ALMA MATER STUDIORUM - UNIVERSITY OF BOLOGNA

Department of Physics and Astronomy "Augusto Righi"

Second-cycle/ Master's Degree

in

ASTROPHYSICS AND COSMOLOGY

CONSTRAINING THE CHEMICAL CONDITIONS OF IRAS 4A2 PROTOSTAR AT PLANET-FORMING SCALES

DEFENDED BY

Jenny Frediani

SUPERVISORS

Prof. Leonardo Testi

Dr. Marta De Simone

Graduation session of September

Academic year [2022/2023]

This thesis work was carried out
at ESO Headquarters
in Garching bei München, Germany,
from 27 th February to 3 rd June 2023.

Contents

Abstract

The organic chemistry, with its chains and rings of carbon (C), builds life as we know it on the planet Earth. Primitive Solar System objects, such as comets and asteroids, show traces of prebiotic organic compounds. Therefore, it is natural to think that the physical and chemical conditions to produce and store complex organic molecules was wide spread in the early Solar System. Studying planet-forming environments may thus enlighten how common these conditions are across the Galaxy, and eventually leading to life in other planetary systems similar to our.

There is recent evidence that the planet formation may already start during the very early stage of star formation (Class 0/I protostars $\sim 10^{4-5}$ yr old). Many complex organic molecules, afterwards called Interstellar Complex Organic Molecules (iCOMs), have been discovered in Class 0/I hot, dense, inner regions at < 100 AU, the so-called "hot corinos". Constraining the abundance and the spatial distribution of iCOMs in such peculiar environments is of crucial importance to reconstruct the path followed by organic chemistry at the dawn of Solar-like exoplanetary systems.

In this thesis project, I present the first results of the ALMA Large Program FAUST (PI. S. Yamamoto) observations towards the well-known IRAS 4A2 hot corino located in the Perseus/NGC 1333 cloud ($\sim$ 300 pc from us). Exploiting the excellent angular resolution ($<$ 50 AU) and sensitivity of ALMA interferometric radiotelescope, FAUST data allowed me to sample, for the first time, the planet-forming scales of this Solar-like young protostar, resolving its hot corino emission and leaving room for investigation of its chemical content.

On the one hand, I imaged a sample of O-bearing and N-bearing iCOMs (including methanol CH_3OH, methyl formate $HCOOCH_3$, acetaldehyde CH_3CHO, formamide, NH_2CHO, and glycolaldehyde CH_2OHCHO), and compared their spatial distribution. On the other hand, using a radiative transfer analysis under the Local Thermal Equilibrium (LTE) and optically thin lines assumptions, I derived the physical properties of the iCOMs emitting gas, namely, gas temperature and molecular species column density. I found a spatial segregation of iCOMs and temperature non uniformity in the hot corino region, with glycolaldehyde and methanol emitting respectively from the most compact (r $\sim$ 20 AU) and extended (r $\sim$ 40 AU) scales.

Combining together the information on the spatial segregation of iCOMs and that on the gas temperature traced by these molecules, I derived the first observational and model independent gas temperature profile of IRAS 4A2 Class 0 protostar at scales $<$ 50 AU, validating some crucial expectations for disk-like protostellar temperature structures.

Keywords— astrochemistry, iCOMs, hot corino, young solar analogs, protostars, star formation, planet formation, ALMA, FAUST, radio astronomy, molecular spectroscopy

CHAPTER 1

Scientific Background

Figure 1.1: Artistic view of some intriguing exoplanets discovered in the last decades. *Figure adapted from: NASA/JPL-CALTECH.*

[1]Planetary science has witnessed an unprecedented revolution in the past decades, as the scientific community started realizing the existence of many thousands of exoplanets[2] (some of them, artistically depicted in Fig. 1.1) orbiting around other stars, beyond the Sun and its eight known-so-far planets, Earth included (Winn & Fabrycky 2015). What perhaps has been even more revolutionary was to realize that many of these exoplanets seem to be "unconventional" compared to the Solar System standards. As example, the puzzling family of *Hot Jupiters*, which are similar in size to Jupiter but orbit very close to their host stars, often much closer than Mercury is to the Sun (Dawson & Johnson 2018). At the same time, another subgroup of exoplanets show "familiar" properties to us, in the sense that they are rocky and apparently at the right temperature to sustain liquid water, like the Earth. These exoplanets are now referred in the literature as potentially *habitable* exoplanets (Kasting et al. 1993; Winn & Fabrycky 2015). Yet,

[1]Main references for this chapter: Caselli & Ceccarelli (2012); Öberg (2016); van Dishoeck (2018); van Dishoeck & Bergin (2020); Öberg & Bergin (2021); Ceccarelli et al. (2022).

[2]https://exoplanets.nasa.gov

whether they are able to form with water and host an increasingly complex organic chemistry, to build Earth-like life compounds, remains a mistery.

As we gain knowledge on exogenous planetary systems, it is becoming clear that the planetary composition represents a crucial tool to constrain the origins of individual planets, as well as families of them, and to assess the fundamental question of life in the Universe. Investigate the composition of planets requires a detailed understanding of the elemental composition of the gas, and of solids from which they form.

We therefore have to take a step back and study the environments where planets form.

In a nutshell, they form in disks of gas and dust around young stars, that is, newborn stars which have not ignited yet the hydrogen burning in their cores through nuclear fusion. Combining together theory and direct observations of these disks with data collected from our own Solar System, we find that disk compositions are in their turn a product of both chemical processes occuring *in situ*, and heredity of molecular material from the previous evolutionary stages. This implies that we need to get a comprehensive view of the astrochemical origins of planetary compositions that includes the chemical evolution from the onset of molecular clouds formation, to the final stages of planet formation in disks (Caselli & Ceccarelli 2012; Öberg & Bergin 2021).

Such a view can be achieved in the context of Astrochemistry, which studies the formation, excitation, and destruction of molecules in astronomical environments and their influence on the surrounding astronomical objects. In this direction, molecules are also recognized as excellent diagnostics of the physical conditions and processes characterizing the sky regions where they are observed.

From the pure chemical perspective, the *interstellar medium* (ISM), provides a unique environment in which molecular behavior can be studied under extreme conditions (of low density and low temperature), given that the ISM is permeated by high-energy radiation fields (Tielens 2005; Draine 2011).

At the time of writing, more than 300 different molecules have been detected in the interstellar space (in the Galaxy and extragalactic sources), spanning from the simplest (2 atoms only), most abundant ones, such as molecular hydrogen H_2 and carbon monoxide CO, to aggregates of more than twelve atoms, like the exotic benzonitril c–C_6H_5CN (McGuire 2018, 2022).

Astrochemistry comprehends atomic and molecular spectroscopic data, deep knowledge of chemical processes, also involving different classes of molecules and reactions that are not normally considered on the Earth, and laboratory experiments. The latter implies actual experiments as well as quantum chemical and molecular dynamics calculations. From this point of view, the key is to have enough chemical physics knowledge of what we can, and what cannot, do in the laboratory. The view that emerges is that of a vibrant, multidisciplinary field of Astrophysics, which claims understanding not only in chemistry, physics and mathematics, which have always been integral parts of this field, but also in biology, geochemistry and informatics, hence involving a raising number of personalities with different scientific background.

When it comes to Astrochemistry concerning molecules, which are the main characters of my thesis, we can question a lot of things. When and where are molecules created and excited, i.e. brought to emit and absorb radiation? What do they tell us about the physical properties and the dynamics of the stellar nurseries across the Universe, the molecular clouds? What is their inheritance path through the various phases of stellar evolution, from birth to death of stars? How far does chemical complexity go in space? And, putting it on a different level, can forming planetary systems harbour interstellar molecules to build life elsewhere?

In this thesis project, I try to give a little contribution to solve this interrogative by approaching, from an astrochemical perspective, the molecular complexity of the inner envelope surrounding a Solar-like young star, which could be inherited by forming planets.

Considering the vastness of the involved topics, and the natural brevity of a thesis work, in the following sections I will only give a state-of-the-art overview on the stages of Solar-like star and planet formation, in § 1.1, and then have a better look at the current picture on the origins and

differentiation of the chemical enrichment we observe around some young stars similar to the Sun (§ 1.2). Then, I will introduce the specific protostar, for which I conduced an observational data analysis with the ALMA/FAUST Large Program (for major details, jump to § 2.3), and what can be seen in its sky background (§ 1.3). Finally, I will expose the objectives of the thesis (§ 1.4), to make sense of the analysis and results discussion made in the next chapters.

1.1 The Zoo Of Protostellar Environments

Stars are found to cover a large range of masses, from $\sim 0.08\ M_\odot$ to $\sim 100\ M_\odot$, and the process of star formation proceeds differently depending on the mass of the final star. However, since planets must be less massive than stars, for the purpose of predicting planetary compositions we are particularly interested in studying the formation of low-mass stars, or Solar-like stars ($\leq 2\ M_\odot$) (Carroll & Ostlie 2007). It is actually quite convenient to have to deal with them, because they constitute the vast majority of stars in the Milky Way. Moreover, considering how much time we think life needed to take place on the Earth (nearly 500 Myr after its formation (Maas et al. 1992)), their long life times ($\sim$ 10 Gyr for a 1 $M_\odot$ star (Carroll & Ostlie 2007)) are likely a pre-requisite for the origins of life.

Consequently, the most solid theory on the planet formation concerns planets forming around young Solar-like stars, also called Young Solar Objects (YSOs; Lada (1987); André & Montmerle (1994); André et al. (2000)) and we will refer to this theory hereafter.

In short, as artistically summarized in Fig. 1.2, planetary systems like our own form from the condensation of matter in diffuse clouds to dense bodies as stars, planets, asteroids, comets and residual interstellar dust, experiencing dramatic changes in terms of both physical and chemical structure in less than 10^7 yr (André et al. 2000; Luhman et al. 2016).

Physical and chemical changes proceed hand in hand. In other words, as the primordial molecular cloud evolves into a *protostellar envelope*, *protoplanetary disk* and finally into a planetary system, the chemical composition of the primordial gas becomes increasingly more complex (Caselli & Ceccarelli 2012; Ceccarelli et al. 2022).

This interplay between increasing physical and chemical complexity will become more clear in the following sections, where we will specifically follow the major phases of star and planet formation for an isolated Solar-type star.

Classes of YSOs

Traditionally, the different evolutionary stages of young solar objects were classified on an observational basis in a backward timeline, starting from the most evolved, optically visible YSOs, towards less evolved, heavily obscured objects at optical wavelengths (Lada 1987; André & Montmerle 1994; André et al. 2000). In the NIR-MIR range, four classes of YSOs can be distinguished, according to the slope (α_{IR}) of the *Spectral Energy Distribution* (SED, $\lambda F_\lambda(\lambda)$) within 2.2 μm and 10-25 μm. In this time-backward sorting scheme, at ages of 10^{6-7} yr we find the Class III ($\alpha_{IR} < -1.5$) and Class II ($-1.5 < \alpha_{IR} < 0$) sources, that correspond to *Pre-Main Sequence* (PMS) stars. The protostar dominates the SED (optical-NIR ranges) with its black body emission, while the protoplanetary disk component emission (IR excess) is more accentuated in the less evolved Class II sources. At younger ages of 10^{4-5} yr, Class I and Class 0 sources ($\alpha_{IR} > 0$) are dominated in their SEDs by the emission of the infalling envelope (mm-submm), but for the more evolved Class I YSOs, it starts to rise the IR excess connected to the protoplanetary disk, as well as showing up the (cold) blackbody of the future star.

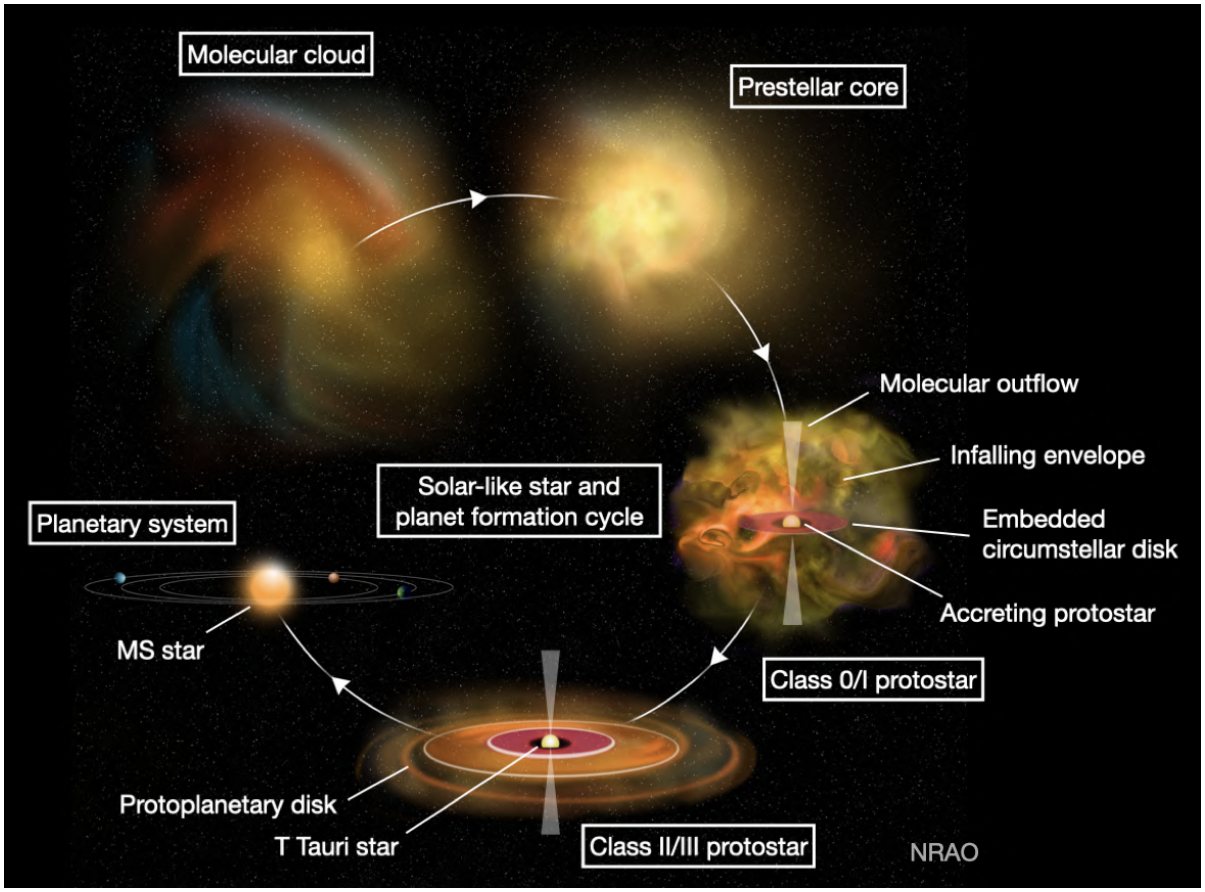

Figure 1.2: Sketch of the cycle of solar-like star and planet formation, evolving from a dense molecular cloud in the interstellar medium to the formation of a new planetary system around a young MS star. *Figure adapted from: NRAO.*

1.1.1 Where It All Begins: Molecular Clouds Cores

In order to understand how planets can form, we have first to explain how stars form.

The beginning of star formation is the assembly of a dense molecular cloud with even denser cloud *cores*. Molecular clouds represent over-dense regions ($> 10^2$ n_H cm^{-3}, hydrogen nuclei per cubic centimeter) in the ISM, enriched in molecular gas with respect to surrounding diffuse medium. The ISM consists, by mass, of ∼99% gas, mainly atomic or molecular hydrogen with traces of heavier elements, and 1% sub-micron (μm) dust grains. Outside over-dense regions, gas and grains are almost completely exposed to the *interstellar radiation field* (ISRF), which has a substantial UV emission component (at wavelengths shorter than 200 nm), that readily dissociates most molecules. The diffuse ISM is therefore characterized mainly by atoms and ions (Tielens 2005; Draine 2011; Stahler & Palla 2004; Klessen & Glover 2014).

When we start approaching a molecular cloud, its edges are still exposed to the ISRF, as well as to radiation of any nearby stars, resulting in so-called *photon dominated regions* (PDRs), which undergo a UV radiation that is 10^3–10^5 times the standard ISRF. Here, the combination of this intense radiation field with the dense, relatively cold gas and dust result in a particular gas-phase chemistry, that can form simple molecules behind the UV front, such as carbonium ion C^+ and *polycyclic hydrocarbons* (PAHs)[3]. The initial chemical composition of dense clouds is actually driven by the balance between molecular formation and photodissociation, which is the process through which a molecular bond is broken following the absorption of a photon ($AB+h\nu \rightarrow A+B$). Typical molecular (covalent) bond strengths are above > 5 eV, hence UV photons are required.

Entering the cloud, we find that its interiors are well shielded from external UV radiation by the optically thick dust grains population, especially toward the centre of the cloud, creating a favorable environment for the formation of macromolecules.

Also, we start to see even denser sub-structures, referred to as dense cores, or "starless cores", with typical sizes of about a tenth of a parsec. These cores are characterized by orders of magnitude higher densities than the parent cloud, $\sim 10^6$ n_H cm^{-3}, and lower temperatures, of the order of 10 K (Keto et al. 2015).

A lower temperature results from the general optical thinness of clouds towards the emission of radiation by chemical species in the gas phase, which act as very efficient coolants. At the same time, a small fraction of the gas is heated, or more specifically, ionized by cosmic rays, i.e high energy atomic nuclei, which can easily penetrate in clouds and interact with hydrogen, the main gas component, to produce ionising agents like secondary UV radiation, and free electrons (Shen et al. 2004; Caselli et al. 2012). However, it is still largely debated which kind of radiation, X-rays, cosmic rays, or free electrons, plays the most important role for ice chemistry in molecular clouds and in the following stages.

In any case, under such conditions of high density (but still very low compared to terrestrial conditions!), low temperature, and gas ionization, ion–molecule reactions in the gas-phase drive an efficient chemistry resulting, for example, in the formation of large carbon chains and smaller molecules (Ascenzi et al. 2019). Moreover, at grain temperatures of $\sim$ 10 K, all volatile molecules[4], except for H_2 and He, which tend to remain in the gas phase own to their very low atomic mass, start to stick onto grains upon collision, *freezing-out*. The freeze-out process, also called *adsorption*, combined with atom reactions on the grain surfaces, produce heavier species through H atom addition (*hydrogenation*), mainly H_2O and CO_2, but also more complex organic species, like CH_4 and CH_3OH, coating the grains with a chemically rich icy mantle (Caselli 1999).

[3] Organic compounds made of multiple aromatic rings, which are special cyclic structures with delocalized electrons.

[4] Volatile species are classified upon their sublimation temperatures. The lower the sublimation temperature, depending on surface and pressure, the higher the volatility (e.g. 20-40 K for methane CH_4, carbon monoxide CO and diatomic nitrogen N_2; $\geq$ 100 K for water H_2O, carbon dioxide CO_2, and methanol CH_3OH). Silicates and carbonaceous grains have the lowest volatility, and are classified as *refractories*.

The molecular cloud phase is therefore responsible for distributing O, C, and N elements into major volatile reservoirs, and for forming the first generation of organic molecules, which will constitute the basic "recipe" of the organic chemistry during the later stages of star and planet formation (Herbst & van Dishoeck 2009).

In addition to that, the cold temperatures of molecular clouds and cloud cores trigger the *isotopic fractionation* process, through which molecules get enriched or depleted in heavy, stable isotopes (i.e. same atoms, different number of neutrons). Deuterated species are very important, since they mainly form at this stage, and can be therefore used to trace the history of chemical species in the ISM. For example, their fingerprint has been found in the Earth's oceans and in the ices of comets, representing our most important evidence to connect solar system chemistry to interstellar chemistry (Caselli & Ceccarelli 2012; Ceccarelli 2004).

If a dense core becomes sufficiently massive, it begins to collapse under its own gravity, overcoming turbulence, thermal and magnetic pressure that stir the environment. The starless cores that are fated to collapse are referred to as *prestellar cores*.

1.1.2 The Dividing Line: Prestellar Cores

Not all starless cores give birth to stars. The "dividing line" for dynamical stability is represented by the critical density for gas cooling by gas-dust collisions, equal to 10^5 H_2 molecules per cm^{-3}. Starless cores with central densities below this critical density, start to accumulate matter towards their center under self-gravity, which is weakly counteracted by magnetic pressure and turbulence. These are the prestellar cores, in which the future star and planetary systems will actually form (Stahler & Palla 2004; Tielens 2005).

Prestellar cores display different gradients in both physical and chemical properties, as sketched in Fig. 1.3. Sizing about 10^4 AU, they span a range of number densities from a few times 10^3 cm^{-3}, bordering with the surrounding molecular cloud, to about 10^7 cm^{-3} within the central 1000 AU, where the gas and dust temperatures drop to about 7 K (Keto & Caselli 2010). Indeed, similarly to what we have seen for molecular clouds, prestellar cores are optically thin to the main sources of radiative cooling, i.e. warmed dust grains re-emitting in the sub-mm, and excited molecules emitting in the radio band. Moreover, their innermost regions are well shielded from the external UV field by the dust grains population. As a consequence, an outward positive gradient in temperature builds up in prestellar cores despite their overall contraction.

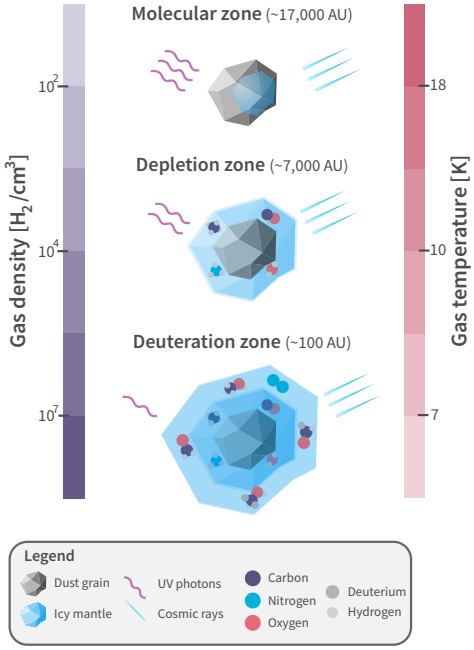

Figure 1.3: Ice mantle evolution in prestellar cores. From top to bottom, the physical and chemical changes (inspired to the prototypical prestellar core L1544, Keto & Caselli (2010)) as moving from the molecular outer-edge zone, via the depletion zone, to the central deuteration zone. Note that the color code of atoms in the legend is kept throughout the thesis. *Credits: A. Houge.*

The physical stratification in terms of density and temperature defines three major zones from the chemical point of view (Fig 1.3): the outer-edge "molecular zone", the intermediate "depletion

zone", and the inner "deuteration zone" (Caselli & Ceccarelli 2012). Starting from the outer edge, a prestellar core is exposed to the surrounding UV field, and the chemistry is regulated by photoprocesses. The ice mantles have just begun to stratify: oxygen, carbon, and nitrogen atoms are distributed via H addition to build respectively water (H_2O), methane (CH_4), and ammonia (NH_3). In addition to these chemical species, starting from 2012 special organic molecules, the *interstellar Complex Organic Molecules*, or iCOMs (Herbst & van Dishoeck 2009; Ceccarelli et al. 2017), have been observed towards prestellar cores, with emission lines coming from the inner edge of the molecular zone. These molecules represent the most complex end products of organic chemistry in the ISM, as well as the building blocks of prebiotic chemistry. It is therefore crucial to understand how do these molecules form, and be inherited by forming planets to develop life as we know it on the Earth (see § 1.2 for a dedicated section).

The presence of gaseous iCOMs in these cold objects actually still poses serious challenges to the current astrochemical models and formation theories (please refer to (Ceccarelli et al. 2022) for the latest review on the topic).

Moving inwardly toward the depletion zone, where the core starts to depart from the molecular cloud within which it is embedded, UV photons are absorbed by dust grains, and CO, the second most abundant volatile after H_2, rapidly disappear from the gas phase to stick onto the grains, in what is referred to in the literature as "catastrophic" freeze-out (Caselli 1999; Bergin & Tafalla 2007; Ceccarelli et al. 2007). This has two important implications. First, the frozen CO can be hydrogenated and deuterated, thanks to the high mobility of H and D atoms even at low temperatures, forming complex species such as formaldehyde (H_2CO), methanol (CH_3OH), and their deuterated forms. Secondly, since CO is very abundant, its freezing-out (along with other neutrals such as oxygen) opens the way to new chemical gas phase paths, in particular, for deuterium fractionation.

Indeed, deeper into the prestellar core, in the deuteration zone, the low temperatures favor the exothermic reaction between H_3^+ ion and HD hydrogen deuteride with respect to the inverse reaction ($H_3^+ + HD \rightarrow H2D^+ + H_2$), because the absence of gaseous CO cuts out other reactions that could occur with H_3^+ (Watson 1974). Overall, hydrogenated and deuterated species of CO, and large amounts of multiply deuterated H_2CO and CH_3OH species are produced and stored in the icy mantles.

Deep down in the deuteration zone, dust grains are expected to build very thick icy mantles (Caselli & Ceccarelli 2012): indeed, the freeze-out timescale ($\propto 10^9/n_H$ yr) is significantly shorter than the free-fall timescale of collapse of a core ($\propto 4\times10^7\sqrt{n_H}$; Tielens & Hagen (1982); Ioppolo et al. (2011)). It is important to underline that once species stick to grain surfaces, they cannot thermally evaporate in prestellar cores. This is because the thermal evaporation rate has an exponential behaviour ($\propto e^{E_B/kT_{dust}}$) in which the typical binding energies ($E_B/k \geq 1000$ K) of adsorbed species prevaricate the dust temperature ($T_{dust} \leq 10$ K). Moreover, they cannot get photodissociated, as central regions are extremely well shielded by dust extinction of radiation. The only way to return in the gas phase at this stage, which affects only a small amount of adsorbed molecules, is via non-thermal desorption mechanisms, mainly driven by cosmic rays (Leger et al. 1985; Wakelam et al. 2021).

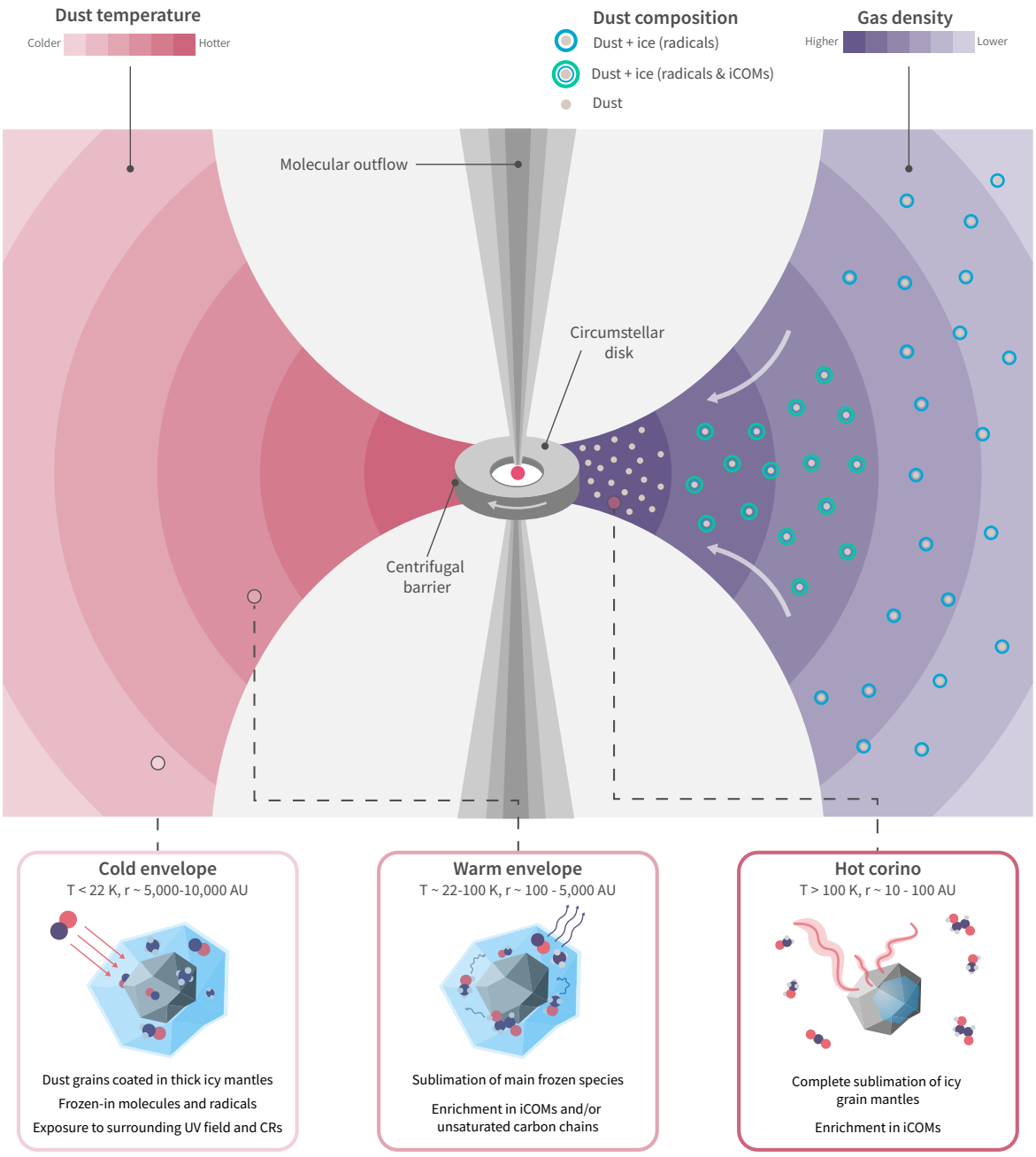

Figure 1.4: Schematic structure of a Class 0/I protostar outlining the chemical composition of dust grains at different scales. The complex interplay between physical and chemical evolution gives rise to a stratified protostellar environment moving from the largest scales of the infalling envelope ($\leq$ 10000 AU) and the molecular outflow down to the planet-forming scales of the circumstellar disk ($\leq$ 100 AU) surrounding the central accreting object. At scales comparable to the circumstellar disk, the hot and dense hot corino region hosts an incredible chemical complexity own to the complete sublimation of the pre-built icy mantles of dust grains. *Credits: A. Houge.*

1.1.3 Star Cradles: Protostellar Envelopes And Circumstellar Disks

At the onset of the collapse, the matter in the prestellar core free falls towards the center, and a protostar is formed.

Observed as a Class 0 or Class I source ($\sim 10^4$-10^5 yr from the collapse, André & Montmerle (1994); André et al. (2000)), at this stage the protostar is constituted of four main components: the central object, that will eventually become a star; the circumstellar disk, made of matter in rotation to conserve the angular momentum of the parent cloud, and in which planets and minor bodies can start forming (Tychoniec et al. 2020; Segura-Cox et al. 2020); the slowly infalling envelope, from which the central object accretes matter; an outflow of ejected material, departing from the innermost $\sim$ 1-10 AU scales of the disk. Indeed, the angular momentum in the circumstellar disk prevents material from directly falling onto the central stellar core, and, in order to remove the angular momentum and allow accretion, a fraction of matter gets violently ejected outward through highly supersonic and collimated jets ($\geq$ 100 km s^{-1}) and slower molecular outflows (1-30 km s^{-1}) (Frank et al. 2014). When the outflowing material encounters the quiescent gas of the outer envelope and the molecular cloud embedding the system, it creates shocks, which partially *sputter* (destroy) and vaporize grain mantles and grains' solid cores (Flower & Pineau des Forets 1994). But not all bad comes to hurt: molecular outflow shocks constitute special space laboratories in which several iCOMs are observed. Indeed, the quiescent medium gets perturbed via shock waves that heat and compress it triggering several processes (which would not occur otherwise in quiescent environments): besides sputtering, also molecular dissociation and warm gas chemical reactions, contributing to the chemical enrichment of the gas surrounding young protostars.

Since there is no active fusion reaction in the protostar, its luminosity is associated to the conversion of the gravitational energy of the collapse into radiation, according to: $L_{acc} = GM_* \dot{M}/R_*$, where M_* and R_* are the mass and radius of the central protostar, while $\dot{M}$ is the mass accretion rate (Stahler & Palla 2004). Despite the fact that the exact structure of protostellar envelopes is unknown, it is clear that as the collapse proceeds both the density and temperature increase toward the center, having now also an internal heating source, the protostar. Similarly, the velocity of the infalling gas generally increases as approaching the center with an $r^{-1/2}$ power law. Again, the chemical structure and evolution of the protostellar environment are tightly linked to the physical structure (e.g. density and temperature profiles), as well as the presence of ionising agents (e.g. UV photons and cosmic rays).

From this point of view, it is convenient to divide the chemo-physical structure of a Class 0/I protostar into four major zones, covering scales from the outermost envelope layers down to the central accreting object (Caselli & Ceccarelli 2012), as sketched in Fig. 1.4:

1. A cold envelope (10000 $\rightarrow$ 5000 AU), characterised by dust temperatures $\leq$ 22 K. Similarly to the chemistry in dense prestellar cores, several molecules, including CO, are largely stocked into the grain icy mantles. Although with a much lower abundance than in inner regions, iCOMs are present here, and it is likely that the same process is responsible for the presence of iCOMs in this and in prestellar cores.

2. A lukewarm region (5000 $\rightarrow$ 100 AU), having dust temperatures in between 22 - 100 K, whose extremes corresponding to the sublimation temperatures of the main grain mantles components, in order, CO and H_2O. In this region, more complex species can be present either in the form of carbon chains, iCOMs or both. Indeed, there is a sort of competition between the chemistry induced by CO depletion, and the one associated to the sublimation of methane (CH_4), having a sublimation temperature of $\sim$ 60 K). If abundant enough,

when injected in the gas phase CH_4 triggers a rich chemistry of unsaturated carbon-chains, such as butadiynyl (C_4H).

3. A hot ($T_{dust} > 100$ K), dense ($> 10^7$ cm^{-3}), compact inner envelope (100 $\rightarrow$ 10 AU) surrounding the central object, the *hot corino*, where water-rich ices completely sublimate into the gas-phase, greatly enriching this region in iCOMs.

4. A circumstellar disk ($\sim$ at scales of the inner envelope), embedded within the thick and massive outer infalling envelope. The dust absorption and the line emission from the infalling envelope tend to mask the line emission from the disk itself, in particular weak iCOM lines. This initial disk stage is expected to be hot, and may remove any recollection of the chemistry of previous stages Öberg & Bergin (2021).

Hot Corino vs WCCC Sources

Depending on whether the observed spectrum of a Solar-type Class 0/I protostar is dominated by iCOMs fingerprints (their emission/absorption lines), or unsaturated carbon-chains, we call them respectively hot corinos and WCCC sources (*Warm Carbon Chain Chemistry* sources; Sakai & Yamamoto (2013)). As often happens in science, they represent the two extreme cases, and hybrid sources, where the WCCC and hot corino chemistry appear on different scales, have been found.

There is no general agreement on what determines the chemical nature of a protostar, and the existence of hybrid sources.

At present, two major hypothesis are at play, *(i)* a different timescale of formation of grain mantles, and *(ii)* a different illumination of grains from UV photons. These two hypothesis have in common the idea that the chemical distinction in iCOMs/unsaturated carbon chains is induced by which molecule dominates as carbon reservoir in dust ices, methanol (CH_3OH) or methane (CH_4). When CH_3OH is more abundant, it triggers the production of iCOMs, and we observe an hot corino. If CH_4 is the most abundant of the two, the chemistry is instead dominated by the formation of unsaturated carbon chains, and we recognize the protostar as a WCCC source.

So, where do these two molecules come from? Methanol is thought to be formed on the grain surfaces by addition of H atoms to CO; methane, by addition of H atoms to C (Rimola et al. 2014). Therefore, going to the root of the problem, we must understand whether CO or C is more abundant on grain surfaces. And this is where the two hypotheses really come at play: *(i)* since methanol forms after CO freezes out at relatively late times ($\geq 10^5$ yr), while the formation of methane is immediate, the formation time of icy grain mantles determines whether they are enriched with one molecule rather than the other; *(ii)* a different UV illumination can affect the relative abundance of CO and C, depleting CO through photo-dissociation in case of strong UV field.

Hot corino sources represent very intriguing protostars, as they incubate many, and in abbundance, iCOMs. Since Class 0/I YSOs containing hot corinos could host the seeds of planets, investigating the chemical complexity in hot corinos could therefore shed light where come from the roots of life in planetary systems like our own (§ 1.2).

1.1.4 Work in Progress: The Protoplanetary Disk Stage

After $\sim$ 0.5 Myr since the birth of the protostar/disk/outflow system, the envelope dissipates, the circumstellar disk enlarges and stratifies into a so-called protoplanetary disk, while at the centre the accreting object has become a *T Tauri star*, or pre-main sequence star. Depending on the exact evolutionary features of the system, we have entered the Class II/Class III phase.

As suggested by the name, PMS have not reached the MS yet, but their luminosity is due to an interplay between radiative conversion of gravitational energy and partial fusion chain reactions of H in their forming core (Carroll & Ostlie 2007).

Protoplanetary disks, being only a few % the stellar mass, and composed by a mixture of dust and gas with expected Keplerian motion, are very complex structures: they are characterized by a temperature > 1000 K in the inner disk and upper layers, to ≤ 10 K in the outer midplane (the plane dividing the disk into upper and lower halves along its vertical axis); the density varies as well, from $> 10^{13}$ cm^{-3} in the inner midplane down to 10^5 cm^{-3} in the upper outer layer. Also the UV radiation field varies across a protoplanetary disk, reaching 10^5 times the interstellar radiation field in the surface layers at 10 AU. Even dust grains are interested by dramatic evolutionary changes: they grow from gas-rich, small grains (~ 0.1 μm), to dry *pebbles*, the bricks of planets ($\sim$ 0.1-1 m), which are largely decoupled from the gas, and settle to the midplane, where they start radially drifting. If they encounter a pressure bump, they get trapped and can grow to even larger *planetesimal* sizes (Testi et al. 2014a; Miotello et al. 2023). In fact, considered as one of the most remarkable discoveries of the last decade, protoplanetary disks show gaps in the dust continuum emission maps, a possible evidence of ongoing planetary formation (Isella & Turner 2016; Fedele et al. 2018).

As one might imagine at this point, chemistry is not free from change: effectively, different types of chemistry show up in different parts of the disk. Imagining to slice a vertical section of the disk, at a sufficient distance from the central hot protostar, and moving from top to bottom of this disk slice, we would find (Miotello et al. 2023):

1. The *hot surface layer*, chemically shaped by the UV photons population. As in photo-dissociation regions at the edges of molecular clouds, only ionised and neutral atoms are present, because molecules are photo-dissociated.

2. The *warm molecular layer*, where molecules are in the gas-phase and iCOMs are detected. In fact, here the UV photons are largely absorbed and the dust temperature is high enough for several molecules to remain gaseous. Yet, in general, the detection of iCOMs is challenging because the dust along the line of sight might heavily absorb their line emission, and they are not expected to be much abundant, due to the small vertical size and depth (in radius) of this disk layer.

3. The *cold midplane*, where molecules freeze-out onto grain mantles, leaving only H and D bearing species in the gas-phase. The freeze-out does not occur at the same disk location for all molecules, determining the so-called *snowline* of individual molecules (such as of H_2O, CO, CH_4 and NH_3). Each snowline flags the disk radius and height where the gas-solid transition of a molecule occurs, as the dust temperature is lower than the specific molecule's freezing temperature. However, a small fraction of molecules can be found in the gas-phase beyond their snowlines (outward in the disk), due to the UV and X-ray protostar's illumination or cosmic rays irradiation. Snowlines are thought to influence planet formation in different ways, in particular, by determining the elemental content of solids and gas at different disk locations (hence potentially the future exoplanetary atmospheres), and changing the grain growth process, accelerating the assembly of pebbles.

A special mention goes to protoplanetary disks surrounding *FU Orionis* stars (Fu Ori), YSOs whose luminosity fluctuations are so intense that the illumination of the surrounding disk can vary by up to 2-3 orders of magnitude. The precise mechanism behind FU Ori outbursts is not completely understood, but there are several leading theories in this regard, including accretion disk instabilities, magnetic field interactions, fragmentation of accretion disks, and episodic mass accretion (Connelley & Reipurth 2018).

The increased luminosity warms up the dust population, sublimating the icy mantles in relatively large portions of the disk, therefore outwardly pushing the snowlines. In turn, this makes

possible to observe the molecular content just released from the grains, and detect iCOMs in the disk at larger scales.

1.1.5 Open Ending: From Planetesimal To Planet Formation

We are at the end of this long journey, but this is not as clear as we would like. In principle, the sub-micron dust grains coagulate into larger rocks, called *planetesimals*, the bricks of the future planets, as well as moons, comets and asteroids. While the grains progressively glue together, some of their icy mantles are likely preserved, so that at least part of the protostellar chemical history may be conserved in the building blocks of rocky bodies in the new planetary system (Turner et al. 2014).

However, the transition between protoplanetary disks and planetary systems is far from being understood. This uncertainty spreads also on the chemical thread connecting the earliest and the latest stages of star and planet formation.

In fact, prior to planet formation, the disk chemical composition is hidden from our view, own to the massive presence of dust. We are getting some clues about it from the observation of warmer, younger Class 0 disks, where less of the material is frozen out and fewer planetesimals have formed. Also, by studying the composition of comets and other icy bodies in the Solar System, which provide hints of the chemistry in our natal solar nebula disk, assuming that these bodies are largely unchanged since their formation (4.5 billion years ago), alluring, yet scattered, similarities have been found between cometary and interstellar abundances (Ceccarelli 2014). Overall, it is not possible at the moment to outline a general trend for iCOMs inheritance and abundances moving from Class 0 and Class I sources to the latest stages of star and planet formation.

As a matter of fact, the chemical composition and origin of exoplanet atmospheres, from *Super-Earths* to *mini-Neptunes* and giant Jovian planets, clearly represent the next frontier for Astrochemistry.

Up to now, only CO and H_2O have been robustly detected, with indications of CH_4, NH_3 and/or HCN (hydrogen cyanide) (Madhusudhan et al. 2016). Atmospheric features are often obscured by clouds and hazes in the optical-NIR windows. In the case of giant planet atmospheres, it is clear that individual molecule abundances will be fully reset, and even the overall abundance ratios of carbon-oxygen, oxygen-hydrogen etc., likely preserved in the atmosphere, could be affected if part of the atmospheric material is cycled to the planetary core (Piso et al. 2015; Eistrup, Christian et al. 2018).

The atmospheres of rocky Earth-like planets may have an even more complicated history. Built well outside the water snowline, primitive terrestrial planets consist of planetesimals that are practically half rock and half ice. Migrating inward, closer to the water snowline, water ocean planets should form from the fusion of iced water. However, inside the snowline, much closer to the host star, rocky planets are usually thought to be very dry (look at Mercury or Venus), therefore, they should lose most of their original volatiles, their initial atmosphere, through different mechanisms. Yet, a secondary atmosphere could layer on the surface of rocky planets through impacts of icy planetesimals in the late stages of their formation. Minor bodies like comets could thus deliver water as well as organics to young planets. The survival of these organics is not obvious, but if they do, then there would finally be a direct link between interstellar molecules and the building blocks for life on planets.

1.2 The Quest Of The Chemical Complexity In Hot Corinos

The first complex organic molecule, formamide (NH_2CHO), was detected in 1971 towards Sgr B2 (Rubin et al. 1971), a star nearby the Galactic center, that will have helped characterizing our supermassive black hole, Sgr A* (Gillessen et al. 2009).

Since then, the interstellar complex organic molecules (iCOMs; Herbst & van Dishoeck (2009); Ceccarelli et al. (2017)), have become of primary interest in the scientific community. They are specifically defined as "complex" (in the context of the ISM) molecules with at least 6 atoms of which carbon, hydrogen and atoms of other species, connected by single covalent bonds (i.e. sharing one pair of electrons, as found in the most common terrestrial organic molecules).

Why are we so interested in them? Because iCOMs represent the tip point of organic chemistry in the interstellar space, and they may be lead to the emergence of life acting as building blocks for even more complex, prebiotic molecules, throughout their creation and inheritance path from protostars to planetary systems.

Between the '80s and the '90s, several iCOMs started to get observed towards protostellar envelopes surrounding massive young stars. The proprieties of the line emission indicated that these iCOMs had to emit from compact ($\leq$ 0.01 pc), hot ($\geq$ 100 K), and dense ($\geq 10^7$ cm^{-3}) protostellar regions, hereafter called *hot cores* (Blake et al. 1987; Kurtz et al. 2000; van der Tak 2005).

The first, simple, and intuitive interpretation that was given is that iCOMs result from the sublimation of some molecular species from the icy grain mantles, the "mother" or "primary" species, and the gas-phase synthesis of others, called "daughter" or "secondary" species.

Then, at the beginning of the new millennium, a similar environment was recognized in the envelope of the prototype Class 0 YSO named IRAS 16293-2422 (Ceccarelli et al. 2000b; Cazaux et al. 2003). It turned out not to be an isolated case: in the following years, many low-mass hot cores have been discovered (about 25 up to now) and, to distinguish them from the high-mass hot cores, they were called hot corinos (Bottinelli et al. 2004a,b; Ceccarelli 2004; Bottinelli et al. 2007; Lahuis et al. 2006; Jørgensen et al. 2012).

Hot corinos are not simply "rescaled" versions of hot cores: they are characterized by smaller sizes, lower temperatures and densities, but also different chemistry.

The fact that hot corinos often belong to binary/multiple systems does not help at finding a theoretical model able to describe their physical structures. Indeed, at scales $\geq$ 100 AU, a roughly spherical envelope heated by the internal new born star is probably a correct description. However, at smaller scales ($<$ 100 AU), that is, the hot corino/circumstellar disk region, the envelope may not be spherical, also due to the presence of multiple sources.

Moreover, the binary/multiple components of Class 0/I sources are often reported of showing a noticeably different chemistry (e.g. Jørgensen et al. (2011) and Manigand et al. (2020) for IRAS 16293-2422; López-Sepulcre, A. et al. (2017) for IRAS 4A). However, recent evidence points out that this chemical difference may be just apparent, induced by a different intervening dust opacity on the line of sight of observation of the components (De Simone, M. et al. 2020).

1.2.1 The Diatribe On iCOMs Formation

Looking back at the previous sections, the reader might have noticed that we have described what iCOMs are, when and where they can be observed in a protostellar environment, but never described how they form. The thing is, the iCOMs formation process is an open and debated question, and is not the only one: above all, we do not know either how iCOMs, once formed around a protostar, can be preserved and inherited by the forming planets.

The state-of-the-art literature evokes two major iCOM formation routes, which are the GAS+GRAIN CHEMISTRY and the ONLY GRAIN CHEMISTRY paradigms, following the nomen-

clature of (Ceccarelli et al. 2022). The first step, in common between the two paradigms, occurs during the molecular cloud (§ 1.1.1) and prestellar core (§ 1.1.2) evolutionary stages, when organic ices start to layer on the surface of dust grains. Gaseous atoms and small molecules collide with the grains and stick to their surface, getting hydrogenated and oxygenated (enriched in H and O atoms). The bulk of icy mantles is done, mainly made of H_2O, CO_2 (carbon dioxide), CH_3OH, NH_3 and CH_4 (Boogert et al. 2015; McClure et al. 2023). At the second step, the two formation paths diverge.

In the GAS+GRAIN CHEMISTRY paradigm (Millar et al. 1991; Charnley et al. 1992; Balucani et al. 2015; Skouteris et al. 2017), thermal and non-thermal processes partially or completely sublimate the grain mantles, depending on the temperature and the species binding energy E_B. Grain mantles components can be realised in the gas-phase and observed via thermal and non-thermal desorption. The former occurs when the dust temperature is greater than the desorption energy E_{des} (linked to E_B) of a given species, e.g. water sublimates at around 100 K. The latter consists in energetic processes that overcome the species binding energy, e.g. sputtering, chemical desorption (occuring when a surface reaction releases enough heat in a reaction to overcome the binding energy of the produced species), or photo/cosmic rays desorption (causing a molecule to pass in the gas-phase due to its bombardment with energetic particles).

Once in the gas-phase, the desorbed mantle components react with the other gaseous species and form, through ion-neutral and neutral-neutral reactions, also new iCOMs.

The ONLY GRAIN CHEMISTRY paradigm (Garrod, R. T. & Herbst, E. 2006; Garrod, R. T. et al. 2007) takes a slightly more complex way. The grain mantles do not sublimate, and are instead processed by UV photons and cosmic-rays irradiation (in order, *photolysis* and *radiolysis*), which bombard the icy mantles forming radicals in their bulk (e.g. OH hydroxide and CH_3 methyl group by photodissociation of CH_3OH). When the dust temperature reaches about 20-30 K, the torpid radicals become mobile. At this point, iCOMs can be formed in two ways. As first option, radicals can meet each other on the grain surface by diffusion, hopping over their potential barriers, react and form iCOMs. Alternatively, non-diffusive processes can also cause the radicals to meet on the grain surface and react to form iCOMs, e.g. radicals forming by reactions occurring close to another reactant, producing an immediate follow-on reaction (Jin & Garrod 2020; Garrod et al. 2022). The final population of iCOMs is released in the gas-phase by thermal and non-thermal processes, and they can be observed through their emission and absorption lines, as for the other paradigm.

Probably, both gas-phase and grain-surface chemistry are at work in the interstellar space, yet dominating the formation of different iCOMs in different environments and at different times of Solar-type star forming regions.

Independently of the formation route, we are not currently able to explain how iCOMs can be removed from the grain mantles and injected into the gas-phase, where they are detected, in cold and dense protostellar regions, such as in prestellar cores and in protoplanetary disks (§ 1.1.4).

Non-thermal processes could be the answer, but they depend on several poorly known parameters: on top of all, and common to almost all non-thermal processes, is the species binding energy (E_B), the energy necessary to release the frozen iCOM into the gas. Constrain the spatial location, abundance, and the gas temperature at which these molecules emit observable radiation is therefore crucial to help constraining their formation routes.

1.3 The Source Under Investigation: IRAS 4A Binary

The IRAS 4 system is part of the southern filament of the NGC 1333 star cluster (Dhabal et al. 2018) belonging to the Perseus Giant Molecular Cloud, or Per MCld (α: 3^h 35^m [J2000], δ: $+31°13'$ [J2000] (Kirk et al. 2006), at 299 $\pm$ 17 pc away from us (Zucker et al. 2018) in the Perseus Constellation.

On The Background: Perseus Giant Molecular Cloud And NGC 1333 Cluster

Giant Molecular Clouds are among the largest and most massive structures in galaxies, and we have seen that they play a crucial role in the process of star formation. Inside them, the dense cloud cores, possibly leading to the formation of stars, show up as dark patches against a background network of bright filaments of dense gas.

The Perseus Molecular Cloud is estimated to have a total mass $\geq 1 \times 10^4$ $M_\odot$, and an angular extension of $\sim 1.5° - 5°$ (Bachiller & Cernicharo 1986; Carpenter 2000; Kirk et al. 2006).

Most of the young stars in Per MCLd reside in IC 348 and NGC 1333 star forming regions, which contain about 480 and 200 objects respectively, with ages of 1–3 Myr (Luhman et al. 2016).

In particular, NGC 1333 currently stands out as the most active region of star formation in the Perseus molecular cloud, with an estimated star formation rate (averaged over the last million years) close to $10^{-4} M_\odot$ yr^{-1} (Walsh et al. 2007). Originally, NGC 1333 was identified as a bright reflection nebula in the western portion of the cloud (Sargent 1979; Loren 1976), but today the term is used to denote the young stellar cluster in addition to the reflection nebula.

NGC 1333 cluster seems to contain about 150 YSOs, with a median age of $\sim 10^6$ year and a total mass of circa 100 $M_\odot$ (Walsh et al. 2006).

Zooming In: The IRAS 4 System

The IRAS 4 system is composed by the Class 0 protostar IRAS 4B, and the Class 0 binary system IRAS 4A, characterized by 4A1, the brightest component at mm wavelengths, and the lower luminosity companion 4A2, separated by a distance of 1".8.

The name comes from the system detection in IRAS maps by (Jennings et al. 1987). Since then, the system has been repeatedly imaged at mm and radio wavelengths, enabling to disentangle the IRAS 4A1 and 4A2 components, plus two components in IRAS 4B, 4B and 4B', separated by 10" (Choi et al. 1999; Smith et al. 2000; Looney et al. 2000; Di Francesco et al. 2001; Reipurth et al. 2002).

IRAS 4A2 is the second hot corino ever discovered (Bottinelli et al. 2004a; Taquet et al. 2015; López-Sepulcre, A. et al. 2017; De Simone, M. et al. 2020) just after IRAS 16293-2422 (Ceccarelli et al. 2000b; Cazaux et al. 2003). Interferometric IRAM/PdBI (*Plateau de Bure Interferometer*)[5] observations suggested that iCOM emission originates from 4A2 rather than 4A1 (Taquet et al. 2015; De Simone, M. et al. 2017; López-Sepulcre, A. et al. 2017). However, De Simone, M. et al. (2020) targeting IRAS 4A with the VLA interferometer (*Very Large Array*)[6], discovered that moving at cm wavelengths iCOMs abundances are similar between 4A1 and 4A2, also own to a different optical depth of the dust in front of the two components. An hot corino has been observed also towards IRAS 4B (Bottinelli et al. 2007; De Simone, M. et al. 2017).

[5] https://web-archives.iram.fr/PDBI/
[6] https://public.nrao.edu/telescopes/vla/

At millimeter wavelengths, 4A1 is three times brighter than 4A2, but we only know the bolometric luminosity of the whole IRAS 4A system, being of 9.1 $L_{\odot}$ (Kristensen, L. E. et al. 2012; Karska, A. et al. 2013), because the two components are not resolved in the FIR range (where the luminosity peaks), and therefore the relative luminosity of the two components is unknown.

Interestingly enough, the IRAS 4A system is associated with a large-scale ($\sim$ few arcmin) bipolar outflow (Choi 2005): two blue-shifted southern lobes, of which a fast collimated jet driven by 4A1, and a slower and precessing jet driven by 4A2 (Santangelo, G. et al. 2015); a red-shifted northern lobe, where Taquet, V. et al. (2020) were able to to distinguish as well an outflow driven by 4A1 from one driven by 4A2.

Fig. 1.5 tries to frame the subject of this thesis project in its visual and scientific beauty. The NGC 1333 star forming region constitutes a small, shining aggregation of stars in the Perseus Giant Molecular Cloud, that is, indeed, a giant cloud of dust, gas, and stars in the Perseus constellation. Zooming onto the IRAS 4A system, IRAM (Lefloch et al. 2017), VLA (Choi 2005) and NOEMA-SOLIS[7] (De Simone, M. et al. 2020, 2022) observations have revealed, together with other works earlier cited, a complex, and chemically copius outflowing structure.

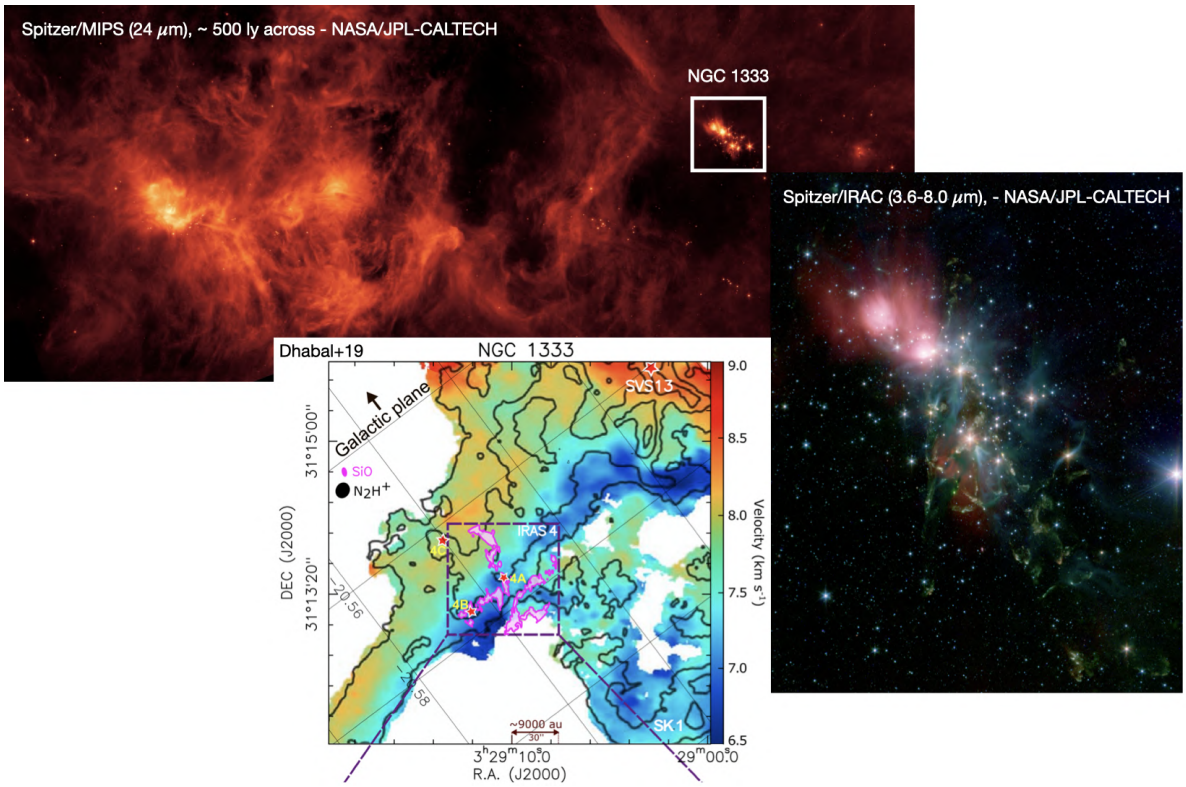

Figure 1.5: The Perseus Giant Molecular Cloud, NGC 1333, and the IRAS 4A system. In the upper left panel, the MIR-FIR Spitzer Space Telescope image of the Perseus Molecular Cloud. The light white rectangle frames the most active star forming region in the cloud, NGC 1333. In the center right panel, NIR Spitzer Space Telescope Image of NGC 1333. In the lowest panel, N_2H+ ion velocity map by Dhabal et al. (2019) tracing the outflowing structure of IRAS 4A. *Figure adapted from: NASA/JPL-CALTECH & De Simone, M. et al. (2022).*

[7] https://solis.osug.fr

1.4 Thesis Objectives

The organic chemistry that builds life on our planet, unmatched elsewhere, very likely originates in the process of formation of the Solar System (Caselli & Ceccarelli 2012; Ceccarelli et al. 2022; Öberg & Bergin 2021). Among its primitive building blocks components, comets and asteroids show traces in their composition of fundamental building blocks of life, such as amino acids (Ceccarelli 2014; Caselli & Ceccarelli 2012). The formation of Solar-like stars consists in different phases, starting from a cold yet dense molecular cloud core that eventually collapses to form a protostar/disk/outflow system. This matures in few million years until the growth of dust grains in the protoplanetary disk leads to a new planetary system. Throughout all these evolutionary phases, the physical and the chemical evolution proceed hand in hand (Öberg & Bergin 2021). However, iCOMs (Herbst & van Dishoeck 2009; Ceccarelli et al. 2017), the most complex ISM organic molecules and precursors of life compounds, have been discovered already in the inner envelope, i.e. the "hot corinos" (Ceccarelli 2004), surrounding the youngest Solar-like protostars, the Class 0/I YSOs (André & Montmerle 1994; André et al. 2000), which have been recently re-discovered as possible forming planets hosts (Tychoniec et al. 2020; Segura-Cox et al. 2020). This finding begs a natural question: is this chemical complexity preserved and inherited by planets, their atmospheres, and other minor bodies in new planetary systems? In order to answer this question, it is clear that is crucial to fully characterize, from a chemical point of view, the early phases of the formation of Solar-like stars, and to connect the chemical evidences with the physical properties of these environments.

