

UNIVERSITY OF COPENHAGEN

MASTERS THESIS

**Linking Galactic and extragalactic star
formation through high-J CO emission**

Markus Thor RASMUSSEN

Supervisor: Lars E. KRISTENSEN

Niels Bohr Institute

March 31, 2021

UNIVERSITY OF COPENHAGEN

Abstract

Linking Galactic and extragalactic star formation through high-J CO emission

by Markus Thor RASMUSSEN

Stars form from collapsing gas and dust in dense molecular clouds. Observations of such clouds in the Milky Way has revealed deeply embedded molecular outflows launched by forming stars. Specific molecular emission lines from such outflows is empirically proven to trace active star formation inside these molecular clouds. The same type of emission is also observed in distant galaxies but further analysis of the specific emission lines is required to determine if the extragalactic emission traces active star formation. If so, we can ultimately link star formation conditions of our own Galaxy with distant high- z galaxies, thereby shedding light on how star formation proceeds in these distant galaxies, at a time in the Universe when most stars formed.

In this thesis the Carbon Monoxide ($^{12}\text{C}^{16}\text{O}$) ladder of rotational transitions are studied. Emission data observed by the Herschel Space Observatory of CO rotational transitions ranging from $J=4-3$ to $J=48-47$ emission have been collected for 190 objects (87 galactic protostellar objects and 103 extragalactic objects). The collected data has been organized in a MySQL database and analysed to find common characteristics, revealing embedded star formation in extragalactic sources.

From the analysis of the CO ladder in galactic protostellar objects, the thesis finds 3 distinct physical components of a star-forming region. One of these is a warm component (~ 300 K), shown to be well represented in the entire sample with a unique origin in molecular outflows and traced by high-J CO emission ($J_{\text{up}} > 9$) making these lines a promising tracer of embedded and active star formation.

To quantitatively link the emission properties of nearby Galactic star-forming regions to those of more distant galaxies, the theoretical model Galaxy-in-a-Box has been calibrated using CO $J=10-9$ observations from the MySQL database. This allows us to constrain the physical parameters, such as, star-formation efficiency, core-collapse timescale and initial mass function, even for galaxies with limited angular resolution. This thesis work is a stepping stone to extend measurements of star formation to high- z galaxies and ultimately improve our global knowledge of star formation.

Contents

Abstract	iii
1 Introduction	1
1.1 Star formation	1
1.1.1 Molecular clouds	2
1.1.2 Evolutionary stages of stars	3
Outflows	5
1.2 Tracer of molecular outflows	5
1.2.1 Molecular Hydrogen	5
1.2.2 Carbon Monoxide	6
1.3 Herschel Space Observatory	7
1.4 Scientific goals	9
2 Methods	11
2.1 Database of Carbon Monoxide emission	11
2.1.1 MySQL	11
2.1.2 Structure of the database	12
2.1.3 Instruments	12
Herschel PACS	12
Herschel SPIRE	15
Herschel HIFI	16
2.2 Evaluating data from protostars	17
2.3 Galaxy in a Box	18
Spatial distribution	18
Age distribution	19
Initial Mass Function	20
2.3.1 Mass distribution	21
2.3.2 Emission	23
2.3.3 From cluster to galaxy	23
Cluster mass distribution	23
Spatial cluster distribution	24
2.4 Optimization	25
2.4.1 Input/output fileformats	26
2.4.2 Elementwise calculations with NumPy	26

2.4.3	Mass distribution variable	27
2.4.4	Print results	28
3	Results	29
3.1	Database	29
3.2	Line Luminosity vs Luminosity	30
3.3	SLEDs	32
3.3.1	Raw Data	32
3.3.2	The Spline fit	34
3.4	Galaxy in a Box	39
	Initial mass function	39
	Star formation efficiency	40
	Free fall timescale	42
	Total cluster emission	42
4	Discussion	45
4.1	Physical origin of CO emission	45
4.1.1	The Cold Component	45
4.1.2	The Warm Component	46
4.1.3	The Hot Component	47
4.1.4	The Case for Active Galactic Nuclei	47
4.2	Spectral Line Energy Distributions	47
4.2.1	Comparing the Object Categories	47
	Class 0 vs Class 1	47
	Class 2	49
4.2.2	Extragalactic Objects	49
4.2.3	Temperature Components in the Data Sample	50
4.3	Galaxy-in-a-Box	52
	Star Formation Tracer Checklist	52
4.3.1	Significance of results	53
	Initial Mass Function	53
	Star Formation Efficiency	54
	Free Fall Timescale	55
4.4	Outlook	56
5	Summary and Conclusions	57

Introduction

When glancing at the clear night sky a number of shining bright dots stands in sharp contrast to the pitch black background, one cannot help but wonder how these shining bright objects mankind came to know as stars came to life?

This has been one of the big questions of modern astronomy. The last decades advancements of modern telescopes has helped achieved some answers to this question. Empirical data has provided a detailed understanding of the physical condition present when the nearby stars of our Galaxy are born. The level of detail in our understanding of star formation fades as the distance to the regions increases and thereby limits our angular resolution. This is a problem to our large-scale knowledge because studies find stars to have formed at a much higher rate billions of years ago (Bouwens et al. 2015).

In this thesis I am going to present a possible solution to this problem, by using our current knowledge of local star forming regions to find characteristic emission features with origin in molecular outflows and investigate their legibility as tracers of active star formation. A quality tracer of active star formation can be used as a link between local and distant star formation and thus contribute to an increased understanding of how stars have been born through the history of the Universe.

In this first chapter I will describe the foundation of star formation, the physical conditions, the evolutionary stages of a young star, molecular outflows and potential tracers.

1.1 Star formation

In order to describe how the stars of our universe formed, it is required to also think about when they formed. Bouwens et al. (2015) describes an enormous survey of more than 10,000 high redshift galaxies from which an amount of star formation was estimated from a correlation with UV luminosity density, and displayed as a function of time/redshift. In figure 1.1 the result of the work is shown. It is seen that star formation in no way seem to be peaking at present day around $z=0$, but rather at a redshift in the $z=2-3$ range or approximately 10 billion years ago. The about one order of magnitude higher Star Formation Rate (SFR) 10 Gyrs ago should make the

time period of high interest to anyone wanting to explain the origin of the stars of the universe. To increase the knowledge of such large-scale star formation a better understanding of individual protostellar objects in our own Galaxy should be well identified and the research of them should be used to find good tracers which can probe star formation and extend our knowledge of it to extragalactic objects all the way to high- z where the SFR seems to have peaked and thereby obtain a broader understanding of the formation of the stars in the universe.

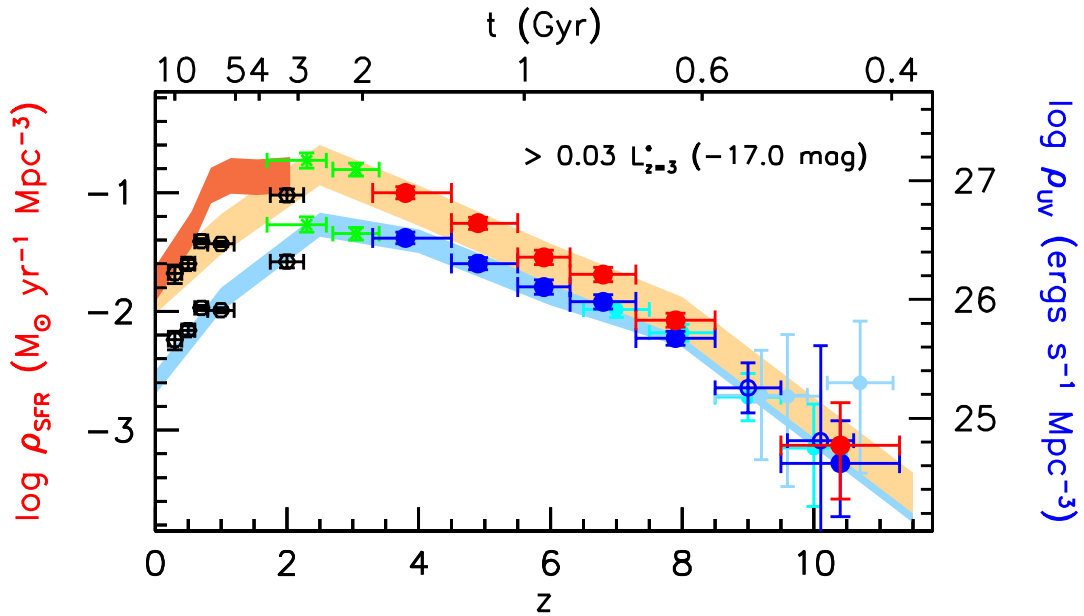


FIGURE 1.1: Logarithm of the star formation rate pr area (left axis) as a function of time/redshift, derived from the UV luminosity densities (right axis). The red line and shaded area is corrected for dust extinction while the blue is left uncorrected.

Credit: [Bouwens et al. \(2015\)](#)

1.1.1 Molecular clouds

In order to begin the extension of the knowledge of star formation to the largest scales of the universe a proper look at the current knowledge and observations made very close by, right in our own Galaxy has to be examined. Decades of observations in our local neighborhood has shown that the majority of stars are formed in interstellar clouds, called "molecular clouds" made of cold molecular gas and dust. The largest of these molecular clouds are called Giant Molecular Clouds (GMC) which are large (≈ 50 pc) low density clouds ($\approx 100 \text{ cm}^{-3}$) consisting primarily of molecules such as H_2 ([Stahler & Palla 2004](#)) which is by far the most common molecule in the universe. GMCs account for most local star formation concentrated in denser clusters as they over time collapse into stars ([Lada & Lada 2003](#)) until the clouds are depleted of gas. This collapse only happens under certain conditions described by the Virial Theorem

(Stahler & Palla 2004):

$$\frac{1}{2} \frac{\partial^2 I}{\partial t^2} = 2\mathcal{T} + 2\mathcal{U} + \mathcal{W} + \mathcal{M} \quad (1.1)$$

Which is describing the acceleration of the moment of inertia I in a GMC, on the right hand side of the equal sign the four different energies contributing to the inertia is added together, the contributors are the total kinetic energy also called bulk motion \mathcal{T} , the thermal energy \mathcal{U} , the gravitational potential energy \mathcal{W} and the magnetic field \mathcal{M} . When the four contributor add up to zero like:

$$2\mathcal{T} + 2\mathcal{U} + \mathcal{W} + \mathcal{M} = 0$$

the cloud is in virial equilibrium and therefore supported against gravitational collapse and not likely forming any stars. This means the usual star forming regions is dominated by the gravitational potential energy \mathcal{M} which causes the cloud to collapse.

1.1.2 Evolutionary stages of stars

When a molecular cloud starts to collapse and thereby starts the star formation process a protostellar object is developing. The developmental stages of a protostellar object can be divided into four categories originally purposed by Shu et al. (1987) and gradually specified as further knowledge of these often tricky to observe objects were increased. The objects are gradually developing so they might be in-between to developmental stages at a given point in time but the overall stages have some important differences. A cartoon overview of the stages can be seen in figure 1.2 and each stage will be further described in this section.

Class 0

The first evolutionary stage of a protostar is Class 0. Class 0 objects are always deeply embedded in an envelope of cold gas and dust ($M_{env} \gg M_*$) which historically has made observing the protostar itself challenging however observing molecular outflows from the environment close to the protostar can be done since all Class 0 objects are shown to host such outflows. Class 0 objects are developing through an in fall of mass from the envelope towards the central objects, which is illustrated in figure 1.2. In the top left of the figure marked Class 0, it shows a thick envelope surrounding the central object and contributing to the observations hereof. This must be taken into account when analyzing observations of Class 0 objects.

Class 1

Class 0 objects slowly evolve into the next evolutionary stage Classified as Class 1 Objects. There is not a perfectly fixed line between Class 0 and Class 1 objects but at the Class 1 stage much less of the surrounding envelope has either been accreted by the protostar or the protostellar disk or ejected via outflows, also illustrated in figure 1.2

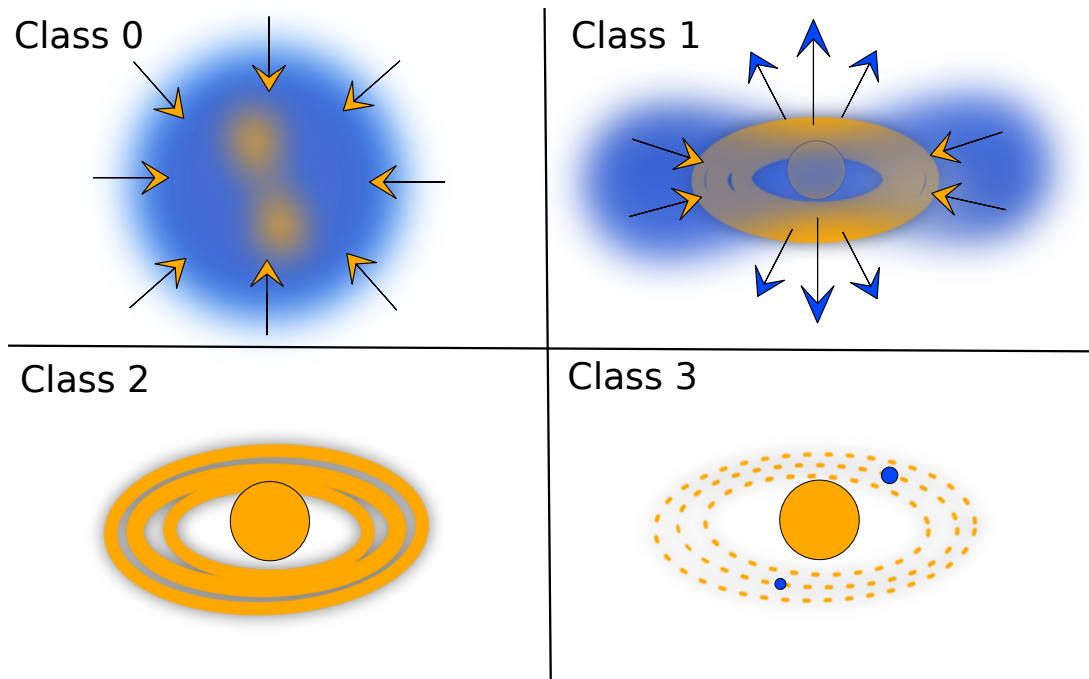


FIGURE 1.2: Cartoon of the four developmental stages of a protostar/pre-main sequence star. Top left: Class 0 object; Blue semi-transparent parts are representing envelope of gas and dust, the orange arrows are related to infall/accretion and the orange semi-transparent droplets in the center illustrates the molecular outflow from the object. Top right: Class 1 object; Blue arrows are representing outflow-and jets and the orange arrows are related to infall/accretion. Bottom left: Class 2 object defined ring structure surrounds the pre-main-sequence star. Bottom right: Class 3 object; the PMS star are surrounded by a protoplanetary disk of planetesimals and protoplanets. The cartoon is not to scale.

in the top right marked Class 1. The lower amount of surrounding envelope causes Class 1 objects embedded to a significantly smaller degree than Class 0 objects which causes these objects to be more easily observable due to a lower optical thickness. In addition to the molecular outflows observed in Class 0 objects, Class 1 objects also ejects material via jets.

Class 2

When the envelope surrounding the protostar is depleted leaving only the central star and a surrounding disk, like illustrated on the bottom left cartoon of figure 1.2, the object is no longer labeled a protostar but rather a pre-main-sequence star. Class 2 objects rarely host outflows and are characterised by its lack of envelope and exposed and defined disk.

Class 3

Much like Class 2 objects Class 3 objects completely lacks an envelope, however unlike the Class 2 objects with their defined disk, the disk of Class 3 objects have started to collapse into planetesimals and planets, as illustrated in figure 1.2 in the bottom

right cartoon. Class 3 is the last stage before the main sequence. Class 3 objects have a low amount of molecular gas since they are depleted of both envelope and outflows. They are of little interest to this thesis and will not be discussed any further.

Outflows

The main focus of this thesis will be on tracing the molecular outflows of protostars of Class 0, 1 and 2. Molecular outflows are a necessary part of star formation. They carry out the unavoidable task of removing angular momentum and envelope material from the local environment around the protostar itself allowing the formation of a disk (Mottram et al. (2014)). Molecular outflows are observed in Class 0 which shows the largest molecular outflows, and like Class 1 objects always have outflows, Class 2 objects on rare occasions have both outflows and jets. The molecular footprint of these protostellar outflows are uncommon under most other physical condition which makes outflows an interesting target when wanting to map star formation. The outflows are not yet perfectly understood, but some feasible and well described theories are in great correlation with existing empirical data and will be described here and discussed in this thesis. In figure 1.3 a cartoon of possible physical components of a molecular outflow are shown. The figure is split in two sections at the central protostellar objects (blue spherical objects). Both sides show an outflow consisting of an exterior of cold entrained outflow closest to the surrounding envelope (not shown in the figure) followed by cavity walls. On the left hand side fast moving cavity shocks (Mottram et al. (2014)) are shown as green arrows. Together with narrow high velocity well collimated spot shocks and jets (Orange) most commonly observed with Class 0 objects. On the right hand side cavity shocks and spot shocks are exchanges for strong disk winds driven the rapid movement inside the cavity walls of the outflow. The disk wind are likely to be driven by magneto-hydrodynamic processes (Yvart et al. (2016)).

1.2 Tracer of molecular outflows

To trace a molecular outflow a molecule or a series of molecules present in the outflow has to be targeted and observed, but what is a good tracer of the environment in and around a molecular outflow? The initial thought is to go with the most abundant molecules.

1.2.1 Molecular Hydrogen

Molecular Hydrogen (H_2) is the most abundant molecule (Stahler & Palla 2004), however several complications makes H_2 one of the most difficult to observe and therefore also a poor choice for tracing star formation. A main reason for H_2 being hard to detect is to be found in the rotational energy of the symmetric molecule. The rotational

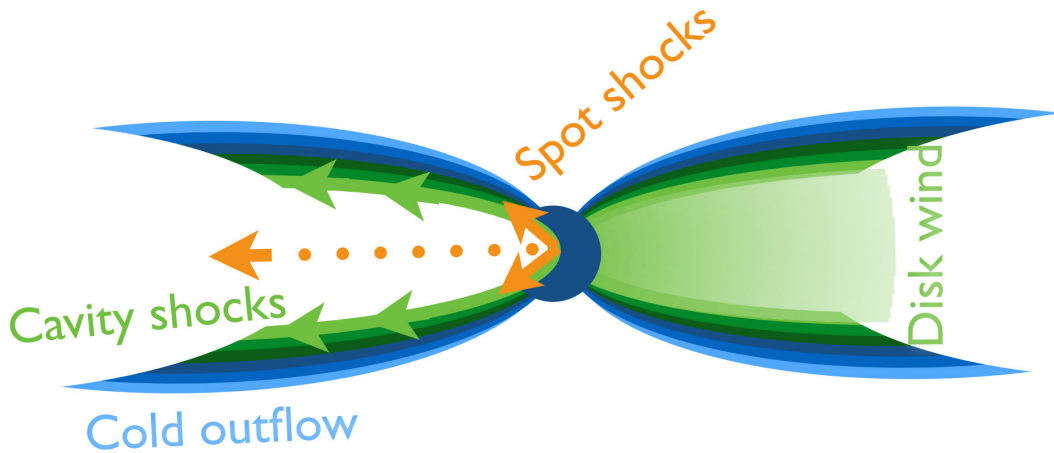


FIGURE 1.3: Cartoon of the possible physical components of a protostellar system. The cartoon sees past the envelope and only shows the deeply embedded protostellar object and outflow. The two sides of the cartoon illustrate different physical scenarios explaining observation. On the left the outflows host cavity shocks, and on the right side the motions are driven by disk winds. The observed feature of spot shocks are shown in orange. Credit: Modified from [Kristensen et al. \(2017\)](#)

energy for a diatomic molecule is given by:

$$E_{\text{rot}} = \frac{\hbar^2}{2I} J(J+1) \quad (1.2)$$

Where I is the moment of inertia, and J is the rotational quantum number, which is a dimensional number linked to the state of the molecule and therefore always an integer, 0 for ground state, 1 for the first state over ground and so on. Due to molecular hydrogen having the smallest moment of inertia of any molecule it is seen from equation 1.2 that it will also have a widely spaced energy levels J . H_2 has no dipole moment which means the first allowed transition of the molecule is from the electric quadrupole transition going from $J=2$ to $J=0$ which has a energy change of $E/k_b = 510$ K corresponding to an emitted photon with wavelength $28.2 \mu\text{m}$. Even the lowest transition of H_2 as mentioned here requires an approximate temperature of 300 K to collisionally populated ([Dionatos 2015](#)) as it would have to be in a deeply embedded region of a protostar. Add to this the likelihood of H_2 being optically thick and causing self shielding the H_2 molecule must be deemed unsuited to be a standalone tracer of molecular outflows.

1.2.2 Carbon Monoxide

Carbon Monoxide (CO) is the second most abundant molecule and has some physical/chemical qualities which makes it a strong tracer for star forming regions. First and foremost CO unlike H_2 has a permanent dipole moment which allows transitions of $J \pm 1$, making transition be like $J=1-0$, $2-1$, $3-2$, etc, which paired with a drastically

higher moment of inertia makes the packs the energy levels of CO much tighter together than H₂. In table 1.1 the first 48 transitions of CO are listed with corresponding photon wavelength in micrometers and upper level energy given as equivalent temperature $E_{\text{up}}/k_{\text{b}}$, the table values are from The Cologne Database for Molecular Spectroscopy (CDMS; Endres et al. (2016)). It is seen from table 1.1 that the rotational transitions of CO are close together in wavelength making it easy to observe a large part of the CO ladder in one observation, and since the transitions also spans a large range of upper level energies it is well suited to trace different physical conditions. The lowest J transitions of CO are very easily excited even at low temperatures which for practical purposes causes them to always be optically thick (Dionatos 2015). While optical thin emission will almost always be possible to obtain moving to high enough J or observing different isotopologues than the most abundant ¹²C¹⁶O which is the most common isotopologue and the one in focus in this thesis. From now on every mention of CO or ¹²CO is referring to ¹²C¹⁶O.

