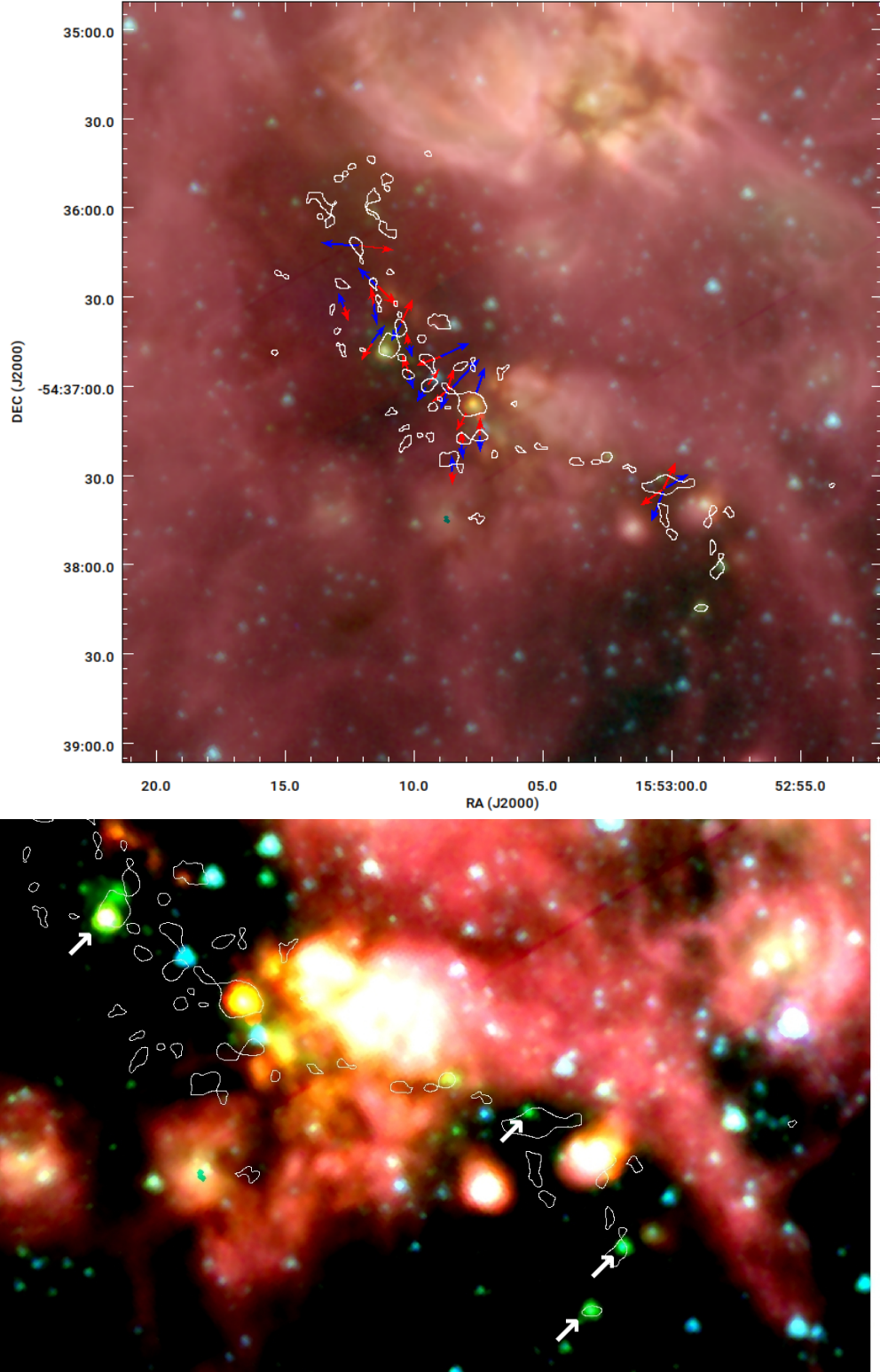


Figure 6.2: Upper panel: Spitzer three-color composite image with the 3.6, 4.5 and 8 μm emissions coded in blue, green and red, respectively, produced by the Python package `astropy.visualization` and the `make_lupton_rgb` function within. The white contours are the identified continuum cores at 1.3 mm. The red and blue arrows shows the red/blue-shifted outflow direction traced by SiO emission. Bottom panel: the zoom-in Spitzer three-color image generated by the `SAOImage-ds9` RGB tool, in which the EGOs are green and marked with white arrows, clearly shows the overlap of EGO with the sub-mm cores.

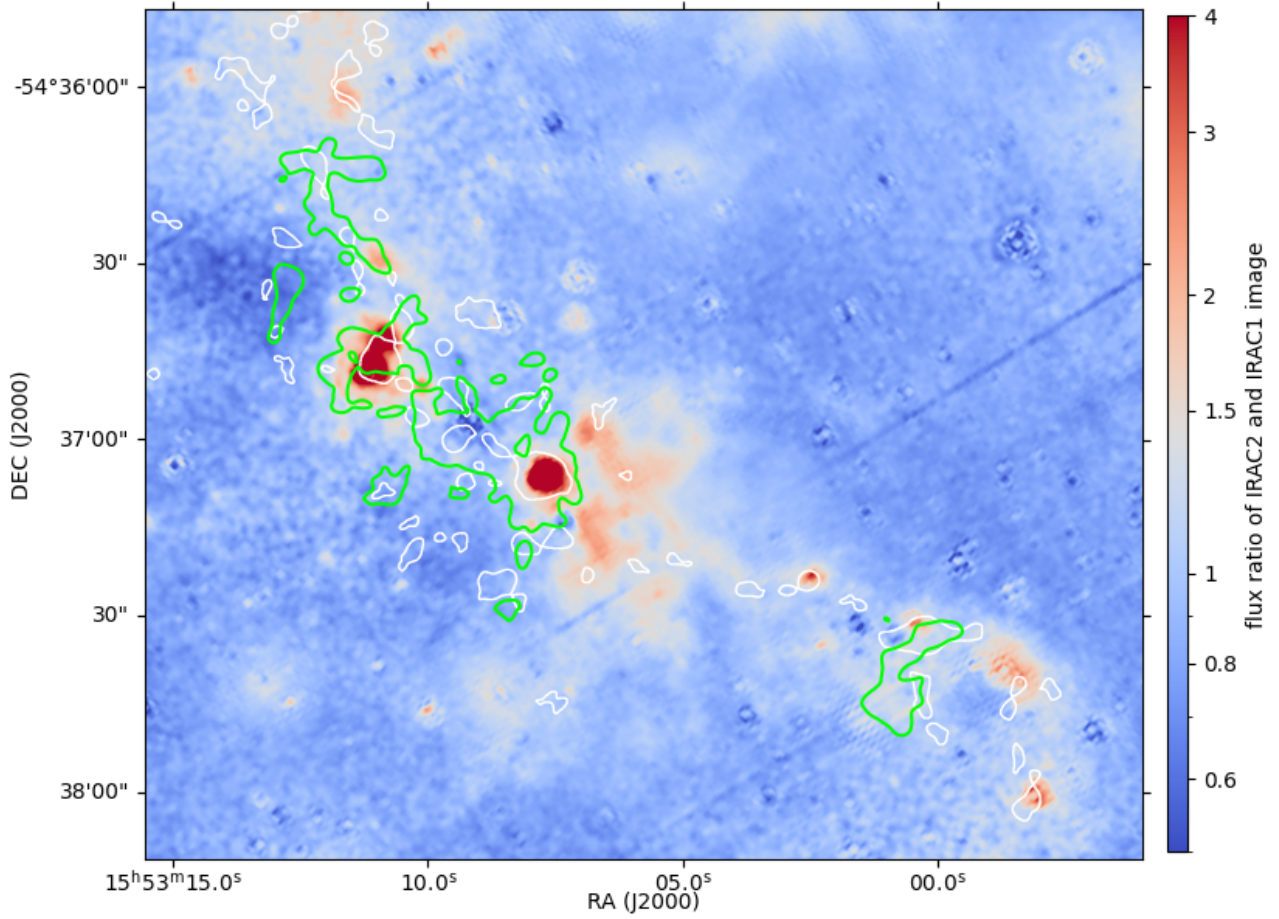


Figure 6.3: The flux ratio of IRAC 2 over IRAC 1 image. The continuum cores are shown in white contour, and the SiO moment 0 map is shown in green contour. A flux ratio exceeding 1.5 (in red color) refers to higher possibility of outflow and shocks, e.g., the regions surrounding the hot core are very likely to be affected by outflow.

6.1 Extended green objects (EGO)

Due to this common color-coding of the $4.5\ \mu\text{m}$ band as green, sources with extended emission features in this band are commonly considered as *extended green objects* (EGOs) Cyganowski et al. (2008), or *green fuzzies* (Chambers et al., 2009). The $4.5\ \mu\text{m}$ enhancement is associated with the presence of shock-excited H₂ ($v = 0-0$, S (9,10,11)) lines and/or CO ($v = 1-0$) band heads according to Cyganowski et al. (2008). There are also hypotheses stating that EGOs correspond to massive young stellar objects (MYSOs) at the earliest evolutionary stage, with ongoing outflow activity and active rapid accretion (Chen et al., 2013).

If we look into the ALMA submillimeter continuum cores as white contours on the Spitzer infrared image (shown in the bottom zoom-in image of Fig. 6.2, there are at least four cores showing coherent EGOs (marked with white arrows), in which two of them show significant SiO outflow at the same time. These two cores are also the most luminous cores throughout the filament. We did not find these EGOs in other catalogues, e.g., Cyganowski et al. (2008), therefore, we claim these four sources as new detection. In the future, we will perform a proper color-color diagram analysis following e.g. Kuhn et al. (2021) and based on that to identify young stellar objects and assign their class by the SED index. We will also include MIPS GAL 24 μm data to obtain a more robust sample of YSOs to associate with our cores.

In addition to that, we also derived the flux ratio map of IRAC band 2 and 1 following (Veena et al., 2018). It has been found that the $[4.5]/[3.6]$ ratio is around 1.5 or larger for jets and outflows, whereas it is lower for stellar sources (Takami et al., 2010). In our Fig. 6.3, the excess $[4.5]/[3.6]$ ratio (>1.5) are shown in red color, and a higher ratio refers to a higher possibility of being affected by outflow or shock. We notice that the two cores mentioned above again stand out in the flux ratio map (>4), while many other high-ratio positions are not correlated with any cores. We also show the SiO moment 0 contour on top of the flux ratio map and confirm that the two cores are covered by SiO emission as well. Hence we can safely draw the conclusion that the hot core is in a much more evolved stage while the outflow is still strong, and the other massive core is at an early accretion stage with outflow activity.

6.1.1 Possible PDR

Interestingly, we note that the moment 0 map of a cyclic molecule $c\text{-C}_3\text{H}_2$ shows an elongated bow-shock feature that fits with the north-eastern edge of the infrared blobs, as seen in Fig. 6.4. Rizzo et al. (2005) detected $c\text{-C}_3\text{H}_2$ toward the edge of an ultracompact HII region of Mon R2, which is the interface between the ionized gas and the dense cold envelope. This interface is chemically identified as a Photo-dissociation Region (PDR). Thus, $c\text{-C}_3\text{H}_2$ is treated as a PDR tracer.

One possible scenario in our case is that the infrared blob is generated from the HII region, which compresses the gas to the shock front and creates the PDR region. In this case, a new question arises of where is the power source of the extended infrared blob, which is not obvious from the known data. We have searched for existing radio continuum emission for G327 and there is no detection with ATCA 3/6 cm data, as seen in Fig. 6.5. Considering the intensity of RRL at 86 GHz is around 0.04 K, the RRL intensity is anti-proportional to frequency, and the strength ratio of RRL over radio continuum is around 0.1, we get the centimeter continuum should peak at around 4 K. This explains why no radio source was detected in the ATCA 3cm/6cm observation, since the noise level of around 7K far exceeds the signal. In 2022 we

proposed a new ATCA observation to prove the existence of the HII region, and hence a PDR associated with the IR blob. In addition, we are going to examine the Spitzer 24 μm and Herschel 70 μm maps, since the bow shock heated dust emission often appears in these bands, too.

An alternative scenario is that the infrared emission comes from another cloud that is not at the same distance as G327, which cannot be excluded with the known data.