In this thesis project, by exploiting for the first time the ALMA/FAUST dataset (resolution $< 0.''2$) collected for the well known NGC 1333/IRAS 4A2 hot corino (< 300 pc), I can spatially resolve an hot corino region, which has very few precedents (Codella et al. 2016; Bianchi et al. 2022; Maureira et al. 2022; Lee et al. 2023), and study in detail the chemical complexity and thermal structure of a Class 0 YSO.

Aiming to assess the chemo-physical complexity of hot corinos, throughout this project I was guided by the following question marks: are different complex organic molecules tracing spatially distinct regions within the planet-forming zone around the young IRAS 4A2 protostar? If so, are the probed scales associated with a uniform or varying gas temperature? Eventually, can we infer a gas temperature profile of the planet-forming zone from the observation of these molecules?

I therefore pursued to achieve the following practical objectives, processing and analyzing my dataset with the *Common Astronomy Software Applications* (`CASA`; version: 6.5.6), the *Cube Analysis and Rendering Tool for Astronomy* (`CARTA`; version: 4.0.0)[8], and `python` (version: 3.9.7)[9]:

- Provide an up-to-date chemical characterization of IRAS 4A2 hot corino through the detection of 5 common, abundant, and bright iCOMs found in Solar-type star forming regions (Ceccarelli et al. 2022): methanol (CH_3OH), acetaldehyde (CH_3CHO), methyl formate ($HCOOCH_3$), glycolaldehyde (CH_2OHCHO), and formamide (NH_2CHO).
- Image and compare the emission of iCOMs emission lines with similar excitation temperature to constrain their spatial segregation in the hot corino region.
- Derive two fundamental physical properties of the hot corino, iCOMs abundances (column densities), and gas kinetic temperature of the gas probed by these molecules, and compare the results with the literature.
- Produce a radial temperature profile of IRAS 4A2 at compact scales < 50 AU, comparable to Solar System scales, exploiting the observed iCOMs emission lines.

[8] https://carta.readthedocs.io/en/latest/
[9] https://www.python.org

The thesis is organized as following:

1. Chapter 1 gives a scientific background and context of the thesis and the target source, and closes with this section on the objectives of this work.

2. Chapter 2 contains the methodology adopted to achieve the objectives set. Specifically, § 2.1 introduces the radio astronomy window and the observational requirements in terms of angular resolution and sensitivity, § 2.2 illustrates the principles and methods of data reduction for radio interferometry, § 2.3 presents the ALMA observatory and the Large Program ALMA/FAUST exploited for the thesis project, and § 2.4 overviews the basics of molecular spectroscopy and the molecular line analysis technique used in this thesis.

3. Chapter 3 presents and discusses the results of this work, in terms of: data deconvolution, imaging, and spectral analysis (§ 3.1 and § 3.2); imaging and comparison of iCOMs emission (§ 3.3); derivation of the physical properties of the hot corino (§ 3.4); outline of the temperature profile of the hot corino (§ 3.5).

4. Chapter 4 summarizes the main conclusions and the future perspectives of this work (§ 4.1).

In addition, at the end of the thesis there is the complete list of referred bibliographic records, and the appendices containing subsidiary tables and figures to the results (Appendix A, B, C, D, E).

At last, in the non-numbered chapters, the reader will find the abstract of the thesis project at the beginning, and the acknowledgments at the end of the document.

CHAPTER 2

Adopted Methodology

2.1 The Framework: Radio Astronomy

When we frame astronomical observations in the window of radio electromagnetic radiation, we talk about Radio Astronomy. The radio window is defined by where the Earth's atmosphere is transparent to radio waves, which occurs in a frequency (15 MHz - 1.5 THz)/wavelength (20 m - 0.2 mm) range whose limits are not well-defined (see Fig. 2.1), due to variations e.g. with altitude. The upper limit is roughly set by atmospheric molecules that absorb radiation while undergoing vibration motions, mainly water vapor (H_2O) and molecular oxygen (O_2). At the lower bound of the radio window, instead, the atmosphere becomes opaque to radiation because of free electrons in the ionosphere, which causes radio waves to be reflected back into space[1].

Considering the electromagnetic radiation in the radio window to travel in straight lines, we define as *brightness* I_ν (also called *intensity* or *specific intensity*) the power P we receive from a source per unit solid angle dΩ, area dσ, and frequency dν:

$$I_\nu(\theta,\phi) = \frac{dP}{cos\theta\ d\Omega\ d\sigma\ d\nu}\ [\ erg\ s^{-1}\ Hz^{-1}\ sr^{-1}\ ] \tag{2.1}$$

where θ refers to the angle between the normal to dσ and the direction to dΩ. If we integrate the brightness of a source over its subtended solid angle Ω in the sky, we obtain the *flux density* S_ν of that source:

$$S_\nu = \int I_\nu(\theta,\phi)\ cos\theta\ d\Omega\ [\ erg\ cm^{-2}\ Hz^{-1}\ ]. \tag{2.2}$$

Since astrophysical radio sources tend to have a small density flux, we use a more practical unit to measure it, the so-called *Jansky* (Jy):

$$1\ Jy = 10^{-23}\ erg\ s^{-1}\ cm^{-2}\ Hz^{-1}. \tag{2.3}$$

.

Together with the optical/near-infrared regions of the electromagnetic spectrum, the radio window can be observed from the ground, as done by single-aperture, single-mount radio telescopes, such as the *Robert C. Byrd Green Bank Telescope* (GBT)[2], and the IRAM-30m[3] telescope. Radio telescopes are practically receiving antennas, e.g. they convert electromagnetic radiation in the sky into electrical currents in conductors, and are equipped with large reflectors to collect more radiation and to focus the electromagnetic waves in phase onto their feed antennas (Wilson et al. 2013).

[1] The ionosphere comprehends different layers of the upper Earth's atmosphere, ionized by solar radiation. The electromagnetic radiation cannot be transmitted through the atmosphere if its frequency is below the *plasma frequency*, which is proportional to the square root of the electron density of the plasma (Chapin et al. 2011)

[2] https://greenbankobservatory.org

[3] https://iram-institute.org/about/

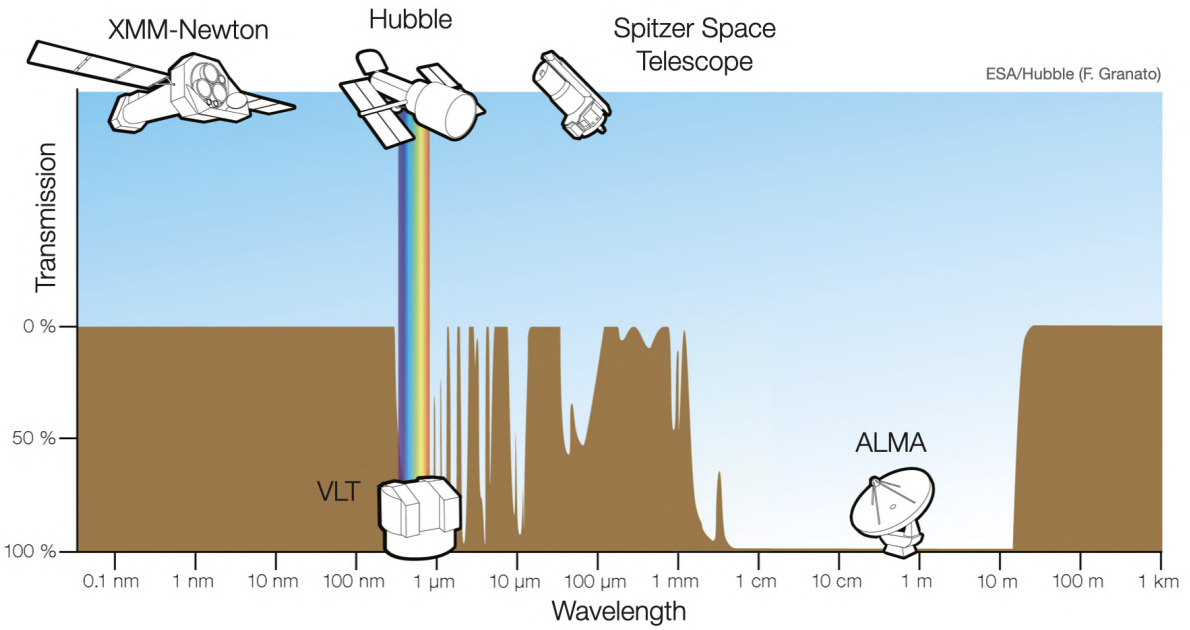

Figure 2.1: Transmission of electromagnetic radiation through the Earth's atmosphere. The brown curve shows how transparent the atmosphere is to radiation from space as a function of the wavelength. From the ground, the major observable windows are at visible wavelengths (rainbow) and at radio wavelengths ($\sim$ 1 mm - 10 m). ALMA operates in a spectral window where the opacity depends strongly on how high and dry the observing site is. In space, observations can be done outside the atmospheric windows, as illustrated by *XMM-Newton* (ESA), *Hubble* (NASA) and the *Spitzer Space Telescope* (NASA).

2.1.1 Looking For Angular Resolution... And Sensitivity

[4]Any optical system images a point source at infinity (for all practical purposes, a star in the sky) as a bright spot surrounded by faint concentric rings. This is caused by the diffraction phenomenon, which is the spreading of light waves as they pass the edge of an opaque body; in a telescope, we mean at the edge of the aperture and at any obstacle within the telescope aperture. The same argument applies to extended objects, which can be considered to the first order as a collection of point sources. The distribution of light intensity in the image of a point source is called *Point Spread Function* (PSF). In the ideal case of a perfect optical system with an unobstructed circular aperture and monochromatic light, the PSF can be determined analytically and is given by the well-known *Airy function*. The Airy pattern arises only due to the diffraction phenomenon, and for this reason an optical system with such a PSF is called diffraction-limited system. The ability of an optical system to distinguish details in the image it produces is called *angular resolution* $\Delta\theta$ (or *resolving power*). For a telescope, it can be addressed as the smallest angle between two point sources for which distinct images are produced. In particular, for a diffraction-limited system, the *Rayleigh criterion*, schematized in Fig 2.2, states that this angle is the one for which the central intensity peak of one image falls upon the first minimum of the other. In other words:

$$\Delta\theta = 1.22\frac{\lambda}{D} \tag{2.4}$$

where D is the diameter of the telescope aperture and λ is the wavelength of the incoming radiation.

[4]Main references for this subsection: Bely (2003); Taylor (1996).

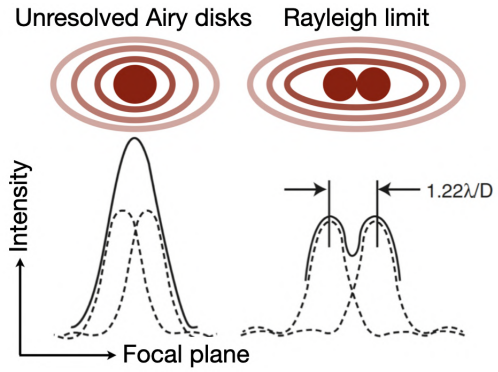

Figure 2.2: Illustration of the Rayleigh criterion. The two axis refer to the image plane. In red is drawn the way two distinct sources are imaged by an optical system. The black solid and dotted lines outline the PSF profiles in the unresolved *(left)* and resolved *(right)* case. *Figure adapted from Bely (2003).*

One thing, complementary to the angular resolution, that we want to achieve when performing any kind of observation is a good detector *sensitivity*, or *signal-to-noise ratio* (SNR, S/N). Modern astronomical detectors work so that the recorded signal is proportional to the number of photons received. The problem is that there are several "background" sources of radiation, besides the object of interest (natural sources in the sky, atmospheric emission, thermal emission from the telescope, and side effects in the detector itself), that affect the detector. Besides, also the global telescope's efficiency (atmosphere transparency, detector efficiency, optics efficiency, etc...) is a factor to take into account.

What actually matters to the problem of background sources are the fluctuations around the mean value of the background, the *background noise*. Assuming that background processes are the result of independent events (i.e., photons hitting the detector) occuring at a constant rate, they can be described by Poisson statistics. Which implies that their standard deviation, the *root mean square* (rms) of values of the background fluctuation, is simply the square root of the average number of events multiplied by the integration time (under root) (Taylor 1996). The standard way to express the significance of a detected signal and the precision of its measurement is to take the ratio between the signal and the fluctuation of the background, forming in fact the S/N. It basically tells us which is the "strength" of the measured signal in units of its standard deviation. In other words, assuming a Gaussian distribution for the errors inherent in a measurement (which is generally the case), if the resulting measurement has $S/N = 1$, it means there is a 68% probability that the signal is real. But at the same time, there will be about one chance in three that it is not real! Hence, a S/N of 1, or "1σ" detection, is not regarded as a credible detection of a signal. This is why we usually rely upon "3σ" detections (3 over 1000 chance of non significance), or, even better, "5σ" detections (1 in 10^5 chance of non significance).

The S/N formulation, whose exact shape is beyond the scopes of the thesis, incorporates all "instrumental features" of the observation, such as the detector's efficiency, and the telescope instrument optics. Moreover, it is this formulation that permits the determination of the faintest detectable magnitude and the exposure time required for a given observation.

2.2 Pills Of Interferometry

[5]In principle, according to (2.4), for a given wavelength is possible to improve the angular resolution by increasing the diameter of the telescope aperture. Nevertheless, it is impractical to build single telescopes with diameters larger than ~ 300 m, and with smaller diameter telescopes the resolution would not reach the mas (e.g. with the IRAM-30m telescope at $\lambda = 1$ mm, the resolution would be $\approx 0.03"$). Yet, there is another way of collecting light from astronomical sources, that allows to achieve better angular resolutions: it consists of splitting the collecting area of photons into two or more widely separated apertures. In this way, we can achieve a resolving power $\Delta\theta \simeq \lambda/b >> \lambda/D$, so as to disentangle sub-arcsecond features, by opportunely combining the output of at least two reflectors of diameter D separated by a distance, or *baseline* b (being $b > D$). These facilities are referred to as interferometers.

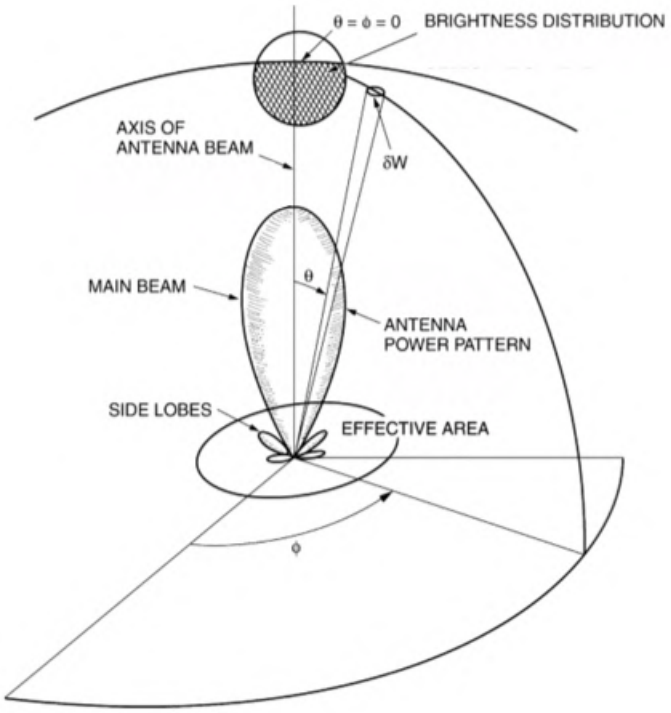

Figure 2.3: The antenna power pattern. *Figure adapted from Robert A. Meyers, Encyclopedia of Physical Science and Technology (Third Edition), 2003.*

2.2.1 The Power Of Interferometers

As anticipated at the end of the previous section, in order to practically better the angular resolution, we can build km-sized arrays of antennas, ergo, interferometers, and reconstruct an equivalent sized unique antenna. In an interferometric array, each antenna collects a certain amount of energy, that is, an electric field pattern, so that the power pattern of the interferometer is the square of the total collected electric field. Referring to Fig. 2.3, we call *main beam* the angular region containing the principal response of the interferometer. Outside this region there are the so called *sidelobes*, which represent unwanted smaller peaks where the interferometer's power is

[5]Main references for this section and subsections therein: Wilson et al. (2013); Condon & Ransom (2016); Ishak (2019).

wasted, because the antennas also collect energy from directions in the sky not concerning the targeted source. These sidelobes must be minimized in order to produce high-fidelity and high-resolution images, and this is basically the scope of the cleaning procedure applied on the raw output of interferometers (see § 2.2.5).

In this context, the angular resolution therefore measures the main beam width, that is, the central peak of the power pattern between its first two zeros. Defined the power pattern as a function of the sky plane direction cosine $l \equiv sin(\theta)$ (being θ the angle subtended by the source direction from the aperture axis), these zeros are identified by the condition $l \pm 1$. It also means that the beam resolving power defines the *minimum recoverable scale* by the interferometer.

Since any interferometer with $N >> 2$ antennas can be treated as $N(N-1)/2$ independent two-element interferometers, it is worth to understand in detail how this basic interferometer works.

Following the notation shown in Fig. 2.4, we can start by defining as $\hat{s}$ the unit vector directed towards a distant point source, and as $\vec{b}$ the baseline vector pointing from antenna "1" to antenna "2"; θ is the angle between $\vec{b}$ and $\hat{s}$, while $b\ sin\theta$ is the projected baseline. Plane waves coming from the target point source must travel an extra distance $\vec{b}\cdot\hat{s} = b\cos\theta$ to reach antenna "1" with respect to antenna "2", so that the output of the first is the same as the output of the second, but retarded in time by the so called *geometric delay*:

$$\tau_g = \frac{\vec{b}\cdot\hat{s}}{c} \tag{2.5}$$

Considering now for simplicity a *quasi-monochromatic interferometer* (sensitive only to radiation coming from a very narrow frequency band $\Delta\nu \ll 2\pi/\tau_g$, centered on $\nu = \omega/(2\pi)$), the output voltages of the two antennas are the ones indicated in Fig. 2.4. These voltages get amplified, multiplied (through *cross-correlation*), and time averaged by the so called *cosine correlator*. This delivers an output voltage, or response R:

$$R = \left(\frac{V^2}{2}\right) cos(\omega\tau_g) \tag{2.6}$$

characterized by an amplitude $A = V^2/2 \propto$ flux density of the point source, and a phase $\phi = \omega\tau_g$ dependent on the geometric delay and the frequency.

Convolution And Cross-Correlation

The convolution is a mathematical operation defined as:
$f * g \equiv \int_{-\infty}^{\infty} f(u)\ g(x-u)\ du$
where the function $f(x)$ gets multiplied by the time-reversed kernel function g shifted by an amount u. For an antenna or imaging system, the kernel identifies with the beam. Noticeably, the convolution theorem states that the Fourier transform of the convolution of two functions is the product of their individual FTs: $f * g \Leftrightarrow F \cdot G$.
The cross-correlation is a very similar operation to convolution, except that the kernel g is not time-reversed: $f \bigstar g \equiv \int_{-\infty}^{\infty} f(u)\ g(u-x)\ du$.
Closely related to the convolution theorem, the cross-correlation theorem states that the Fourier transform of this operation performed between two functions is equal to the product of the individual Fourier transforms, where one of them has been complex conjugated: $f \bigstar g \Leftrightarrow F \cdot G$.

Since the Earth's rotation changes the source direction $\hat{s}$ with respect to the baseline vector, the correlator output voltage varies in time as a sinusoid. These sinusoidal components are called

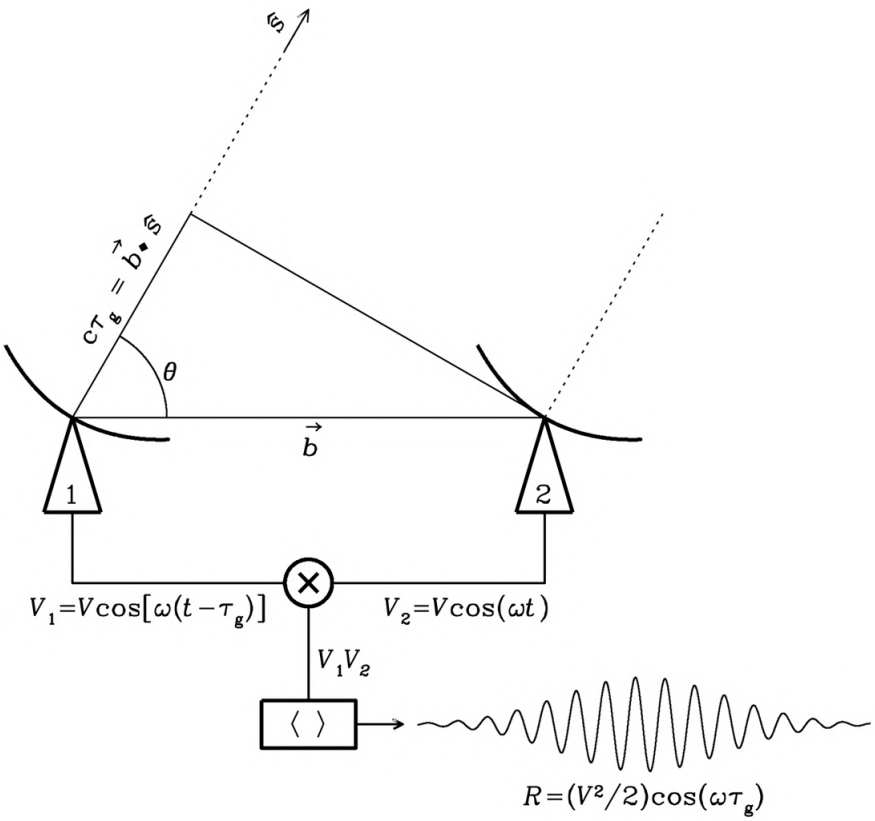

Figure 2.4: A two-element quasi-monochromatic multyplying interferometer.
Figure adapted from Condon & Ransom (2016).

fringes, and their phase ϕ can be written down as:

$$\phi = \omega\tau_g = \frac{\omega}{c}\ b\ cos\theta \tag{2.7}$$

Given the dependency of the fringe phase on θ, the fringe period $\Delta\phi = 2\pi$ will correspond to an angular distance $\Delta\theta = \lambda/(b\ sin\theta)$.

Since we do not measure the absolute phases of signals received by individual antennas, but the phase of the fringes (the phase difference of pairs of antennas), it is needed to set a reference antenna and calibrate the phases of all the other antennas with respect to the reference one, which is fixed at null phase.

To sum up, the response of a two-element interferometer is unique and the total power pattern is obtained by convolving the power pattern of the two-element interferometer with the product of the voltage patterns of the individual antennas.

An interferometer of N antennas, instead, will produce $N(N-1)/2$ independent responses, one for each antenna pair, so that the total power pattern, the so-called instantaneous *synthesized beam* (exemplified in Fig. 2.5) is the arithmetic mean of these responses, and it approaches a Gaussian profile with increasing N.

Assuming that the antennas are identical, the product between voltage patterns is the power pattern of the individual antennas: it is called *primary beam* of the two- or N-interferometer, and in Gaussian approximation is equal to $\theta_{PB} \approx \lambda/D$. In the end, for an interferometric array, the resolving power is $\Delta\theta \approx \lambda/b_{max}$, where b_{max} is the maximum baseline separating the dishes in the array, much larger than the single dish diameter D.

We can identify other two main advantages of interferometric observations over single dish observations, and we can express them in terms of *effective collecting area* (A_e), and the *field of view* (θ_{FOV}), or FOV.

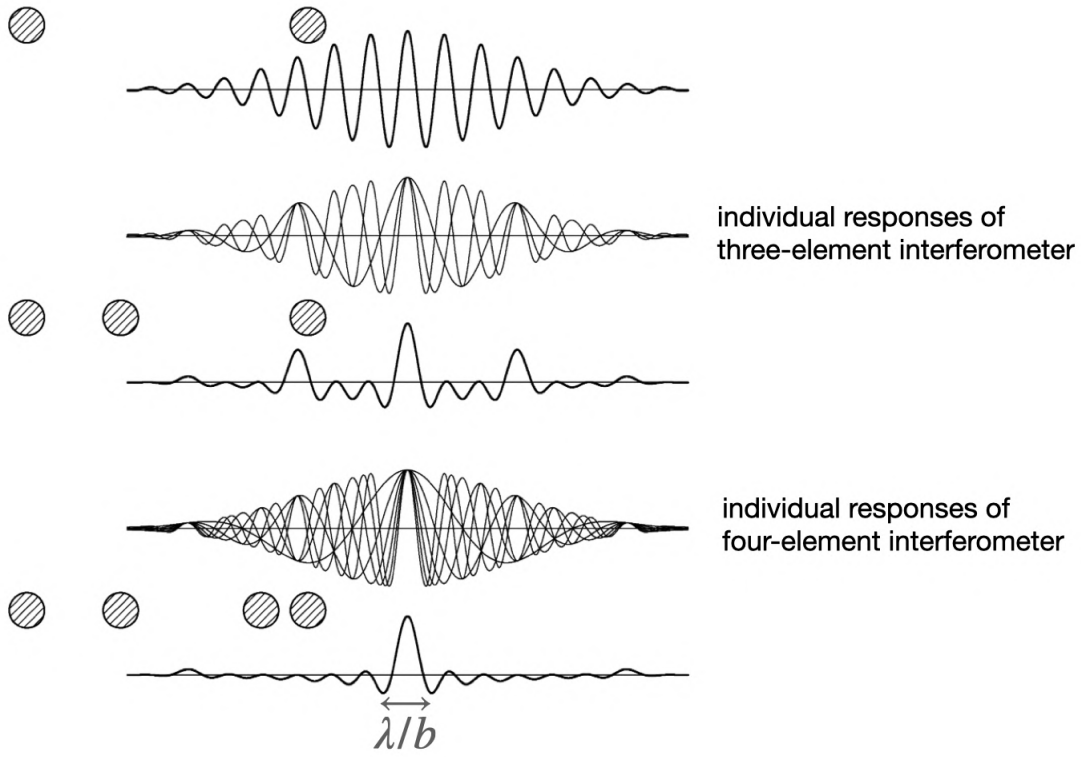

Figure 2.5: The instantaneous synthesized beam of N-antenna interferometers. From top to bottom, the instantaneous point-source responses of a two-, three-, and a four-element interferometer with overall projected baseline length b. In the case of the four-element interferometer, the synthesized main beam is nearly Gaussian, with angular resolution $\theta \approx \lambda/b$. However, sidelobes are still significant, and there is an overturned "bowl" caused by the lack of spacings shorter than the diameter of individual antennas. *Figure adapted from Condon & Ransom (2016).*

The effective collecting area quantifies the capability of collecting photons onto the detector area, and it is formally defined as the ratio of twice the received power by an antenna P_ν to the power per unit detector area, that is, the source flux density S_ν of an unpolarized source. For a N-elements interferometer, $A_e \lesssim N\pi D^2/4$, that can be larger than for a single dish radiotelescope, in which case $A_e \lesssim \pi D^2/4$.

Thirdly, the interferometric FOV is defined as the angular response contained within the two external nulls of the $sinc^2(l)$ power pattern of a single antenna; for an array of interferometers, it can be expressed as $\theta_{FOV} \approx 1.2\ \lambda/D$. The FOV should not be confused with the primary beam: the former describes the maximum angle over which the interferometer can observe multiple points in the sky at the same time (depending on e.g. array configuration, observed frequency, and correlator design); the latter defines the region of space that a single antenna can observe effectively at any given time, and is typically characterized by the antenna pattern full-width half maximum (FWHM), i.e. measure of the width of the beam at half of its maximum sensitivity.

Another important scale to be defined, together with the FOV and the minimum recoverable scale, is the *maximum recoverable scale*, defined as λ/b_{min}, i.e., the largest detectable angular scale by the interferometer, being b_{min} the minimum array baseline. The minimum and the maximum recoverable scales are related by the *sampling theorem*: if we want to reconstruct a continuous function extending up to a maximum spatial frequency u_{max} (remember the definition $u \equiv x/\lambda$), we must use uniformly spaced samples separated by an interval $\Delta x < (2\ u_{max})^{-1}$. If this theorem is not satisfied, we produce *aliasing*, that is, the signal frequency gets misidentified and there is a loss of information about the signal.

The Power Of Single-Dishes

Having a maximum recoverable scale, interferometric observations miss something. Indeed, the flux changes on scales larger than their maximum recoverable scale. This is due to the fact that the diameter D of an antenna must be smaller than the smallest baseline b_{min} to prevent overlapping. Consequently, the interferometer can not cover the 0-spacings in angular frequency. This issue with interformetric observations is referred as *missing short spacings problem*, and it can be addressed only by using a single-dish radio telescope, having a diameter $D > b_{min}$.

Interferometric Temperatures

Radioastronomers find it convenient to refer to the power of various signals from a radio telescope in terms of the equivalent temperature of a matched termination on the receiver. According to the *Rayleigh-Jeans approximation* of black-body spectrum ($h\nu << kT$), any signal power P can be calculated as:

$$P = kT\Delta\nu \tag{2.8}$$

so that it is delivered in a bandwidth $\Delta\nu$ to a matched resistor at temperature T. The receiver bandwidth can be made explicit as: $[\nu_{RF} - \frac{\Delta\nu}{2}, \nu_{RF} + \frac{\Delta\nu}{2}]$, being ν_{RF} the frequency of observation in the radio band. The bandwidth can be split in different sub-bands, called *spectral windows* (SPWs), with each SPW divided in multiple *channels*, which are small windows in frequency.

We call *antenna temperature* the component of the power resulting from a cosmic source observed with an antenna of given effective area A:

$$T_A = \frac{P_A}{k\ \Delta\nu} \tag{2.9}$$

Instead, we call system noise temperature (T_{sys}) the total noise power from all sources referenced to the input of a radiometer, which measures the time-averaged power of the input noise in the bandwidth, and that is connected to the output of a radiotelescope. It is obtained by considering the sum of different terms, including the antenna temperature, and the noise associated to the atmosphere. It can be shown (Condon & Ransom 2016) that the error on the measure of T_{sys} is (*ideal radiometer equation*):

$$\sigma_{T_{sys}} \approx T_{sys}/\sqrt{\Delta\nu\ \tau} \tag{2.10}$$

being τ the total observing time. It means that the weakest detectable signals have to be larger than this rms, that is, they can be as low as a certain fraction of T_{sys}, depending on the type of receiver, bandwidth and observing time (typically, signals as faint as $\sim 5 \times 10^{-4}\ T_{sys}$ are detectable). In the limit where the temperature contributed by a point source is much smaller than T_{sys}, for an interferometer of N antennas the point source rms noise is:

$$\sigma_{rms} = \frac{2k\ T_{sys}}{A_e\sqrt{N(N-1)\Delta\nu\ \tau}} \propto \frac{1}{\sqrt{\Delta\nu\ \tau}} \tag{2.11}$$

where k stands for the Boltzmann constant ($k = 1.380649 \times 10^{-23}\ m^2\ kg\ s^{-2}\ K^{-1}$), A_e is the effective area. Given that high sensitivity means low value of σ_{rms}, the point source sensitivity of a two element interferometer is $\sqrt{2}$ better than that of each antenna.

2.2.2 What We "Collect": Visibilities

For an extended source, characterized by a certain sky brightness distribution $I(\hat{s})$, where $\hat{s}$ is the source direction in the sky, the response R of the interferometer is obtained by treating the source as the sum of independent point sources. In this sense, the correlator delivering the interferometric response performs an operation of convolution between the source brightness distribution of the source with the total power pattern of the interferometer, integrating this product over the sky. The output measurements of this operation are the so called *visibilities*:

$$\mathcal{V} = \int I(\hat{s}) e^{\left(-i2\pi\vec{b}\cdot\hat{s}/\lambda\right)} d\Omega \tag{2.12}$$

where $\vec{b}$ indicates the baseline vector, and $d\Omega$ the solid angle subtended by an element of the radiosource. In general, a visibility is a complex number $\mathcal{V} = A\ e^{-i\phi}$, characterized by an amplitude $A = \sqrt{R_c^2 + R_s^2}$ and a phase $\phi = \arctan(R_s/R_c)$.

R_c and R_s respectively correspond to the responses of the sine and cosine correlators, which take into account the odd and the even components of $I(\hat{s})$. Indeed, as for any real function, we can always write the intensity as the sum of two real functions having specific symmetries. The combination of cosine and sine correlators is called *complex correlator*, responsible for the overall interferometric response.

In particular, given a pair of i-th and j-th antennas, the complex correlator produces one visibility measurement for each baseline, integration time, polarization, and channel, according to:

$$\mathcal{V}^{ij}(\tau) = \lim_{T\to\infty} \int_{-T/2}^{+T/2} \mathcal{V}^i(t)(\mathcal{V}^j)^*(t+\tau)\, dt \tag{2.13}$$

where $(V^j)^*$ indicates the complex conjugate of the j-th visibility, T stands for the integration time, and τ is the *delay tracking error*, e.g. the difference between the geometric delay and the instrumental one, which is introduced by the correlator in one of the two antennas.

2.2.3 To Start With: Calibration

The first thing we must do once we are delivered with the dataset of visibilities of the requested target is to calibrate them. The *calibration* procedure is needed to correct for effects that may interfere with the measurements due to the instrument, and due to local temporary conditions affecting the observations. It practically consists in determining the so-called *complex gains* $\mathcal{G}^{ij}$, which corrupt the true visibilities, and then applying this series of corrections and scaling factors to the target data. In this sense, the observed visibilities can be defined as:

$$\mathcal{V}^{ij}_{obs}(\nu, t) = \mathcal{G}^{ij}(\nu, t) \cdot \mathcal{V}^{ij}_{true}(\nu, t) \tag{2.14}$$

The complex gains depend on many factors, to name the most important, the atmosphere absorption and variability. In the calibration process, two main assumptions are taken, *(i)* the corruptions are antenna-based, that is, the gains can be factorized for each antenna, and *(ii)* the temporal and frequency dependence of the corruption can be determined independently (they are lightly coupled), so that the gains can be factorized in a time-dependent and in a frequency-dependent term.

Since, as the visibilities, the gains are complex numbers, we must determine their amplitudes and phases. Yet, in order to resolve the system of equations enclosed in (2.14), it is necessary to fix a reference antenna, but also to observe sources for which the true visibilities are known. These sources are called *calibrators*, and they are chosen to be bright point-like sources, of known flux, observed at the phase center, so as to be characterized by constant amplitude and zero phase. We can distinguish two major classes: *flux* or *primary calibrators*, and *phase* or *secondary calibrators*. The former set the flux density scale for the target visibilities. The latter serve to correct for the time-dependent response of each interferometric antenna, induced by the atmosphere's variability. Therefore, an equation similar to (2.14) must be resolved also for calibrators, where the true visibilities are replaced by the model visibilities model of known amplitude and phase.

Self-Calibration

In order to correct for the atmospheric variability, the secondary calibrator gets regularly observed before and after the target, so that its amplitude and phase corrections can be applied to the target by linear interpolation.

However, there are some cases in which the interpolation is not enough to find solutions for the target. The atmospheric variability with time could be not well sampled, or these variations are too strong/rapid. Also, accounting for the "spatial" variability of the atmosphere, the secondary calibrator may be too far from the target, inducing a phase and amplitude offset. Observing the calibrator more often is hence not an efficient solution. But if the target displays a good S/N in the continuum, we can proceed with the so called *self-calibration*. This procedure can be used as an additional calibration step, if the phase calibration has found bad solutions for the target, or also if it has not been possible to do the phase calibration at all. The name of this process comes from the fact that we use the target source itself as a model (this is why we need a good S/N), and that visibilities get calibrated upon this model. For the scopes of this work, we can conclude by saying that the self-calibration is looped until the rms associated to the restored target image does not improve significantly anymore.

2.2.4 What We See: Images

What we are actually interested when performing interferometric observations is to obtain images of the target(s), that is, producing maps of the true source brightness distribution $I(\hat{s})$ starting from the observed sampled visibilities.

The visibilities can be conveniently expressed in a cartesian coordinate system called (u, v) plane (or *Fourier plane*), that contains the projected components of the inteferometer baseline vector $\vec{b}$ (in units of λ). The coordinates of this reference system, u and v, form a plane perpendicular to $\hat{s_0}$ (the centre source direction), and perpendicular to w, which is a third coordinate pointing in the source direction $\hat{s}$. It is also possible to define a (l, m) plane, which represents the tangential plane to the celestial sphere in the point $\hat{s_0}$. Fig. 2.6 summarizes the theoretical planes at play. Focusing on directions close to $\hat{s_0}$, and considering the *small angle approximation* ($\theta \lesssim 1$ rad, $l \approx \theta$), it is possible to switch from the *(l, m)* to a plane formed by rectilinear coordinates x and y respectively parallel to u and v, the *(x, y)* plane. In this plane a very important result holds (if the field of view is small and we can neglect the wz term): the complex visibilities express the 2D Fourier Transform of the source brightness distribution:

$$\mathcal{V}(u, v) \approx \int\int I_\nu(x, y) e^{[-i2\pi(ux+\nu y)]}\, dx\, dy \tag{2.15}$$

Conversely, the source brightness distribution is the 2D anti-Fourier Transform of the measured visibilities:

$$\mathcal{I}(x, y) \approx \int\int V(u, v) e^{[+i2\pi(ux+\nu y)]}\, du\, dv \tag{2.16}$$

These integrals are not exact due to the fact that the brightness distribution is modified by the primary beam of individual antennas, resulting in a decreasing intensity moving from the center of the FOV towards its edges. This effect can be corrected by applying the so called *primary beam correction*.

The Fourier Transforms

The Fourier transform is a reversible, linear transform of any complex function $f(x)$ of the real variable x that is both integrable over $[-\infty, +\infty]$, and contains only finite discontinuities. This transform is defined as a function of the real variable s:
$F(s) \equiv \int_{-\infty}^{\infty} f(x) e^{-2\pi i s x}\, dx$. The inverse transform, or anti-Fourier transform, is:
$f(x) \equiv \int_{-\infty}^{\infty} F(s) e^{+2\pi i s x}\, ds$.

Since each baseline (a two-element interferometer) samples a single point in the (u, v) plane, and so delivers a single visibility, in order to increase the sampling of the visibility function $\mathcal{V}(u, v)$ as much as possible, we need to have many baselines. This is obtained with the so-called *aperture synthesis procedure*, which practically consists in letting the Earth's rotation doing its job. The fact that Earth rotates, indeed, allows to increase the number of baselines and so the number of sampled points in the Fourier plane. In the end, the total number of (u, v) discrete points sampled by the interferometer array constitutes the so called *u-v coverage*, or *sampling function*, for which the visibilities are set to have $A = 1$ and $\phi = 0$.

To wrap up, in order to produce a map of the true sky brightness distribution of the source $I(l, m)$ starting from the observed sampled visibilities, we first need to calibrate the visibilities, and then perform a deconvolution (through the `CLEAN` algorithm described in § 2.2.5) of the "dirty map", that is, the anti-FT of the observed visibilities, from the *dirty beam*. Indeed, the sidelobes of the interferometric beam are quite significant (this is why the name "dirty"), so the beam must be "cleaned" from them to deliver high-fidelity images of the target(s).

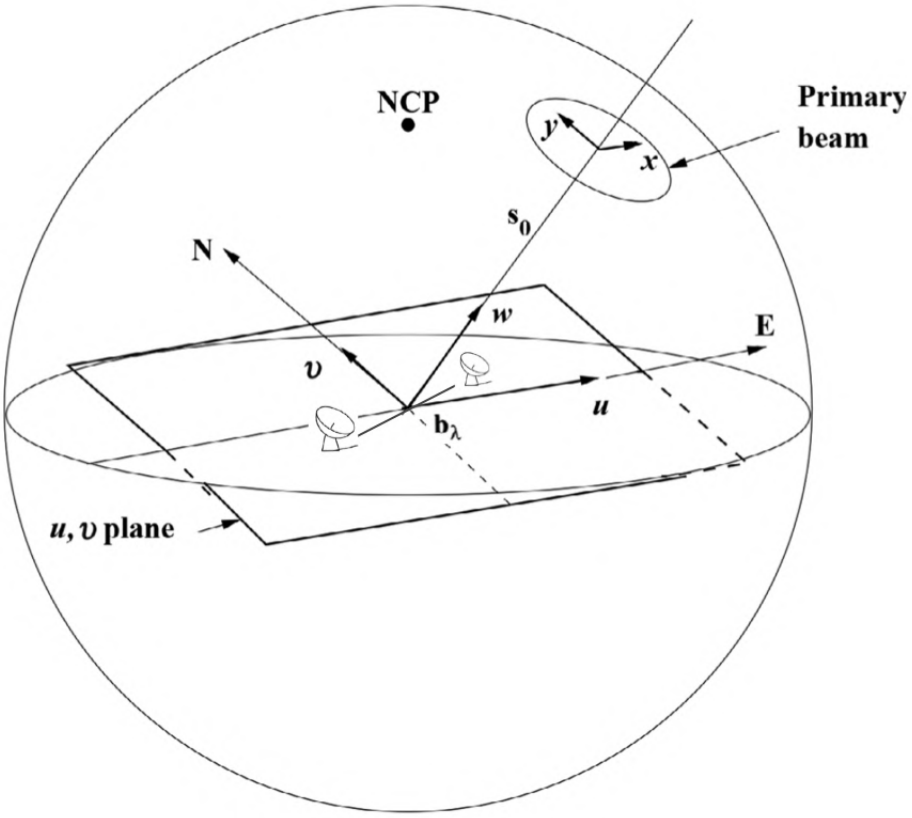

Figure 2.6: The interferometric planes. The geometrical relationship between a two-element interferometer, a celestial source with phase center in direction $\hat{s}_0$, the *(u, v)* plane, the celestial sphere, and the *(x, y)* plane; NCP is the North Celestial Pole. *Figure adapted from Ishak (2019).*

In light of the above, we can better contextualize the sampling theorem, introduced in § 2.2.1. The continuous function we want to sample is the visibility function, which extends from the minimum u_{min} to the maximum u_{max} baseline projected on the *(u,v)* plane. The criterion of this theorem for good sampling will therefore be related to the pixel size Δx, which must be at least half of $1/u_{max}$, in our case, half of the beam (but it is common procedure to take Δx as 1/4 or 1/5 of the beam).

A Polyhedric Output: Datacubes

Once data are calibrated, the images of a source sky brightness distribution at each wavelength are reassembled in different ways, creating multiple images or multiple image planes, and folded in a *spectral cube*, or *datacube*.

This can be visualized with a 3D graph of which two spatial axis (y and z, usually declination and right ascension), and one spectral axis (x, frequency or wavelength), as illustrated in Fig. 2.7. To each location *(y, z)* is associated an intensity as a function of the frequency/wavelength. Slicing the cube in parallel to the spectral axis, we pick the spectrum relative to a given pixel; with a "slice" perpendicular to the spectral axis, instead, the channel image at a certain frequency/wavelength. From a datacube we can therefore obtain both spectral and continuum info about a celestial object. For the latter, collapsing all data in a single channel and excluding spectral lines, we are left with a continuous emission that traces the background source(s) of the observations. In the case of an astrophysical radio source (in the sub-mm-IR range), this background emission comes from heated dust grains (e.g. by the external radiation field and/or the central accreting protostar, see Chapter 1).

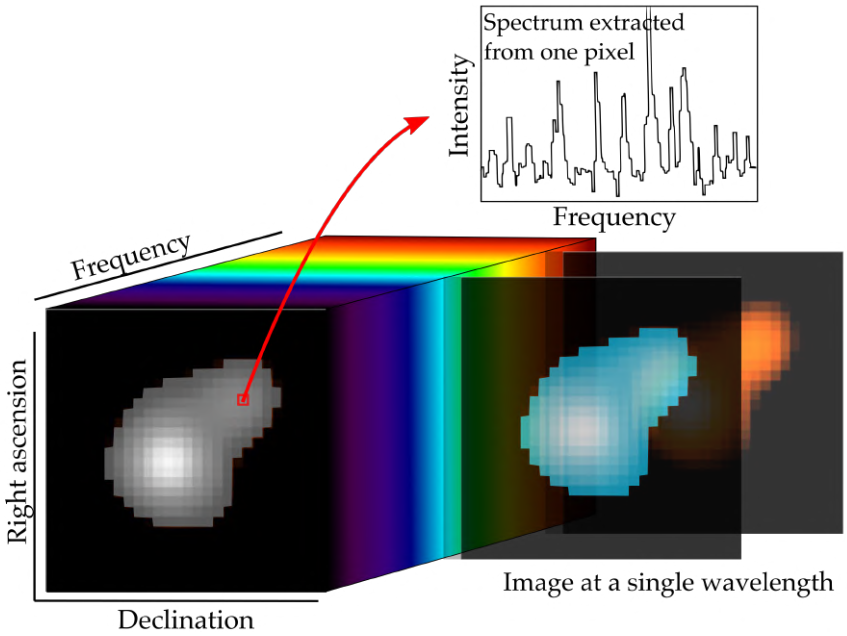

Figure 2.7: Artistic illustration of a datacube. *Credits: Marta De Simone.*

Producing an image of each channel, the rms tends to be larger than for a continuum image, having a much smaller $\Delta\nu$ in (2.11).

Given a datacube's spectrum, the y- and x-axis are normally in units of Jy beam^{-1} for the flux and MHz/GHz for the frequency. Actually, a spectrum can be more useful if the flux is converted in units of *brightness temperature* T_B (K).

Indeed, the radiative transfer theory tells us, under the LTE (see § 2.4.1), Rayleigh-Jeans and $\tau \to \infty$ (being τ the optical depth; (2.21)) assumptions, which can be all valid in the mm and sub-mm range, that the brightness and the thermodynamic temperature of a black body are strictly proportional (Wilson et al. 2013). Therefore, it has become usual in radio astronomy to measure the brightness I_ν of an extended source by its correspondent brightness temperature, which is precisely expressed in Kelvin. Yet, what we measure from a spectrum is the flux density S_ν, which is in Jy beam^{-1}, that is related to I_ν through (2.2). Hence, we need to convert Jansky into Kelvin, taking into account the beam solid angle of the observations. One can show (Wilson et al. 2013) that S_ν and T_B are linked by:

$$T_B = \frac{\lambda^2}{2k\ \Omega}\ S_\nu \tag{2.17}$$

where λ is the radiation wavelength, k the Boltzmann constant, and Ω the beam solid angle. Now, given a deconvolved image, the clean beam is modelled as a Gaussian with peak of unity and θ_{maj} and θ_{min} as, respectively, the half-power beam widths along the major and minor axes. The area of this Gaussian beam is defined by:

$$\Omega = \frac{\pi\ \theta_{maj}\ \theta_{min}}{4\ ln(2)} \tag{2.18}$$

Substituting Ω in (2.17),and converting all constants to a pre-factor finally results in (Condon & Ransom 2016):

$$T_B = 1.222 \times 10^3\ \frac{I_\nu}{\nu^2\ \theta_{maj}\ \theta_{min}}\ [K] \tag{2.19}$$

where the brightness is in mJy beam^{-1}, the frequency in GHz, and the half-power beam widths in arcsec.

To sum up, knowing the size of the source, the observed flux and the frequency of observation, it is possible to convert datacubes' units from Jy beam^{-1} to Kelvin. However, this is true only under the theoretical assumptions stated at beginning and considering a Gaussian clean beam.

2.2.5 What We Do: Cleaning Algorithms

The calibrated visibilities are just the starting point of performing imaging. The key point is to subtract the sidelobes from the interferometric response through a deconvolution process. The `CLEAN` algorithm (Högbom 1974; Clark 1980; Schwab & Cotton 1983) represents the standard method used, usually in the image plane, in order to retrieve the true brightness distribution $I(u, v)$ of a given celestial source. This kind of algorithm assumes that the radio source we want to image is an ensemble of a finite number of point-like sources without background. We can distinguish two major steps:

1. *Clean.* The algorithm scans the dirty map looking for the brightest points, which are modelled as Dirac delta δ functions. Then, it iteratively subtracts from the dirty map the so called *clean components*, e.g. fractions of the dirty beam at the peak position of the δ functions found. What is left from this operation at each iteration is a residual image, in which the brightest points have been removed with all their sidelobes. Yet, after each subtraction fainter source components can pop up, so the algorithm goes on cleaning the dirty image from the new deltas until a stopping criterion is reached. At this point, the final residual image is expected to contain only noise, with no source components left.

2. *Restore.* The algorithm takes all the clean components collected in the previous step, adds the residual noise, and convolves them with a clean beam of choice, that is usually an elliptical gaussian fit of the central lobe of the dirty beam (no sidelobes!).

The CASA `tclean`[6] task has been designed to Fourier transform interferometric visibilities, and cleaning the resulting images. It is characterized by several adjustable parameters, among which, also in relation to this work, we want to draw attention to: pixel size, image size, stopping criteria, visibilities weighting, masking options, and deconvolving algorithms.

Pixel And Image Size

As stated in the previous section, in order to practically satisfy the sampling theorem, and successfully reconstruct the target brightness distribution in the image plane, we need a good sampling of the interferometer resolving power, hence, a pixel size to be 1/3 to 1/5 of it. Once computed the pixel unit of our image, we can derive the image size by dividing the interferometric field of view by the pixel size. It is common practice to take a greater value than the one we can derive, in order to retrieve also the most external sidelobes of the FOV.

Stopping Criteria

We can identify two major criteria to stop the iterative cleaning process. The first one is based on the maximum number of iterations to perform (`niter` parameter), corresponding to the total number of clean components to be subtracted. The second criterion is based on the minimum peak flux to expect for a clean component (controlled via the `threshold` parameter). There is no magical combination of values for these two criteria to achieve good image fidelity. It is matter of a trial-and-error process in which we reasonably decrease the threshold, and increase the number of iterations. If the stopping criterion is the number of iterations, but the residuals still contain source components, it is the case to increase this parameter without modifying the threshold. A good prescription for the threshold is to set it as much as (or higher than) the theoretical rms expected when observing the target, so as to prevent the algorithm from creating artificial artefacts in images.

Weighting

The measured visibilities contain noise and they are not uniformly distributed across the *u-v* plane; some u-v ranges, typically the short spacings, are sampled more than others. It is therefore critical to apply a certain weighting to the gridded visibilities, i.e. multiplying each visibility value by a weight before distributing it on the *(u-v)* grid. There are three main weighting schemes, among which we can choose:

- *Natural.* Each grid cell is given a weight proportional to the number of visibilities in it. This weighting scheme provides the maximum imaging SNR at the expense of angular resolution.
- *Uniform.* The weights correspond to the original data weights divided by the total weight associated to the density of *(u-v)* points in each cell. The weight grid results in a PSF with a narrow main lobe and suppressed sidelobes, hence higher angular resolution. However, since every cell is given an equal weight, heavily sampled areas of the *(u,v)* plane get down-weighted, decreasing the imaging sensitivity.
- *Briggs.* Intermediate scheme generating a PSF, through an adjustable parameters called `robust`, which can vary smoothly between "natural" (`robust` $= +2$) and "uniform" (`robust`

[6] https://casadocs.readthedocs.io/en/stable/api/tt/casatasks.imaging.tclean.html

= −2). High signal-to-noise data are weighted by density of *u-v* points in order to narrower the PSF, while low signal-to-noise data are weighted to optimize for sensitivity.

Masking

Given the large number of degrees of freedom in interferometric images, and the high non-linearity of the cleaning process, we can try to constrain the possible model solutions for cleaning by masking regions that contain emission. In this sense, *masks* are used to leave behind fewer undeconvolved residuals after the cleaning, hence increasing the image fidelity.

A mask is presented as a normal, single polarization image file, where all zero values are interpreted as being not masked, and all non-zero values are interpreted as masked. In `CASA`, we can either create masks by hand during the cleaning process (manually constraining where `CLEAN` finds model components), or we can opt for, by setting the `auto-multithresh` parameter and its subparameters, an automated mask algorithm by multiple thresholds, the so-called `AUTO-MULTITHRESH` algorithm (Kepley et al. 2020).

Given the large data volumes produced by current interferometric arrays, manually masking images has become an extremely time-consuming process. For this reason, and for the sake of the thesis adopted methodology, I will review the operation of the `AUTO-MULTITHRESH` algorithm.

First, it identifies regions that are either above a signal-to-noise limit or a sidelobe level, whichever is higher (decided by the `noisethreshold` and the `sidelobethreshold` parameters; see below). The mask regions that are smaller than some fraction of the beam get removed ("pruned") to avoid including noise peaks in the masked region. Any extended low SN emission region surrounding the initial mask is included in the mask by expanding the noise/sidelobe mask into this kind of region[7]. The mask is also expanded to include a buffer region around the masked emission by convolving the mask with a Gaussian, and then retaining only regions that are above a certain percentage of the peak in the smoothed mask. In the case of datacubes, the algorithm masks each channel independently. The algorithm operates on the residual image at the beginning of every minor cycle, so that the mask updates as the clean progresses.

The subparameters for `AUTO-MULTITHRESH` are defined in terms of fundamental image properties rather than exact numerical values (e.g., in arcseconds or Jy beam^{-1}) We can identify five primary auto-multithresh parameters:

- *Noisethreshold*. It sets the signal-to-noise threshold above which the emission is masked when the mask is created.
- *Sidelobethreshold*. It sets a threshold based on the sidelobe level above which the emission is masked when the mask is created.
- *Minbeamfrac*. It sets the minimum size a region must be to be retained in the mask; any mask smaller than this gets pruned.
- *Lownoisethreshold*. It sets the threshold into which the initial mask is expanded in order to include low signal-to-noise regions in the mask.
- *Negativethreshold*. It sets the signal-to-noise threshold for absorption features to be masked.

`AUTO-MULTITHRESH` also comes with a pipeline of standard values for all parameters (optimized for different configurations of the ALMA arrays) that have been extensively tested (Kepley et al. 2020).

[7] Absorption regions can also be masked using a similar method, but the resulting masks are not pruned or extended into low signal-to noise regions.

Deconvolvers

Image reconstruction in `CASA` divides in an outer loop of *major cycles*, which implements transforms between the *u-v* and image planes (such as, model image $\hookrightarrow$ model visibilities, residual visibilities $\hookrightarrow$ residual image), and an inner loop of *minor cycles*, which operates in the image plane only (to find a new flux component and update the model and residual images).

To each minor cycle is associated a deconvolver algorithm (`hogbom, clark, multiscale, mtmfs, mem, clarkstokes, asp`), which picks flux components from the whole image or within a mask. The two most used ones are `hogbom` (Högbom 1974), and `multiscale` (Cornwell et al. 2008). The former models the sky brightness distribution as Dirac delta functions, and is therefore ideal when imaging compact sources. The latter is helpful in presence of extended emission, because it subtracts from residual images specific multiscale components. The set of scale sizes to be used in this latter deconvolving algorithm should represent the sizes, that is, diameters in units of number of pixels, of source features in the image being reconstructed. It is recommended to use as smallest scale size that of a point source (0), as second the size of the synthesized beam, as third 3-5 times the synthesized beam, and so on[8].

The `tclean` task provides in output several useful files that we list here on the basis of the file extension name:

- *Image.* Final restored image.
- *Pb.* Primary beam model.
- *Mask.* Map containing the scanned regions in which `tclean` can search for the emission of the source.
- *Model.* Model image, containing only flux components.
- *Psf.* Interferometer dirty beam.
- *Residual.* Residual image left at the end of the deconvolution process.

In addition, if `tclean` runs with the parameter `pbcor = True`, then the primary beam correction is automatically applied on the output restored image, and a new image with extension "image.pbcor" is created It is also possible to correct for it after imaging using the `CASA` task `impbcor`[9].

2.3 Not Any Radio Interferometric Facility: The ALMA Observatory

ALMA (*Atacama Large Millimeter/Submillimeter Array*, Fig. 2.8)[10] represents the largest, ground-based, international astronomical radio telescope, located on the Chajnantor Plateau in the Atacama Desert of northern Chile, at an altitude of 5000 m. This achievement, started to take life in 2003 as the result of an international association between Europe (ESO[11]), North America (NRAO[12]) and East Asia (NAOJ[13]), in collaboration with the Republic of Chile. The ALMA inauguration took place in 2013.

[8] For numerical stability, the largest scale must be $<$ than the image, or mask, size as well as $\lesssim$ the scale corresponding to the lowest measured spatial frequency.

[9] https://casa.nrao.edu/docs/taskref/impbcor-task.html.

[10] https://www.almaobservatory.org/en/home/

[11] https://www.eso.org/public/

[12] https://public.nrao.edu

[13] https://www.nao.ac.jp/en/

Figure 2.8: View of the ALMA antennas on the Chajnantor Plateau under the Milky Way. *Credits: ALMA(ESO/NAOJ/NRAO).*

This interferometric radio telescope is composed of 66 antennas, operating on wavelengths from 0.32 to 3.6 mm. Its main array has 50 antennas, each with 12-meter diameters. This is complemented by the *Atacama Compact Array* (ACA), or *Morita array*, consisting of four antennas with 12-meter diameters and 12 antennas with 7-meter diameters. ALMA's antennas can be configured in different ways on the plateau thanks to movable pads, spaced at distances varying from 150 meters to 16 kilometers, to enhance spatial resolution (more extended arrays), or give better sensitivity to extended sources (more compact arrays).

The frequency range available to ALMA is divided into ten different receiver bands, ranging from band 1, starting at 35 GHz, to band 10, ending at $\sim$ 950 GHz. In particular, band 2 (67 - 90 GHz) is under development. Each band is designed to cover a tuning range that is approximately tailored to the Earth's atmospheric transmission windows.

ALMA's spatial resolution, referring to the FWHM of the synthesized beam, ranges from 0.5" (950 GHz) to 4.3" (110 GHz) in the most compact 12-m array configurations ($\sim$ 160 m), and from 20 mas (230 GHz) to 43 mas (110 GHz) in the most extended 12-m array configuration ($\sim$ 16 km). ALMA can deliver datacubes with up to 7680 frequency channels. The width of these channels can range between 3.8 kHz and 15.6 MHz, but the total bandwidth cannot exceed 8 GHz. For example, at 110 GHz, the highest spectral resolution implies a velocity resolution of 0.01 km/s; at 110 GHz.

With the final aim of "discovering our cosmic origins", the ALMA observatory has made sensational discoveries in all main science themes, spanning from the scales of cosmology and high redshift galaxies down to the Solar system.

In 2014, ALMA achieved its highest angular resolution and observed the protostellar HL Tau system, revealing a series of concentric and bright rings, separated by gaps. These represent the most unmistakable evidence for ongoing planet formation within a protoplanetary disk, and that planets form faster than previously thought (ALMA 2015). In 2018, ALMA has joined a global collaboration of radio observatories, the *Event Horizon Telescope* (EHT)[14], to form an Earth-size interferometer and allow astronomers to reveal the first image of a black hole (EHT 2019). As a final remark, in 2021, astronomers used ALMA to map the nearby Universe and reveal the diversity of star-forming galaxies (Leroy et al. 2021).

2.3.1 The ALMA FAUST Large Program

[15]The data used in this thesis project are part of the international ALMA Large Program FAUST (PI. S. Yamamoto), or *Fifty AU Study of the chemistry in the disk/envelope systems of Solar-like protostars*[16], approved in ALMA Cycle 6 (2018) and provided with 106.2 h and 45.6 h time respectively in 12m and 7m ACA configurations.

The goal of the project is to reveal and quantify the variety of chemical composition of the envelope/disk system at planet-forming scales ($\sim$ 50 au) in a sample of well-known 13 Class 0 and I protostars ($L_{bolometric} < 25\ L_{\odot}$) located in nearby star forming regions ($\leq$ 300 pc), and representative of the chemical diversity observed at larger scales (100-2000 AU).

The FAUST approach is to identify three zones (expected to possess distinct physical and chemical properties; see Fig. 2.9): the *envelope*, that is, the infalling–rotating material around the central protostar on scales of a few 100 AU; the *centrifugal barrier*, where the gas transits before entering the disk on scales of about 50 AU; the *disk*, at scales smaller than $\sim$ 50 AU, in which the gas settles in the rotationally supported disk. The idea is therefore to disentangle these zones with the help of their kinematic signatures and ALMA line images at 50 AU spatial resolution.