While the first few transitions of CO are observable from ground, the rest of them are not because the atmosphere is completely opaque at the wavelength of the photons emitted from those transitions. This has the natural consequence that the high J CO transitions in our local universe cannot be observed with ground-based telescopes and must be observed from space. At the current time when writing this thesis no operational space based telescope can observe the sub-mm wavelengths emitted by high-J CO emission and there are currently no plans for a new space-based telescope with this ability. This means that existing observations has to be further explored if CO is to be used as a star formation tracer.

1.3 Herschel Space Observatory

The Herschel Space Observatory (Herschel; Pilbratt et al. (2010)) was an ESA space observatory launched on 14 May 2009 and after almost 4 years of operation in the second lagrangian point (L2) of the Sun-Earth system it was deactivated on 29 April 2013 after running out of the liquid helium used for cooling the instruments. With its 3,5 meters Herschel is to date the largest known telescope to have ever flown in space. The 3,5 meter large mirror of Herschel gave an unprecedented capability for far-infrared and sub-millimetre observations which made Herschel perfect for observing the formation of stars and molecular emission lines such as CO in the interstellar medium. On board Herschel had three instruments which are individually described in section 2.1.3. The three instrument in combination covers a spectral range from 51-671 μm making Herschel the perfect instrument for observing the CO ladder. From table 1.1 it is apparent that Herschel observes the CO rotational transitions from J=4-3 all the way to 48-47, which is the transitions discussed in this thesis.

TABLE 1.1: Properties of rotational transitions of CO

Line [$J_{\text{up}} - J_{\text{low}}$]	Wavelength [μm]	$E_{\text{up}}/k_{\text{b}}$ [K]
1-0	2600.75763	5.53
2-1	1300.40366	16.6
3-2	866.96337	33.2
4-3	650.25152	55.3
5-4	520.23103	83.0
6-5	433.55623	116.2
7-6	371.65039	154.9
8-7	325.22516	199.1
9-8	289.12033	248.9
10-9	260.23980	304.2
11-10	236.61328	365.0
12-11	216.92729	431.3
13-12	200.27248	503.1
14-13	185.99929	580.5
15-14	173.63143	663.4
16-15	162.81163	751.7
17-16	153.26671	845.6
18-17	144.78419	945.0
19-18	137.19633	1049.8
20-19	130.36893	1160.2
21-20	124.19335	1276.0
22-21	118.58072	1397.4
23-22	113.45760	1524.2
24-23	108.76281	1656.5
25-24	104.44495	1794.2
26-25	100.46053	1937.4
27-26	96.77251	2086.1
28-27	93.34912	2240.3
29-28	90.16300	2399.8
30-29	87.13042	2564.8
31-30	84.41072	2735.3
32-31	81.80581	2911.2
33-32	79.35981	3092.5
34-33	77.05870	3279.2
35-34	74.89006	3471.3
36-35	72.84285	3668.8
37-36	70.90724	3871.7
38-37	69.07441	4080.0
39-38	67.33645	4293.7
40-39	65.68626	4517.7
41-40	64.11741	4737.1
42-41	62.62410	4966.9
43-42	61.20106	5201.9
44-43	59.84349	5442.4
45-44	58.54705	5688.1
46-45	57.30773	5939.2
47-46	56.12190	6196.0
48-47	54.98622	6457.2

1.4 Scientific goals

In this thesis available Herschel CO emission data from both galactic protostars and prestellar objects as well as for extragalactic sources will be collected in order to address the value of using high-J CO as a tracer of deeply embedded star formation inside molecular clouds and explore to what extent the CO ladder traces the same physical parameters on both galactic and extragalactic scales. If indications can show CO to be a good tracer of embedded star formation in extragalactic sources this will be a stepping stone towards determination of whether or not CO can be used as a star formation tracer in large redshift galaxies where it would potentially strengthen our knowledge of large-scale and most rapid star formation in the history of the universe.

Methods

Deeply embedded protostars and their molecular outflows have been observed through emission lines such as those from high J CO transitions. These are in this thesis used as a tracer for galactic and extragalactic star formation. In order to convincingly use high J CO emission as a tracer of distant star formation a significant amount of high J CO observations of confirmed star formation, along with data of similar CO emission of extragalactic objects has been evaluated. Through the development of this thesis a database has been created a database which contains both galactic and extragalactic line emission data of CO rotational transitions ranging from J=4-3 to J = 48-47. In this chapter the process of collecting the data and making the database is described, as is the methods for comparing galactic and extragalactic CO emission and star formation. The chapter finishes off describing the theoretical model Galaxy-in-a-Box, and my contributions to the model used to predict star formation in galaxies where resolving individual stars are not possible.

2.1 Database of Carbon Monoxide emission

In order to do the best possible job describing star formation with high-J CO emission lines a sufficient amount of observational data have to be the foundation which is why an emission database has been created. The data has been inserted and organized in a MySQL database, the included data has been obtained from many different sources all cited within the database. From some sources the data has been manually typed in from the PDF files of the published articles. Rather than reading off the data values manually when access to electronic data products was provided the data was transferred to the database using *Python* with an installation of the *mysql-connector* package. The *Herschel* Space Observatory (Pilbratt et al. 2010) is particularly good at observing CO line emission and is therefore the telescope used for the observations included in the database.

2.1.1 MySQL

The choice of making the database in MySQL was done for multiple reasons: manipulating and selecting specific parts of the database are very easy, quick and convenient in MySQL. An advantage of MySQL over a .csv or .ascii file is that it is not necessary to load the entire database or file into the computers memory in order to do initial

analysis of the data. This means the users of the database for instance can call all records of CO J = 10-9 and then only read these specific records (in SQL language a row is called a record) into memory which is much more efficient for big data sets. Another pro of using MySQL is its rigidity, when making a column in MySQL it is required to specify what kind of data is going in the column. Only having one type of data allowed limits the amount of typing/input mistakes and makes it easier for the end-user to know exactly what the content of each column is and for less manipulation of the data to be required by the end user.

2.1.2 Structure of the database

In table 2.1 a overview of which data is contained in the database is shown. The database has 17 columns of information for each record. An example of 3 records from the database can be seen in table 2.2. Some values of the database is either invalid or missing, all values like such are assigned the MySQL value "NULL", an example of this is artificially added to the resolution column of the third record of table 2.2. Another thing worth noting in the figure is the id numbers being higher than the number of records in the database this origins in the way data was inserted into and sorted inside the database. Data was entered into the database in different ways, some were manually type in from articles. When manually type in the data was evaluated before entering the database in order to decrease the amount of manual typing. When possible the data was imported through digital data products/archives such as ViZieR or data files. When this was possible all the data was entered into the database and then evaluated inside MySQL which made for a significantly easier evaluation process but also meant that a lot of other non-CO molecular transitions had to be removed from the database again along with invalid values or values without interest. The reason for the high id numbers in the example records table 2.2 is the initially imported records later removed. Because the id number must always be unique and can never be reused in the same database, the removed records still count towards the id although removed, so later imported records will therefore increment from the highest ever used id when new records are added.

2.1.3 Instruments

Herschel PACS

The records in the database assigned with PACS in the *Instrument* column was observed with the spectrometer of Herschel Photodetector Array Camera & Spectrometer (PACS; Poglitsch et al. (2010)) PACS is a 5x5 array of spatial pixels (*Spaxels*) each spaxel with a pixel size of 9".4 x 9".4. The spectral range of PACS is 55-210 μm in two spectral bands 51-105 μm and 103-210 μm which can operate simultaneously however with a spectral gap caused by spectral leakage in the wavelengths

TABLE 2.1: Overview of the content of the MySQL database

Column Name	Column Content
Telescope	Telescope used for the observation in this record
Instrument	The instrument used for the observation
Object	Name of the astronomical object observed
Type	The type of object here refers to Class 0, 1 or 2 protostars or an extragalactic object.
Molecule	The specific isotopologue of the molecule
Transition	The rotational transition in the form : $J_{\text{up}} - J_{\text{low}}$
RA	Right Ascension
DEC	Declination
Distance	Distance in parsecs
Wavelength	Restframe wavelength of transition described in transition column in units of micrometers μm
Resolution	Angular resolution given in arc seconds
Flux	Integrated flux of emission line corresponding to transition in <i>Transition</i> column. Units of the flux are specified in column F_{units}
Errorflux	Error of the flux described in <i>Flux</i> column in the same units, describe in F_{units}
F_units	Units of value in <i>Flux</i> Column. Either $\text{K} * \text{km} * \text{s}^{-1}$ or $10^{-20} * \text{W} * \text{cm}^{-2}$
Luminosity	Luminosity in solar luminosities L_{\odot}
Reference	NASA ADS or arXiv link to the article/publication from which the data was obtained.
id	Unique identification number for each record

ranges 98-103 μm and 190-220 μm . Figure 2.1 shows the spectral ranges of the Herschel instruments plotted against the wavelengths of the ^{12}CO transitions here represented by the upper energy level J_{up} . The PACS spectral coverage is illustrated in the figure by the orange box. PACS has a resolving power $R=1000-4000$ corresponding to $\Delta v = 75 - 300 \text{ km s}^{-1}$. For a point source PACS has a line sensitivity of $\approx 4 - 10 \times 10^{-18} \text{ W m}^{-2}$. All of which makes PACS a useful instrument when wanting to observe CO emission lines for the rotational transitions ranging from $J=14-13$ to the upper limit of transitions contained in the database $J=48-47$ with the exception of the two lines $J=27-26$ and $J=26-25$ which emits its light in the PACS spectral gap.

The PACS data contained in the database comes from multiple surveys: *Water in star-forming regions with Herschel* (WISH; Karska et al. (2013)), *The William Herschel Line*

TABLE 2.2: Example of records in database

Column Name	Records		
Telescope	Herschel	Herschel	Herschel
Instrument	PACS	HIFI-5a	SPIRE
Object	Serpens-SMM4	BHR 71	NGC 5104
Type	Class 1	Class 0	Extragalactic
Molecule	12CO	12CO	12CO
Transition	33-32	10-9	6-5
RA	277.4878845	12.0136810000	00.20034625
DEC	1.2212413000	-65.1475830000	13.1841541
Distance	429.000000000	178.000000000	9080000.00000
Wavelength	79.3618087768554	260.23990053	433.567138671875
Resolution	9.40	20.00	NULL
Flux	1.085197047	9.8999996185302723	1.3099999427
Errorflux	0.488436700	0.45268407526662	0.0900000
F_units	10^{-20} W cm $^{-2}$	K km s $^{-1}$	10^{-20} W cm $^{-2}$
Luminosity	22.50000000	11.50000000	78433.0000000
Reference	Green et al. 2016: https://ui.adsabs.harvard.edu/abs/2016AJ...151...75G/abstract	San José-García et al. 2013: https://ui.adsabs.harvard.edu/abs/2013A26A...553A.125S/abstract	Lu et al. 2017: http://arxiv.org/abs/1703.00005
id	162416	325928	325191

Notes. Resolution in the third example column/record, is left as NULL to showcase how an empty value is output from the database. Units of all values listed values are found in table 2.1.

Legacy (WILL; (Mottram et al. 2017; Karska et al. 2014)), *Ultra Luminous Infrared Galaxies* (ULIRGs; Mashian et al. (2015)), *CO in ProtoStars* (COPS), *Dust, Ice and Gas in Time* (DIGIT) and *FU orionis Objects Surveyed with Herschel* (FOOSH) the latter three being contained in the COPS-DIGIT-FOOSH (CDF) archive (Green et al. 2016). For WILL and WISH data was already reduced to a single point source per object and I simply manually typed the values into the database using MySQL queries whereas for the CDF archive all 25 spaxels were included. In Green et al. (2016) it is described that the central spaxel is located in the center of the objects, which is where we would expect

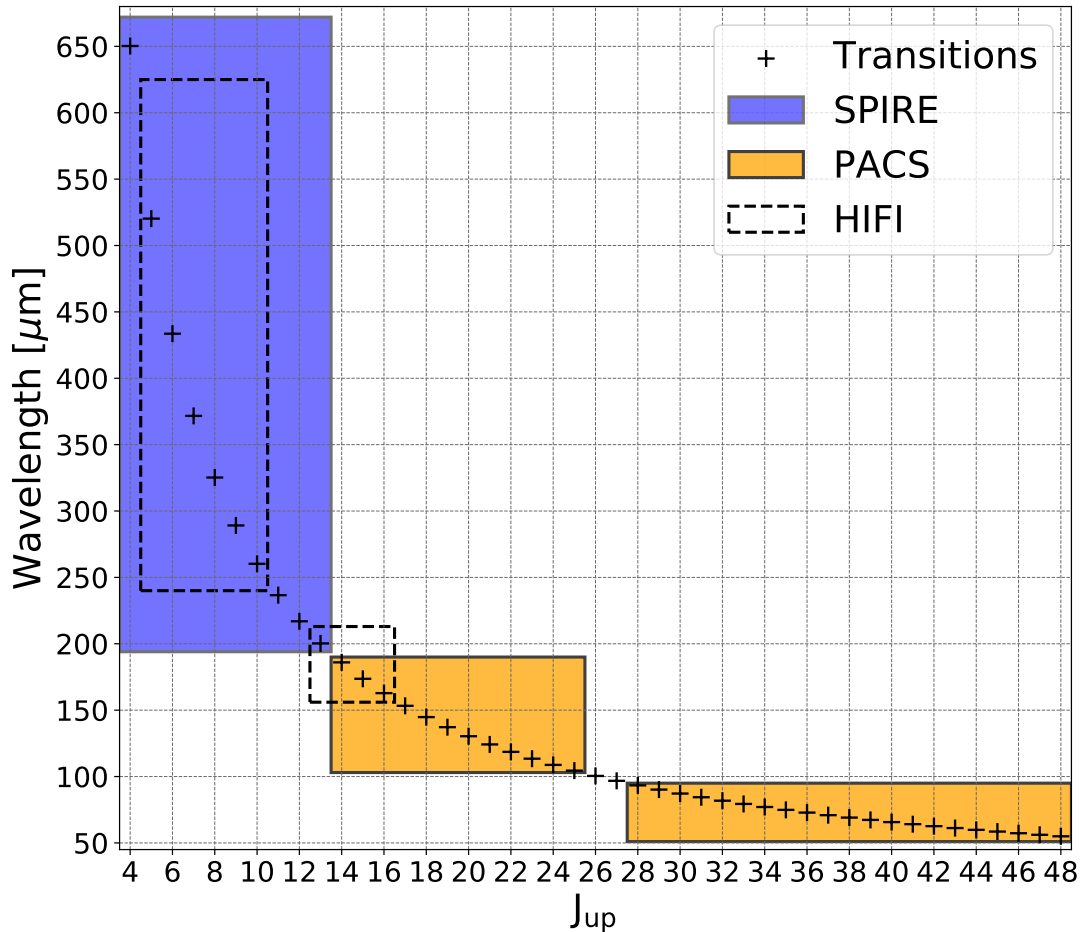


FIGURE 2.1: On the x-axis the upper energy level J_{up} , against wavelength in μm on the y axis. The plus points in the plot represents the ^{12}CO transitions with upper level energy as shown on the x-axis which emits light at wavelength described on the y-axis. The spectral coverage of three Herschel instruments are shown as boxes, where blue is SPIRE, orange is PACS and the transparent with black dashed edges is HIFI. The specific values for wavelengths of the transitions are listed in figure 1.1

to see outflows and higher temperatures and therefore it probes the high-J CO emission of the full 5x5 array. All CO emission data from the CDF archive was initially entered into the database and then later sorted, removing all non-center spaxels and only leaving the center spaxel of each object and each transition in the final iteration of the database (see figure 2.2).

Herschel SPIRE

The Herschel Spectral and Photometric Imaging Receiver (SPIRE; Griffin et al. (2010)) consists of a photometer and an imaging Fourier-transform spectrometer (FTS). Since spectrally resolved emission lines from CO transitions is used in the database the FTS is the one of interest to this thesis. The FTS uses two arrays of overlapping bands with spectral range 194 – 313 μm and 303 – 671 μm which works simultaneously and

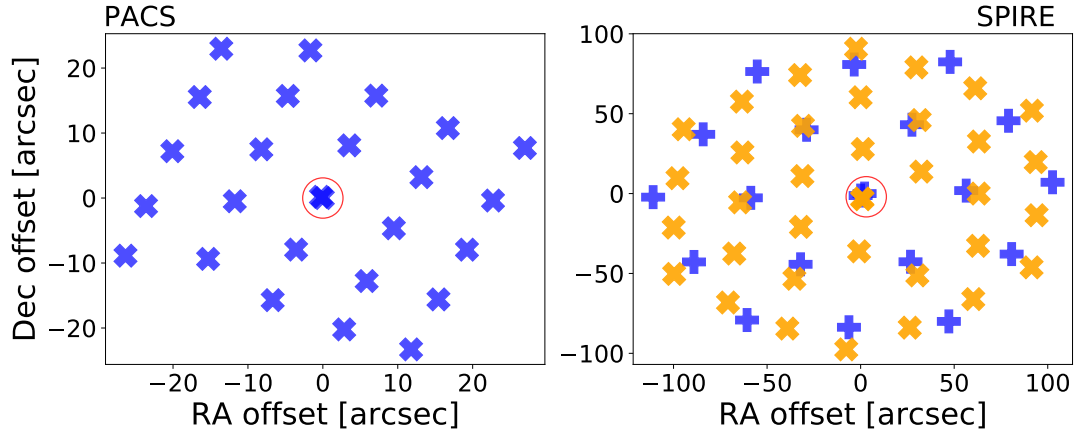


FIGURE 2.2: LEFT: Example of PACS 5x5 array from an CO 14-13 observation of the Class 0 protostar BHR 71. The circled point is the center point which is the one included into the database. RIGHT: Diamond-shaped SPIRE arrays consisting of a total of 37 spaxels (Orange) or 19 spaxels (Blue). Here to are points missing from the red array because they were already invalidated by the observing team. The points circled is the center points included in the database.

covers all rotational CO transitions in the entire spectral band. The resolution ranging between $R=370-1300$ corresponding to a sensitivity of these to bands being of $\approx 1.5 \times 10^{-17} \text{ W m}^{-2}$ and $\approx 2 \times 10^{-17} \text{ W m}^{-2}$. As illustrated in figure 2.1 this makes SPIRE a useful instrument when wanting to observe CO emission lines for the rotational transitions ranging from $J=4-3$ to $J=13-12$. In figure 2.1 the SPIRE spectral coverage is illustrated by the blue box

The SPIRE data included in the database comes from multiple surveys: Luminous Infrared Galaxies (LIRGs, Lu et al. (2017)), ULIRGs(Lu et al. 2017; Rosenberg et al. 2015; Mashian et al. 2015),The Herschel comprehensive (U)LIRGS Emission Survey (HerCULES; Rosenberg et al. (2015)), COPS, DIGIT, FOOSH the latter three from the CDF archive (Green et al. (2016)). The LIRG and (U)LIRGs data from Lu et al. (2017) was automatically inserted record by record from the electronic data product provided through VizieR using a custom Python script. The (U)LIRGs and HerCULES data provided by Rosenberg et al. (2015) was manually typed into the database, and as with the PACS part of the CDF archive the full array of spaxels was included in the database, to later be removed and only left with the central spaxel (see figure 2.2) for each object and each transition.

Herschel HIFI

Herschel Heterodyne Instrument for the Far-Infrared (HIFI; de Graauw et al. (2010)) is a instrument on-board Herschel, it is a of set of 7 heterodyne receivers which collectively covers the spectral range $157-625 \mu\text{m}$. HIFI consist of two spectrometers the Wide Band Spectrometer (WBS) and the High Resolution Spectrometer (HRS). Both spectrometers measure the horizontal and vertical polarisation simultaneously, HRS

with configurable spectral resolution of 0.125 - 1.0 MHz, and WBS with constant spectral resolution of 1.1 MHz. The HIFI data are included from (San José-García et al. 2013).

2.2 Evaluating data from protostars

In order to be able to compare the emission from galactic protostellar objects and extragalactic sources contained in the created database, Spectral Line Energy Distributions (SLEDs) are chosen for displaying the data since this is typically the preferred distribution for extragalactic sources over the also commonly used rotational diagram. The main parameter of interest here is the total flux of emission from each rotational transition of the CO molecule.

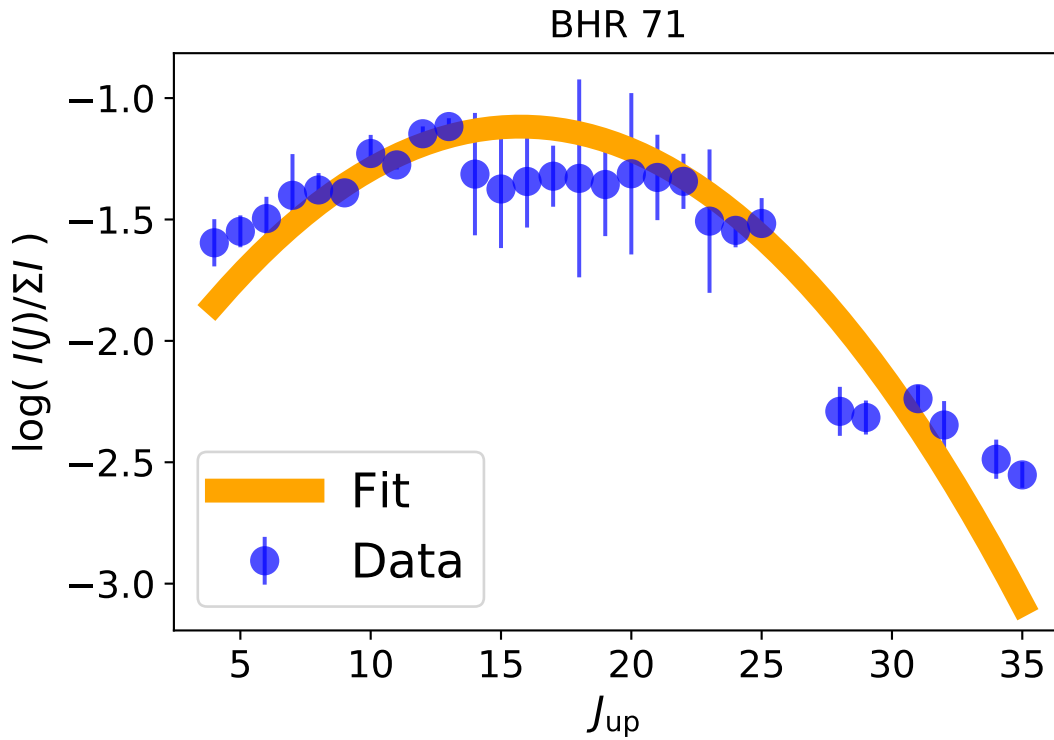


FIGURE 2.3: Normalized intensities for transitions $^{12}\text{CO } J_{\text{up}}$ from 4-3 to 36-35. Intensities with a $S/N < 5$ excluded. Errorbars on the data are multiplied by 5 to make them more visible. The fit is a weighted spline fit

In figure 2.3 a plot of the logarithm of the relative intensity of the emission from one specific transition relative to the total amount of observed emission from CO rotational transitions. All datapoints with a $S/N < 5$ is excluded in order to avoid including data too highly influenced by noise. The errors displayed in the errorbars on the figure is calculated using errorpropagation.

$$\sigma_f = \left| \frac{df}{dx} \right| \cdot \sigma_x$$

In this case $f = \log(x)$ so the uncertainty here becomes

$$\sigma_f = \frac{1}{x \ln(10)} * \sigma \quad (2.1)$$

σ being the uncertainty on the flux obtained from the database.