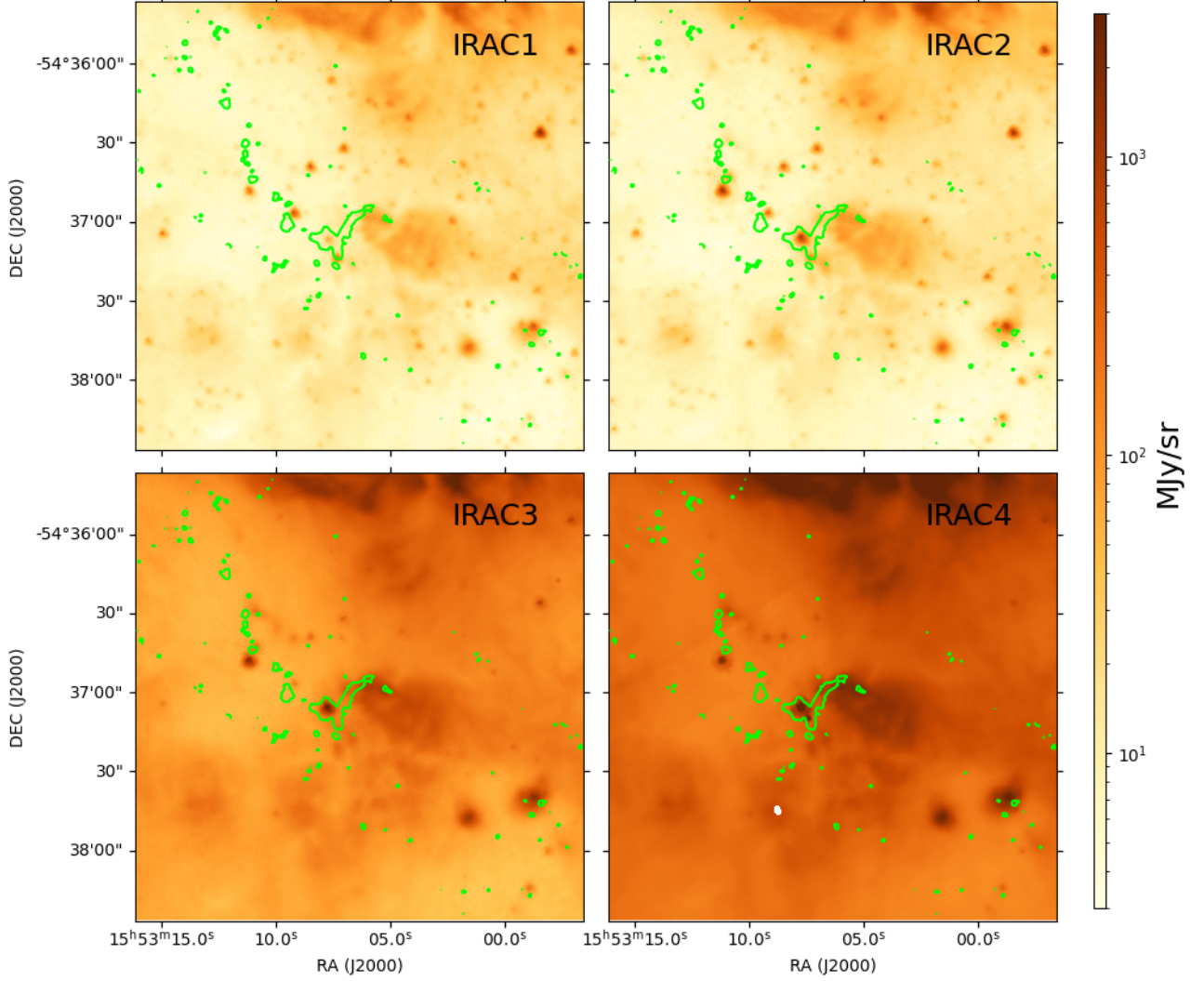


Figure 6.4: The moment 0 map of c-C₃H₂ 216279 GHz line as green contour, on top of the IRAC four bands images. The contour shows a bow-shock feature that fits with the north-eastern edge of the infrared blobs appeared in all the IRAC bands.

6.2 Hot core and a possible face-on disk

There are several other observations that focus on the G327 hot core, not only in submillimeter, but also in mid-IR. We obtained the data from both the ALMA archive and from our collaborators, which are shown in Fig. 6.6. The 0.87mm continuum (orange contour, derived from ALMA Project 2018.1.01449.S) shows a circular core with a smooth intensity gradient from the center to the edge. It has a resolution of 1", half of what we have in band 6 (2"). Luckily, the G327 hot core was covered by the ALMAIMF large program with even higher resolution. The 1mm continuum (0.53", color background in Fig. 6.6) shows an elongated core in the

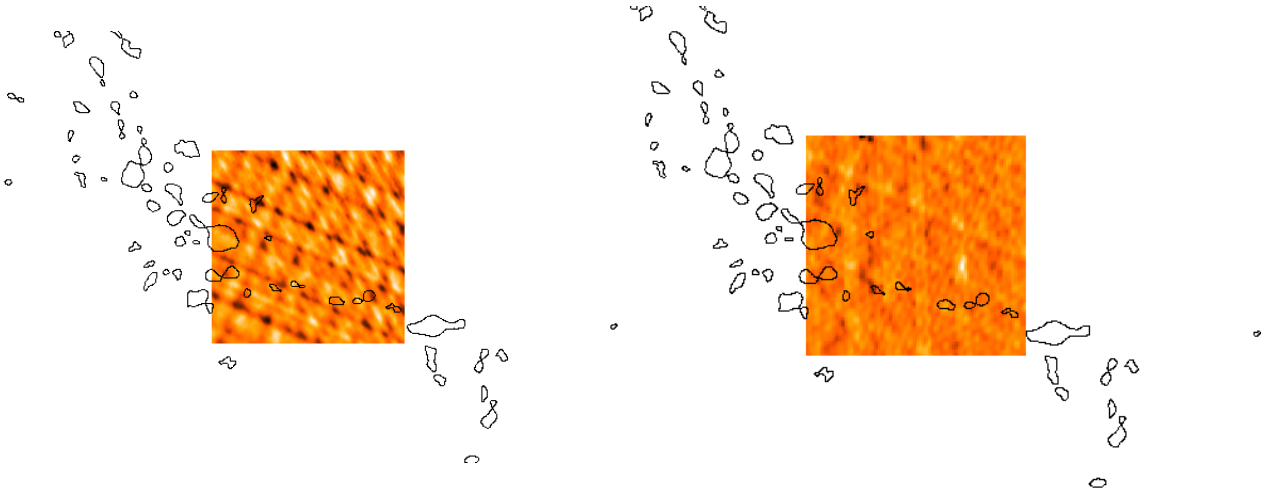


Figure 6.5: Left: 3cm ATCA observation with no detection, overlaid with the identified sub-mm continuum cores. Right: 6cm ATCA observation also without detection.

west-east direction, whose peak is east-shifted by $0.2''$ with respect to the 0.87mm peak. This core is resolved into two separate cores in the 3mm continuum ($0.4''$, cyan contour), indicating core fragmentation on a physical scale of 2000 au ($\sim 0.6''$). This fragmentation is seen not only because of the higher proposed resolution, but also due to the dust opacity decreasing with wavelength. Thus 3mm emission is optically thinner than 1mm and 0.87mm emission and shows more details like sub-cores and disks behind the heavy envelop of dust.

However, we don't have a consistent resolution of these observations at different wavelengths; otherwise we would apply a modified black-body function to fit the spectral energy distribution (SED), hence determining the respective dust opacity and spectral index. In addition, with only three data points that spread in a narrow wavelength range and all locate at the Rayleigh-Jeans approximation side, i.e. the low-frequency side, the SED fitting will have a large error bar for the temperature. As suggested by [Kim et al. \(2019\)](#), we need ALMA band 9/10 high-resolution data to observe the optically thick dust and get closer to the turning-over point of SED; therefore, it will fit the whole modified-black-body function instead of the R-J approximation and give a better limit to the dust properties within the hot core. We have proposed new observations in ALMA Cycle 9/10, which have been partially observed. We will analyze these results in the future.

6.2.1 VLT PAH continuum

By comparing the high-resolution continuum in submillimeter and infrared, we can look deeper into the hot core and analyze the evolutionary stages. One recent high-resolution infrared observation was done by the Very Large Telescope (VLT) in Chile ([Elbe et al. in preparation](#)). The VLT spectrometer and imager for the mid-infrared (VISIR) provides two continuum image filters centered on polycyclic aromatic hydrocarbons (PAH), named PAH1 (at $8.7\text{ }\mu\text{m}$, $0.25''$) and PAH2 (at $11.3\text{ }\mu\text{m}$, $0.33''$). [Elbe et al. \(in prep\)](#) applied the Lucy-Richardson deconvolution method towards the continuum image, using a PSF reference to further enhance the image resolution following [Kervella et al. \(2011\)](#). The preliminary results are shown in [Fig. 6.7](#) with multiple hot spots in both PAH1 and PAH2 images, indicating possible core/disk fragmentation. We need to be more careful when comparing the VLT and ALMA images, since the coordinates in the VLT image require further examination. Plotting the PAH2 image ($0.33''$)

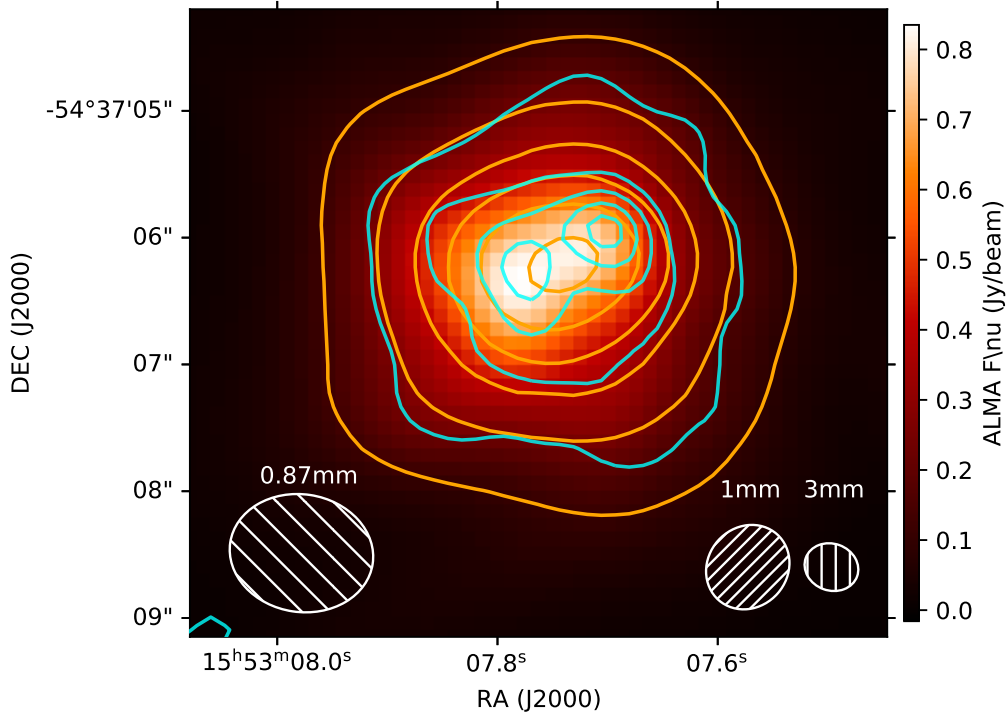


Figure 6.6: The comparison of hot core continuum from ALMA archive data. The background image is the ALMA 1mm continuum tracing the dust emission (ALMAIMF Project 2017.1.01355.L). The overlaid contours are ALMA 3mm continuum (cyan) and 0.87mm continuum (orange, from ALMA Project 2018.1.01449.S).

with ALMAIMF 3mm continuum ($0.4''$), we found the infrared sources are correlated with one of the sub-mm cores, while there is no infrared emission for the other sub-mm core. This may be explained by different extinction and optical depth of the two sub-mm cores, which we could not derive yet with the limited resolution and frequency bands of the observations. We also found an offset of $0.1''$ (330 au) between the peaks of PAH1 and PAH2, which is in doubt and needs to be proved through further VLT observation and calibration for the coordinates.

6.2.2 Possible disk fragmentation

Combining all the knowledge we have for the G327 hot core, we found evidence of possible disk fragmentation. First, we explain the unique geometry of the G327 hot core, which allows us to see through the envelop and directly on the star-forming hot cores.