[14]https://eventhorizontelescope.org
[15]Main reference for this subsection: Codella et al. (2021), plus references therein.
[16]http://faust-alma.riken.jp/index.html

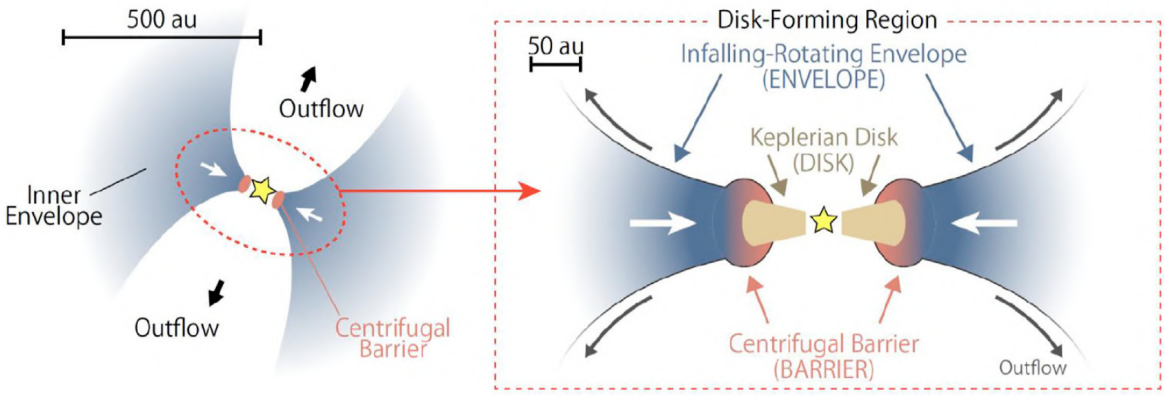

Figure 2.9: Schematic structure of the surrounding environment of solar-like Class 0/I protostars. Highlighted in different colored labels, the three main zones, the infalling and rotating envelope, the centrifugal barrier, and the rotating and accreting disk aimed to be disentangled in the FAUST LP.*Figure adapted from Codella et al. (2021).*

In fact, FAUST observations target four groups of chemical species, able to: *(i)* disentangle the components of the 50-2000 AU envelope/disk system; *(ii)* characterize the organic complexity in each system, as it might be inherited at later stages by the nascent planetary system; *(iii)* probe their ionization structure, as the degree of ionization is a crucial parameter for any planet formation theory and is poorly constrained; *(iv)* measure their molecular deuteration, which is a precious tool to understand the physical conditions when these species are/were deuterated, and also to link the interstellar medium to the Solar System history.

In this regard, the sources are selected to represent different chemical classes, from WCCC only to hot corino only sources. They have also been selected in distance and luminosity so as to obtain a uniform linear resolution of 50 AU (i.e. 250-350 mas in angular resolution), to disentangle the different system zones, and almost uniform brightness sensitivity.

The list of selected targets, including my target, Class 0 IRAS 4A2, their physical properties, the observational setups, and the target groups of chemical species are summarized in Tab. 2.1.

Table 2.1: (Top) Target sources properties and requested angular resolution. (Bottom) List of the frequency windows and of the species covered by the selected ALMA bands, divided in four groups (see text) (Codella et al. 2021).

Source	Region	Class	L_{bol} [$L_\odot$]	Distance [pc]	Resolution ["]
L1527	Tau	0/I	1.9	137	0.36
IRAS 15398-3359	Lup	0	0.92	155	0.32
CB68	Isolated	0	0.86	137	0.36
L1551 IRS5	Tau	I	24.5	147	0.34
L483	Aquila	0	13	200	0.25
Elias29	Oph	I	13	137	0.36
VLA 1623A	Oph	0	2.6	137	0.36
GSS30 (LFAM1)	Oph	I	8.7	137	0.36
RCrA IRS7B	Cr A	0	1.4	135	0.37
NGC 1333 IRAS4C	Per	0	0.49	235	0.21
BHB07-11 (B59)	Pipe	0/I	2.2	145	0.35
IRS63	Oph	I	1.6	147	0.34
NGC 1333 IRAS4A	Per	0	9.1	299	0.21
Frequency [GHz]	**ALMA Band**	**Zones**	**Molecular complexity**	**Ionization**	**Deuteration**
Setup 1: 214 - 219, 229 - 234	6	c-CH_3H_2, CH_3OH	CH_3CHO, NH_2CHO	DCO^+	N_2D^+, D_2CO
		SO, SiO	CH_3CHO, $HCOOCH_3$		
Setup 2: 242.5 - 247.5, 257.5 - 262.5	6	c-CH_3H_2, CS	CH_3OH, NH_2CHO	$H^{13}CO^+$	HDCO
		CH_3OH, SO	CH_3CHO, CH_3OCH_3		CH_2DOH
			$HCOOCH_3$		
Setup 3: 85 - 89, 97 - 101	3	CH_3OH, $HC_3N(^{13}C)$	CH_3OH, $13CH_3OH$	N_2H^+	c-C_3HD

2.4 Into The Forest Of ISM Molecular Spectral Lines

The core of this work is to analyze molecular transitions lines. In this section I will therefore review first of all what is a molecular transition line, and then how it relates to the physical properties of the gas where molecules can be found. This requires to crunch some of the theory of the radiative transfer and thermodynamics of gas systems (§ 2.4.1 and § 2.4.2), revise the quantum physics basics of how molecules can produce a transition line (§ 2.4.3), and illustrate one of the main methods based on the radiative transfer theory, that can be used to derive the physical properties of the gas emitting molecular lines (§ 2.4.4), and exploited in this thesis project.

2.4.1 The Radiative Transfer Equation

[17]When propagating through a medium, the radiation can lose or gain energy by emission or absorption by the medium, and its specific intensity I_ν (defined at (2.1)) is not constant. The *radiative transfer equation* precisely describes the change of I_ν if radiation is emitted ($\epsilon_\nu\ ds$) or absorbed ($-k_\nu\ I_\nu\ ds$):

$$\frac{dI_\nu(s)}{ds} = -k_\nu I_\nu(s) + \epsilon_\nu \tag{2.20}$$

through a slab of material of thickness ds. k_ν is defined as the *linear absorption coefficient*, in units of inverse of a length, while ϵ_ν is the *emissivity* (in units of energy per time per area).

There are several limiting cases for which the solution of this differential equation is simple.

First of all, introducing the concept of *optical depth*:

$$\tau_\nu(s) = \int_{s_0}^{s} k_\nu(s')\ ds' \tag{2.21}$$

as the degree to which light is absorbed as it passes through a medium of thickness $s - s_0$, the equation of transfer (2.20) can be written as:

$$\frac{dI_\nu(\tau_\nu)}{d\tau_\nu} = I_\nu - \frac{\epsilon_\nu}{k_\nu} \tag{2.22}$$

Defined $S_\nu = \epsilon_\nu / k_\nu$ as the *source function*, i.e. the rate of photons that manage to pass through the medium with respect to the rate of photons that get absorbed per unit distance, the general solution of the transfer equation becomes:

$$I_\nu(\tau_\nu(s)) = I_\nu(0)e^{-\tau_\nu(s)} + \int_0^{\tau_\nu(s)} S_\nu(\tau_\nu)\ e^{-\tau}\ d\tau \tag{2.23}$$

where $I_\nu(0)$ indicates the background radiation intensity ($\tau_\nu(s_0) = 0$).

The solution of the transfer equation can be read as follows: the intensity of the observed radiation, emerging from a given medium of thickness s, is given by the sum of the background radiation intensity, absorbed by the full length of the medium, and of all the contributions emitted by the medium layers, each one absorbed by the path corresponding to its depth in the medium.

If we make the simplifying assumption that S_ν is spatially constant (the medium affects the passing radiation in the same way everywhere), therefore, not depending on τ_ν, we can solve the integral in (2.23) obtaining:

$$I_\nu(\tau_\nu(s)) = I_\nu(0)e^{-\tau_\nu(s)} + S_\nu(1 - e^{-\tau_\nu(s)}) \tag{2.24}$$

Of this solution, we can identify two limiting cases:

[17]Main reference for this subsection: Wilson et al. (2013).

- The *optically thick case* ($\tau_\nu(s) \gg 1$), for which $I_\nu = S_\nu$, so no background is visible and the observed intensity is only the radiation that can come out of the outermost layer of the medium.
- The *optically thin case* ($\tau_\nu(s) \ll 1$), which means we can see the radiation coming from behind the medium, with the addition of a faint emission contribution. It is the case of the emission lines, produced by atomic/molecular species host by an intervening medium (between observer and background) like an interstellar cloud.

Another semplification to solve the radiative transfer equation is the condition of *Local Thermodynamic Equilibrium* (LTE). It means there is a complete equilibrium of the radiation with the medium, so that the brightness distribution B_ν is described by the *Planck function* (the spectral distribution of the radiation of a black body in thermodynamic equilibrium, i.e. power per unit frequency interval), dependent only on the thermodynamic temperature T of the gas. The source function will thus represent thermal radiation, because is governed by the gas kinetic temperature through the Planck function. The source function S_ν under LTE leads to the *Kirchhoff 's law*:

$$S_\nu = \frac{k_\nu}{\epsilon_\nu} = B_\nu(T) = \frac{2h\nu^3}{c^2} \frac{1}{e^{h\nu/kT} - 1} \tag{2.25}$$

Only in the optically thick case ($I_\nu = S_\nu$) thermal radiation approaches blackbody radiation. In fact, when a medium becomes optically thick, it means it is very efficient at absorbing and re-emitting radiation. In such a medium, radiation is more likely to be absorbed and then re-emitted by collisions between particles in the medium. In this way, radiation gets "thermalized", that is, it achieves a distribution in frequency/wavelength that resembles that of a blackbody at the same temperature.

2.4.2 Spectral Lines Basics

[18]All atoms and molecules have a series of discrete energy levels in which their electrons orbit. The spectral lines are to be interpreted as the emission, or absorption, of light in the form of photons, carrying a "packet" of energy $h\nu$, due to transitions of electrons from one energy level to another.

Einstein Coefficients

Each atom, or molecule, can absorb, jumping to a higher energy level, or emit light, dropping to a lower energy level, only at a series of characteristic wavelengths. Not all transitions are equally likely; each transition is in fact characterized by a precise wavelength (or frequency) and a probability of that transition occurring.

In particular, atoms and molecules can move from a level u, of higher energy, to a level l, of lower energy, or vice versa, in three fundamental ways:

- *Spontaneous emission.* Decay from an excited atomic state to one state of lower energy, with emission of a photon;
- *Stimulated emission.* Transition from an excited atomic state to a state of lower energy in which the species is excited ("stimulated") by the close passage of a photon (not interacting otherwise), with the emission of a photon with the same energy of the external one;

[18]Main references for this subsection: Wilson et al. (2013); Tennyson (2011).

- *Absorption.* Opposite process to photon emission: excitation to a state of higher energy due to the absorption of a photon.

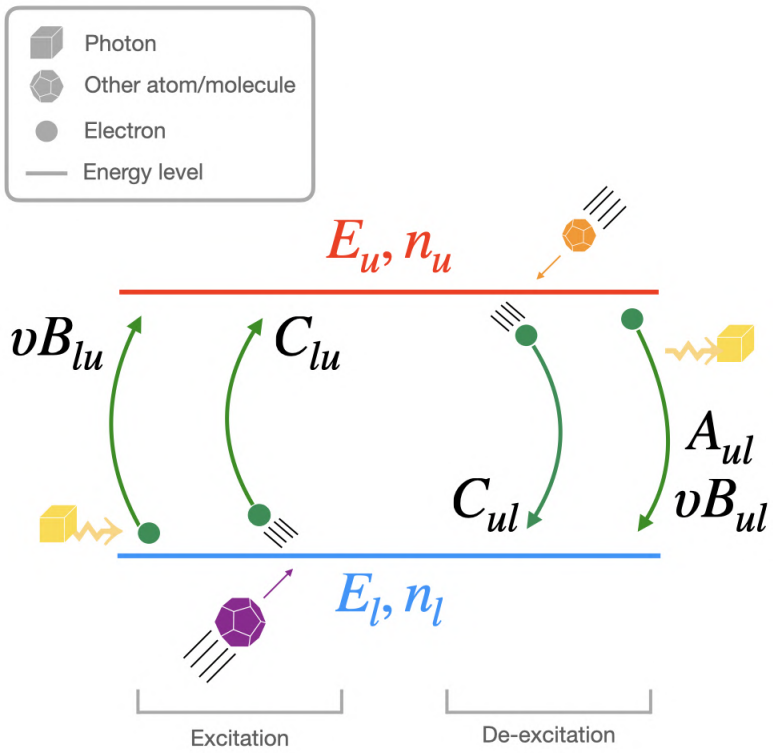

Figure 2.10: Schematics of a two-level system (u and l) of atoms in thermal equilibrium with radiation and intervening processes of excitation/de-excitation. The energy levels are indicated in blue and red solid lines with energies and number densities being respectively E_l, n_l and E_u, n_u. The absorption of a photon of energy $h\nu_{ul}$ *(yellow box)* or the collision with an energetic particle *(purple polygon)* can cause excitation. The collision with a less energetic particle *(orange polygon)*, the passage of an external photon or simply the presence of an excited atom in the upper energy level can cause de-excitation. The Einstein coefficients associated to each process are labelled in black.

In reality, the same excitation and de-excitation processes can be carried out by collisions with other atoms/molecules.

Albert Einstein was the first to understand that the Kirchhoff's law (2.25) had to imply a microscopic connection between emission and absorption of light interacting with the matter. Considering a cavity with atoms of a given species and radiation in thermal equilibrium and two non-degenerate atomic levels l and u with energies E_l and E_u and particle number densities n_l and $n_u (E_u > E_l)$, he derived the so-called *Einstein coefficients* to relate the stimulated emission with the spontaneous emission in presence of both radiative and collisional excitation/de-excitation (Einstein 1916); see Fig. 2.10 for a visual reference. These coefficients express the probability per unit time for which a transition can occur between the two levels of the cavity:

- *Radiative excitation.* The probability of spontaneous absorption of a photon at frequency ν_{ul} is related to the B_{lu} Einstein coefficient as $n_l B_{lu} \upsilon(\nu_{ul})$. ν_{ul} is the frequency associated to the energy difference of the two levels ($\frac{E_u - E_l}{h}$), while $\upsilon(\nu_{ul})$ is the average *energy density* of the radiation field[19]

- *Collisional excitation.* The probability of excitation via collision with more energetic particles is determined by the C_{lu} Einstein coefficient as $n_l C_{lu}$.

[19] Formally defined as: $\upsilon = \frac{1}{c} \int_{4\pi} I_\nu \, d\Omega$.

- *Radiative de-excitation.* In case of emission of a photon at frequency ν_{ul} either by spontaneous or stimulated emission, the probability of these two processes is related to $n_u A_{ul}$ and $n_u B_{ul} \upsilon(\nu_{ul})$ respectively, with A_{ul} and B_{ul} Einstein coefficients.
- *Collisional de-excitation.* For collisions with other, less energetic molecules/atoms, the probability of collisional de-excitation is related to the C_{ul} Einstein coefficient as $n_u C_{ul}$.

All Einstein coefficients are obtained by opportunely equating the variation with time of the number densities of the upper and lower energy levels (dn_u/dt and dn_l/dt) for collisional and radiative excitation/de-excitation.

The collision rates C_{ul} and C_{lu} also depend on the number density of the colliding partner n_c as $C_{ul} = \sigma_{ul}\ n_c$, where σ_{ul} indicates the velocity integrated collision cross section (in cm^{-2}), known from calculations or lab experiments. In principle, any collision with interstellar medium particles contributes to the collision coefficients, however, as a matter of fact the main contributors are free electrons, H_2, and He atoms.

Statistical Balance

If the cavity containing atoms and radiation is in a stationary state, the number of absorbed and emitted photons must be equal, hence their respective Einstein coefficients, which traduces in the *statistical equilibrium equation*:

$$n_u B_{ul} \upsilon(\nu_{ul}) + C_{ul} + A_{ul} = n_l B_{lu} \upsilon(\nu_{ul}) + n_l C_{lu} \tag{2.26}$$

This equation can also be easily generalized for a N multi-level system summing over all different pairs of lower and upper energy states ($\sum_{l \neq u}$).

Similarly to the radiative transfer equation, the generalized statistical equilibrium equation is not straightforward to solve.

Going back to the simpler system of two energy levels only, and supposing that collisions represent the dominant process both in excitation and de-excitation of the levels, (2.26) reduces to:

$$n_u C_{ul} = n_l C_{lu} \tag{2.27}$$

Since atoms/molecules reach thermal equilibrium with the environment by collisions (LTE), the ratio between the populations of the two energy levels of the system follows the *Boltzmann distribution*:

$$\frac{n_u}{n_l} = \frac{g_u}{g_l} e^{-h\nu_{ul}/kT} \tag{2.28}$$

assigned a statistical weight g_u and g_l to the u and l level in order, which take into account their degeneracies, and being k the Boltzmann constant, T the thermodynamic equilibrium temperature.

Equating the ratio C_{lu}/C_{ul} to (2.28), we obtain the *detailed balance* of the collision Einstein coefficients.

However, if collisional processes dominate onto radiative ones (negligible radiation field), but the collisions between atoms/molecules are not frequent enough to reach LTE conditions, the energy levels' populations will depend explicitly on the number of collisions and on the collision coefficients ($C_{ul} = \sigma_{ul}\ n_c$). Thus, simplifying (2.26) and exploiting the equation of detailed balance, (2.28) becomes:

$$\frac{n_u}{n_l} = \frac{g_u}{g_l} e^{-h\nu_{ul}/kT} \left(1 + \frac{n^*}{n_c}\right)^{-1} \tag{2.29}$$

in which a new quantity is introduces, the so-called *critical density* n^*:

$$n^* = \frac{A_{ul}}{\gamma_{ul}} \tag{2.30}$$

whose expression can be generalized as well for a multi-level system, by summing over the number of levels N ($\sum_1^N$, identified "1" as the fundamental energy level).

At this density, half of the atoms/molecules which get excited to level u are also collisionally de-excited back to level l. Very importantly, at densities higher than n^*, the emission of a specific line will be efficiently overcome by collisional de-excitations. In such case, collisions govern the energy level populations, hence making them distributed like the Boltzmann distribution, and LTE will hold.

On the other hand, if the gas density is smaller than the critical density ($n \ll n^*$), the photon emission will dominate the de-excitation from high energy levels for that particular line. In this case, the n_u/n_l ratio depends on the gas temperature and on the density of the colliders, the LTE condition is not valid, and the energy levels are *sub-thermally populated*.

The Boltzmann Distribution

[a] In statistics, the Boltzmann distribution defines the probability distribution that a system, like a gas particle system, will be in a certain state as a function of that state's energy and the temperature of the system; it is expressed in the form:
$p_i = \frac{1}{Q(T)} e^{-E_i/kT}$. $Q(T)$ is the normalization denominator, called *partition function*, and results from the sum of all possible j-th states in which the system can be (N in total), that is:
$Q(T) = \sum_j^N e^{-E_i/kT}$; p_i is the probability of the system being in state i, E_i the energy of that state; k the Boltzmann constant, and T the temperature of the system. This distribution shows that states with lower energy will always have a higher probability of being occupied. It can also yield the ratio of probabilities of two states i and j being occupied, which will depend on the states' energy difference only (as, in our case, it is shown at (2.28)).

[a] Reference: Landau & Lifshitz (1969).

The Einstein coefficients are not independent one from the other. It is easy to find the relations linking them by combining the Boltzmann distribution (2.28) with the statistical balance (2.26) in presence, this time, of a non-negligible radiation field (e.g. a low density gas strongly immersed in a strong radiation field). In this way, we discover that the average energy density of the radiation field $\upsilon(\nu_{ul})$ has the same functional shape of the energy density of a black body at the same temperature of the gas system:

$$\upsilon(\nu) = \frac{8\pi h\nu^3}{c^3} \frac{1}{e^{h\nu/kT} - 1}$$

In the end, the coefficient for radiative excitation is correlated to that of radiative "stimulated" de-excitation (stimulated emission) through:

$$g_l B_{lu} = g_u B_{ul} \tag{2.31}$$

and another correlation exists between the coefficient of radiative de-excitation for stimulated emission and the one for spontaneous emission:

$$B_{ul} = \frac{c^2}{2h\nu_{ul}^3} A_{ul} \tag{2.32}$$

It is worth to notice that this latter relation inherits the property of the Boltzmann distribution of being independent of the temperature of the gas, and only depends on the energy difference of the

two levels u and l. Moreover, it tells us that spontaneous emission becomes more probable than stimulated emission for increasing ν_{ul}. In fact, the maser emission, or *microwave amplification by stimulated emission of radiation*, is literally dominated by stimulated emission, and naturally occurs in the microwave range, where frequencies are lower.

For completeness, we will mention the following additional way to use the Einstein coefficients. When we considered the radiative transfer in § 2.4.1, the material properties were expressed in terms of emissivity ϵ_ν and absorption coefficient k_ν, which are both macroscopic parameters; they must be related to microscopic properties of the matter in the cavity. If line radiation is considered, the Einstein coefficients can be directly linked to the properties of the transition responsible for the spectral line.

Different Temperatures, Same Gas

In the interstellar medium atomic and molecular energy levels are generally not in LTE condition (the mean particle density is about 1 atom per cm^3!). If it was the case, the so-called *excitation temperature* T_{ex} would indicate the kinetic temperature of the gas with thermally populated levels. But if the levels are thermally populated, we know that they follow the Boltzmann distribution (2.28), hence, inverting this equation and inserting the definition of critical density n^*, we can write down T_{ex} as:

$$T_{ex} = T \left(\frac{kT}{h\nu_{ul}} \ln \left(1 + \frac{n^*}{n_c} \right) + 1 \right)^{-1} \tag{2.33}$$

which means, for densities larger than the critical density (i.e., LTE conditions), T_{ex} is exactly the kinetic temperature of the gas T. On the contrary, for densities lower than the critical density, not in LTE, $T_{ex} < T$. As a matter of fact, the definition of T_{ex} only holds for a two-level system. In a multi-level system, each transition has its own excitation temperature. However, if all levels are in LTE, the excitation temperature is the same for all transitions and is called *rotational temperature* T_{rot}. The LTE condition is therefore really precious, because in this case the excitation temperature (T_{ex}) and the rotational temperature (T_{rot}) are the same and equal to the gas kinetic temperature ($T_{ex} = T_{rot} = T$).

2.4.3 Molecular Excitation

[20] Compared to atoms, molecules have a complicated structure. For a start, molecules are formed at least by two atoms (*diatomic*), but in the interstellar medium have been found up to now at least 200 molecules formed from more than two atoms (*polyatomic*), including iCOMs (McGuire 2018); CDMS[21]). Since molecular binding energies are generally lower than ionisation energies of atoms (necessary to extract a bond electron), molecules are only observed in cooler environments than atoms.

The structure, and hence the spectra, of molecules are more complicated than atoms in two ways: *(i)* there is no single charge centre about which electrons move, hence lower symmetry, making hard to make computations of electronic wavefunctions (describing, in quantum physics, the physical state of electrons); *(ii)* the nuclei of protons and neutrons themselves move, giving rise to both rotational and vibrational motions of the atoms within the molecule. These precisely motions give rise to discrete molecular spectra.

[20] Main references for this subsection: Gordy & Cook (1970); Tennyson (2011); Condon & Ransom (2016); Griffiths & Schroeter (2018).

[21] https://cdms.astro.uni-koeln.de/classic/molecules

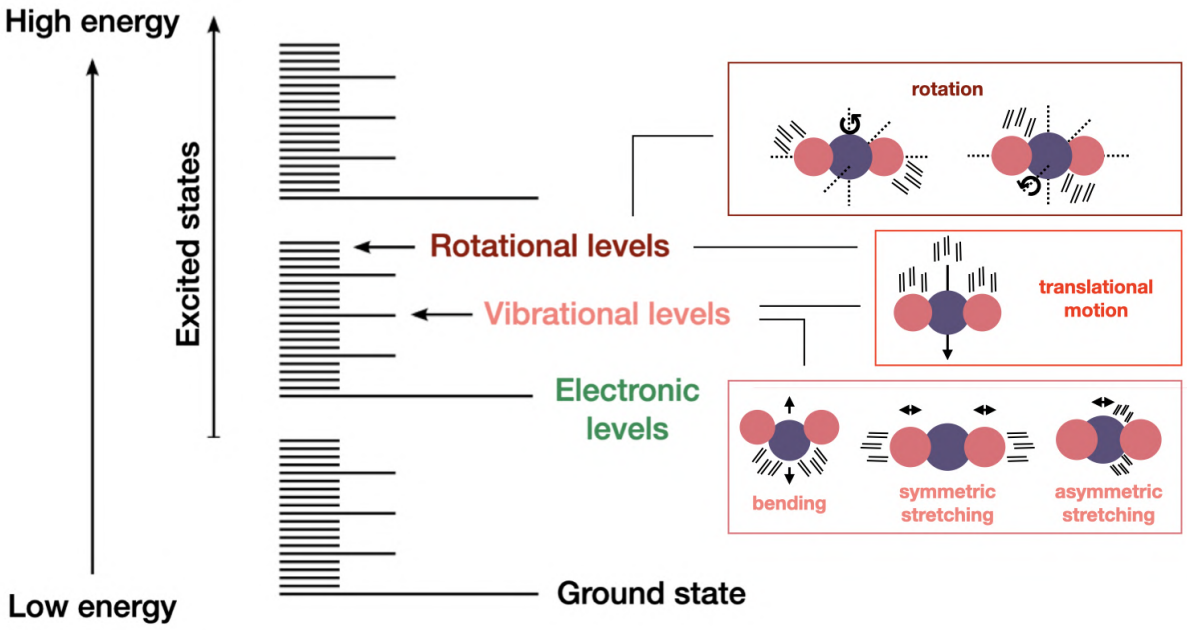

Figure 2.11: Schematics of the ladder of molecular energy levels, with examples of translational, vibrational, and rotational motions for carbon dioxide (CO_2).

Fortunately, to a very good approximation, known as the *Born–Oppenheimer approximation*, the motions of the electrons and the nuclei can be separated, because the electrons move much faster than the nuclei, relaxing immediately to any change in the positions of the nuclei.

Thanks to this approximation, we can classify the transitions in a molecule according to the different energies at play (see Fig. 2.11 for a visual recap):

- *Electronic transitions*, with typical energies of a few eV, associated to spectral lines in the optical - UV ranges;
- *Vibrational transitions*, caused by oscillations of the nuclei with respect to their equilibrium positions. The typical energies involved are 0.1 – 0.01 eV, corresponding to spectral lines in the IR;
- *Rotational transitions*, generated by the rotation of the nuclei with typical energies of $\sim 10^{-3}$ eV, thus corresponding to cm-mm wavelength spectral lines.

In general, molecular spectral lines arise from a transition between two states described by different electronic, vibrational and rotational quantum numbers (which in general describe the properties and the electronic structure of the molecule).

If a transition involves different electronic states, the corresponding spectral line will lie in the optical-UV range.

However, to a less good approximation one can consider separately the vibrational and rotational motions of nuclei. In a molecular spectra, a particular transition can involve changes in more than one type of motion. Yet, if we restrict to the centimeter/millimeter/sub-mm wavelength ranges, only transitions between different rotational levels and perhaps different vibrational levels will be involved, resulting in a simpler description of the molecular energy levels.

Schrödinger Equation Formalism For Diatomic Molecules

I will now present the mathematical formulation of spectra for diatomic molecules, because they illustrate the key features of molecular spectra.

The Schrödinger equation for a diatomic molecule with nucleus A, of mass M_A and nuclear charge Z_A, and nucleus B, of mass M_B and nuclear charge Z_B, and N electrons can be written as:

$$\left(-\frac{\hbar^2}{2M_\mathrm{A}}\nabla_\mathrm{A}^2-\frac{\hbar^2}{2M_\mathrm{B}}\nabla_\mathrm{B}^2-\frac{\hbar^2}{2m_e}\sum_{i=1}^{N}\nabla_i^2+V_e-E\right)\Psi\left(R_\mathrm{A},R_\mathrm{B},r_i\right)=0 \tag{2.34}$$

where the first two terms indicate respectively the kinetic energy of nuclei A and B, the third term gives the kinetic energy of the electrons (in all three terms, ∇^2 is the Laplacian operator), V_e is the potential and E the total energy of the system. The potential is given by the various Coulomb interactions (inverse-square attraction of oppositely charged particles) within the molecule. The function multiplying the pharenthesis on the left hand side of this equation is the wavefunction of the system, $\Psi\left(R_\mathrm{A},R_\mathrm{B},r_i\right)$, in which R_A identifies the vector connecting the molecular centre-of-gravity with nucleus A, same for R_B but with nucleus B; r_i is the i-th interdistance between nuclei and electrons. Within the Born–Oppenheimer approximation, the system wavefunction can be explicated as the product of electronic, Ψ_e, and nuclear, Ψ_n, wavefunctions:

$$\Psi\left(R_\mathrm{A},R_\mathrm{B},r_i\right)=\Psi_e(r_i)\Psi_n(R_A,R_B) \tag{2.35}$$

The electronic and the nuclear wavefunctions satisfy separate Schrödinger equations. Solving the one for the electronic wavefunction, we obtain the E_e eigenvalue, the electronic energy at each internuclear separation R, and the electronic potential $V(R)$ governing the motion of the nuclei. The Schrödinger equation satisfied by the nuclear wavefunction is:

$$\left(-\frac{\hbar^2}{2M_\mathrm{A}}\nabla_\mathrm{A}^2-\frac{\hbar^2}{2M_\mathrm{B}}\nabla_\mathrm{B}^2+V(R)-E\right)\psi(R_\mathrm{A},R_\mathrm{B})=0 \tag{2.36}$$

In this form, (2.36) encloses all the motions of the nuclei: vibrations, rotations, and the translation of the molecule through space. Yet, this equation has the same functional shape of a two-body equation, hence, it can be separated into an equation for the translational motion of the centre-of-gravity of the system, and an equation for the internal roto-vibrational motion of the molecule.

Neglecting the translational motion (which gives a continuum and including it in the total energy would hide the discrete spectrum of roto-vibrational energy levels), the Schödinger equation for nuclear motion becomes:

$$\left(-\frac{\hbar^2}{2\mu}\nabla^2+V(R)-E\right)\psi_n(R)=0 \tag{2.37}$$

where $\hbar$ is the Planck's constant divided by 2π, and $\mu=M_AM_B/(M_A+M_B)$ is the *reduced mass* of the two-body problem in exam. Also, R is now a 3D vector that gives the internuclear distance R and the orientation (θ,ϕ) of the molecule's axis relative to the laboratory z-axis.

As we have already anticipated, vibrational and rotational motions cannot be separated rigorously. Yet, as a good first approximation, we can split them as:

$$\psi_n(R)=\psi_v(R)\psi_r(\theta,\phi) \tag{2.38}$$

where $\psi_v(R)$ describes the vibrational motion, while $\psi_r(\theta,\phi)$ the rotational motion.

Considering first the rotational motion, solving (2.37) for $\psi_r(\theta,\phi)$, we can finally obtain one of the first main results concerning molecular spectra:

$$E_r=\frac{\hbar^2}{2\mu R^2}J(J+1) \tag{2.39}$$

This is the *rotational energy*, and can assume discrete values according to the *rotational quantum number* J ($J=0,1,2...$). This quantum number takes into account the rotation of the diatomic

molecule perpendicularly to the bond axis (motion quantized in turn by the N number), and the angular momentum of the electrons (quantized with L). Projecting J onto the molecular axis, we get a total of $2J+1$ degenerate values for the rotational state.

Switching now to the radial equation resulting from solving (2.37) for $\psi_v(R)$, it actually has no general algebraic solution, because the potential $V(R)$ is not a simple function of R. However, for low values of *vibrational energy* E_v, that summed with E_r gives the total energy of the molecule E, the potential can be approximated with a parabola, and the radial equation becomes the quantum mechanical equation of the harmonic oscillator. The solutions of this well-known equation, its eigenvalues, are the discrete *vibrational energies*:

$$E_v = \hbar\omega\left(\nu + \frac{1}{2}\right) \tag{2.40}$$

with ω as angular frequency. The *vibrational quantum number* takes, similarly to J, integer values $v = 0, 1, 2....$ It means that the vibrational levels supported by an harmonic potential are evenly spaced.

In reality, molecules are not harmonic oscillators. Their potential is elastic, but the repulsion at short binding distances of molecules is stronger than the attraction at long binding distances, and so it is called *anharmonic potential*. Unlike the case of harmonic potential, anharmonic energy levels get closer to each other with increasing vibrational quantum number. Moreover, while the harmonic potential supports an infinite number of energies, the discrete levels actually stop due to *dissociation*, when molecules split in atoms/ions/radicals. Also, a more complex expression describes the energy of roto-vibrational levels.

It is interesting to notice that, since molecules have "packed" energy levels, they act very well as thermometers of the surrounding environment, yet depending on the molecule and its transitions. This is particularly true for rotational levels. For a diatomic rigid rotator (simplest assumption for a rotating molecule) with the so-called *rotational constant* B, the rotational level with the largest population is given, as a function of the quantum number J, by $J \approx= \sqrt{kT/2B} - 1/2$. Doing the maths, at the cold interstellar medium temperatures of $10 - 20$ K, most molecules have already several of their levels occupied. Vibrational levels, instead, are more widely separated than rotational ones, hence they are only sensitive to higher temperatures.

Diatomic Transitions Selection Rules

Selection rules for molecular transitions determine whether a particular transition between two energy levels is *allowed* (bright and intense molecular line, very likely observable) or *forbidden* (weak molecular line, occurring only in very peculiar conditions, unlikely to observe) based on conservation laws and the characteristics of the molecular wavefunctions. They are derived applying the principles of quantum mechanics and of symmetries on molecular systems.

A diatomic molecule can undergo a pure rotation transition (involving rotational states only) if it possesses a *permanent dipole moment* μ (usually determined with laboratory experiments), which means to possess an asymmetric charge distribution. This is true for *heteronuclear* diatomic molecules (NaCl sodium chloride, just to give an example), but not for *homonuclears*, like H_2. These molecules do not have a permanent dipole moment, so they cannot undergo rotational dipole transitions, but can be observed through weak quadrupole transitions and vibrational transitions. Also, the strength of a rotational transition depends on μ^2, therefore molecules with large permanent dipoles have intense transitions. Dipole radiative transitions occur with a change in the rotational quantum number J of one unit: $\Delta J \pm 1$, and the parity of the initial and final states must be opposite for them to occur.

The potential outside a molecule originated by charge distribution of electrons in its orbitals can be mathematically described by a multipole expansion, and the same applies for the external

vector potential generated by the internal molecular magnetic field. In particular, the electric quadrupole (consisting of two electric dipoles of equal and opposite charges, with null net charge and dipole moment) or the magnetic dipole ($\sim$ analogue of the electric dipole) of nuclei can interact with electrons or with other nuclei, giving rise to the so-called *hyperfine structure*. In fact, under the occurrence of these electric and magnetic nuclei-electrons interactions, small shifts in otherwise degenerate energy levels appear, splitting those energy levels and making possible to observe "hyperfine" transitions.

In the case of hyperfine structure, the *total angular momentum quantum number* $F = J + I$ must be conserved. In this definition, I is the *nuclear spin quantum number*, describing the intrinsic angular momentum (*spin*) of nuclei. To have *allowed* transitions, the quantic selection rule to be respected is: $\Delta F = \pm 1, 0$ but not $\Delta F = 0 \longleftrightarrow 0$.

As we can imagine, there exist selection rules also for vibrational transitions (and, by the way, for electronic transitions).

Within the harmonic oscillator approximation, the fundamental electric dipole selection rule not to violate is: $\Delta v = \pm 1$. More realistically (anharmonic potential), any change in Δv is allowed, but in practice, $\Delta v = \pm 1$ leads to the strongest transitions, because the intensity of individual transitions drops rapidly with increasing Δv.

Vibrational transitions are usually accompanied by a change in rotational state, i.e., rotational transitions. Indeed, for a certain molecular vibrational state, there are many rotational states compacted inside.

A diatomic molecule can vibrate only in one direction, so it has only one vibrational mode ($v = 0$). Clearly, the situation gets more complex for polyatomic molecules, for which a plethora of different vibrational modes may exist, each of them with its own ladder of vibrational states, maybe degenerate. For polyatomic molecules with a permanent electric dipole moment, transitions with $\Delta J = 0, \pm 1$ can occur.

From Diatomic To Polyatomic Molecules

What happens if we try to scale to polyatomic molecules? As first general comment, the more complex the molecule is, more transitions it has, which are also weaker when observed, because the total energy needs to be redistributed among all its transitions. Besides, their potential functions are much more complicated than those of diatomic molecules, hence, it is generally not feasible to solve their wave equations using the same methods applied to diatomic molecules. Instead, it is custom to find a first solution for the polyatomic molecule with the assumption that it is completely rigid, and then use the perturbation theory, which basically means to add a weak disturbance to a simpler system of known exact solution. In this way, centrifugal distortion, i.e. the fact a molecule acts as if it has a spring for a chemical bond, and hyperfine electric and magnetic interactions can be taken into account.

Polyatomic molecules are conveniently classified and studied on the basis of their moment of inertia, i.e. their rotational motion around x-, y-, and z- axis in 3D space.

In particular, molecules in which two of the principal moments of inertia are equal are called *symmetric-top molecules* (e.g. ammonia NH_3). In this sense, a *linear molecule* (of two or more atoms) can be treated as a special case of a symmetric-top in which the angular momentum about the symmetry axis is zero.

When none of the three principal moments of inertia of a molecule is zero, we talk about *asymemetric-top molecules* (e.g. water H_2O and acetaldehyde CH_3CHO). In this case, considerable complexity is encountered in their pure rotational spectrum. Differently from linear and symmetric-top molecules, the rotational frequencies can no longer be expressed in convenient equations.

At the end of the story, the nuclear framework is not rigid: bond distances and angles will

vary because of the centrifugal force produced by rotation. The rotating molecule will no longer be in its equilibrium configuration but in a distorted configuration, raising the name of molecules as *distortable rotators*. As a consequence of this distortion, the moments of inertia can no longer be considered constant and independent of the rotational state, with tangible effects in the rotational spectra. The theory of centrifugal distortion is considerably more complex for asymmetric molecules, and the effects are in general larger for them than for linear or symmetric-top molecules. Fortunately, although we must consider the influence of centrifugal distortion in order to accurately account for the positions of rotational transitions, its effects still represents only a small fraction of the rotational energy, which is mostly accounted by the rigid rotor term. This is way, in many cases, it can be treated as a small perturbation of the rigid rotor.

Quantum Numbers Labels In Astronomical Databases

Resigned ourselves to the complexity of molecules, in theory and in practice, we shall highlight for the aims of this thesis project that the quantum numbers associated to any molecular transition (whose combination allows to detect it) are usually found in the following order[22]: J (or N); K_a and K_c (or $\pm K$); v; F_1, F_2...F. This sequence is shown first for the upper state, and then for the lower state involved in the transition.

N is the total rotational quantum number excluding electrons and nuclei spins. For singlet molecules (all electrons are paired, a single spectral line arises), J is equal to N.

K_a (indicating molecule's prolate symmetry) and K_c (indicating molecule's oblate symmetry) are the projections of J/N onto two of the momentum inertial axes. For symmetric top molecules, only $\pm K$ is given ($\pm$ designating parity with respect to inversion through the molecular centre of mass of all particles' coordinates in the laboratory coordinate system) For linear molecules, instead of K, L or I may be used. v specifies different vibrational/electronic states. Finally, F_1, F_2... F quantum numbers describe the the total angular momentum in the presence of an external magnetic field. In relation to the other quantum numbers, $F = J + N$.

To be precise, only this final quantum number (F), which includes all effects of rotation, electronic and nuclear spins, uniquely labels a molecular transition. In fact, it is the selection rule $\Delta F = 0, \pm 1$ that holds strictly for any molecular transition. Effects mediated by vibration, electronic and nuclear spin, can mix J/N, K, v, or F_i quantum numbers, making them ambiguous labels for molecular transitions.

2.4.4 A Way To Solve The Radiative Transfer Equation: Rotational Diagram Method

[23]The radiative transfer equation (2.20), despite its apparent simplicity, is not easy to solve at all, due to the coupling between each energy level population and the intensity of radiation in the medium, unless we make some reasonable assumptions. For the sake of the methodology I adopted in this work, I will present the *Rotational Diagram method* (RD method), the most basic molecular line spectra analysis method, which consists in solving the radiative transfer equation assuming LTE condition and optically thin spectral lines (§ 2.4.1).

In this sense, we can simplify the general solution of the radiative transfer equation given, apart from the LTE condition validity, a constant source function S_ν, and the absence of a background

[22]https://cdms.astro.uni-koeln.de/classic/

[23]Main references for this subsection: Blake et al. (1987); Turner (1991); Taylor (1996).

source (with intensity $I_\nu(0)$) as:

$$I_\nu\left(\tau_\nu(s)\right) = B_\nu(T)\left(1 - e^{-\tau_\nu(s)}\right) \tag{2.41}$$

Since, in general, the intensity of an emission line is defined as the observed intensity subtracted by the intensity of the background, the optical depth in (2.41) assumes the significance of *line opacity*.

How does relate this theoretical solution to observations? We have seen in § 2.2.4 that what we practically measure is the brightness temperature T_B, in this context, the individual molecular lines T_B. We therefore need to convert them into column densities, but this requires information about the line optical depths and excitation conditions. In order to retrieve these information, we have to observe many transitions of an individual molecule, so as to estimate any deviation from thermal excitation along with the line opacities, which works well if the lines are not optically thick.

Considering the typical profile of a spectral line, a Gaussian with a FWHM Δv (line width, in velocity units), and the Boltzmann distribution (2.28) (which requires LTE to hold), the optical depth can be written as:

$$\tau_\nu = \frac{c^2}{8\pi\nu^2\Delta\nu}\int n_u A_{ul}\left(e^{h\nu/kT} - 1\right)\, ds \tag{2.42}$$

Using now the Rayleigh-Jeans approximation for the black body emission, and the definition of brightness temperature as a function of intensity (which replaces the flux density in (2.17)), T_B can be expanded as:

$$T_B = \frac{hc^3}{8\pi k\nu^2\Delta\nu}A_{ul}\int n_u ds\left(\frac{1 - e^{-\tau}}{\tau}\right) \tag{2.43}$$

From this equation, it is possible to define a crucial property of spectral lines, the *velocity-integrated intensity*:

$$W = \int T_B\, dv \approx T_B\, \Delta v \tag{2.44}$$

At this point, by inverting (2.43), we can obtain an explicit expression of the column density of the upper state, which is formally defined as the integration of the upper level number density n_u along the line of sight s ($N_u = \int n_u\, ds$), as a function of the observed brightness temperature:

$$\frac{N_u}{g_u} = \frac{3kc\int T_B\, dv}{8\pi^3 S\mu^2\nu^2} \tag{2.45}$$

where k and c are in order the Boltzmann constant and the speed of light, ν the frequency of the line, and $S\mu^2$ the product between the *line strength* (S) and the squared *transition electric dipole moment* (μ^2), both referred to a particular spectral line. The line strength is widely used in theoretical work to quantify line intensities (in SI units, $m^2\ C^2$), and is listed in comprehensive numerical tables derived from both experimental and theoretical methods. The transition electric dipole has to do with the magnitude of the interaction with the electric field that causes a transition between two states, and is measured in Debye ($1\ D = 3.336 \times 10^{-30}$ C m in SI; McNaught & Wilkinson (1997); Drake (2006)). (2.45) can be re-written, isolating N_u, as:

$$N_u = \frac{8\pi k\nu^2}{hc^3}\frac{W}{A_{ul}} \tag{2.46}$$

If we did not assume lines to be optically thin ($\tau_\nu \ll 1$), we could not integrate the radiative transfer equation to obtain this result.

We now have an equation for the number of molecules in the upper energy state u of the two-level system, but it can make sense only if related to the total number of molecules per unit area

along the line of sight, N_{tot}. Summing over all possible upper energy levels to which a molecule can transit to produce a line, the Boltzmann distribution (2.28) becomes:

$$\frac{n_u}{g_u} = \frac{n_{tot}}{Q(T)} e^{E_{up}/kT} \tag{2.47}$$

where $n_{tot} = \sum n_u$, is the total number density of u states, $Q(T)$ is the partition function at a given temperature, and E_{up} ($\equiv E_u$) the *upper state energy*, such that the photon emitted with a transition from level l to level u ($u > l$) has energy $h\nu_{ul} = E_{up} - E_l$.

The partition function can be calculated if we know the energies of the states accessible to the system of interest; its values can be found in different astronomical spectroscopy database (as done in this project using CDMS and JPL databases; see § 3.2.2).

Making the final assumption of dealing with an isothermal cloud (i.e. the population distribution is thermalized at a single rotational temperature T_{rot}), and integrating n_u and n_tot along the line of sight to obtain respectively N_u ($\int n_u \, ds$) and N_{tot} ($\int n_{tot} \, ds$), (2.39) gives (in logarithm scale):

$$ln\left(\frac{N_u}{g_u}\right) = ln\left(\frac{N_{tot}}{Q(T)}\right) - \frac{E_{up}}{kT} \tag{2.48}$$

where the two unknowns are N_{tot}, the total column density of the molecular species, and T_{rot}, the molecular rotational temperature, which equals the gas kinetic temperature only if all energy levels are under LTE.

Therefore, if we are able to observe several molecular line transitions, we can construct the *Rotational Diagrams* (§ 3.4.2 to see the diagrams at work). This name relates to the fact that we are measuring the energies of transitions between rotational and vibrational energy levels of polar molecules (having a net electric dipole) in the gas phase. These diagrams are also called *Boltzmann plots*, since, more trivially, the energy levels in LTE are populated according to the Boltzmann distribution.

In practice, the Rotational Diagram is a plot of the column density per statistical weight of a number of excited molecular energy levels, $y = ln(N_u/g_u)$ as a function of their upper level energies, E_{up} (often divided by kT to convert them from cm^{-1} to K).

Supposing to have drawn the Rotational Diagram for a set of transitions of a given molecular species, how can we extract its total column density and rotational temperature ($\approx$ gas kinetic temperature in LTE)? In fact, (2.48) can be seen as the equation of a straight line fitting the data points (i.e. the molecular transitions) of the form: $y = a + b \cdot x$, where $y = ln\left(\frac{N_u}{g_u}\right)$, $x = E_{up}/k$, and a and b respectively the intercept and the slope of the fitting line. In this way, the rotational temperature corresponds to the inverse of the slope:

$$T_{rot} = -1/b \tag{2.49}$$

while the total column density of the species can be evaluated from the intercept as:

$$ln\left(\frac{N_{tot}}{Q(T)}\right) = a$$

$$e^{ln\left(\frac{N_{tot}}{Q(T)}\right)} = e^a$$

$$\frac{N_{tot}}{Q(T)} = e^a$$

$$N_{tot} = e^a \cdot Q(T) \tag{2.50}$$

If the assumptions of LTE conditions and optically thin lines are fully satisfied, the data points should follow a perfect straight line, and T_{rot} would give a good estimate of the kinetic gas temperature T. Otherwise, a scattering along the line implies that non-LTE effects and/or optically thick lines are present. The Rotational Diagram method therefore provides a precious,

yet first-order approximation, of the value of the gas temperature probed by a molecule and the overall abundance of this molecule.

Another important comment to make is that the column density of the upper energy level N_u depends on a another observable (apart from T_B), that is the telescope beam θ_b (product of beam's major and minor axis). Indeed, N_u is averaged over the beam area of our observations. If the emitting source, of estimated angular size θ_s, is smaller than the telescope beam, we must correct N_u, and therefore N_{tot}, for the *beam dilution* of the source radiation, applying the so-called *filling factor*:

$$ff = \frac{\theta_s^2}{\theta_s^2 + \theta_b^2} \leq 1 \tag{2.51}$$

which divides N_u in (2.46) (Goldsmith & Langer 1999).

Moreover, if the spectral lines cover a large range of frequencies, the filling factor differs from line to line, and the rotational temperature will have to be corrected for beam dilution as well.

When analyzing a Rotational Diagram in practice, it is important to have in mind that the line intensity, i.e. how much bright a line appears in a spectrum, is tuned by two line parameters that we have introduced in this section, the upper state energy E_{up}, and $S\mu^2$ (2.45). E_{up}, the excitation energy of the upper level in a two-level system, is connected to the temperature, assuming LTE holds and therefore that the population in this level follows the Boltzmann statistics $\propto e^{E_{up}/kT}$. Consequently, we expect that lower excitation temperatures T_{ex} (2.33) are related to lower excitation energies, hence a brighter line, and vice versa. On the other hand, since the Boltzmann distribution favors populating lower energy levels, E_u must be high enough to have a good number of molecules in the excited state, which traduces in requiring the factor $e^{E_{up}/kT}$ not to be too small. Therefore, it is convenient to consider transitions with a maximum E_{up} of 300 K, because nearly all molecules will be in excited levels of smaller energy. With the same column density, a line with higher $S\mu^2$ will be brighter and more intense (much molecules contributing to it). In conclusion, a combination of low excitation energy E_{up} and high $S\mu^2$ parameter is a good and practical indicator of bright and intense spectral lines.

Uncertainties

The uncertainties on the Rotational Diagram and on the estimates of T_{rot} and N_{tot} we can extract from it need a bit of space.

First of all, the upper state energies E_{up}, the x-values of rotational diagrams, are assumed to be without error. The errors on the y-values, being $y = ln\left(\frac{N_u}{g_u}\right)$ and $W \pm dW$ ((2.44); K · km/s), informally called *line integrated area*, instead, need to be taken into account and can be calculated as a sum in quadrature:

$$dy = \frac{\sqrt{(dW)^2 + \sigma_{fluc}^2}}{W} \tag{2.52}$$

With the second term at the numerator (σ_{fluc}), we enclose continuum subtraction and calibration errors. It is set as an arbitrary fraction of the line integrated area.

Now, supposing that the $ln\left(\frac{N_u}{g_u}\right)$ values are linearly correlated to the upper state energies E_{up}, that is, assuming it subsists a *linear regression* of the form $y = a + b \cdot x$ between the two quantities, we can use the *least squares method* to find the coefficients a and b which minimize the discrepancy between this theoretical line and the set of observed data points. In particular, if we deal with different uncertainties (dy_i) for different y-values, we can adopt the *weighted least squares method*, in which the formulas of the χ^2, a and b get modified introducing a weighting

sum of the inverse of the N errors on the y-values: $w_i = 1/dy_i$. To be specific, we have:

$$\chi^2 = \sum_{i=1}^{N} \frac{w_i\ (y_i - a - bx_i)^2}{dy_i^2} \tag{2.53}$$

$$a = \frac{\sum wx^2 \sum wy - \sum wx \sum wxy}{\Delta} \tag{2.54}$$

$$b = \frac{\sum w \sum wxy - \sum wx \sum wy}{\Delta} \tag{2.55}$$

being x_i and y_i the x and y coordinates of data points, and $\Delta = (\sum_{i=1}^{N} w \cdot \sum_{i=1}^{N} wx^2) - \sum_{i=1}^{N} (wx)^2$. The extremes of the sum (i = 1, N) and the i-th subscript in (2.54) and (2.55) have been omitted for simplicity.

According to this weighted method, the uncertainties on a and b are respectively:

$$da = \sqrt{\frac{\sum_{i=1}^{N} wx^2}{\Delta}} \tag{2.56}$$

$$db = \sqrt{\frac{\sum_{i=1}^{N} w}{\Delta}} \tag{2.57}$$

Finally, knowing da and db, the uncertainty on T_{rot} can be evaluated as:

$$T_{rot} = -1/b \rightarrow \frac{d}{db}(T) = \frac{d}{db}\left(-\frac{1}{b}\right)$$

$$\frac{dT}{db} = \frac{1}{b^2}$$

$$\frac{d}{dT}(T) = \frac{d}{dT}\left(-\frac{1}{b}\right)$$

$$\frac{b^2}{db}(T) = \frac{d}{dT}\left(-\frac{1}{b}\right)$$

$$\frac{dT}{T} = -\frac{db}{b}$$

$$dT = -\frac{db}{b} \cdot T \tag{2.58}$$

and the one on N_{tot} as:

$$N_{tot} = e^a \cdot Q(T) \rightarrow \frac{dN}{da} = \frac{d}{da}\left(e^a \cdot Q(T)\right)$$

$$\frac{dN}{da} = Q(T) \cdot e^a \cdot 1$$

$$\frac{d}{dN}N = \frac{d}{dN}\left(e^a \cdot Q(T)\right)$$

$$\frac{d}{dN}N = \frac{1}{e^a Q(T) da}\left(e^a \cdot Q(T)\right)$$

$$\frac{dN}{N} = da$$

$$dN = da \cdot N \tag{2.59}$$

The Least Squares Method χ^2

This method is a statistical procedure to verify whether two variables are correlated by a linear functional relation. It is based on three assumptions: the errors on the measurements of the independent variable (x, dx) are negligible with respect to those on the dependent variable (y, dy); the errors on the dependent variable are equal or comparable; these errors are distributed on a Gaussian distribution around its mean value.
Given the probability of obtaining the i-th value of y, the χ^2 variable is defined as: $\chi^2 = \sum_{i=1}^{N} \frac{(y_i - a - bx_i)^2}{dy_i^2}$. Therefore, the values of a and b that best approximate the observed data are those that maximize the overall probability of obtaining y, which traduces in minimizing the sum of the squares χ^2. Thus, deriving the χ^2 wrt a and b and fixing to 0 these derivatives leads to a system of linear equations whose solutions are precisely the seeked a and b coefficients.
To calculate their errors, one first estimates the characteristic error on y (σ_y), and then propagates this error on the two coefficients. σ_y is usually computed using the sum of squared deviations, $\sigma_y = \sqrt{\frac{1}{N-2} \sum_i^N (y_i - a - bx_i)^2}$, where we subtract from N the two true (theoretical) values of a and b, which can only be guessed (with the least squared method).

CHAPTER 3

Results And Discussion

3.1 Data Deconvolution And Imaging

IRAS 4A2 has been observed with ALMA in the context of the LP FAUST in a setup of three different frequency ranges, called "setup 1" (217.8 - 233.8 GHz), "setup 2" (245.6 - 267.9 GHz), and "setup 3" (94.4 - 104.3 GHz). Of these, I had data available at 234 GHz ("setup 1") and 245 GHz ("setup 2"), being these frequencies the center rest ones of the continuum SPW in the two frequency arrangements. At 245 GHz, the setup comprehends both the 12m and the 7m ACA ALMA arrays, while at 234 GHz the observations lack the 7m ACA configuration. At both frequencies, the setup is arranged in the angular resolution configuration named "TM1", which allows to reach an angular resolution of 0.21 arcsec.

The basic information about the starting visibilities dataset can be recovered using the `listobs` `CASA` task. Tab. 3.1 below reports the main information about the two setups.

Table 3.1: Information on FAUST/ALMA observations performed towards IRAS 4A2. The total integration time is calculated for target and execution block. The notations "wb" and "nb" stand for "wide-band" and "narrow-band".

	234 GHz	**245 GHz**
observing period	15/10/2018 - 21/09/2019	21/09/2019 - 27/09/2019
total integration time	3h 8min	2h 6min
antennas	45 - 50	
SPWs (IDs)	25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49	
wide-band SPW	49	33
total bandwidth	1875 MHz (wb), 59 MHz (nb)	
channel bandwidth	0.977 MHz (wb), 0.122 MHz (nb)	
execution blocks	4	3
source list	IRAS4A (target), J0237+2848	
	J0336+3218, J0328+3139	
	J0510+1800, J0423-0120	

For the thesis goals, I was interested in imaging the wide-band ($\sim$ 1 GHz) spectral windows only, in order to identify as many spectral lines as possible of each iCOM (see § 3.2.2). Yet, since the observations are performed in chuncks, or execution blocks (EBs), at different times, the

measured velocities differ from EB to EB because they are acquired by ALMA in the TOPO rest-frame[1]. I therefore needed to convert topocentric velocities into the LSRK frame[2] and perform accordingly a frequency rebinning of the visibilities. For this reason I used the `mstransform`[3] `CASA` task to *(i)* split the full visibility dataset and create a new dataset containing target visibilities relative to the wide-band SPWs only, *(ii)* combine the wide-band SPW of each execution block, and *(iii)* rebin the splitted visibilities on a channel-by-channel basis in frequency. As a sanity check, I produced and looked at the dirty map of these new datacubes (no threshold parameters set in `tclean`) to verify that the rebinning worked correctly.

The datacubes I analyzed were already continuum-subtracted and self-calibrated. In FAUST, the continuum-subtraction is done by identifying all spectral line free channels, and averaging them to estimate the continuum baseline, which is self-calibrated and subtracted from the entire datacube. Therefore, I directly proceeded by cleaning and imaging the visibilities using the `CASA` `tclean` task and the `CLEAN` algorithm. I illustrated the basics of both of them in § 2.2.5, so here I will just present the tuning parameters I used and the results in terms of image, model image, and residuals at the two observed frequencies (reported in Tab. 3.2). In order, I chose the pixel size (`cell`) to be 1/4 of the beam: it is not the same at 234 GHz and 245 GHz because the setup at the two frequencies have slightly different angular resolution. The image size in pixels (`imsize`) has been chosen accordingly.

Table 3.2: `tclean` **parameters used for deconvolving and imaging of the two analyzed datacubes.**

	234 GHz	**245 GHz**
imsize	1200	1500
cell	0.053"	0.04"
niter	10 000	
threshold	1.0 mJy	1.3 mJy
weighting	robust 0.5	
deconvolver	multiscale [0, 5, 15, 30, 60]	
sidelobethreshold	3.0	1.5
noisethreshold	5.0	5.0
lownoisethreshold	1.5	1.25
minbeamfrac	0.3	0.01

The reported threshold parameters come from several cleaning attempts; the values for the `threshold` parameter, in particular, result to be slightly higher than the expected theoretical noise of 0.9196 mJy beam^{-1} (at 234 GHz), and 1.1 mJy beam^{-1} (at 245 GHz). I opted for a `briggs` weighting scheme with an intermediate value of the `robust` parameter of 0.5 since ALMA/FAUST observations are not particularly unbalanced towards bad PSF profile or very low sensitivity. I also opted for a `multiscale deconvolver` algorithm, because the images of the source show complicated spatial structure, given an interplay between compact emission towards IRAS 4A2, strong absorption towards IRAS 4A1, and large scale emission in certain ranges of channels. As regarding the usage of masks, I decided to perform an automatic masking with the `AUTO-MULTITHRESH` algorithm described in § 2.2.5, to perform an overall efficient and quick

[1]The tropocentric velocities are the ones observed from the ground, with no correction applied for the Earth's rotation, revolution, nor for the peculiar motion of the Sun.

[2]Kinematic Local Standard of Rest frame, having a velocity offset relative to the solar system barycenter.

[3]https://casadocs.readthedocs.io/en/v6.5.6/api/tt/casatasks.manipulation.mstransform.html

cleaning taking care of both compact and extended emission. At 234 GHz, it resulted to be enough for my scopes to fix the `AUTO-MULTITRESH` pipeline parameters for the ALMA 12m $b75 >$ 300 m configuration (being $b75$ the 75th percentile of baselines). For the 245 GHz setup, instead, I decreased `minbeamfrac` to 0.01 (with respect to the standard value of 0.3) because, thanks to the high angular resolution, I was confident to retain true source components in smaller beam fractions. I also decreased `sidelobethreshold` and `lownoisethreshold` with respect to their standard values, keeping a signal-to-noise threshold of 5σ, in order to pick up fainter and fainter emission components from the source and boost the expansion of the mask into low signal-to-noise regions.