The data from BHR 71 shown in this figure is observed with partly Herschel PACS and Herschel SPIRE which explains the big jump in relative intensity going from $J_{\text{up}} = 14$ to $J_{\text{up}} = 15$.

The reason for choosing the intensity relative to the total emission from all the observed transitions rather than the more commonly used relative intensity which is relative to $J=1-0$, is both that CO $J=1-0$ is not observed with Herschel but more importantly that the CO $J=1-0$ line is almost always optically thick and therefore it would imply a larger optical thickness upon the rest of the transitions than necessary. This is especially important when wanting to compare the CO emission from galactic protostars to that of the extragalactic sources in the database where the lowest J transitions would almost certainly be optically thick and therefore not be usable to trace embedded protostars.

2.3 Galaxy in a Box

Galaxy in a box (hereafter referred to as GiB) is an expansion of the model *Cluster in a box* (hereafter referred to as CiB; Kristensen & Bergin (2015)). CiB is a model which predicts/reproduces emission from embedded protostars in stellar clusters given the ability for the user to interpret important parameters of star formation from their observations. GiB is a not yet published expansion of the model aiming to reproduce observations of CO emission from embedded protostars in an entire galaxy.

To a large part, the GiB model is similar to CiB. The GiB distributes clusters as created in the CiB model throughout a template model of a galaxy. To make it a little more clear exactly what the CiB model does, I am going to go through the model and the equations providing its foundation: The CiB model simulates a star-forming cluster. Using a template cluster with a spatial, age and mass distribution. These distributions are described in detail below. In table 2.3 all variable input parameters, a brief description and their default values are listed, the variables are described in detail in this section.

Spatial distribution

The cluster is in this model assumed to be a perfect sphere with a maximum radius, from Adams et al. (2014) given by:

$$R_{\text{max}} = R_0 \left(\frac{N}{N_0} \right)^{-\alpha},$$

where R_0 is the unit of the distance as default set to 1 pc, N is the number of stars, $N_0=300$ and alpha is set to 1/3. The radial probability function is given by:

$$\frac{dP}{dr} = \frac{3-p}{R_{\max}} \left(\frac{r}{R_{\max}} \right)^{2-p} \left(\frac{M}{M_{\min}} \right)^{-0.15},$$

where p is the powerlaw index as default set to 1.5, M is the mass of the star and M_{\min} is the minimal mass of a star in the cluster. The last term is set to ensure the probability of high mass protostars to appear more often towards the center of the cluster. The position of each protostar is then given by these two equation so that the position is $x = r \cos(\phi)$ and $y = r \sin(\phi)$, ϕ being a randomly assigned angle from a normal distribution.

Age distribution

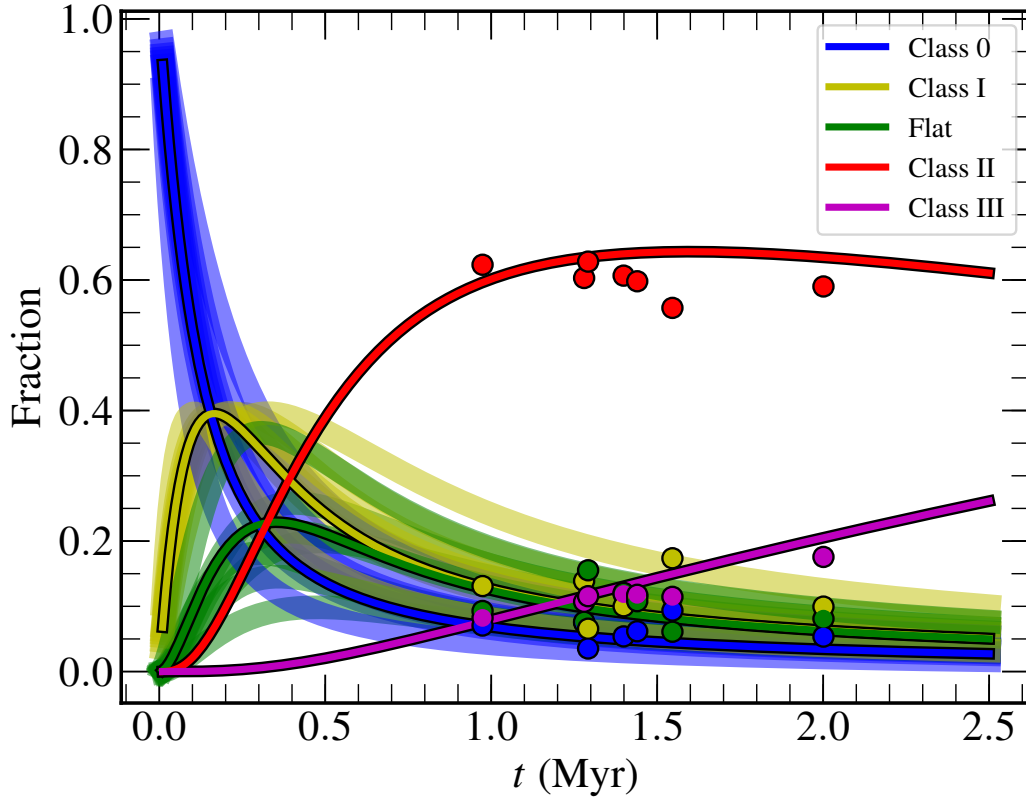


FIGURE 2.4: Fractions of each protostellar class as function of time. Dots show to the observed fraction. Thin lines are representing the result of the studies entire population corresponding to equation 2.2, with the mentioned values for λ_D . The thick lines are representing different sub-populations of their data. Credit: [Kristensen & Dunham \(2018\)](#)

The age distribution depends on several factors the free fall timescale defined as:

$$t_{\text{ff}} = \frac{0.5427}{\sqrt{G\rho}} \text{Myr},$$

where $G = 4.43 \times 10^{-3} \text{pc}^3 M_{\odot}^{-1} \text{Myr}^{-2}$ and ρ is the density of the cluster in units of $M_{\odot} \text{pc}^{-3}$. The age of the cluster is then estimated to be proportional to this free-fall time, with a random factor between 0 and 1 representing the maturity of the cluster, 0 being "newborn" and 1 being completely collapsed and a variable scale factor here used as inverse of the one described in: [Utomo et al. \(2018\)](#)

$$\epsilon_{\text{ff}} = \frac{t_{\text{dep}}}{t_{\text{ff}}}$$

representing the depletion time in units of free fall timescales, or in other word how many free fall times it takes for a cluster to finish star formation. The scale factor is added to account for the cluster collapse time being influenced by other energies than just the gravitational potential energy, \mathcal{M} , as assumed for the free fall timescale from the virial theorem, equation 1.1.

The distribution of the different stages of star formation is taking from [Kristensen & Dunham \(2018\)](#) in which they reveal a new methodology for inferring a half-life for each protostellar class. The half-life concept is similar to the same concept known from nuclear decay, where one nucleus decays to another with an associated half-life. To assume a similar type of half-life for stages of a protostellar life the assumptions that the evolution of stars is continuous and sequential going through stages: Class 0, Class I, Flat, Class II and ends in Class III with constant star formation rate and half-life. With these main assumptions the population of protostars in Class D is given by:

$$N_D(t) = \frac{\lambda_{\text{SF}}}{\lambda_D} \left[1 - \exp(-\lambda_D * t) \right], \quad (2.2)$$

with t being the age of the cluster, $\lambda_{\text{SF}} = N(t)/t$ being the number of stars created to date over the age of the cluster (Star Formation Rate, SFR), and λ_D being the "decay" constant associated with the half-life $\lambda_D = \ln(2)/t_{1/2}$. In [Kristensen & Dunham \(2018\)](#) these decay constant are estimated from observations of 7 clouds to be $\lambda_0 = 14.7 \text{Myr}^{-1}$, $\lambda_1 = 7.9 \text{Myr}^{-1}$ and $\lambda_{\text{flat}} = 8.0 \text{Myr}^{-1}$ which is the values used in the CiB and GiB models. The evolution of the species as a function of time is shown in figure 2.4. The Class III objects does not have an estimated λ_{III} but it is rather assumed that the remaining population not falling into the other categories are Class III objects meaning $N_{\text{III}} = N_{\text{tot}} - N_0 - N_I - N_{\text{II}} - N_{\text{flat}}$ which is a simplification but does not raise concern since the main interest of this thesis are the ones included in the database meaning Class 0, I and II objects.

Initial Mass Function

The Initial Mass Function (IMF) follows that described in [Chabrier \(2003\)](#).

$$\xi = A_1 \times \exp \left[- \frac{(m - \log(mc))^2}{2 \times \sigma^2} \right], \quad m \leq 1 M_{\odot} \quad (2.3)$$

$$\xi = A_2 \times (10^m)^{x_0 - \beta}, \quad m \geq 1 M_\odot \quad (2.4)$$

Where the two normalisation constants are as default set to $A_1 = 0.158$ and $A_2 = 4.43 \times 10^{-2}$, m is the mass of the protostar, $m_c = 0.079 M_\odot$, $\sigma = 0.69 M_\odot$, and x_0 is the powerlaw index set to $x_0 = -1.3$, β is the IMF type which is a parameter to control the slope of initial mass function for mass larger than $1 M_\odot$, through this thesis 0, 1 or 2 are corresponding to standard, top heavy and bottom heavy IMF and also referred to as such.

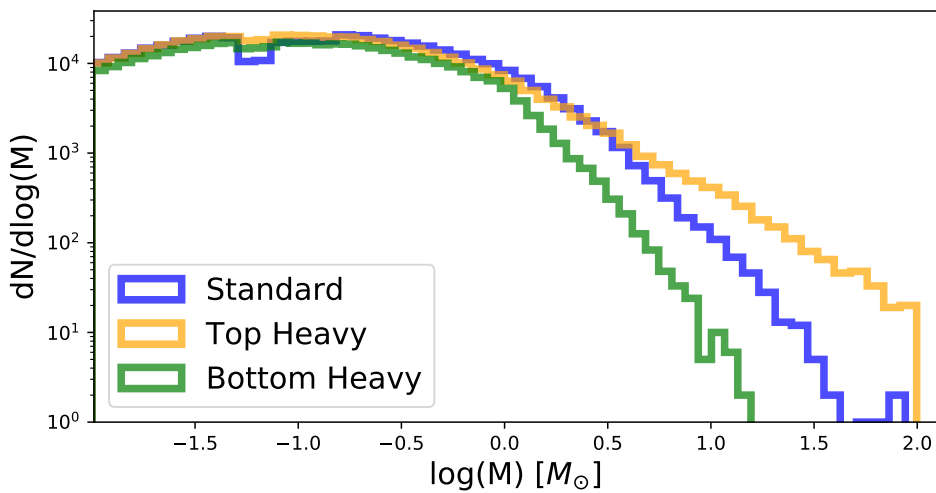


FIGURE 2.5: Plot of the mass distribution of stars in the CiB model. Standard ($\beta=0$), Top heavy ($\beta=1$), Bottom heavy ($\beta=2$). bin-width = 0.18

The β parameter is put to make it particularly easy for the user to change the distribution of the more well studied protostellar objects with a mass above $1 M_\odot$. In figure 2.5 the effects of the β parameter can be seen to steepen or flatten the power-law tail of the IMF. This makes the GiB/CiB model capable of investigating the influence of different mass distributions.

2.3.1 Mass distribution

The mass of each star in the cluster is assigned to be in the range between $m_{\min} = 0.01 M_\odot$ and $m_{\max} = 100 M_\odot$. The Molecular Cloud Mass itself is determined by the variable parameter M_{cm} , described in 2.3.3. The model is made so that a chunk of stars within the defined mass range is added to cluster with a distribution corresponding to the IMF type chosen by the parameter β . Stars are added until the defined Molecular Cloud Mass (M_{cm}) is reached so that the star forming part of the star cluster has a mass given by the total mass multiplied by the Star Formation Efficiency (SFE):

$$M = M_{cm} \times SFE \quad (2.5)$$

TABLE 2.3: Overview of Input Parameters in Galaxy-in-a-Box

Variable Name	Input Description	Default
β	Slope of powerlaw tail of initial mass function for masses larger than $1 M_{\odot}$: 0=standard,1=top and 2=bottom heavy IMF	0
ϵ_{ff}	Free fall time scaling factor, SFE per free fall timescale.	1
SFE	Star Formation Efficiency	0.03
Mcm	Molecular cloud mass in solar masses, in galaxy-in-a-box it is determined by randomly generated array of masses.	-
L_bol	Bolometric luminosity of observed object given in units solar luminosity L_{\odot}	-
Dist	Distance to observed object in units of parsecs.	-
M_env	Envelope mass of observed object given in units solar mass M_{\odot}	-
Flux	Emission line flux from object in specific transition and molecule of interest. Units of K km s^{-1}	$^{12}\text{CO J}=10-9$
N_{cl}	number of clusters in galaxy	10^4
mmin	Minimum mass of a protostar in this model, in units of solar masses M_{\odot} .	0.01
mmax	Maximum mass of a protostar in this model, in units of solar masses M_{\odot} .	100
n₀	Initial number of stars.	300
R₀	Step size of Distance from center of cluster.	1 pc
p	powerlaw index of radial probability function.	1.5
alpha	Powerlaw index of maximum cluster radius.	0.33
tol	tolerance of intensity/envelope mass fit.	0.6827
A_{inner}	Scaling factors for spatial galaxy distribution, for inner and outer part of the galaxy.	4
A_{inner}		1.6
B_{inner}	Arm sweep of spatial distribution of galaxy for both inner and outer part of galaxy.	6
B_{outer}		2.5
N_{inner}	Winding number of spatial distribution of galaxy in inner and outer part of galaxy	12
N_{outer}		32

The SFE is put here as a free parameter in order to define how much of the mass is effectively forming stars which is important since the mass not actively taking part in star formation is not of interest in this contexts since it is assumed to not emit light corresponding to the transition of interest. The values of SFE are chosen to be

between 3 % which is about an average SFE seen in the lifetime of a GMC, and 30 % which is an order of magnitude bigger but is necessary when considering the model in extragalactic objects at high redshift as shown in figure (1.1).

2.3.2 Emission

When the individual stars in a cluster is placed in the CiB model the emission from the cluster is possible to estimate. The interest here being the emission from high-J CO transitions but in principle any type of emission should be possible to estimate assuming the emission scales with envelope mass. The model takes the input observations of flux and distance and turn them into line intensity using the relation:

$$I_{\text{CO}} = \frac{4\pi \times \text{flux} \times (d/d_0)^2}{A_{\text{beam}}}, \quad (2.6)$$

where d is distance to the flux source, d_0 is the distance to the modelled cluster and A_{beam} is the area of the beam used to observe the flux. When the line intensity is calculated the model correlates the line intensity with envelope mass (see figure 2.6). The model first tries a linear fit; and compares the p -value to a tolerance value (default $1\sigma=0.6827$), the p -value corresponds to the chance of the observed values coming from the fit distribution. If the p -value is less than the set tolerance it will check to see if a power-law fit is more suiting.

When a proper fit is found, the model uses the masses assigned to the protostars in the model cluster to determine an intensity for each protostellar object. Only Class 0 and Class 1 protostars are included in the model since these are the ones with outflows. A scale factor of 0.1 is multiplied to the Class 1 objects since these typically has an order of magnitude lower outflow activity (Bontemps et al. 1996).

2.3.3 From cluster to galaxy

In order to make the model go from CiB to GiB a distribution of cluster masses and spatial distribution within the galaxy is required. This will be described in detail below.

Cluster mass distribution

The model has a default cluster mass distribution which can easily be replaced to fit the users wishes. The default one is defined from Krumholz (2015) with one sligth change being the addition of a lower mass threshold:

$$\frac{d\mathcal{N}}{dM} = \begin{cases} \mathcal{N}_u \left(\frac{M_u}{M}\right)^\gamma, & M \leq M_u \\ 0, & M > M_u \\ 0, & M \leq M_l \end{cases} \quad (2.7)$$

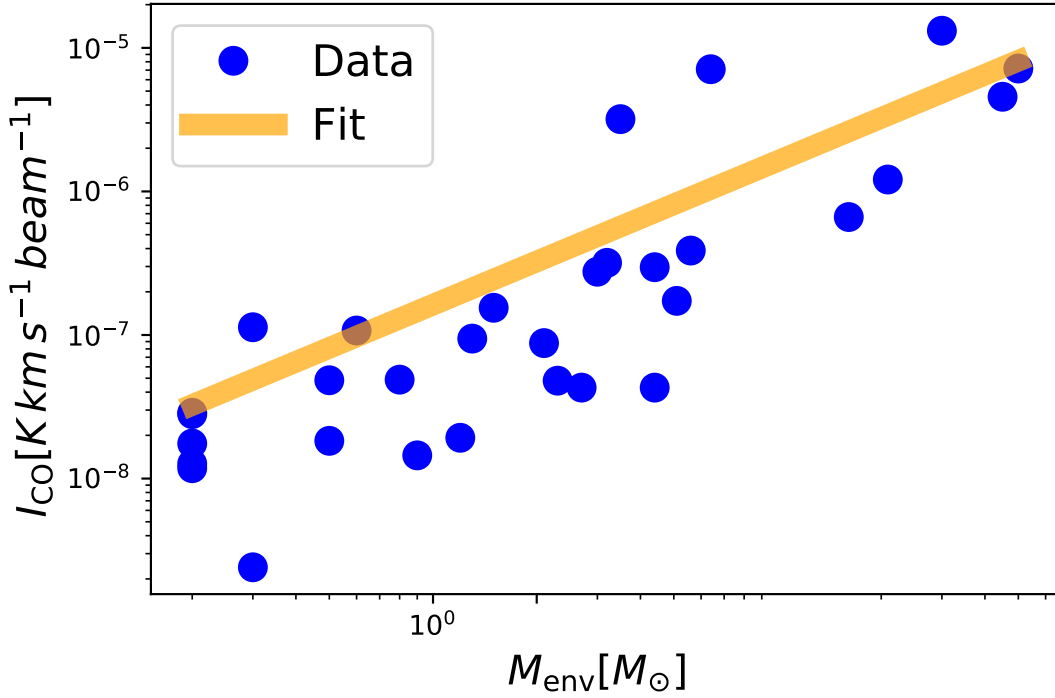


FIGURE 2.6: Correlation plot of line intensity I_{CO} and envelope mass, M_{env} as calculated from equation 2.6 of the line CO J=10-9 imported from the database, fit shown is done by the GiB model

Here \mathcal{N} is the number of clusters with Mass M , M_u is the highest mass occurring in the distribution default set to be $10^6 M_{\odot}$, M_l is the lower mass threshold default set to be $10^4 M_{\odot}$ and γ is the slope of power-law as default set to -1.64 (Roman-Duval et al. 2010).

With this mass distribution or a user input mass distribution the GiB model can produce the desired amount of clusters for creating a galaxy. As default the model is set to make 10,000 clusters but this can be changes via the variable N_{cl} .

Spatial cluster distribution

The spatial distribution of cluster in the galaxy simulated by the GiB model is that of a spiral galaxy, since the most star formation is seen in spiral galaxy and merger galaxies (which the updates to GiB might incorporate better). The distribution is divided into two parts, one describing the inner and one describing the outer region of the galaxy both given by a spiral arm shape determined from the formula for describing spiral galaxy shape from Ringermacher & Mead (2009):

$$r(\phi) = \frac{A}{\log \left(B \times \tan(\phi / (2N)) \right)} \quad (2.8)$$

Where A is a scaling factor by default set to 4 for the inner region and 5 for the outer region. B is arm sweep by default set to 2.5 for the inner, and 3.5 for the outer part of the galaxy. N is the winding number by default set to 12.0 and 20.5 for the inner and

outer part. ϕ is the angle. The structure of the spiral galaxy with inner and outer part can be seen in figure 2.7.