For G327, the H_2 column density is about $2 \times 10^{25} \text{ cm}^{-2}$, which corresponds to a large visual extinction A_V of approximately 10,000. Non-detection of ionized gas in centimeter continuum observations (e.g., RMS survey: Lumsden et al. (2013) and ATCA data shown in Fig. 6.5) excludes that G327 is a much more evolved object. Usually in this case, we cannot observe the central prestellar object or the disk around it. Thus, we were very surprised when we found that Spitzer detected a point source at the very center of the core in all its IRAC bands ($3.6 \mu\text{m}$ to $8 \mu\text{m}$), which means small extinction at the position of protostars. At a shorter wavelength at $2 \mu\text{m}$ in the VVV survey, the point source is no longer detected, since the extinction is larger at a shorter wavelength. The most likely scenario is that we are looking at an almost face-on disk and see the IR source through the outflow cone. This scenario is corroborated by several pieces of evidence:

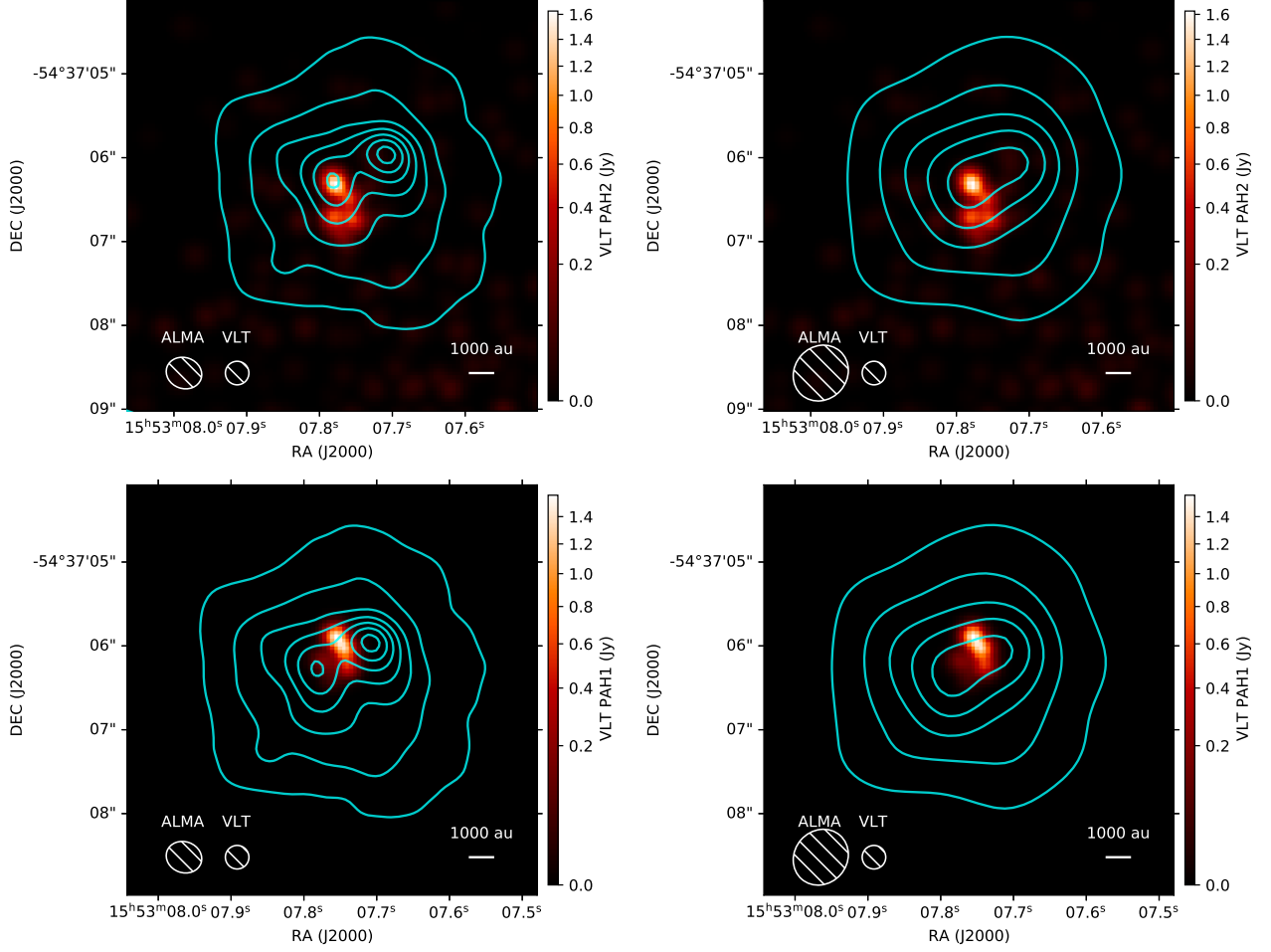


Figure 6.7: **Upper left:** The background color image is the VLT/VISIR 11.3 μm continuum observed with the PAH2 filter (Elbe et al. in prep), overlaid by ALMA 3mm continuum contour tracing the dust emission in cyan (from ALMAIMF Project 2017.1.01355.L). The comparison of resolution between ALMA (0.4" at 3 mm) and VLT (0.33" at 11.3 μm) are shown as white ellipses. The physical scale is shown as white line with a length of 1000 au. The infrared sources are correlated with one of the sub-mm cores. **Upper right:** PAH2 image (0.33") overlaid by ALMAIMF 1mm continuum (0.53"). **Bottom left:** PAH1 image (0.28" at 8.7 μm) overlaid by ALMAIMF 3mm continuum, indicating an offset between the PAH1 and sub-mm continuum cores. This deviation needs to be proved through further VLT observation and calibration for the position. **Bottom right:** PAH1 image overlaid by ALMAIMF 1mm continuum.

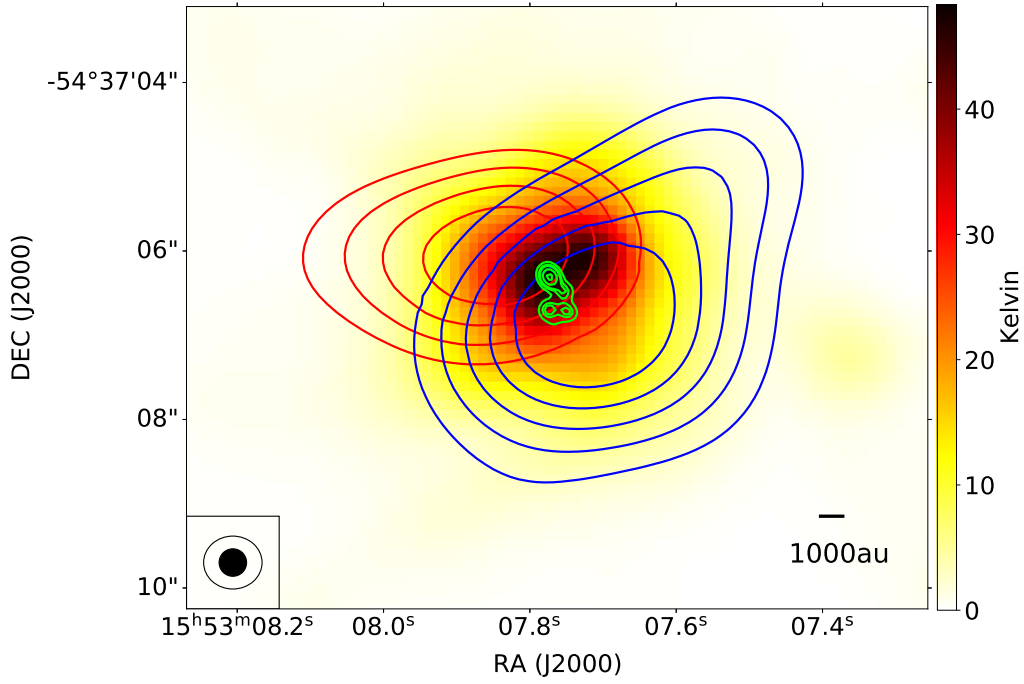


Figure 6.8: ALMA 1.3 mm continuum (false color, beam shown as a hollow circle), overlaid by CO(2-1) red and blue outflow lobes which are overlapping at the same position. The outflow is compact and only slightly elongated in the northeastern-southwestern direction. The infrared sources are newly detected by VLT/VISIR at $11.3 \mu\text{m}$ (green contours, beam shown as a filled circle, VLT Project 109.237A.003), indicating we see through the outflow cone and hence a face-on disk.

(1) The ALMA observation made for this thesis (resolution $\sim 1.9''$, or $\approx 6000 \text{ AU}$) does resolve a very compact outflow in CO (2–1) powered by the main dense cores, though it still remains unresolved (left panel of Fig. 1).

(2) Recent VLT VISIR observations (right panel of Fig. 6.7) show multiple hot spots with a separation of $\sim 0.2''$ (660 AU) at $11.3 \mu\text{m}$, which is a strong indication of high-mass disk fragmentation.

(3) ALMA 3 mm continuum shows two blobs with a distance of $\sim 0.6''$ (2000 AU), indicating core fragmentation and two potential disks.

(4) The fact that we only detected a slight velocity gradient through the disk ($\sim 2 \text{ km/s}$ in species not tracing the outflow, e.g., CH_3CN), is compatible with the fact that G327 is an almost face-on disk.

(5) The narrow line widths ($\sim 5 \text{ km/s}$ for e.g., CH_3OH , OCS , DCN and HC_3N) also indicate a small inclination.

(6) The geometry of the G327 magnetic field found by Beuther et al. (2020) is in favor of rotation in the plane of the sky.

Above all, G327 is one of the strongest ($\sim 2 \times 10^5 L_{\text{odot}}$) and chemically richest hot molecular cores in the Galaxy (Schilke et al., 2006; Wyrowski et al., 2006; Bisschop et al., 2013). The fact that it is actually one of the closest available high-mass cores, at a distance of 3.3 kpc, makes it possible to resolve the compact outflow and fragmentation of star-forming cores. All together,

this constitutes very strong evidence for the existence of large, face-on disks, but the final proof does require higher-resolution observations in both submillimeter and infrared, which are going to be done by ALMA, VLT, and James Webb Space Telescope (JWST).

7 Summary and outlook

The high-mass star-forming region G327.3-0.6 is a 3 pc filament hosting a hot molecular core, several high-mass cold cores, as well as other lower mass cores. In this study, we have analyzed the region following four aspects, including the continuum, the spectral line fitting, the spectral structure in large scale, and the corresponding infrared data.

Firstly, we start with the continuum analysis. The ALMA 12m and 7m array data were self-calibrated to improve the SNR by a factor of two. Then the high-resolution continuum map (2'' at 1.3 mm) was combined with the ALMA ACA data and single dish data to correct the missing flux in the interferometer. The combined image was analyzed by multiple algorithms to identify cores. The dendrogram algorithm, together with the background subtraction, performed the best and determined a list of 66 compact cores. We used minimum spanning tree to build a network of the core distribution, and derived a Q-value of 0.49, which favored sub=structures in our source. The two-point correlation function did not support double peaks of the nearest-neighbor separations. For each distinct core, the prestellar mass was derived from the continuum flux. We fitted the core mass function with an index of -0.83, which is a hint of high-mass star-forming regions. The mass-radius relation follows Larson's power law with an index of 2.26, indicating more high-mass cores in our samples. The mass segregation ratios are calculated to be close to unity, leading to the conclusion of a uniform mass cluster. The study of fragmentation in the filament and core scale is based on the theory of cylindrical fragmentation and (non)thermal Jeans fragmentation. Both scales show a dominant role of turbulence support rather than thermal pressure, while the filament/cores are not preferred to continue fragmenting.

For spectral analysis, the average spectra were extracted within the core area. First, we fitted the hot core spectrum using toolbox XCLASS. 26 molecules and 39 isotopes were fitted together, and an error estimation using MCMC method was done towards CH₃OH, proving the reliability of its high temperature (270 K) and small source size (0.485'' 1500 au). We identified the new vibrational state (v13+v21=1) lines of ethyl cyanide with the most recent laboratory data, and the fitting is improved with two components of cold (53 K) and hot (218 K) layer. Fitting the entire hot core region gives a gradient of the line parameters. For the other cores, formaldehyde was first fitted under the LTE assumption. We tested to fit with band 7 low resolution data, to fit two components, and to fit under non-LTE, all of which tell us that the three para-H₂CO lines are not enough to limit the temperature error. Despite that, the derived H₂CO abundance ratio has a peak distribution at the literature value (10^{-10} – 10^{-9}), although it has a much wider range (10^{-12} – 10^{-7}) because of high sensitivity and the hot environment. Methanol was also fitted for each core, whose temperatures are not the same as those of formaldehyde.

Then we used statistical methods to examine the evolutionary status of the cores. Principal component analysis (PCA) and clustering algorithm were applied to 65 cold cores. Eight principal components were adopted, and the cores were assigned to four groups according to their characteristic lines. Different groups possibly trace different evolutionary stages of early star formation: Group 1 shows weak lines of extended molecules, located far from the main filament; Group 2 and 3 have similar and richer lines, while Group 3 is closer to the hot core and its line strength is slightly stronger than Group 2; Group 4 traces the most significant cores on the filament and presents the richest chemistry among the four groups. The cores in Group 4 have a temperature at around 100 K and usually show intermediate mass, which tends to enter a hot core phase. Hence, PCA has been shown to be profitable in classifying cores.

For the whole filament, we derived the moment-zero maps of 42 transitions of 24 molecules. We used two methods, PCA and histogram of gradient (HOG), to examine the correlation between the molecules. We then investigate the eigenvectors of PCA, the mean resultant vector (MRV), the projected Rayleigh statistic (PRS), and alignment measurement (AM) of HOG. We found that the transitions with similar energy levels tend to show highly correlated moment zero maps, as well as the molecules with the same special atom (e.g., sulfur and deuterium). The most distinct molecules are CCD and H_2CO . N_2D^+ is also different from all the other images. Compared to the continuum image, we conclude that DCN, H_2CO , CH_3OH , HC_3N , OCS, and ^{13}CS are the most relevant molecules in descending order, while $\text{c-C}_3\text{H}_2$, CCD, $\text{C}_2\text{H}_5\text{CN}$, and HNC are the least correlated. In addition, the kinematics of the entire filament was explored using the DCN central velocity map (moment 1 map), which shows hints of global collapse towards the hot core.

As for the infrared counterparts, we looked into Spitzer RGB image and found 4 new extended green objects (EGOs) associated with the sub-mm cores. We also found 16 GLIMPSE high-reliability point sources appear at 13 sub-mm cores out of 66. Comparing the infrared image with the moment-zero map of a photon-dissociation region (PDR) tracer $\text{c-C}_3\text{H}_2$, we probably find a new PDR towards the west of the $\text{c-C}_3\text{H}_2$ arc. For the hot core, it is very possible to be a face-on disk and to fragment into multiple systems.

In the future, the central hot core is worth further exploration. The resolution ($2'' \sim 6500 \text{ au}$) of our study is not enough to resolve the prestellar object or the disk. Several days before submitting this thesis, a new observation has been performed by ALMA band 6 with a resolution of $0.1''$, and a future $0.02''$ resolution observation with the longest ALMA baseline will be done hopefully in August. Moreover, a new ALMA proposal of observing the hot core at a resolution of $0.08''$ at band 7 and 9 was submitted in Cycle 10, which will be very useful in understanding the optical depth and spectral index of the disk. In addition, a new VLT-VISIR near-infrared spectral observation was proposed to check the existence of disk fragments. We are going to resubmit the JWST MIRI spectral proposal in October this year. All of the mentioned observations will largely fruit us as the best template of early high-mass star formation.