From Fig. 3.1 to Fig. 3.3 are shown some representative results of my adopted deconvolving, cleaning, and auto-masking procedures. As it can be noticed from Fig. 3.1, I have successfully removed the majority of the sidelobes. Please notice that the contours reported in all dirty/cleaned images refer to an rms estimated in channels images, not the full datacubes, and equal to the mean of the noise levels in three equal size boxes I fixed in the image so as to be far from the source emission. Also the "flatness" (in terms of absence of source emission) of the residuals associated to the cleaned channel images (refer to center images in Fig. 3.2, and Fig. 3.3) is an indicator of the goodness of my adopted cleaning. Moreover, the right-hand side of Fig. 3.2 and Fig. 3.3 demonstrate the capacities of the `CASA AUTO-MULTITHRESH` algorithm for auto-masking different emission patterns, as well as the complexity of the IRAS 4A system, as mentioned in § 1.3. As a final comment, in some channel image, like in Fig. 3.3, the residuals can still show traces of extended emission. This kind of regions will be fine tuned with a manual masking a posteriori when the extended emission will be analyzed in a future work.

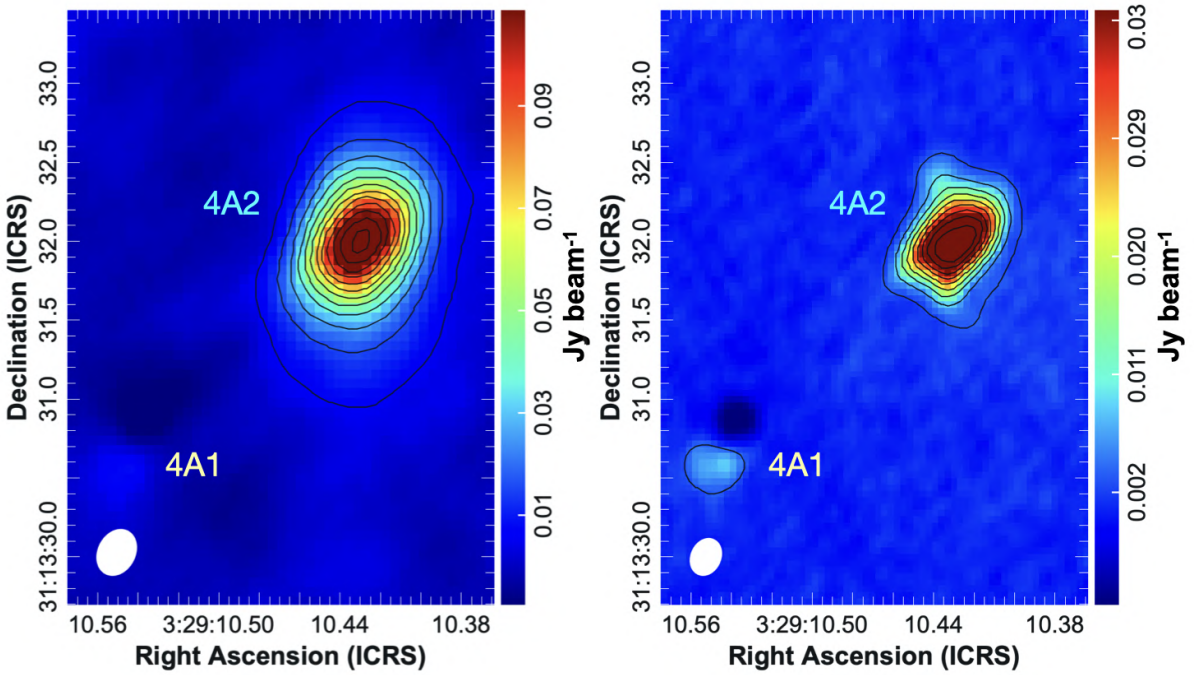

Figure 3.1: Zoomed channel image of IRAS 4A before and after cleaning in datacube centered at 245 GHz. *(Left)* No cleaning has been performed; visibilities are weighting with the `natural` scheme. The beam is 0.24" x 0.17", -28°. *(Right)* Image resulting from the adopted cleaning procedure (`robust` weighting) and the application of the primary beam correction a posteriori. The beam is 0.17" x 0.12", -24°. In both figures, the 5σ step contours are plotted in black solid line, starting from $3\sigma \sim 6$ mJy beam^{-1} *(left)* and $3\sigma \sim 3$ mJy beam^{-1} *(right)*; the colorbar indicates the map intensity in units of Jy beam^{-1}.

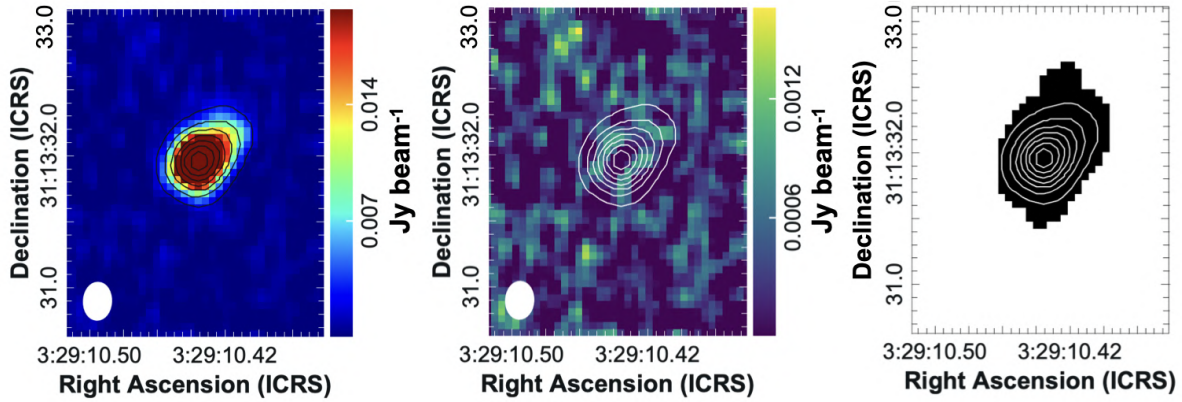

Figure 3.2: Channel image composition of IRAS 4A2 in datacube centered at 234 GHz, showing compact emission pattern, residuals and auto-masking.
(Left) Primary beam corrected and cleaned image in colorbar (Jy beam^{-1}). *(Center)* Final residual image in colorbar (Jy beam^{-1}). *(Right)* Mask calculated internally by the `AUTO-MULTITHRESH` algorithm; not masked regions are in white (0), masked regions in black (1). In all images, in white solid line are shown the cleaned image contours of 5σ step starting from 3$\sigma \sim$ 3 mJy beam^{-1}. The beam is 0.21" x 0.14", -3°.

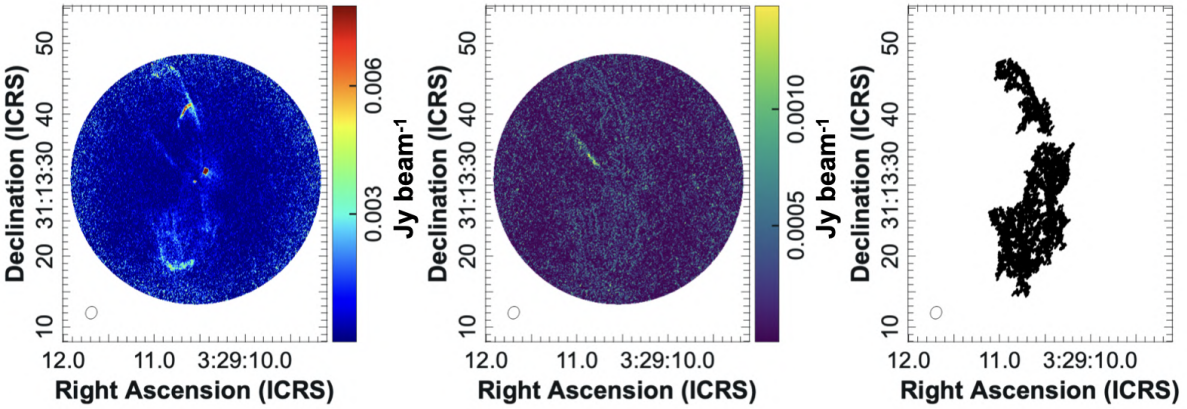

Figure 3.3: Channel image composition of IRAS 4A2 in datacube centered at 245 GHz, showing extended emission pattern, residuals and auto-masking.
(Left) Primary beam corrected and cleaned image in colorbar (Jy beam^{-1}). *(Center)* Final residual image in colorbar (Jy beam^{-1}). *(Right)* Mask calculated internally by the `AUTO-MULTITHRESH` algorithm; not masked regions are in white (0), masked regions in black (1). In all images, in white are shown the cleaned image contours of 5σ step starting from 3$\sigma \sim$ 3 mJy beam^{-1}. The beam (*bottom left*) is 0.17" x 0.12", -24°, and magnified by 3x for visualization purposes.

3.2 Spectra Analysis

Once cleaned the datacubes, I am ready to extract the spectrum towards the target source, IRAS 4A2 (§ 3.2.1), and identify the iCOMs I want to investigate (§ 3.2.2). To do that, I used the CARTA *spectral profiler* tool and molecular spectroscopy catalogs (CDMS[4] and JPL[5]).

3.2.1 Spectrum Extraction

In order to extract and study the spectrum of my target source at both frequencies, I used the map of the continuum produced by the FAUST team in a spectral line-free region centred on 252.9554 GHz. As depicted in Fig. 3.4, it has been obtained with a cleaning procedure adopting a `robust = 0.6` for the `briggs` weighting scheme. The dust emission clearly traces the presence of the two protostars, IRAS 4A1, the brighter spot on the lower left of the map, and IRAS 4A2, my target, contoured on the upper right. I chose to extract the spectrum towards IRAS 4A2 from the continuum peak shown in Fig. 3.5. I reserve as a possibility for a future work to consider a different extraction region depending on the convolved size of emission found for each iCOM (see 3.3.3). I identified the continuum peak directly in the CARTA interface of the continuum map (Fig. 3.4), following the intensity distribution of the target visible in the software interface. This point sets the target coordinates to 3:29:10.431 (R.A.) and +31:13:32.00 (Dec.) in J2000 coordinates. The output spectra, of which the one extracted from the datacube at 245 GHz is reported in Fig. 3.5, show the amazing peculiarity of hot corinos: a dense forest of emission lines, generated by the excitation of rotational, vibrational, and roto-vibrational levels of more or less complex molecules, including iCOMs. The detection of the vast majority of detected spectral lines of the iCOMs I wanted to investigate is robust, since their S/N is well above the 3σ noise level.

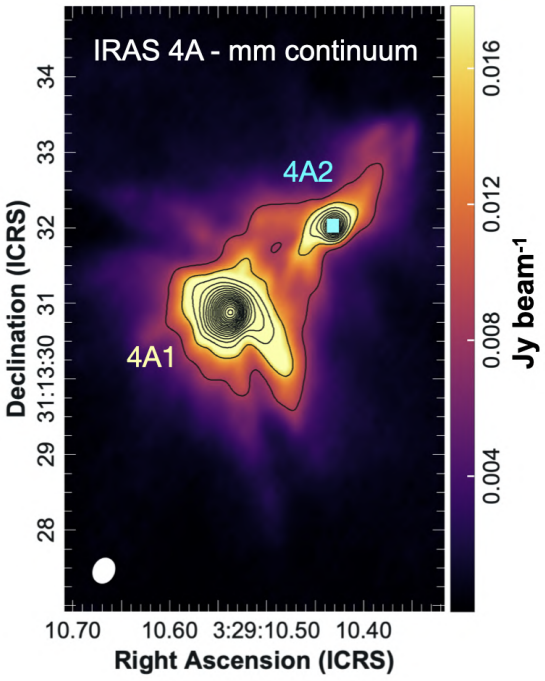

Figure 3.4: Dust continuum image of IRAS 4A system at 252.9554 GHz. Continuum image in colorbar (Jy beam^{-1}) cleaned with `robust = 0.6` and primary beam corrected. The 4σ step contours in black start from $5\sigma \sim 9$ mJy beam^{-1}. The beam is 0.20" x 0.14", -26°. The point used for spectral extraction is indicated with a light blue square and is ~ 86 mJy beam^{-1}.

In order to quantify the noise associated to the two datacubes, I used the `imstat`[6] task of CASA, which allows to retrieve some useful statistical information about an input image or image region. Having to deal with datacubes, one has to avoid the spectral regions dense of lines to properly estimate it, that is evidently motive of struggle with such a line-rich source as IRAS 4A2.

[4] https://cdms.astro.uni-koeln.de/classic/
[5] https://spec.jpl.nasa.gov
[6] https://casadocs.readthedocs.io/en/latest/api/tt/casatasks.information.imstat.html

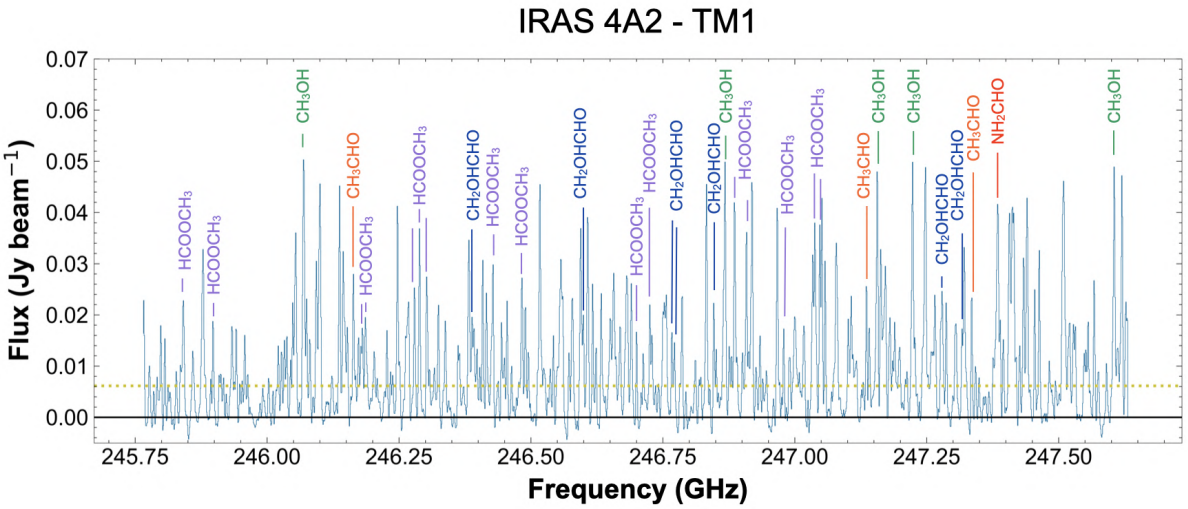

Figure 3.5: IRAS 4A2 spectrum extracted from continuum emission peak at 245 GHz. The spectrum is plotted in blue. The green dotted line marks 5 times the rms of the datacube $\sigma = 1.23$ mJy beam^{-1}, the black solid line the baseline. The rest-frequency of the datacube is 246.700 GHz. The iCOMs lines identified in the datacube are labelled in different colors.

Using the line-free channel ranges used for the continuum identification for this source, I could estimate the noise level of the extracted spectra. The rms values are respectively of 1.16 K (at 234 GHz) and of 1.23 K (at 245 GHz), and represent my guideline to detect iCOMs lines with a S/N $> 3\sigma$, making me confident about their emission and therefore of their identification.

About Flux Recovery

In order to verify whether my interferometric observations are altered by filtering of extended-scale emission, I extracted two spectra towards IRAS 4A2 from two different datacubes using a circular beam (0.6", 3x synthesized beam), and compared the spectral fluxes integrated on selected channel ranges. One spectrum comes from my cleaned datacube at 245 GHz, the other from a datacube in which the 12m TM1 and TM2 datasets (of different large recoverable scale, larger for TM1 having better angular resolution) plus the 7m ACA configuration are combined together. This latter datacube has been cleaned and primary beam corrected with `briggs` weighting (`robust = 0.5`), as I did, but adopting a different pixel size, therefore image size, and applying a different frequency rebinning, such that there are three fewer channels than in my datacube. The spectra are compared in Fig. 3.6, and in Tab. 3.3 is reported the estimated flux loss (column "% Loss") with respect to the difference between integrated fluxes calculated over five different channel ranges (see "Channel" in table). What comes out is that I lose about 25-30% of the flux of my target on a scale comparable to the beam and to the primary beam. The missing flux problem is a known issue for interferometric observations. It mainly depends on the morphology of the target source, and on the configuration of the interferometer. In this regard, other works dealing with sources similar to mine and with spectral datacubes have found that for extended emission the level of missing flux can reach indeed $\sim$ 20-30 % (Fontani et al. 2014; Codella et al. 2015; Plunkett et al. 2023). However, my ALMA/FAUST data have the 7m ACA configuration only at 245 GHz, so I could not perform the same comparison at 234 GHz. Hence, given that I want to *(i)* use the largest available bandwidth to detect as many iCOMs transitions as possible, and *(ii)* use the highest available angular resolution to resolve iCOMs emission, I used the TM1 configuration only ($\sim$ 0.17 $\sim$ 50 AU).

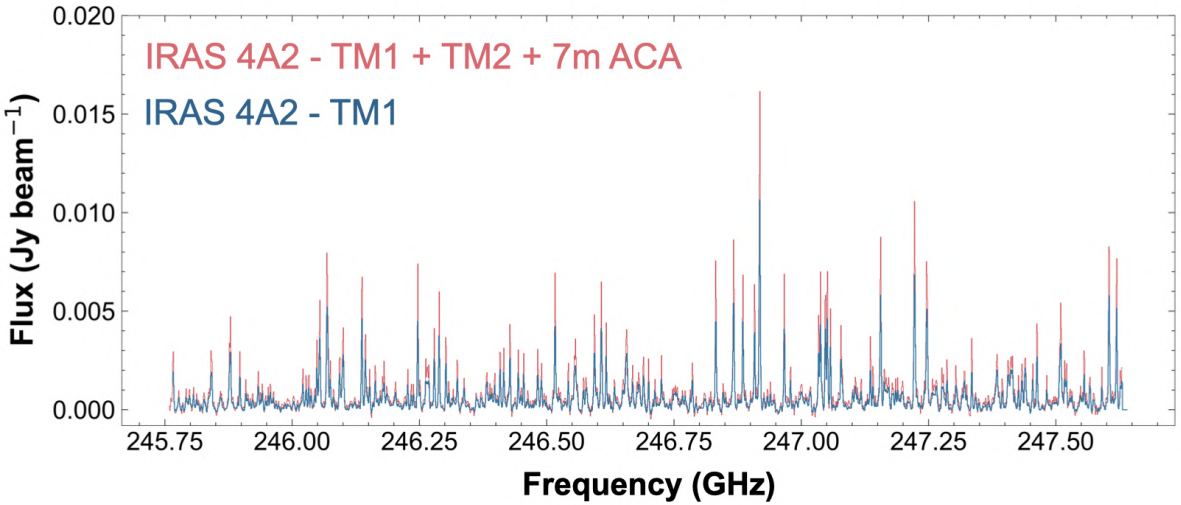

Figure 3.6: Spectral comparison of IRAS 4A2 between joint data (12m-TM1+TM2, 7m ACA) and cleaned data (12m-TM1 only) at 245 GHz. The spectrum extracted from the joint datacube is in red, and the one extracted from my cleaned datacube is overplotted in blue.

Table 3.3: Flux comparison in IRAS 4A2 extraction region between joint datacube (TM1 + TM2 + 7m ACA) and TM1 cleaned datacube at 245 GHz. ν_0 and E_{up} refer respectively to rest-frame frequency and upper state energy of identified iCOMs' lines. $F_{\nu,\ joint}$ refers to the integrated flux estimated in the joint datacube; $F_{\nu,\ TM1}$ to our cleaned datacube at 245 GHz. [a] identifies formaldehyde (H_2CO) transition of QNs: $4_{1,4}$-$3_{1,3}$.

Line	**Emission Type**	ν_0 **[GHz]**	E_{up} **[K]**	**Channel**	$F_{\nu,\ joint}$ **[mJy beam^{-1}]**	$F_{\nu,\ TM1}$ **[mJy beam^{-1}]**	**% Loss**
CH_3OH	$\sim$ 2-3 beam	246.075	537	307-314	29.440	21.050	$\sim$ 28%
CH_2OHCHO	$\sim$ beam	246.395	344	636-640	4.305	3.011	$\sim$ 30%
H_2CHO[a]	$\sim$ primary beam	38	246.925	1178-1183	44.469	32.683	$\sim$ 27%
NH_2CHO	$\sim$ beam	247.391	78	1652-1661	14.674	11.037	$\sim$ 25%
CH_3OH	$\sim$ 2-3 beam	247.611	447	1880-1887	25.594	19.307	$\sim$ 25%

3.2.2 iCOMs Identification

As outlined in the previous section, IRAS 4A2 spectra are extremely enriched in more or less complex species, including iCOMs. I decided to focus on a sample of five iCOMs, which *(i)* are all prebiotic molecules known to be in the IRAS 4A2 hot corino environment (e.g., Bottinelli et al. (2004a); Taquet et al. (2015); De Simone, M. et al. (2017); López-Sepulcre, A. et al. (2017); Sahu et al. (2019); Belloche et al. (2020)) and in other hot corinos (e.g, Jørgensen et al. (2016); Manigand et al. (2020); Bianchi et al. (2020); Codella et al. (2022); Imai et al. (2022)), also allowing me to compare my results with the literature; *(iii)* are all very common in solar-like star forming regions (Ceccarelli et al. 2022); *(iii)* show a relatively bright emission in each of their lines (in my sample, at least $\approx 10\sigma$ significance for the brightest transitions and for a channel spacing of 1.3 $km\ s^{-1}$).

These molecules are[7]:

 *Methanol* (CH_3OH), one of the simplest iCOMs and progenitor of most of them (Caselli & Ceccarelli 2012).

 *Glycolaldehyde* (CH_2OHCHO), the precursor of the simplest monosaccharide[8] sugar, key molecule in the formation path of life compounds such as in the *ribonucleic acid* (RNA) (Larralde et al. 1995; Jørgensen et al. 2012), and relatively abundant in YSOs (De Simone, M. et al. 2017).

 *Acetaldehyde* (CH_3CHO), one of the most detected iCOMs, whose formation route is still largely debated (Vazart et al. 2020).

 *Methyl formate* ($HCOOCH_3$), simplest example of an *ester*[9], whose formation route is still under debate as well (Balucani et al. 2015).

 *Formamide* (NH_2CHO), the simplest *amide*[10], currently thought to form via gas-phase reactions (Kahane et al. 2013; Spezia et al. 2016; Codella, C. et al. 2017), despite not a huge agreement yet. This iCOM is also claimed to be a major building block of metabolic and genetic life compounds (Saladino et al. 2012; López-Sepulcre et al. 2019).

Except for glycolaldehyde, FAUST observations have been tuned to observe the rest of my sample iCOMs as complexity probes of the disk/envelope system of YSOs like in IRAS 4A2 (Codella et al. 2021). Given my main goal of deriving the abundance of these species in the hot corino gas and infer its temperature and temperature structure, I have not taken into account the low abundant isotopologues nor deuterated species of these molecules.

[7] *Figures credits: Adrien Houge.*

[8] Monosaccharides are organic compounds made up of carbon, hydrogen and oxygen. They are carbohydrates that cannot be broken down by the intervention of water.

[9] This kind of compound is derived from the formic acid (HCOOH), where a methyl group (CH_3) is attached to the carboxyl group (C(=O)OH).

[10] Compound with the general formula $R\text{–}C(=O)\text{–}NR'R''$, where R, R', and R'' represent any group, such as hydrogen atoms.

I looked for the iCOMs in the forest of IRAS 4A2 spectral lines starting with a spectral line query to the SPLATALOGUE[11] service within the CARTA spectral profiler.

SPLATALOGUE is one of the largest and most used database for astronomical spectroscopy. It collects the linelist catalogues of several other databases, including CDMS and JPL, which are actually the primary sources of molecular data, and the ones I refer to in this work. In fact, to be confident that the spectral parameters were up to date we verified them directly on the individual CDMS and JPL catalogs, which are frequently updated.

All these databases share the same working principles: one can search for a chemical species, specify a frequency range and/or energy range, and, apart from other search filters, submit the request. The output is a list of line parameters, including: the rest-frame frequency (in MHz or cm^{-1}), the resolved quantum numbers of the transition, the catalog base 10 logarithm of the integrated intensity at 300 K (in nm^2 MHz), the lower state energy (in cm^{-1}), and the upper state degeneracy g_{up}. In the SPLATALOGUE interface in CARTA, I also set the systemic velocity (in km s^{-1}) of the target source, in my case, 6.8 km s^{-1} Choi (2001). In fact, given that spectra are affected by the Doppler effect, I needed to verify that the rest frequency of each line is at their rest velocity.

Using the SPLATALOGUE query in CARTA, it is very difficult to disentangle (also visually) "genuine" spectral lines. What I mean by that are lines which *(i)* have catalog line counterparts (plotted with SPLATALOGUE) peaking at the channel of maximum T_B for that line; *(ii)* their S/N is above the 5σ of datacube noise; *(iii)* are well-isolated and/or not contaminated by other species. Moreover, in case of multiple spectral lines with the same spectroscopic parameters (upper energy state E_{up} and $S\mu^2$ parameter), if we do not see all of them in the spectrum and/or they do not have similar brightness temperature (adopting a tolerance of a factor of 1-1.5 of difference), it means we are not actually observing those transitions, therefore the queried molecule. In the end, we need to see all predicted transitions; if we do not (unless the lines are too weak, or contaminated), the molecule we are looking for is not there.

What we can do to simplify our quest is to identify first the brightest lines, which are usually those characterized by the vibrational quantum number $v = 0$, and fix a reasonable threshold on E_{up} and $S\mu^2$ parameters to pick up strong lines. I have filtered for $E_{up} \leq 500$ K (as expected in hot corinos assuming that the molecular excitation temperature is indicative of the surrounding gas temperature), and $S\mu^2 > 20$ nm^2 MHz.

In Appendix A there are listed the spectral parameters of all identified transitions for each sampled iCOM; Tab. 3.4 refers to CH_3OH only and it is shown here as example. The first column (ν_0) contains the rest-frame frequency of each line, doppler-shifted with respect to the observed frequency in the spectra according to the systemic velocity of the IRAS 4A system. In the second column are listed the resolved Quantum Numbers (QNs) cataloged for all spectral lines; if not explicated in subscript to the QNs, they are pure rotational transitions ($v = 0$).

I therefore calculated $S_{ij}\mu^2$, $log_{10}(A_{ij})$, and E_{up} spectral parameters according to PICKETT et al. (1998) (following CDMS and JPL catalogs' documentation). I derived the 10 base logarithm of the Einstein coefficient A_{ij} by taking the logarithm of the following equation:

$$A_{ij} = I(300\ K)\ \nu_0^2\ \frac{Q(300\ K)}{g_{up}}\ e^{-E_l \cdot 1.44/T}(1 - e^{h\nu_0}/kT) \cdot 2.7964 \times 10^{-16}\ [s^{-1}] \qquad (3.1)$$

where $Q(300\ K)$ stands for the partition function of the parent molecule of a given transition calculated at 300 K, consistently with the line catalog intensity I, that is evaluated at the same temperature. Also, the lower state energy E_l is multiplied by 1.44 to convert it from cm^{-1} to K; the temperature T is fixed at 300 K; h is the Planck constant in cgs units (6.626196×10^{-27} erg $\cdot$ s); k the Boltzmann constant in the same measurement units system.

[11] https://splatalogue.online/

I evaluated the $S\mu_{ij}^2$ values from:

$$S_{ij}\mu^2 = \frac{g_{up}\ A_{ij}}{1.16395 \times 10^{-20}\ \nu_0^3}\ [D^2] \tag{3.2}$$

having the derived A_{ij}, and the rest-frame frequency in MHz; $S_{ij}\mu^2$ is here evaluated in cgs units of squared electic dipole moment (debye, D^2).

Finally, I evaluated the upper state energy E_{up} first in units of cm^{-1} given:

$$E_{upper} = E_{low} + \left(\frac{\nu_0}{c}\right) \times 10^9\ [cm^{-1}] \tag{3.3}$$

in which E_{low} is in cm^{-1}, ν_0 in GHz and c, the speed of light, in cm s^{-1}. To get the energies in K, I multiplied these results by a numerical factor of 1.44.

Table 3.4: Spectral properties of methanol lines identified in IRAS 4A2 extracted spectra.

CH_3OH - N_{lines} = 9, CDMS catalog (PICKETT et al. 1998)					
ν_0 [MHz]	QNs	$S_{ij}\mu^2$ [D^2]	$Log_{10}(A_{ij})$	E_{up} [K]	g_{up}
232945.797	$10_{3,7}$-$11_{2,9}$ E	12.15	-4.672	191	84
233795.666	$18_{3,15}$-$17_{4,14}$ A	21.89	-4.658	447	148
234683.370	$4_{2,3}$-$5_{1,4}$ A	4.48	-4.727	61	36
234698.519	$5_{4,2}$-$6_{3,3}$ E	1.85	-5.198	123	44
246074.605	$20_{3,17}$-$20_{2,18}$ A	78.14	-4.084	537	164
246873.301	$19_{3,16}$-$19_{2,17}$ A	73.78	-4.082	491	156
247161.950	$16_{-2,15}$-$15_{-3,13}$ E	19.33	-4.589	338	132
247228.587	$4_{2,2}$-$5_{1,5}$ A	4.34	-4.673	61	36
247610.918	$18_{3,15}$-$18_{2,16}$ A	69.51	-4.081	447	148

I compared the identified iCOMs transitions with the literature, in particular, with the works cited at the beginning of this Section having data at my same frequencies. As a result of this survey, I have observed all methanol lines already detected towards IRAS 4A2 and other hot corinos, and the same applies to the spectral line of formamide (NH_2CHO) with resolved QNs $12_{0,12}$-$11_{0,11}$. Similarly to Codella et al. (2022), I also detected NH_2CHO line $12_{2,10}$-$11_{2,9}$ at 260189.848 MHz, using the cleaned narrow-band spectral window with ID 35 at 245 GHz.

As regarding the other molecules of our sample, namely, glycolaldehyde (CH_2OHCHO), acetaldehyde (CH_3CHO), and methyl formate ($HCOOCH_3$), some of the transitions are blended together due to the limited spectral resolution of FAUST/ALMA observations, resulting in less agreement with the literature. Taking glycolaldehyde as an example, the doublet $23_{0,23}$-$22_{1,22}$ and $23_{1,23}$-$22_{0,22}$ detected by Imai et al. (2022), in my datacube at 234 GHz is blended in one single channel. CH_3CHO line $14_{1,14}$-$13_{1,13,v=1}$ E detected by López-Sepulcre, A. et al. (2017) at 260530.40 MHz, instead, was out of every narrow-band SPW to enable me to attempt a detection. $HCOOCH_3$ is the most problematic, since it is an incredibly abundant iCOM in hot corinos, with plenty of rotational and vibrational relatively bright lines that have been detected so far. Yet, just to mention one case, the lines $19_{9,11}$-$18_{9,10}$ A and $19_{9,10}$-$18_{9,9}$ A listed in Okoda et al. (2023), appear to be blended in one channel of my extracted spectrum at 234 GHz.

In light of the discussion above, I am confident with the lines detection, of at least 3σ significance, and with the statistical sample of transitions for each targeted iCOM: summing the emission lines detected at the two observed frequencies, in total I got 9 for methanol, 12 for glycolaldehyde, 5 for acetaldehyde, 24 for methyl formate, and 6 for formamide.

3.3 Resolve And Compare iCOMs Spatial Distribution

In this thesis project, I aimed at constraining something more than the abundance and the gas temperature probed by a sample of fundamental iCOMs. Indeed, I am able to resolve the emission of iCOMs in a hot corino region ($<$ 50 AU). This allows me also to image these complex organic molecules, in particular, looking at integrated intensity maps of their emission lines (§ 3.3.1). But, most importantly, in order to "observe" and quantify whether there is a spatial segregation of iCOMs at different scales in the hot corino, comparing their integrated emission and fitting these emission maps to extract information about the size of emission (§ 3.3.2 and § 3.3.3). In the end, combining together these two pieces of information, that is, iCOMs spatial segregation on the one hand, and physical properties (namely, gas temperature and species abundance) probed by these molecules on the other hand (§ 3.4), I can attempt to outline a temperature profile of IRAS 4A2 that is purely based on observations, and at scales never reached nor accurately modelled before (§ 3.5).

3.3.1 Integrated Intensity Maps

I produced integrated intensity maps, or moment-0 maps, of a selection of of identified iCOMs lines at the two observed frequencies. Namely, the ones that are the brightest (with $T_B > 15$ K), most isolated in the spectra, and associated to the compact emission of IRAS 4A2 hot corino only. In this way, I can then compare the moment-0 maps of lines of different iCOMs at similar excitation energy E_{up} (likely tracing the same gas), and obtain a reliable estimate of the emitting size radius of each iCOM (see § 3.3.3).

To do it, I used the `CASA` task `immoments`[12]. Given the shape of a datacube, in general one collapses the cube into a *m*-moment image by taking a linear combination of the individual image planes stacked along the spectral direction:

$$M_m(x_i, y_i) = \sum_k^N w_m(x_i, y_i, \nu_k)\ I(x_i, y_i, \nu_k) \tag{3.4}$$

for pixel i and channel k in the datacube I. In my case, I produced 0-moment images.

Spectral moments help extract useful information about the distribution of flux as a function of wavelength across an astronomical object image. The most commonly derived spectral moment distributions correspond to *(i)* m = 0, the total integrated flux within a certain frequency range (my case), to estimate the overall brightness of a source, *(ii)* m = 1 (*centroid*), the average frequency where most of the source flux is concentrated, *(iii)* m = 2 (*velocity dispersion*), measure of the spread of the spectral distribution around the centroid, *(iv)* m = 3, 4 (*skewness* and *kurtosis*) measure of the asymmetry of the spectral distribution (skewness), and of the "clumpiness" of the distribution (kurtosis). In my case, the selected lines are integrated on average over four channels ($\sim$ 4.8 km s^{-1}), according to the typical line width of hot corino emission.

The comparison of the integrated emission leads to a first hint of spatial segregation of iCOMs: different molecules seem to trace different scales.

[12] https://casadocs.readthedocs.io/en/stable/api/tt/casatasks.analysis.immoments.html

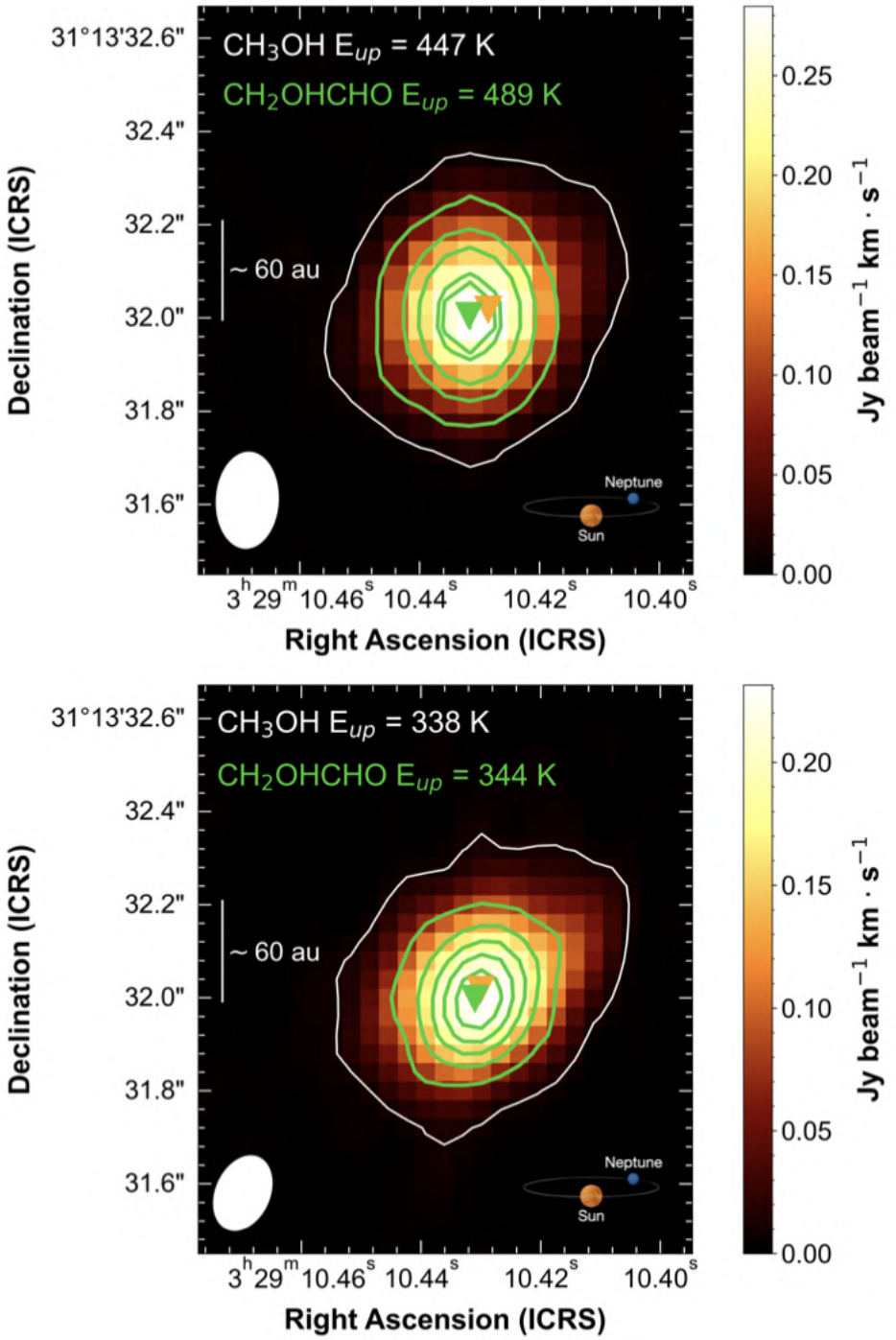

Figure 3.7: Evidence of spatial segregation of methanol and glycolaldehyde in ALMA/FAUST extracted spectra towards IRAS 4A2. Comparison of integrated emission at 234 GHz *(upper panel)* and 245 GHz *(bottom panel)*. Methanol emission is shown in colour scale plus a 3σ white contour of 10 mJy beam^{-1} km/s (234 GHz) and 13 mJy beam^{-1} km/s (245 GHz). Glycolaldehyde emission is in 5σ step green contours starting from $3\sigma \sim 7$ mJy beam^{-1} km/s (234 GHz) and ~ 9 mJy beam^{-1} km/s (245 GHz). In both panels, the triangles in orange and green indicate the emission peaks of methanol and glycolaldehyde. The restoring beam is 0.21" x 0.14", −3° *(upper panel)*, and 0.17" x 0.12", −24° *(bottom panel)*. The sketch on the bottom right shows the Solar System size.

In particular, methanol ($18_{3,15}$-$17_{4,14}$ A at 234 GHz and $16_{-2,15}$-$15_{-3,13}$ E at 245 GHz) and glycolaldehyde ($38_{11,27}$-$38_{10,28}$ at 234 GHz and $34_{5,30}$-$34_{4,31}$ at 245 GHz) display integrated emission patterns both centred on IRAS 4A2 (within the errorbars of the maps flux peaks estimated with `imfit` `CASA` task, see § 3.3.2), but with clear different extension, comparable to Solar System scales. In Appendix B are reported all moment 0 maps comparisons performed at both observed frequencies. At the stage of this "qualitative" investigation of spatial segregation of iCOMs, methanol seems to trace the largest hot-corino scale, followed by methyl formate, then acetaldehyde, formamide and finally glycolaldehyde, which appears to be emitted from the most compact hot gas scale in my sample. Looking more in detail, formamide seems to be associated with a very compact emission pattern as well; methyl formate, on the other side, could compete with methanol on large scales; acetaldehyde apparently settles down in the middle of these extremes at both frequencies. Fig. 3.7 resumes the evidences of different spatial segregation for methanol and glycolaldehyde. It should be noticed that the centre of emission evaluated at 3σ is just indicative of the spatial segregation, because it fixes a value that can be similar for lines of different intensity.

Therefore, in order to quantify these different scales of emission and secure their measurements, I performed a gaussian fit to model the source in the iCOMs integrated intensity maps, and extract, above all fitting parameters, the radius of their emitting sizes (§ 3.3.2).

3.3.2 Two Sides Of The Same Coin: Gaussian Fitting on *u-v* and image plane

In general, a gaussian fit allows to model a source with gaussian shape. For me, the (single) source to model is IRAS 4A2, as seen through the emission of the sampled iCOMs, while the gaussian modelling is justified by assuming a gaussian beam for the observations. Having in mind the interchange of visibility and brightness distributions through Fourier transforms, (2.15) and (2.16), this fit can be performed either on the image plane, in my case, the moment-0 maps I produced, or on the *(u-v)* plane, that is, corresponding to the visibility datasets of the same imaged transitions in their moment 0. The *u-v* model fitting in principle can give more accurate results than the image plane fitting, because we are not limited by the size of the beam. Yet, while for the latter we just need to specify the image or image region where to fit (focusing on the target source), in order to fit *u-v* data we need to be sure we are picking up only the visibilities of the target, which in turn means to restrict to bright lines associated to IRAS 4A2 emission only. I was able to perform both image and *u-v* fitting for the selected lines at 245 GHz, instead, at 234 GHz, it was not possible to safely employ the modelling on the *u-v* plane. In fact, at this frequency iCOMs lines tended to be contaminated and associated with strong absorption and emission towards IRAS 4A1. I actually proved that at 245 GHz no significant difference (within errorbars) intervenes between the two methods in terms of resulting radius of emission (as shown in Fig. 3.10). So, to stay consistent at the two observed frequencies, I proceeded with the data analysis considering only image plane fitting parameters. I implemented the gaussian fitting on moment-0 maps and *u-v* data using `CASA` tasks `imfit`[13] and `uvmodelfit`[14] respectively.

[13] https://casa.nrao.edu/docs/taskref/imfit-task.html
[14] https://casadocs.readthedocs.io/en/stable/api/tt/casatasks.manipulation.uvmodelfit.html

Imfit

I specified a rectangular region on the image plane (through the `box` parameter of `imfit`) where to search for the source and model it. I created this rectangular region in the `CASA` Viewer[15] with ICRS coordinates $\sim 52'' \times 31''$ and used it for all moment-0 maps. I also computed the residual image associated to each gaussian fit. Fig. 3.8 and 3.9 show, as example, the fitted moment-0 maps and the corresponding image fit residuals of methanol and glycolaldehyde lines compared in Fig. 3.7; the other fits are collected in Appendix C.

Again, as visible at a first look in the left panels of Fig. 3.8 and 3.9, these two molecules seem to trace clearly different scales. However, in addition to the previous investigation, now we can see this difference also in terms of the fitted elliptical gaussian, convolved with the beam, which is contoured in blue solid line in figures. In this way, we can can be sure that the observed spatial difference is not due to an image representation bias.

Willing to obtain the least-biased estimate of emitting radius starting from the two fitted axis of these elliptical gaussians, I wanted to consider deconvolved (from the beam) fitted images size. That is why I analyzed the fit residual images (right panel in figures). For all selected lines, the residuals are irrelevant at a 3σ level, therefore justifying me to derive an emitting radius of emission deconvolved from the beam.

Uvmodelfit

Switching to the Fourier plane, before performing the fit I had to recast the target visibilities. In particular, since `uvmodelfit` works with source coordinates in offsets, I changed the Phase Center of my visibilities dataset according to IRAS 4A2 estimated ICRS coordinates from the continuum peak (used for spectral extraction, refer back to § 3.2.1), exploiting `fixvis`[16] `CASA` task.

Moreover, in order to limit the amount of visibilities to fit, hence helping the algorithm in the fitting procedure, I used the `split`[17] `CASA` task to derive, starting from the newly phase centered target dataset, smaller datasets collecting at most visibilities corresponding to the channels of integration of lines' emission (different for each transition). Specifically, I averaged channels to form one output channel fixing the `width` parameter, and averaged also in time considering a standard time bin of 30 s (`timebin` parameter).

Having now set up everything for `uvmodelfit`, I gave in input to this task every splitted visibilities dataset and, fixing a gaussian component model type (`comptype = 'G'`), I gave in input also the starting guess values for component parameters (through `sourcepar`). In case of a gaussian model type, the parameters that get first guessed and estimated are, in order: flux (Jy), source coordinates (in offset east and offset north, both in arcsec), FWHM along the major axis (arcsec), the ratio of minor to major axis (< 1), and the position angle (deg). For me, in terms of first guesses, the peak flux is the maximum one associated to each iCOM's line (visible in the `CARTA` spectral profiler), the offsets are fixed to 0 (because the visibilities are already phase centered on our target), while major axis, ratio of minor to major axis and position angle are set accordingly to the beam properties of the datacube.

I performed each fit using 20 iterations (`niter`) and let vary all fit parameters except for the source offset coordinates (controlling the `varypar` boolean parameter). The results coming from these fitting procedures are presented and discussed in the following subsection.

[15] https://casa.nrao.edu/casadocs/casa-5.1.0/image-cube-visualization/viewer-basics
[16] https://casa.nrao.edu/docs/taskref/fixvis-task.html
[17] https://casa.nrao.edu/docs/taskref/split-task.html

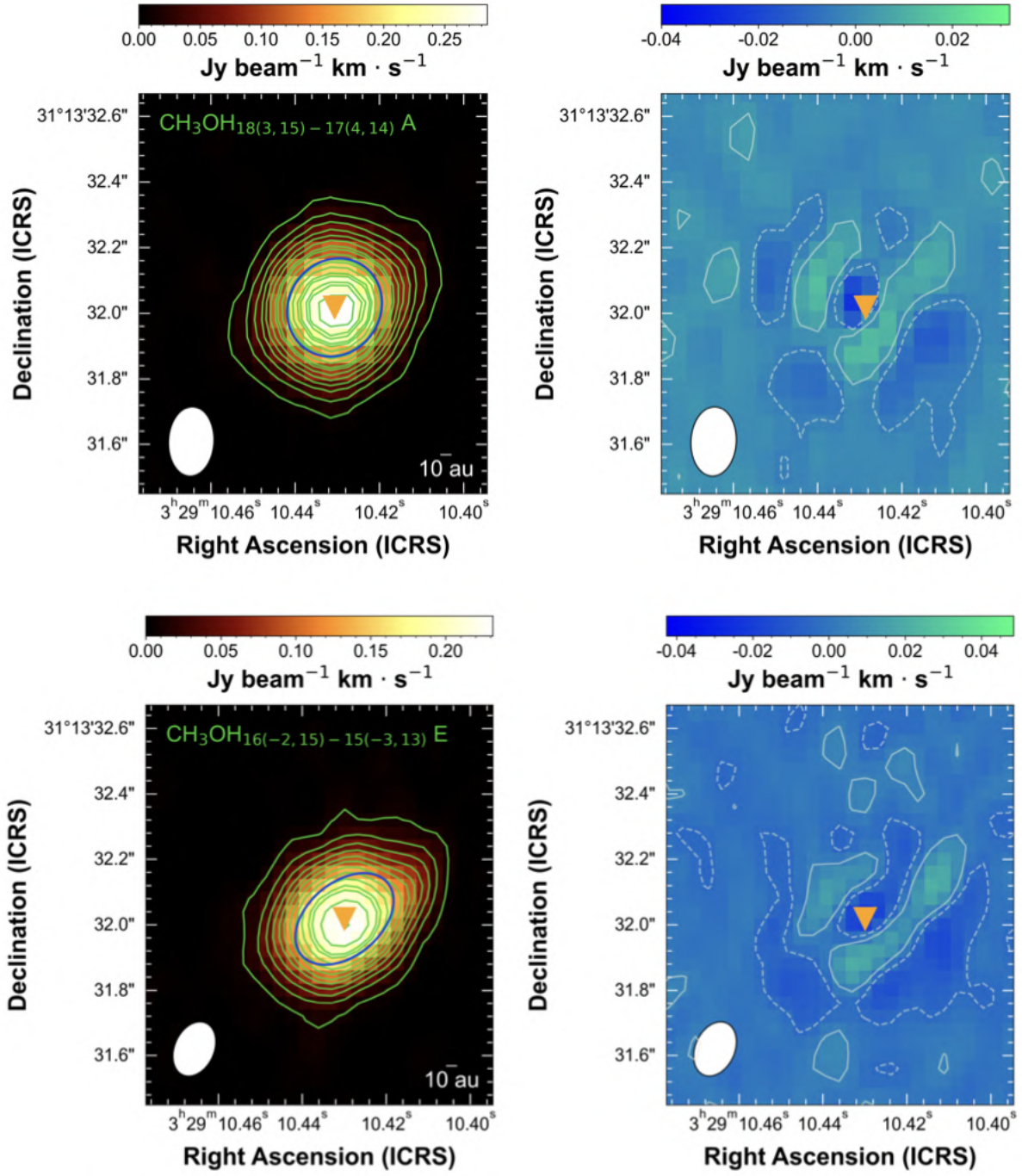

Figure 3.8: Gaussian fitted moment-0 map (left) and fit residuals (right) of selected methanol lines. Top panels refer to 234 GHz data (beam: 0.21" x 0.14", –3°), bottom panels to 245 GHz data (beam: 0.17" x 0.11", –24°). Methanol integrated emission is shown in color scale and green contours starting from 3σ of $\sim$ 10 mJy beam^{-1} km/s *(top)* and 11 mJy beam^{-1} km/s *(bottom)* with step of 5σ. The fitting elliptical gaussian convolved with the beam is overlayed in blue solid line. The corresponding fit residual images are in color scale and contoured with $\pm1\sigma$ of $\sim$ 3 mJy beam^{-1} km/s *(top)* and $\sim$ 4 mJy beam^{-1} km/s *(bottom)*. In all panels, the triangles in orange indicate the emission peaks of methanol.

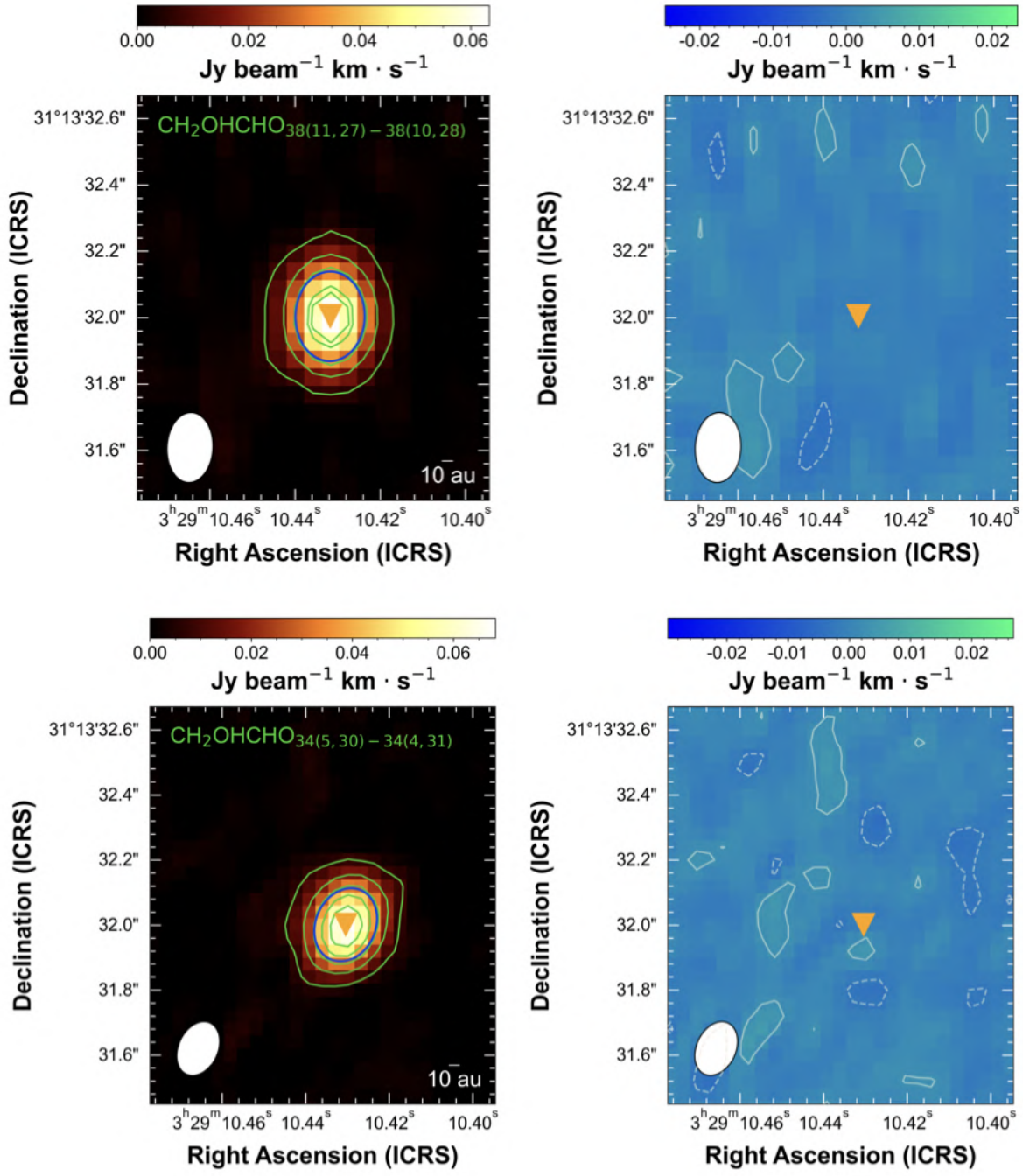

Figure 3.9: Gaussian fitted moment-0 map (left) and fit residuals (right) of selected glycolaldehyde lines. Top panels refer to 234 GHz data (beam: 0.21" x 0.14", −3°), bottom panels to 245 GHz data (beam: 0.17" x 0.12", −24°). Glycolaldehyde integrated emission is shown in color scale and green contours starting from 3σ of 7 mJy beam^{-1} km/s *(to and bottom)* with step of 5σ. The fitting elliptical gaussian convolved with the beam is overlayed in blue solid line. The corresponding fit residual images are in color scale and contoured with $\pm 1\sigma$ of $\sim$ 2 mJy beam^{-1} km/s *(top and bottom)*. In all panels, the triangles in orange indicate the emission peaks of glycolaldehyde.

3.3.3 iCOMs Footprints: Size Of Emission

In Tab. 3.5 and 3.6 are reported the source parameters as estimated by the 2D gaussian fitting procedure over integrated intensity maps at 234 GHz and 245 GHz respectively. The source coordinates (R.A and Dec. in table) are all consistent with my estimate coming from the dust continuum peak (see § 3.2.1. More importantly, having in hand the estimate of source sizes deconvolved from the beam, in terms of major axis and minor axis (θ_M and θ_m in table), I computed the beam-deconvolved effective radius of emission of each selected iCOM's line using the formula:

$$r_{eff} = d_{pc} \frac{\sqrt{\theta_M \cdot \theta_m}}{2} \; [AU] \tag{3.5}$$

where $d_{pc} \sim 300$ pc is the distance of IRAS 4A as estimated by Zucker et al. (2018), while the square root indicates the geometric average (Taylor 1996) of the source size along major and minor axis. This geometric average is divided by half because the size components are diameters.

I calculated the uncertainty on this estimate of the radius as:

$$\Delta r = \Delta(\sqrt{\theta_M \cdot \theta_m}) \; 1/2 \; d_{pc} \; [AU] \tag{3.6}$$

using the rules of propagation of errors on a square root and given the multiplication by constant factors (Taylor 1996). $\Delta(\sqrt{\theta_M \cdot \theta_m})$ is the absolute error on the square root (geometric average), that is:

$$\Delta(\sqrt{\theta_M \cdot \theta_m}) = \delta(\sqrt{\theta_M \cdot \theta_m}) \cdot \sqrt{\theta_M \cdot \theta_m} = \frac{\Delta\theta_M/\theta_M + \Delta\theta_m/\theta_m}{2} \cdot \sqrt{\theta_M \cdot \theta_m}$$

in which the relative error (δ) is calculated using the errors on the source sizes as estimated by `imfit` (reported in Tab. 3.5 and 3.6).

Table 3.5: Image plane gaussian fitted parameters for selected transitions at 234 GHz. The error on the estimated parameters is indicated in round brackets.

Molecule	QNs	R.A. [h: m: s]	Dec. [°,′,″]	θ_M ["$\times 10^{-3}$]	θ_m ["$\times 10^{-3}$]	P.A. [°]
CH_3OH	$4_{2,3}$-$5_{1,4}$ A	03:29:10.43084 (0.002")	+031.13.32.02215 (0.002")	351 (8)	232 (7)	129 (3)
	$5_{4,2}$-$6_{3,3}$ E	03:29:10.43066 (0.002")	+031.13.32.02293 (0.002")	315 (8)	229 (8)	121 (4)
	$10_{3,7}$-$11_{2,9}$ E	03:29:10.43062 (0.002")	+031.13.32.02319 (0.002")	331 (7)	224 (6)	130 (3)
	$18_{3,15}$-$17_{4,14}$ A	03:29:10.430612 (0.0013")	+031.13.32.017780 (0.0014")	257 (6)	207 (8)	107 (6)
$HCOOCH_3$	$18_{4,14}$-$17_{4,13}$ E	03:29:10.43191 (0.002")	+031.13.32.01001 (0.002")	313 (8)	229 (8)	117 (4)
	$19_{4,16}$-$18_{4,15}$ A	03:29:10.43203 (0.002")	+031.13.32.00967 (0.002")	322 (8)	231 (8)	118 (4)
	$19_{11,9}$-$18_{11,8}$ E	03:29:10.43202 (0.003")	+031.13.32.00975 (0.003")	277 (11)	220 (13)	105 (10)
CH_3CHO	$12_{2,11}$-$11_{2,10,\ vt=1}$ E	03:29:10.431107 (0.0011")	+031.13.32.006375 (0.0014")	206 (4)	182 (6)	91 (12)
	$12_{4,8}$-$11_{4,7,\ vt=2}$ E	03:29:10.43116 (0.0018")	+031.13.31.99791 (0.002")	186 (11)	165 (13)	52 (26)
NH_2CHO	$11_{3,9}$-$10_{3,8}$	03:29:10.430827 (0.0008")	+031.13.32.008971 (0.0012")	165 (6)	139 (5)	149 (9)
	$11_{4,7}$-$10_{4,6}$	03:29:10.430572 (0.0008")	+031.13.32.005905 (0.0012")	165 (6)	153 (5)	154 (19)
CH_2OHCHO	$26_{10,17}$-$26_{9,18}$	03:29:10.431511 (0.0013")	+031.13.32.001410 (0.002")	154 (10)	144 (8)	16 (79)
	$38_{11,27}$-$38_{10,28}$	03:29:10.43176 (0.0017")	+031.13.32.00390 (0.003")	169 (12)	155 (4)	13 (89)

Table 3.6: Image plane gaussian fitted parameters for selected transitions at 245 GHz. The error on the estimated parameters is indicated in round brackets. [a] refers to the transition detected in the narrow SPW centred at 260 GHz.

Molecule	QNs	R.A. [h: m: s]	Dec. [°,′,″]	θ_M ["$\times 10^{-3}$]	θ_m ["$\times 10^{-3}$]	P.A. [°]
CH_3OH	$4_{2,2}$-$5_{1,5}$ A	03:29:10.42972 (0.003")	+031.13.32.01928 (0.003")	353 (9)	239 (6)	130 (3)
	$16_{-2,15}$-$15_{-3,13}$ E	03:29:10.42992 (0.002")	+031.13.32.01795 (0.0019")	302 (7)	224 (6)	122 (4)
	$18_{3,15}$-$18_{2,16}$ A	03:29:10.42997 (0.002")	+031.13.32.01694 (0.002")	310 (8)	224 (6)	125 (4)
$HCOOCH_3$	$20_{13,8}$-$19_{13,7}$ E	03:29:10.43125 (0.003")	+031.13.32.00364 (0.003")	258 (13)	204 (13)	107 (11)
	$20_{6,15}$-$19_{6,14,\ v=1}$A	03:29:10.43142 (0.004")	+031.13.32.00354 (0.004")	247 (16)	204 (16)	106 (17)
CH_3CHO	$14_{0,14}$-$13_{1,13}$ A	03:29:10.43074 (0.003")	+031.13.31.99903 (0.002")	249 (10)	192 (11)	95 (9)
	$13_{0,13}$-$12_{0,12,\ vt=2}$ E	03:29:10.43029 (0.0016")	+031.13.31.99569 (0.0018")	160 (8)	157 (9)	73 (75)
NH_2CHO	$12_{0,12}$-$11_{0,11}$	03:29:10.430255 (0.0010")	+031.13.32.001735 (0.0012")	158 (5)	136 (3)	148 (10)
	$12_{2,10}$-$11_{2,9}$ [a]	03:29:10.43056 (0.007")	+031.13.31.98662 (0.008")	146 (36)	132 (52)	61 (55)
CH_2OHCHO	$30_{2,28}$-$30_{1,29}$	03:29:10.43121 (0.002")	+031.13.31.99542 (0.003")	149 (14)	129 (17)	30 (37)
	$34_{5,30}$-$34_{4,31}$	03:29:10.43041 (0.0019")	+031.13.32.00123 (0.002")	153 (9)	139 (7)	158 (51)
	$36_{11,26}$-$36_{10,27}$	03:29:10.43080 (0.0018")	+031.13.31.9974 (0.003")	143 (12)	115 (13)	16 (19)

In Fig. 3.10 I compare the effective radii estimates at 245 GHz coming from image plane and Fourier plane. As anticipated and as now visible in this figure, these estimates are consistent between the two fitting methods.

