Tracing star formation at high redshift using the Milky Way as a template

Katarzyna Magdalena Dutkowska

Dissertation submitted for the degree of

PHILOSOPHIÆ DOCTOR

Niels Bohr Institute
Faculty of Science
University of Copenhagen

December 31st 2022

This thesis has been submitted to the PhD School of The Faculty of Science,
University of Copenhagen

Title	Tracing star formation at high redshift using the Milky Way as a template
Author	Katarzyna M. Dutkowska
Department	Niels Bohr Institute, Faculty of Science, University of Copenhagen
Academic Advisor	Assoc. Prof. Lars E. Kristensen
Assessment Committee	Prof. Jes K. Jørgensen (chair) Assoc. Prof. Angela Adamo Senior Astronomer, Dr. Mario Tafalla

Typeset using LaTeX and the `memoir` documentclass.
Layout and typography inspired by Friedrich Wiemer's thesis
using the class `classicthesis` developed by André Miede.

Dla M.

Abstract

A key question in astrophysics is how most stars form, both in the nearby and the more distant Universe, near the peak of cosmic star formation. All stars form in molecular clouds and numerous surveys have provided detailed molecular inventories of such clouds within the Galaxy. Thanks to state-of-the-art facilities such as the Atacama Large Millimeter/submillimeter Array (ALMA), we are now routinely observing the distant Universe in molecular line emission. As we observe these molecules across the Universe, we start to fill the informational gap between high-z galaxies and the Milky Way.

How can we effectively compare these regimes and utilize our Galactic knowledge? We know that most stars form in giant molecular clouds and that these form clusters, where all stellar masses are present. Furthermore, relatively few high-mass stars easily outshine the entire low-mass population in a cluster. Moreover, the younger the protostar, the deeper it is embedded in gas and dust. Therefore, if we want to observe the entire forming population of stars, we need to use reliable tracers of active star formation that are common and bright enough to be easily observed and not obscured by dust extinction. One of the best tracers in our Galaxy also observed in the distant Universe is water: emission from this molecule lights up in the outflows driven by the youngest accreting protostars, and water emission thus serves as a tracer of forming stars. Furthermore, a linear empirical relation exists between the mass of the accreting star and the water intensity, making water emission a low-contrast tracer of active and current star formation. With this in mind, we can use what we know about the local star formation to constrain spatially unresolved star-formation processes in more distant galaxies and understand how molecular line emission can be used to quantitatively trace active star formation at different cosmic times.

The work presented in this thesis marks the first steps of creating an observationally based tool simulating molecular emission from star-forming regions in galaxies. The result of this work is the "galaxy-in-a-box" model, simulating water emission associated with the earliest and most active stages of clustered star formation. The initial results from running the galaxy-in-a-box model demonstrate that water emission is sensitive to the star-formation efficiency and the age of clusters. Further investigations have shown that a straightforward application of Galactic star-formation laws is not enough to match the observations, which can overestimate the amount of ongoing star formation while underestimating the expected emission.

To probe the origin of water emission further, water emission in shocked regions was studied, as this is where most of the Galactic water emission originates. This part included radiative transfer modeling of results from sophisticated shock models using the Paris-Durham shock code. The initial results show that water emission is highly dependent on the density of the medium, as well as the velocity of shock waves. The former can be especially

important in the high-redshift regime, where the interstellar medium is denser. These results carve out a path for a novel use of molecular emission as a star formation tracer across cosmic times.

Dansk resumé

Et af de vigtige spørgsmål inden for astrofysik er hvordan stjerner dannes, især i højdepunktet for kosmisk stjernedannelse. Stjerner dannes i kæmpe molekylære skyer, som findes i både det nære og fjerne univers. Undersøgelser har givet detaljerede molekylære data om sådanne skyer i Mælkevejen. Moderne faciliteter som Atacama Large Millimeter/submillimeter Array (ALMA) gør det muligt for os at observere molekylær linjeemission i det fjerne univers. Ved at observere disse molekyler kan vi begynde at forstå forbindelsen mellem høj-z-galakser og Mælkevejen når det kommer til stjernedannelse.