As for principal component analysis, it works for a small sample in our study, but a larger sample is necessary for a better classification. ALMA Evolutionary Study of High Mass Protocluster Formation in the Galaxy (ALMAGAL) provides us with a good chance. It is a large-scale (1000+ sources) super-high-resolution ($< 0.5''$) survey being carried out using ALMA this year. The same method could be applied to the new ALMAGAL data to quickly segregate high-mass and line-rich sources. Another long-term goal of this study is to determine the evolutionary states of high/low mass protostars by PCA, thereby facilitating an understanding of molecules as tracers of star-forming cores. The method developed in this thesis will be very practical to answer the key questions in the ALMAGAL proposal: What are the physical processes governing the formation of high-mass stars and clusters? And how do they evolve over time? I will extend this work in the future, also focusing on the application of machine learning techniques towards star formation.

References

- Emilio J. Alfaro and Carlos G. Román-Zúñiga. Primordial mass and density segregation in a young molecular cloud. *MNRAS*, 478(1):L110–L115, July 2018. doi: 10.1093/mnras/sly075.
- Richard J. Allison, Simon P. Goodwin, Richard J. Parker, Simon F. Portegies Zwart, Richard de Grijs, and M. B. N. Kouwenhoven. Using the minimum spanning tree to trace mass segregation. *MNRAS*, 395(3):1449–1454, May 2009. doi: 10.1111/j.1365-2966.2009.14508.x.
- Y. Ao, C. Henkel, K. M. Menten, M. A. Requena-Torres, T. Stanke, R. Mauersberger, S. Aalto, S. Mühle, and J. Mangum. The thermal state of molecular clouds in the Galactic center: evidence for non-photon-driven heating. *A&A*, 550:A135, February 2013. doi: 10.1051/0004-6361/201220096.
- Javier Ballesteros-Paredes, Philippe André, Patrick Hennebelle, Ralf S. Klessen, J. M. Diederik Kruijssen, Mélanie Chevance, Fumitaka Nakamura, Angela Adamo, and Enrique Vázquez-Semadeni. From Diffuse Gas to Dense Molecular Cloud Cores. *Space Science Reviews*, 216(5):76, June 2020. doi: 10.1007/s11214-020-00698-3.
- John Bally. Protostellar Outflows. *ARA&A*, 54:491–528, September 2016. doi: 10.1146/annurev-astro-081915-023341.
- R. A. Batchelor, J. L. Caswell, W. M. Goss, R. F. Haynes, S. H. Knowles, and K. J. Wellington. Galactic plane H₂O masers: a southern survey. *Australian Journal of Physics*, 33:139–157, March 1980. doi: 10.1071/PH800139.
- R. A. Batchelor, M. G. McCulloch, and J. B. Whiteoak. Observations of the $J = 1-0$ transition of HCO⁺ in southern molecular clouds. *MNRAS*, 194:911–917, March 1981. doi: 10.1093/mnras/194.4.911.
- A. Belloche, H. S. P. Müller, K. M. Menten, P. Schilke, and C. Comito. Complex organic molecules in the interstellar medium: IRAM 30 m line survey of Sagittarius B2(N) and (M). *A&A*, 559:A47, November 2013. doi: 10.1051/0004-6361/201321096.
- P. Bergman. PhD thesis, Chalmers Univ. Technology, 1992.
- H. Beuther, P. Schilke, K. M. Menten, F. Motte, T. K. Sridharan, and F. Wyrowski. High-Mass Protostellar Candidates. II. Density Structure from Dust Continuum and CS Emission. *ApJ*, 566(2):945–965, February 2002. doi: 10.1086/338334.
- H. Beuther, C. Gieser, S. Suri, H. Linz, P. Klaassen, D. Semenov, J. M. Winters, Th. Henning, J. D. Soler, J. S. Urquhart, J. Syed, S. Feng, T. Möller, M. T. Beltrán, Á. Sánchez-Monge, S. N. Longmore, T. Peters, J. Ballesteros-Paredes, P. Schilke, L. Moscadelli, A. Palau, R. Cesaroni, S. Lumsden, R. Pudritz, F. Wyrowski, R. Kuiper, and A. Ahmadi. Fragmentation and kinematics in high-mass star formation. CORE-extension targeting two very young high-mass star-forming regions. *A&A*, 649:A113, May 2021. doi: 10.1051/0004-6361/202040106.
- Henrik Beuther, Juan D. Soler, Hendrik Linz, Thomas Henning, Caroline Gieser, Rolf Kuiper, Wouter Vlemmings, Patrick Hennebelle, Siyi Feng, Rowan Smith, and Aida Ahmadi. Gravity and Rotation Drag the Magnetic Field in High-mass Star Formation. *ApJ*, 904(2):168, December 2020. doi: 10.3847/1538-4357/abc019.
- S. Bhatnagar and T. J. Cornwell. Scale sensitive deconvolution of interferometric images. I. Adaptive Scale Pixel (Asp) decomposition. *A&A*, 426:747–754, November 2004. doi: 10.1051/0004-6361:20040354.

- S. E. Bisschop, P. Schilke, F. Wyrowski, A. Belloche, C. Brinch, C. P. Endres, R. Güsten, H. Hafok, S. Heyminck, J. K. Jørgensen, H. S. P. Müller, K. M. Menten, R. Rolfs, and S. Schlemmer. Dimethyl ether in its ground state, $v = 0$, and lowest two torsionally excited states, $v_{11} = 1$ and $v_{15} = 1$, in the high-mass star-forming region G327.3-0.6. *A&A*, 552: A122, April 2013. doi: 10.1051/0004-6361/201118719.
- I. A. Bonnell, M. R. Bate, C. J. Clarke, and J. E. Pringle. Competitive accretion in embedded stellar clusters. *MNRAS*, 323(4):785–794, May 2001. doi: 10.1046/j.1365-8711.2001.04270.x.
- Ian A. Bonnell and Matthew R. Bate. Star formation through gravitational collapse and competitive accretion. *MNRAS*, 370(1):488–494, July 2006. doi: 10.1111/j.1365-2966.2006.10495.x.
- A. C. A. Boogert and P. Ehrenfreund. Interstellar Ices. In Adolf N. Witt, Geoffrey C. Clayton, and Bruce T. Draine, editors, *Astrophysics of Dust*, volume 309 of *Astronomical Society of the Pacific Conference Series*, page 547, May 2004. doi: 10.48550/arXiv.astro-ph/0311163.
- J. Brand. CO in southern sources. In W. B. Burton and F. P. Israel, editors, *Surveys of the Southern Galaxy*, volume 105 of *Astrophysics and Space Science Library*, pages 223–228, January 1983. doi: 10.1007/978-94-009-7217-9_29.
- D. S. Briggs. High Fidelity Interferometric Imaging: Robust Weighting and NNLS Deconvolution. In *American Astronomical Society Meeting Abstracts*, volume 187 of *American Astronomical Society Meeting Abstracts*, page 112.02, December 1995.
- Francis Graham-Smith Burnard F. Burke. *An introduction to Radio Astronomy (second edition)*. Cambridge University Press, 2002.
- G. Busquet, J. M. Girart, R. Estalella, M. Fernández-López, R. Galván-Madrid, G. Anglada, C. Carrasco-González, N. Añez-López, S. Curiel, M. Osorio, L. F. Rodríguez, and J. M. Torrelles. Unveiling a cluster of protostellar disks around the massive protostar GGD 27 MM1. *A&A*, 623:L8, March 2019. doi: 10.1051/0004-6361/201833687.
- Gemma Busquet, Robert Estalella, Aina Palau, Hauyu Baobab Liu, Qizhou Zhang, Josep Miquel Girart, Itziar de Gregorio-Monsalvo, Thushara Pillai, Guillem Anglada, and Paul T. P. Ho. What Is Controlling the Fragmentation in the Infrared Dark Cloud G14.225-0.506?: Different Levels of Fragmentation in Twin Hubs. *ApJ*, 819(2):139, March 2016. doi: 10.3847/0004-637X/819/2/139.
- T. Calinski and J. Harabasz. A dendrite method for cluster analysis. *Communications in Statistics*, 3:1–27, September 1974. URL <https://www.tandfonline.com/doi/abs/10.1080/03610927408827101>.
- Sean J. Carey, P. A. Feldman, R. O. Redman, M. P. Egan, J. M. MacLeod, and S. D. Price. Submillimeter Observations of Midcourse Space Experiment Galactic Infrared-Dark Clouds. *ApJL*, 543(2):L157–L161, November 2000. doi: 10.1086/317270.
- Annabel Cartwright and Anthony P. Whitworth. The statistical analysis of star clusters. *MNRAS*, 348(2):589–598, February 2004. doi: 10.1111/j.1365-2966.2004.07360.x.
- J. L. Caswell, R. A. Batchelor, R. F. Haynes, and W. K. Huchtmeier. 22 GHz observations of southern H₂O sources. *Australian Journal of Physics*, 27:417, June 1974. doi: 10.1071/PH740417.

- E. T. Chambers, J. M. Jackson, J. M. Rathborne, and R. Simon. Star Formation Activity of Cores within Infrared Dark Clouds. *ApJS*, 181(2):360–390, April 2009. doi: 10.1088/0067-0049/181/2/360.
- S. Chandrasekhar and E. Fermi. Problems of Gravitational Stability in the Presence of a Magnetic Field. *ApJ*, 118:116, July 1953. doi: 10.1086/145732.
- Xi Chen, Cong-Gui Gan, Simon P. Ellingsen, Jin-Hua He, Zhi-Qiang Shen, and Anita Titmarsh. Newly Identified Extended Green Objects (EGOs) from the Spitzer GLIMPSE II Survey. I. Catalog. *ApJS*, 206(1):9, May 2013. doi: 10.1088/0067-0049/206/1/9.
- Yu Cheng, Jonathan C. Tan, Mengyao Liu, Shuo Kong, Wanggi Lim, Morten Andersen, and Nicola Da Rio. The Core Mass Function in the Massive Protocluster G286.21+0.17 Revealed by ALMA. *ApJ*, 853(2):160, February 2018. doi: 10.3847/1538-4357/aaa3f1.
- R. Choudhury, P. Schilke, G. Stéphan, E. Bergin, T. Möller, A. Schmiedeke, and A. Zernickel. Evolution of complex organic molecules in hot molecular cores. Synthetic spectra at (sub-)mm wavebands. *A&A*, 575:A68, March 2015. doi: 10.1051/0004-6361/201424499.
- D. Colombo, E. Rosolowsky, A. Ginsburg, A. Duarte-Cabral, and A. Hughes. Graph-based interpretation of the molecular interstellar medium segmentation. *MNRAS*, 454(2):2067–2091, December 2015. doi: 10.1093/mnras/stv2063.
- T. J. Cornwell. Multiscale CLEAN Deconvolution of Radio Synthesis Images. *IEEE Journal of Selected Topics in Signal Processing*, 2(5):793–801, November 2008. doi: 10.1109/JSTSP.2008.2006388.
- T. Csengeri, A. Belloche, S. Bontemps, F. Wyrowski, K. M. Menten, and L. Bouscasse. Search for high-mass protostars with ALMA revealed up to kilo-parsec scales (SPARKS). II. Complex organic molecules and heavy water in shocks around a young high-mass protostar. *A&A*, 632:A57, December 2019. doi: 10.1051/0004-6361/201935226.
- C. J. Cyganowski, B. A. Whitney, E. Holden, E. Braden, C. L. Brogan, E. Churchwell, R. Indebetouw, D. F. Watson, B. L. Babler, R. Benjamin, M. Gomez, M. R. Meade, M. S. Povich, T. P. Robitaille, and C. Watson. A Catalog of Extended Green Objects in the GLIMPSE Survey: A New Sample of Massive Young Stellar Object Outflow Candidates. *AJ*, 136(6):2391–2412, December 2008. doi: 10.1088/0004-6256/136/6/2391.
- Scott D. Brown Don van Ravenzwaaij, Pete Cassey. A simple introduction to Markov Chain Monte-Carlo sampling. *Psychonomic Bulletin & Review*, 25:143–154, 2018.
- Bruce T. Draine. *Physics of the Interstellar and Intergalactic Medium*. 2011.
- M. L. Dubernet, B. K. Antony, Y. A. Ba, Yu L. Babikov, K. Bartschat, V. Boudon, B. J. Braams, H. K. Chung, F. Daniel, F. Delahaye, G. Del Zanna, J. de Urquijo, M. S. Dimitrijević, A. Domaracka, M. Doronin, B. J. Drouin, C. P. Endres, A. Z. Fazliev, S. V. Gagarin, I. E. Gordon, P. Gratier, U. Heiter, C. Hill, D. Jevremović, C. Joblin, A. Kasprzak, E. Krishnakumar, G. Leto, P. A. Loboda, T. Louge, S. Maciot, B. P. Marinković, A. Markwick, T. Marquart, H. E. Mason, N. J. Mason, C. Mendoza, A. A. Mihajlov, T. J. Millar, N. Moreau, G. Mulas, Yu Pakhomov, P. Palmeri, S. Pancheshnyi, V. I. Perevalov, N. Piskunov, J. Postler, P. Quinet, E. Quintas-Sánchez, Yu Ralchenko, Y. J. Rhee, G. Rixon, L. S. Rothman, E. Roueff, T. Ryabchikova, S. Sahal-Bréchet, P. Scheier, S. Schlemmer, B. Schmitt, E. Stempels, S. Tashkun, J. Tennyson, VI G. Tyuterev, V. Vujčić, V. Wakelam,