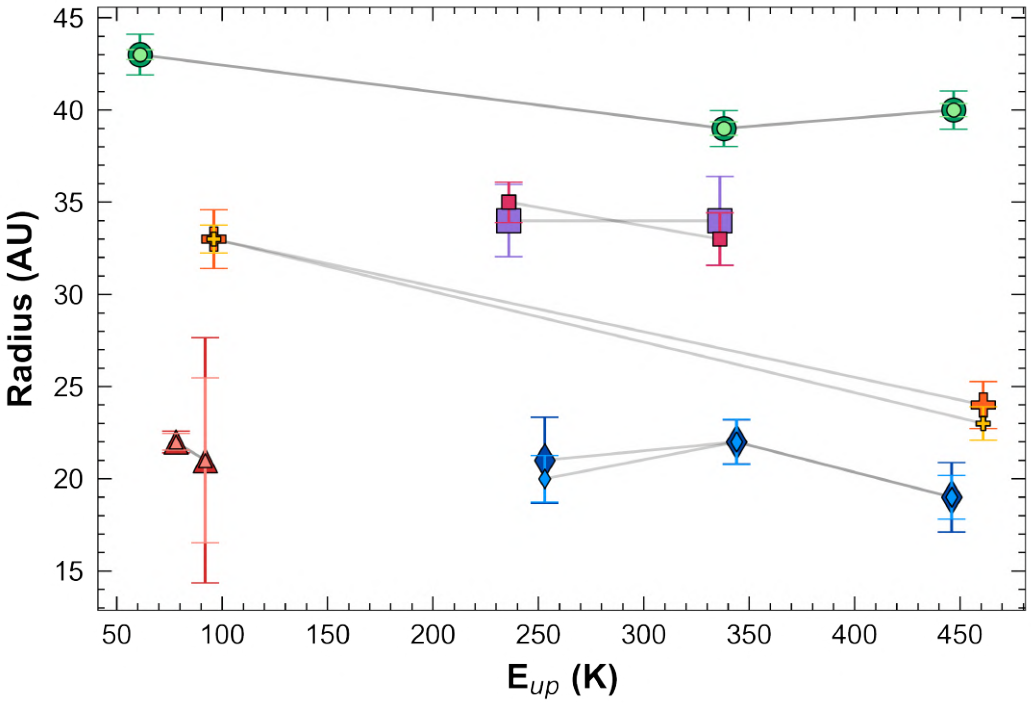

Figure 3.10: Trend with excitation energy of iCOMs effective radii of emission as derived from image plane and (u-v) plane gaussian fitting at 245 GHz. Selected iCOMs transitions for which I estimated the radius are plotted with circles (CH_3OH), squares ($HCOOCH_3$), crosses (CH_3CHO), triangles (NH_2CHO) and diamonds (CH_2OHCHO). Light-colored and smaller markers correspond to radii estimates from *(u-v)* plane fitting. Dark-colored and bigger markers to radii estimates from image plane fitting. The data points errorbars are computed according to (3.6).

Tab. 3.7 summarizes another striking result of this thesis project: the iCOMs I sampled, all fundamental building blocks of prebiotic chemistry, emit from scales of the hot corino gas totally comparable to the Solar System size between Uranus's orbits ($\sim$ 20 AU) and the inner main region of the *Kuiper Belt* ($\sim$ 50 AU; Kuiper (1951)). As quantitative confirmation of what can be seen in Fig. 3.7, methanol and glycolaldehyde emitting sizes define the two extremes, upper and lower respectively, of the spatial segregation of my sample iCOMs in the hot corino environment of IRAS 4A2.

Table 3.7: Emitting radius ranges of sampled iCOMs derived from image plane fitting.

	Radius [AU]	
iCOM	**234 GHz**	**245 GHz**
CH_3OH	34 - 43	39 - 43
$HCOOCH_3$	37 - 41	34
CH_3CHO	26 - 29	24 - 33
NH_2CHO	23 - 24	21 - 22
CH_2OHCHO	22 - 24	19 - 22

I can exclude a dependence of these estimates on the fitting method, at least for transitions detected at 245 GHz. At this frequency, indeed, the emitting size ranges derived from the Fourier plane are practically equal to the ones derived starting from the image plane. I registered a small discrepancy only for acetaldehyde (from *u-v* fitting: 23 - 33 AU), and methyl formate (from *u-v* fitting: 33 - 35 AU). Still, within uncertainties the two fitting methods give perfectly consistent results. In the following analysis I therefore consider `imfit` results only.

Is there any trend with the frequency of observation and/or the excitation energy of the selected transitions? In Fig. 3.11, I put all together the emitting radii as estimated with `imfit` at both observed frequencies, and performed a linear fit on the emitting sizes found for each iCOM. For most of my sample, the slope b of the fitting line does not deviate significantly from 0 within errorbars ($\leq 1\%$ deviation), which means it is reasonable to assume a uniform size. Acetaldehyde and methyl formate display a steeper decrease (in order, $2 \pm 0.5\%$ and $4 \pm 1\%$ deviations from 0) and in size with increasing excitation energy. Formamide shows an apparent increase of the size with the energy of $5 \pm 3\%$, but within the large errorbars I am still legitimated to consider a uniform size also for this iCOM.

Then, if it is reasonable to assume iCOMs have a uniform emitting size with frequency and with the excitation energy of their transitions, I can take an average of their radii ranges and obtain mean molecular sizes. And finally, by combining these average emitting sizes with average gas temperatures traced by these iCOMs (discussed in the following section), in § 3.4.2 I managed to derive a temperature profile of IRAS 4A2 hot corino on scales < 50 AU.

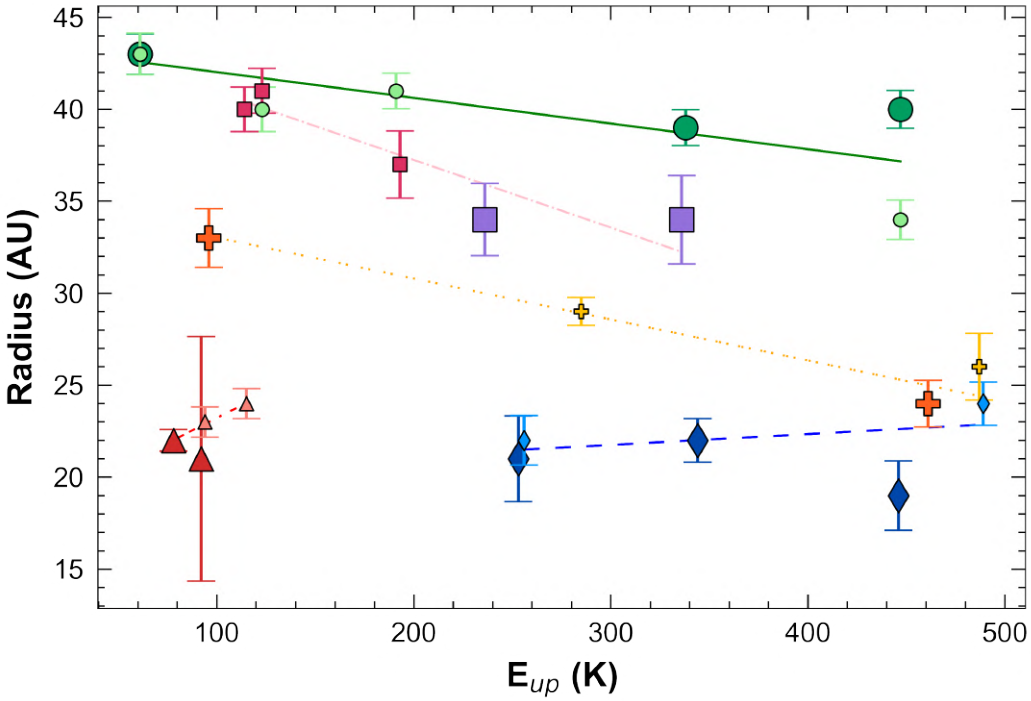

Figure 3.11: Trend with excitation energy of iCOMs effective radii of emission at observed frequencies and plus size fitting. Selected iCOMs transitions for which the radius has been estimated are plotted with circles (CH_3OH), squares ($HCOOCH_3$), crosses (CH_3CHO), triangles (NH_2CHO) and diamonds (CH_2OHCHO). Light-colored and smaller markers correspond to radii estimates at 234 GHz. Dark-colored and bigger markers to radii estimates at 245 GHz. The data points errorbars are computed according to (3.6). The straight lines of different color and linestyle show the linear fit performed on the estimated radii at both frequencies for each iCOM.

3.4 Constraining The Physical Properties Of The Hot Corino Gas

Molecular spectral lines constitute the perfect tools to infer physical properties of the gas hosting their parent molecules. In this thesis project, I was interested in constraining the abundance, in terms of column density of my sample of iCOMs, as well as the gas temperature probed by these molecules in the IRAS 4A2 hot corino region.

3.4.1 Line Gaussian Fitting

The profile of a spectral line depends on several parameters, hence appearing to be more or less broad. The type of broadening characterizing my lines is the Doppler broadening, originated from the random distribution of speeds (with respect to our line of sight) of molecules in thermal equilibrium with the hot corino gas (Böhm-Vitense 1989; Carroll & Ostlie 2007). The Doppler broadening is known to be shaped as a Gaussian distribution (Taylor 1996):

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{(x-\overline{x})^2}{2\sigma^2}} \tag{3.7}$$

where $f(x)$ encloses the velocity distribution, $\overline{x}$ is the mean velocity, and σ the standard deviation. The full width at half maximum of this distribution is related to the standard deviation by: FWHM $= 2\sqrt{2\ ln(2)}\ \sigma$. In light of this brief theoretical insert, I fitted in `CARTA` iCOMs identified spectral lines with a Gaussian model in a small spectral window centered on each line (or double/triple lines if the case). In addition, I fitted the spectral *baseline*, that `CARTA` automatically removes in the fitting process of a given spectral window. In fact, to extract meaningful information from spectral lines, in general one first needs to subtract the dust (continuum) emission, and correct for any instrumental effect that produces fluctuations in this continuum. The choice of the degree of the fitting polynomial is crucial to eliminate large-scale (with respect to the spectral line widths) variations of the continuum, and to avoid introducing small-scale fluctuations, which in both cases would result in an alteration of the intensity of a spectral line. In my case, I wanted to remove any residual fluctuation in the continuum-subtracted data, and take care of small-scale fluctuations.

Fig. 3.12 and Tab. 3.8 resume representative results of this section; see Appendix D, and Appendix E for full reference.

First of all, spectra are reported in brightness temperature, according to (2.19), as a function of radio velocity. We can estimate this velocity as if the rest-frequency was centered on each of the fitted lines. Having in mind the Doppler effect formulation, the systemic velocity is equal to:

$$v_{sys} = \left(\frac{\nu_{0,\ line} - \nu_{0,\ datacube}}{\nu_{0,\ datacube}} \cdot c\right) + v_{obs} \tag{3.8}$$

given the rest frequency of the line ($\nu_{0,\ line}$), the rest frequency of the datacube ($\nu_{0,\ datacube}$), the observed velocity of the line (v_{obs}), and the light speed in units of km s^{-1}. From this point of view, as a result of the fitting procedure, all lines show a velocity component (listed in the "Velocity" column in table) consistent with the estimated source systemic velocity (Choi 2001), within the uncertainties dictated by the channel width of 1.2 km/s.

In order to practically fit the identified iCOMs lines, I first attempted an autodetection by `CARTA`, and then proceed by manually adding line components to the fit that were not autodetected. I performed the fit not only on the iCOMs lines I wanted, but also on any neighboring lines of other molecular species, in order to reduce as much as possible any contamination. The fit parameters are: `center` (line peak velocity) (in km s^{-1}), `amplitude` (in K), and `FWHM` (in km s^{-1}). These parameters can be fixed at any moment during the fitting procedure. This is

particularly useful when dealing with weak lines. In my case, it was reasonable to think that all transitions trace the same gas, since I saw no significant deviations in the line profile or the spatial distribution, legitimating fixing some parameters of the Gaussian model associated to weak lines (e.g., speed of the peak emission and FWHM). The same argument applies to bright lines that are (not heavily) contaminated: in some cases I fixed the FWHM of these lines on the basis of that of other bright lines of the same iCOM which were less contaminated. In the very few cases I fixed the emission peak velocity, I checked that accounting for the Doppler effect it was consistent with the source systemic velocity. Clearly, where I fixed FWHM and/or peak emission velocity, there is no error associated to these parameter estimates, as reported for some lines in Tab. 3.8, and Appendix E.

The kind of information coming from line components fitting is in terms of "height" (amplitude), "width" (FWHM), velocity, and "area" (integrated area of emission, W, in units of K · km s^{-1}) of the lines, as well as the uncertainties associated to these parameters. As explained in § 2.4.4, the integrated area of emission (W) is a key piece of information we need to extract from a spectral line, in order to use the Rotational Diagram method to infer the physical properties (T_{rot} and N_{tot}) of the gas producing that line.

In all spectra showing the line fits (like in Fig. 3.12 and 3.13), I overplotted the Gaussian model profile relative to the iCOMs line(s) adapting (3.7) as following:

$$f(v_{spec}) = b + T_B \cdot e^{-\frac{(v_{spec}-v_{obs})^2}{2(FWHM/2\sqrt{2ln(2)})^2}} \tag{3.9}$$

The two velocity arrays, v_{spec} and v_{obs}, respectively refer to the x-axis values of the spectral windows and the observed peak emission velocity of the line, both corrected for the Doppler effect; b is the baseline, which shifts the Gaussian model up or down; T_B is the line profile amplitude, that modulates the height of the peak of the Gaussian; the mean $\overline{x}$ and the standard deviation σ of the distribution are evaluated respectively from the "center" of the line (v_{obs}) and its FWHM.

Together with gaussian models, in the spectra of line fits is also shown in dotted green line the 5σ threshold, given the rms estimated for each datacube. In this way we can have a nice visual confirmation of the robustness of detection of all lines, which are at least at of 10σ significance.

As regarding the baseline, in my case I did not find necessary to fit the baseline with a polynomial of order higher than 0 in any spectral window used to fit iCOMs lines. However, in some of them `CARTA` struggled to individuate the baseline. Since the datacubes are continuum-subtracted, I do not expect the baseline to deviate too much from the "0" of their extracted spectra. Therefore, instead of arbitrarily lowering the baseline, I considered to be more correct to manually set it to 0. In these cases, as visible in some rows of Tab. 3.8 and in Appendix E, the estimated error on the continuum subtraction ($\sigma_{cont.sub.}$ in table) is null.

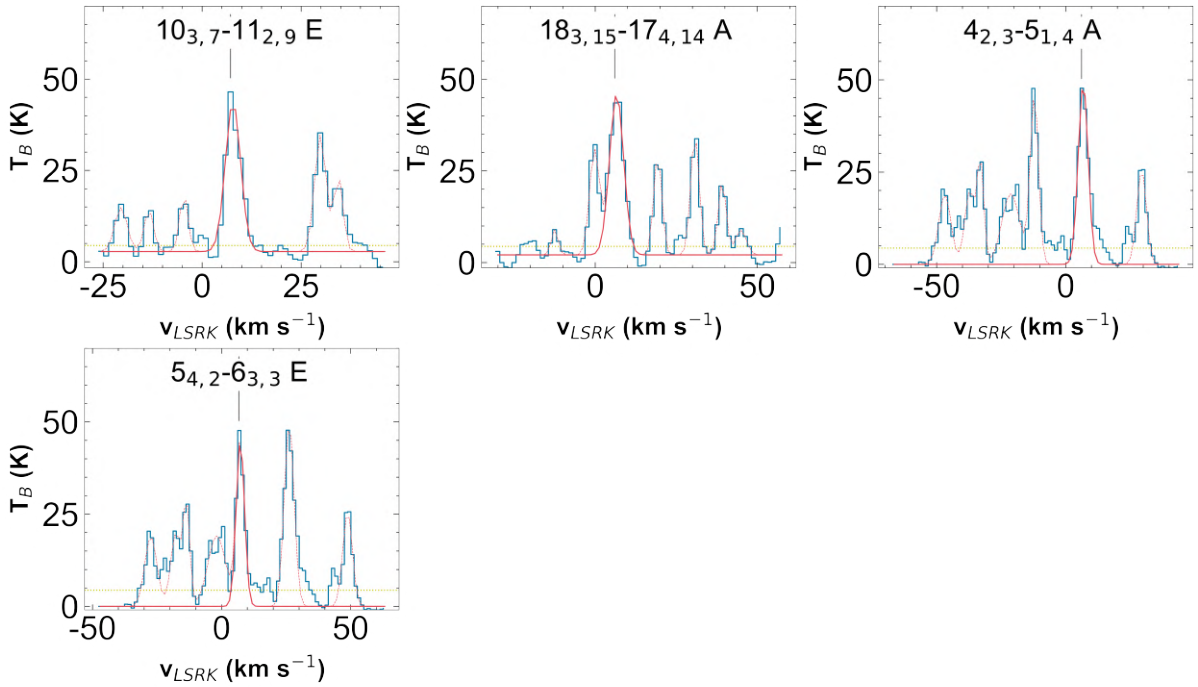

Figure 3.12: Gaussian fits of identified CH_3OH lines in IRAS 4A2 extracted spectra at 234 GHz. The spectral windows are reported in brightness temperature (T_B) as a function of radio velocity in LSRK frame (v_{LSRK}). The spectrum is in blue solid line. In red is plotted the fitting model (dashed line) and the gaussian profile of the fitted lines (solid line). Lines are labelled in black on top of each subfigure. The dotted green line stands for the threshold of 5σ, given σ = 1.16 K.

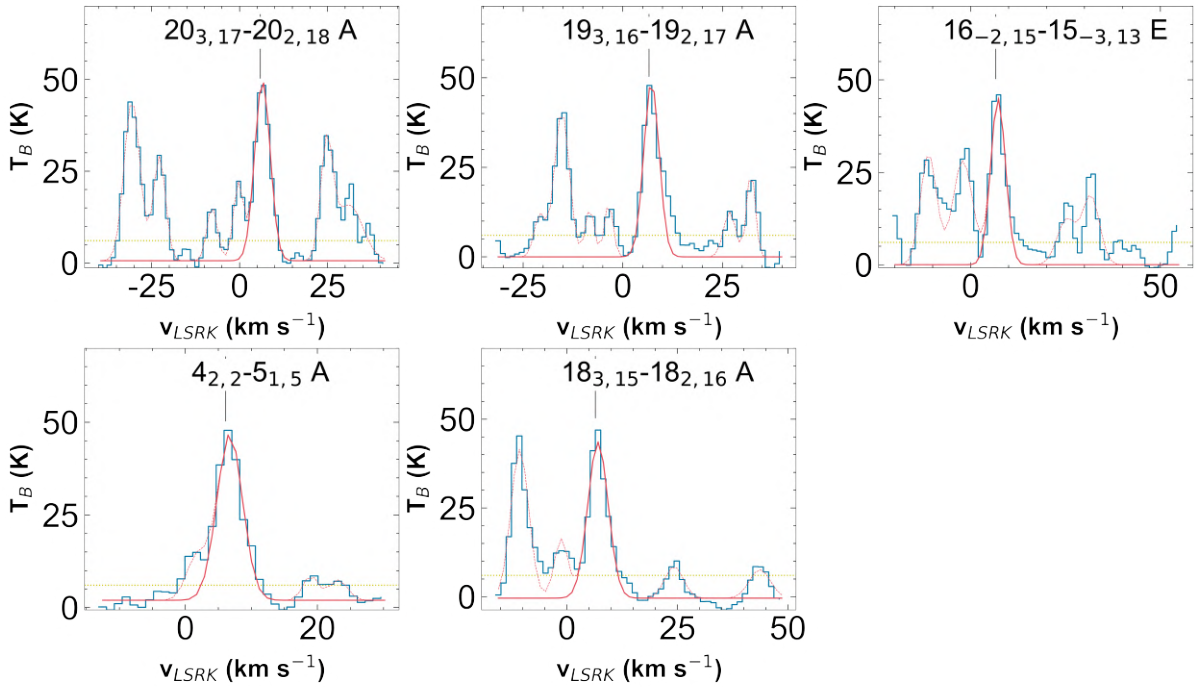

Figure 3.13: Gaussian fits of identified CH_3OH lines in IRAS 4A2 extracted spectra at 245 GHz. The spectral windows are reported in brightness temperature (T_B) as a function of radio velocity in LSRK frame (v_{LSRK}). The spectrum is in blue solid line. In red is plotted the fitting model (dashed line) and the gaussian profile of the fitted lines (solid line). Lines are labelled in black on top of each subfigure. The dotted green line stands for the threshold of 5σ, given σ = 1.23 K.

Table 3.8: Gaussian fit parameters of identified CH_3OH lines in IRAS 4A2 extracted spectra. In round brackets are reported the errors associated to the parameters estimated in CARTA. Rest-frame frequencies (ν_0) and resolved quantum numbers (QNs) are taken from the CDMS catalog. The upper state energy (E_{up}) is evaluated following (3.3).

ν_0 [MHz]	QNs	E_{up} [K]	Area [K · km/s]	Velocity [km/s]	FWHM [km/s]	T_{peak} [K]	$\sigma_{cont.sub.}$ [K]
232945.797	$10_{3,7}$-$11_{2,9}$ E	191	196 (10)	7.19 (0.10)	4.5 (0.3)	41	0.5
233795.666	$18_{3,15}$-$17_{4,14}$ A	447	237 (11)	5.99 (0.10)	5.1 (0.3)	44	0.4
234683.370	$4_{2,3}$-$5_{1,4}$ A	61	237 (11)	6.16 (0.13)	4.5 (-)	49	-
234698.519	$5_{4,2}$-$6_{3,3}$ E	123	183 (15)	6.67 (0.14)	3.8 (0.3)	45	-
246074.605	$20_{3,17}$-$20_{2,18}$ A	537	267 (11)	5.91 (0.08)	5.2 (0.2)	49	0.6
246873.301	$19_{3,16}$-$19_{2,17}$ A	491	236 (11)	6.61 (0.13)	4.5 (-)	49	-
247161.950	$16_{-2,15}$-$15_{-3,13}$ E	338	224 (19)	6.65 (0.19)	4.6 (0.5)	45	-
247228.587	$4_{2,2}$-$5_{1,5}$ A	61	217 (8)	6.12 (0.07)	4.55 (0.19)	45	2.0
247610.918	$18_{3,15}$-$18_{2,16}$ A	447	250 (15)	6.46 (0.13)	5.4 (0.3)	44	0.4

3.4.2 Rotational Diagrams

After measuring the main line parameters, above all, the integrated intensities W, we can finally go back to (2.48):

$$ln\left(\frac{N_u}{g_u}\right) = ln\left(\frac{N_{tot}}{Q(T)}\right) - \frac{E_{up}}{kT}$$

and build the rotational diagrams of each iCOM to extract, under the assumptions of LTE and optically thin lines, their rotational temperature T_{rot} (K) and their total column density N_{tot} (cm^{-2}).

If the assumptions of LTE conditions and optically thin lines are fully satisfied, and if uncertainties can be neglected, the data points in a rotational diagram must follow a perfect straight line. However, in reality the observed data points always come with errorbars that scatter them around the straight line. The main uncertainties propagate from the calibration, the measurements of the integrated lines intensities (W), and from the estimate of the line fluxes. In turn, the line flux uncertainty depends on the formal error we commit with the gaussian fit, which is usually very small, and the error committed on the deconvolution of the flux of the line of interest from the neighbouring lines, which is more difficult to handle. In my case, all diagrams reported in Fig. 3.14 show a degree of scattering along the fit that is less significant than 3σ, meaning the two assumptions of the method are reasonable. Only a couple of transitions of methanol detected at 234 GHz clearly deviate from the fit by more than 3σ. A more refined analysis than the Rotational Diagram approach (e.g. the non-LTE Large Velocity Gradient method, Wilson et al. (2013), or the LTE Population Diagram method, Goldsmith & Langer (1999)), including non-LTE effects and/or optically thick lines, can shed light on the origin of the scattering of molecular lines in RDs.

For formamide (NH_2CHO), the range of excitation energies of the identified lines turned out to be too small to extract meaningful information from the fitting line (see left panel of Fig. 3.15). Given the peculiarity of this molecule, it is analyzed separately in the paragraph below.

The main results (summarized in Tab. 3.9) are a range of temperatures for the IRAS 4A2 hot corino ranging from 157 K (methanol) to 238 K (glycolaldehyde and methyl formate), and a range of column densities of 2.0×10^{17} cm^{-2} for glycolaldehyde up to 7.8×10^{17} cm^{-2} for methanol. Taking into account also the formamide, it is the least abundant molecule in my sample, being the estimated column density range of 1.7 - 2.6×10^{16} cm^{-2}.

Table 3.9: Rotational temperature (T_{rot}) and total column density (N_{tot}) derived for sampled iCOMs from their rotational diagrams, with uncertainties. [a]At fixed temperature, see text.

iCOM	N_{lines}	T_{rot} [K]	N_{tot} [$\times 10^{17}$ cm^{-2}]
CH_3OH	9	157 (9)	7.8 (1.0)
$HCOOCH_3$	24	228 (27)	2.9 (0.4)
CH_3CHO	5	214 (26)	2.4 (0.4)
NH_2CHO[a]	6	238 (47)	1.7 - 2.4
CH_2OHCHO	12	238 (47)	2.0 (0.6)

All column densities are corrected for the beam dilution (discussed in 2.4.4). In particular, since methanol, acetaldehyde and methyl formate appear to be resolved, I used a different filling factor at 234 GHz and 245 GHz according to their estimated deconvolved sizes and the beam of the observations at these frequencies. Glycolaldehyde and formamide, instead, are only partially resolved (because they are associated with very compact emission), hence, at both frequencies

I applied the same filling factor, corresponding to the estimated size of these molecules at 245 GHz (and the beam size), where I achieve the highest angular resolution in the data.

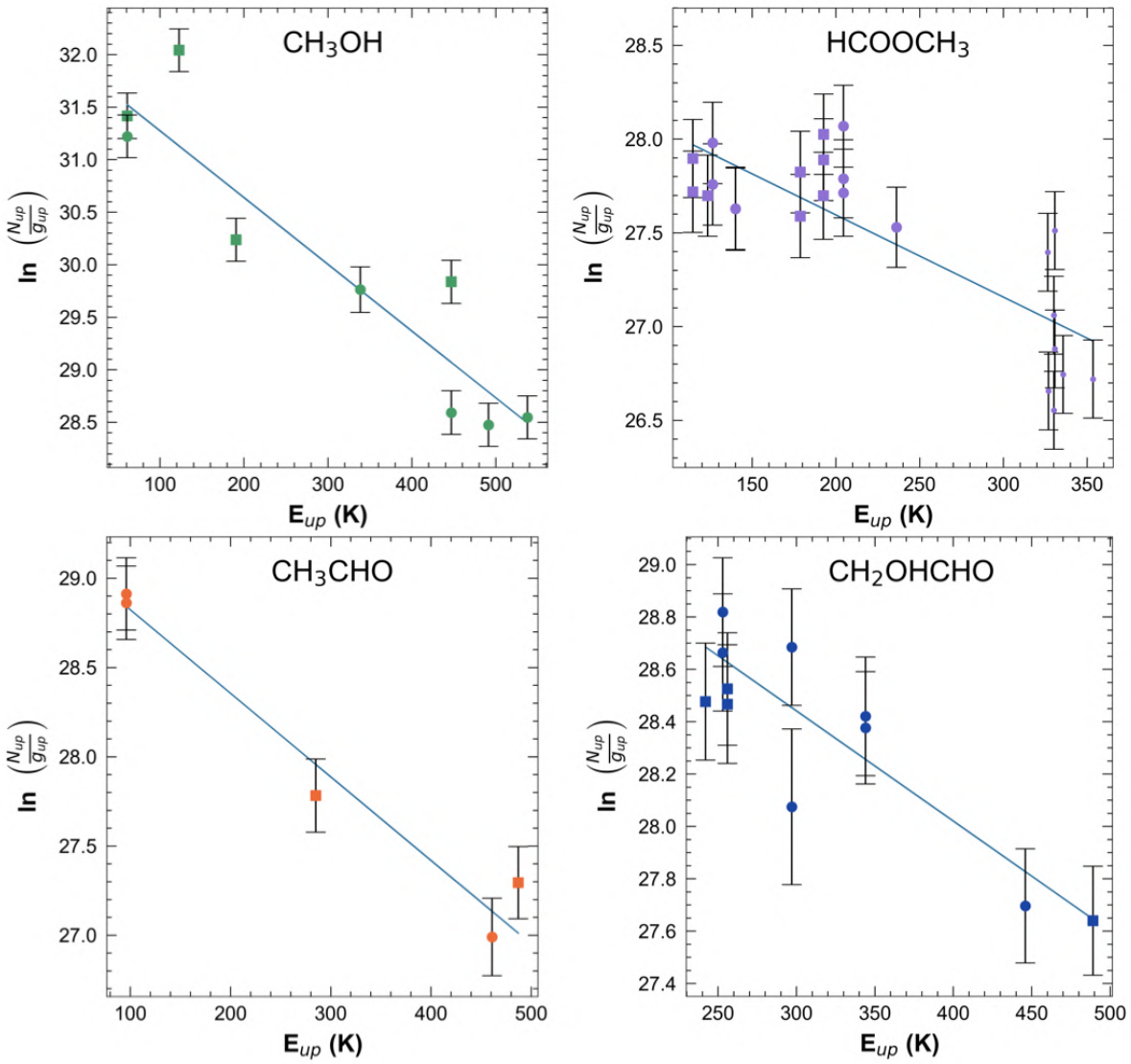

Figure 3.14: Rotational diagrams of methanol, methyl formate (top row), acetaldehyde and glycolaldehyde (bottom row). The data points correspond to transitions identified at 234 GHz *(squares)* and 245 GHz *(circles; smaller markers for vibrational trasitions with v=1)*. The errorbars are evaluated according to (2.52). In blue solid line it is shown the fitting line. The column densities are corrected for beam dilution (see text).

I also compared our results with the values in the literature, specifically, Taquet et al. (2015); López-Sepulcre, A. et al. (2017), De Simone, M. et al. (2017, 2020). In order to perform a meaningful comparison, I rescaled their column densities to account for different assumed sizes (because they were not able to resolve the iCOMs emission) following Goldsmith & Langer (1999):

$$N_{tot,rescaled} = N_{tot} \frac{ff}{ff_{rescaled}} \tag{3.10}$$

N_{tot} stands for the column density reported in their works according to their beam and their estimated/assumed sizes, while $N_{tot,rescaled}$ is the column density I compared with mine, rescaled using the ratio between the "original" filling factor (ff), and the filling factor resulting from fixing our estimated size ($ff_{rescaled}$).

Table 3.10: Column densities derived in the literature rescaled using estimated iCOMs sizes in this work.
[a] refers to the same work, but adopting Population Diagram method (Goldsmith & Langer 1999).

iCOM	Reference	Beam ["]	Size ["]	T_{rot} [K]	N_{tot} [cm^{-2}]	$N_{tot,rescaled}$ [cm^{-2}]
CH_3OH	Taquet+15	$\sim$ 2 (PdBi)	0.5	300 (151)	6.3 (3.1) $\times$ 10^{17}	1.9 $\times$ 10^{18}
	Taquet+15[a]		0.2	140 (-30, 30)	1.6 (-0.8, 0.6) $\times$ 10^{19}	8.0 $\times$ 10^{18}
	De Simone+20	$\sim$ 1 (VLA)	0.24	120 - 190	1.0 $\times$ 10^{19}	6.3 $\times$ 10^{18}
$HCOOCH_3$	Taquet+15	$\sim$ 2 (PdBi)	0.5	141 (90)	5.2 (3.3) $\times$ 10^{16}	1.6 $\times$ 10^{17}
	Taquet+15[a]		0.2	60 (30)	5.0 (-1.8, 5.0) $\times$ 10^{17}	2.3 $\times$ 10^{17}
	Sepulcre+17	0.68 x 0.37	0.3	196 (-35, 15)	3.5 (-0.4, 0.2) $\times$ 10^{16}	$\sim$ 3.5 $\times$ 10^{16}
CH_3CHO	Sepulcre+17	0.66 x 0.35 (ALMA)	0.3	100 - 200	1.0 - 1.9 $\times$ 10^{16}	1.5 - 2.8 $\times$ 10^{16}
NH_2CHO	Taquet+15	$\sim$ 2 (PdBi)	0.5	100 - 300	2.1 - 8.5 $\times$ 10^{15}	2.1 - 8.1 $\times$ 10^{16}
	Sepulcre+17	0.69 x 0.37 (ALMA)	0.3	100 - 200	1-2 $\times$ 10^{15}	2.1 - 4.2 $\times$ 10^{15}
CH_2OHCHO	De Simone+17	0.70 x 0.42 (CALIPSO/IRAM)	0.4	236 (74)	88 (70) $\times$ 10^{14}	2.6 $\times$ 10^{16}
	Sepulcre+17	0.65 x 0.35 (ALMA)	0.3	100 - 200	5.8 - 6.8 $\times$ 10^{15}	1.4 - 1.4 $\times$ 10^{16}

In comparison, as summarized in Tab. 3.10, I derived column densities consistent within the errors for glycolaldehyde and acetaldehyde, and given that several caveats are hidden in these estimates. As regarding methanol and methyl formate, instead, considering the impact of line optical depth in LTE conditions (Taquet et al. 2015), and non-LTE conditions at cm wavelengths (De Simone, M. et al. 2020), the rescaled column densities should be larger than the ones I obtained in LTE conditions and assuming optically thin lines. Investigate the effect of line opacity on the observed lines of these iCOMs could shed light on the observed difference between rescaled column densities from previous works and mine.

The Case Of Formamide

As visible in the left panel of Fig. 3.15, fitting the data points relative to the identified transitions of this iCOM leads to too high uncertainty in the determination of the rotational temperature and the column density. In order to get at least a limit on its abundance, I supposed that formamide probes gas at the same temperature of glycolaldehyde (238 $\pm$ 41 K), because these two iCOMs are associated with a similar emitting scale with radius of the order of $\sim$ 20 AU (see § 3.3.3). Going back to (2.48), we can see that fixing the temperature means fixing the slope (b) of the fitting line. Therefore, following the least squares method exposed in § 2.4.4, the derivative of the χ^2 variable wrt b will be equal to 0 by default, and we will need to solve just for $\delta\chi^2/\delta a = 0$, which traduces in:

$$-2\sum_{i=1}^{N}(y_i - a - bx_i) = 0$$

Expanding this sum on the data and dividing by the (generic) number of measurements N, we find:

$$a = \overline{y} - b\overline{x} \tag{3.11}$$

For the uncertainty on the intercept, assuming that the errors on x are negligible with respect to the errors on y, for the propagation of errors (Taylor 1996):

$$\Delta a = \Delta\overline{y} + \Delta b \tag{3.12}$$

in which Δb is known (propagating the error on the temperature, Δb = (ΔT_{rot} / T_{rot}) $\cdot$ b), and $\Delta\overline{y}$ is simply the standard deviation of the y-values (remember that y = $ln\left(\frac{N_{up}}{g_{up}}\right)$) divided by the square root of N.

The resulting rotational diagram at fixed temperature is plotted in the right panel of Fig. 3.14. The column density ranges from 1.7 (at 191 K) to 2.6 (at 285 K) $\times$ 10^{16} cm^{-2}. Compared to the literature, the rescaled column density from López-Sepulcre, A. et al. (2017) would be 4 $\times$ 10^{15} cm^{-2} (but according to the upper limit of their assumed 100-200 K temperature range), while rescaling the one found by Taquet et al. (2015) (at the upper limit of their assumed 100-300 K range), I get 8.1 $\times$ 10^{16} cm^{-2}. This difference could be explained, apart from the eventual impact of non-LTE effects and line opacity, by taking into account uncertainties and caveats in these measurements.

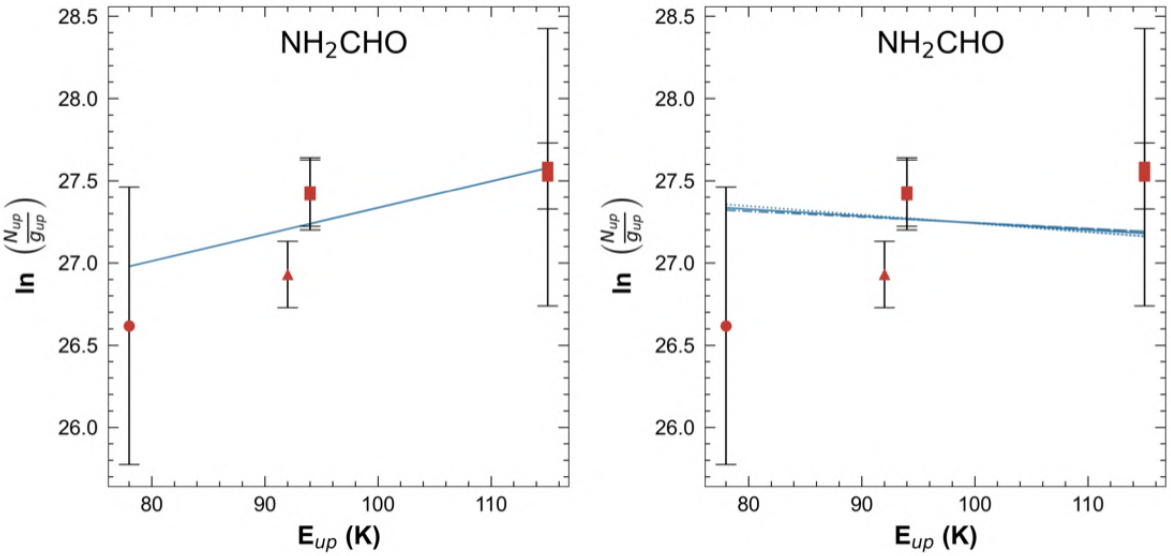

Figure 3.15: Rotational diagram of formamide. *(left)* No fixed temperature. The data points correspond to transitions identified at 234 GHz *(squares)*, 245 GHz *(circle)*, and the identified line in the narrow SPW centred at 260 GHz *(triangle)*. *(Right)* The same data points are fitted at fixed temperature within errorbar (238 ± 47 K) in solid (T_{rot} = 238 K), dotted (T_{rot} = 191 K), and dash-dotted lines (T_{rot} = 285 K). In both figures, the errorbars are evaluated according to (2.52), and the column densities are corrected for beam dilution (see text).

3.5 IRAS 4A2 Temperature Profile

One of the current hot topics in star formation is about the structure and dynamics of protostellar envelopes (refer to Chapter 1). In order to address these issues, the physical parameters within the envelope, mainly, the density and temperature profiles and the velocity field, are needed (Li et al. 2014). Up to now, several efforts have been made in the direction of modelling protostellar envelopes (Shu et al. 1987; Ceccarelli et al. 1996; Motte, F. & André, P. 2001), both in large surveys (Beckwith et al. 1990; Jørgensen et al. 2002; Andrews & Williams 2007; Crimier, N. et al. 2010) and for individual Class 0/I protostars (Ceccarelli et al. 2000a; Schöier et al. 2002; Maret et al. 2002; Jacobsen, S. K. et al. 2018; van 't Hoff et al. 2018; van ´t Hoff, Merel L. R. et al. 2020). However, the theoretical modelling does not go hand in hand with observational evidences. In fact, sensitive, high angular resolution maps of the circumstellar environment of YSOs, resulting from observations of large (sub)millimeter radiotelescopes, can potentially provide crucial tests of our theoretical ideas on star formation (Jørgensen et al. 2002; van ´t Hoff, Merel L. R. et al. 2020).

Exploiting the ALMA/FAUST mm data at scales $<$ 50 AU, with this work I can characterize the gas temperature in the inner envelope of Class 0 protostar at scales never explored before with direct observations, which in my case consist of emission lines of key prebiotic molecules, my sample of iCOMs.

Fig. 3.16 and 3.17 represent the final result of this thesis project, and they are built upon all results obtained throughout this chapter. Out of the five iCOMs in my sample, it was not possible to include formamide (NH_2CHO), whose rotational diagram has been derived fixing in input the rotational temperature. For the other molecules, I derived independently the rotational temperature, approximating the hot corino gas temperature, and put it as a function of their mean emitting radii, given the observed spatial segregation. The temperatures with errorbars are exactly the ones listed in Tab. 3.9. For the radii and their uncertainties (both evaluated in AU), instead, I considered a weighted average (Taylor 1996) of the radii estimated for each iCOM with `imfit` at all frequencies (the ranges are listed in Tab. 3.7).

In this regard, I obtained mean sizes of: 0.3" (methanol); 0.28" (methyl formate); 0.2" (acetaldehyde); 0.16" (formamide); 0.14" (glycolaldehyde). These estimates are the ones I took into account to evaluate the rescaled filling factor $ff_{rescaled}$ in (3.10).

Fig. 3.16 shows the gas temperature versus radius in log-log scale in basis 10. Indeed, plotting a function like $f(x) = log_{10}\, x$, for the properties of logarithms: $log_{10}\, x = ln(x)/ln(10)$. Therefore, the error on the logarithms of temperature and radius is like, given a generic variable v:

$$\Delta v = 1/ln(10)\ (\Delta v/v)$$

The log scale is justified by the fact that the gas temperature is likely well-coupled with the dust temperature (Ceccarelli et al. 1996; Maret et al. 2002; Crimier, N. et al. 2010), and because the temperature is expected to follow a power-law drop with the radial distance from the central protostar (Beckwith et al. 1990; Motte, F. & André, P. 2001; Andrews & Williams 2007), that is, $log_{10}T \propto \beta log_{10}r$, defined β as the exponent of this T-r power-law relation.

Fitting my T-r datapoints with a straight line, and linearizing the logarithmic functional relation between the two variables, I finally get:

$$T(r) = r^{\,-0.7(0.3)}\ 10^{\,3.4(0.4)} \tag{3.13}$$

being the temperature in K and the radius in AU.

Thus, on the basis of observed iCOMs emission lines, and at scales smaller than 50 AU, if the emitting radius is now intended as radial distance from IRAS 4A2, the temperature in its hot corino region falls with the radial distance as a single power-law with exponent $\beta \approx$ -0.7. It should be stressed that no model or any kind of assumption on the geometry of the hot corino

have been made in order to retrieve this relation. However, methyl formate ($HCOOCH_3$, violet square in figure) stands out as outlier of this power-law behaviour, apparently tracing a very hot gas ($\sim$ 228 K) on a very large scale ($\sim$ 37 AU). Given the uniqueness of this gas temperature profile, there is no straightforward answer to explain it. In this *T-r* functional relation, the variable that could play a major role in producing the "offset" observed with that iCOM is the rotational temperature, whose estimate heavily depends on the Rotational Diagram method assumptions. In this context, in fact, the uncertainties on the emitting radius estimates are rather "intrinsic" of the performed observations, given the methodology of the gaussian fitting. Therefore, it could be that the LTE and/or optically thin lines assumptions implied in the RDs are too simple. A more refined molecular lines analysis technique, such as the Population Diagram method (Goldsmith & Langer 1999) essentially drops the assumption of optical thinness of molecular lines. Therefore, it could help understanding whether methyl formate is tracing colder gas than what I measure, because its low excitation (low E_{up}) emission lines are saturated ($\tau >> 1$). Saturated lines would be characterized by a larger profile broadening than normal and truncated profiles, thus biased integrated intensities (W), ultimately changing the rotational temperature estimate for the parent molecule in its rotational diagram. Alternatively, if this iCOM is still optically thin (and in LTE) and actually traces hotter gas than methanol, but at almost the same outer spatial scale ($\sim$ 40 AU), this finding would leave room for interesting investigation on the formation route of this molecule and the physical structure of the hot corino region.

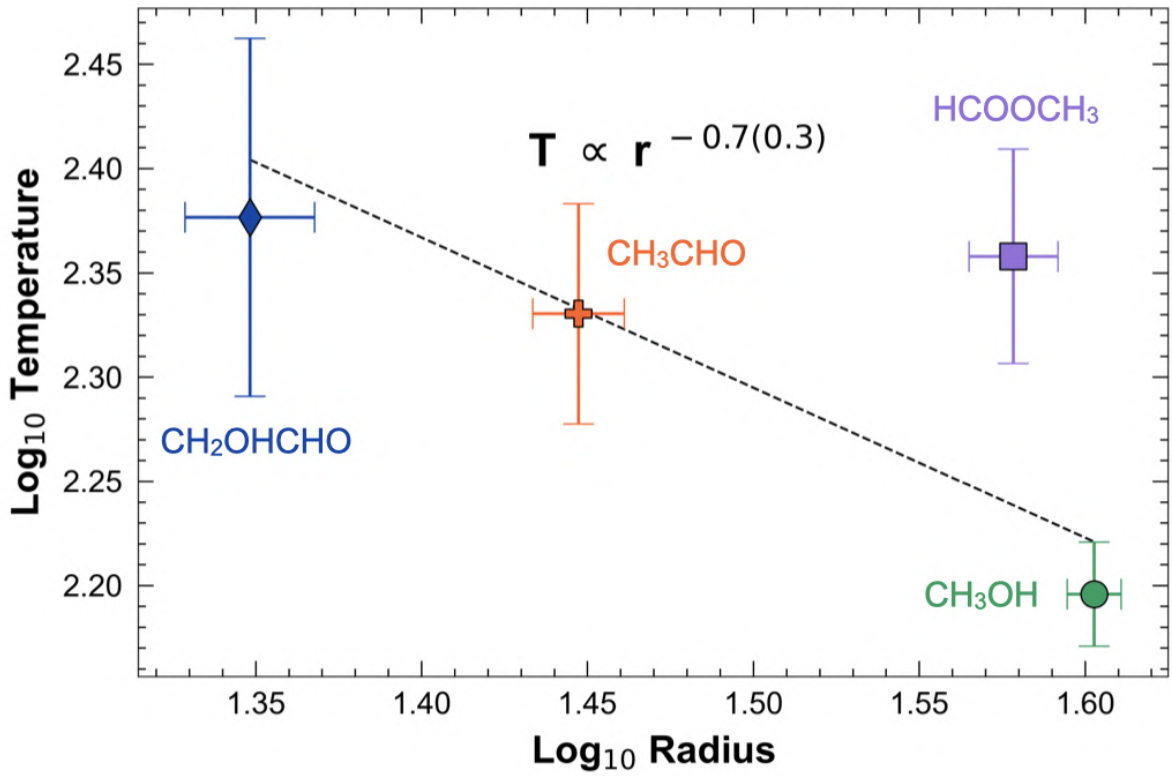

Figure 3.16: Trend of gas temperature as a function of iCOMs emitting size radius in log-log scale. The datapoints with errorbars correspond to the pair of rotational temperature and mean emitting radius values as estimated respectively from the Rotational Diagram method and image plane fitting for all sampled iCOMs except for formamide (see text). The black dotted line indicates the power-law relation between gas temperature and radius obtained by linearly fitting the datapoints, and labelled in bold in the plot.

How does this result relate with the literature? I addressed this final question by comparing in Fig. 3.17 my temperature profile in linear scale with temperature profiles estimated in previous works. Specifically, with the one derived for the IRAS 4A system (Maret et al. 2002), Cep E-mm intermediate mass Class 0 protostar (Crimier, N. et al. 2010), L1527 Class 0/I protostar (van 't Hoff et al. 2018), and the IRAS 16293A Class 0 protostar (van ´t Hoff, Merel L. R. et al. 2020).

Differently from this work, all these profiles are models that have been compared with observations and extrapolated at small scales down to the beam size of their observations (or even below, as in Maret et al. (2002)), in any case larger than the ALMA/FAUST beam for IRAS 4A2.

Therefore, comparing the power-law exponent I found with the underlying *T*-*r* theoretical scaling relation of these models, either valid for a collapsing envelope or a protostellar environment with a disk component, I find that my profile is consistent with the scenario of a disk-like structure, having a "fast" scaling relation with $\beta \approx -$ (0.4 - 0.75; Andrews & Williams (2005); Jacobsen, S. K. et al. (2018); van ´t Hoff, Merel L. R. et al. (2020). This is expected, since, thanks to the ALMA/FAUST resolution, I can resolve down to the typical dimensions of disks. A collapsing envelope, on the other hand, is expected to possess a "slower" *T*-*r* scaling relation, with $\beta \approx -$ (0.4 - 0.5); (Ceccarelli et al. 2000a; Schöier et al. 2002; Crimier, N. et al. 2010).

Given the bolometric luminosity of IRAS 4A ($\sim$ 9 $L_\odot$; there is no current estimate of IRAS 4A1 and IRAS 4A2 individual luminosities (Kristensen, L. E. et al. 2012; Karska, A. et al. 2013), and that disk temperature should increase with the bolometric luminosity (Beckwith et al. 1990), we actually do not expect in principle to measure such an higher mean temperature compared to the other profiles. For example, in the case of IRAS 16293A, the total system is known to have a bolometric luminosity $\sim$ 21 $L_\odot$ (Jørgensen et al. 2016), yet, van ´t Hoff, Merel L. R. et al. (2020) extrapolates a lower temperature than my profile. Cep E-mm, that is definitely more luminous than IRAS 4A ($\sim$ 78 - 100 $L_\odot$; Palla et al. (1993); Lefloch, B. et al. (1996); Froebrich et al. (2003); Noriega-Crespo et al. (2004)), displays as well a not usual small discrepancy in temperature with my profile. A possible explanation is that, as it is being recently shown by advanced numerical simulations (Lebreuilly et al., in prep.), that protostellar disks' temperatures are generally underestimated.

Nevertheless, my result represents the first pure observational constraint on the gas temperature structure in a young protostellar environment at compact scales. A confrontation with more accurate models and simulations, as well as other high-resolution observations at disk scales are needed to better constrain the issue of temperature structure of YSOs, as well as other protostellar physical properties especially at the most compact scales.

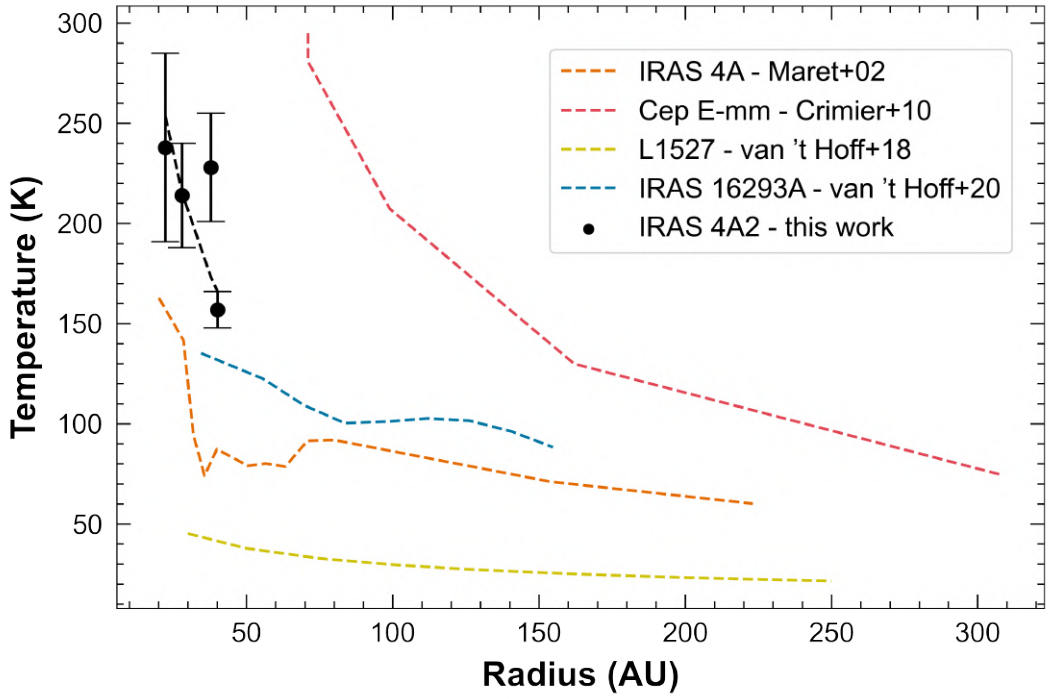

Figure 3.17: Comparison of protostellar gas temperature profiles. The observational profile derived in this work is plotted with black filled circles and errorbars on temperature, plus the derived power-law with $\beta \approx 0.7$ (black dotted line). The dotted lines of different colors reproduce on scales between beam/model starting point up to 300 AU the temperature profiles as derived in the literature for IRAS 4A (Maret et al. 2002), Cep E-mm (Crimier, N. et al. 2010), L1527 (van 't Hoff et al. 2018), and IRAS 16293A (van ´t Hoff, Merel L. R. et al. 2020).

CHAPTER 4

Conclusions

It is time to sum up. This thesis project ultimately aimed to characterize the chemical richness in the inner envelope of a young Solar-like star, at scales where planets can form, and link the observed chemical characterization to the physical structure of the protostar at the same scales.

In this regard, I asked myself three questions: *(i) are different complex organic molecules tracing different scales in the planet-forming region around a young protostar? (ii) If so, are the probed scales associated with a uniform or varying gas temperature? (iii) Eventually, can we infer a gas temperature profile at planet-forming scales from the observation of these molecules?*

In order to answer these questions, I have studied the well known IRAS 4A2 hot corino using (sub-)mm interferometric observations of the state-of-the-art ALMA observatory, pushed at resolving scales < 50 AU in the context of the Large Program FAUST. Having in hand such an high angular ($< 0''.2$) and spectral (~ 1.2 km s^{-1}) resolution dataset, when I looked at the extremely line-rich spectra of the source dataset, it was clear that I could not attempt to identify all possible molecules. I therefore focused on a sample of 5 common (in Solar-like star forming regions), abundant, and bright complex organic molecules, which are all claimed to be crucial building blocks of prebiotic compounds of life on the Earth. My subsequent analysis split in two, to then reunite in the final step to derive a last, innovative result.

On the one hand, I imaged and compared the integrated emission of selected transitions detected for my sample of molecules that had similar excitation energy. In this way, comparing emission patterns likely excited at the same temperature, there was little doubt about what I was seeing: different molecules emit from different scales in the hot corino region. Yet, as science teaches, I needed to make what I saw objective and reproducible, and a Gaussian fit performed on image and Fourier planes confirmed the previous qualitative result. Not only are the molecules spatially segregated, but they are confined to a spatial scale totally comparable to that of the Solar System within the Uranus's orbit and the inner edge of the Kuiper Belt ($\sim$19 - 43 AU).

On the other hand, I derived two new constrains on the gas kinetic temperature and the column densities as probed by my sample of iCOMs in IRAS 4A2, by performing the well known molecular line analysis technique based on Rotational Diagrams. From this point of view, my observations, compared to the literature, suggest that non-LTE but especially line opacity effects could play an important role in the hot corino gas, and that can be investigated with more refined molecular lines analysis.

Thirdly, I joined together these two pieces of information, the iCOMs spatial segregation on one hand, and the gas temperature probed by each molecule on the other hand, to outline the first direct and model independent observational temperature profile of a Class 0 protostar at planet-forming scales < 50 AU, and established agreement with theoretical expectations for disk-like protostellar temperature structures, despite the exceptional behaviour of methyl formate ($HCOOCH_3$), which must be investigated.

In conclusion, to answer the three questions posed at the outset: *(i) some of the most common and crucial interstellar complex organic molecules likely reside in different regions of the*

inner envelope of a Class O Solar-like protostar, corresponding to scales where a planetary system like our own can form; (ii) the scales of emission probed by these molecules are associated with a different (approximate) gas temperature that is tendentially higher for bigger molecules at more compact scales; (iii) it is possible to derive the gas temperature profile only based on the observed molecular transitions in the planet-forming region.

4.1 Future Perspectives

With this thesis project, I resolved for the first time the emission coming from the hot corino prototype IRAS 4A2. Adding to that my observational constrain on its temperature structure in the planet-forming region, I provided two dowels to settle our understanding of hot corinos, the earliest furnaces of complex organic molecules, and of the physical structure of the youngest protostars.

Yet, it can be said that I have only scratched the surface. For example, little can be inferred from the temperature profile I derived regarding the formation route of my sampled iCOMs. Moreover, claiming the presence of prebiotic precursors in the planet-forming region at the earliest stage of a Solar-like protostar, does not imply iCOMs will actually be preserved and inherited by the forming planets at the same scales.

Forward-thinking, some interesting avenues for future studies come to mind, that I will list from "smallest to largest":

1. *Characterize and compare 4A1 and 4A2 binary components of the IRAS 4A system from a chemical, kinematical, and spatial perspective, coupling mm FAUST data of IRAS 4A2 with other mm and cm observations of the binary system.* IRAS 4A2 is the less massive, lower (mm) luminosity binary companion of IRAS 4A1. FAUST observations, dedicated to disentangle and chemically characterize the envelope/disk system of Class 0/I YSOs like IRAS 4A2, are limited in angular resolution and sensitivity by three setups and $\sim$ 1.3 mm observations only. A good idea may be to couple this dataset with other mm and cm high angular and sensitivity observations of the IRAS 4A1 companion as well, coming from e.g. VLA and IRAM/NOEMA[1], unveiling its nature of hot corino (De Simone, M. et al. 2020). In this way, I could perform a comparative analysis of the chemical and temperature profiles of the two IRAS 4A protostars, aiming to understand whether they harbor significantly different initial conditions for planet formation, and shed light on the formation of their parent protostellar system.

2. *Perform a systematic comparison of temperature profiles among different Class 0/I protostars down to planet-forming scales using mm observations of iCOMs lines, accurate theoretical models and numerical simulations.* The star and planet formation process consists of a complex and non-stop interplay between physical and chemical evolution. The chemical characterization of protostellar environments must be therefore complemented with a solid understanding of their physical structure. Class 0/I protostars are particularly problematic because their young disks embedded in dust are difficult to observe, especially at the smallest scales, and are shaped by non-linear processes e.g. fragmentation, turbulence, outflows, and magnetic fields. On the most basic level, I could populate the observed temperature profile of IRAS 4A2 with a larger sample of iCOMs, to further verify the expected *T-r* scaling law and look if there is any other "exception" to this relation, as represented by methyl formate, and if so, why. On a higher level of investigation, starting with the other targets in the FAUST archive, and potentially implementing it with other (sub)mm high angular resolution observations, taken with e.g. IRAM/NOEMA and ALMA, I may try

[1] https://iram-institute.org/science-portal/noema/

to derive observational temperature profiles of these targets, similarly to the one I derived for IRAS 4A2, and compare them also with the state-of-the-art models and simulations. I could also attempt to extend the same analysis to other crucial physical properties, having observable chemical counterparts, such as envelope densities and disk masses.