For at sammenligne disse forskellige afstandsregimer har vi brug for et sporstof for aktiv stjernedannelse der er udbredt, lysstærkt og ikke formørket af støv. Vandemission, der observeres både i Mælkevejen og det fjerne univers, fungerer som et effektivt spor af stjernedannelse. Denne emission findes i udstrømningerne fra unge, dannende protostjerner. Det viser sig at der er et lineært empirisk forhold mellem massen af den dannende stjerne og intensiteten af vandemissionen. Dette gør vandemissionen til et lavkontrast-sporstof af aktiv og nuværende stjernedannelse. Som resultat er det muligt for os at bruge den lokale stjernedannelse til at begrænse uopløste stjernedannelsesprocesser i mere fjerne galakser, samt forstå, hvordan molekylær linjeemission kan bruges til at spore aktiv stjernedannelse på forskellige kosmiske tidspunkter.

Arbejdet i denne afhandling har ført til udviklingen af den såkaldte "galaxy-in-a-box"-model, der simulerer vandemission forbundet med de tidligste og mest aktive stadier af stjernedannelse i hobe. De indledende resultater viser, at vandemission er følsom over for effektiviteten af stjernedannelsen og alderen på en hob. Undersøgelser har vist, at det alene at anvende galaktiske stjernedannelseslove ikke er tilstrækkeligt til at forklare observationer, som har en tendens til at overestimere mængden af igangværende stjernedannelse og underestimere den forventede emission.

For at forstå oprindelsen af vandemission nærmere har denne afhandling også set på vandemission i chockede regioner, som er den dominerende kilde til galaktisk vandemission. Dette inkluderede modellering af strålingstransport af resultater fra chokmodeller ved hjælp af Paris-Durham chokkoden. De indledende resultater viser, at vandemission er meget afhængig af både tætheden og hastigheden af chokbølger. Tætheden spiller en særlig vigtig rolle i høj-z-regimet, hvor det interstellare medium er mere tæt. Disse resultater viser, at molekylær vandemission kan bruges som sporstof af stjernedannelse på tværs af kosmiske tidspunkter.

Acknowledgments

Everything's got to end sometime. Otherwise, nothing would ever get started.
–The (Eleventh) Doctor Who

When I started my Ph.D., the only thing I expected was that everything would be different afterward. And I was right. The past years were many things that touched the backbone of who I am. Despite the wonder of Copenhagen, the Ph.D. itself turned out to be more than just a *hyggelig* career path. Choosing this career path can be very hard if you are a first-gen academic, the impostor syndrome lives with you rent-free, and/or the majority of your Ph.D. passes during the global lockdown. But these challenges and unexpected bumps on the way can help you grow as a scientist and as a person. This happened in my case, but I can't take full credit for turning these downs into ups.

I want to start by thanking my supervisor, Lars Kristensen. I can confidently say that Lars was the best possible supervisor for me. He helped me to find my way in academia and evict all self-doubts occupying my personal space. During all these years, Lars was always there for me and helped me hold the steer whenever it started to slip from my hands. He gave me all the support I needed to fly from the nest and become independent in science. Thank you for sharing your incredible expertise, which made me a better scientist. Thank you for all your "Socratic" questions, which trained my thinking. And, of course, thank you for all the doors you opened for me and the offered opportunities. And finally, thank you for the mutual trust. It meant the world to me. And, speaking of supervisors, I would also like to thank Agata Karska. Agata supervised me during my master's studies, which for personal reasons, turned out to be the most challenging and painful time of my life. Agata's trust in me helped me to get through all difficulties. Thank you for all the opportunities you gave me, for sharing your experience, and for being the first to show me how science can be done.

I also want to acknowledge my collaborators. In particular, Benjamin Godard, Antoine Gusdorf, and Jorge Villa for guiding me through the world of interstellar shocks. Thank you for your hospitality and support. You taught me a great deal, for which I am deeply grateful. I thank my discussion partners from DARK: Marianne Vestergaard, Anne Klitsch, and Oliver Damkjær. Our meetings were one of the highlights of my Ph.D.