- N. A. Walton, O. Zatsarinny, C. J. Zeippen, and C. M. Zwölf. The virtual atomic and molecular data centre (VAMDC) consortium. *Journal of Physics B Atomic Molecular Physics*, 49(7):074003, April 2016. doi: 10.1088/0953-4075/49/7/074003.
- Davide Elia, S. Molinari, E. Schisano, M. Pestalozzi, S. Pezzuto, M. Merello, A. Noriega-Crespo, T. J. T. Moore, D. Russeil, J. C. Mottram, R. Paladini, F. Strafella, M. Benedettini, J. P. Bernard, A. Di Giorgio, D. J. Eden, Y. Fukui, R. Plume, J. Bally, P. G. Martin, S. E. Ragan, S. E. Jaffa, F. Motte, L. Olmi, N. Schneider, L. Testi, F. Wyrowski, A. Zavagno, L. Calzolari, F. Faustini, P. Natoli, P. Palmeirim, F. Piacentini, L. Piazzo, G. L. Pilbratt, D. Polychroni, A. Baldeschi, M. T. Beltrán, N. Billot, L. Cambrésy, R. Cesaroni, P. García-Lario, M. G. Hoare, M. Huang, G. Joncas, S. J. Liu, B. M. T. Maiolo, K. A. Marsh, Y. Maruccia, P. Mège, N. Peretto, K. L. J. Rygl, P. Schilke, M. A. Thompson, A. Traficante, G. Umana, M. Veneziani, D. Ward-Thompson, A. P. Whitworth, H. Arab, M. Bandieramonte, U. Beciani, M. Brescia, C. Buemi, F. Bufano, R. Butora, S. Cavuoti, A. Costa, E. Fiorellino, A. Hajnal, T. Hayakawa, P. Kacsuk, P. Leto, G. Li Causi, N. Marchili, S. Martinavarro-Armengol, A. Mercurio, M. Molinaro, G. Riccio, H. Sano, E. Sciacca, K. Tachihara, K. Torii, C. Trigilio, F. Vitello, and H. Yamamoto. The Hi-GAL compact source catalogue - I. The physical properties of the clumps in the inner Galaxy ($-71.0^\circ < \ell < 67.0^\circ$). *MNRAS*, 471(1): 100–143, October 2017. doi: 10.1093/mnras/stx1357.
- Christian P. Endres, Marie-Aline Martin-Drumel, Oliver Zingsheim, Luis Bonah, Olivier Pirali, Tianwei Zhang, Álvaro Sánchez-Monge, Thomas Möller, Nadine Wehres, Peter Schilke, Michael C. McCarthy, Stephan Schlemmer, Paola Caselli, and Sven Thorwirth. SOLEIL and ALMA views on prototypical organic nitriles: $\text{C}_2\text{H}_5\text{CN}$. *Journal of Molecular Spectroscopy*, 375:111392, January 2021. doi: 10.1016/j.jms.2020.111392.
- ALMA Partnership et. al. First Results from High Angular Resolution ALMA Observations Toward the HL Tau Region. 2015.
- II Evans, Neal J., Michael M. Dunham, Jes K. Jørgensen, Melissa L. Enoch, Bruno Merín, Ewine F. van Dishoeck, Juan M. Alcalá, Philip C. Myers, Karl R. Stapelfeldt, Tracy L. Huard, Lori E. Allen, Paul M. Harvey, Tim van Kempen, Geoffrey A. Blake, David W. Koerner, Lee G. Mundy, Deborah L. Padgett, and Anneila I. Sargent. The Spitzer c2d Legacy Results: Star-Formation Rates and Efficiencies; Evolution and Lifetimes. *ApJS*, 181(2):321–350, April 2009. doi: 10.1088/0067-0049/181/2/321.
- Jr. Frank J. Massey. The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*, 46:68–78, March 1951. URL <https://doi.org/10.2307/2280095>.
- Roth M. Freeman, W. T. Orientation Histograms for Hand Gesture Recognition. *Tech. Rep. TR94-03, MERL - Mitsubishi Electric Research Laboratories, Cambridge, MA, 02139*, 1994.
- F. F. Gardner and J. B. Whiteoak. Observations of the $J = 1-0$ transition of CS at 49 GHz in southern molecular clouds. *MNRAS*, 183:711–725, June 1978. doi: 10.1093/mnras/183.4.711.
- E. Gibb, A. Nummelin, W. M. Irvine, D. C. B. Whittet, and P. Bergman. Chemistry of the Organic-Rich Hot Core G327.3-0.6. *ApJ*, 545(1):309–326, December 2000. doi: 10.1086/317805.
- C. Gieser, D. Semenov, H. Beuther, A. Ahmadi, J. C. Mottram, Th. Henning, M. Beltran, L. T. Maud, F. Bosco, S. Leurini, T. Peters, P. Klaassen, R. Kuiper, S. Feng, J. S. Urquhart, L. Moscadelli, T. Csengeri, S. Lumsden, J. M. Winters, S. Suri, Q. Zhang, R. Pudritz,

- A. Palau, K. M. Menten, R. Galvan-Madrid, F. Wyrowski, P. Schilke, Á. Sánchez-Monge, H. Linz, K. G. Johnston, I. Jiménez-Serra, S. Longmore, and T. Möller. Chemical complexity in high-mass star formation. An observational and modeling case study of the AFGL 2591 VLA 3 hot core. *A&A*, 631:A142, November 2019. doi: 10.1051/0004-6361/201935865.
- A. Ginsburg, T. Csengeri, R. Galván-Madrid, N. Cunningham, R. H. Álvarez-Gutiérrez, T. Baug, M. Bonfand, S. Bontemps, G. Busquet, D. J. Díaz-González, M. Fernández-López, A. Guzmán, F. Herpin, H. Liu, A. López-Sepulcre, F. Louvet, L. Maud, F. Motte, F. Nakamura, T. Nony, F. A. Olguin, Y. Pouteau, P. Sanhueza, A. M. Stutz, A. P. M. Towner, ALMA-IMF Consortium, M. Armante, C. Battersby, L. Bronfman, J. Braine, N. Brouillet, E. Chapillon, J. Di Francesco, A. Gusdorf, N. Izumi, I. Joncour, X. Walker Lu, A. Men'shchikov, K. M. Menten, E. Moraux, J. Molet, L. Mundy, Q. Nguyen Luong, S. D. Reyes-Reyes, J. Robitaille, E. Rosolowsky, N. A. Sandoval-Garrido, B. Svoboda, K. Tatematsu, D. L. Walker, A. Whitworth, B. Wu, and F. Wyrowski. ALMA-IMF. II. Investigating the origin of stellar masses: Continuum images and data processing. *A&A*, 662:A9, June 2022. doi: 10.1051/0004-6361/202141681.
- Alyssa A. Goodman, Erik W. Rosolowsky, Michelle A. Borkin, Jonathan B. Foster, Michael Halle, Jens Kauffmann, and Jaime E. Pineda. A role for self-gravity at multiple length scales in the process of star formation. *Nature*, 457(7225):63–66, January 2009. doi: 10.1038/nature07609.
- S. P. Goodwin, D. Nutter, P. Kroupa, D. Ward-Thompson, and A. P. Whitworth. The relationship between the prestellar core mass function and the stellar initial mass function. *A&A*, 477(3):823–827, January 2008. doi: 10.1051/0004-6361:20078452.
- W. M. Goss and P. A. Shaver. Galactic Radio Sources I: High Resolution 5000 MHz observations. *Australian Journal of Physics Astrophysical Supplement*, 14:1, January 1970.
- Andrés E. Guzmán, Patricio Sanhueza, Yanett Contreras, Howard A. Smith, James M. Jackson, Sadia Hoq, and Jill M. Rathborne. Far-infrared Dust Temperatures and Column Densities of the MALT90 Molecular Clump Sample. *ApJ*, 815(2):130, December 2015. doi: 10.1088/0004-637X/815/2/130.
- Jerrold h. Zar. Probabilities of Rayleigh's test statistics for circular data. *Behavior Research Methods*, 6(4):450–450, July 1974. doi: 10.3758/BF03200403.
- J. Hatchell, G. A. Fuller, T. J. Millar, M. A. Thompson, and G. H. Macdonald. SCUBA imaging of high mass star formation regions. *A&A*, 357:637–650, May 2000.
- Mark Heyer, Coleman Krawczyk, Julia Duval, and James M. Jackson. Re-Examining Larson's Scaling Relationships in Galactic Molecular Clouds. *ApJ*, 699(2):1092–1103, July 2009. doi: 10.1088/0004-637X/699/2/1092.
- Mark H. Heyer and F. Peter Schloerb. Application of Principal Component Analysis to Large-Scale Spectral Line Imaging Studies of the Interstellar Medium. *ApJ*, 475(1):173–187, January 1997. doi: 10.1086/303514.
- P. T. P. Ho and C. H. Townes. Interstellar ammonia. *ARA&A*, 21:239–270, January 1983. doi: 10.1146/annurev.aa.21.090183.001323.
- J. A. Högbom. Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines. *A&AS*, 15:417, June 1974.

- Michiel Hogerheijde. *The molecular environment of low-mass protostars*. PhD thesis, University of California, Berkeley, Department of Astronomy, June 1998.
- Philip F. Hopkins. Variations in the stellar CMF and IMF: from bottom to top. *MNRAS*, 433(1):170–177, July 2013. doi: 10.1093/mnras/stt713.
- T. R. Hunter, C. L. Brogan, C. J. Cyganowski, and K. H. Young. Subarcsecond Imaging of the NGC 6334 I(N) Protocluster: Two Dozen Compact Sources and a Massive Disk Candidate. *ApJ*, 788(2):187, June 2014. doi: 10.1088/0004-637X/788/2/187.
- John D. Ilee, Catherine Walsh, Alice S. Booth, Yuri Aikawa, Sean M. Andrews, Jaehan Bae, Edwin A. Bergin, Jennifer B. Bergner, Arthur D. Bosman, Gianni Cataldi, L. Ilse Cleeves, Ian Czekala, Viviana V. Guzmán, Jane Huang, Charles J. Law, Romane Le Gal, Ryan A. Loomis, François Ménard, Hideko Nomura, Karin I. Öberg, Chunhua Qi, Kamber R. Schwarz, Richard Teague, Takashi Tsukagoshi, David J. Wilner, Yoshihide Yamato, and Ke Zhang. Molecules with ALMA at Planet-forming Scales (MAPS). IX. Distribution and Properties of the Large Organic Molecules HC₃N, CH₃CN, and c-C₃H₂. *ApJS*, 257(1):9, November 2021. doi: 10.3847/1538-4365/ac1441.
- K. J. Johnston, B. J. Robinson, J. L. Caswell, and R. A. Batchelor. Water Sources Associated with OH Emission in the Southern Milky Way. *Astrophysics Letters*, 10:93, January 1972.
- D. Johnstone, A. M. S. Boonman, and E. F. van Dishoeck. Astrochemistry of sub-millimeter sources in Orion. Studying the variations of molecular tracers with changing physical conditions. *A&A*, 412:157–174, December 2003. doi: 10.1051/0004-6361:20031370.
- Dylan L. Jow, Ryley Hill, Douglas Scott, J. D. Soler, P. G. Martin, M. J. Devlin, L. M. Fissel, and F. Poidevin. An application of an optimal statistic for characterizing relative orientations. *MNRAS*, 474(1):1018–1027, February 2018. doi: 10.1093/mnras/stx2736.
- J. Kainulainen, H. Beuther, R. Banerjee, C. Federrath, and T. Henning. Probing the evolution of molecular cloud structure. II. From chaos to confinement. *A&A*, 530:A64, June 2011. doi: 10.1051/0004-6361/201016383.
- Rousseeuw P. J. Kaufman, L. Finding Groups in Data: An Introduction to Cluster Analysis. *John Wiley & Sons*, 1990.
- P. Kervella, G. Perrin, A. Chiavassa, S. T. Ridgway, J. Cami, X. Haubois, and T. Verhoelst. The close circumstellar environment of Betelgeuse. II. Diffraction-limited spectro-imaging from 7.76 to 19.50 μm with VLT/VISIR. *A&A*, 531:A117, July 2011. doi: 10.1051/0004-6361/201116962.
- Seongjoong Kim, Hideko Nomura, Takashi Tsukagoshi, Ryohei Kawabe, and Takayuki Muto. The Synthetic ALMA Multiband Analysis of the Dust Properties of the TW Hya Protoplanetary Disk. *ApJ*, 872(2):179, February 2019. doi: 10.3847/1538-4357/ab0304.
- H. Kirk, D. Johnstone, J. Di Francesco, J. Lane, J. Buckle, D. S. Berry, H. Broekhoven-Fiene, M. J. Currie, M. Fich, J. Hatchell, T. Jenness, J. C. Mottram, D. Nutter, K. Pattle, J. E. Pineda, C. Quinn, C. Salji, S. Tisi, M. R. Hogerheijde, D. Ward-Thompson, and JCMT Gould Belt Survey Team. The JCMT Gould Belt Survey: Dense Core Clusters in Orion B. *ApJ*, 821(2):98, April 2016. doi: 10.3847/0004-637X/821/2/98.
- Kobayashi, K. *Complex Organic Molecules*, pages 352–352. Springer Berlin Heidelberg, 2011. ISBN 978-3-642-11274-4. doi: 10.1007/978-3-642-11274-4_337. URL https://doi.org/10.1007/978-3-642-11274-4_337.