3. *Explore the chemical link between a representative sample of Class 0/I protostars at mm and cm wavelenghts and of protoplanetary disks in the IR.* While we are now able to characterize the chemistry of the youngest protostars at planet-forming scales, we lack a characterization of the more evolved stages, namely the protoplanetary disks around Class II/III stars. Hence, we do not know if and how iCOMs survive along the star formation process and how they could be acquired by the forming planets in protoplanetary disks. We could therefore plan, on the one hand, to characterize the chemistry of protostellar environments using FAUST, ALMA, VLA, and IRAM/NOEMA mm and cm archival/new data. On the other hand, study the properties of protoplanets and disks substructures at infrared wavelengths through observations obtained with VLT-ERIS[2], VLT-SPHERE[3], and JWST[4]. This kind of research project would also act as the perfect ground for the preparation of the science cases for the next-generation, cutting-edge facilities, the *Extremely Large Telescope* (ELT)[5], and the *Square Kilometre Array* (SKA)[6]. ELT observations at infrared wavelengths will allow us to directly detect young giant protoplanets and their atmospheres, as well as probing the inner few astronomical units of protoplanetary discs in the nearest star-forming regions ($\simeq$ 500 ly). SKA, on its turn, will enable us to search for larger, more complex iCOMs species at cm wavelengths, where spectral line blending and line confusion are reduced with respect to the mm/sub-mm range.

The road to unveil the mysteries of hot corinos, protostars, planet formation, and ultimately of the life in the Universe is still long ahead of us, but we can enjoy the ride and the panorama, and I am thrilled to be on board.

[2] https://www.eso.org/public/teles-instr/paranal-observatory/vlt/vlt-instr/eris/
[3] https://www.eso.org/public/teles-instr/paranal-observatory/vlt/vlt-instr/sphere/
[4] https://webb.nasa.gov
[5] https://elt.eso.org
[6] https://www.skao.int

APPENDIX A

Identified iCOMs Lines

Table A.1: Spectral properties of methyl formate lines identified in IRAS 4A2 extracted spectra.

$HCOOCH_3$ - N_{lines} = 24, JPL catalog (PICKETT et al. 1998)					
ν_0 [MHz]	QNs	$S_{ij}\mu^2$ [D^2]	$Log_{10}(A_{ij})$	E_{up} [K]	g_{up}
233854.286	$19_{11,9}$-$18_{11,8}$ A	67.42	-3.891	193	78
233867.193	$19_{11,9}$-$18_{11,8}$ E	67.41	-3.891	193	78
233845.233	$19_{11,8}$-$18_{11,7}$ E	67.41	-3.891	193	78
233753.960	$18_{4,14}$-$17_{4,13}$ E	91.55	-3.735	114	74
233777.521	$18_{4,14}$-$17_{4,13}$ A	91.55	-3.735	114	74
233226.788	$19_{4,16}$-$18_{4,15}$ A	96.11	-3.740	123	78
234124.883	$19_{10,10}$-$18_{10,9}$ A	73.29	-3.853	179	78
234134.600	$19_{10,10}$-$18_{10,9}$ E	73.29	-3.853	179	78
245903.680	$20_{13,8}$-$19_{13,7}$ E	61.69	-3.885	236	82
246285.400	$20_{11,9}$-$19_{11,8}$ E	74.46	-3.802	204	82
246295.135	$20_{11,10}$-$19_{11,9}$ A	74.47	-3.802	204	82
246308.272	$20_{11,10}$-$19_{11,9}$ E	74.46	-3.802	204	82
246891.611	$19_{4,15}$-$18_{4,14}$ E	97.08	-3.661	126	78
246914.658	$19_{4,15}$-$18_{4,14}$ A	97.09	-3.661	126	78
247044.146	$21_{3,19}$-$20_{3,18}$ E	108.10	-3.656	140	86
247053.453	$21_{3,19}$-$20_{3,18}$ A	108.11	-3.656	140	86
		v = 1			
245846.914	$21_{3,19}$-$20_{3,18}$ E	108.41	-3.662	327	86
246184.177	$20_{8,13}$-$19_{8,12}$ E	89.75	-3.721	354	82
246187.016	$21_{2,19}$-$20_{2,18}$ A	108.07	-3.661	327	86
246461.167	$22_{2,21}$-$21_{2,20}$ A	114.70	-3.654	331	90
246488.433	$22_{1,21}$-$21_{1,20}$ A	114.70	-3.653	331	90
246706.504	$22_{2,21}$-$21_{2,20}$ E	115.11	-3.651	330	90
246731.729	$22_{1,21}$-$21_{1,20}$ E	115.11	-3.651	330	90
246985.225	$20_{6,15}$-$19_{6,14}$ A	96.84	-3.684	336	82

Table A.2: Spectral properties of acetaldehyde lines identified in IRAS 4A2 extracted spectra.

CH_3CHO - N_{lines} = 5, JPL catalog (PICKETT et al. 1998)					
ν_0 [MHz]	QNs	$S_{ij}\mu^2$ [D^2]	$Log_{10}(A_{ij})$	E_{up} [K]	g_{up}
233048.516	$12_{2,11}$-$11_{2,10vt=1}$ E	141.06	-3.381	285	50
234707.132	$12_{4,8}$-$11_{4,7,vt=2}$ E	128.54	-3.412	487	50
247142.155	$14_{0,14}$-$13_{1,13}$ A	21.91	-4.178	96	58
247341.332	$14_{0,14}$-$13_{1,13}$ E	21.83	-4.179	96	58
246169.217	$13_{0,13}$-$12_{0,12,vt=2}$ E	159.72	-3.289	461	54

Table A.3: Spectral properties of formamide lines identified in IRAS 4A2 extracted spectra.

NH_2CHO - N_{lines} = 6, JPL catalog (PICKETT et al. 1998)					
ν_0 [MHz]	QNs	$S_{ij}\mu^2$ [D^2]	$Log_{10}(A_{ij})$	E_{up} [K]	g_{up}
233735.603	$11_{4,8}$-$10_{4,7}$	374.56	-3.093	115	69
233746.504	$11_{4,7}$-$10_{4,6}$	374.52	-3.093	115	69
233897.318	$11_{3,9}$-$10_{3,8}$	399.43	-3.064	94	69
234316.254	$11_{3,8}$-$10_{3,7}$	399.43	-3.064	94	69
247391.356	$12_{0,12}$-$11_{0,11}$	469.13	-2.958	78	75
260189.848	$12_{2,10}$-$11_{2,9}$	457.77	-2.903	92	75

Table A.4: Spectral properties of glycolaldehyde lines identified in IRAS 4A2 extracted spectra.

CH_2OHCHO - N_{lines} = 12, CDMS catalog (PICKETT et al. 1998)					
ν_0 [MHz]	QNs	$S_{ij}\mu^2$ [D^2]	$Log_{10}(A_{ij})$	E_{up} [K]	g_{up}
233587.449	$26_{10,16}$-$26_{9,17}$	73.42	-3.687	256	53
233709.657	$26_{10,17}$-$26_{9,18}$	73.41	-3.687	256	53
234554.566	$38_{11,27}$-$38_{10,28}$	123.01	-3.620	489	77
234704.065	$25_{10,15}$-$25_{9,16}$	69.27	-3.690	242	51
246395.021	$34_{4,30}$-$34_{3,31}$	63.54	-3.795	344	69
246605.634	$34_{5,30}$-$34_{4,31}$	63.52	-3.794	344	69
246687.251	$35_{11,24}$-$35_{10,25}$	108.29	-3.574	426	71
246773.215	$30_{2,28}$-$30_{1,29}$	31.76	-4.041	253	61
246778.410	$30_{3,28}$-$30_{2,29}$	31.76	-4.041	253	61
246852.728	$36_{11,26}$-$36_{10,27}$	112.77	-3.568	446	73
247285.434	$32_{3,29}$-$32_{2,30}$	47.60	-3.890	297	65
247323.158	$32_{4,29}$-$32_{3,30}$	47.60	-3.890	297	65

APPENDIX B

Moment 0 Maps Comparison

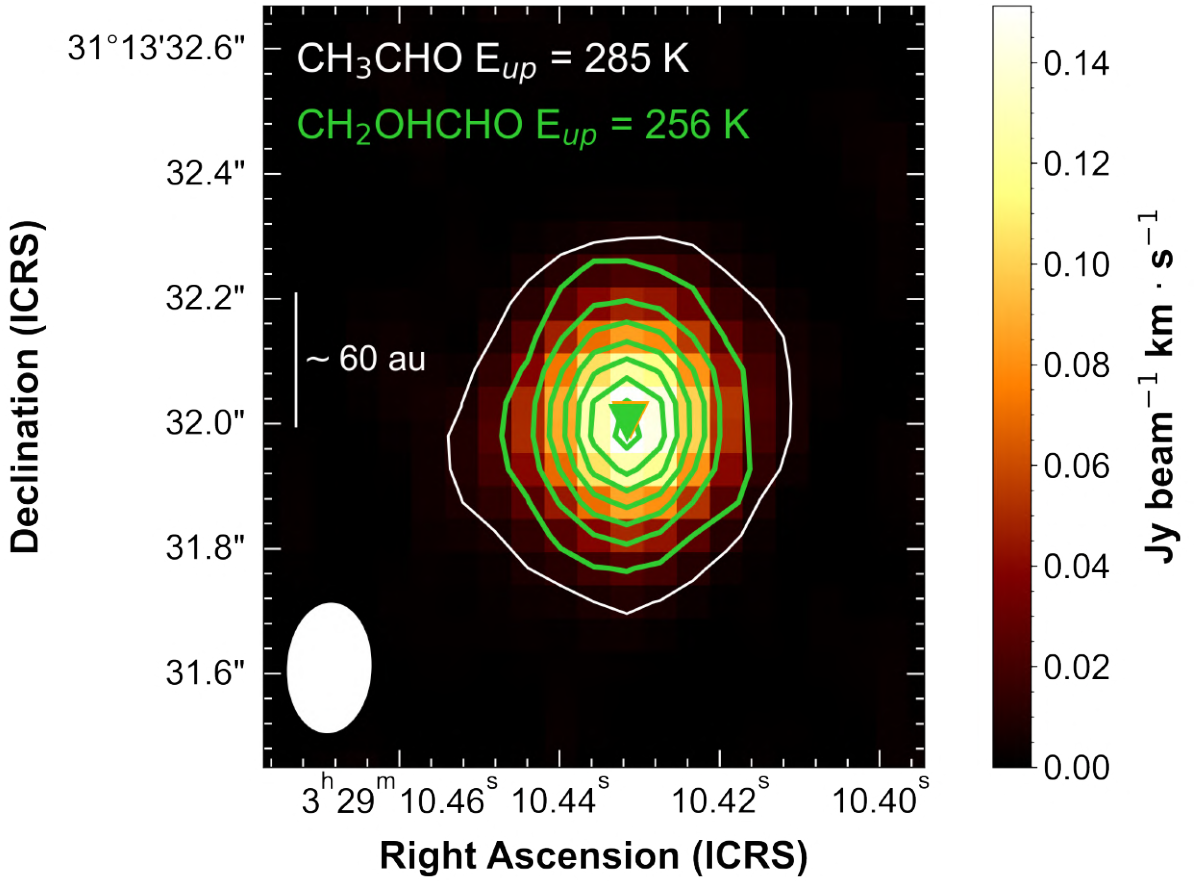

Figure B.1: Comparison of integrated emission of acetaldehyde (CH_3CHO) and glycolaldehyde (CH_2OHCHO) at 234 GHz. Acetaldehyde emission is shown in colour scale plus a 3σ white contour of 8 mJy beam^{-1} km s^{-1}. Glycolaldehyde emission is in 5σ steps green contours starting from $3\sigma \sim 6$ mJy beam^{-1} km s^{-1}. The triangles in orange and green indicate the emission peaks of acetaldehyde and glycolaldehyde. The restoring beam is 0.21" x 0.14", −3°.

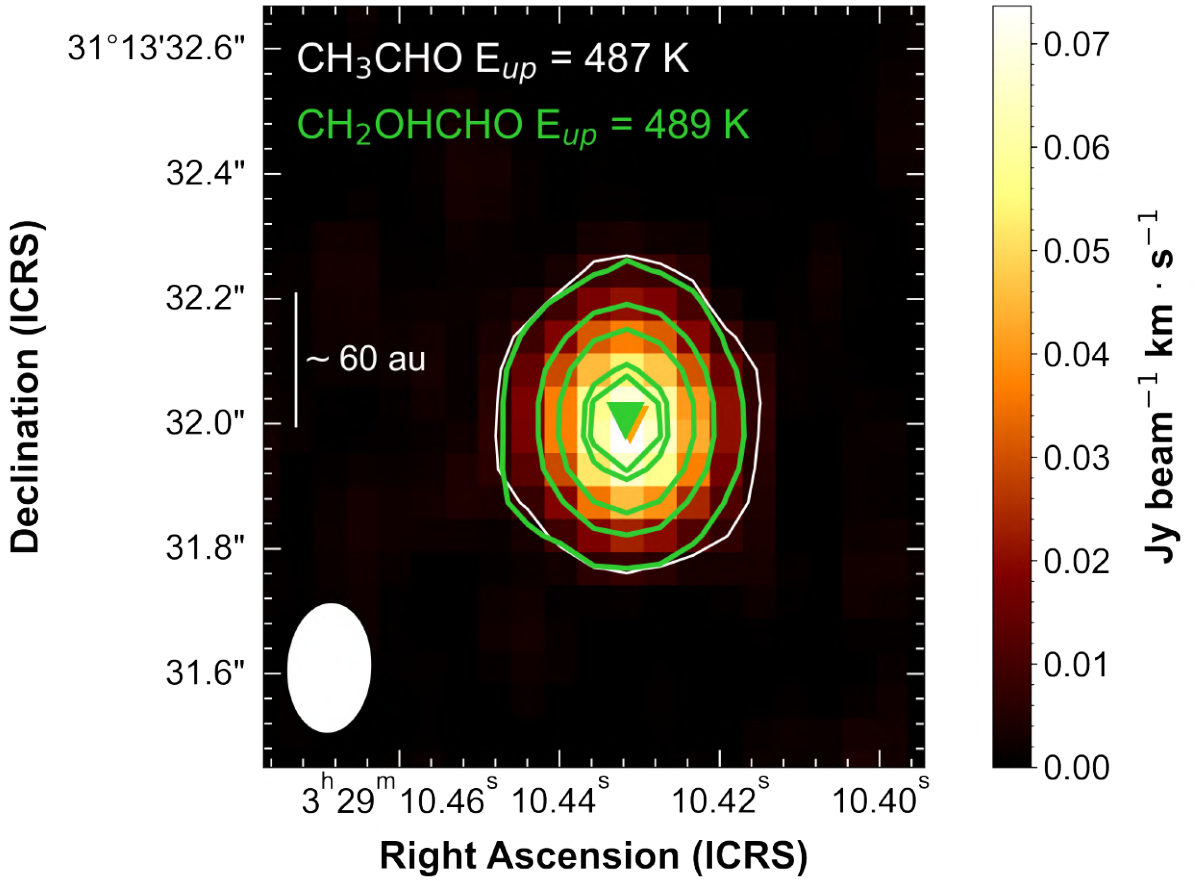

Figure B.2: Comparison of integrated emission of acetaldehyde (CH_3CHO) and glycolaldehyde (CH_2OHCHO) at 234 GHz. Acetaldehyde emission is shown in colour scale plus a 3σ white contour of 7 mJy beam^{-1} km s^{-1}. Glycolaldehyde emission is in 5σ steps green contours starting from $3\sigma \sim 7$ mJy beam^{-1} km s^{-1}. The triangles in orange and green indicate the emission peaks of acetaldehyde and glycolaldehyde. The restoring beam is 0.21" x 0.14", $-3°$.

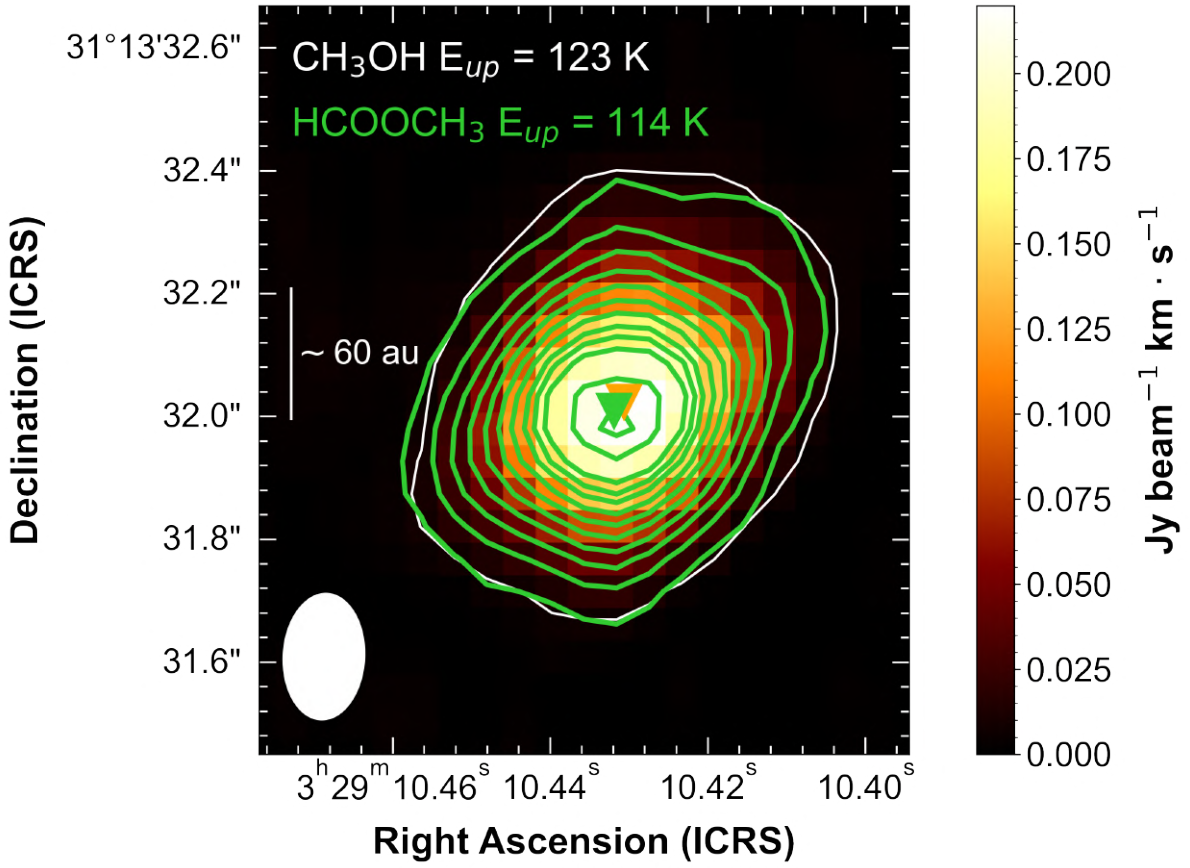

Figure B.3: Comparison of integrated emission of methanol (CH_3OH) and methyl formate ($HCOOCH_3$) at 234 GHz. Methanol emission is shown in colour scale plus a 3σ white contour of 10 mJy beam^{-1} km s^{-1}. Methyl formate emission is in 5σ steps green contours starting from $3\sigma \sim 7$ mJy beam^{-1} km s^{-1}. The triangles in orange and green indicate the emission peaks of methanol and methyl formate. The restoring beam is 0.21" x 0.14", –3°.

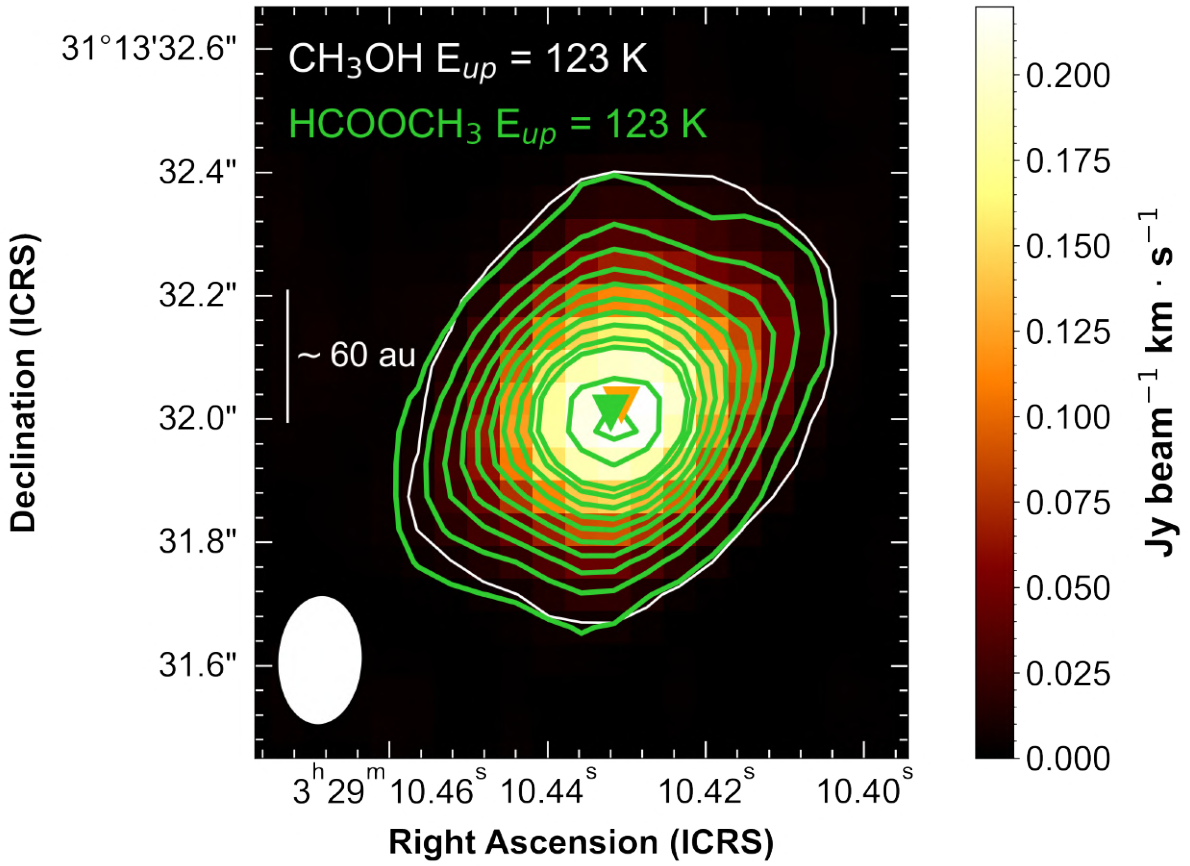

Figure B.4: Comparison of integrated emission of methanol (CH_3OH) and methyl formate ($HCOOCH_3$) at 234 GHz. Methanol emission is shown in colour scale plus a 3σ white contour of 8 mJy beam^{-1} km s^{-1}. Methyl formate emission is in 5σ steps green contours starting from $3\sigma \sim 7$ mJy beam^{-1} km s^{-1}. The triangles in orange and green indicate the emission peaks of methanol and methyl formate. The restoring beam is 0.21" x 0.14", –3°.

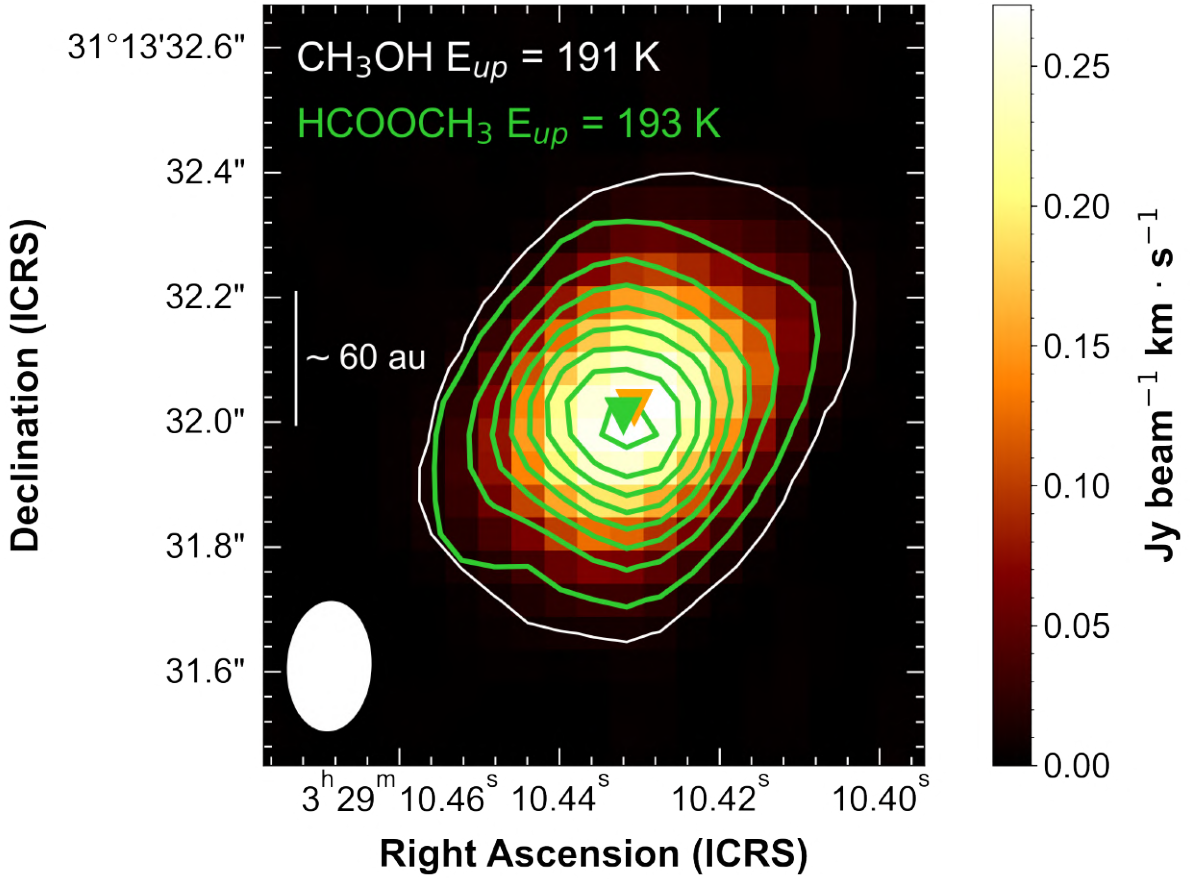

Figure B.5: Comparison of integrated emission of methanol (CH_3OH) and methyl formate ($HCOOCH_3$) at 234 GHz. Methanol emission is shown in colour scale plus a 3σ white contour of 9 mJy beam^{-1} km s^{-1}. Methyl formate emission is in 5σ steps green contours starting from $3\sigma \sim 8$ mJy beam^{-1} km s^{-1}. The triangles in orange and green indicate the emission peaks of methanol and methyl formate. The restoring beam is 0.21" x 0.14", –3°.

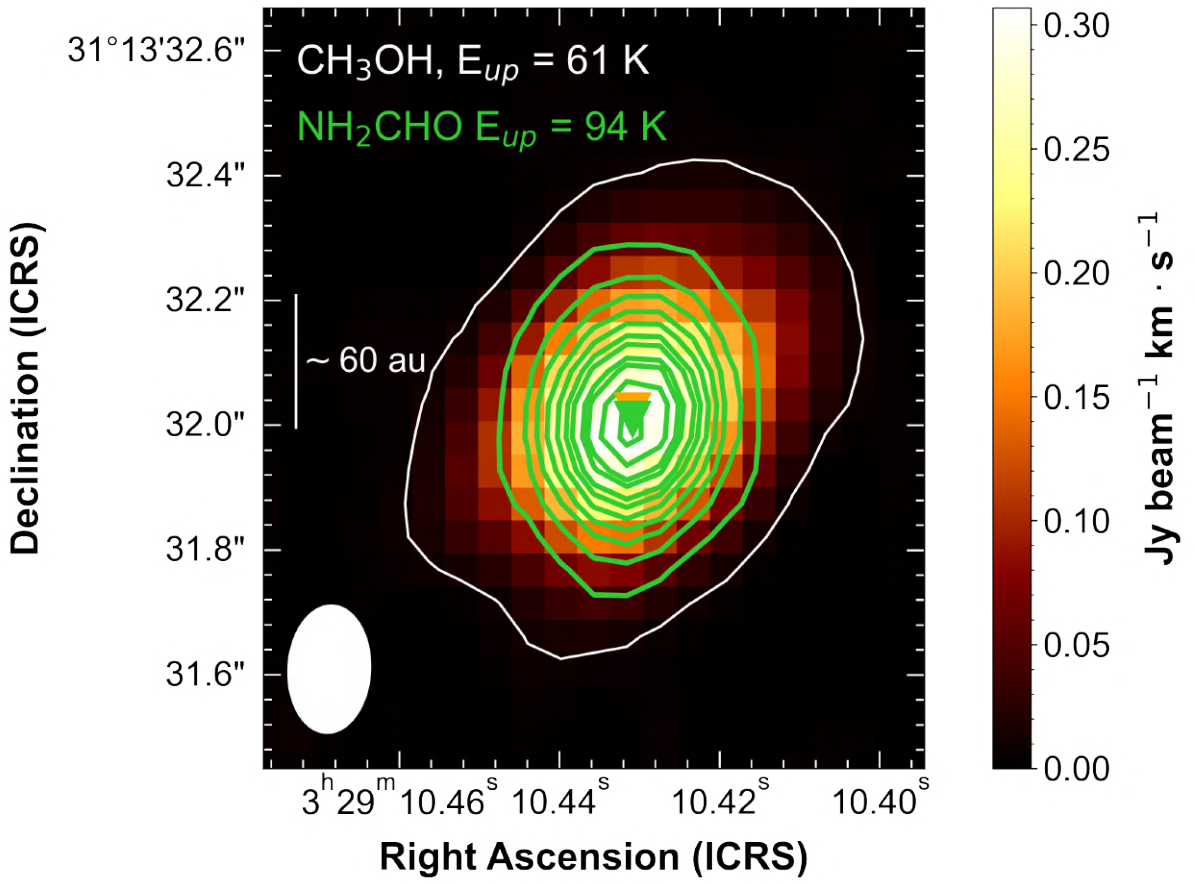

Figure B.6: Comparison of integrated emission of methanol (CH_3OH) and formamide (NH_2CHO) at 234 GHz. Methanol emission is shown in colour scale plus a 3σ white contour of 11 mJy beam^{-1} km s^{-1}. Formamide emission is in 5σ steps green contours starting from $3\sigma \sim 9$ mJy beam^{-1} km s^{-1}. The triangles in orange and green indicate the emission peaks of methanol and formamide. The restoring beam is 0.21" x 0.14", −3°.

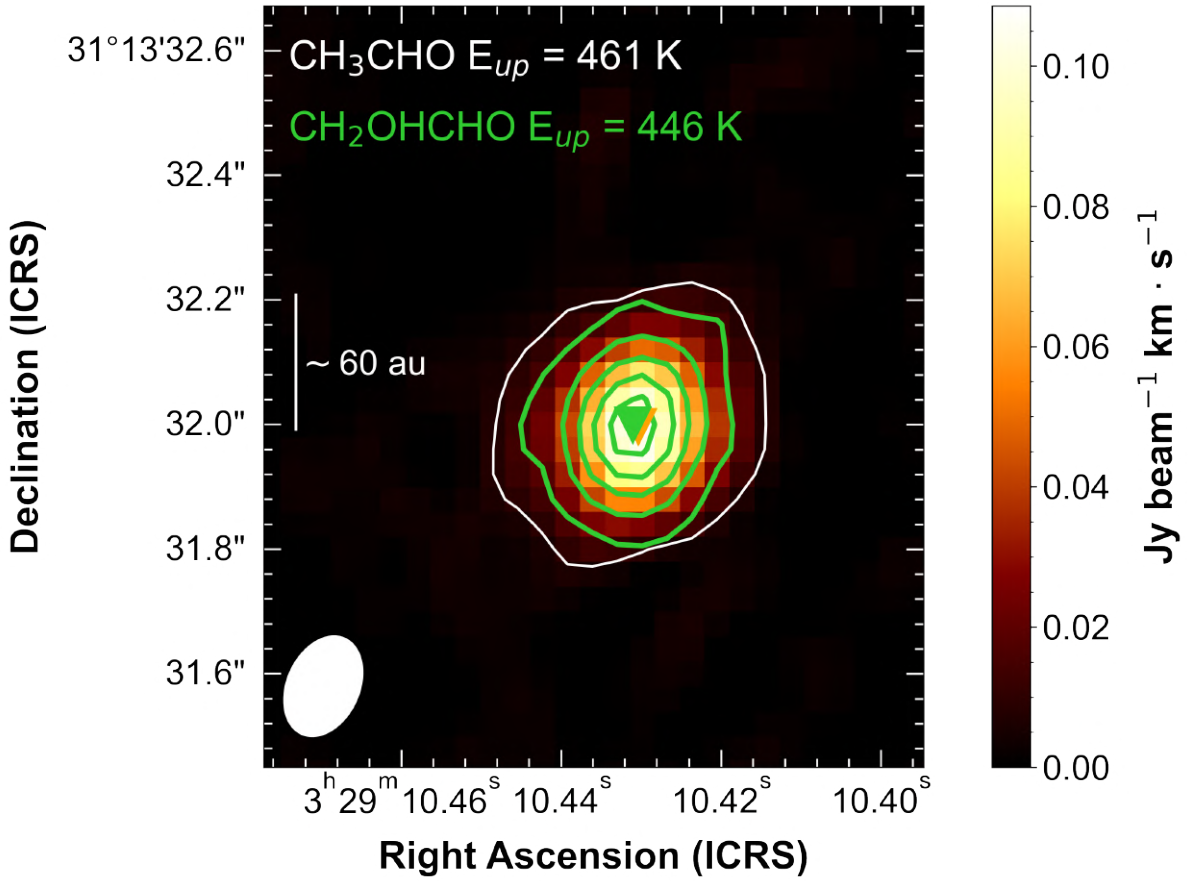

Figure B.7: Comparison of integrated emission of acetaldehyde (CH_3CHO) and glycolaldehyde (CH_2OHCHO) at 245 GHz. Acetaldehyde emission is shown in colour scale plus a 3σ white contour of 8 mJy beam^{-1} km s^{-1}. Glycolaldehyde emission is in 5σ steps green contours starting from $3\sigma \sim 9$ mJy beam^{-1} km s^{-1}. The triangles in orange and green indicate the emission peaks of acetaldehyde and glycolaldehyde. The restoring beam is 0.17" x 0.12", −24°.

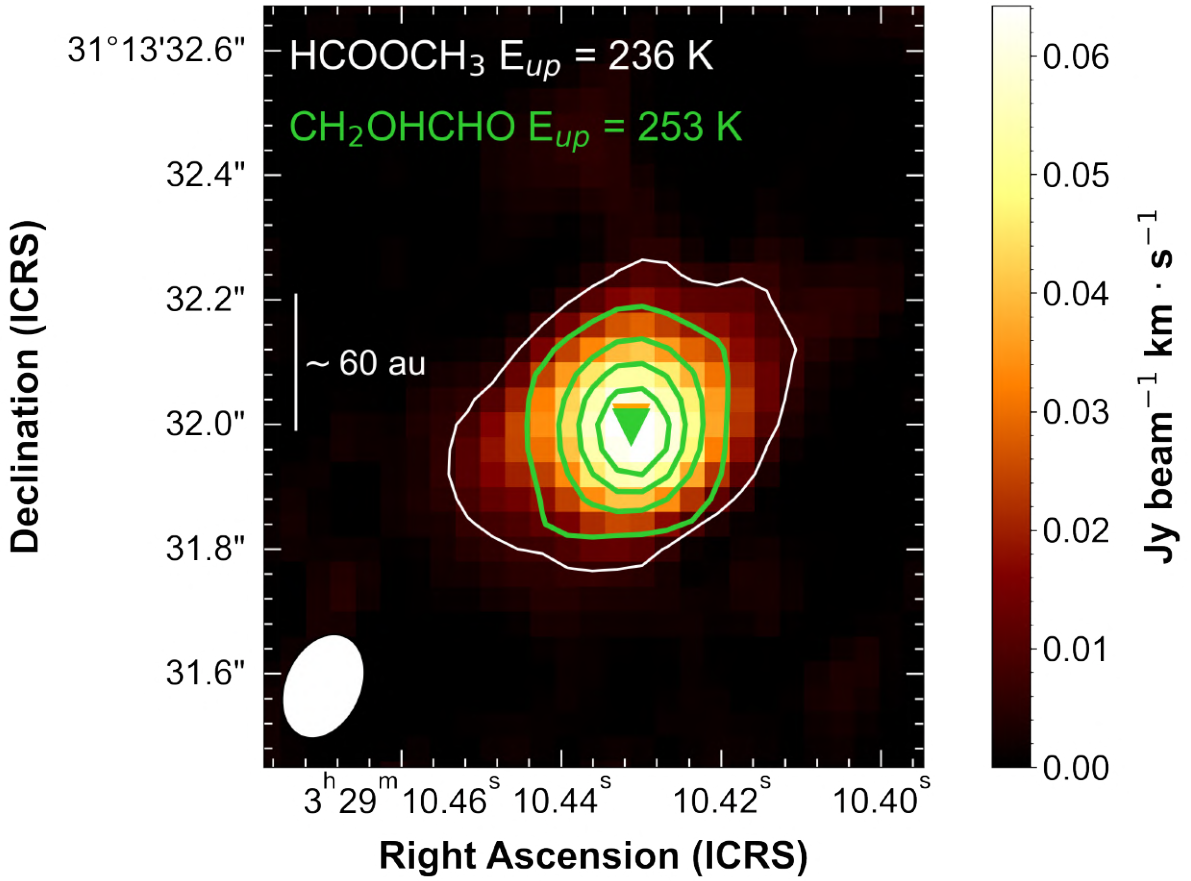

Figure B.8: Comparison of integrated emission of methyl formate ($HCOOCH_3$) and glycolaldehyde (CH_2OHCHO) at 245 GHz. Methyl formate emission is shown in colour scale plus a 3σ white contour of 7 mJy beam^{-1} km s^{-1}. Glycolaldehyde emission is in 5σ steps green contours starting from $3\sigma \sim 9$ mJy beam^{-1} km s^{-1}. The triangles in orange and green indicate the emission peaks of methyl formate and glycolaldehyde. The restoring beam is 0.17" x 0.12", −24°.

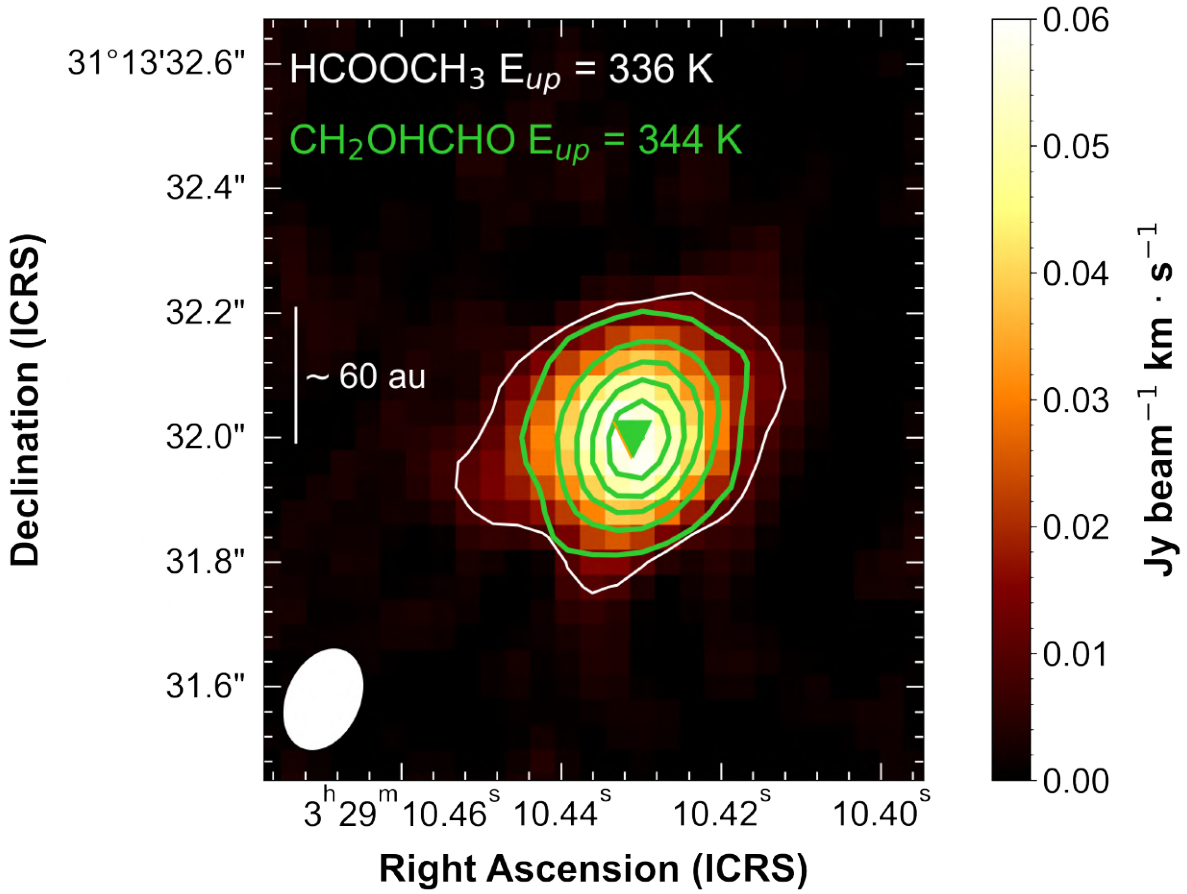

Figure B.9: Comparison of integrated emission of methyl formate ($HCOOCH_3$) and glycolaldehyde (CH_2OHCHO) at 245 GHz. Methyl formate emission is shown in colour scale plus a 3σ white contour of 9 mJy beam^{-1} km s^{-1}. Glycolaldehyde emission is in 5σ steps green contours starting from $3\sigma \sim 7$ mJy beam^{-1} km s^{-1}. The triangles in orange and green indicate the emission peaks of methyl formate and glycolaldehyde. The restoring beam is 0.17" x 0.12", −24°.

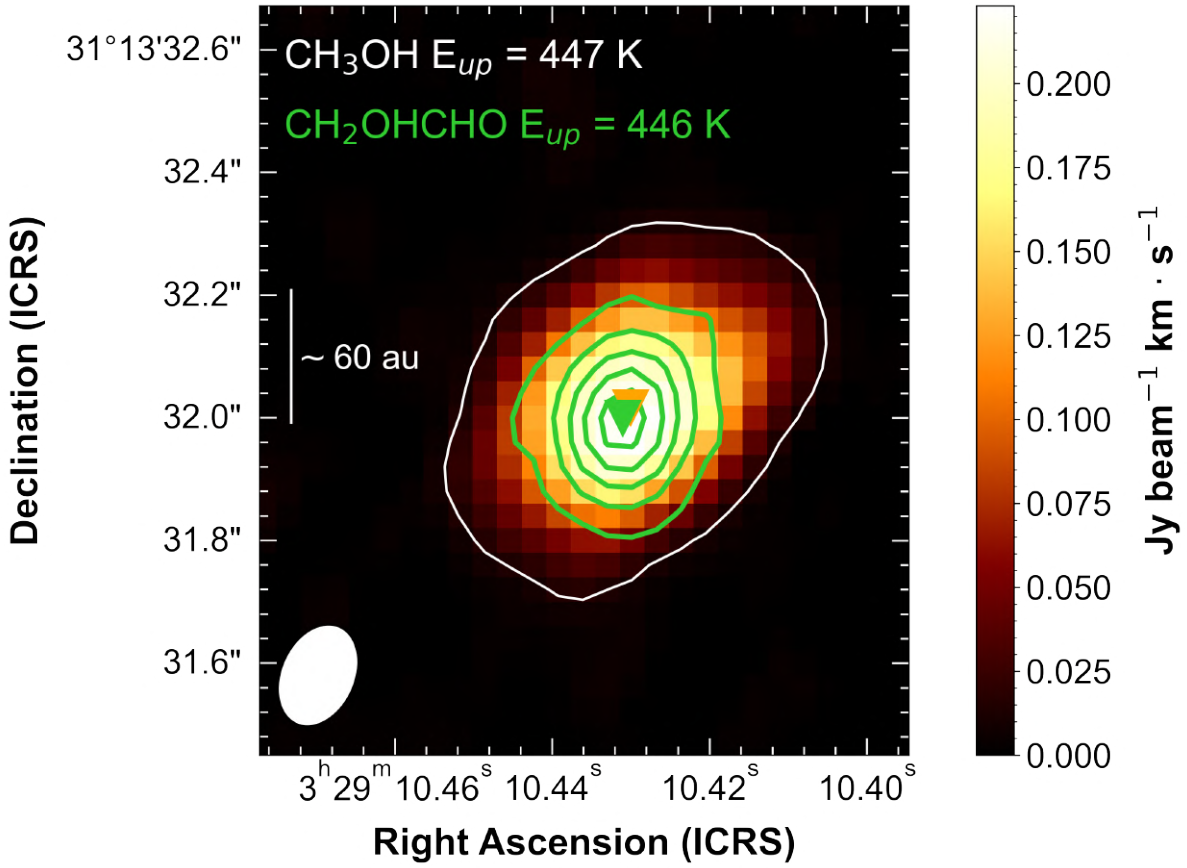

Figure B.10: Comparison of integrated emission of methanol (CH_3OH) and glycolaldehyde (CH_2OHCHO) at 245 GHz. Methanol emission is shown in colour scale plus a 3σ white contour of 13 mJy beam^{-1} km s^{-1}. Glycolaldehyde emission is in 5σ steps green contours starting from $3\sigma \sim 9$ mJy beam^{-1} km s^{-1}. The triangles in orange and green indicate the emission peaks of methanol and glycolaldehyde. The restoring beam is 0.17" x 0.12", −24°.

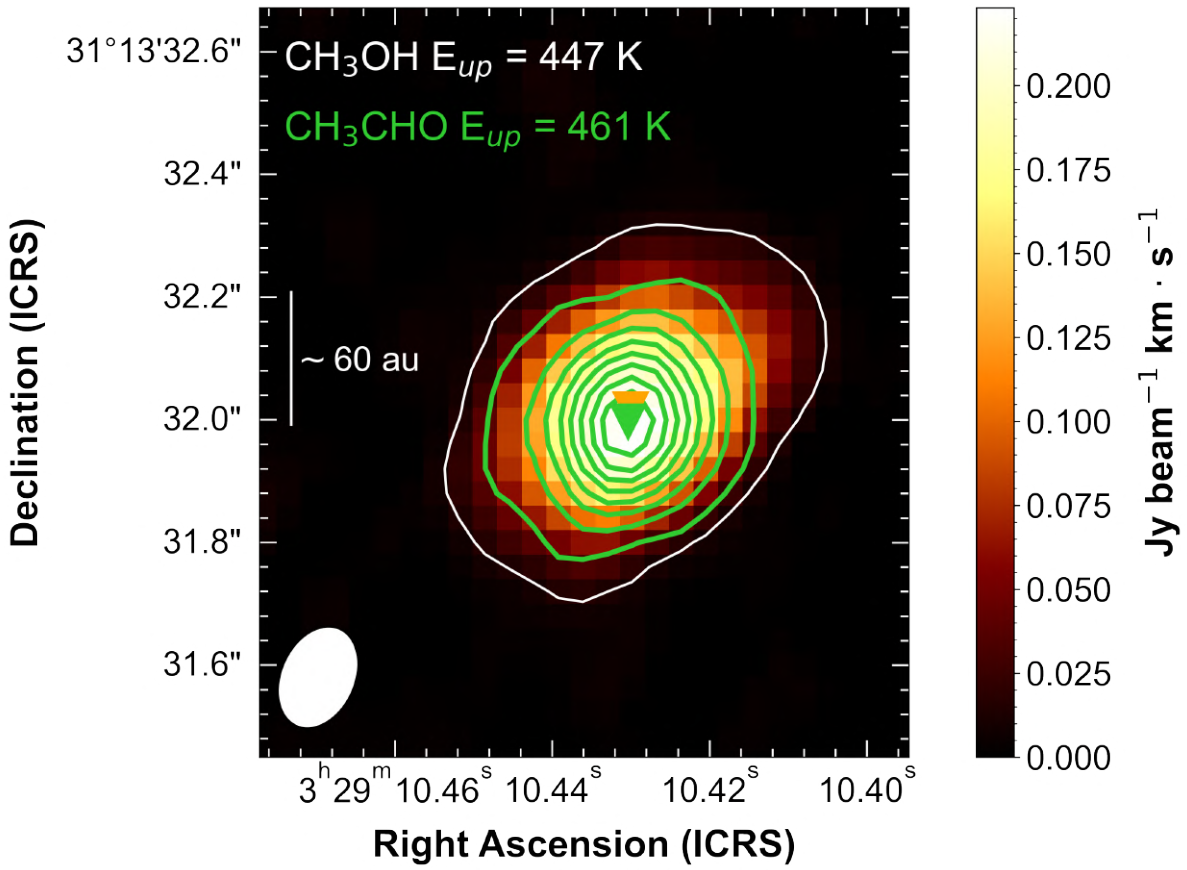

Figure B.11: Comparison of integrated emission of methanol (CH_3OH) and acetaldehyde (CH_3CHO) at 245 GHz. Methanol emission is shown in colour scale plus a 3σ white contour of 13 mJy beam^{-1} km s^{-1}. Acetaldehyde emission is in 5σ steps green contours starting from $3\sigma \sim 8$ mJy beam^{-1} km s^{-1}. The triangles in orange and green indicate the emission peaks of methanol and acetaldehyde. The restoring beam is 0.17" x 0.12", −24°.

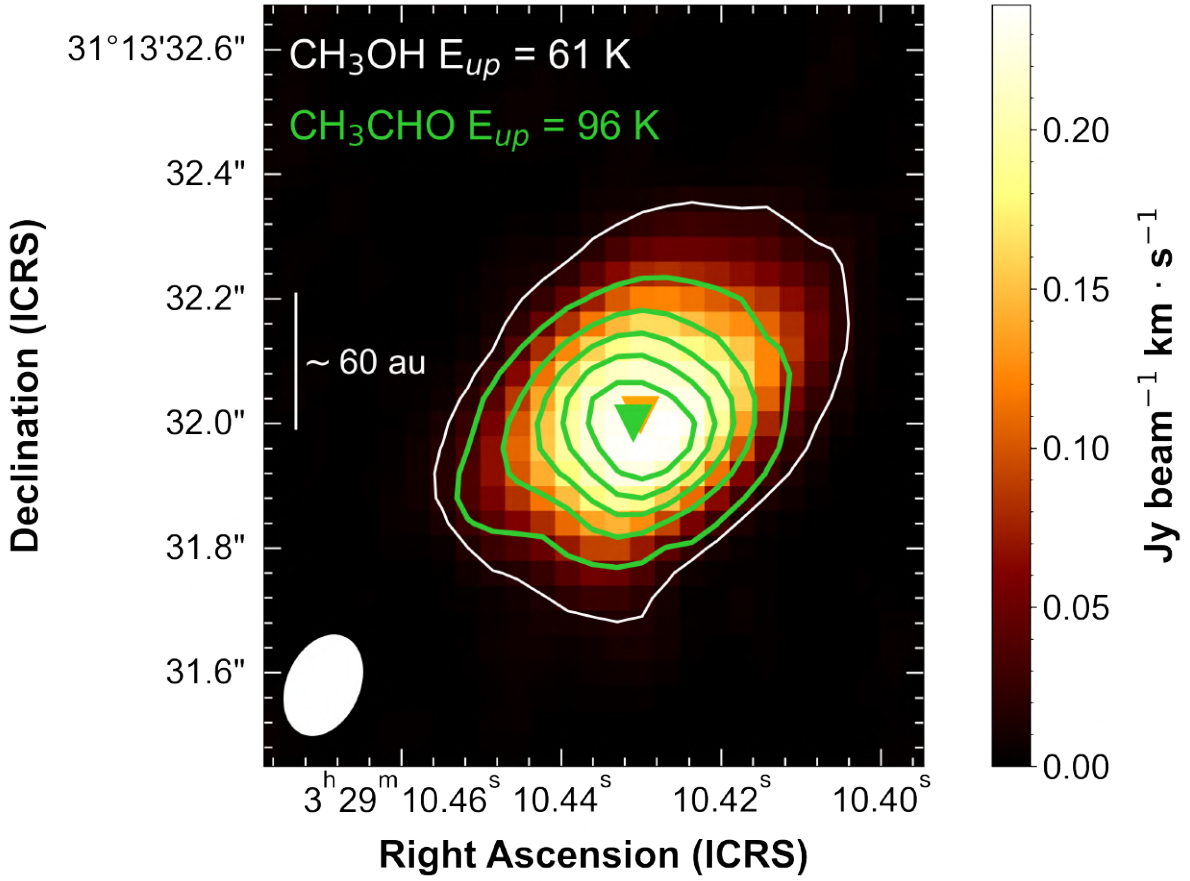

Figure B.12: Comparison of integrated emission of methanol (CH_3OH) and acetaldehyde (CH_3CHO) at 245 GHz. Methanol emission is shown in colour scale plus a 3σ white contour of 15 mJy beam^{-1} km s^{-1}. Acetaldehyde emission is in 5σ steps green contours starting from $3\sigma \sim 10$ mJy beam^{-1} km s^{-1}. The triangles in orange and green indicate the emission peaks of methanol and acetaldehyde. The restoring beam is 0.17" x 0.12", –24°.

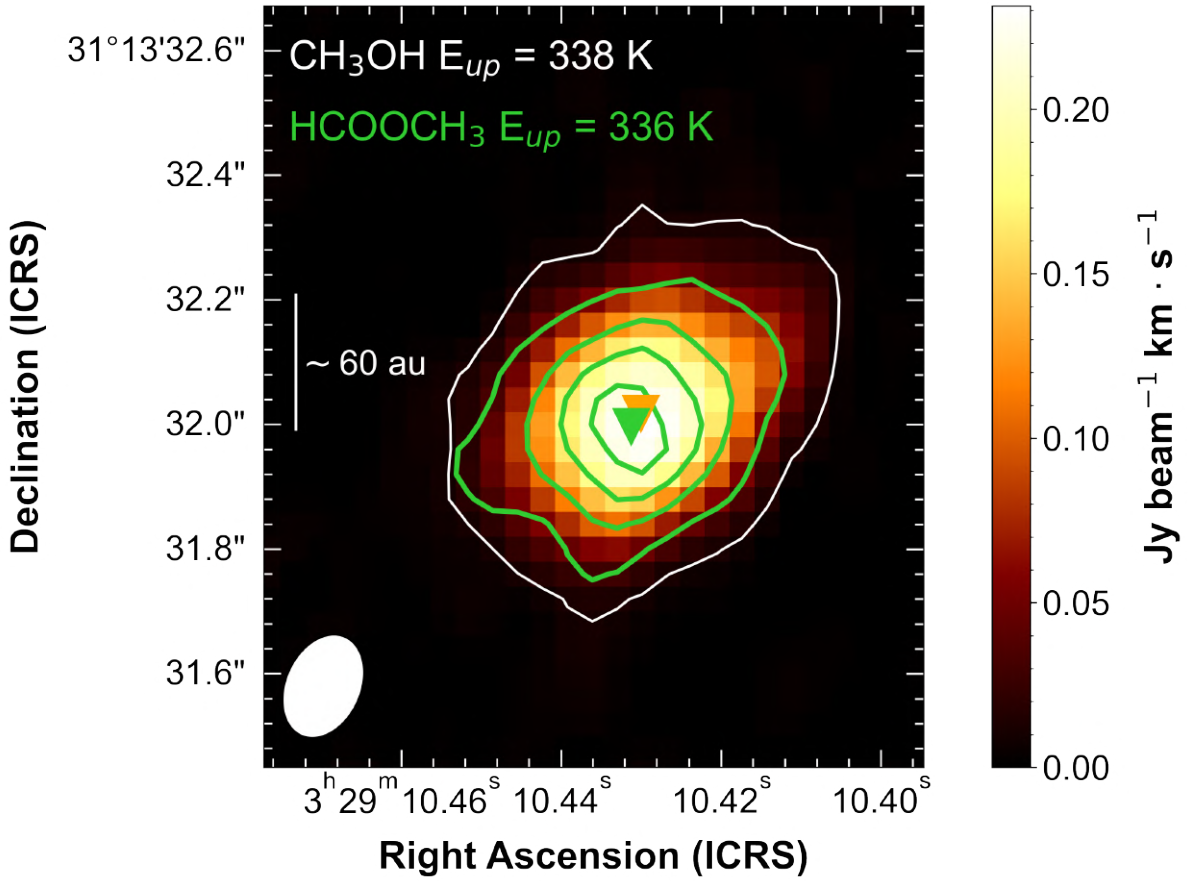

Figure B.13: Comparison of integrated emission of methanol (CH_3OH) and methyl formate ($HCOOCH_3$) at 245 GHz. Methanol emission is shown in colour scale plus a 3σ white contour of 11 mJy beam^{-1} km s^{-1}. Methyl formate emission is in 5σ steps green contours starting from $3\sigma \sim 9$ mJy beam^{-1} km s^{-1} (234 GHz). The triangles in orange and green indicate the emission peaks of methanol and methyl formate. The restoring beam is 0.17" x 0.12", −24°.

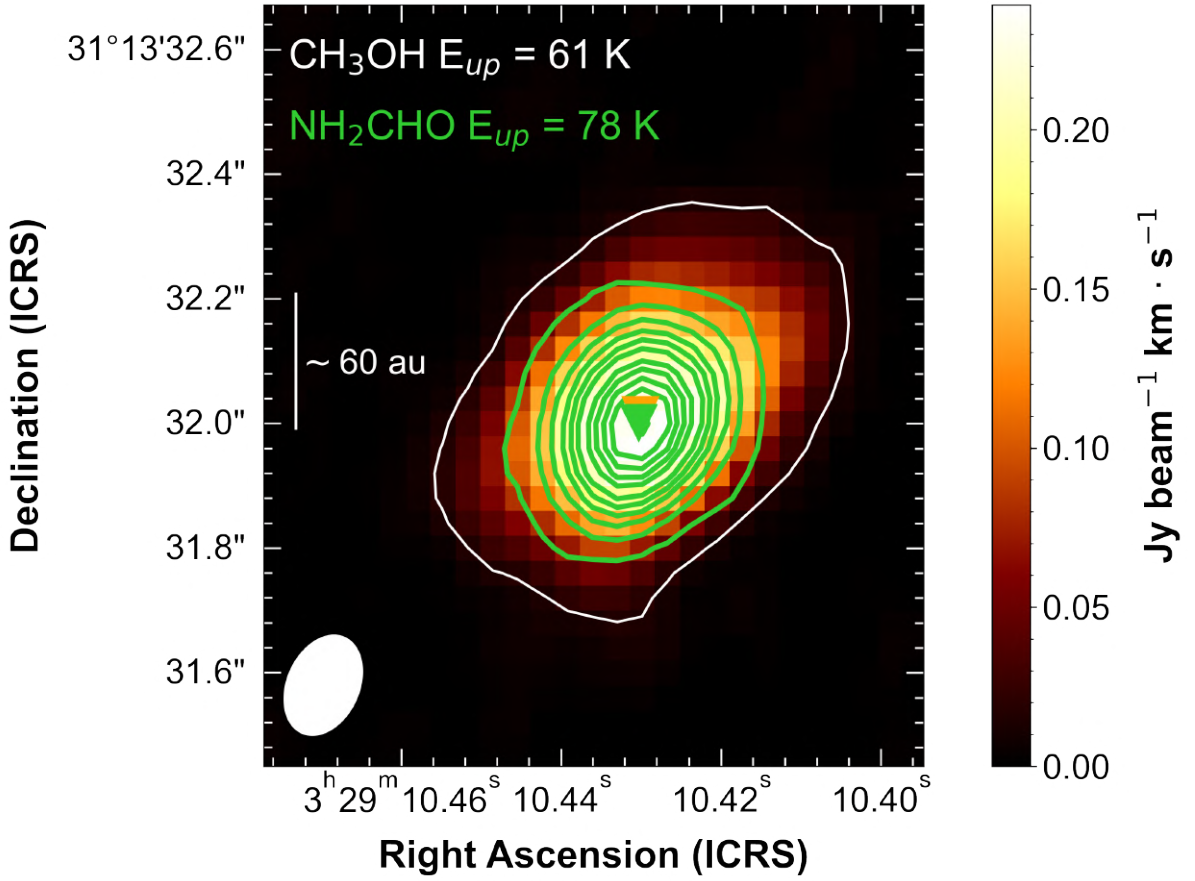

Figure B.14: Comparison of integrated emission of methanol (CH_3OH) and formamide (NH_2CHO at 245 GHz. Methanol emission is shown in colour scale plus a 3σ white contour of 15 mJy beam^{-1} km s^{-1}. Formamide emission is in 5σ steps green contours starting from $3\sigma \sim$ 13 mJy beam^{-1} km s^{-1}. The triangles in orange and green indicate the emission peaks of methanol and formamide. The restoring beam is 0.17" x 0.12", –24°.