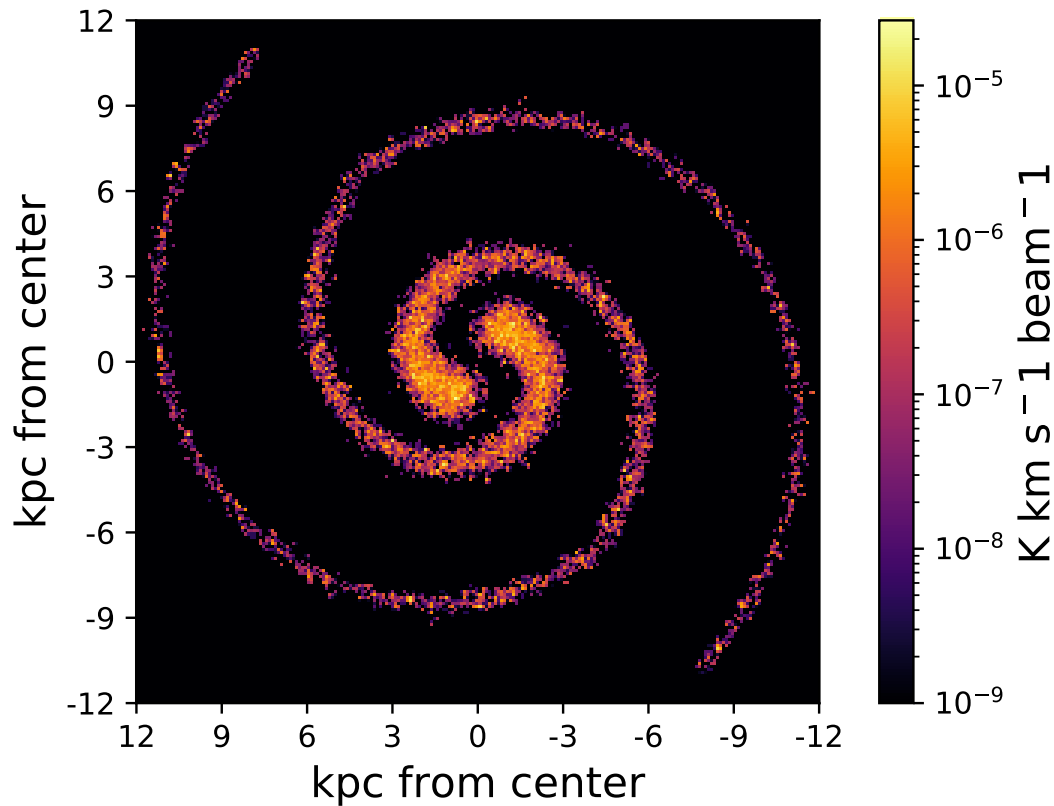


FIGURE 2.7: Spacial distribution of Galaxy in a Box default spiral galaxy. The model was here run with CO J = 10-9 from the database and all remaining values set to default

Like for the stellar cluster in the CiB model the GiB model has a factor for the spatial distribution or more specifically the surface-density which takes into account the gravitational bias to place more mass towards the center of the Galaxy. This surface density is defined as from Casasola et al. (2017):

$$S = \exp(-R * 1/h_0) \quad (2.9)$$

Where R is the radius from center in kiloparsecs and h_0 is the scale length also in kiloparsecs for M51 it is found by Casasola et al. (2017) to be 2.38 kpc.

2.4 Optimization

In order to make the ease of using the GiB model, I have gone through the code and upgraded the performance of the model to make it run in a shorter amount of time. All timestamps discussed in this section is made with my computer (MacBook Pro 2011 with 480 Gb SSD and 16 Gb of RAM) and will vary significantly depending on the machine used, which is why I have given the relative numbers to all times as

well. All changes I have made I have initially tested outside of GiB in smaller sections, I have thoroughly tested these changes do not alter the results but only improve the runtime. In order to quantify the improvements I timed a run of both the original and changed section a hundred times in a for-loop (I refer to such test as "100-run test" from hereon). As default the model creates 10,000 stellar cluster meaning it has to loop 10,000 times one time for each stellar cluster so the numbers created by the 100-run test can be multiplied by 100 to get an estimate of the improvement in the full GiB model. I have also tested all changes one by one, running the entire model making sure no errors are carried from the change and through the model script. The total runtime for GiB was originally \approx 1 hour 25 minutes for the model with all values set to default which will also be the case for all other timings discussed. After all the changes described below I made to the model the runtime of GiB is \approx 17 minutes or a relative performance improvement of 80 %.

2.4.1 Input/output fileformats

In the original model the spatial distribution of each cluster is written and saved to a .dat file. In the GiB script the distribution is read to make each of the \approx 10,000 cluster for a galaxy. Consequently the model reads and writes 10,000 temporary files to determine the distribution of individual stars in each cluster. Although it is very useful for the CiB model to be able to manually read off the distribution file which is possible in the .dat format, when going to galactic scale where 10,000 temporary files are made, it is no longer of any value to be able to read each file made in each simulation. Therefore I have decided to change the format to the much more time efficient non human-readable format .npy. The .npy format is the native format Python/Numpy uses for Numpy arrays written in the language C and therefore very efficient for input/output in Python since it is not required for the computer to change the formatting of the data to write or read it to/from a file. For the .npy changes I split the 100-run test into two sections, one for the input/writing the file, and one for the output/reading the file. For input the .npy version outperformed the original with a time of 0.92 seconds vs the .dat version with a time of 18.53 seconds or a relative improvement of 95.0 % . For the output section the .npy version outperformed the original with a time of 0.07 seconds vs the .dat version with a time of 14.29 seconds or a relative improvement of 99.5 %.

2.4.2 Elementwise calculations with NumPy

In the original GiB model a lot of array and value manipulation was done using *for* loops. In order to optimize the model I changed a lot of these manipulations to be done via the Python package NumPy which has element-wise calculations built in to its arrays. An example of this is in the distribution of the individual stars in the cluster, which in the original script looked like:

```
x = numpy.zeros(N)
y = numpy.zeros(N)

for i in range(0,N):
    x[i] = rad_m[i]*cos(phi[i])
    y[i] = rad_m[i]*sin(phi[i])
```

Where it can be seen that a *for* loop has to run N times, and the variables x and y have to be defined to have the correct shape before the loop runs. I have changed this to the significantly more effective NumPy format:

```
x = rad_m[:] * numpy.cos(phi[:])
y = rad_m[:] * numpy.sin(phi[:])
```

These two codes produce the exact same values as long as the variables rad_m (radius) and phi (angle) is defined with length N. N here is the number of stars in the cluster defined by the mass distribution. Running GiB 3 times, one for each IMF type, I found the average value of N to be 10.568 and used this number to estimate the time optimization of the change using the 100-run test. The original code ran in 2.73 seconds, the new code in only 0.01 seconds, an improvement of 99.5 %. In total I have made 5 changes of this type with similar time improvements replacing *for* loops with NumPy operators.

2.4.3 Mass distribution variable

In order to bring down the runtime of the GiB model I took a look at the star mass distribution in each cluster, since this is one of the most computationally time consuming parts of the script (after updating the input/output). Originally it was set to make a fixed "chunk" of 1000 stars from a uniform distribution, and then reject the stars that fell outside of the IMF and keeping those that fell inside. Repeating this until the stars that were kept through the IMF had a mass of the input Molecular Cloud Mass(Mcm). In order to reduce the runtime of GiB I made a test run of a chunksize that varies with molecular cloud mass. I set the new chunksize to be $M_{cm} \times C$, C being a constant determined from the runtime test to be 0.3 (See figure 2.8). In figure 2.8 I plotted 100-run test times for different values of the constant C. I also plotted it for different values of Mcm to make sure it did not significantly alter the effects of the constant C. In the figure the blue line representing $M_{cm}=10^4$ the time is multiplied by 5 to make it easier to look at in the plot which also highlights that the runtime is mostly impacted by the molecular cloud mass so the high Mcm (orange lines) should be of most concern. It is seen from the figure that the fixed chunksize of 1000 was a suboptimal choice for both cases. Adding the variable chunksize= $M_{cm} \times 0.3$ made for an optimization of the mass distribution section of CiB of 45.7 %. Even after this improvement the mass distribution section of the script is still one of the most computationally time consuming,

and if the model was to be further optimized this would be a good place to further improve.

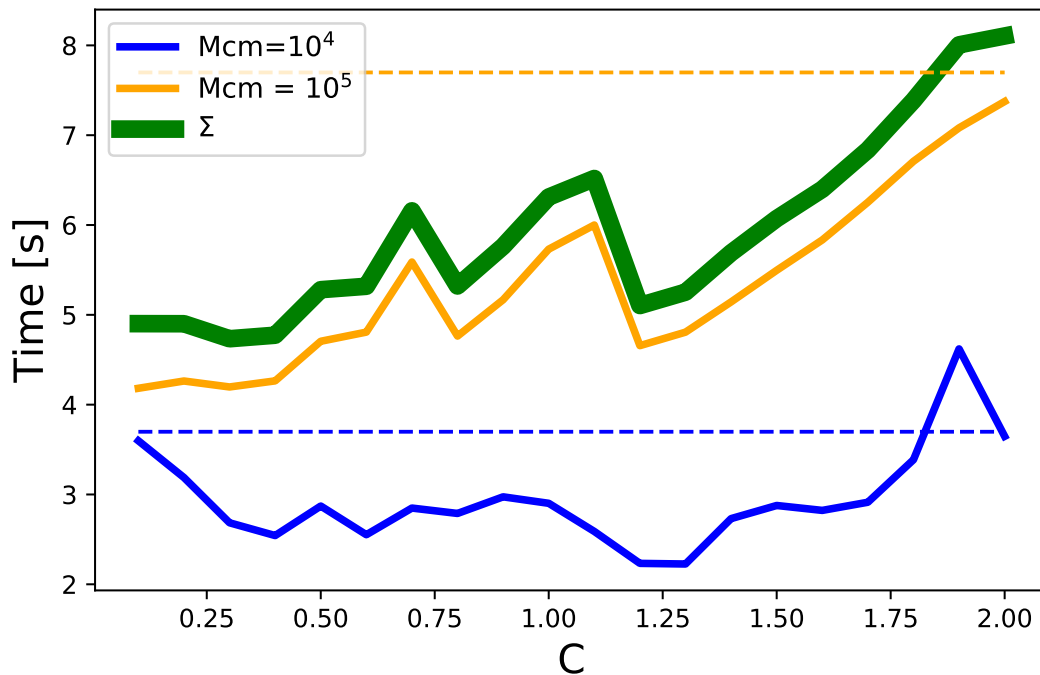


FIGURE 2.8: Plot of 100-run times for the individual star mass distribution function, the x-axis is the value of the constant on the chunksize in the distribution. The green line is the sum of the orange and blue (the time on both blue lines are here multiplied by 5 for illustrative purposes). The stipled lines mark for the fixed chunksize of 1000 for the same Molecular Cloud Masses as the coloured lines.

2.4.4 Print results

In the original CiB code the total mass, emission and number of stars were printed to the console. This was a very good idea since it provided an easy way to read the results. In GiB the same 3 results, total mass, emission and number of stars are written to a file to be usable for further analysis. Originally the GiB model as well as writing to an output file also in the same way as CiB printed the results to the console which is unnecessary when it also writes it to a file. In addition to being a bit unnecessary in might also cause the code to be less resilient when running for a large number of clusters, e.g 3 results times 10,000 clusters is 30,000 lines printed to your screen doing a run of the model. For some machines displaying this much data on the screen in the time it takes to run the model could possible make the machine struggle or crash. I solved this problem implementing an output switch which can be turned on or off by the user before running the code to decide if a print to console is desired. This eliminates the problem with big amounts of cluster, but also allows the user to quickly read the data in the console if desired. The switch is turned on (print) as default for CiB and turned off (no print) as default for GiB.

Results

3.1 Database

In this section I am going to go through the content of the database at the time of submission for this thesis. The database consists of a total of 190 different objects. In figure 3.1 the type of objects are specified and the distribution of them explicitly shown. Contained is 21 (11.0%) Class 0 Protostellar objects, 18 (9.5%) Class 1 Protostellar objects, 38 (20.0%) Class 2 Protostellar objects, 10 (5.2 %) High Mass Protostellar objects and 103 (54.2%) Extragalactic objects.

The 190 objects are all observed in at least one CO transition and maximally 43 CO

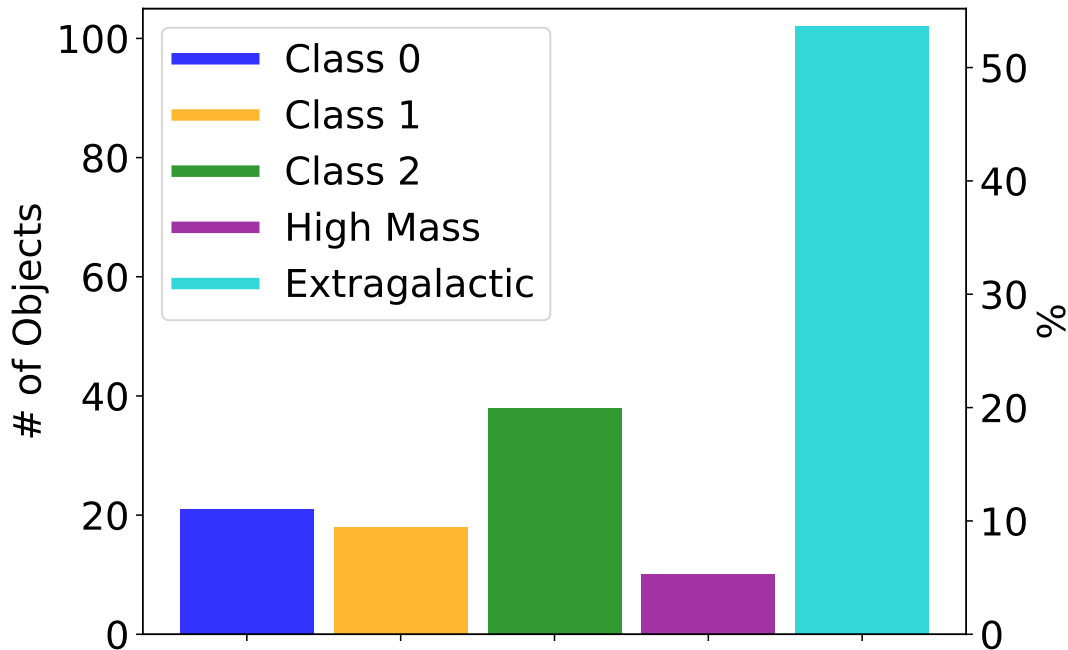


FIGURE 3.1: Distribution of object types in the database.

transitions between CO J=4-3 and CO J=48-47 a total of 2739 emission line detections are included in the database. How many objects has been observed in each transition is shown in figure 3.2. In the figure it is seen that the most observed emission line is CO J=10-9 which is observed in 157 (88.2 %) of the objects. Of the observed emission line the line observed towards the fewest objects is CO J=48-47 which is only observed towards 17 (9.1%) objects. Two different regimes is seen in figure 3.2 below and above

$J_{up}=13$ a switch is seen where the lines stop being observed towards a majority of the database population. The jump between these regimes is caused by the switch from SPIRE (below $J_{up}=13$) to PACS (above $J_{up}=13$). The entire spectral ranges and which transitions each instrument could observe is described in further detail in section 2.1.3 and is illustrated as a broad overview in figure 2.1. In particular the well represented extragalactic objects are often only observed with SPIRE and therefore not observed in a majority of objects outside of the spire range $J_{up}=4$ to $J_{up}=13$. Another feature of this histogram in figure 3.2 which meets the eye is the emission lines from the $J_{up}=26$ and $J_{up}=27$ transitions being absent. The absence of the $J_{up}=26$ and $J_{up}=27$ transitions is due to a spectral leakage in the PACS instrument (Poglitsch et al. (2010)) in the wavelength range 95-110 μm in which the two specific transitions fall. The data in the leaked spectral range are therefore always discarded before published and these two transition has not been observed using Herschel.

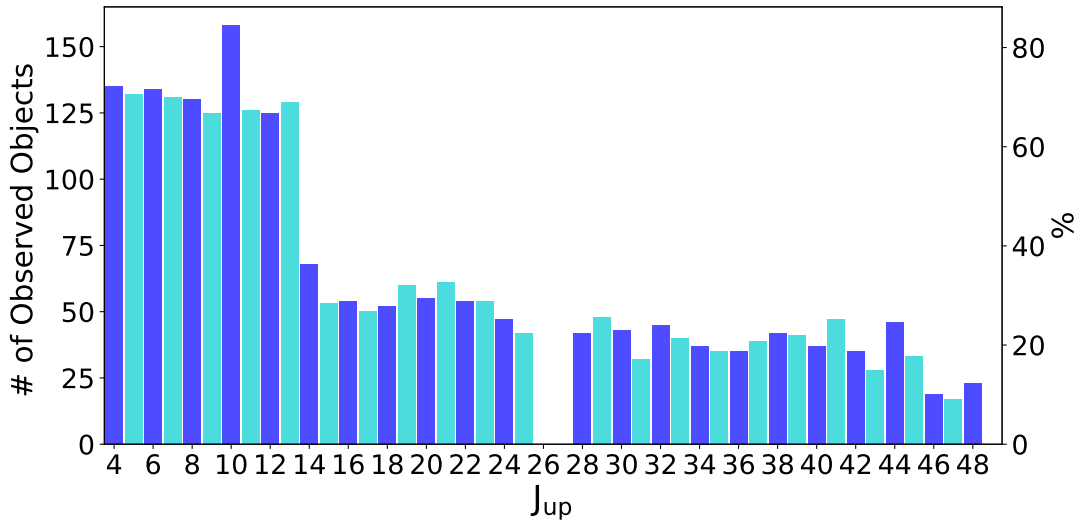


FIGURE 3.2: Distribution of observed emission lines, described here by the upper level J_{up} . On the left y-axis is number of objects observed in the line, on the right axis the percentage of the total database population the line is observed for.

3.2 Line Luminosity vs Luminosity

Figure 3.3 shows the line luminosity of the line $^{12}\text{CO} J=10-9$ for a specific object plotted against the bolometric luminosity of the same object. The line luminosity is found using the formula for line luminosity given in Wu et al. (2005) as:

$$L_{\text{CO}} = 23.5 \times 10^{-6} D^2 \frac{\pi \theta_s^2}{4 \ln 2} \frac{\theta_s^2 + \theta_{\text{beam}}^2}{\theta_s^2} \int T_R dv \quad (3.1)$$

Where D is the distance to the object in parsecs, θ_{beam} is the angular size of the beam, θ_s is the angular size of the source, and $\int T_R dv$ is the flux in K km s^{-1} . A total of 33 galactic objects are included in the figure, all Class 0 and Class 1 objects from the database which is observed in the $^{12}\text{CO} J=10-9$ line with HIFI. Class 2 Objects are

excluded. The objects are here split into four categories; Class 0 and Class 1 are protostars of the mentioned class and IM and HM are corresponding to Intermediate mass young stellar objects (YSO) and HM to High mass YSO. The fit is a powerlaw fit as used in [San José-García et al. \(2013\)](#):

$$L_{\text{CO}} = 10^a \times L_{\text{bol}}^b \quad (3.2)$$

With L_{CO} being the line luminosity of CO J=10-9 and L_{rmbol} is the bolometric luminosity. The parameters of the fit a and b are found using a weighted fit with the `scipy.optimize` package in Python to be $a=-3.89\pm 0.13$ and $b=0.94\pm 0.07$. In this case it is important to be aware of the logarithmic scaling since both axis spans over 5 orders of magnitude which make the fit qualitatively to seem to be sub optimal. A goodness-of-fit analysis results in a coefficient of determination $R^2=0.99$ which tells that the fit captures the features of the data well. One thing to consider is how much brighter the single high mass data point is compared to the rest, this point has a massive impact on the fit but ultimately a strong correlation is measured from the Pearson correlation coefficient $r=0.99$.

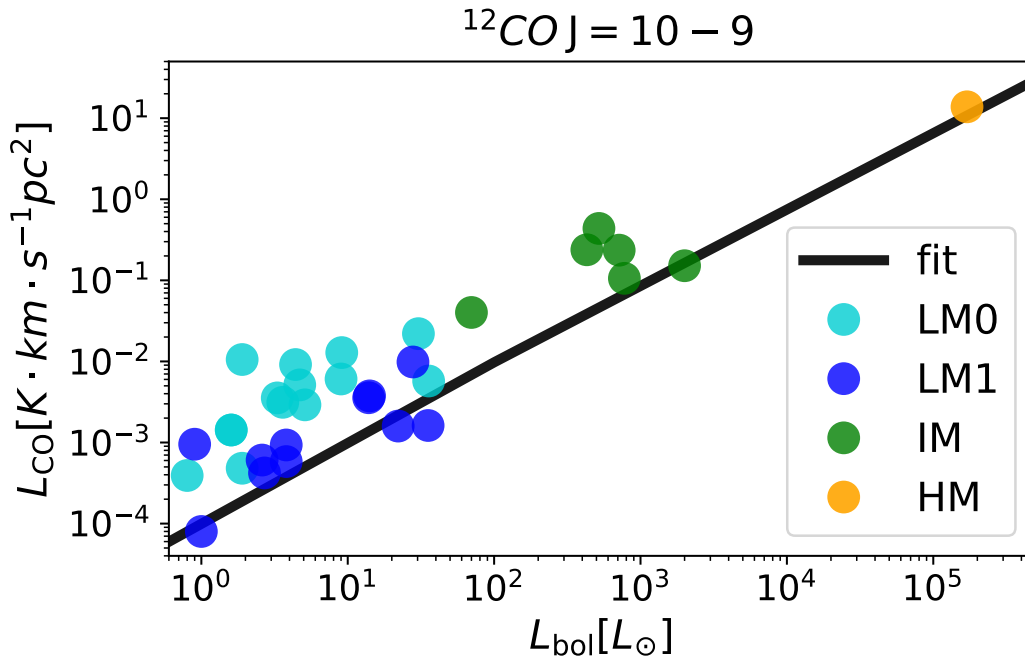


FIGURE 3.3: Line luminosity of ^{12}CO transition J=10-9 on y-axis plotted against bolometric luminosity of object on x-axis. The fit; $L_{\text{CO}} = 10^{-3.89} \times L_{\text{bol}}^{0.94}$ is a powerlaw fit. HM: High Mass object, IM: Intermediate Mass object, LM0: Low mass Class 0 object, and LM1: Low Mass Class 1 object

3.3 SLEDs

3.3.1 Raw Data

In figure 3.4, 3.5, 3.6 and 3.7 SLEDs for respectively Class 0, Class 1, Class 2 and Extragalactic objects with more than two emission lines observed to a satisfying quality here defined as a Signal-to-noise (S/N) ratio of more than 5 are plotted directly from the database to give a visual insight to the emission intensity of each lines compared to one another. The x-axis is here defined as the upper energy level (J_{up}) of the specific rotational transition of ^{12}CO . The y-axis is the logarithm (\log_{10}) of the normalised intensity, normalised with respect to all the emission from the plotted lines of that specific object.