- Jin Koda, Tsuyoshi Sawada, Melvyn C. H. Wright, Peter Teuben, Stuartt A. Corder, Jenny Patience, Nick Scoville, Jennifer Donovan Meyer, and Fumi Egusa. CO(J = 1-0) Imaging of M51 with CARMA and the Nobeyama 45 m Telescope. *ApJS*, 193(1):19, March 2011. doi: 10.1088/0067-0049/193/1/19.
- Jin Koda, Peter Teuben, Tsuyoshi Sawada, Adele Plunkett, and Ed Fomalont. Total Power Map to Visibilities (TP2VIS): Joint Deconvolution of ALMA 12m, 7m, and Total Power Array Data. *PASP*, 131(999):054505, May 2019. doi: 10.1088/1538-3873/ab047e.
- Nico Krieger, Alberto D. Bolatto, Eric W. Koch, Adam K. Leroy, Erik Rosolowsky, Fabian Walter, Axel Weiß, David J. Eden, Rebecca C. Levy, David S. Meier, Elisabeth A. C. Mills, Toby Moore, Jürgen Ott, Yang Su, and Sylvain Veilleux. The Turbulent Gas Structure in the Centers of NGC 253 and the Milky Way. *ApJ*, 899(2):158, August 2020. doi: 10.3847/1538-4357/aba903.
- P. Kroupa. The Local Stellar Initial Mass Function. In S. Deiters, B. Fuchs, A. Just, R. Spurzem, and R. Wielen, editors, *Dynamics of Star Clusters and the Milky Way*, volume 228 of *Astronomical Society of the Pacific Conference Series*, page 187, January 2001. doi: 10.48550/arXiv.astro-ph/0011328.
- Joseph B. Kruskal. On the shortest spanning subtree of a graph and the traveling salesman problem. *American Mathematical Society*, 7:48–50, 1956. URL <http://dx.doi.org/10.1090/S0002-9939-1956-0078686-7>.
- Michael A. Kuhn, Rafael S. de Souza, Alberto Krone-Martins, Alfred Castro-Ginard, Emille E. O. Ishida, Matthew S. Povich, Lynne A. Hillenbrand, and COIN Collaboration. SPICY: The Spitzer/IRAC Candidate YSO Catalog for the Inner Galactic Midplane. *ApJS*, 254(2):33, June 2021. doi: 10.3847/1538-4365/abe465.
- Stephen D. Landy and Alexander S. Szalay. Bias and Variance of Angular Correlation Functions. *ApJ*, 412:64, July 1993. doi: 10.1086/172900.
- R. B. Larson. Turbulence and star formation in molecular clouds. *MNRAS*, 194:809–826, March 1981. doi: 10.1093/mnras/194.4.809.
- A. Lazarian and Thiem Hoang. Radiative torques: analytical model and basic properties. *MNRAS*, 378(3):910–946, July 2007. doi: 10.1111/j.1365-2966.2007.11817.x.
- S. Leurini, F. Wyrowski, F. Herpin, F. van der Tak, R. Güsten, and E. F. van Dishoeck. The distribution of warm gas in the G327.3-0.6 massive star-forming region. *A&A*, 550:A10, February 2013. doi: 10.1051/0004-6361/201218962.
- S. Leurini, F. Herpin, F. van der Tak, F. Wyrowski, G. J. Herczeg, and E. F. van Dishoeck. Distribution of water in the G327.3-0.6 massive star-forming region. *A&A*, 602:A70, June 2017. doi: 10.1051/0004-6361/201730387.
- D. Li, T. Velusamy, P. F. Goldsmith, and William D. Langer. Massive Quiescent Cores in Orion. II. Core Mass Function. *ApJ*, 655(1):351–363, January 2007. doi: 10.1086/509736.
- Guang-Xing Li, Yue Cao, and Keping Qiu. Network of Star Formation: Fragmentation Controlled by Scale-dependent Turbulent Pressure and Accretion onto the Massive Cores Revealed in the Cygnus-X GMC Complex. *ApJ*, 916(1):13, July 2021. doi: 10.3847/1538-4357/ac01d4.

- Adrian Liu, Max Tegmark, Scott Morrison, Andrew Lutomirski, and Matias Zaldarriaga. Precision calibration of radio interferometers using redundant baselines. *MNRAS*, 408(2):1029–1050, October 2010. doi: 10.1111/j.1365-2966.2010.17174.x.
- Mengyao Liu, Jonathan C. Tan, Yu Cheng, and Shuo Kong. The Core Mass Function across Galactic Environments. II. Infrared Dark Cloud Clumps. *ApJ*, 862(2):105, August 2018. doi: 10.3847/1538-4357/aacb7c.
- Tie Liu, Yuefang Wu, Sheng-Yuan Liu, Sheng-Li Qin, Yu-Nung Su, Huei-Ru Chen, and Zhiyuan Ren. Infall and Outflow Motions in the High-mass Star-forming Complex G9.62+0.19. *ApJ*, 730(2):102, April 2011. doi: 10.1088/0004-637X/730/2/102.
- Tie Liu, Neal J. Evans, Kee-Tae Kim, Paul F. Goldsmith, Sheng-Yuan Liu, Qizhou Zhang, Ken'ichi Tatematsu, Ke Wang, Mika Juvela, Leonardo Bronfman, Maria R. Cunningham, Guido Garay, Tomoya Hirota, Jeong-Eun Lee, Sung-Ju Kang, Di Li, Pak-Shing Li, Diego Mardones, Sheng-Li Qin, Isabelle Ristorcelli, Anandmayee Tej, L. Viktor Toth, Jing-Wen Wu, Yue-Fang Wu, Hee-weon Yi, Hyeong-Sik Yun, Hong-Li Liu, Ya-Ping Peng, Juan Li, Shang-Huo Li, Chang Won Lee, Zhi-Qiang Shen, Tapas Baug, Jun-Zhi Wang, Yong Zhang, Namitha Issac, Feng-Yao Zhu, Qiu-Yi Luo, Archana Soam, Xun-Chuan Liu, Feng-Wei Xu, Yu Wang, Chao Zhang, Zhiyuan Ren, and Chao Zhang. ATOMS: ALMA Three-millimeter Observations of Massive Star-forming regions - I. Survey description and a first look at G9.62+0.19. *MNRAS*, 496(3):2790–2820, August 2020. doi: 10.1093/mnras/staa1577.
- M. Lombardi, J. Alves, and C. J. Lada. Larson's third law and the universality of molecular cloud structure. *A&A*, 519:L7, September 2010. doi: 10.1051/0004-6361/201015282.
- S. L. Lumsden, M. G. Hoare, J. S. Urquhart, R. D. Oudmaijer, B. Davies, J. C. Mottram, H. D. B. Cooper, and T. J. T. Moore. The Red MSX Source Survey: The Massive Young Stellar Population of Our Galaxy. *ApJS*, 208(1):11, September 2013. doi: 10.1088/0067-0049/208/1/11.
- Veronique Boiteau Azam Niknafs Malika Charrad, Nadia Ghazzali. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software*, 61:6, November 2014. doi: 10.18637/jss.v061.i06. URL <https://www.jstatsoft.org/article/view/v061i06>.
- Jeffrey G. Mangum and Alwyn Wootten. Formaldehyde as a Probe of Physical Conditions in Dense Molecular Clouds. *ApJS*, 89:123, November 1993. doi: 10.1086/191841.
- Jeffrey G. Mangum, Alwyn Wootten, and Richard L. Plambeck. The Physical Structure of Orion-KL on 2500 AU Scales Using the K-Doublet Transitions of Formaldehyde. *ApJ*, 409:282, May 1993. doi: 10.1086/172663.
- McConnell, R. K. Method of and apparatus for pattern recognition. 1 1986. URL <https://www.osti.gov/biblio/6007283>.
- Christopher F. McKee and Jonathan C. Tan. The Formation of Massive Stars from Turbulent Cores. *ApJ*, 585(2):850–871, March 2003. doi: 10.1086/346149.
- K. M. Menten, C. M. Walmsley, C. Henkel, and T. L. Wilson. Methanol in the Orion region. I. Millimeter-wave observations. *A&A*, 198:253–266, June 1988.

- V. Minier, Ph. André, P. Bergman, F. Motte, F. Wyrowski, J. Le Penneç, L. Rodriguez, O. Boulade, E. Doumayrou, D. Dubreuil, P. Gallais, G. Hamon, P. O. Lagage, M. Lortholary, J. Martignac, V. Révêret, H. Roussel, M. Talvard, G. Willmann, and H. Olofsson. Evidence of triggered star formation in G327.3-0.6. Dust-continuum mapping of an infrared dark cloud with P-ArTéMiS. *A&A*, 501(1):L1–L4, July 2009. doi: 10.1051/0004-6361/200912308.
- T. Möller, C. Endres, and P. Schilke. eXtended CASA Line Analysis Software Suite (XCLASS). *A&A*, 598:A7, February 2017. doi: 10.1051/0004-6361/201527203.
- S. Mühle, E. R. Seaquist, and C. Henkel. Formaldehyde as a Tracer of Extragalactic Molecular Gas. I. Para-H₂CO Emission from M82. *ApJ*, 671(2):1579–1590, December 2007. doi: 10.1086/522294.
- A. Nummelin, J. E. Dickens, P. Bergman, Aa. Hjalmarson, W. M. Irvine, M. Ikeda, and M. Ohishi. Abundances of ethylene oxide and acetaldehyde in hot molecular cloud cores. *A&A*, 337:275–286, September 1998.
- S. S. R. Offner, P. C. Clark, P. Hennebelle, N. Bastian, M. R. Bate, P. F. Hopkins, E. Moraux, and A. P. Whitworth. The Origin and Universality of the Stellar Initial Mass Function. In Henrik Beuther, Ralf S. Klessen, Cornelis P. Dullemond, and Thomas Henning, editors, *Protostars and Planets VI*, pages 53–75, January 2014. doi: 10.2458/azu_uapress.9780816531240-ch003.
- C. Olczak, R. Spurzem, and Th. Henning. A highly efficient measure of mass segregation in star clusters. *A&A*, 532:A119, August 2011. doi: 10.1051/0004-6361/201116902.
- V. Ossenkopf and Th. Henning. Dust opacities for protostellar cores. *A&A*, 291:943–959, November 1994.
- Aina Palau, Javier Ballesteros-Paredes, Enrique Vázquez-Semadeni, Álvaro Sánchez-Monge, Robert Estalella, S. Michael Fall, Luis A. Zapata, Vianey Camacho, Laura Gómez, Raúl Naranjo-Romero, Gemma Busquet, and Francesco Fontani. Gravity or turbulence? - III. Evidence of pure thermal Jeans fragmentation at ~ 0.1 pc scale. *MNRAS*, 453(4):3785–3797, November 2015. doi: 10.1093/mnras/stv1834.
- Richard J. Parker. On the spatial distributions of dense cores in Orion B. *MNRAS*, 476(1): 617–629, May 2018. doi: 10.1093/mnras/sty249.
- T. J. Pearson and A. C. S. Readhead. Image Formation by Self-Calibration in Radio Astronomy. *ARA&A*, 22:97–130, January 1984. doi: 10.1146/annurev.aa.22.090184.000525.
- Adele Plunkett, Alvaro Hacar, Lydia Moser-Fischer, Dirk Petry, Peter Teuben, Nickolas Pingel, Devaky Kunneriath, Toshinobu Takagi, Yusuke Miyamoto, Emily Moravec, Sümeyye Suri, Kelley M. Hess, Melissa Hoffman, and Brian Mason. Data Combination: Interferometry and Single-dish Imaging in Radio Astronomy. *PASP*, 135(1045):034501, March 2023. doi: 10.1088/1538-3873/acb9bd.
- Adele L. Plunkett, Manuel Fernández-López, Héctor G. Arce, Gemma Busquet, Diego Mar-dones, and Michael M. Dunham. Distribution of Serpens South protostars revealed with ALMA. *A&A*, 615:A9, July 2018. doi: 10.1051/0004-6361/201732372.
- Cormac R. Purcell. URL <http://web.science.mq.edu.au/~cpurcell/public/index.php>.