APPENDIX C

Moment 0 Maps And Gaussian Fit Residuals

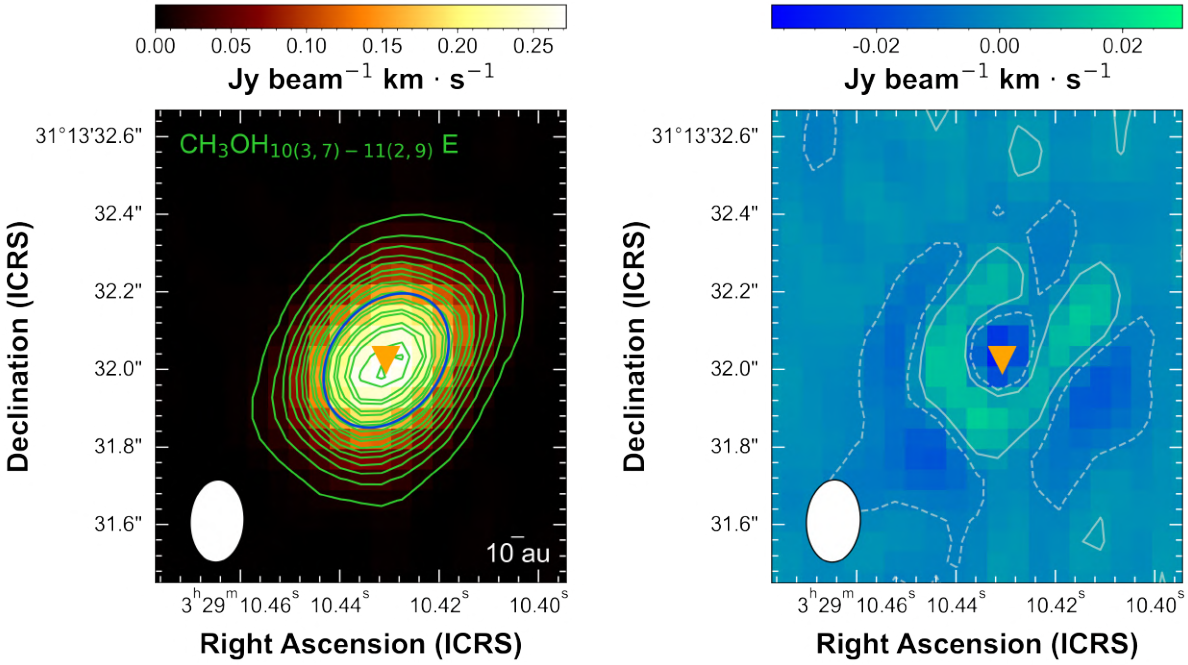

Figure C.1: Gaussian fitted moment-0 map (left) and fit residuals (right) of selected methanol line at 234 GHz. Methanol integrated emission is shown in color scale and green contours starting from 3σ of ~ 9 mJy beam^{-1} km s^{-1} with step of 5σ. The fitting elliptical gaussian convolved with the beam is overlayed in blue solid line. The corresponding fit residual images are in color scale and contoured with $\pm 1\sigma$ of ~ 3 mJy beam^{-1} km s^{-1} (solid and dashed line respectively). In both panels, the triangles in orange indicate the emission peaks of methanol and the beam is 0.21" x 0.14", −3°.

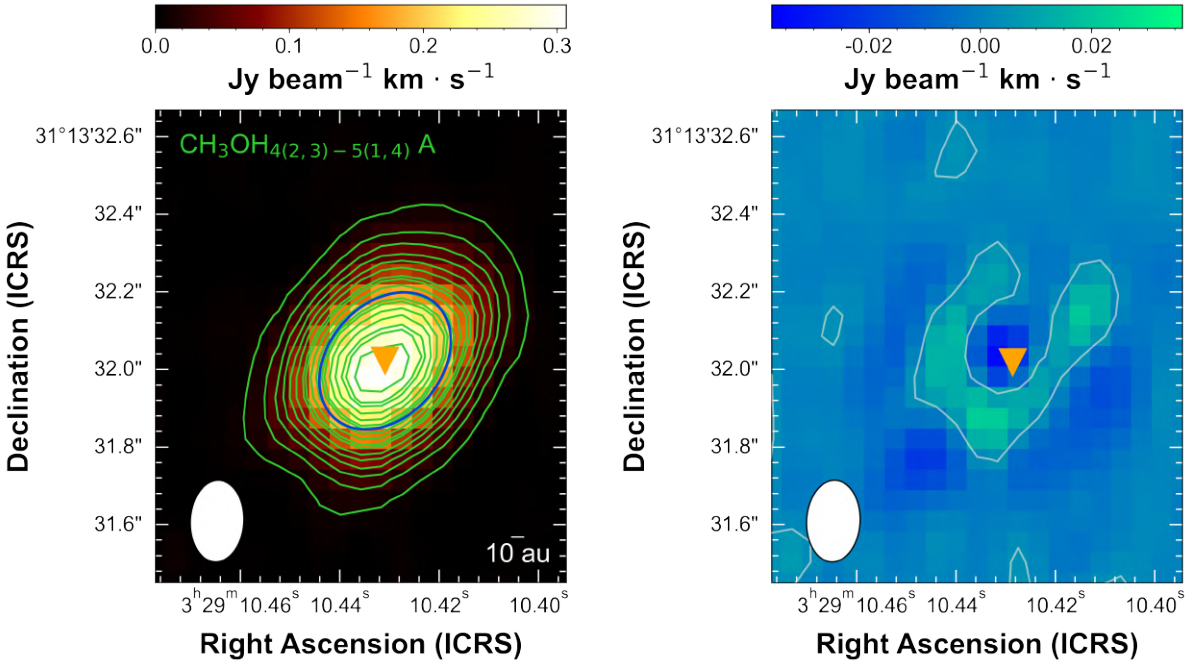

Figure C.2: Gaussian fitted moment-0 map (left) and fit residuals (right) of selected methanol line at 234 GHz. Methanol integrated emission is shown in color scale and green contours starting from 3σ of $\sim$ 11 mJy beam^{-1} km s^{-1} with step of 5σ. The fitting elliptical gaussian convolved with the beam is overlayed in blue solid line. The corresponding fit residual images are in color scale and contoured with $\pm 1\sigma$ of $\sim$ 4 mJy beam^{-1} km s^{-1} (solid and dashed line respectively). In both panels, the triangles in orange indicate the emission peaks of methanol and the beam is 0.21" x 0.14", −3°.

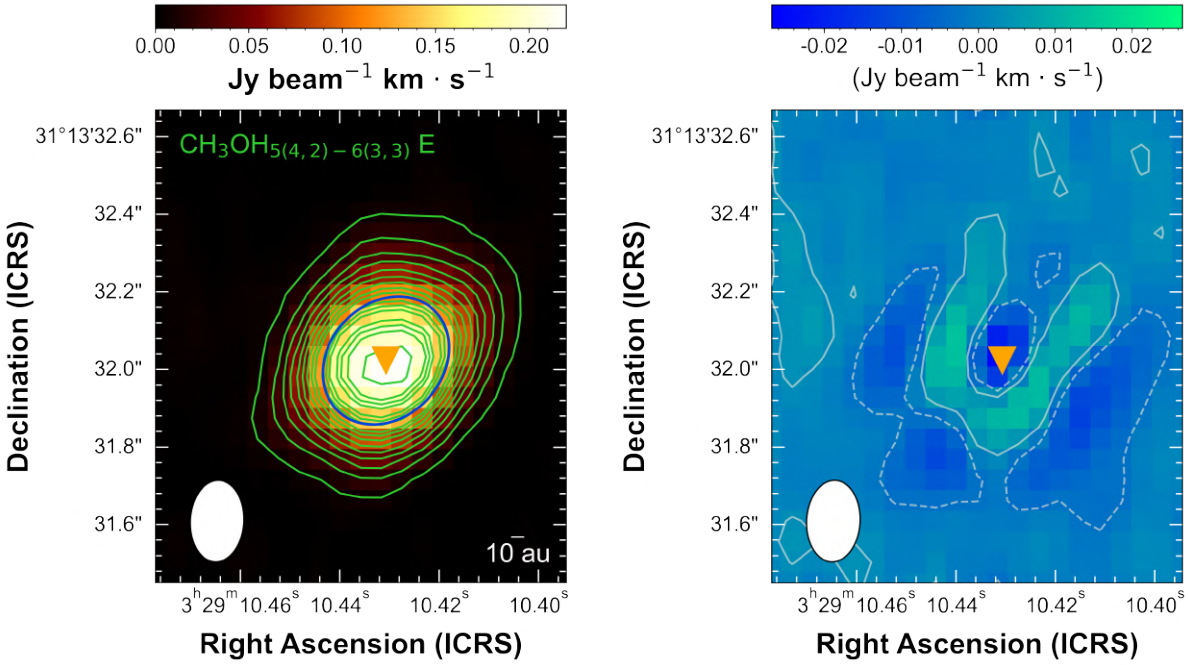

Figure C.3: Gaussian fitted moment-0 map (left) and fit residuals (right) of selected methanol line at 234 GHz. Methanol integrated emission is shown in color scale and green contours starting from 3σ of ~ 8 mJy beam^{-1} km s^{-1} with step of 5σ. The fitting elliptical gaussian convolved with the beam is overlayed in blue solid line. The corresponding fit residual images are in color scale and contoured with $\pm 1\sigma$ of ~ 3 mJy beam^{-1} km s^{-1} (solid and dashed line respectively). In both panels, the triangles in orange indicate the emission peaks of methanol and the beam is 0.21" x 0.14", −3°.

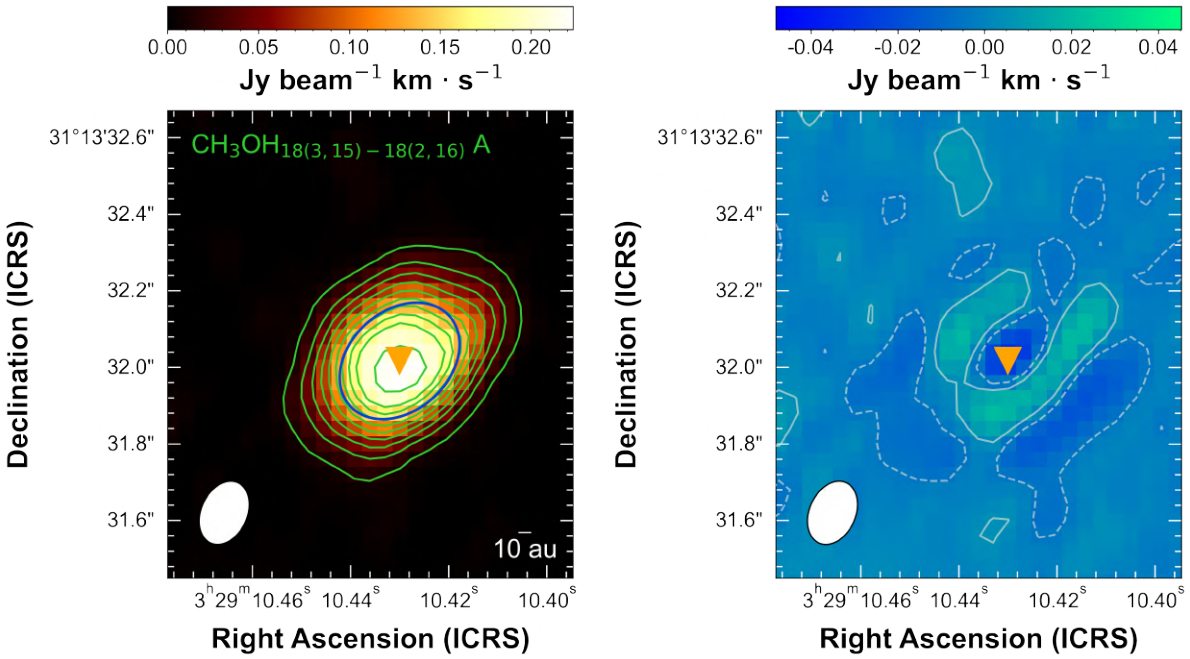

Figure C.4: Gaussian fitted moment-0 map (left) and fit residuals (right) of selected methanol line at 245 GHz. Methanol integrated emission is shown in color scale and green contours starting from 3σ of $\sim$ 13 mJy beam^{-1} km s^{-1} with step of 5σ. The fitting elliptical gaussian convolved with the beam is overlayed in blue solid line. The corresponding fit residual images are in color scale and contoured with $\pm1\sigma$ of $\sim$ 4 mJy beam^{-1} km s^{-1} (solid and dashed line respectively). In both panels, the triangles in orange indicate the emission peaks of methanol and the beam is 0.17" x 0.12", −24°.

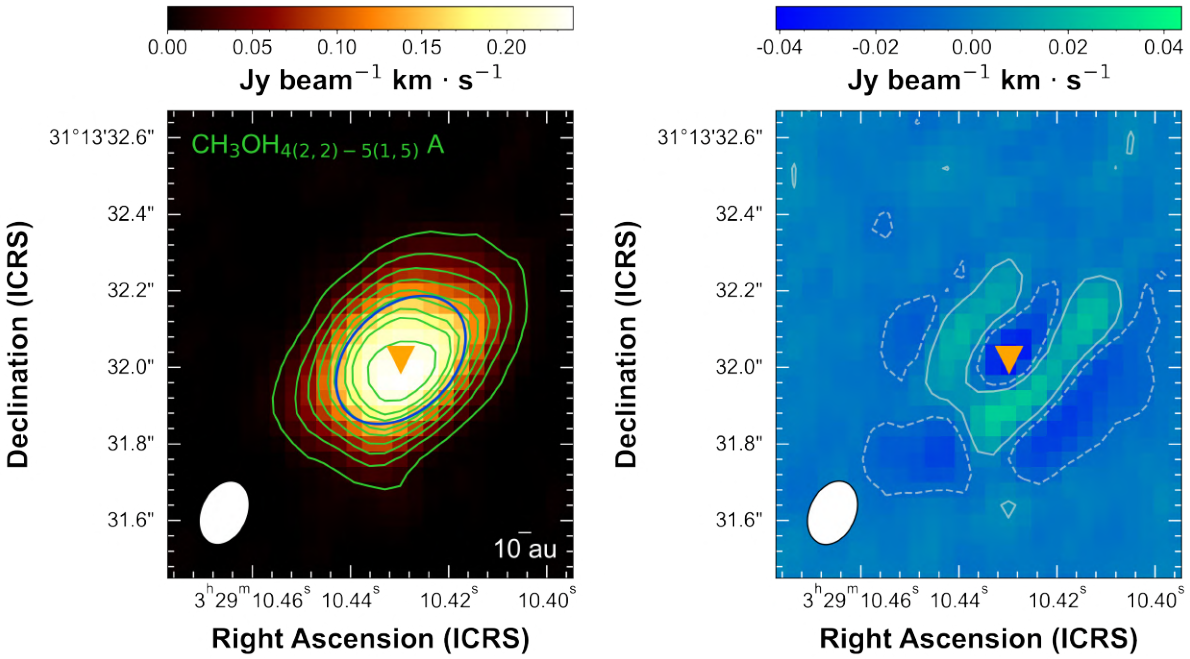

Figure C.5: Gaussian fitted moment-0 map (left) and fit residuals (right) of selected methanol line at 245 GHz. Methanol integrated emission is shown in color scale and green contours starting from 3σ of $\sim$ 15 mJy beam^{-1} km s^{-1} with step of 5σ. The fitting elliptical gaussian convolved with the beam is overlayed in blue solid line. The corresponding fit residual images are in color scale and contoured with $\pm 1\sigma$ of $\sim$ 5 mJy beam^{-1} km s^{-1} (solid and dashed line respectively). In both panels, the triangles in orange indicate the emission peaks of methanol and the beam is 0.17" x 0.12", −24°.

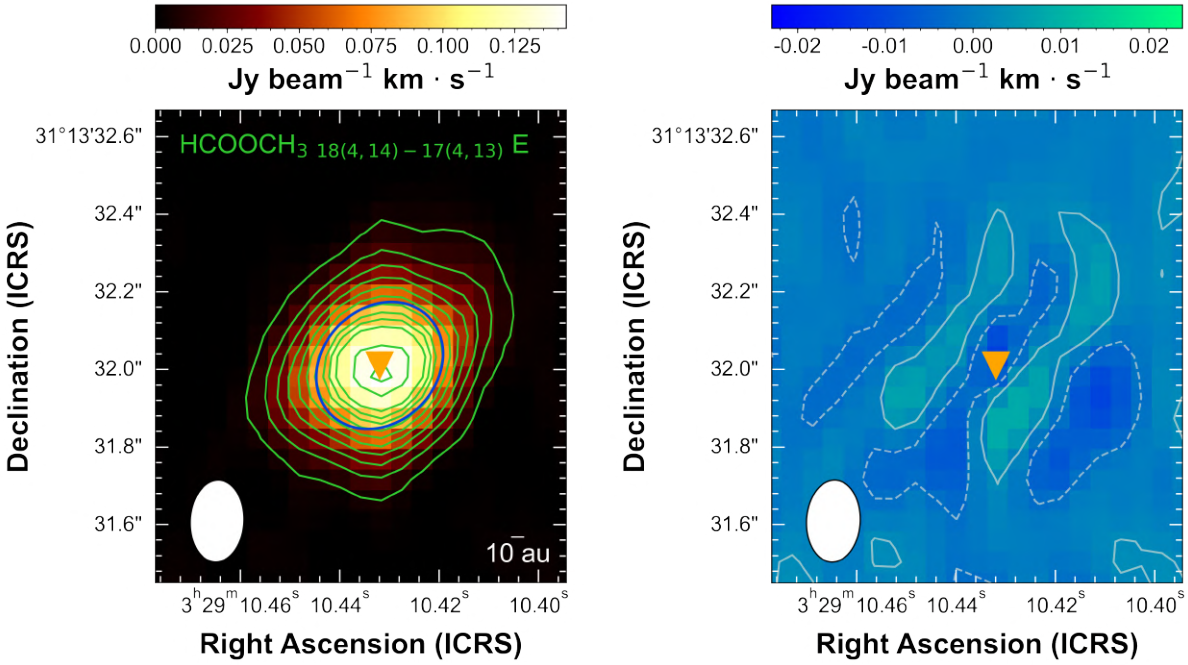

Figure C.6: Gaussian fitted moment-0 map (left) and fit residuals (right) of selected methyl formate line at 234 GHz. Methyl formate integrated emission is shown in color scale and green contours starting from 3σ of ~ 7 mJy beam^{-1} km s^{-1} with step of 5σ. The fitting elliptical gaussian convolved with the beam is overlayed in blue solid line. The corresponding fit residual images are in color scale and contoured with $\pm 1\sigma$ of ~ 3 mJy beam^{-1} km s^{-1} (solid and dashed line respectively). In both panels, the triangles in orange indicate the emission peaks of methyl formate and the beam is 0.21" x 0.14", –3°.

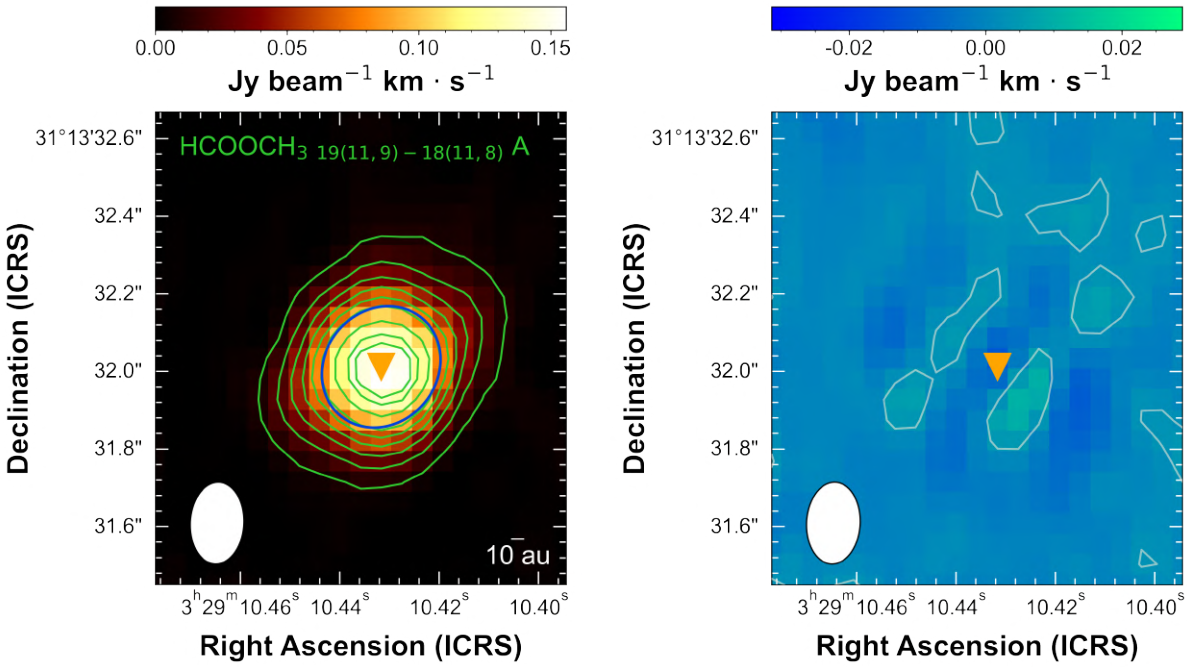

Figure C.7: Gaussian fitted moment-0 map (left) and fit residuals (right) of selected methyl formate line at 234 GHz. Methyl formate integrated emission is shown in color scale and green contours starting from 3σ of $\sim$ 9 mJy beam^{-1} km s^{-1} with step of 5σ. The fitting elliptical gaussian convolved with the beam is overlayed in blue solid line. The corresponding fit residual images are in color scale and contoured with $\pm1\sigma$ of $\sim$ 3 mJy beam^{-1} km s^{-1} (solid and dashed line respectively). In both panels, the triangles in orange indicate the emission peaks of methyl formate and the beam is 0.21" x 0.14", –3°.

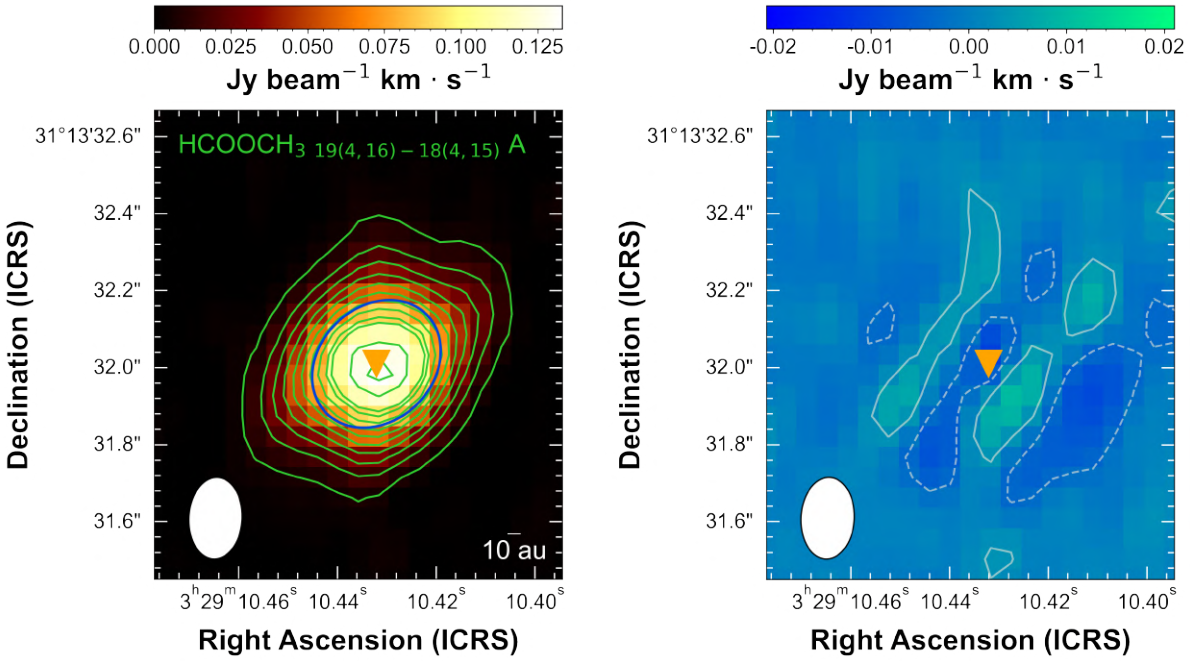

Figure C.8: Gaussian fitted moment-0 map (left) and fit residuals (right) of selected methyl formate line at 234 GHz. Methyl formate integrated emission is shown in color scale and green contours starting from 3σ of ~ 7 mJy beam^{-1} km s^{-1} with step of 5σ. The fitting elliptical gaussian convolved with the beam is overlayed in blue solid line. The corresponding fit residual images are in color scale and contoured with $\pm 1\sigma$ of ~ 3 mJy beam^{-1} km s^{-1} (solid and dashed line respectively). In both panels, the triangles in orange indicate the emission peaks of methyl formate and the beam is 0.21" x 0.14", $-3°$.

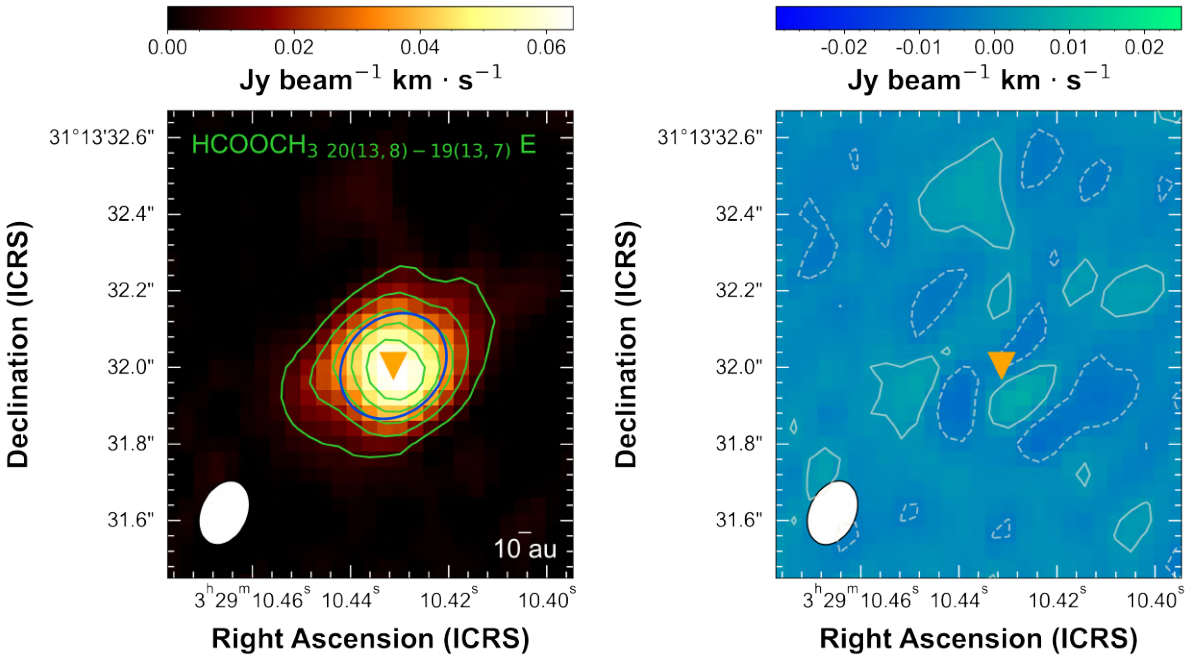

Figure C.9: Gaussian fitted moment-0 map (left) and fit residuals (right) of selected methyl formate line at 245 GHz. Methyl formate integrated emission is shown in color scale and green contours starting from 3σ of ~ 7 mJy beam^{-1} km s^{-1} with step of 5σ. The fitting elliptical gaussian convolved with the beam is overlayed in blue solid line. The corresponding fit residual images are in color scale and contoured with $\pm 1\sigma$ of ~ 2 mJy beam^{-1} km s^{-1} (solid and dashed line respectively). In both panels, the triangles in orange indicate the emission peaks of methyl formate and the beam is 0.17" x 0.12", –24°.

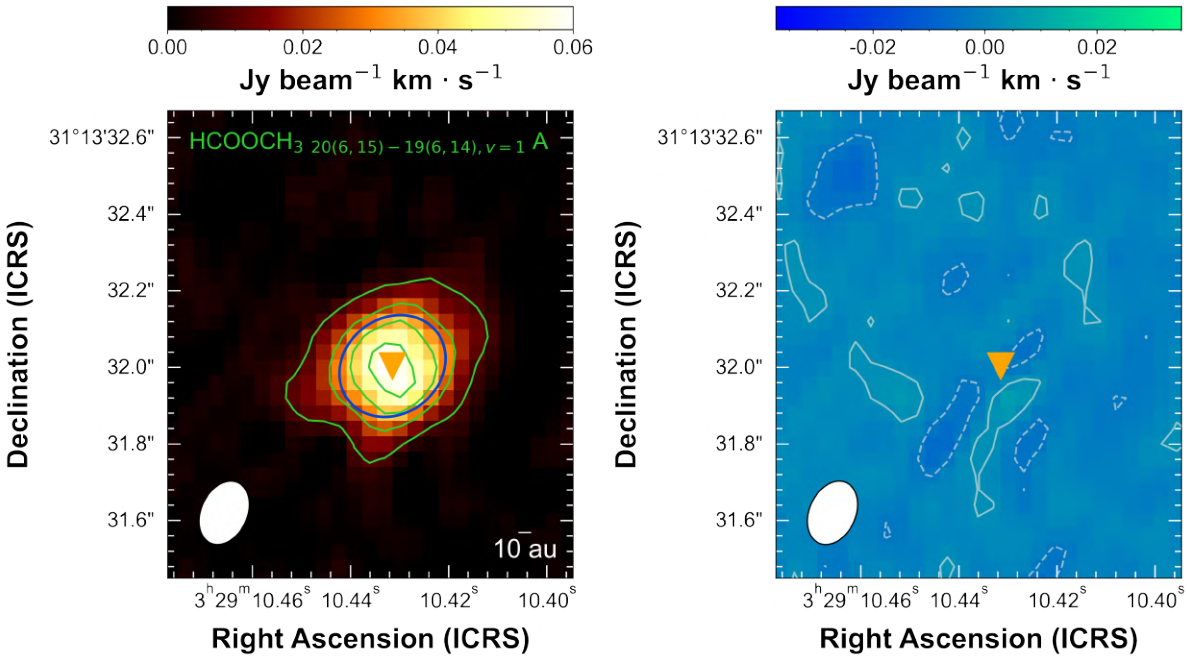

Figure C.10: Gaussian fitted moment-0 map (left) and fit residuals (right) of selected methyl formate line at 245 GHz. Methyl formate integrated emission is shown in color scale and green contours starting from 3σ of $\sim$ 9 mJy beam^{-1} km s^{-1} with step of 5σ. The fitting elliptical gaussian convolved with the beam is overlayed in blue solid line. The corresponding fit residual images are in color scale and contoured with $\pm1\sigma$ of $\sim$ 3 mJy beam^{-1} km s^{-1} (solid and dashed line respectively). In both panels, the triangles in orange indicate the emission peaks of methyl formate and the beam is 0.17" x 0.12", –24°.

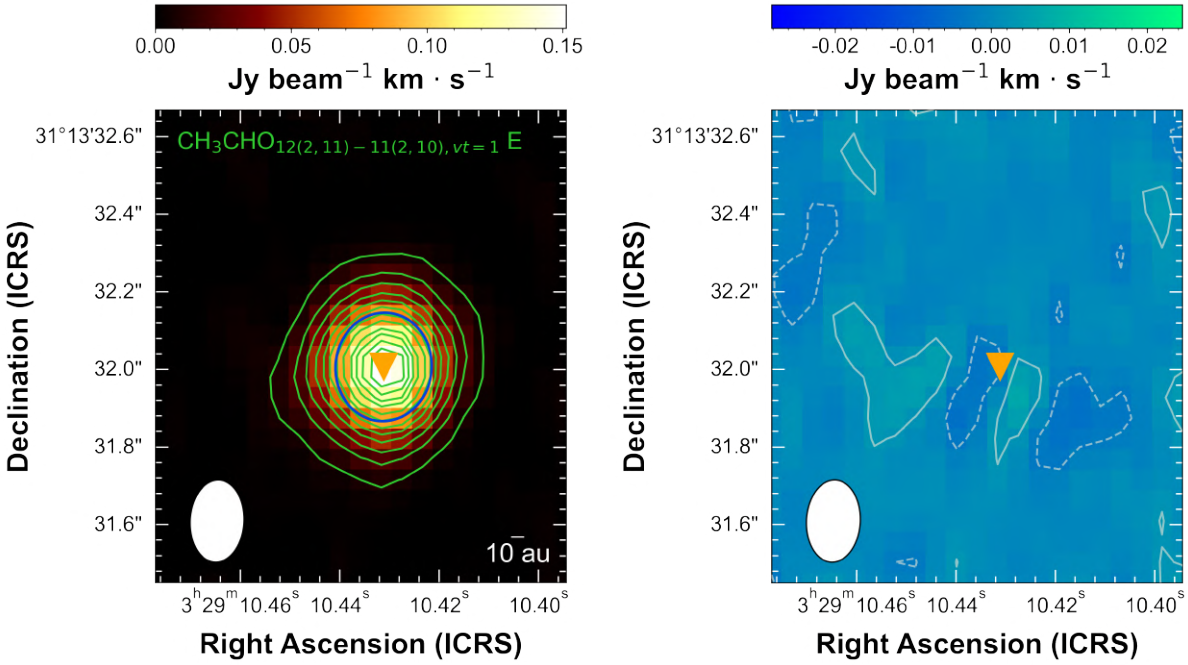

Figure C.11: Gaussian fitted moment-0 map (left) and fit residuals (right) of selected acetaldehyde line at 234 GHz. Acetaldehyde integrated emission is shown in color scale and green contours starting from 3σ of ~ 8 mJy beam^{-1} km s^{-1} with step of 5σ. The fitting elliptical gaussian convolved with the beam is overlayed in blue solid line. The corresponding fit residual images are in color scale and contoured with $\pm 1\sigma$ of ~ 3 mJy beam^{-1} km s^{-1} (solid and dashed line respectively). In both panels, the triangles in orange indicate the emission peaks of acetaldehyde and the beam is 0.21" x 0.14", −3°.

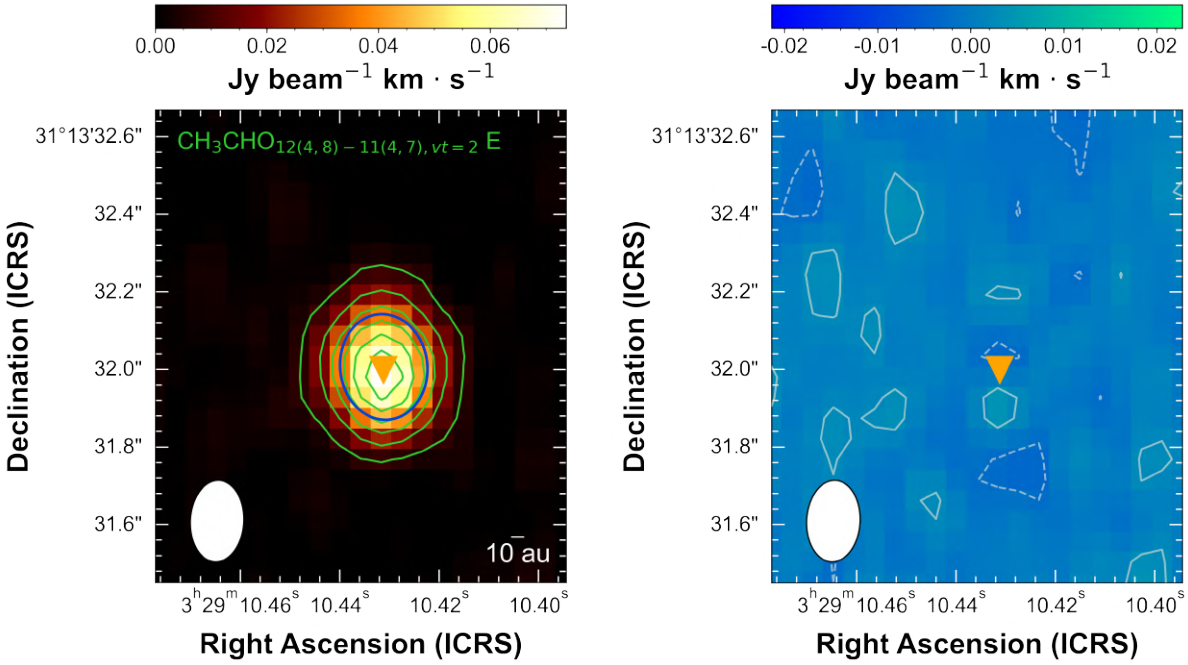

Figure C.12: Gaussian fitted moment-0 map (left) and fit residuals (right) of selected acetaldehyde line at 234 GHz. Acetaldehyde integrated emission is shown in color scale and green contours starting from 3σ of ~ 7 mJy beam^{-1} km s^{-1} with step of 5σ. The fitting elliptical gaussian convolved with the beam is overlayed in blue solid line. The corresponding fit residual images are in color scale and contoured with $\pm 1\sigma$ of ~ 2 mJy beam^{-1} km s^{-1} (solid and dashed line respectively). In both panels, the triangles in orange indicate the emission peaks of acetaldehyde and the beam is 0.21" x 0.14", −3°.

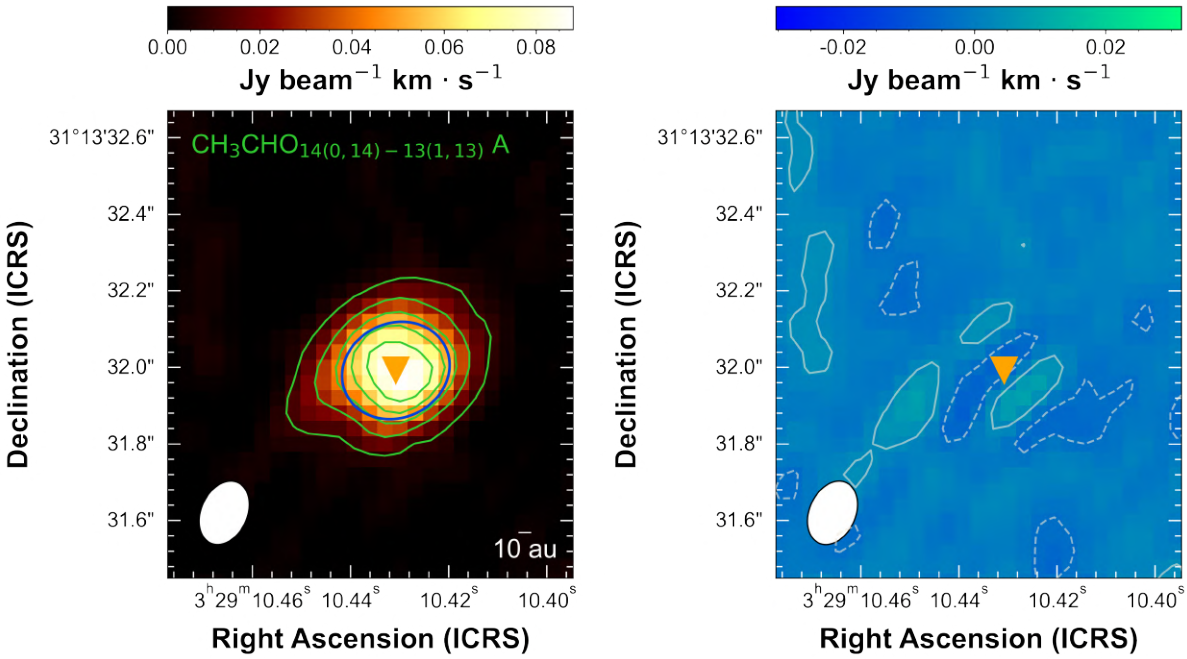

Figure C.13: Gaussian fitted moment-0 map (left) and fit residuals (right) of selected acetaldehyde line at 245 GHz. Acetaldehyde integrated emission is shown in color scale and green contours starting from 3σ of $\sim$ 10 mJy beam^{-1} km s^{-1} with step of 5σ. The fitting elliptical gaussian convolved with the beam is overlayed in blue solid line. The corresponding fit residual images are in color scale and contoured with $\pm 1\sigma$ of $\sim$ 3 mJy beam^{-1} km s^{-1} (solid and dashed line respectively). In both panels, the triangles in orange indicate the emission peaks of acetaldehyde and the beam is 0.17" x 0.12", −24°.

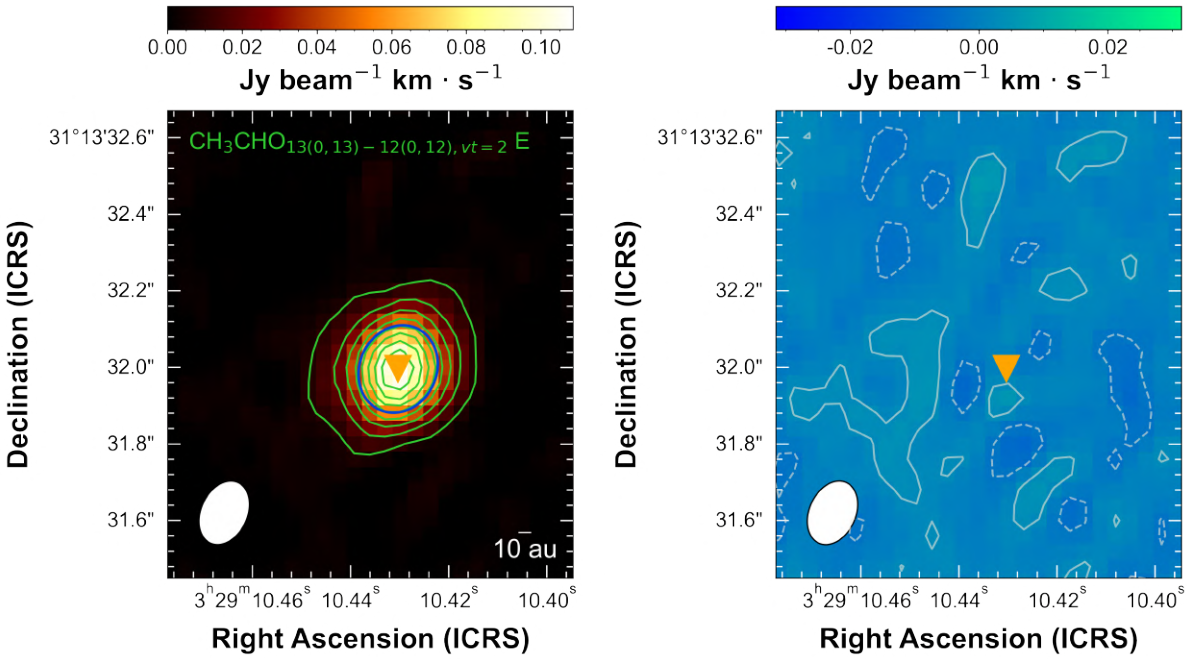

Figure C.14: Gaussian fitted moment-0 map (left) and fit residuals (right) of selected acetaldehyde line at 245 GHz. Acetaldehyde integrated emission is shown in color scale and green contours starting from 3σ of $\sim$ 8 mJy beam^{-1} km s^{-1} with step of 5σ. The fitting elliptical gaussian convolved with the beam is overlayed in blue solid line. The corresponding fit residual images are in color scale and contoured with $\pm 1\sigma$ of $\sim$ 3 mJy beam^{-1} km s^{-1} (solid and dashed line respectively). In both panels, the triangles in orange indicate the emission peaks of acetaldehyde and the beam is 0.17" x 0.12", −24°.

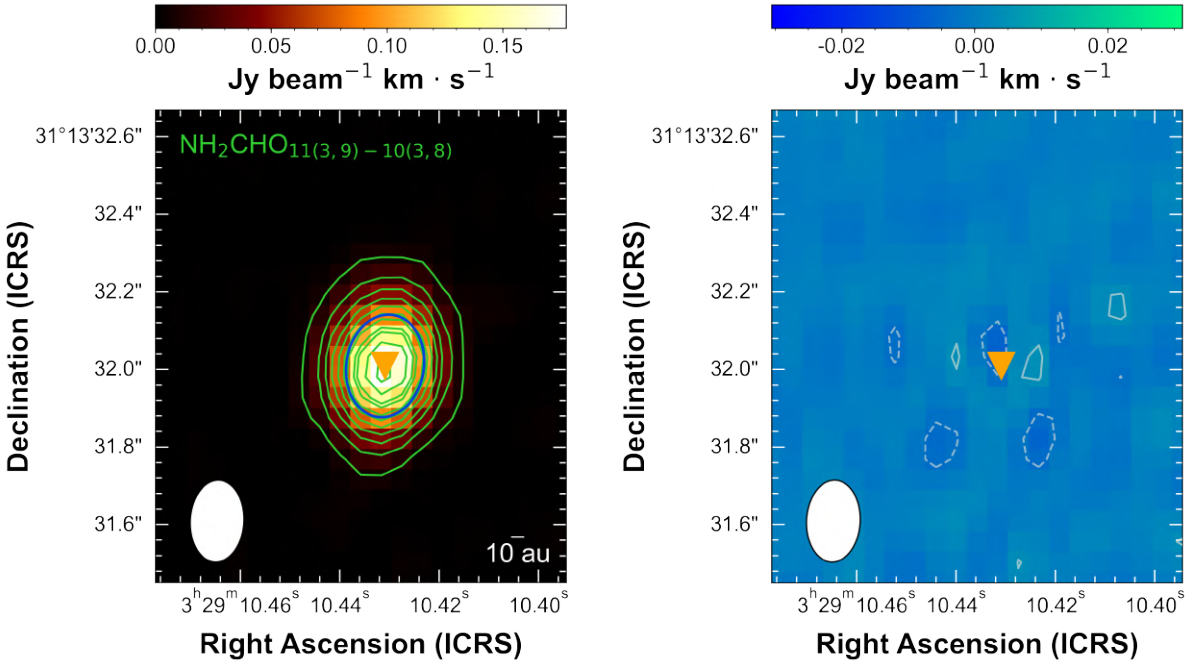

Figure C.15: Gaussian fitted moment-0 map (left) and fit residuals (right) of selected formamide line at 234 GHz. Formamide integrated emission is shown in color scale and green contours starting from 3σ of $\sim$ 9 mJy beam^{-1} km s^{-1} with step of 5σ. The fitting elliptical gaussian convolved with the beam is overlayed in blue solid line. The corresponding fit residual images are in color scale and contoured with $\pm 1\sigma$ of $\sim$ 3 mJy beam^{-1} km s^{-1} (solid and dashed line respectively). In both panels, the triangles in orange indicate the emission peaks of formamide and the beam is 0.21" x 0.14", –3°.

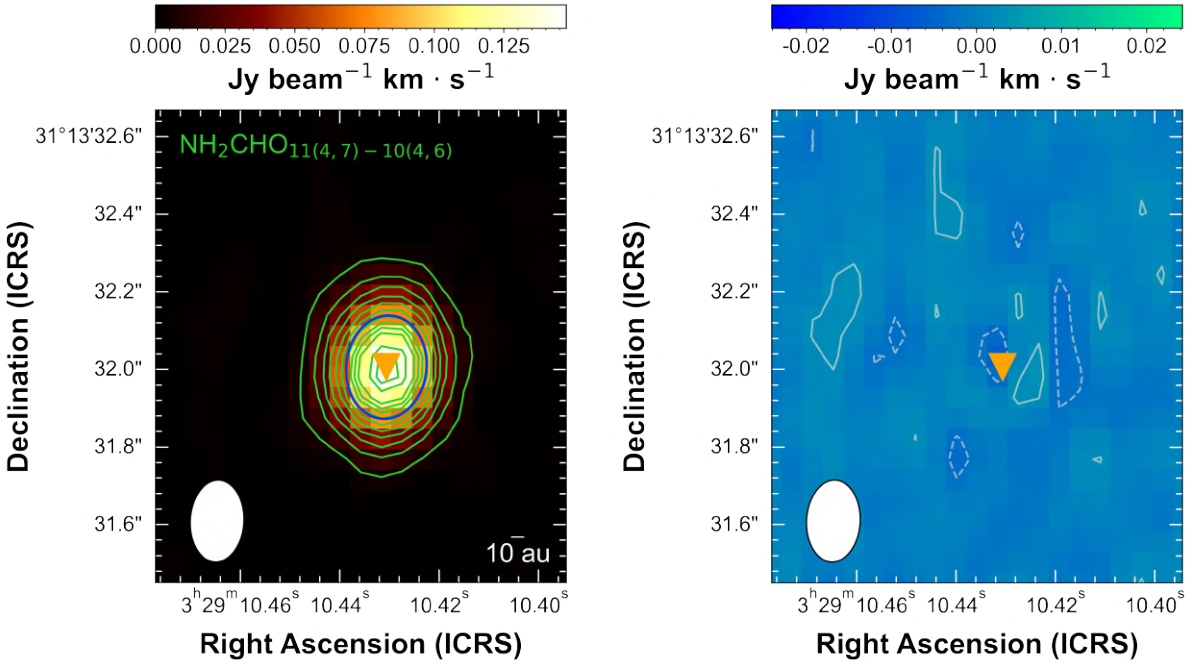

Figure C.16: Gaussian fitted moment-0 map (left) and fit residuals (right) of selected formamide line at 234 GHz. Formamide integrated emission is shown in color scale and green contours starting from 3σ of $\sim$ 7 mJy beam^{-1} km s^{-1} with step of 5σ. The fitting elliptical gaussian convolved with the beam is overlayed in blue solid line. The corresponding fit residual images are in color scale and contoured with $\pm 1\sigma$ of $\sim$ 2 mJy beam^{-1} km s^{-1} (solid and dashed line respectively). In both panels, the triangles in orange indicate the emission peaks of formamide and the beam is 0.21" x 0.14", –3°.

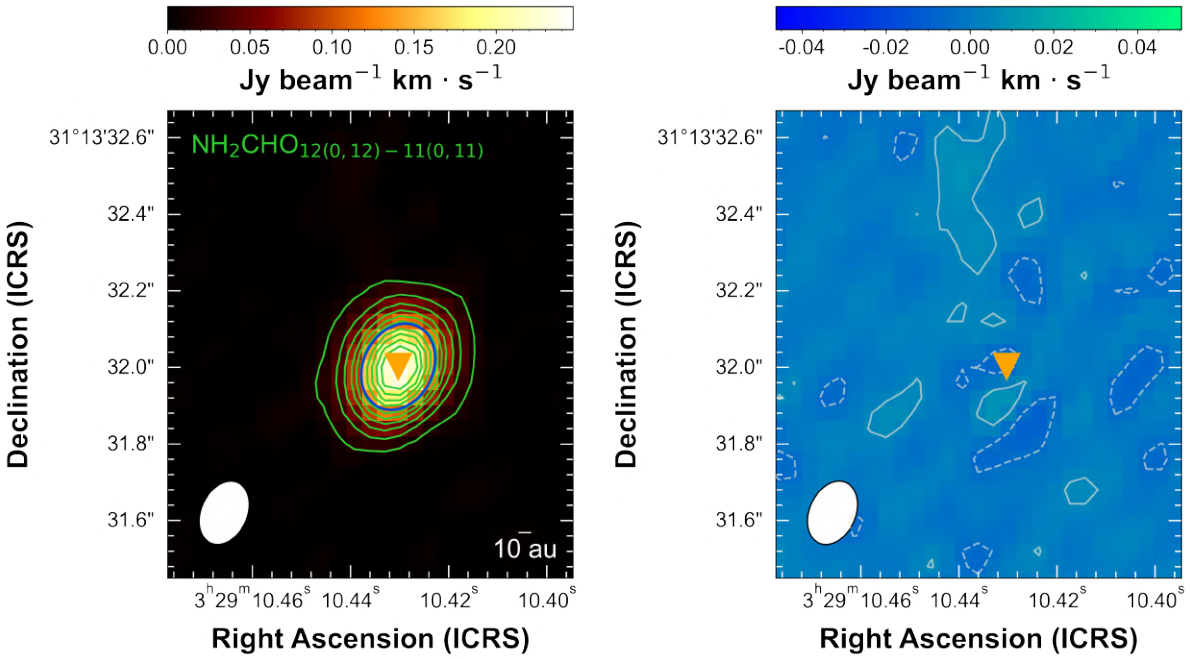

Figure C.17: Gaussian fitted moment-0 map (left) and fit residuals (right) of selected formamide line at 245 GHz. Formamide integrated emission is shown in color scale and green contours starting from 3σ of $\sim$ 13 mJy beam^{-1} km s^{-1} with step of 5σ. The fitting elliptical gaussian convolved with the beam is overlayed in blue solid line. The corresponding fit residual images are in color scale and contoured with $\pm 1\sigma$ of $\sim$ 5 mJy beam^{-1} km s^{-1} (solid and dashed line respectively). In both panels, the triangles in orange indicate the emission peaks of formamide and the beam is 0.17" x 0.12", −24°.

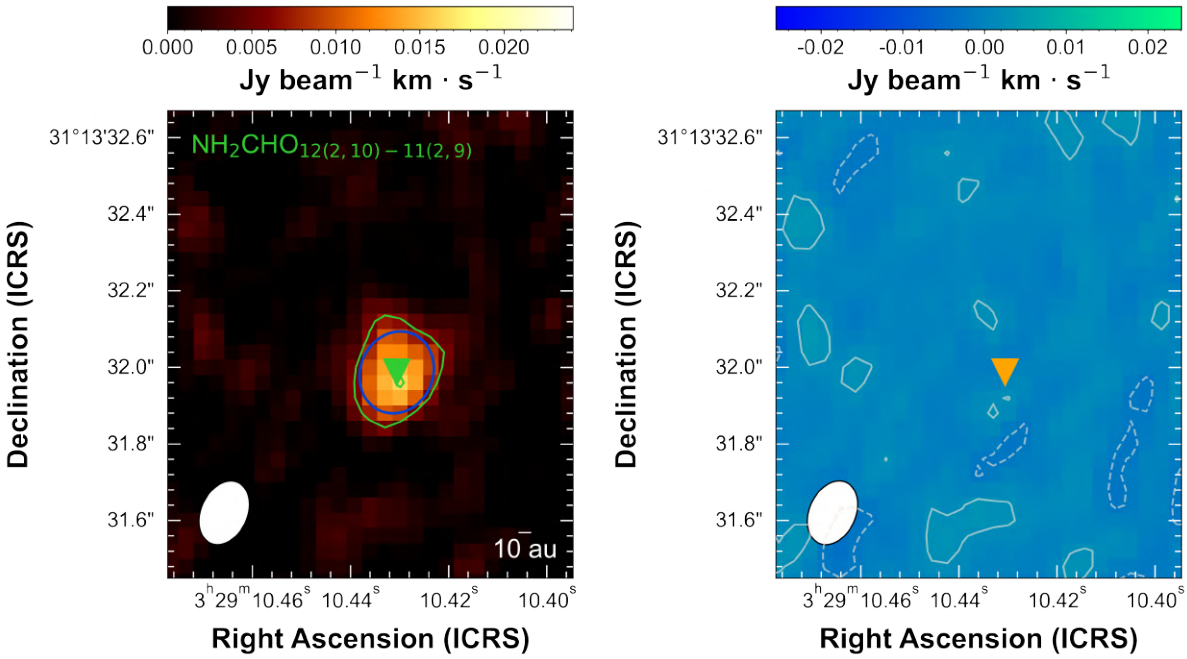

Figure C.18: Gaussian fitted moment-0 map (left) and fit residuals (right) of selected formamide line at 245 GHz. Formamide integrated emission is shown in color scale and green contours starting from 3σ of $\sim$ 6 mJy beam^{-1} km s^{-1} with step of 5σ. The fitting elliptical gaussian convolved with the beam is overlayed in blue solid line. The corresponding fit residual images are in color scale and contoured with $\pm 1\sigma$ of $\sim$ 2 mJy beam^{-1} km s^{-1} (solid and dashed line respectively). In both panels, the triangles in orange indicate the emission peaks of formamide and the beam is 0.17" x 0.12", −24°.

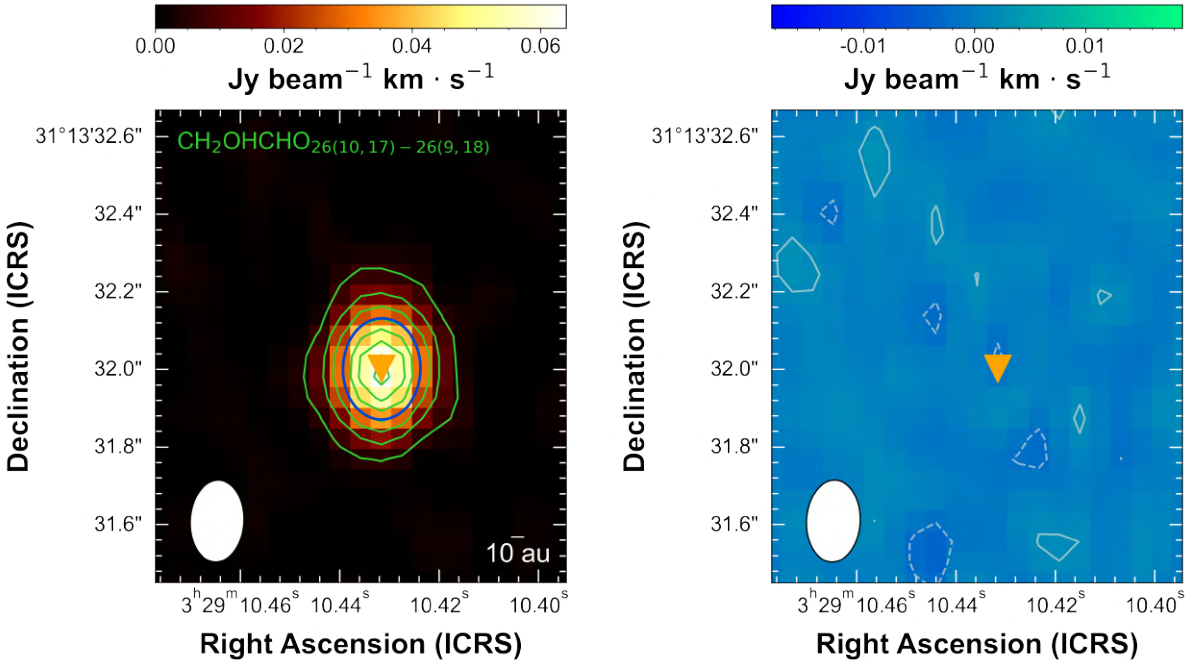

Figure C.19: Gaussian fitted moment-0 map (left) and fit residuals (right) of selected glycolaldehyde line at 234 GHz. Glycolaldehyde integrated emission is shown in color scale and green contours starting from 3σ of ~ 6 mJy beam^{-1} km s^{-1} with step of 5σ. The fitting elliptical gaussian convolved with the beam is overlayed in blue solid line. The corresponding fit residual images are in color scale and contoured with $\pm 1\sigma$ of ~ 2 mJy beam^{-1} km s^{-1} (solid and dashed line respectively). In both panels, the triangles in orange indicate the emission peaks of glycolaldehyde and the beam is 0.21" x 0.14", −3°.