In figure 3.4 16 of the databases total 21 Class 0 objects are represented the two excluded objects not having at least two points with the desired quality of $S/N > 5$. In table 3.1 the median, average and standard deviation of the peaks of the SLEDs in figure 3.4 with 6 or more observed emission lines are displayed as raw data values of Class 0 objects. In figure 3.5 16 of the database total 18 Class 0 objects are represented the excluded objects not having at least two points with the desired quality of $S/N > 5$. In table 3.1 the median, average and standard deviation of the peaks of the SLEDs in figure 3.5 with 6 or more observed emission lines are displayed as raw data values of Class 1 objects.

In figure 3.6 10 of the database total 38 Class 0 objects are represented again with the excluded objects not having at least two points with the desired quality. Compared to Class 0 and Class 1 objects a significantly smaller fraction of the Class 2 objects are represented in the SLED plot, this is because a smaller fraction is reaching the desired quality of $S/N < 5$. The reason for this is Class 2 objects having rare and weak outflows which is the primary source of high J CO emission. This causes a weaker signal for the emission while the noise is still comparable and causes the S/N to be lower for most Class 2 objects. On average the S/N is $\mu_{Class2}(S/N) = 2.632 \pm 0.004$ for Class 2 Objects which is below the desired quality of 5, and also low as compared to Class 0, $\mu_{Class0}(S/N) = 16.14 \pm 0.04$ and Class 1 $\mu_{Class1}(S/N) = 12.25 \pm 0.02$ which are both significantly higher than the desired threshold. In table 3.1 the median, average and standard deviation of the peaks of the SLEDs in figure 3.6 with 6 or more observed emission lines are displayed as raw data values of Class 2 objects.

In figure 3.7, 97 of the database total 103 extragalactic objects are represented the excluded objects not having at least two points with the desired quality of $S/N > 5$. This figure illustrates the same as figure 3.2 that most extragalactic objects has only been observed with in the SPIRE regime and therefore only to $J_{up}=13$, as mentioned and as clear from figure 3.4, 3.5, 3.6 and this figure 3.7 the SPIRE regimes domination is mostly due to the fact that most extragalactic sources have not been observed with PACS very often. A few extragalactic objects has been observed with PACS and a sufficient S/N of 5 in transitions up to $J_{up}=28$ and are also shown in the figure. In table 3.1 the median, average and standard deviation of the peaks of the SLEDs in

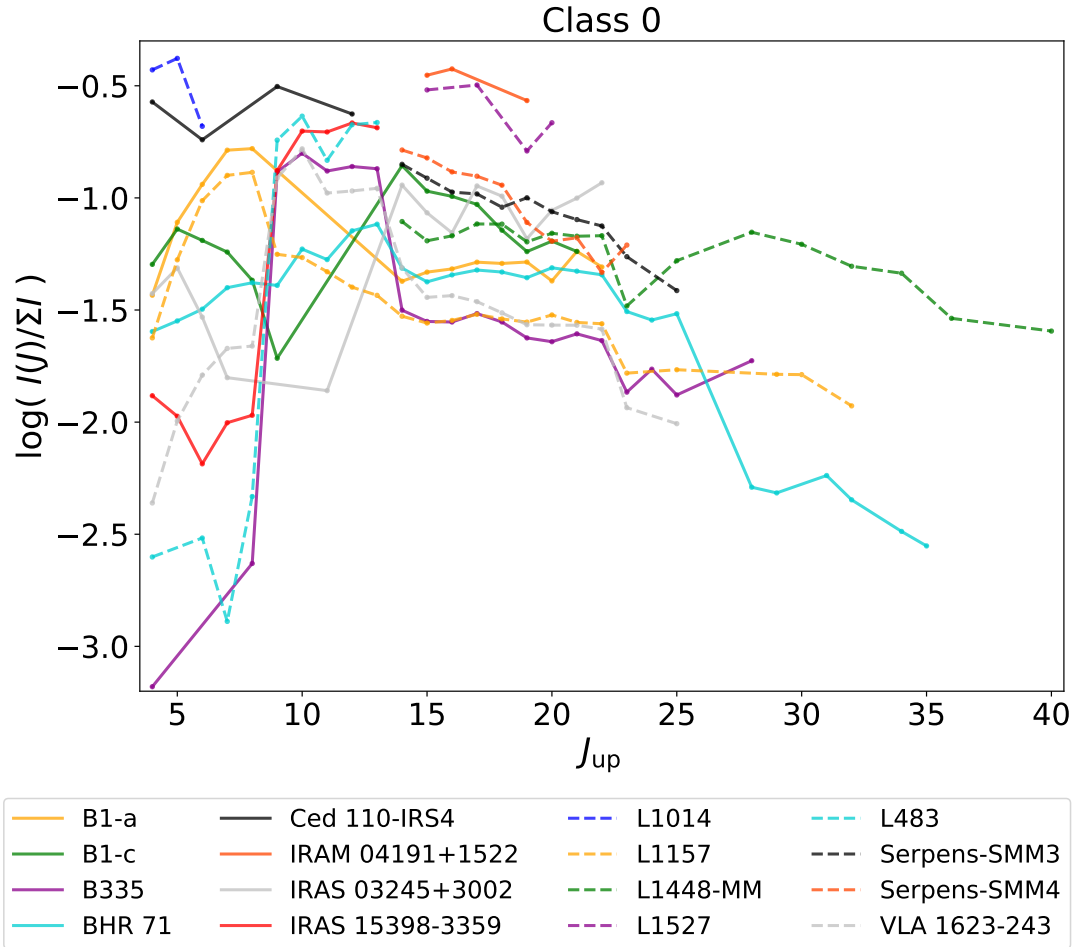


FIGURE 3.4: CO spectral line energy distributions for all Class 0 objects included in the database with at least 2 observed lines with $S/N > 5$.

figure 3.7 with 6 or more observed emission lines are displayed as raw data values of extragalactic objects.

TABLE 3.1: Overview of SLED peak transitions

Object Type	Raw data			Spline fit		
	Median	Average	σ	Median	Average	σ
Class 0	12.5	12.2 ± 0.8	4	14	14.2 ± 1.2	4
Class 1	13.5	14.4 ± 0.9	6	19	18.3 ± 1.5	5
Class 2	21.0	20 ± 4	5	24	21 ± 4	4
Extragalactic	6.0	6.06 ± 0.06	2	6	6.2 ± 0.2	2

Notes. Median, average and standard deviation (σ) values for raw and spline fit data illustrated in figures 3.4, 3.5, 3.6, 3.7, 3.10 and 3.8. The values are found from the peak transition of each object with at least 6 observed transitions with $S/N > 5$.

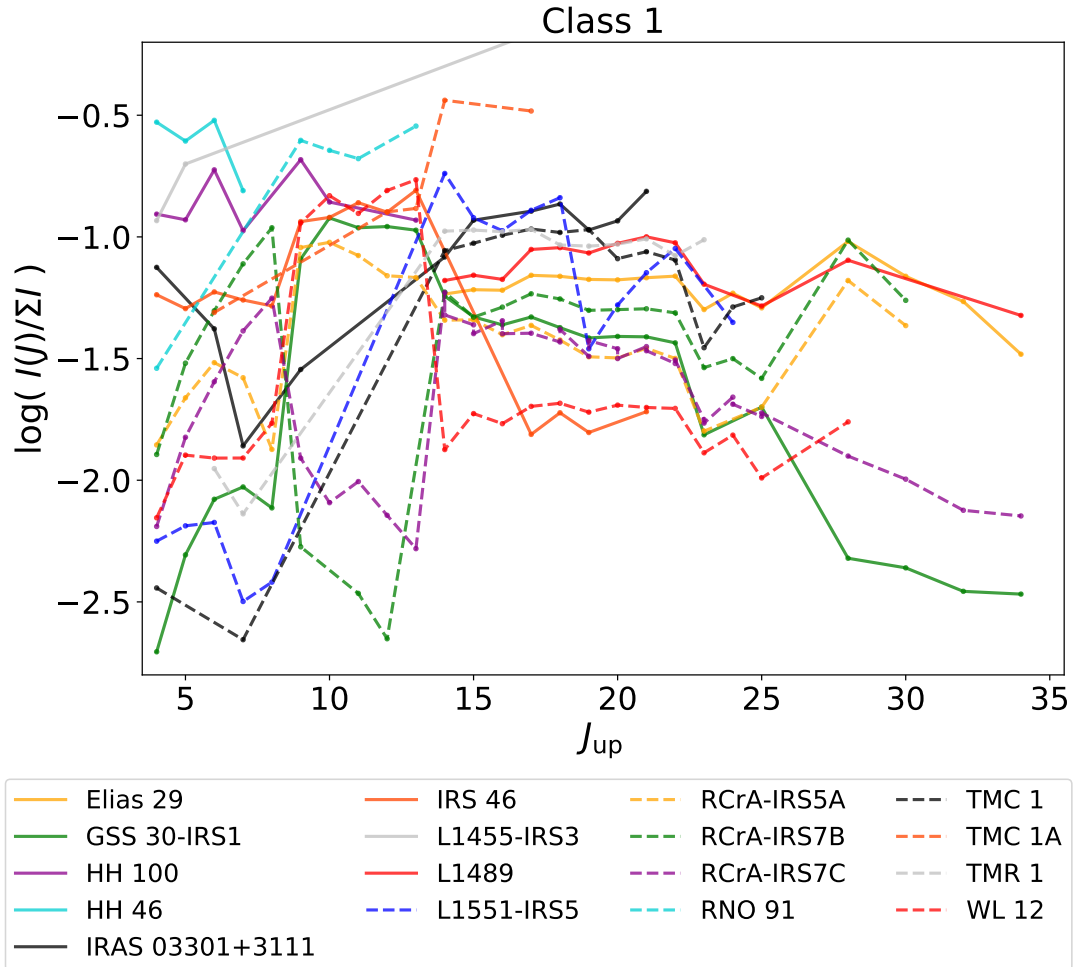


FIGURE 3.5: CO spectral line energy distributions for all Class 1 objects included in the database with at least 2 observed lines with $S/N > 5$.

3.3.2 The Spline fit

In figure 3.8 the same type of SLEDs are plotted for galactic objects with 6 or more observations with $S/N > 5$. As opposed to figure 3.4, 3.5, and 3.6 where the raw values from the database are plotted in figure 3.8 a spline fit has been made to the SLED values for each object. The spline fit was done using the fit function "UnivariateSpline" from the Python package "scipy.interpolate". 3 variables are to be set in the UnivariateSpline function, s, k and w . s being the number of knots the spline fit should do, here set to 2 for a polynomial, k set to 2 being a smoothing factor here chosen from trial and error experiments looking for the best qualitative fit, w being the inverse of the error on the normalized intensity, found using error propagation (see equation 2.1) to be

$$w = \left(\frac{\sigma_I}{I * \ln(10)} \right)^{-1}$$

Where I is the intensity and σ_I is the error on the intensity.

In figure 3.8 it is seen that the overall shape of the SLEDs are similar for the three types of objects with exception of a few objects that stands out. Two of the objects

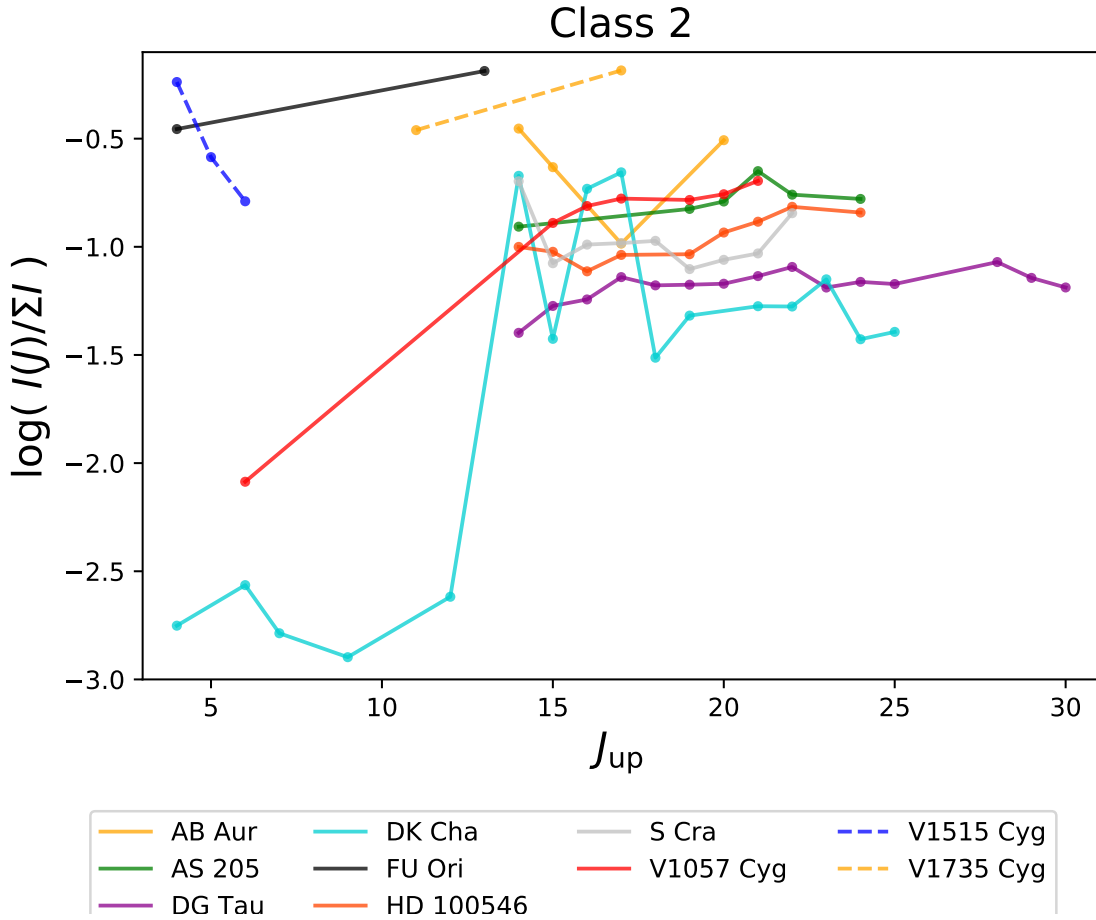


FIGURE 3.6: CO spectral line energy distributions for all Class 2 objects included in the database with at least 2 observed lines with $S/N > 5$.

TABLE 3.2: P-Values for galactic SLEDs

Object Type	Raw data			
	Class 0	Class 1	Class 2	Extragalactic
Class 0	-	1.0	0.08	<0.01
Class 1	1.0	-	0.47	<0.01
Class 2	0.08	0.47	-	
Extragalactic	<0.01	<0.01	<0.01	-
Object Type	Spline fit			
	Class 0	Class 1	Class 2	Extragalactic
Class 0	-	0.04	0.22	<0.01
Class 1	0.04	-	0.88	<0.01
Class 2	0.22	0.88	-	0.05
Extragalactic	<0.01	<0.01	0.05	-

Notes. P-values of both raw data and spline fits for each protostellar class object with respect to each other. Values are calculated using Python package "scipy.stats".

that visibly stands out are the Class 0 object *L1157* and the Class 2 object *S Cra* which

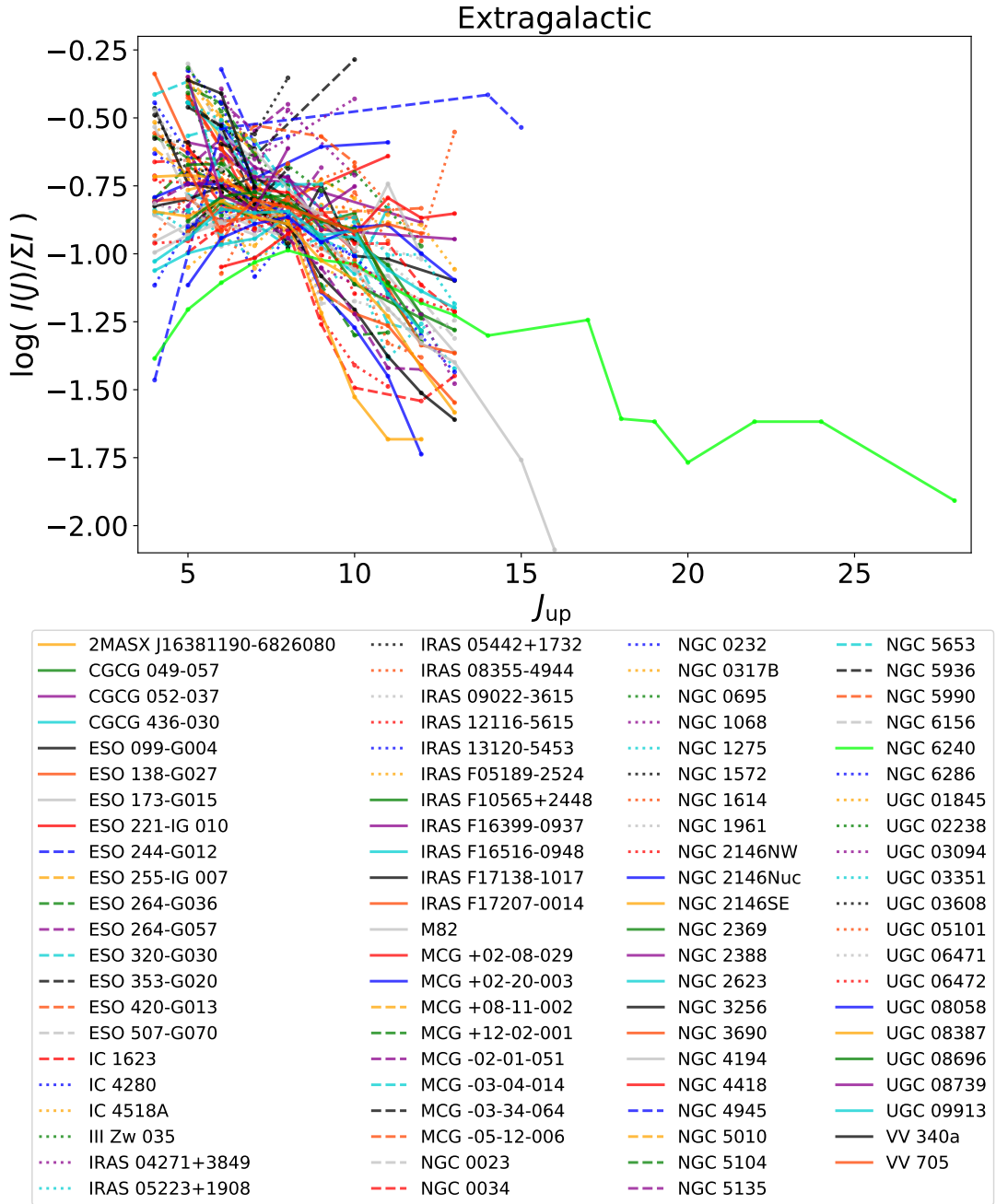


FIGURE 3.7: CO spectral line energy distributions for all Extragalactic objects included in the database with at least 2 observed lines with $S/N > 5$.

are shown individually in figure 3.9. *S Cra* is a Class 2 object and therefore blue in figure 3.8 and can be easily recognized as the only positive parabola, as can also be seen in the right image on figure 3.9, where it can also be qualitatively seen that the fit is well representing the data. There are a couple of cases where the fit does not correctly represent the data. An example of this is seen in the left side of figure 3.9 where the object Class 0 object L1157 is shown. For L1157 the spline fit is dictated by a series of data points ranging from $J_{up} = 9$ to $J_{up} = 14$ which for this object has significantly smaller errors than the the rest of the data, which is why *L1157* appears

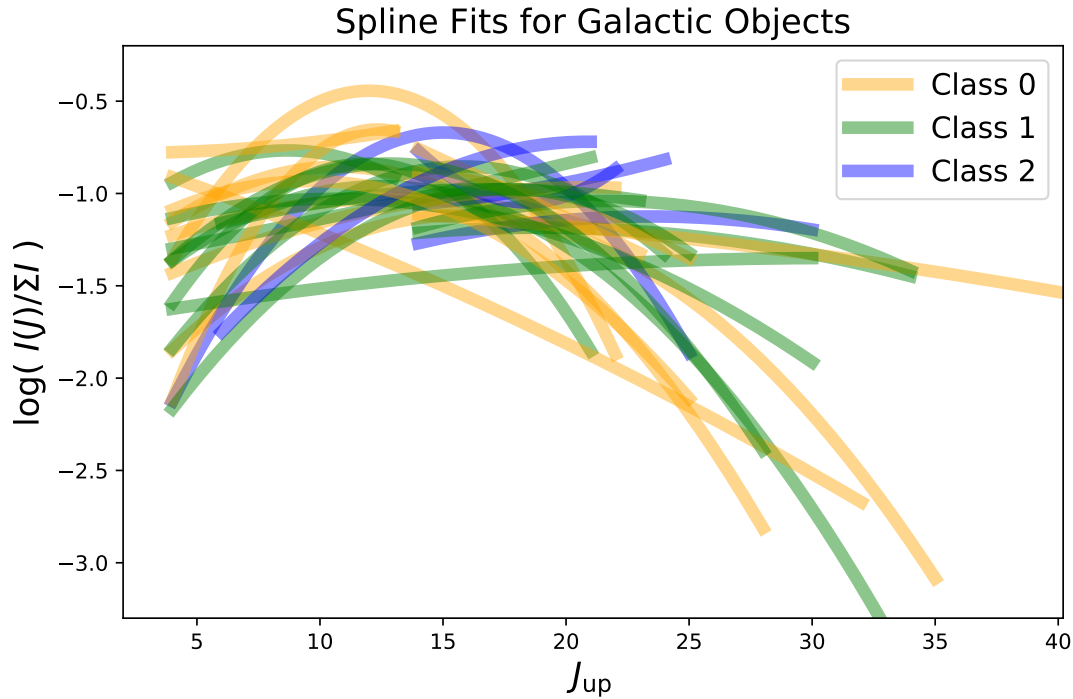


FIGURE 3.8: CO spectral line energy distributions with all galactic objects with at least 6 observed lines with $S/N > 5$. Each object fitted with weighted `scipy.univariate.spline k=2 and s=2`.

like the yellow straight line in figure 3.8, the green straight line in the same figure is caused by a similar overbias with small errors in the spline fit of the SLED of Class 1 object *TMR 1*.