- Sheng-Li Qin, Tie Liu, Xunchuan Liu, Paul F. Goldsmith, Di Li, Qizhou Zhang, Hong-Li Liu, Yuefang Wu, Leonardo Bronfman, Mika Juvela, Chang Won Lee, Guido Garay, Yong Zhang, Jinhua He, Shih-Ying Hsu, Zhi-Qiang Shen, Jeong-Eun Lee, Ke Wang, Ningyu Tang, Mengyao Tang, Chao Zhang, Yinghua Yue, Qiaowei Xue, Shanghuo Li, Yaping Peng, Somnath Dutta, Jixing Ge, Fengwei Xu, Long-Fei Chen, Tapas Baug, Lokesh Dewangan, and Anandmayee Tej. ATOMS: ALMA Three-millimeter Observations of Massive Star-forming regions - VIII. A search for hot cores by using $\text{C}_2\text{H}_5\text{CN}$, CH_3OCHO , and CH_3OH lines. *MNRAS*, 511(3):3463–3476, April 2022. doi: 10.1093/mnras/stac219.
- V. M. Rivilla, J. Martín-Pintado, I. Jiménez-Serra, and A. Rodríguez-Franco. The role of low-mass star clusters in massive star formation. The Orion case. *A&A*, 554:A48, June 2013. doi: 10.1051/0004-6361/201117487.
- J. R. Rizzo, A. Fuente, and S. García-Burillo. A Chemical Study of the PDR around Mon R2. In C. Lidman and D. Alloin, editors, *The Cool Universe: Observing Cosmic Dawn*, volume 344 of *Astronomical Society of the Pacific Conference Series*, page 184, December 2005.
- B. J. Robinson, J. L. Caswell, and W. M. Goss. A 1665 MHz OH Survey of the Southern Milky Way. *Astrophysics Letters*, 9:5, August 1971.
- A. W. Rodgers, C. T. Campbell, and J. B. Whiteoak. A catalogue of $\text{H}\alpha$ -emission regions in the southern Milky Way. *MNRAS*, 121:103, January 1960. doi: 10.1093/mnras/121.1.103.
- R. Rolfs, P. Schilke, F. Wyrowski, K. M. Menten, R. Güsten, and S. E. Bisschop. Structure of evolved cluster-forming regions. *A&A*, 527:A68, March 2011. doi: 10.1051/0004-6361/201015367.
- M. Sadaghiani, Á. Sánchez-Monge, P. Schilke, H. B. Liu, S. D. Clarke, Q. Zhang, J. M. Girart, D. Seifried, A. Aghababaei, H. Li, C. Juárez, and K. S. Tang. Physical properties of the star-forming clusters in NGC 6334. A study of the continuum dust emission with ALMA. *A&A*, 635:A2, March 2020. doi: 10.1051/0004-6361/201935699.
- Edwin E. Salpeter. The Luminosity Function and Stellar Evolution. *ApJ*, 121:161, January 1955. doi: 10.1086/145971.
- Á. Sánchez-Monge, P. Schilke, A. Schmiedeke, A. Ginsburg, R. Cesaroni, D. C. Lis, S. L. Qin, H. S. P. Müller, E. Bergin, C. Comito, and Th. Möller. The physical and chemical structure of Sagittarius B2. II. Continuum millimeter emission of Sgr B2(M) and Sgr B2(N) with ALMA. *A&A*, 604:A6, July 2017. doi: 10.1051/0004-6361/201730426.
- Á. Sánchez-Monge, P. Schilke, A. Ginsburg, R. Cesaroni, and A. Schmiedeke. STATCONT: A statistical continuum level determination method for line-rich sources. *A&A*, 609:A101, January 2018. doi: 10.1051/0004-6361/201730425.
- U. Sawangwit, T. Shanks, F. B. Abdalla, R. D. Cannon, S. M. Croom, A. C. Edge, Nicholas P. Ross, and D. A. Wake. Angular correlation function of 1.5 million luminous red galaxies: clustering evolution and a search for baryon acoustic oscillations. *MNRAS*, 416(4):3033–3056, October 2011. doi: 10.1111/j.1365-2966.2011.19251.x.
- P. Schilke. High-Mass Star Formation. In *EAS Publications Series*, volume 75-76 of *EAS Publications Series*, pages 227–235, May 2015. doi: 10.1051/eas/1575046.
- P. Schilke, C. Comito, S. Thorwirth, F. Wyrowski, K. M. Menten, R. Güsten, P. Bergman, and L. Å. Nyman. Submillimeter spectroscopy of southern hot cores: NGC 6334(I) and G327.3-0.6. *A&A*, 454(2):L41–L45, August 2006. doi: 10.1051/0004-6361:20065398.

- F. Schuller, K. M. Menten, Y. Contreras, F. Wyrowski, P. Schilke, L. Bronfman, T. Henning, C. M. Walmsley, H. Beuther, S. Bontemps, R. Cesaroni, L. Deharveng, G. Garay, F. Herpin, B. Lefloch, H. Linz, D. Mardones, V. Minier, S. Molinari, F. Motte, L. Å. Nyman, V. Reveret, C. Risacher, D. Russeil, N. Schneider, L. Testi, T. Troost, T. Vasyunina, M. Wienen, A. Zavagno, A. Kovacs, E. Kreysa, G. Siringo, and A. Weiß. ATLASGAL - The APEX telescope large area survey of the galaxy at 870 μ m. *A&A*, 504(2):415–427, September 2009. doi: 10.1051/0004-6361/200811568.
- Marta Sewilo, Remy Indebetouw, Steven B. Charnley, Sarolta Zahorecz, Joana M. Oliveira, Jacco Th. van Loon, Jacob L. Ward, C. H. Rosie Chen, Jennifer Wiseman, Yasuo Fukui, Akiko Kawamura, Margaret Meixner, Toshikazu Onishi, and Peter Schilke. The Detection of Hot Cores and Complex Organic Molecules in the Large Magellanic Cloud. *ApJL*, 853(2): L19, February 2018. doi: 10.3847/2041-8213/aaa079.
- F. H. Shu. Self-similar collapse of isothermal spheres and star formation. *ApJ*, 214:488–497, June 1977. doi: 10.1086/155274.
- Janet P. Simpson, Angela Cotera, and Barbara Whitney. Characterizing YSO Evolutionary States from their Mid-Infrared Colors. In *American Astronomical Society Meeting Abstracts #224*, volume 224 of *American Astronomical Society Meeting Abstracts*, page 223.05, June 2014.
- Rowan J. Smith, Robin G. Treß, Mattia C. Sormani, Simon C. O. Glover, Ralf S. Klessen, Paul C. Clark, Andrés F. Izquierdo, Ana Duarte-Cabral, and Catherine Zucker. The Cloud Factory I: Generating resolved filamentary molecular clouds from galactic-scale forces. *MNRAS*, 492(2):1594–1613, February 2020. doi: 10.1093/mnras/stz3328.
- Theodore P. Snow and Benjamin J. McCall. Diffuse Atomic and Molecular Clouds. *ARA&A*, 44(1):367–414, September 2006. doi: 10.1146/annurev.astro.43.072103.150624.
- J. D. Soler, P. Hennebelle, P. G. Martin, M. A. Miville-Deschênes, C. B. Netterfield, and L. M. Fissel. An Imprint of Molecular Cloud Magnetization in the Morphology of the Dust Polarized Emission. *ApJ*, 774(2):128, September 2013. doi: 10.1088/0004-637X/774/2/128.
- J. D. Soler, P. A. R. Ade, F. E. Angilè, P. Ashton, S. J. Benton, M. J. Devlin, B. Dober, L. M. Fissel, Y. Fukui, N. Galitzki, N. N. Gandilo, P. Hennebelle, J. Klein, Z. Y. Li, A. L. Korotkov, P. G. Martin, T. G. Matthews, L. Monceli, C. B. Netterfield, G. Novak, E. Pascale, F. Poidevin, F. P. Santos, G. Savini, D. Scott, J. A. Shariff, N. E. Thomas, C. E. Tucker, G. S. Tucker, and D. Ward-Thompson. The relation between the column density structures and the magnetic field orientation in the Vela C molecular complex. *A&A*, 603:A64, July 2017. doi: 10.1051/0004-6361/201730608.
- J. D. Soler, H. Beuther, M. Rugel, Y. Wang, P. C. Clark, S. C. O. Glover, P. F. Goldsmith, M. Heyer, L. D. Anderson, A. Goodman, Th. Henning, J. Kainulainen, R. S. Klessen, S. N. Longmore, N. M. McClure-Griffiths, K. M. Menten, J. C. Mottram, J. Ott, S. E. Ragan, R. J. Smith, J. S. Urquhart, F. Bigiel, P. Hennebelle, N. Roy, and P. Schilke. Histogram of oriented gradients: a technique for the study of molecular cloud formation. *A&A*, 622:A166, February 2019. doi: 10.1051/0004-6361/201834300.
- Brian E. Svoboda, Yancy L. Shirley, Alessio Traficante, Cara Battersby, Gary A. Fuller, Qizhou Zhang, Henrik Beuther, Nicolas Peretto, Crystal Brogan, and Todd Hunter. ALMA Observations of Fragmentation, Substructure, and Protostars in High-mass Starless Clump Candidates. *ApJ*, 886(1):36, November 2019. doi: 10.3847/1538-4357/ab40ca.

- Michihiro Takami, Jennifer L. Karr, Haegon Koh, How-Huan Chen, and Hsu-Tai Lee. A Detailed Study of Spitzer-IRAC Emission in Herbig-Haro Objects. I. Morphology and Flux Ratios of Shocked Emission. *ApJ*, 720(1):155–172, September 2010. doi: 10.1088/0004-637X/720/1/155.
- X. D. Tang, C. Henkel, C. H. R. Chen, K. M. Menten, R. Indebetouw, X. W. Zheng, J. Esimbek, J. J. Zhou, Y. Yuan, D. L. Li, and Y. X. He. Kinetic temperature of massive star-forming molecular clumps measured with formaldehyde. II. The Large Magellanic Cloud. *A&A*, 600:A16, April 2017a. doi: 10.1051/0004-6361/201630183.
- X. D. Tang, C. Henkel, K. M. Menten, X. W. Zheng, J. Esimbek, J. J. Zhou, C. C. Yeh, C. König, Y. Yuan, Y. X. He, and D. L. Li. Kinetic temperature of massive star forming molecular clumps measured with formaldehyde. *A&A*, 598:A30, February 2017b. doi: 10.1051/0004-6361/201629694.
- A. Richard Thompson, James M. Moran, and Jr. Swenson, George W. *Interferometry and Synthesis in Radio Astronomy, 3rd Edition*. 2017. doi: 10.1007/978-3-319-44431-4.
- J. S. Urquhart, M. G. Hoare, S. L. Lumsden, R. D. Oudmaijer, T. J. T. Moore, J. C. Mottram, H. D. B. Cooper, M. Mottram, and H. C. Rogers. The RMS survey: resolving kinematic distance ambiguities towards a sample of compact H ii regions using H i absorption*. *MNRAS*, 420(2):1656–1672, February 2012. doi: 10.1111/j.1365-2966.2011.20157.x.
- F. F. S. van der Tak, E. F. van Dishoeck, and P. Caselli. Abundance profiles of CH₃OH and H₂CO toward massive young stars as tests of gas-grain chemical models. *A&A*, 361:327–339, September 2000. doi: 10.48550/arXiv.astro-ph/0008010.
- Ewine F. van Dishoeck and Geoffrey A. Blake. Chemical Evolution of Star-Forming Regions. *ARA&A*, 36:317–368, January 1998. doi: 10.1146/annurev.astro.36.1.317.
- V. S. Veena, S. Vig, B. Mookerjee, Á. Sánchez-Monge, A. Tej, and C. H. Ishwara-Chandra. Probing the Massive Star-forming Environment: A Multiwavelength Investigation of the Filamentary IRDC G333.73+0.37. *ApJ*, 852(2):93, January 2018. doi: 10.3847/1538-4357/aa9aef.
- Ke Wang, Qizhou Zhang, Yuefang Wu, and Huawei Zhang. Hierarchical Fragmentation and Jet-like Outflows in IRDC G28.34+0.06: A Growing Massive Protostar Cluster. *ApJ*, 735(1):64, July 2011. doi: 10.1088/0004-637X/735/1/64.
- Ke Wang, Qizhou Zhang, Leonardo Testi, Floris van der Tak, Yuefang Wu, Huawei Zhang, Thushara Pillai, Friedrich Wyrowski, Sean Carey, Sarah E. Ragan, and Thomas Henning. Hierarchical fragmentation and differential star formation in the Galactic ‘Snake’: infrared dark cloud G11.11-0.12. *MNRAS*, 439(4):3275–3293, April 2014. doi: 10.1093/mnras/stu127.
- Derek Ward-Thompson and Anthony P. Whitworth. *An Introduction to Star Formation*. 2011.
- J. B. Whiteoak and F. F. Gardner. Observations of the J= 1-0 transition of HCN in southern molecular clouds. *MNRAS*, 185:33P–624, November 1978. doi: 10.1093/mnras/185.1.33P.
- Thomas L. Wilson, Kristen Rohlfs, and Susanne Hüttemeister. *Tools of Radio Astronomy*. 2013. doi: 10.1007/978-3-642-39950-3.
- F. Wyrowski, K. M. Menten, P. Schilke, S. Thorwirth, R. Güsten, and P. Bergman. Revealing the environs of the remarkable southern hot core G327.3-0.6. *A&A*, 454(2):L91–L94, August 2006. doi: 10.1051/0004-6361:20065347.