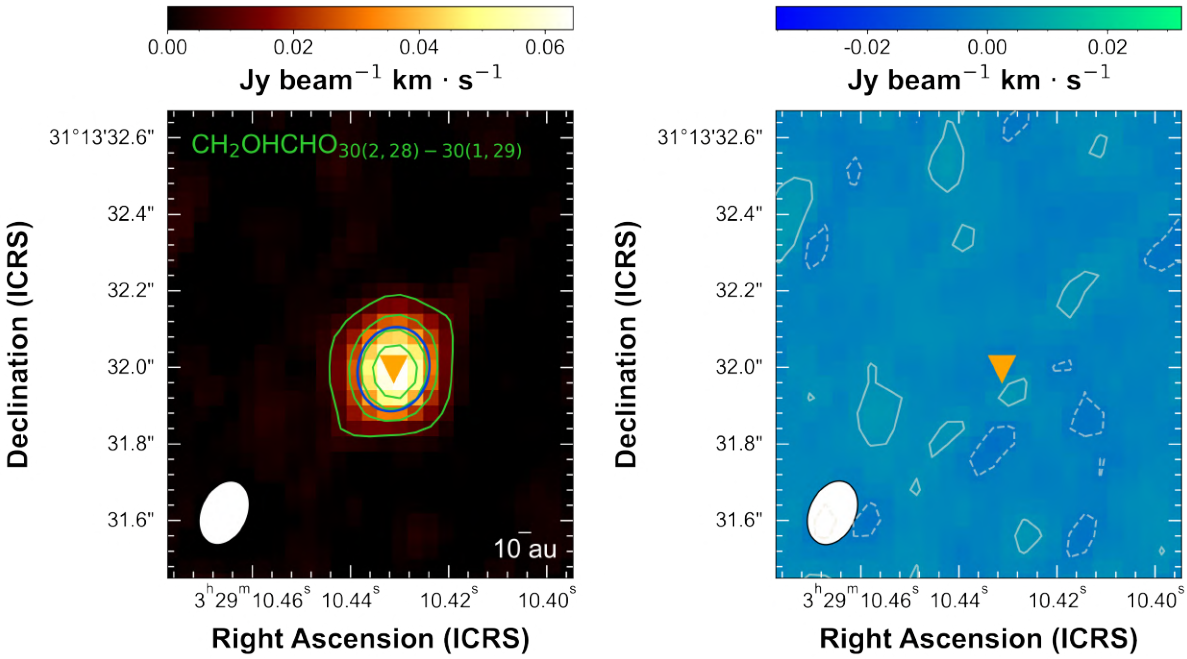

Figure C.20: Gaussian fitted moment-0 map (left) and fit residuals (right) of selected glycolaldehyde line at 245 GHz. Glycolaldehyde integrated emission is shown in color scale and green contours starting from 3σ of $\sim$ 9 mJy beam^{-1} km s^{-1} with step of 5σ. The fitting elliptical gaussian convolved with the beam is overlayed in blue solid line. The corresponding fit residual images are in color scale and contoured with $\pm 1\sigma$ of $\sim$ 3 mJy beam^{-1} km s^{-1} (solid and dashed line respectively). In both panels, the triangles in orange indicate the emission peaks of glycolaldehyde and the beam is 0.17" x 0.12", –24°.

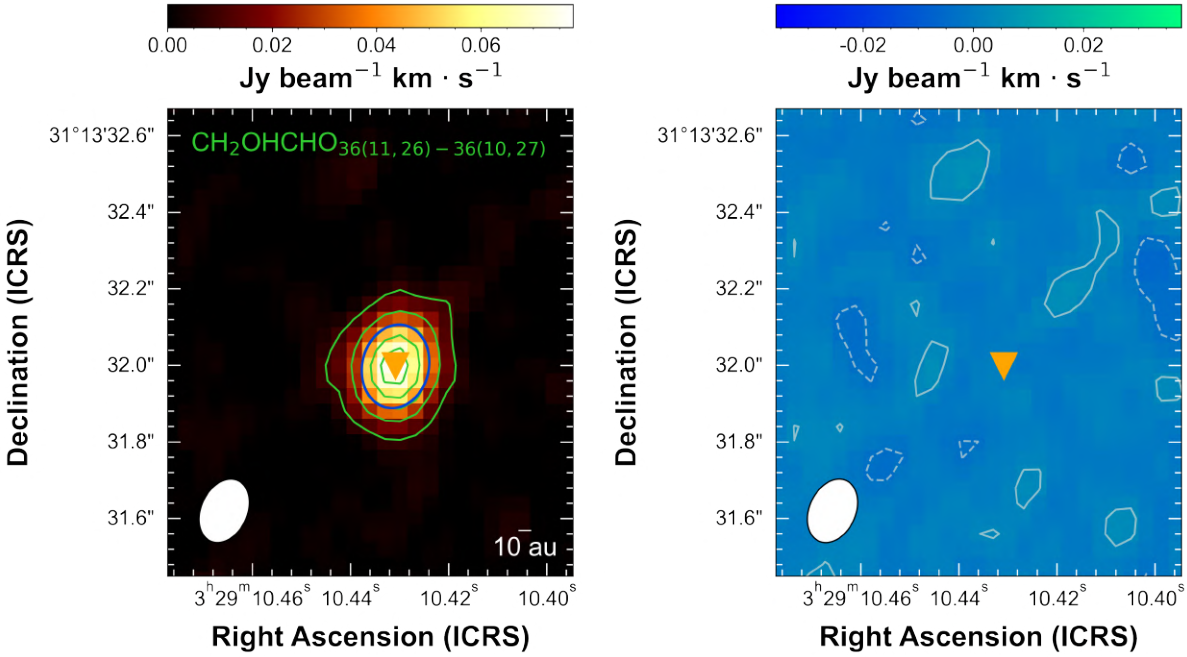

Figure C.21: Gaussian fitted moment-0 map (left) and fit residuals (right) of selected glycolaldehyde line at 245 GHz. Glycolaldehyde integrated emission is shown in color scale and green contours starting from 3σ of $\sim$ 9 mJy beam^{-1} km s^{-1} with step of 5σ. The fitting elliptical gaussian convolved with the beam is overlayed in blue solid line. The corresponding fit residual images are in color scale and contoured with $\pm 1\sigma$ of $\sim$ 3 mJy beam^{-1} km s^{-1} (solid and dashed line respectively). In both panels, the triangles in orange indicate the emission peaks of glycolaldehyde and the beam is 0.17" x 0.12", −24°.

APPENDIX D

Lines Gaussian Fits

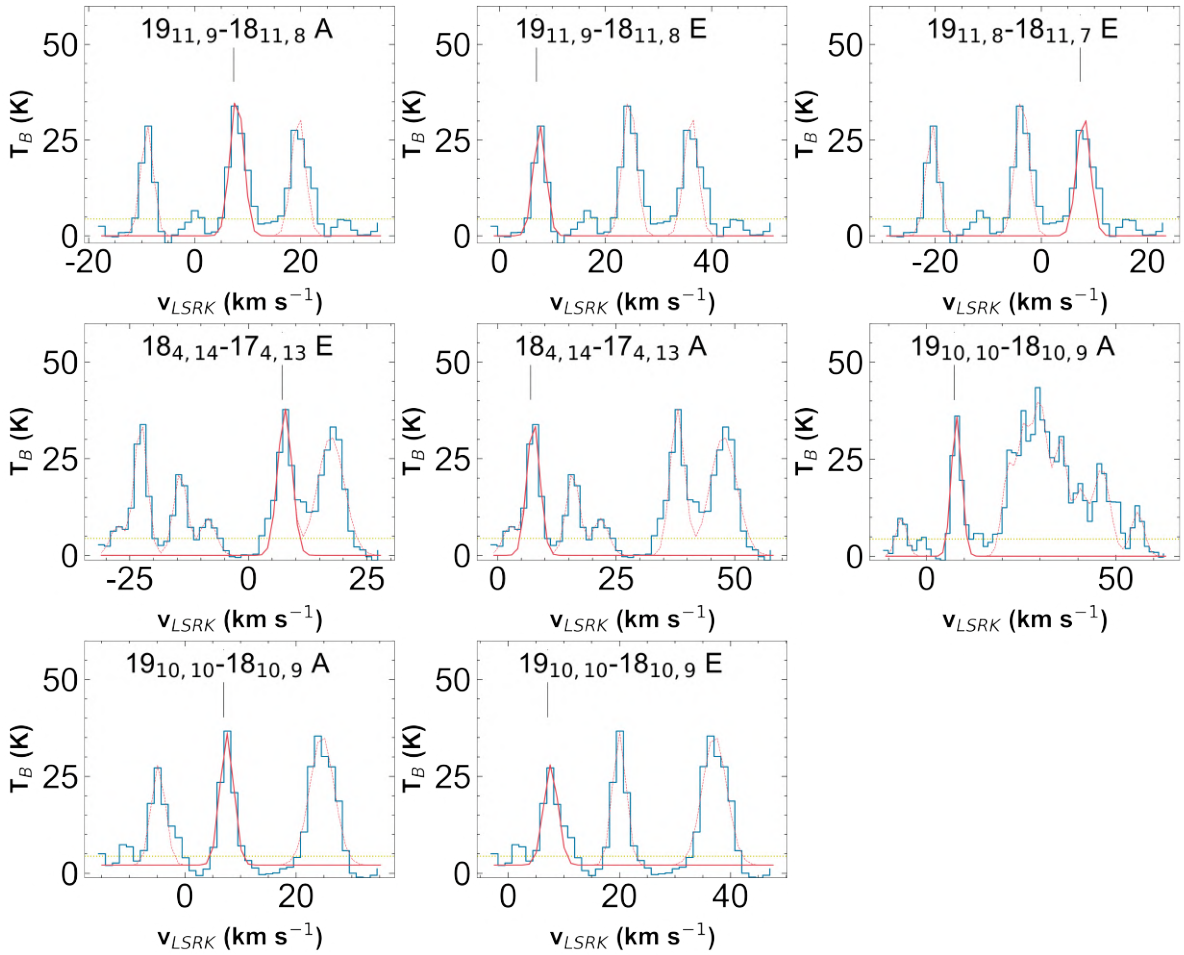

Figure D.1: Gaussian fits of identified $HCOOCH_3$ lines in IRAS 4A2 extracted spectra at 234 GHz. The spectral windows are reported in brightness temperature (T_B) as a function of radio velocity in LSRK frame (v_{LSRK}). The spectrum is in blue solid line. In red is plotted the fitting model (dashed line) and the gaussian profile of the fitted lines (solid line). Lines are labelled in black on top of each subfigure. The dotted green line stands for the threshold of 5σ, given σ = 1.16 K.

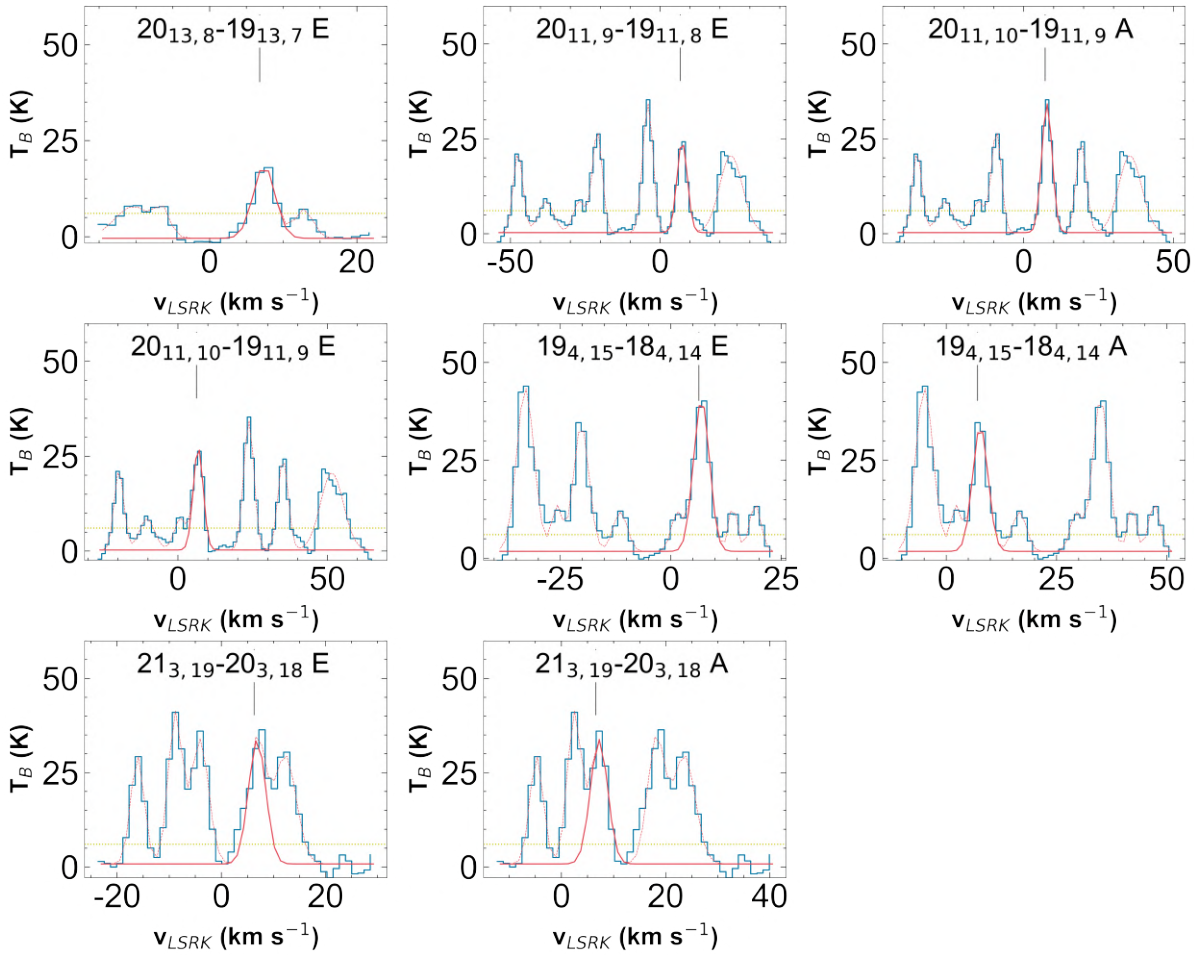

Figure D.2: Gaussian fits of identified $HCOOCH_3$ rotational lines in IRAS 4A2 extracted spectra at 245 GHz. The spectral windows are reported in brightness temperature (T_B) as a function of radio velocity in LSRK frame (v_{LSRK}). The spectrum is in blue solid line. In red is plotted the fitting model (dashed line) and the gaussian profile of the fitted lines (solid line). Lines are labelled in black on top of each subfigure. The dotted green line stands for the threshold of 5σ, given σ = 1.23 K.

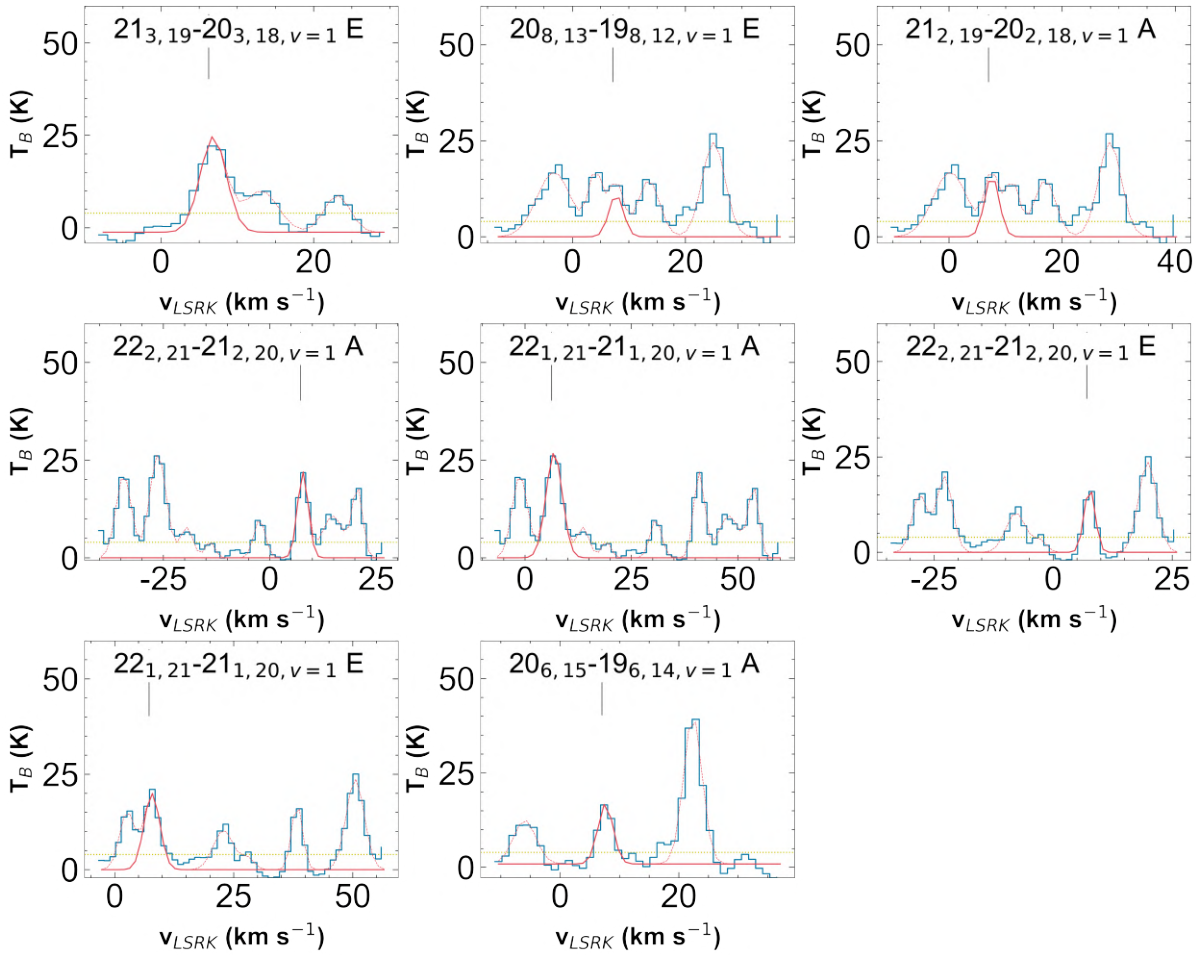

Figure D.3: Gaussian fits of identified $HCOOCH_3$ vibrational lines in IRAS 4A2 extracted spectra at 245 GHz. The spectral windows are reported in brightness temperature (T_B) as a function of radio velocity in LSRK frame (v_{LSRK}). The spectrum is in blue solid line. In red is plotted the fitting model (dashed line) and the gaussian profile of the fitted lines (solid line). Lines are labelled in black on top of each subfigure. The dotted green line stands for the threshold of 5σ, given σ = 1.23 K.

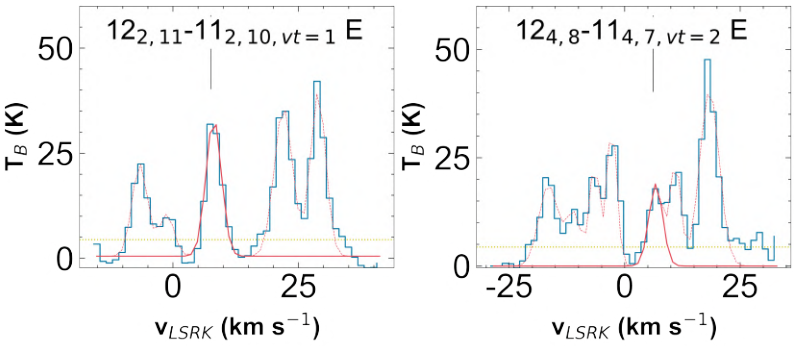

Figure D.4: Gaussian fits of identified CH_3CHO lines in IRAS 4A2 extracted spectra at 234 GHz. The spectral windows are reported in brightness temperature (T_B) as a function of radio velocity in LSRK frame (v_{LSRK}). The spectrum is in blue solid line. In red is plotted the fitting model (dashed line) and the gaussian profile of the fitted lines (solid line). Lines are labelled in black on top of each subfigure. The dotted green line stands for the threshold of 5σ, given σ = 1.16 K.

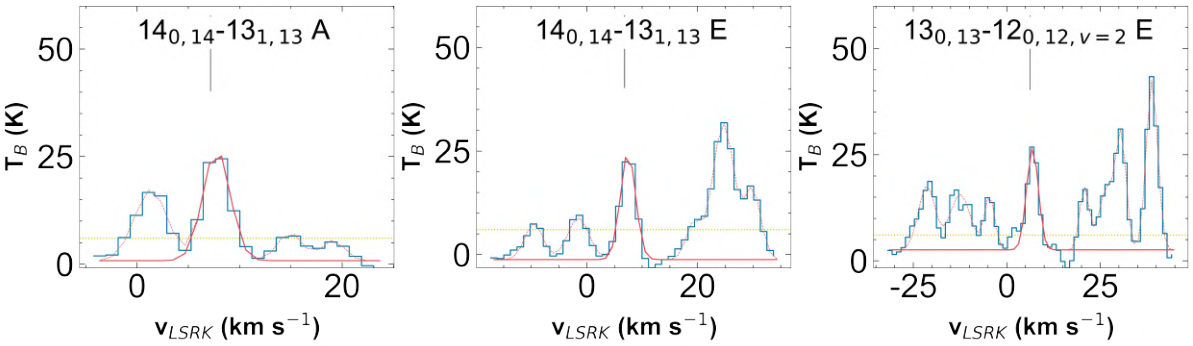

Figure D.5: Gaussian fits of identified CH_3CHO lines in IRAS 4A2 extracted spectra at 245 GHz. The spectral windows are reported in brightness temperature (T_B) as a function of radio velocity in LSRK frame (v_{LSRK}). The spectrum is in blue solid line. In red is plotted the fitting model (dashed line) and the gaussian profile of the fitted lines (solid line). Lines are labelled in black on top of each subfigure. The dotted green line stands for the threshold of 5σ, given σ = 1.23 K.

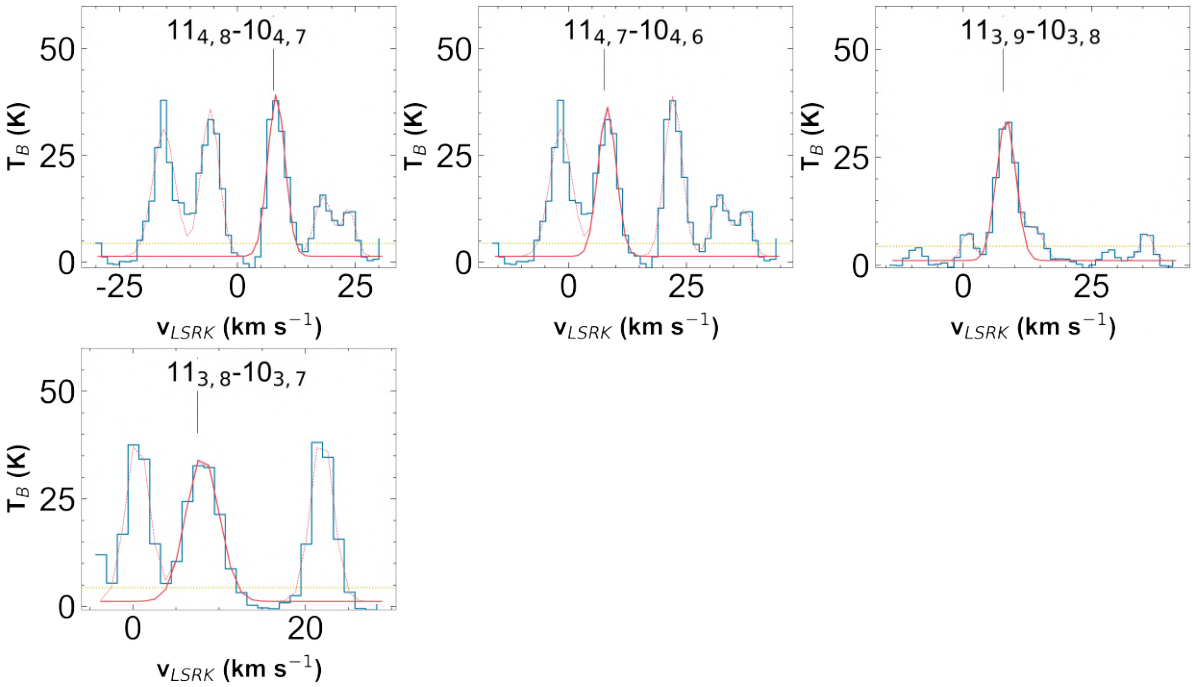

Figure D.6: Gaussian fits of identified NH_2CHO lines in IRAS 4A2 extracted spectra at 234 GHz. The spectral windows are reported in brightness temperature (T_B) as a function of radio velocity in LSRK frame (v_{LSRK}). The spectrum is in blue solid line. In red is plotted the fitting model (dashed line) and the gaussian profile of the fitted lines (solid line). Lines are labelled in black on top of each subfigure. The dotted green line stands for the threshold of 5σ, given σ = 1.16 K.

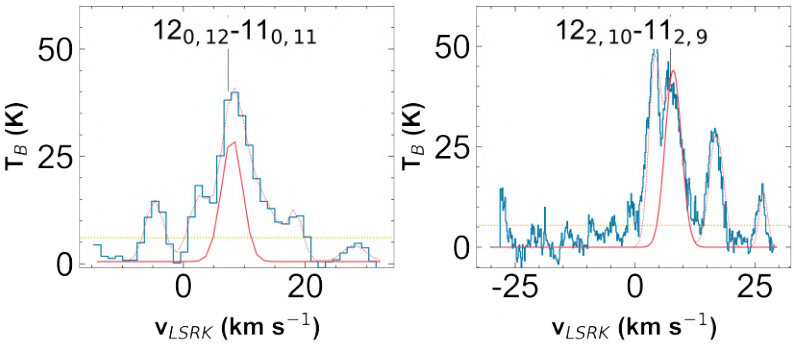

Figure D.7: Gaussian fits of identified NH_2CHO lines in IRAS 4A2 extracted spectra at 245 GHz (left) and 260 GHz (right). The spectral windows are reported in brightness temperature (T_B) as a function of radio velocity in LSRK frame (v_{LSRK}). The spectrum is in blue solid line. In red is plotted the fitting model (dashed line) and the gaussian profile of the fitted lines (solid line). Lines are labelled in black on top of each subfigure. The dotted green line stands for the threshold of 5σ, given σ = 1.23 K.

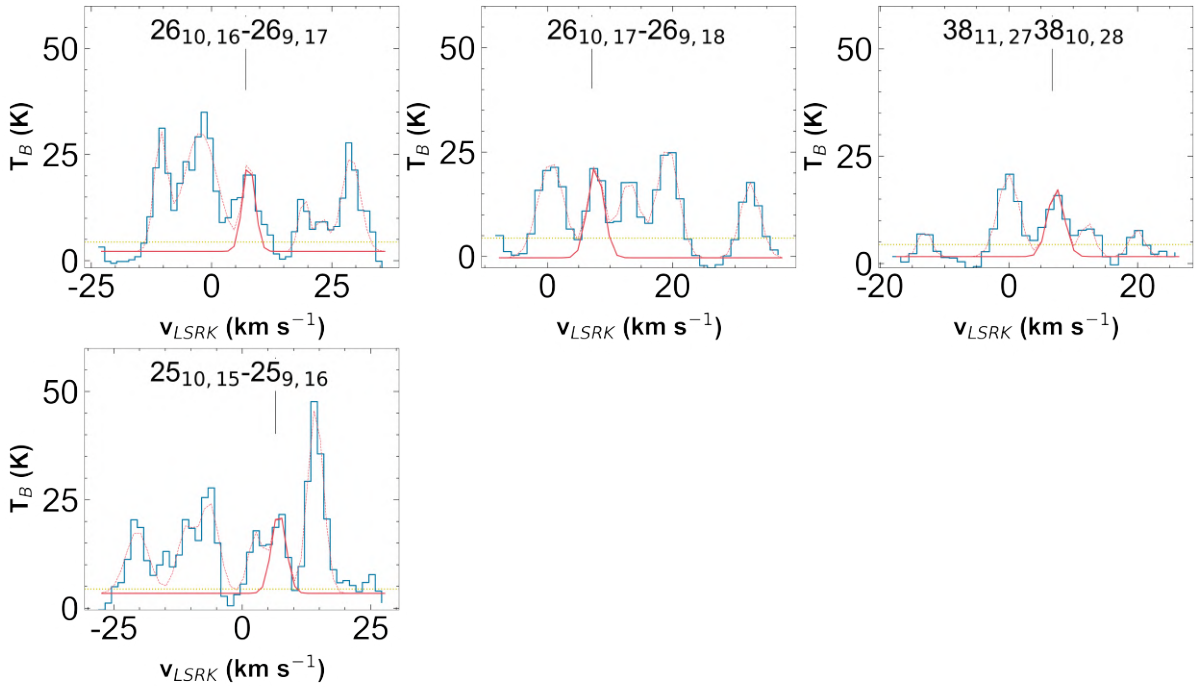

Figure D.8: Gaussian fits of identified CH_2OHCHO lines in IRAS 4A2 extracted spectra at 234 GHz. The spectral windows are reported in brightness temperature (T_B) as a function of radio velocity in LSRK frame (v_{LSRK}). The spectrum is in blue solid line. In red is plotted the fitting model (dashed line) and the gaussian profile of the fitted lines (solid line). Lines are labelled in black on top of each subfigure. The dotted green line stands for the threshold of 5σ, given σ = 1.16 K.

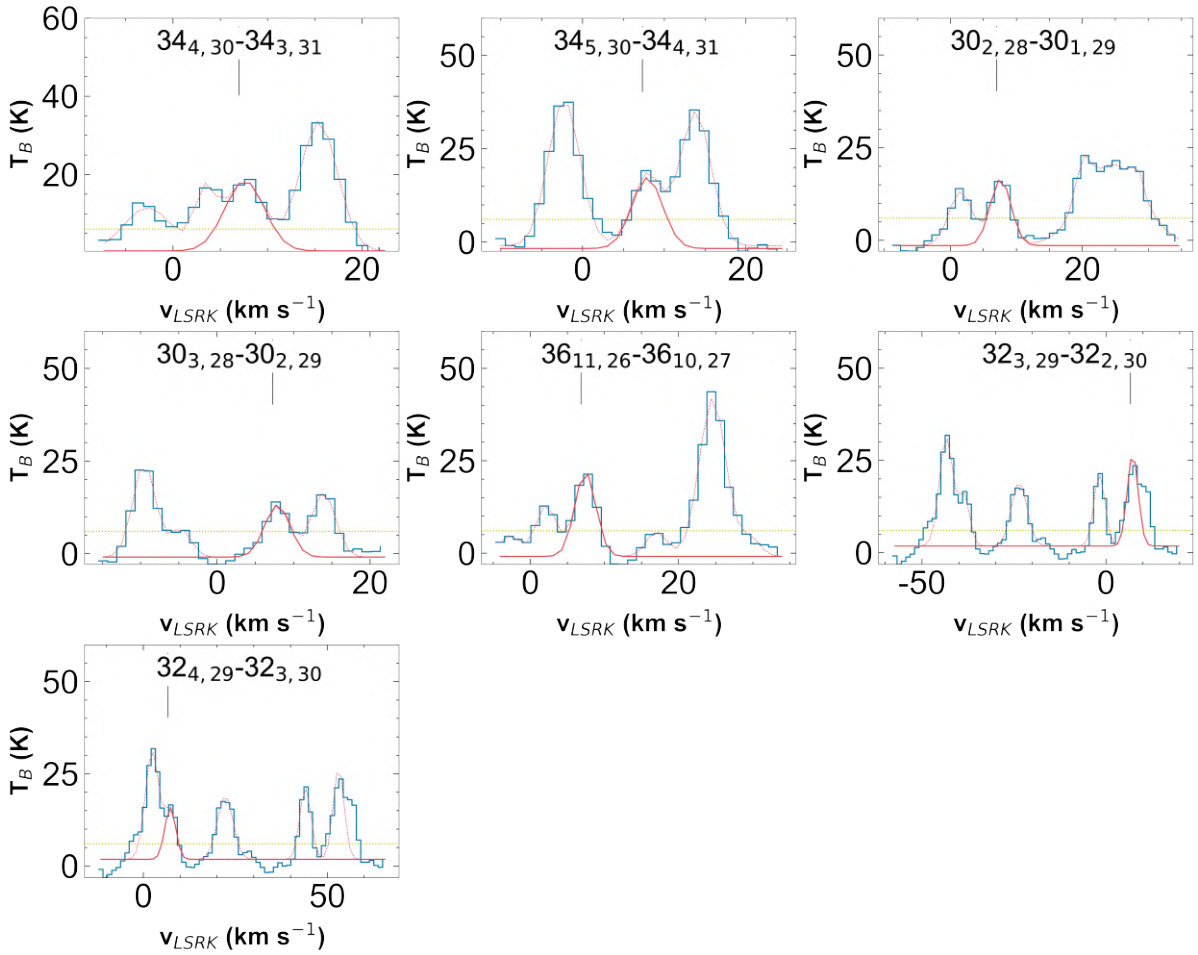

Figure D.9: Gaussian fits of identified CH_2OHCHO lines in IRAS 4A2 extracted spectra at 245 GHz. The spectral windows are reported in brightness temperature (T_B) as a function of radio velocity in LSRK frame (v_{LSRK}). The spectrum is in blue solid line. In red is plotted the fitting model (dashed line) and the gaussian profile of the fitted lines (solid line). Lines are labelled in black on top of each subfigure. The dotted green line stands for the threshold of 5σ, given σ = 1.23 K.

APPENDIX E

Gaussian Fits Parameters

Table E.1: Gaussian fit parameters of identified $HCOOCH_3$ lines in IRAS 4A2 extracted spectra. In round brackets are reported the errors associated to the parameters estimated in CARTA. Rest-frame frequencies (ν_0) and resolved quantum numbers (QNs) are taken from the CDMS catalog. The upper state energy (E_{up}) is evaluated following (3.3).

ν_0 [MHz]	QNs	E_{up} [K]	Area [K · km/s]	Velocity [km/s]	FWHM [km/s]	T_{peak} [K]	$\sigma_{cont.sub.}$ [K]
233854.286	$19_{11,9}$-$18_{11,8}$ A	193	118 (9)	7.36 (0.13)	3.0 (-)	37	-
233867.193	$19_{11,9}$-$18_{11,8}$ E	193	85 (10)	6.98 (0.16)	2.8 (0.4)	29	-
233845.233	$19_{11,8}$-$18_{11,7}$ E	193	103 (9)	7.36 (0.15)	3.0 (-)	32	-
233753.960	$18_{4,14}$-$17_{4,13}$ E	114	141 (8)	7.10 (0.12)	3.5 (-)	35	-
233777.521	$18_{4,14}$-$17_{4,13}$ A	114	118 (10)	6.97 (0.13)	3.2 (0.3)	35	-
233226.788	$19_{4,16}$-$18_{4,15}$ A	123	121 (10)	7.34 (0.13)	3.2 (0.3)	36	-
234124.883	$19_{10,10}$-$18_{10,9}$ A	179	105 (9)	6.95 (0.11)	2.9 (0.3)	34	0.6
234134.600	$19_{10,10}$-$18_{10,9}$ E	179	83 (8)	7.07 (0.15)	3.0 (-)	26	0.6
245903.680	$20_{13,8}$-$19_{13,7}$ E	236	71 (4)	6.80 (0.07)	3.5 (0.19)	19	0.3
246285.400	$20_{11,9}$-$19_{11,8}$ E	204	103 (8)	6.75 (0.13)	4.1 (0.3)	24	0.5
246295.135	$20_{11,10}$-$19_{11,9}$ A	204	147 (8)	7.13 (0.08)	4.1 (0.2)	34	0.5
246308.272	$20_{11,10}$-$19_{11,9}$ E	204	112 (8)	6.20 (0.12)	3.9 (0.3)	27	0.5
246891.611	$19_{4,15}$-$18_{4,14}$ E	126	176 (11)	6.35 (0.10)	4.2 (0.3)	39	0.7
246914.658	$19_{4,15}$-$18_{4,14}$ A	126	141 (11)	7.12 (0.12)	4.1 (0.3)	32	0.7
247044.146	$21_{3,19}$-$20_{3,18}$ E	140	138 (9)	6.27 (0.16)	3.9 (-)	33	0.6
247053.453	$21_{3,19}$-$20_{3,18}$ A	140	138 (8)	6.61 (0.14)	3.9 (-)	33	0.6
				v = 1			
245846.914	$21_{3,19}$-$20_{3,18}$ E	327	57 (18)	6.24 (0.16)	4.0 (-)	26	0.5
246184.177	$20_{8,13}$-$19_{8,12}$ E	354	46 (6)	7.14 (-)	3.0 (-)	11	-
246187.016	$21_{2,19}$-$20_{2,18}$ A	327	52 (6)	6.97 (-)	3.0 (-)	16	-
246461.167	$22_{2,21}$-$21_{2,20}$ A	331	69 (4)	7.17 (0.11)	3.0 (0.3)	22	-
246488.433	$22_{1,21}$-$21_{1,20}$ A	331	130 (6)	6.24 (0.09)	4.5 (0.2)	27	-
246706.504	$22_{2,21}$-$21_{2,20}$ E	330	50 (5)	7.06 (0.13)	2.7 (0.3)	17	-
246731.729	$22_{1,21}$-$21_{1,20}$ E	330	83 (9)	7.23 (0.19)	3.9 (0.5)	20	-
246985.225	$20_{6,15}$-$19_{6,14}$ A	336	51 (7)	7.04 (0.17)	3.0 (0.4)	16	0.5

Table E.2: Gaussian fit parameters of identified CH_3CHO lines in IRAS 4A2 extracted spectra. In round brackets are reported the errors associated to the parameters estimated in CARTA. Rest-frame frequencies (ν_0) and resolved quantum numbers (QNs) are taken from the CDMS catalog. The upper state energy (E_{up}) is evaluated following (3.3).

ν_0 [MHz]	QNs	E_{up} [K]	Area [K · km/s]	Velocity [km/s]	FWHM [km/s]	T_{peak} [K]	$\sigma_{cont.sub.}$ [K]
233048.516	$12_{2,11}$-$11_{2,10vt=1}$ E	285	126 (9)	7.52 (0.11)	3.6 (0.3)	33	0.5
234707.132	$12_{4,8}$-$11_{4,7,vt=2}$ E	487	71 (10)	6.1 (0.3)	3.5 (-)	19	-
247142.155	$14_{0,14}$-$13_{1,13}$ A	96	85 (4)	7.16 (0.05)	3.1 (0.1)	26	0.4
247341.332	$14_{0,14}$-$13_{1,13}$ E	96	89 (3)	6.84 (0.05)	3.2 (0.1)	26	0.3
246169.217	$13_{0,13}$-$12_{0,12,vt=2}$ E	461	95 (8)	6.29 (0.12)	3.7 (0.3)	24	0.5

Table E.3: Gaussian fit parameters of identified NH_2CHO lines in IRAS 4A2 extracted spectra. In round brackets are reported the errors associated to the parameters estimated in CARTA. Rest-frame frequencies (ν_0) and resolved quantum numbers (QNs) are taken from the CDMS catalog. The upper state energy (E_{up}) is evaluated following (3.3).

ν_0 [MHz]	QNs	E_{up} [K]	Area [K · km/s]	Velocity [km/s]	FWHM [km/s]	T_{peak} [K]	$\sigma_{cont.sub.}$ [K]
233735.6031	$11_{4,8}$-$10_{4,7}$	80	174 (13)	7.66 (0.13)	4.3 (0.3)	38	0.8
233746.5043	$11_{4,7}$-$10_{4,6}$	80	165 (10)	7.58 (0.14)	4.5 (-)	35	0.8
233897.3180	$11_{3,9}$-$10_{3,8}$	94	157 (6)	7.76 (0.08)	4.41 (0.19)	33	0.2
234316.2540	$11_{3,8}$-$10_{3,7}$	94	160 (14)	7.48 (0.14)	4.5 (0.4)	34	1.0
247391.356	$12_{0,12}$-$11_{0,11}$	78	133 (109)	7.37 (0.18)	4 (1)	29	0.5
260189.848	$12_{2,10}$-$11_{2,9}$	92	186 (5)	7.37 (-)	4 (-)	44	-

Table E.4: Gaussian fit parameters of identified CH_2OHCHO lines in IRAS 4A2 extracted spectra. In round brackets are reported the errors associated to the parameters estimated in CARTA. Rest-frame frequencies (ν_0) and resolved quantum numbers (QNs) are taken from the CDMS catalog. The upper state energy (E_{up}) is evaluated following (3.3).

ν_0 [MHz]	QNs	E_{up} [K]	Area [K · km/s]	Velocity [km/s]	FWHM [km/s]	T_{peak} [K]	$\sigma_{cont.sub.}$ [K]
233587.4488	$26_{10,16}$-$26_{9,17}$	256	68 (9)	7.1 (0.2)	3.0 (-)	21	0.9
233709.6570	$26_{10,17}$-$26_{9,18}$	256	72 (6)	7.12 (0.11)	3.0 (0.3)	22	0.6
234554.5664	$38_{11,27}$-$38_{10,28}$	489	50 (4)	6.72 (0.14)	3.0 (-)	16	0.4
234704.0652	$25_{10,15}$-$25_{9,16}$	242	65 (11)	6.5 (0.2)	3.3 (0.6)	19	0.8
246395.021	$34_{4,30}$-$34_{3,31}$	344	94 (10)	7.98 (0.17)	4.9 (0.5)	18	0.9
246605.634	$34_{5,30}$-$34_{4,31}$	344	90 (7)	7.32 (0.15)	4.4 (0.4)	19	0.4
246773.215	$30_{2,28}$-$30_{1,29}$	253	70 (4)	7.01 (0.09)	3.7 (0.2)	18	0.4
246778.410	$30_{3,28}$-$30_{2,29}$	253	60 (6)	7.27 (0.16)	3.9 (0.4)	14	0.4
246852.728	$36_{11,26}$-$36_{10,27}$	446	81 (7)	6.83 (0.10)	3.3 (0.2)	23	0.7
247285.434	$32_{3,29}$-$32_{2,30}$	297	92 (9)	6.64 (-)	3.5 (-)	25	0.6
247323.158	$32_{4,29}$-$32_{3,30}$	297	50 (11)	6.73 (-)	3.5 (-)	14	0.6

Bibliography

ALMA, P. 2015, The Astrophysical Journal, 808, L3

André, P. & Montmerle, T. 1994, The Astrophysical Journal, 420, 837

André, P., et al. 2000, in Protostars and Planets IV, ed. V. Mannings, A. P. Boss, & S. S. Russell, 59

Andrews, S. M. & Williams, J. P. 2005, The Astrophysical Journal, 631, 1134

Andrews, S. M. & Williams, J. P. 2007, The Astrophysical Journal, 659, 705

Ascenzi, D., et al. 2019, A&A, 625, A72

Baars, J. W. M., et al. 1977, A&A, 61, 99

Bachiller, R. & Cernicharo, J. 1986, A&A, 166, 283

Balucani, N., et al. 2015, Monthly Notices of the Royal Astronomical Society, Letters, 449, L16

Beckwith, S. V. W., et al. 1990, The Astronomical Journal, 99, 924

Belloche, A., et al. 2020, A&A, 635, A198

Bely, Y. P. 2003, The Design and Construction of Large Optical Telescopes (Springer New York)

Bergin, E. A. & Tafalla, M. 2007, Annual Review of Astronomy and Astrophysics, 45, 339

Bianchi, E., et al. 2022, A&A, 662, A103

Bianchi, E., et al. 2020, Monthly Notices of the Royal Astronomical Society: Letters, 498, L87

Blake, G. A., et al. 1987, The Astrophysical Journal, 315, 621

Böhm-Vitense, E. 1989, Introduction to Stellar Astrophysics: Volume 2 (Cambridge University Press)

Boogert, A. C. A., et al. 2015, Annual Review of Astronomy and Astrophysics, 53, 541

Bottinelli, S., et al. 2004a, The Astrophysical Journal, 615, 354

Bottinelli, S., et al. 2004b, Astrophysical Journal, Letters, 617, L69

Bottinelli, S., et al. 2007, A&A, 463, 601

Carpenter, J. M. 2000, The Astronomical Journal, 120, 3139

Carroll, B. W. & Ostlie, D. A. 2007, An Introduction to Modern Astrophysics, 2nd edn., ed. S. F. P. Addison-Wesley

Caselli, P. 1999, in The Physics and Chemistry of the Interstellar Medium, ed. V. Ossenkopf, J. Stutzki, & G. Winnewisser, 395

Caselli, P. & Ceccarelli, C. 2012, The Astronomy and Astrophysics Review, 20

Caselli, P., et al. 2012, The Astrophysical Journal Letters, 759, L37

Cazaux, S., et al. 2003, The Astrophysical Journal, 593, L51

Ceccarelli, C. 2004, Astronomical Society of the Pacific Conference Series, 323, 195

Ceccarelli, C. 2014, 859

Ceccarelli, C., et al. 2017, The Astrophysical Journal, 850, 176

Ceccarelli, C., et al. 2007, in Protostars and Planets V, ed. B. Reipurth, D. Jewitt, & K. Keil, 47

Ceccarelli, C., et al. 2000a, A&A, 355, 1129

Ceccarelli, C., et al. 2022, PPVII

Ceccarelli, C., et al. 1996, The Astrophysical Journal, 471, 400

Ceccarelli, C., et al. 2000b, A&A, 362, 1122

Chapin, I. S. F., et al. 2011, Principles of Terrestrial Ecosystem Ecology (Springer)

Charnley, S. B., et al. 1992, The Astrophysical Journal, Letters, 399, L71

Choi, M. 2001, The Astrophysical Journal, 553, 219

Choi, M. 2005, The Astrophysical Journal, 630, 976

Choi, M., et al. 1999, Astrophysical Journal, Supplement, 122, 519

Clark, B. G. 1980, A&A, 89, 377

Codella, C., et al. 2016, A&A, 586, L3

Codella, C., et al. 2021, Front. Astron. Space Sci., 8, 227

Codella, C., et al. 2015, Monthly Notices of the RAS, 449, L11

Codella, C., et al. 2022, Monthly Notices of the RAS, 515, 543

Codella, C., et al. 2017, A&A, 605, L3

Collins, P. & Ferrier, R. 1997, Journal of Steroid Biochemistry and Molecular Biology, 4, 351

Condon, J. J. & Ransom, S. M. 2016, Essential Radio Astronomy

Connelley, M. S. & Reipurth, B. 2018, The Astrophysical Journal, 861, 145

Cordiner, M. A., et al. 2012, The Astrophysical Journal, 744, 131

Cornwell, T. J., et al. 2008, IEEE Journal of Selected Topics in Signal Processing, 2, 647

Crimier, N., et al. 2010, A&A, 516, A102

Dalgarno, A. 2008, Annual Review of Astronomy and Astrophysics, 46, 1

Dawson, R. I. & Johnson, J. A. 2018, Annual Review of Astron. and Astrophys., 56, 175

De Simone, M., et al. 2020, Astrophysical Journal, Letters, 896, L3

De Simone, M., et al. 2022, Monthly Notices of the RAS, 512, 5214

De Simone, M., et al. 2017, A&A, 599, A121

Dhabal, A., et al. 2018, The Astrophysical Journal, 853, 169

Dhabal, A., et al. 2019, The Astrophysical Journal, 876, 108

Di Francesco, J., et al. 2001, The Astrophysical Journal, 562, 770

Doglioni, C., et al. 2016, Geoscience Frontiers, 7, 865

Draine, B. T. 2011, Physics of the Interstellar and Intergalactic Medium

Drake, G. 2006, Springer Handbook of Atomic, Molecular, and Optical Physics, Springer Handbook of Atomic, Molecular, and Optical Physics No. v. 1 (Springer)

EHT, C. 2019, The Astrophysical Journal Letters, 875, L1

Einstein, A. 1916, Deutsche Physikalische Gesellschaft, 18, 318

Eistrup, Christian, et al. 2018, A&A, 613, A14

Fedele, D., et al. 2018, A&A, 610, A24

Flower, D. R. & Pineau des Forets, G. 1994, Monthly Notices of the Royal Astronomical Society, 268, 724

Fontani, F., et al. 2014, The Astrophysical Journal Letters, 788, L43

Frank, A., et al. 2014, in Protostars and Planets VI, ed. H. Beuther, R. S. Klessen, C. P. Dullemond, & T. Henning, 451–474

Froebrich, D., et al. 2003, Monthly Notices of the Royal Astronomical Society, 346, 163

Garrod, R. T., et al. 2022, The Astrophysical Journal Supplement Series, 259, 1

Garrod, R. T. & Herbst, E. 2006, A&A, 457, 927

Garrod, R. T., et al. 2007, A&A, 467, 1103

Gillessen, S., et al. 2009, The Astrophysical Journal, 692, 1075

Goldsmith, P. F. & Langer, W. D. 1999, The Astrophysical Journal, 517, 209

Gordy, W. & Cook, R. L. R. L. 1970, Microwave molecular spectra / Walter Gordy, Robert L. Cook, Technique of organic chemistry ; v. 9pt. 2 (New York: Interscience)

Griffiths, D. J. & Schroeter, D. F. 2018, Introduction to Quantum Mechanics, 3rd edn. (Cambridge University Press)

Herbst, E. & van Dishoeck, E. F. 2009, Annual Reviews of Astronomy and Astrophysics, 47, 427

Högbom, J. A. 1974, A&A, supplement, 15, 417

Imai, M., et al. 2022, Astrophysical Journal, 934, 70

Ioppolo, S., et al. 2011, Monthly Notices of the Royal Astronomical Society, 413, 2281

Isella, A. & Turner, N. 2016, arXiv e-prints, arXiv:1608.05123

Ishak, B. 2019, Contemporary Physics, 60, 335

Jacobsen, S. K., et al. 2018, A&A, 612, A72

Jennings, R. E., et al. 1987, Monthly Notices of the RAS, 226, 461

Jin, M. & Garrod, R. T. 2020, The Astrophysical Journal Supplement Series, 249, 26

Jørgensen, J. K., et al. 2011, A&A, 534, A100

Jørgensen, J. K., et al. 2012, The Astrophysical Journal, 757, L4

Jørgensen, J. K., et al. 2002, A&A, 389, 908

Jørgensen, J. K., et al. 2016, A&A, 595, A117

Kahane, C., et al. 2013, The Astrophysical Journal Letters, 763, L38

Karska, A., et al. 2013, A&A, 552, A141

Kasting, J. F., et al. 1993, Icarus, 101, 108

Kepley, A. A., et al. 2020, Publications of the Astronomical Society of the Pacific, 132, 024505

Keto, E. & Caselli, P. 2010, Monthly Notices of the Royal Astronomical Society, 402, 1625

Keto, E., et al. 2015, Monthly Notices of the Royal Astronomical Society, 446, 3731

Kirk, H., et al. 2006, The Astrophysical Journal, 646, 1009

Klessen, R. & Glover, S. 2014, 43

Kristensen, L. E., et al. 2012, A&A, 542, A8

Kuiper, G. P. 1951, Proceedings of the National Academy of Science, 37, 1

Kurtz, S., et al. 2000, in Protostars and Planets IV, ed. V. Mannings, A. P. Boss, & S. S. Russell, 299–326

Lada, C. J. 1987, in Star Forming Regions, ed. M. Peimbert & J. Jugaku, Vol. 115, 1

Lahuis, F., et al. 2006, Astrophysical Journal, Letters, 636, L145

Landau, L. D. & Lifshitz, E. M. 1969, Statistical physics. Pt.1

Larralde, R., et al. 1995, Proceedings of the National Academy of Sciences, 92, 8158

Lee, J.-E., et al. 2023

Lefloch, B., et al. 2017, Monthly Notices of the RAS, 469, L73

Lefloch, B., et al. 1996, A&A, 313, L17

Leger, A., et al. 1985, A&A, 144, 147

Leroy, A. K., et al. 2021, The Astrophysical Journal Supplement Series, 257, 43

Li, Z. Y., et al. 2014, 173

Looney, L. W., et al. 2000, The Astrophysical Journal, 529, 477

López-Sepulcre, A., et al. 2017, A&A, 606, A121

Loren, R. B. 1976, The Astrophysical Journal, 209, 466

Luhman, K. L., et al. 2016, The Astrophysical Journal, 827, 52

López-Sepulcre, A., et al. 2019, ACS Earth and Space Chemistry, 3, 2122

Maas, R., et al. 1992, Geochimica et Cosmochimica Acta, 56, 1281

Madhusudhan, N., et al. 2016, Space Science Reviews, 205, 285

Manigand, S., et al. 2020, A&A, 635, A48

Maret, S., et al. 2002, A&A, 395, 573

Maureira, M. J., et al. 2022, The Astrophysical Journal, 941, L23

McClure, M., et al. 2023, Nature Astronomy, 7, 1

McGuire, B. A. 2018, The Astrophysical Journal Supplement Series, 239, 17

McGuire, B. A. 2022, The Astrophysical Journal Supplement Series, 259, 30

McNaught, A. D. & Wilkinson, A. 1997, Compendium of Chemical Terminology, 2nd edn., IUPAC Nomenclature Books Series ("Color Books") (Oxford: Blackwell Science)

Millar, T., et al. 1991, The Astrophysical Journal, 369, 147

Miotello, A., et al. 2023, in Astronomical Society of the Pacific Conference Series, Vol. 534, Protostars and Planets VII, ed. S. Inutsuka, Y. Aikawa, T. Muto, K. Tomida, & M. Tamura, 501

Motte, F. & André, P. 2001, A&A, 365, 440

Noriega-Crespo, A., et al. 2004, The Astrophysical Journal Supplement Series, 154, 402

Öberg, K. I. & Bergin, E. A. 2021, Physics Reports, 893, 1

Öberg, K. I., et al. 2011, The Astrophysical Journal, 740, 14

Okoda, Y., et al. 2023, arXiv e-prints, arXiv:2303.03564

P., A. 2002, EAS Publications Series, 3, 1

Palla, F., et al. 1993, A&A, 280, 599

Pety, J., et al. 2013, The Astrophysical Journal, 779, 43

PICKETT, H., et al. 1998, Journal of Quantitative Spectroscopy and Radiative Transfer, 60, 883

Piso, A.-M. A., et al. 2015, The Astrophysical Journal, 815, 109

Plunkett, A., et al. 2023, Publications of the ASP, 135, 034501

Pokhrel, R., et al. 2017, The Astrophysical Journal, 853

Reipurth, B., et al. 2002, The Astronomical Journal, 124, 1045

Rimola, A., et al. 2014, A&A, 572, A70

Rubin, R. H., et al. 1971, Astrophysical Journal, Letters, 169, L39

Sahu, D., et al. 2019, The Astrophysical Journal, 872, 196

Sakai, N. & Yamamoto, S. 2013, Chemical Reviews, 113, 8981

Saladino, R., et al. 2012, Biochimie, 94, 1451, rNA in all its forms

Santangelo, G., et al. 2015, A&A, 584, A126

Sargent, A. I. 1979, The Astrophysical Journal, 233, 163

Schöier, F. L., et al. 2002, A&A, 390, 1001

Schwab, F. R. & Cotton, W. D. 1983, The Astronomical Journal, 88, 688

Segura-Cox, D., et al. 2020, Nature, 586, 228

Shen, C. J., et al. 2004, A&A, 415, 203

Shu, F. H. 1977, The Astrophysical Journal, 214, 488

Shu, F. H., et al. 1987, Annual Review of Astron. and Astrophys., 25, 23

Skouteris, D., et al. 2017, Monthly Notices of the Royal Astronomical Society: Letters, 468, L1

Smith, K. W., et al. 2000, Monthly Notices of the RAS, 319, 991

Spezia, R., et al. 2016, The Astrophysical Journal, 826, 107

Stahler, S. W. & Palla, F. 2004, The Formation of Stars

Taquet, V., et al. 2015, The Astrophysical Journal, 804, 81

Taquet, V., et al. 2020, A&A, 637, A63

Taylor, J. R. 1996, An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements, 2nd edn. (University Science Books)

Tennyson, J. 2011, Astronomical Spectroscopy, 2nd edn. (WORLD SCIENTIFIC)

Testi, L., et al. 2014a, in Protostars and Planets VI, ed. H. Beuther, R. S. Klessen, C. P. Dullemond, & T. Henning, 339–361

Testi, L., et al. 2014b, in Protostars and Planets VI, ed. H. Beuther, R. S. Klessen, C. P. Dullemond, & T. Henning, 339–361

Tielens, A. G. G. M. 2005, The Physics and Chemistry of the Interstellar Medium

Tielens, A. G. G. M. & Hagen, W. 1982, A&A, 114, 245

Turner, B. E. 1991, The Astrophysical Journal, Supplement, 76, 617

Turner, N. J., et al. 2014, in Protostars and Planets VI, ed. H. Beuther, R. S. Klessen, C. P. Dullemond, & T. Henning, 411–432

Tychoniec, L., et al. 2020, A&A, 640

van der Tak, F. F. S. 2005, in Massive Star Birth: A Crossroads of Astrophysics, ed. R. Cesaroni, M. Felli, E. Churchwell, & M. Walmsley, Vol. 227, 70–79

van Dishoeck, E. F. 2018, IAU Symposium, 332, 3

van Dishoeck, E. F. & Bergin, E. A. 2020, arXiv e-prints, arXiv:2012.01472

van 't Hoff, M. L. R., et al. 2018, IAU Symposium, 332, 121

van ´t Hoff, Merel L. R., et al. 2020, A&A, 633, A7

Vazart, F., et al. 2020, Monthly Notices of the Royal Astronomical Society, 499, 5547

Wakelam, V., et al. 2021, A&A, 652, A63

Walsh, A. J., et al. 2006, The Astrophysical Journal, 637, 860

Walsh, A. J., et al. 2007, The Astrophysical Journal, 655, 958

Watson, W. D. 1974, The Astrophysical Journal, 188, 35

Whitney, B. A., et al. 2003, The Astrophysical Journal, 591, 1049

Wilson, T. L., et al. 2013, Tools of Radio Astronomy

Winn, J. N. & Fabrycky, D. C. 2015, Annual Review of Astronomy and Astrophysics, 53, 409

Zucker, C., et al. 2018, The Astrophysical Journal, 869, 83

Öberg, K. I. 2016, Chemical Reviews, 116, 9631, pMID: 27099922

Öberg, K. I. & Bergin, E. A. 2021, Physics Reports, 893, 1, astrochemistry and compositions of planetary systems

Acknowlegments

It is hard to believe I reached the end of this thesis, and the end of this chapter of my life. These pages have seen me evolving from a carefree, naive Master's student to a (still) naive newborn researcher, and definitely a better human being.

I thank from the bottom of my heart my supervisors, Leonardo and Marta, who entered my life quite casually, but immediately gave me wings to fly beyond my insecurities and towards the real passion for the scientific research. You also gave me the unexpected opportunity to visit ESO, an experience which blown my mind, and made me having such a nice time with lots of unforgettable people. I will never forget you two as well, with all our talks, meetings, and also nights out together. You exemplify for me what a scientific community should be: passionate, motivating, and human (after all...). On the same line, I thank Claudio and Linda for your irreversible kindness.

I must thank all my friends scattered here and there in Italy, who have always supported me in my moments of absence or crisis, and who have always been there as well for me to laugh about and have fun.

Actually, I thank my family for the same reasons, plus a particular thank to my mum for taking care of me in days in which I would have definitely forgotten to eat, and to my brother Luca for the endless talks about good films, music, and videogames.

Last but definitely not least, I thank Mr Adrien Houge. I don't know what the future holds for us, but I can only thank you for being part of my life and this work from its conception.

And I must thank myself, for keep going.