The same type of spline with the same values was used for the extragalactic objects in figure 3.10 where the opacity in the plot corresponds to sum of the errors of the normalised intensity for all emission lines used in the spline fit for the object. Furthermore all the strongest emission lines (here expressed by upper level energy J_{up}) are found for each object, the upper energy level for each object of a given object type (Class 0,1,2 and extragalactic) is used to calculate the median and average value

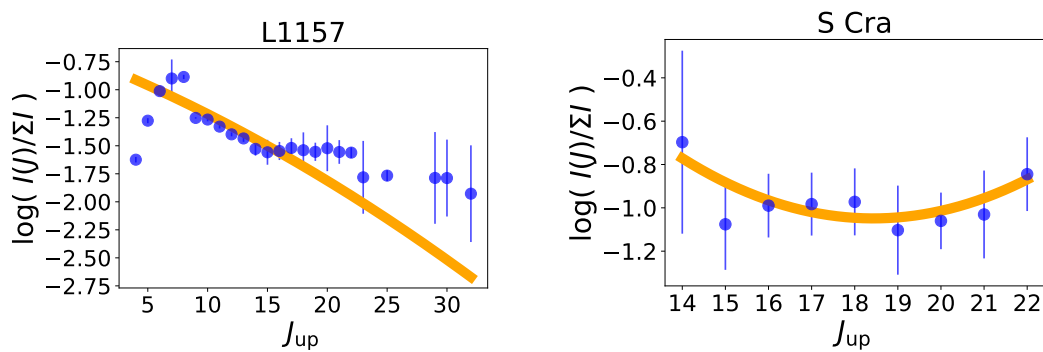


FIGURE 3.9: SLEDs of individual objects Class 0 object L1157 (LEFT) and Class 2 object S Cra (RIGHT). The blue points are representing the raw data with uncertainties multiplied by 5, while the orange shows the spline fit of the raw data.

of each object type. In table 3.1 the median, average and standard deviation values of the strongest emission lines also referred to as the peak transition is showcased both calculated from the raw data values from the database and from the spline fits. To ensure the numbers are comparable the raw data values calculations are also done with just objects which has at least 6 observed emission lines.

In table 3.2 each galactic object type has been compared and the P-values for each comparison can be seen, the calculations of the P-values has been done with the Python package "scipy.stats" and with the same data as for the calculations of medians and averages. The P-Value here compares two data sets and makes an estimate of how likely it is that the two data sets have origin in the same distribution. The P-Value is a number between 0 and 1, where 0 is no chance and 1 is 100 percent chance the two data sets have origin in the same distribution according to the statistical test.

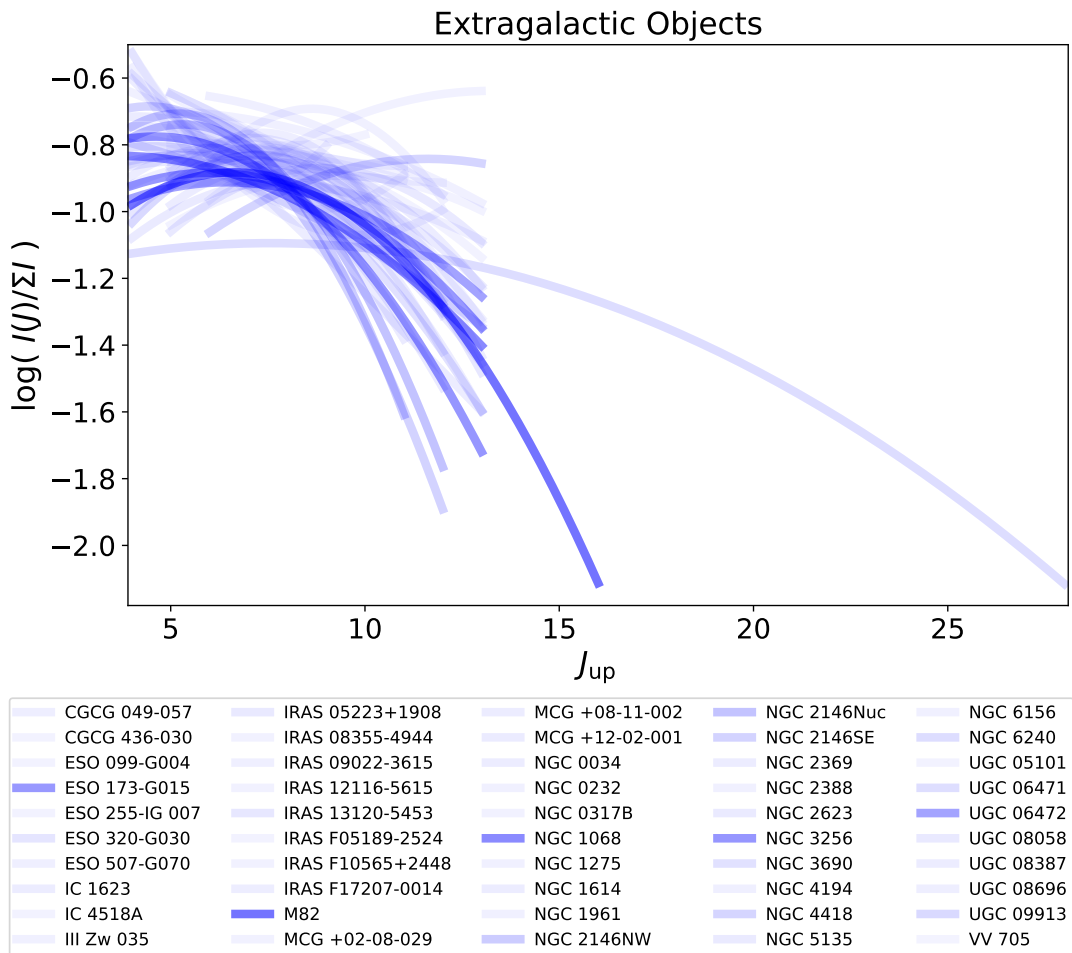


FIGURE 3.10: All lines with $S/N > 5$ fitted with a spline for extragalactic objects with at least 6 observed lines with $S/N > 5$. The opacity correspond to S/N higher opacity is higher S/N

All spline fit of the extragalactic SLEDs from figure 3.10 along with spline fits of the SLEDs from figure 3.8 are plotted together in figure 3.11 with the blue lines being from extragalactic sources, the green lines from galactic sources, including both

Class 0, Class 1 and Class 2 objects. The remaining lines are all stippled and are a set of statistical radiative transfer models for an estimated physical condition in the observed region with varying temperatures. The computer software for performing statistical equilibrium RADEX (van der Tak et al. (2007)) was used to make theoretical calculations of virtual data representing different physical parameters in the observed region that could be compared to the observational data in the database. The collisional rate coefficients are used from Yang et al. (2010) and Neufeld (2010) as tabulated in the Spring 2020 version of the LAMDA database (Schöier et al. (2005)). The number of collision partners are set to 1 which is H₂, which is set to have a density 10^8 which is high enough to ensure the conditions are thermally dominated and thereby the slopes of our SLEDs are dictated by the temperature. The CO column density in RADEX was set to 10^{12} cm⁻² to ensure the produced data appears as optically thin so information is not lost due to optical thickness. The line width is set to 1 km s⁻² and a total of 5 different temperatures are run through RADEX and displayed in figure 3.11, the temperatures tested are T=10 K (Brown stippled line), T=20 K (black stippled line), T=100 K (yellow stippled line), T=300 K (red stippled line) and T=500 K (magenta stippled line). The transition for which the emission reaches its peak for each temperature are highlighted by the arrows of the same color as the RADEX curves on figure 3.11 and found to be at $J_{\text{up}}=3$ for T=10 K, $J_{\text{up}}=4$ for T=20 K, $J_{\text{up}}=9$ for T=100 K, $J_{\text{up}}=16$ for T=300 K and $J_{\text{up}}=21$ for T=500 K.

3.4 Galaxy in a Box

In this section the results of different runs of Galaxy-in-a-Box will be presented. All results presented are with the input of observations of CO J= 10-9 from the database. The parameters are described in detail in the GiB section of the Method chapter (2.3). For the results referred to as standard, all the default parameters are used meaning standard IMF, SFE of 3 %, free fall time scaling factor of 1 (no scaling) and 10,000 stellar clusters. For the simulated galaxy image a beam size of 1 arcsec has been used and the modelled galaxy is at a distance of 12 Mpc.

Initial mass function

In figure 3.12 different IMFs has been run through the GiB model. The Standard, Top heavy and bottom heavy IMF. Corresponding to a slope of -1.3, -2.3 and -3.3 respectively. The Vertical lines in the figure corresponds to the median flux of one cluster from an IMF type corresponding to the shown color, standard IMF gives a median value of 1.14×10^{-5} K km s⁻¹, for a top heavy imf the median value is 1.47×10^{-5} K km s⁻¹ and for bottom heavy 1.07×10^{-5} K km s⁻¹. Corresponding to an increase in median flux of 29.0 % from standard to Top heavy IMF and a decrease of 6.2 % in the median flux from one cluster with a bottom heavy IMF. It can be seen that all three runs seems to have the same overall shape, and merely an offset from one another. The medians and standard deviations can also be read off of table 3.3

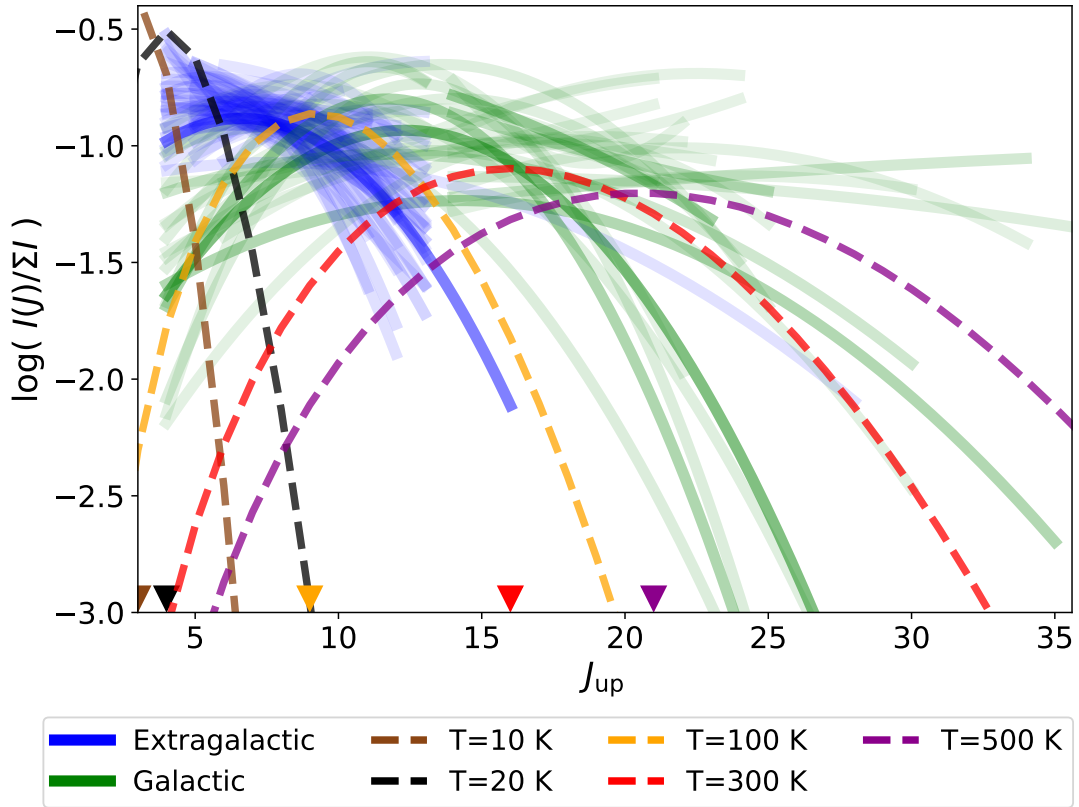


FIGURE 3.11: ^{12}CO lines with $S/N > 5$ plotted for extragalactic objects and galactic objects. The opacity correspond to S/N higher opacity is higher S/N . Theoretical models made from RADEX for gas variable temperature T . The coloured arrows in the bottom are pointing at the peak J_{up} for the RADEX model with the temperature corresponding to the color of the arrow.

along with the P-values of the distributions compared to the standard model with standard IMF. It can be seen from a combination of figure 3.12 and table 3.3 that a standard IMF and a bottom heavy one can be hard to distinguish with a P-value of 0.049 it cannot be definitively stated from these two model runs only if they are from the same distribution or not. Looking at the top heavy IMF in figure 3.12 it is qualitatively possible to distinguish the top heavy distribution from the other two, which is also quantitatively confirmed by the P-value below 0.01.

Star formation efficiency

In figure 3.13 two different SFEs has been run through the GiB model. The standard SFE= 0.03 (3%) and another with SFE = 0.3 (30%) and all other parameters as default like the standard model. The vertical dotted lines in the figure shows the median cluster emission of the model run with values corresponding to its color. The two distributions qualitatively seem to hold the same overall shape, however with a large offset, the run with SFE = 0.3 produces a median CO J=10-9 emission from a single cluster of $1.12 \times 10^{-4} \text{K km s}^{-1}$ which is 982 % higher than the standard model run. In table 3.3 the median, standard deviation and P-value with the distributions compared

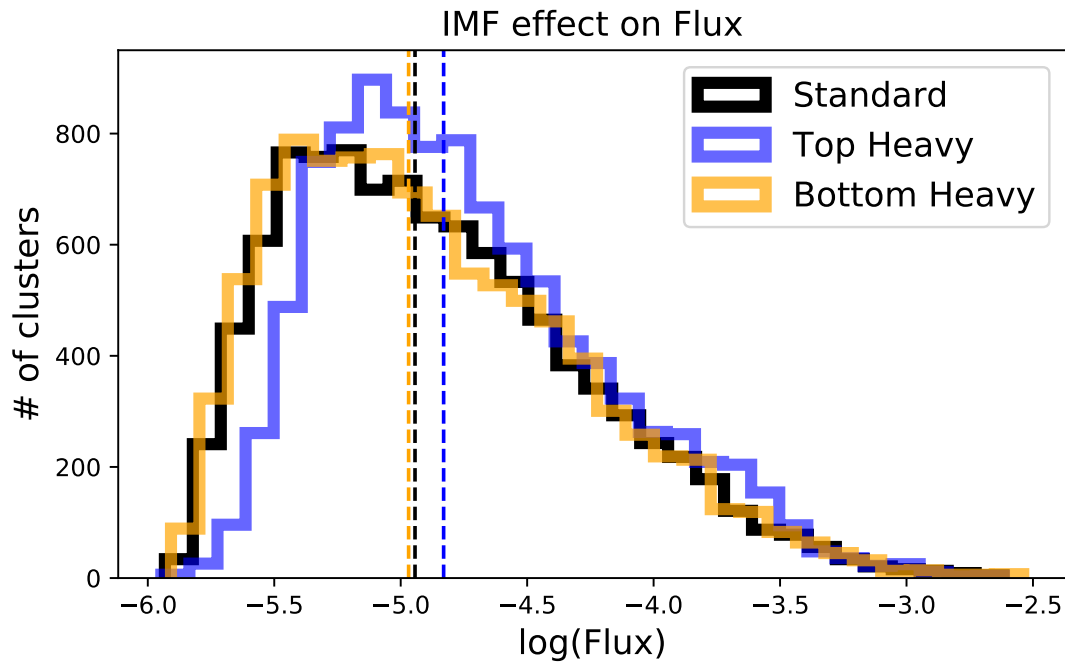


FIGURE 3.12: Galaxy-in-a-box cluster emission distribution with 3 different slopes of the IMF, standard=0, Top Heavy = 1, Bottom heavy=2. Observational data from the database CO J=10-9

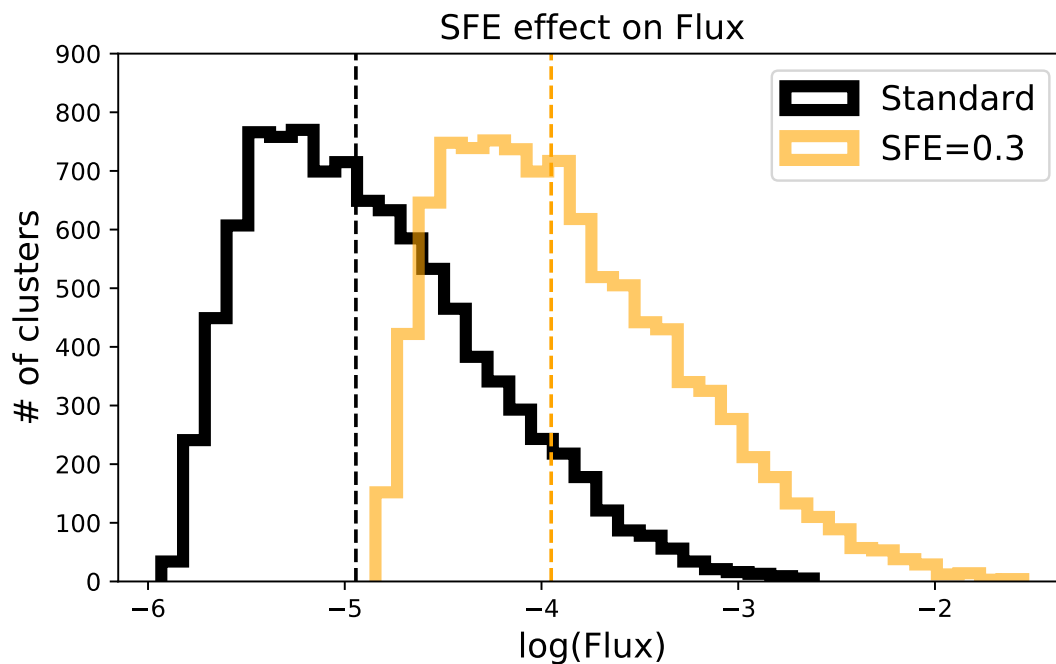


FIGURE 3.13: Galaxy-in-a-box cluster emission distribution default values (standard, SFE=0.03) compared to a SFE=0.3. Observational data from the database CO J=10-9

to the standard model run. It is seen from the P-value below 0.01 and from figure 3.13 that the two distributions are both qualitatively and quantitatively separable.

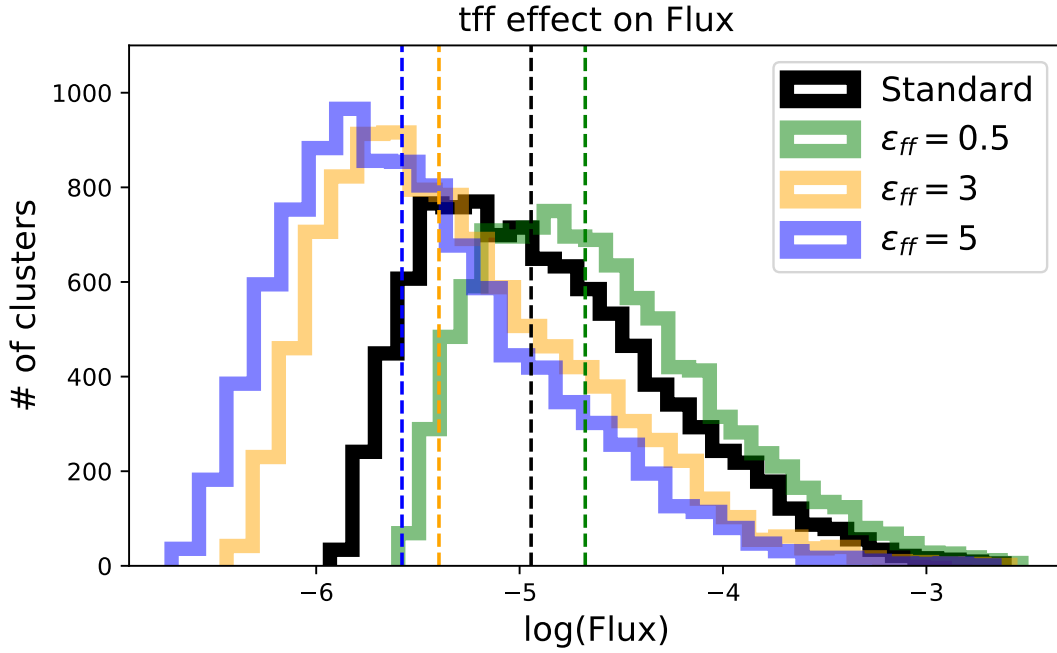


FIGURE 3.14: Galaxy-in-a-box cluster emission distribution with four different values of the free fall timescale parameter $tff=1$. Observational data from the database CO J=10-9

Free fall timescale

In figure 3.14 different values of the free fall scaling factor (tff) has been run through the GiB model. The standard $tff=1$ (no scaling), $tff=0.5$, $tff=3$, and $tff=5$, again with all other parameters as default like the standard model. The vertical dotted lines in the figure shows the median cluster emission of the model run with values corresponding to its color. The run with $tff = 0.5$ produces a median CO J=10-9 emission from a single cluster of $2.10 \times 10^{-5} \text{K km s}^{-1}$ which is 84.2 % higher than the standard model run, $tff=3$ has median cluster emission of $4.01 \times 10^{-6} \text{K km s}^{-1}$ which is 64.9 % lower than the standard run, $tff=5$ has median cluster emission of $2.64 \times 10^{-6} \text{K km s}^{-1}$ which is 76.9 % lower than the standard run. In table 3.3 the median, standard deviation and P-values from comparison to the standard model can be found. From the P-values all below 0.01 it is seen that the different tff values changes the distributions enough to be able to clearly distinguish the distributions from one another.

Total cluster emission

In figure 3.15 the total emission of the model galaxies simulated by the GiB model with different values of parameters described above are shown. The far left blue bar represents the model run with $SFE=0.3$ where the total galaxy emission in CO J=10-9 is 4.61K km s^{-1} but is multiplied by $1/2$ in the figure for illustrative purposes, and therefore looks half as bright as is. The yellow bar corresponds to $tff=0.5$ and the total galaxy emission in CO J=10-9 can be directly read of as 0.72K km s^{-1} , the green bar corresponds to $tff=3$ with total galaxy emission in CO J=10-9 can be directly read of as

TABLE 3.3: Overview of cluster flux distribution, GiB

Parameter	Median flux [10^{-5} K km s $^{-1}$]	Standard deviation [10^{-5} K km s $^{-1}$]	P-Value
Standard	1.14	11.99	-
Top heavy	1.47	11.64	< 0.01
Bottom heavy	1.07	13.00	0.049
SFE = 0.3	11.2	129.03	< 0.01
tff = 0.5	2.1	17.55	< 0.01
tff = 3	0.40	7.80	< 0.01
tff = 5	0.26	5.62	< 0.01

Notes: Median values, standard deviations and P-values as compared to the standard model from different cluster flux distributions of GiB model runs with a changed parameter as described in the first column and shown in figure 3.12, 3.13 and 3.14.