- Jinghua Yuan, Yuefang Wu, Tie Liu, Tianwei Zhang, Jin Zeng Li, Hong-Li Liu, Fanyi Meng, Ping Chen, Runjie Hu, and Ke Wang. Dense Gas in Molecular Cores Associated with Planck Galactic Cold Clumps. *ApJ*, 820(1):37, March 2016. doi: 10.3847/0004-637X/820/1/37.
- Jinghua Yuan, Jin-Zeng Li, Yuefang Wu, Simon P. Ellingsen, Christian Henkel, Ke Wang, Tie Liu, Hong-Li Liu, Annie Zavagno, Zhiyuan Ren, and Ya-Fang Huang. High-mass Star Formation through Filamentary Collapse and Clump-fed Accretion in G22. *ApJ*, 852(1):12, January 2018. doi: 10.3847/1538-4357/aa9d40.
- Guo-Yin Zhang, Jin-Long Xu, A. I. Vasyunin, D. A. Semenov, Jun-Jie Wang, Sami Dib, Tie Liu, Sheng-Yuan Liu, Chuan-Peng Zhang, Xiao-Lan Liu, Ke Wang, Di Li, Zhong-Zu Wu, Jing-Hua Yuan, Da-Lei Li, and Yang Gao. Physical properties and chemical composition of the cores in the California molecular cloud. *A&A*, 620:A163, December 2018. doi: 10.1051/0004-6361/201833622.
- Qizhou Zhang, Ke Wang, Xing Lu, and Izaskun Jiménez-Serra. Fragmentation of Molecular Clumps and Formation of a Protocluster. *ApJ*, 804(2):141, May 2015. doi: 10.1088/0004-637X/804/2/141.

Appendix

Table 7.1: Hot core fitting results

Molecule	size					Temperature			Column Density			Line width			Velocity						
										cm ⁻³			km s ⁻¹			km s ⁻¹					
H ₂ S	v=0	1	n	0.453	n	3.0	1000.0	200	y	1.00e+12	1.00e+20	1.68e+18	y	1.0	10.0	7.13	y	-10.0	10.0	2.09	c
C-13-S	v=0	1	n	0.453	n	3.0	1000.0	200	y	1.00e+12	1.00e+20	4.15e+16	y	1.0	10.0	4.91	y	-10.0	10.0	0.99	c
SiO	v=0	1	n	0.453	n	3.0	1000.0	200	y	1.00e+12	1.00e+20	3.85e+15	y	1.0	10.0	9.53	y	-10.0	10.0	3.87	c
SiS	v=0	1	n	0.453	n	3.0	1000.0	200	y	1.00e+12	1.00e+20	1.08e+16	y	1.0	10.0	4.94	y	-10.0	10.0	4.94	c
OCS	v=0	1	n	0.453	y	3.0	1000.0	263	y	1.00e+12	1.00e+20	7.64e+18	y	1.0	10.0	4.47	y	-10.0	10.0	0.84	c
DCN	v=0	1	n	0.453	y	3.0	1000.0	368	y	1.00e+12	1.00e+20	5.15e+15	y	1.0	10.0	5.48	y	-10.0	10.0	0.86	c
CH ₃ OH	v=0	1	n	0.453	y	3.0	1000.0	269	y	1.00e+12	1.00e+20	1.63e+19	y	1.0	10.0	5.70	y	-10.0	10.0	0.89	c
HCCCN	v=0	1	n	0.453	y	3.0	1000.0	215	y	1.00e+12	1.00e+20	5.01e+17	y	1.0	10.0	5.62	y	-10.0	10.0	1.31	c
HNCO	v=0	1	n	0.453	y	3.0	1000.0	313	y	1.00e+12	1.00e+20	4.59e+17	y	1.0	10.0	10.51	y	-10.0	10.0	0.79	c
NH ₂ D	v=0	1	n	0.453	y	3.0	1000.0	147	y	1.00e+12	1.00e+20	2.46e+17	y	1.0	10.0	8.01	y	-10.0	10.0	6.17	c
SO ₂	v=0	1	n	0.453	y	3.0	1000.0	96	y	1.00e+12	1.00e+20	2.50e+18	y	1.0	10.0	8.64	y	-10.0	10.0	2.63	c
CH ₃ OCHO	v=0	1	n	0.453	y	3.0	1000.0	280	y	1.00e+12	1.00e+20	3.60e+18	y	1.0	10.0	5.59	y	-10.0	10.0	0.89	c
CH ₃ CHO	v=0	1	n	0.453	y	3.0	1000.0	140	y	1.00e+12	1.00e+20	1.45e+17	y	1.0	10.0	4.28	y	-10.0	10.0	0.53	c
CH ₃ OCH ₃	v=0	1	n	0.453	y	3.0	1000.0	44	y	1.00e+12	1.00e+20	4.73e+18	y	1.0	10.0	5.23	y	-10.0	10.0	1.09	c
C ₂ H ₅ CN	v=0	1	n	0.453	y	3.0	1000.0	247	y	1.00e+12	1.00e+20	9.48e+17	y	1.0	10.0	4.29	y	-10.0	10.0	0.74	c
C ₂ H ₃ CN	v=0	1	n	0.453	y	3.0	1000.0	174	y	1.00e+12	1.00e+20	3.97e+17	y	1.0	10.0	5.22	y	-10.0	10.0	0.17	c
C ₂ H ₅ OH	v=0	1	n	0.453	y	3.0	1000.0	261	y	1.00e+12	1.00e+20	1.45e+18	y	1.0	10.0	8.95	y	-10.0	10.0	3.28	c
CH ₃ COCH ₃	v=0	1	n	0.453	y	3.0	1000.0	187	y	1.00e+12	1.00e+20	1.32e+18	y	1.0	10.0	4.68	y	-10.0	10.0	0.70	c
CH ₃ SH	v=0	1	n	0.453	y	3.0	1000.0	3	y	1.00e+12	1.00e+20	4.60e+16	y	1.0	10.0	8.44	y	-10.0	10.0	2.46	c
c-C ₆ H ₅ CHO	v=0	1	n	0.453	y	3.0	1000.0	112	y	1.00e+12	1.00e+20	1.07e+18	y	1.0	10.0	18.81	y	-10.0	10.0	0.54	c
HCOOH	v=0	1	n	0.453	y	3.0	1000.0	148	y	1.00e+12	1.00e+20	2.24e+18	y	1.0	10.0	5.18	y	-10.0	10.0	5.01	c
CH ₃ OOH	v=0	1	n	0.453	y	3.0	1000.0	170	y	1.00e+12	1.00e+20	1.68e+18	y	1.0	10.0	13.07	y	-10.0	10.0	4.38	c
H ₂ CS	v=0	1	n	0.453	n	3.0	1000.0	200	y	1.00e+12	1.00e+20	1.06e+19	y	1.0	10.0	11.51	y	-10.0	10.0	1.27	c
C ₂ H ₃ NC	v=0	1	n	0.453	y	3.0	1000.0	283	y	1.00e+12	1.00e+20	2.03e+16	y	1.0	10.0	7.47	y	-10.0	10.0	2.07	c
c-C ₂ H ₄ O	v=0	1	n	0.453	y	3.0	1000.0	464	y	1.00e+12	1.00e+20	3.17e+17	y	1.0	10.0	18.03	y	-10.0	10.0	0.35	c
HOCHCHCHO	v=0	1	n	0.453	y	3.0	1000.0	474	y	1.00e+12	1.00e+20	6.00e+17	y	1.0	10.0	18.62	y	-10.0	10.0	7.14	c

Table 7.2: Hot core isotopologue ratio

Isotopologue	Iso-master	Iso-ratio (master/iso)
CH ₃ OH;v12=1	CH ₃ OH;v=0	8.854346217988603E-01
¹³ CH ₃ OH;v=0	CH ₃ OH;v=0	6.000000000000000E+01
CH ₃ O-18-H;v=0	CH ₃ OH;v=0	5.000000000000000E+02
CH ₃ OCHO;v18=1	CH ₃ OCHO;v=0	1.497031024243430E+00
CH ₃ O ¹³ CHO;v=0	CH ₃ OCHO;v=0	6.000000000000000E+01
CH ₃ O ¹³ CHO;v18=1	CH ₃ OCHO;v=0	6.000000000000000E+01
CH ₃ CHO;v15=1	CH ₃ CHO;v=0	8.795699407663076E-01
CH ₃ CHO;v15=2	CH ₃ CHO;v=0	1.000000000000000E+00
CH ₃ CHO;v15=2-0	CH ₃ CHO;v=0	1.000000000000000E+00
CH ₃ CHO;v15=2-1	CH ₃ CHO;v=0	1.000000000000000E+00
CH ₃ CHO;v15=1-0	CH ₃ CHO;v=0	1.000000000000000E+00
¹³ CH ₃ CH ₂ CN;v=0	C ₂ H ₅ CN;v=0	6.000000000000000E+01
C ₂ H ₅ CN-15;v=0	C ₂ H ₅ CN;v=0	3.000000000000000E+02
CH ₃ ¹³ CH ₂ CN;v=0	C ₂ H ₅ CN;v=0	6.000000000000000E+01
C ₂ H ₅ ¹³ CN;v=0	C ₂ H ₅ CN;v=0	6.000000000000000E+01
¹³ CH ₃ ¹³ CH ₂ CN;v=0	C ₂ H ₅ CN;v=0	3.600000000000000E+02
¹³ CH ₃ CH ₂ ¹³ CN;v=0	C ₂ H ₅ CN;v=0	3.600000000000000E+02
C ₂ H ₃ CN;v11=1	C ₂ H ₃ CN;v=0	4.874359582645277E-01
C ₂ H ₃ CN;v15=1	C ₂ H ₃ CN;v=0	4.987518097799171E-01
CH ₂ ¹³ CHCN;v=0	C ₂ H ₃ CN;v=0	6.000000000000000E+01
¹³ CH ₂ CHCN;v=0	C ₂ H ₃ CN;v=0	6.000000000000000E+01
CH ₂ CH ¹³ CN;v=0	C ₂ H ₃ CN;v=0	6.000000000000000E+01
H ₂ CCH ¹³ CN;v=0	C ₂ H ₃ CN;v=0	6.000000000000000E+01
HCCCN;v7=1	HCCCN;v=0	1.060184605915290E+00
HCCCN;v7=4	HCCCN;v=0	1.060184605915290E+00
H ¹³ CCCN;v7=1	HCCCN;v=0	6.000000000000000E+01
HC ¹³ CCN;v7=1	HCCCN;v=0	6.000000000000000E+01
HC ¹³ CCN;v7=2	HCCCN;v=0	6.000000000000000E+01
HCC ¹³ CN;v7=1	HCCCN;v=0	6.000000000000000E+01
HCC ¹³ CN;v7=2	HCCCN;v=0	6.000000000000000E+01
HCCCN;v4=1	HCCCN;v=0	9.455812460998602E-01
HCCCN;v5=1	HCCCN;v=0	1.812334895614113E+00
HCCCN;v6=1	HCCCN;v=0	2.250687501077173E+00
HCCCN;v7=2	HCCCN;v=0	1.178034445086011E+00
H ¹³ CCCN;v7=2	HCCCN;v=0	6.000000000000000E+01
OC ³³ S;v=0	OCS;v=0	1.270000000000000E+02
¹⁸ OCS;v=0	OCS;v=0	5.000000000000000E+02
O ¹³ CS;v=0	OCS;v=0	6.000000000000000E+01
³⁴ SO ₂ ;v=0	SO ₂ ;v=0	2.300000000000000E+01

Acknowledgement

Above all, this thesis was supported by the Collaborative Research Centre 956, sub-projects A6, funded by the Deutsche Forschungsgemeinschaft (DFG) - project ID 184018867.

Words are too pale to express my thanks to many people who have helped me, supported me, and inspired me during my study of PhD. I'm grateful to the happiness and tears on the way to PhD, while I can only note down very few of the memory here.

- I would like to especially thank my supervisor, Prof. Dr. Peter Schilke for his great guidance and patience throughout my PhD years.
- Thanks to Dr. Álvaro Sánchez-Monge for always being there to answer my random questions and cheering me up. His clear instructions and explanations impressed me a lot, as well as the personality of generosity and selflessness, which becomes a lifelong role model for me.
- Thanks Dr Thomas Moeller for the handful tool XCLASS, as well as the clear manual and quick response whenever I encountered with problems.
- Thanks Dr. Bethany Jones for their detailed correction of the thesis.
- Thanks Dr. Wonju Kim and Prof. Dr. Gary Fuller for sharing the experience in an advanced stage of academia and encouraging me to go further, especially during the Covid-19 quarantine.
- For all the group members, thank you for accompanying, the relaxing lunch time, and the inspiring environment.
- Thanks to my family for keeping asking me about the progress of research, which guides me to a quicker graduation and deeper understanding of popular science education. I love you all.
- Thanks to the people I met along the way, without you I wouldn't be here. You have my gratitude.

Erklärung zur Dissertation

gemäß der Promotionsordnung vom 12. März 2020

***Diese Erklärung muss in der Dissertation enthalten sein.
(This version must be included in the doctoral thesis)***

„Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel und Literatur angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Werken dem Wortlaut oder dem Sinn nach entnommen wurden, sind als solche kenntlich gemacht. Ich versichere an Eides statt, dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen und eingebundenen Artikeln und Manuskripten - noch nicht veröffentlicht worden ist sowie, dass ich eine Veröffentlichung der Dissertation vor Abschluss der Promotion nicht ohne Genehmigung des Promotionsausschusses vornehmen werde. Die Bestimmungen dieser Ordnung sind mir bekannt. Darüber hinaus erkläre ich hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten der Universität zu Köln gelesen und sie bei der Durchführung der Dissertation zugrundeliegenden Arbeiten und der schriftlich verfassten Dissertation beachtet habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen. Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht.“

Teilpublikationen:

Datum, Name und Unterschrift

19.06.2023, Tianwei Zhang

Tianwei Zhang