0.20 K km s $^{-1}$, the dark-magenta bar corresponds to tff=5 with total galaxy emission in CO J=10-9 can be directly read of as 0.13 K km s $^{-1}$, the turquoise bar corresponds to a top-heavy IMF with total galaxy emission in CO J=10-9 can be directly read of as 0.49 K km s $^{-1}$, the red bar corresponds to bottom-heavy IMF with total galaxy emission in CO J=10-9 can be directly read of as 0.43 K km s $^{-1}$, finally the black bar corresponds to standard GiB model with all parameters set to default and has total galaxy emission in CO J=10-9 can be directly read of as 0.43 K km s $^{-1}$, which is also illustrated with the black horizontal dotted line.

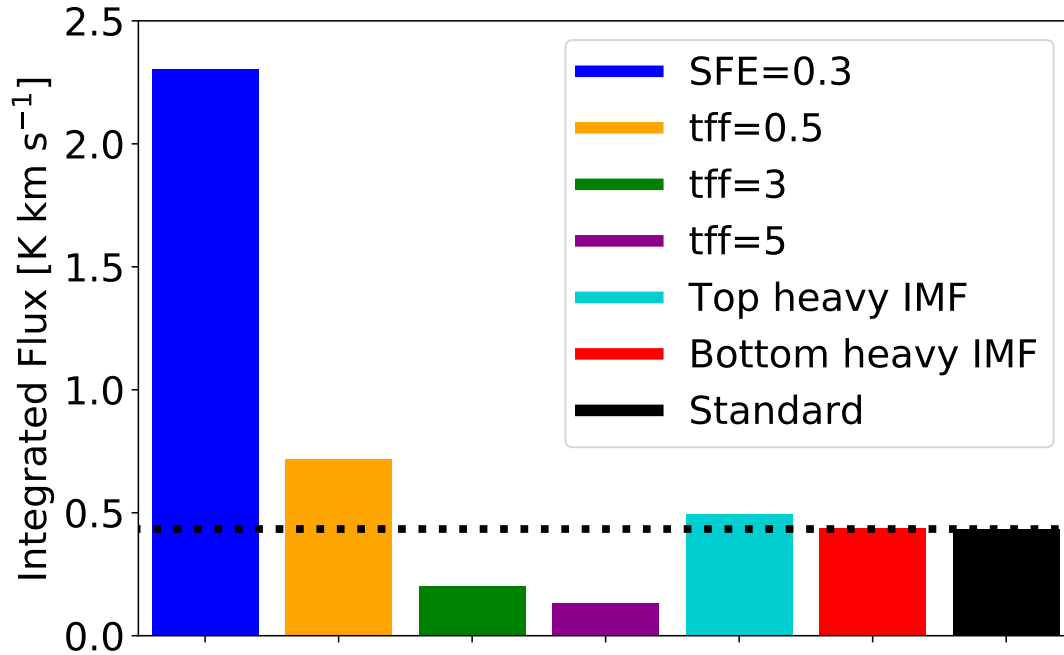


FIGURE 3.15: Plot of total emission from model galaxy from GiB model run with changes in individual parameters altered for each run. The blue bar representing SFE = 0.3 has been multiplied by 1/2 for illustration purposes.

Discussion

4.1 Physical origin of CO emission

The CO emission data collected to the database spans from $J_{\text{up}} = 4$ to $J_{\text{up}} = 48$ which spans a large array of excitation energies ($E_{\text{up}}/k_{\text{b}} \approx 55 - 6500\text{K}$). Trying to describe all of these emission lines from the same isolated region with specific and exact physical condition does not seem to be reasonable the CO ladder is therefore divided into three distinct components each linked to a specific set of physical conditions at the origin of the emission. The three components are discussed in separately in this section.

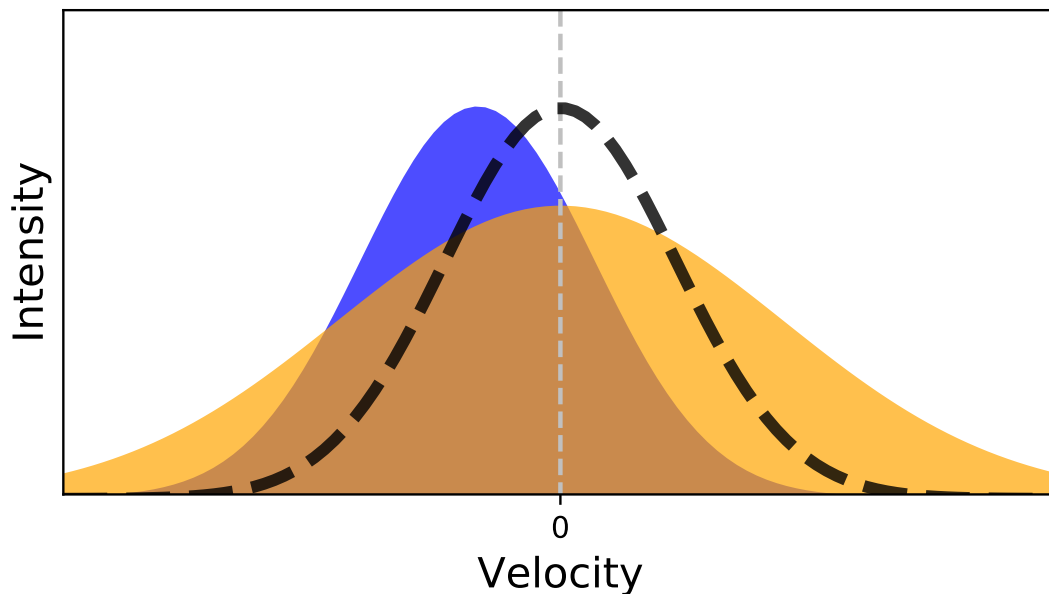


FIGURE 4.1: Not to scale; Illustration of different spectral line components; Black dashed line is a narrow centrally peaked component, Orange is a broad centrally peaked line, and the blue is a narrow offset peak.

4.1.1 The Cold Component

The cold CO gas regime is defined in two overall groups (Yang et al. 2018) a component $>100\text{ K}$ and one $\approx 100 - 200\text{ K}$ both found from fit done on rotational diagrams with multiple CO transitions. For this thesis the two regimes are regarded as a joined cold component due to the lack of ground based low- J ($J_{\text{up}} < 4$) observation data included in the provided data sample/database for this thesis. Low- J CO transitions

are generally linked more to the cold component of the gas than the high-J transitions due to lower excitation temperature. The low-J emission lines tend to have a narrow centrally peaked line profile like the one shown with the black dashed line in figure 4.1 when observed with HIFI facilitating the required spectral resolution for such analysis. Ranging from $J_{\text{up}}=1$ to $J_{\text{up}}=10$ the amount of the emission lines coming from the narrow component is gradually going from 100 % to 42 % (Yıldız et al. 2013) increasing with J_{up} . This indicates that a change happens around the transition $J = 10-9$ where the narrow centrally peaked component of the line is no longer responsible for a majority of the emission however even higher J transitions likely also overlaps to some extent. A narrow line profile centrally peaked indicates a slow moving gas at the emissions origin which in the case of star forming regions could be from cold entrained outflow (illustrated in figure 1.3) closest to the surrounding envelope or more likely from the quiescent envelope or the surrounding cloud (Yıldız et al. 2013, 2015).

San José-García et al. (2013) shows a broadening of the emission lines going from $J=3-2$ (Figure 4.1 black dashed line) to $J=10-9$ (Figure 4.1; Orange) consistent with an understanding that low J CO emission traces slower gas, and moving towards higher J corresponds to faster velocity of the gas at the origin. San José-García et al. (2013) also shows a correlation between the the narrow centrally peaked $J=3-2$ and the broad $J=10-9$ indicating a shared physical mechanism between the entrained outflow and the surrounding envelope material.

4.1.2 The Warm Component

The warm component of the CO emission is here defined as a rotational temperature found to be somewhere in the range between 300-400 K (Kristensen et al. 2017; Yang et al. 2018; Karska et al. 2013; Green et al. 2013). The ^{12}CO transitions suiting to a rotational temperature in the warm range is not perfectly fixed but observational data has with significance found the range to start outside the PACS regime ($J_{\text{up}} = 14$) (Green et al. (2013)). It is probable that it is not a hard cap situation where a specific transition is only linked to one isolated physical origin especially in the areas between the cold and warm component. As described in section 4.1.1 the shapes of the spectral lines has been found to consist of multiple components, this is also the case for $J_{\text{up}} > 10$ which leads to the likelihood of each line having components originating under different physical conditions. Towards higher J multiple studies finds the spectral line broadening to increase with higher J (San José-García et al. 2013; Kristensen et al. 2017) with the broad component (Figure 4.1; Orange) starting to be dominate from somewhere in the range of $J_{\text{up}} = 9$ to $J_{\text{up}} = 14$. This broad component (FWHM $\sim 20 \text{ km s}^{-1}$ Kristensen et al. (2017)) indicates a higher velocity of the molecules emitting the observed light. The higher velocity has been explained with a two different physical phenomenons both related to molecular outflows around the protostellar object. The first physical phenomenon which explains the high velocity component of the warm emission cavity shocks along the outflow cavity walls (illustrated in figure 1.3)

as described in [Mottram et al. \(2014\)](#) such cavity shocks would show up in the spectral line analysis as a broad line component with little or no offset like the orange graph in figure 4.1. Another physical phenomenon which has been proven capable of producing broad high velocity component in the warm molecular gas is disk winds (illustrated in figure 1.3) which has been described with promise in [Yvart et al. \(2016\)](#) for the water molecule however to my knowledge these results has yet to be reproduced for CO and is therefore left out of the further discussion in this thesis but is something to be taken into account if it turns out to be viable explanation.

4.1.3 The Hot Component

At higher CO transitions $J_{\text{up}} > 15$ a third component of the emission is found towards some objects. The third component is described as the hot component and is found to have origin in spot shocks inside the outflow cavity [Kristensen et al. \(2017\)](#), the spots shocks are illustrated in figure 1.3. The spot shocks are determined from the characteristic narrow (FWHM $\sim 5 \text{ km s}^{-1}$ ([Kristensen et al. 2017](#))) and typically with an offset of $>1 \text{ km s}^{-1}$, described as high velocity features of the spectral line, illustrated with the blue graph of figure 4.1. The narrow offset component is most frequently observed in the high J end of the CO ladder and linked to hot rotational temperature above 600 K.

4.1.4 The Case for Active Galactic Nuclei

Active galactic nuclei (AGN) is responsible for an enormous amount of radiative heating in galaxies in which they are present. This has the potential to excite CO molecules in their environment and thereby blur the picture of CO originating only in star forming regions. In [Lu et al. \(2017\)](#) the effects of AGNs on the CO ladder are studied with particular attention to the SPIRE end of the ladder ($3 < J_{\text{up}} < 14$). It is found that galaxies with AGN has a significantly different shape of SLEDs as compared to non-AGN galaxies when looking at high J ($J_{\text{up}} > 10$). It also concludes that the emission in the mid J transitions, $3 < J_{\text{up}} < 10$ is predominantly originating in relation to star formation and unaffected by the AGN of the observed galaxies.

4.2 Spectral Line Energy Distributions

4.2.1 Comparing the Object Categories

Class 0 vs Class 1

Knowing the possible physical origins of the different components of the CO emission allows for an in depth physical analysis an comparison of the different object categories, before being able to effectively compare galactic and extragalactic sources, making sure the galactic sample and categories behave as expected from theoretical and observational knowledge is required. Table 3.1 contains medians, averages

and standard deviations of the peak rotational transition for each galactic object type (Class 0, Class 1 and Class 2) as found from the SLEDs in figures 3.4, 3.5, 3.6 and 3.8 as well as from the extragalactic sources in figures 3.10 and 3.7 which will be discussed later in this section. Initially comparing Class 0 and Class 1 objects it is seen that the median is respectively 12.5 and 13.5 for the raw data sets, and the averages 12.2 ± 0.8 and 14.4 ± 0.9 . Neither the medians or the averages are outside of the standard deviations of either 4 for Class 0 and 6 for Class 1 and when looking at the P-Value of 1 in the statistical comparison between the two objects (Table 3.2) it is seen that it is hard to distinguish Class 0 and Class 1 objects just from the data SLEDs alone. An interesting trend does however appear in the comparison of the raw data between the two object types, Class 0 objects seem to peak at lower J than Class 1 objects which is also what is expected from the theoretical definitions of Class 0 and Class 1 objects. Class 0 objects is expected to be deeply embedded inside a large envelope where as the Class 1 objects are further in their development and therefore should have already accreted most of the envelope and therefore be embedded to a much smaller extent. Since the lower J transitions are associated with the molecular gas in the entrained outflow and in the quiescent envelope (Yildiz et al. 2015) it is expected that the deeply embedded Class 0 objects are emission lines in the low J regime are much brighter than it is for Class 1 objects. This trend is enhanced after spline fitting the raw data here an the medians are 14 for Class 0 and 19 for Class 1 while the averages are 14.2 ± 1.2 for Class 0 and 18.3 ± 1.5 for Class 1.

An interesting question arises here: why does the spline fit so dramatically push the peaks to a higher J? A large part of the explanation is that the shape of the entire SLED transition range influences the peak of the spline while for the raw data peak is entire driven by the single extreme points. Another part of the explanation is possible calibration differences between SPIRE and PACS. In figure 2.3 a downward intensity jump going from the SPIRE to the PACS regimen is observed which opens a possibility that SPIRE could generally be calibrated in a way that produces higher intensity than that of PACS. Since there is effectively no overlap between the two instruments it is hard to determine if this is a feature of the actual observation or a calibration difference however it is something seen in multiple objects most pronounced in objects such as B335 (figure 3.4) and WL 12 (Figure 3.5). The peak of the spline fit might be moved into the higher J PACS regime even though the peak data point in the raw data is in the SPIRE region if this makes the overall fit better. The spline fit leaves a more overall shape of the SLED with less attention to the details. From a qualitatively look at figure 3.8 where the spline fitted SLEDs are plotted it seem hard to tell the difference between Class 0 and Class 1 objects with the possible exception that the Class 0 objects seems to peak at lower J.

The statistical test comparison between the spline fitted Class 0 and Class 1 objects estimates a P-value meaning the chance of Class 0 and Class 1 objects originating from the same distribution to be 0.04 or 4 % which indicates a good chance that the statistical model can distinguish the to categories from one another after smoothing

out the data with a spline fit.

Class 2

A total of 38 Class 2 objects are included in the MySQL database constructed during the work with this thesis, of the 38 included Class 2 objects only 10 show up as an SLED in figure 3.6 where only objects with $S/N > 5$ for at least two transitions were included. An easy conclusion to make from this is that the signal to noise must be smaller from the Class 2 objects than from the Class 0/1 objects. The noise is not significantly different from the Class 0/1 objects however the signal from the Class 2 sources were generally speaking much lower. From a physical standpoint the weaker emission can easily be explained by Class 2 objects being further in their development and therefore having less envelope material left and therefore also less material to eject through outflows.

This physical understanding of Class 2 objects is backed up by our data sample, in table 3.1 the error on the average values for Class 2 objects both in the raw and spline fit data are 4 which is very large compared to the other object categories and a direct consequence of the low statistics due to a lower signal in the CO transitions from Class 2 objects. The P-values from statistical tests for Class 2 objects presented in table 3.2 shows from the raw data an 8 % chance that the Class 2 and Class 0 objects could come from the same distribution and a 47 % chance of Class 1 and Class 2 originating from the same distribution. For the spline fit data it gives a 22 % chance of Class 2 and Class 0 and an 88 % chance of Class 2 and Class 0 originating from the same distribution. All these values are very inconclusive and due to low statistics hardly adds any information other than Class 2 objects as expected have more in common with Class 1 rather than Class 0 objects. The conclusion from the analysis of the CO ladder of Class 2 objects must be that unresolved CO emission is a poor tracer of the physical environment related to Class 2 objects.

4.2.2 Extragalactic Objects

This section has so far compared apples with apples alike when comparing YSO in different stages of their evolution. The observations are made with the same method; the telescope pointed in the direction of the YSO and then the central spacial pixel (Figure 2.2) containing only a small aperture around the object (Green et al. 2016) isolated to capture as much of the protostellar object as possible while limiting the amount of signal originating from outside the region of the targeted object. The extragalactic objects also collected in the database are a different breed in this context, the observations are no longer isolated to a small region around the active star formation like with the galactic objects. The galactic objects are still unresolved and observed like a point source but now rather than having one YSO we have an entire galaxy contained inside each observation.

Making a naive statistical comparison between the observed extragalactic objects and the different categories of galactic objects the P-values of table 3.2 shows less than 0.01 for both raw and spline fit data no potential chance that the extragalactic data originating in the same distribution as the galactic objects with the single exception of the spline fits of the extragalactic objects compared with Class 2 objects which gives a P-value of 0.05 corresponding to a 5 % chance the objects could come from the same distribution, once again this effect is likely associated with the low statistics of the Class 2 objects and directly contradicts with Class 2 objects peaking at a transition line an average 14-15 J higher than the extragalactic objects (3.1) which is the largest gap between any two object categories in this sample.

The extragalactic objects are quantitatively different from the galactic objects from the investigated sample with a very low both median value J_{peak} of 6 for both raw data and spline fit and just around the same average value of 6.06 ± 0.06 and 6.2 ± 0.2 for raw data and spline data respectively, both of these with a low uncertainty of the average value due to a high number of observed extragalactic objects. With a high certainty it can therefore be concluded that extragalactic objects peak at significantly lower J than isolated Class 0 and Class 1 objects which makes a lot of sense from a physical point of view.

When looking at the physical origin of the different components described in section 4.1 the cold quiescent envelope and surrounding cloud are dominating the contribution in the lower J transition range around $J=6-5$ (Yıldız et al. 2015) and lower which is also around the peak emission of the extragalactic objects. This could possibly indicate a majority of the observed CO emission from the extragalactic objects are tracing gas originating in cold molecular clouds which can be explained by a significant higher amount of cold molecular cloud in a Galaxy in comparison to the amount of molecular outflows from star formation. When this same trend is not observed in the sample of galactic objects it is unsurprisingly because of the data selection deliberately choosing the least amount of surrounding cloud to maximize the signal from the YSO itself.

4.2.3 Temperature Components in the Data Sample

To translate the overall shapes of the SLEDs into temperature components the radiative transfer model RADEX was used to plot theoretical expectation for different temperature components on top of the spline fit data in figure 3.11. A qualitative look at the figure indicates that multiple components are required to most accurately describe the SLEDs of both the galactic and extragalactic data. In the galactic data (marked green) a small fraction of the emission seems to follow a component corresponding to cold surrounding gas or quiescent envelope with temperature around 10-20 K (marked brown and black) which is also backed by the significantly higher J_{peak} of the observational data than from the RADEX models for $T=10$ K and $T=20$ K. Moving to slightly higher temperature the component $T=100$ (marked orange) K is marking cold gas likely to originate in the envelope of YSO and in the entrained

outflow of such objects. Some galactic objects show similarity to the $T = 100$ K model in the transition range $J_{\text{up}} < 10$ where the $T=100$ K model has its peak emission, however no galactic objects seem to have as steep of a descent in relative intensity when moving towards higher J from $J_{\text{up}} = 10$ as the $T = 100$ K model which tells us that for $J > 10$ we need a higher temperature component to explain the behaviour of the galactic SLEDs. When moving to even higher temperature $T = 300$ K (marked red) has a peak intensity at $J = 16-15$ known to be strongly related to the cavity shocks in the inner part of outflows as concluded from a broad spectral line profile for the $J = 16-15$ line in [Kristensen et al. \(2017\)](#). A temperature close to 300 K seems to be responsible for the strongest emission from the galactic objects which seems to be the highest in the range between $J_{\text{up}} = 14$ and $J_{\text{up}} = 19$.

The 300 K temperature range does seem to be responsible for the highest intensity component for most of the galactic objects however it does not seem to be the highest temperature component present, at least not for all galactic objects in sample. A 500 K RADEX model has also been run and is illustrated with the magenta dashes line in [figure 3.11](#). Comparing this 500 K line to the galactic objects it can be seen that some objects does not see a decline in intensity as steep as would be the case if a 500 K component were to be the hottest. This indicates that some of the sampled objects has a hot component originating from hot high velocity spot shocks as described in [4.1.3](#).

The CO SLEDs of extragalactic objects also tells a story about the active star formation of the observed region, in the cases of these extragalactic objects an entire galaxy. Comparing te extragalactic objects (In [figure 3.11](#) marked with blue) SLEDs with the theoretical models from RADEX it is reasonable to state the extragalactic objects tend to have a stronger resemblance to the 3 lower temperature ($T=10$ K, $T=20$ K and $T=100$ K) RADEX models and like the galactic objects seem to consist of more than one physical component. In the $J_{\text{up}} < 7$ most objects seem to have a high amount of relative emission in each transition corresponding to a cold temperature < 100 K corresponding to either cold envelope, surrounding molecular cloud or cold entrained outflow. Surrounding molecular cloud should be relatively much more abundant within an aperture spanning an entire galaxy than that of an aperture of a local YSO simply because the surveys used for this sample is specifically targeting outflows and active star formation like the case for the surveys it is targeted much more specifically and close by the protostellar object and therefore much of the surrounding molecular cloud is excluded already in the observing process while for the entire galaxies all the molecular clouds are included in the observations and in this data sample. This explains the higher relative intensity of low J transition of extragalactic objects in comparison to galactic YSOs and also decreases the effectiveness of low J CO emission as an accurate tracer for active star formation in unresolved galaxies.

When turning the eyes towards the $J_{\text{up}} > 7$ regime of the extragalactic SLEDs it is seen that most objects are either increasing or gently decrease suggesting a temperature component around 100 K (yellow in [figure 3.11](#)) or above. A temperature of around 100 K is likely to originate in outflows from protostellar objects and therefore

has a much better chance of accurately tracing star formation in unresolved galaxies.

4.3 Galaxy-in-a-Box

The Galaxy-in-a-Box model is set into the world to try and accurately predict the amount of star formation in a galaxy from the molecular emission. For this particular thesis I have collected high J CO emission data with the same aim of tracing star formation in distant galaxies. After collecting and analysing the CO ladder and SLEDs a choice of specific transition (s) to use for the GiB model had to be made. As discussed in the previous section the low J CO emission $J < 7$ or so are not ideal tracers of active star formation since cold molecular clouds are not necessarily involved in active star formation. Optical thickness is another thing to be taken into account to a much higher extent when trying to map entire galaxies, [Yang et al. \(2018\)](#) finds the optical thickness of CO transitions to be related to upper level energy as:

$$\log(\tau) = -(2.6 \pm 0.6) \times 10^{-3} \times E_u + (1.4 \pm 0.1) \quad (4.1)$$

Where E_{up} is the upper level energy of the transition. They conclude that the optical thickness approaches 1 when E_{up} reaches 522.3 K which in corresponds to the $J_{\text{up}} = 13$ transition, the highest transition observable with SPIRE which has made the majority of the extragalactic observations in the collected sample.

Star Formation Tracer Checklist

The general analysis of the CO ladder suggest that the most ideal line for tracing active star formation should be

- (i) High enough to only trace warm gas related to outflow disk winds or cavity shock to avoid including cold molecular clouds not currently forming stars.
- (ii) Not too high J so that too shallow populations of the upper level species makes emission too weak to easily observe, since not all objects is proven to have hot components related to spot shocks. Transitions around $J=16-15$ seems to be the optimal here since a majority of objects show decreased intensity at higher J.
- (iii) The transition should have the lowest possible optical thickness. Everything above $J=13-12$ is assumed to be optically thin, and below $J=13-12$ the optical thickness is shown to decrease exponentially with upper level energy, E_{up} .
- (iv) The transition should be well studied and observed towards a sufficient amount of protostellar objects to provide good statistics.

From the reasons outlined above the line $^{12}\text{CO } J=10-9$ was chosen to run initial testing of the compatibility of high J CO and GiB. The $J=10-9$ is the best fit for an ideal star formation tracer at this point in time. As outlined in figure 3.2 it is the

most observed line in this sample satisfying checklist point (iv), it is proven to be in upper level energy range to trace warm outflow emission and generally speaking a bright CO line satisfying (i) and (ii) however not the the most ideal line according to these demands since it is shown to have a narrow spectral line component up to 42 % from a quiescent gas (Yıldız et al. 2013). Another complication not mentioned in the checklist is the influence of AGNs, the GiB model does not yet take into account AGNs when modelling a galaxy. Future iterations of the model will possibly be able to take this into account which will allow for higher J transitions to be used to model galaxies with AGNs but until then caution should be taken when choosing CO lines to use in the model. As mentioned in section 4.1.4 the lines of CO with $J_{\text{up}} < 10$ is found to be unaffected by AGNs (Lu et al. 2017) whereas $J_{\text{up}} > 10$ emission is altered from galaxies hosting an AGN. This means that for the current iteration of GiB the upper limit for J_{up} is 10 unless the modelled galaxy has no AGN or is possible to spatially resolve. With J=10-9 being the highest J transition possible without having to address AGNs it is also the possible line with the lowest expected optical thickness (equation 4.1; Yang et al. (2018)). Although J=10-9 is the transition of choice in this thesis all the observed lines included in the database could as easily be utilized in the model and for future work it could be interesting to look at J_{up} higher than 10 to get a lower optical thickness and thereby trace an even higher relative amount of deeply embedded emission.

4.3.1 Significance of results

In order to evaluate the results of GiB model with the chosen ^{12}CO J=10-9, a standard model was defined and different physical parameters were individually varied to isolate their impacts on the results as much as possible. An overview of the most important results from the initial runs of GiB is highlighted in table 3.3 and figure 3.15 and will be discussed in this section.

Initial Mass Function

The effect of the three different IMFs are tested in the work of this thesis a standard, a bottom heavy and a top heavy IMF. The distribution of these three are shown in figure 2.5. Common for the tested IMFs are the low mass end ($< 1 M_{\odot}$) which is not very well studied and therefore left constant like defined in Chabrier (2003). The other end of the IMF ($> M_{\odot}$) is defined from equation 2.4 as

$$\xi = A_2 \times (10^m)^{-1.3-\beta}, \quad m \geq 1 M_{\odot}$$

Where the only thing changed in the model is β which controls the IMF type, for the standard $\beta = 0$, for top heavy $\beta = 1$ and bottom heavy $\beta = 2$. The flux distributions of the 10,000 stellar clusters simulated from each run of GiB are shown in figure 3.12 for all three IMF types. From the figure it is very hard to visibly tell a difference between the standard and the bottom heavy IMF whereas the top heavy is visibly standing

TABLE 4.1: Overview of mass in low and high mass regime

IMF type	Mass of objects $< 1M_{\odot}$ [% of total mass]	Mass of objects $> 1M_{\odot}$ [% of total mass]
Standard	60.4	39.6
Top heavy	33.2	67.8
Bottom heavy	73.1	26.8

Notes: Percentages of mass distributed in high mass and low mass regime for different IMFs. The numbers are calculated from the CiB model from distributions similar to those in figure 2.5.

out. The medians shown as vertical dashed lines in the figure are also written out in table 3.3 together with standard deviations and P-values for comparing the different flux distributions for GiB simulated stellar clusters. From the median values a difference of 6.2 percent between standard and bottom heavy IMF, in contrast to a 29 percent difference in medians of standard and top heavy IMF is adds to the trend of the bottom heavy and standard IMF being more similar to each other than to the top heavy. A look at statistical tests further adds to this trend as the top heavy and standard IMF has a P-value less than 0.01 corresponding to a less than 1 percent chance the two IMF types could originate in the same distribution or in other words they are distinguishable through statistical test and therefore has to be taken into account in GiB when trying to model a galaxy. On the other hand comparing bottom heavy with standard IMF gives a P-value of 0.05 corresponding to a 1 in 20 chance the two GiB results could originate in the same distribution. 1 in 20 is a chance to high to say anything definitively so a natural question arises, Why does the IMF being top heavy change the flux of clusters measurably more than a bottom heavy?

The answer is likely to be found in table 4.1 where the percentages of mass originating in objects less and more massive than $1 M_{\odot}$ are shown. From this it can be seen that for both the standard and bottom heavy IMF results in the clusters being dominated by the mass distribution of the $< 1 M_{\odot}$ end of the mass distribution as opposed to the top heavy IMF which is heavily dominated by high mass stars $> 1 M_{\odot}$. This is important since the only end of the distribution change in the model (both CiB and GiB) is the most well studied $> 1 M_{\odot}$ end of the IMF only resulting in a minority change in both the bottom heavy and standard IMF since the majority of the mass is located in the less studied low mass end of the IMF which is held constant in the model. Ultimately this explains the shown tendency that the GiB model cannot distinguish a bottom heavy IMF from the standard one, but could potentially tell the difference between a top heavy and standard/bottom heavy IMF.

Star Formation Efficiency

SFE is the parameter of the CiB/GiB model which determines how much of the total mass of a cluster will form stars. Since the emission of interest to this thesis is linked

to active star formation the SFE is expected to be an extremely important parameter in the model. In essence the model is developed with the purpose of estimating the active star formation taking place in a given galaxy and without a parameter for the SFE the model estimates molecular gas rather than active star formation.

The result of the initial tests of GiB with CO J=10-9 as the star formation tracer of choice it is clearly seen from figure 3.13 that the distribution of flux from cluster is shifted to be approximately an order of magnitude larger while taking the same shape when going from a SFE of 0.03 to 0.3. The same trend is backed by the quantitative numbers drawn from the figures presented in table 3.3 which shows a median flux from a cluster going from 1.14×10^{-5} K km s⁻¹ when SFE=0.03 to almost an order of magnitude higher when increasing SFE one order of magnitude to 0.3 where the median flux from a cluster becomes 11.2×10^{-5} K km s⁻¹. This seems to be a direct correlation between SFE and median flux which also makes physical sense since a higher SFE means a higher percentage of the mass actively taking part in star formation and therefore should be traced by the CO J=10-9 emission line. Likewise the P-Values of the SFE=0.3 as compared to the standard model is less than 0.01 which further acts as an evidens to suggest that the model can distinguish clusters and galaxies with different SFEs.

Free Fall Timescale

The free fall time scaling factor ϵ_{ff} , representing the core collapse timescale in units of free fall timescales, has also been isolated and varied to tests its effect on the results of GiB. The test runs shows a clear tendency of the median flux of a cluster and in extension the total flux of a galaxy to increase with lower values of ϵ_{ff} as seen by a median cluster emission of 1.14×10^{-5} K km s⁻¹ for the standard $\epsilon_{ff} = 1$, as compared to 0.26×10^{-5} K km s⁻¹ for $\epsilon_{ff} = 5$ and 2.1×10^{-5} K km s⁻¹ for $\epsilon_{ff} = 0.5$. From a physical point of view this makes perfect sense, if the timescale over which the percentage of the total mass of a cluster effectively is converted to stars given by the SFE are longer as with $\epsilon_{ff} = 5$ the amount of star formation taking place at the time of observation should be smaller since less stars has had the time to form in the lifetime of the cluster.

Without the proper caution one could assume the ϵ_{ff} parameter should scale linearly with the total emisison of a cluster however this is not seen to be the case. The reason for this is to be found in the age of the clusters. The clusters will throughout their lifetime form the total mass of the cluster times the SFE as defined in equation 2.5 which means the SFE scales directly and linearly with the emission while the timescale factor ϵ_{ff} does not necessarily because of the dependency on time. Think of a totally collapsed cluster with SFE = 0.03, this cluster will have converted 3 % of its initial mass into star regardless of the timescale over which is collapsed which is why the ϵ_{ff} parameter in some cases loses its influence when the SFE is fixed in the model.

4.4 Outlook

Rotational Transitions of CO with $J_{\text{up}} > 10$ has shown to be a useful tracer of star formation in extragalactic objects since it traces gas warm enough to not be present in significant amounts the surrounding envelope and molecular cloud. The theoretical model implementing correlations from observations of exactly CO 10-9 Galaxy-in-a-Box, has through the tests done in this thesis proven to be able to distinguish different SFEs of a galaxy from each other as well as different star forming timescales, ϵ_{ff} . The model has shown to be less sensitive to changes in the IMF which keep the determining parameters of the models results the SFE and the closely related core collapse timescale. The GiB model can therefore be a useful tool for estimating the amount of active star formation in galaxies, even when resolving the galaxies individual stellar cluster or star forming regions are not possible.

Since the wavelength $260.24 \mu\text{m}$ emitted from the CO 10-9 transition is not observable from ground, because the atmosphere is completely opaque for this and the wavelength corresponding to all higher J transitions and no current space based telescope can observe these wavelengths either, the only option to use the methodology described in this thesis is to go to high redshift. When CO 10-9 is observed at a sufficiently high redshift it is observable from ground, this has been proven by observations done by [ALMA Partnership et al. \(2015\)](#) of CO 10-9 in a galaxy at redshift $z=3.042$ with Atacama Large Milimeter/submillimeter Array (ALMA) and [Yang et al. \(2017\)](#) observing the same galaxy plus an additional three galaxies in CO 10-9 and 11-10 at redshifts between $z=3.1271$ and $z=3.5650$ with the Institute for Radio Astronomy in the Millimeter Range (IRAM) 30 meter telescope. This means that the observations can be made and the GiB can be used to estimate the active star formation in the at the high redshift range where the star formation was at its highest according to ([Bouwens et al. 2015](#)) as shown in figure 1.1, and GiB will be able to estimate important star formation parameters such as SFE of the galaxies.

Naively translating the GiB model build on observation of galactic protostellar objects to high redshift galaxies for which good angular resolution is difficult to achieve might be a over simplification so before inferring star formation in high z galaxies based on GiB it is a good idea to test it on closer galaxies. This means the next steps for the model is to test it up against real world galaxies. The database build for this thesis does contain > 100 galaxies observed in the SPIRE range of the CO ladder which includes CO 10-9, these galaxies would be a great initial test set for GiB, but unfortunately fell outside of the possible scope of this thesis.

Summary and Conclusions

Emission data observed by the Herschel Space Observatory of ^{12}CO rotational transitions ranging from $J=4-3$ to $J=48-47$ emission have been collected for 190 objects (87 galactic protostellar objects and 103 extragalactic objects). The collected data was organized in a MySQL database and analysed for characteristics to reveal embedded star formation. The main conclusions are as follows:

- From Spline fitting of the CO ladder SLEDs of 16 Class 0 objects are found to have a median peak intensity at $J=14-13$ and average peak intensity of $J_{\text{up}} = 14.2 \pm 1.2$. From 17 Class 1 objects the median peak intensity is found to be at $J=19-18$, average being $J_{\text{up}} = 18.3 \pm 1.5$.
- From observations of 38 Class 2 objects, only 10 were found to have quality ($S/N > 5$) observations of CO, and no significant features were found to be common ranking high- J CO as a poor tracer of the physical environment around Class 2 objects.
- It is indicated from comparisons of observed SLEDs and RADEX models that high J CO are tracing multiple star forming components of Class 0 and Class 1 Objects. A cold component with typical temperatures < 100 K linked to quiescent envelope or entrained outflow. A warm component with temperature ≈ 300 K with origin in disk winds or cavity shocks. The third component being a hot component $T_{\text{rot}} > 600$ K and related to spot shocks inside the cavity walls of the outflow only observed towards a selection of objects.
- From observation of 103 extragalactic objects the 50 objects with highest quality (S/N) were chosen and spline fits were applied to the SLEDs to reveal a median peak intensity of the emission from the CO ladder transitions to be $J=6-5$, and the average 6.2 ± 0.2 . This indicates much more cold molecular cloud are traced in extragalactic objects when compared to galactic protostellar objects at all tested evolutionary stages.
- From analysis of the CO ladder SLEDs and comparison to studies of rotational diagrams the ideal CO transition for tracing star formation has been found to fulfill: (i) Trace warm outflow gas to avoid surrounding envelope/molecular cloud. (ii) Be a strong line, at sufficiently high J the transitions become less abundant. (iii) Have lowest possible optical thickness, everything in the PACS

range ($> J=13-12$) assumed optically thin. (iv) The line is well studied and observed towards many objects.

- A strong ($r=0.99$) powerlaw correlation $L_{\text{CO}} = 10^{-3.89 \pm 0.13} \times L_{\text{bol}}^{0.94 \pm 0.07}$ is found between line luminosity of the CO 10-9 transition and the bolometric luminosity of YSOs across 5 orders of magnitude, consistent with a direct proportionality $L_{\text{CO}} \propto L_{\text{bol}}$, indicating a possible use of CO 10-9 as a tracer of dense molecular gas.
- The Galaxy-in-a-Box model has been optimized to decrease the time required to run the model. The model originally ran in ~ 85 minutes and was cut down to ~ 17 minutes, both with all parameters set to default, a relative performance improvement of 80 %.
- The created MySQL database has successfully been implemented with the model Galaxy-in-a-Box for the CO 10-9 rotational transition. The database can easily supply the observational data input for Galaxy-in-a-box allowing the model to simulate emission of all transitions contained in the database.
- The Galaxy-in-a-box model has through initial tests proven to be able to differ between a top-heavy IMF from a standard or bottom-heavy IMF, but not differ variations to the non dominating end of the IMF, as is the case for bottom-heavy and standard IMF both dominated by low-mass stars.
- The model Galaxy-in-a-Box has been tested for the CO 10-9 line where it has proven to effectively differ variations in the total SFE and the SFE per free fall timescale proxied by the parameter ϵ_{ff} in both resolved and unresolved galaxies.

The next step in the pursuit to investigate star formation through cosmological timescales at redshifts where star formation rate is shown to be an order of magnitude larger than observed in our local universe. From studying SLEDs of galactic protostars high-J CO are shown to accurately trace physical conditions unique to star formation. The same SLEDs has shown to trace the same unique star formation features in extragalactic objects for which the model Galaxy-in-a-box shows promise to be able restrain Star formation efficiency and core collapse timescale of the stellar clusters in the galaxies. The ability to constrain these parameters in extragalactic objects all the way to high redshifts would increase our knowledge of star formation on cosmological timescales. The model requires further testing before being able to go to cosmological timescales, a natural stepping stone would be to test the model on the 103 extragalactic sources in the MySQL database created for this thesis. If the model is proven to accurately estimate emission and star formation parameters in these objects it can be a powerful tool in combination with high-z observations from ground based telescopes like ALMA.

Bibliography

- Adams, F. C., Fatuzzo, M., & Holden, L. 2014, *The Astrophysical Journal*, 789, 789
- ALMA Partnership, Vlahakis, C., Hunter, T. R., et al. 2015, *The Astrophysical Journal Letters*, 808, 808
- Bontemps, S., Andre, P., Terebey, S., & Cabrit, S. 1996, *Astronomy and Astrophysics*, 311, 311
- Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2015, *The Astrophysical Journal*, 803, 803
- Casasola, V., Cassarà, L. P., Bianchi, S., et al. 2017, *Astronomy and Astrophysics*, 605, 605
- Chabrier, G. 2003, *Publications of the Astronomical Society of the Pacific*, 115, 115
- de Graauw, T., Helmich, F. P., Phillips, T. G., et al. 2010, *Astronomy and Astrophysics*, 518, 518
- Dionatos, O. 2015, 102, 102
- Endres, C. P., Schlemmer, S., Schilke, P., Stutzki, J., & Müller, H. S. P. 2016, *Journal of Molecular Spectroscopy*, 327, 327
- Green, J. D., Evans, II, N. J., Jørgensen, J. K., et al. 2013, *The Astrophysical Journal*, 770, 770
- Green, J. D., Yang, Y.-L., Evans, II, N. J., et al. 2016, *The Astronomical Journal*, 151, 151
- Griffin, M. J., Abergel, A., Abreu, A., et al. 2010, *Astronomy and Astrophysics*, 518, 518
- Karska, A., Herczeg, G. J., van Dishoeck, E. F., et al. 2013, *Astronomy and Astrophysics*, 552, 552
- Karska, A., Herpin, F., Bruderer, S., et al. 2014, *Astronomy and Astrophysics*, 562, 562
- Kristensen, L. E. & Bergin, E. A. 2015, *The Astrophysical Journal Letters*, 807, 807
- Kristensen, L. E. & Dunham, M. M. 2018, *Astronomy and Astrophysics*, 618, 618

- Kristensen, L. E., van Dishoeck, E. F., Mottram, J. C., et al. 2017, *Astronomy and Astrophysics*, 605, 605
- Krumholz, M. R. 2015, arXiv e-prints, 1511, 1511
- Lada, C. J. & Lada, E. A. 2003, *Annual Review of Astronomy and Astrophysics*, 41, 41
- Lu, N., Zhao, Y., Díaz-Santos, T., et al. 2017, *The Astrophysical Journal Supplement Series*, 230, 230
- Mashian, N., Sturm, E., Sternberg, A., et al. 2015, *The Astrophysical Journal*, 802, 802
- Mottram, J. C., Kristensen, L. E., van Dishoeck, E. F., et al. 2014, *Astronomy and Astrophysics*, 572, 572
- Mottram, J. C., van Dishoeck, E. F., Kristensen, L. E., et al. 2017, *Astronomy and Astrophysics*, 600, 600
- Neufeld, D. A. 2010, *The Astrophysical Journal*, 708, 708
- Pilbratt, G. L., Riedinger, J. R., Passvogel, T., et al. 2010, *Astronomy and Astrophysics*, 518, 518
- Poglitsch, A., Waelkens, C., Geis, N., et al. 2010, *Astronomy & Astrophysics*, 518, 518
- Ringermacher, H. I. & Mead, L. R. 2009, *Monthly Notices of the Royal Astronomical Society*, 397, 397
- Roman-Duval, J., Jackson, J. M., Heyer, M., Rathborne, J., & Simon, R. 2010, *The Astrophysical Journal*, 723, 723
- Rosenberg, M. J. F., van der Werf, P. P., Aalto, S., et al. 2015, *The Astrophysical Journal*, 801, 801
- San José-García, I., Mottram, J. C., Kristensen, L. E., et al. 2013, *Astronomy and Astrophysics*, 553, 553
- Schöier, F. L., van der Tak, F. F. S., van Dishoeck, E. F., & Black, J. H. 2005, *Astronomy and Astrophysics*, 432, 432
- Shu, F. H., Adams, F. C., & Lizano, S. 1987, *Annual Review of Astronomy and Astrophysics*, 25, 25
- Stahler, S. W. & Palla, F. 2004, *The Formation of Stars*, by Steven W. Stahler, Francesco Palla, 865 pp. ISBN 3-527-40559-3. Wiley-VCH, 2004.
- Utomo, D., Sun, J., Leroy, A. K., et al. 2018, *The Astrophysical Journal Letters*, 861, 861
- van der Tak, F. F. S., Black, J. H., Schöier, F. L., Jansen, D. J., & van Dishoeck, E. F. 2007, *Astronomy and Astrophysics*, 468, 468

- Wu, J., Evans, II, N. J., Gao, Y., et al. 2005, *The Astrophysical Journal Letters*, 635, 635
- Yang, B., Stancil, P. C., Balakrishnan, N., & Forrey, R. C. 2010, *The Astrophysical Journal*, 718, 718
- Yang, C., Omont, A., Beelen, A., et al. 2017, *Astronomy and Astrophysics*, 608, 608
- Yang, Y.-L., Green, J. D., Evans II, N. J., et al. 2018, *The Astrophysical Journal*, 860, 860
- Yıldız, U. A., Kristensen, L. E., van Dishoeck, E. F., et al. 2015, *Astronomy and Astrophysics*, 576, 576
- Yıldız, U. A., Kristensen, L. E., van Dishoeck, E. F., et al. 2013, *Astronomy and Astrophysics*, 556, 556
- Yvart, W., Cabrit, S., Pineau des Forêts, G., & Ferreira, J. 2016, *Astronomy and Astrophysics*, 585, 585