Being a part of StarPlan was a fantastic experience. First, I would like to thank the sub-mm group members, past and present: Jes Jørgensen, Giulia Perotti, Elizabeth Artur de la Villarmois, Charl van der Walt, Rebecca Pitts, Sébastien Manigand, Sigurd Jensen, Will Rocha, Manar el Akel, Sacha Gavino, and Rajeeb Sharma. Thanks for all the stimulating conversations, curiosity, and support. I would also love to thank the rest of my colleagues at StarPlan for all the talks we had and time spent together during events, lunches, and (most importantly) coffee breaks, in particular: Ninna Jensen,

Siw Egdalen, Isaac Onyett, Mafalda Costa, Catarina Fernandes, Mattia Tagliavento, Laura Bouvier, Hannah Calcutt, Elsa Amsellem, Rajika Kuruwita, Georgy Makhatadze, Jean Bollard, Rick Verberne, Flavia Amadio, Beatriz Campos Estrada, Marie-Luise Steinmeyer, Michael Küffmeier, Lene Bentzen, Sylvia Blaauw, Mirek Groen, Elishevah van Kooten, Célia Bresson, Courtney Rundhaug, Jesper Roslund, Lu Pan, Zhengbin Deng, Aaron Schneider, Nikitha Saji, Urs Schäfer, and Antoine Petit. If I forgot about someone, please forgive me.

I would also like to thank my friends from all over the world. Thank you for all the conversations, cheerleading, and turning a blind eye when I wasn't responding or calling back for more than a week. In particular, I would like to thank Michał for sharing the path with me. I am also very grateful for my family. I want to thank my mum for accepting me for who I am and for the freedom to choose my own path. Even if she is not among us anymore, her spirit inside me lights me up and motivates me to be a better person every day. Thank you for letting me choose the path of happiness. I want to thank my father for being my greatest supporter. Thank you for teaching me that I can always be better and that I can achieve everything. Finally, I want to thank my brothers for being there for me when I need them and openness to who I am and what I do. I know it is not always easy, which makes me even more grateful. Last but not least, I want to thank my partner for always being there for me. You are my best friend and the best person I have ever met. Thank you for caring for our relationship and for me during these years. Thank you for being so kind and understanding. Thank you for keeping me sane. I wouldn't make it without you.

Contents

1

Introduction

"Every new discovery is just a reminder we're all small and stupid."
—Everything Everywhere All At Once

Probably everyone at some point in their life asked why we are here. Understanding the nature of our existence is a common human longing. Throughout history, people have developed various ways of dealing with the uncomfortable presence of the unknown and tools to slowly get us higher on the ladder of knowledge, with a great hope that one day we will get high enough to see the answer.

Despite the endless number of unanswered questions, we know a few things already. Whatever the pathway for existence was, we were all born out of cosmic matter from the Big Bang. It took billions of years to transform this matter into the form known as life. We also know that the essential part of the emergence of life was, and still is, the Sun. Hence, whatever shaped the Earth and life on Earth was a part of the cause-and-effect relationship of the formation of the Sun and its natal environment.

One thing to ask is in what environment the Sun was born. Luckily, the answer to this question is known. Our host star, like the majority of stars in the Milky Way, formed in massive star-forming clusters (e.g., Lada and Lada, 2003; Adams, 2010). Then, the second thing would be to understand the chemistry of the star formation process and how it happened to be complicated enough to form the seeds of life. Here, the past years brought immense progress in the field of astrochemistry, thanks to which we are now routinely observing star-forming regions in molecular emission (e.g., Herbst and van Dishoeck, 2009; Jørgensen *et al.*, 2020). Hence, we have built strong fundamentals to answer the question of how the building blocks of life could be delivered to a planet like ours. However, this requires an assumption: the star formation process is somewhat universal and that what happens in other regions of the Galaxy and the Universe can be applied to the Solar System.

The work presented in this thesis is the fruit of taking a step back to understand how most stars form and assess how much we can trust in our assumptions. Hence, the main goal was to create a tool that would use the Milky Way as a template for star formation and compare it with other galaxies. This thesis predominantly focuses on developing the galaxy-in-a-box model, built relying on the three main components of a galactic star-forming environment, i.e., giant molecular clouds, star-forming clusters, and protostars. The model was created with the intention to be used as a tool that can be compared with observations. Hence, the model assigns

molecular emission to each simulated protostar. In this thesis, the molecule of focus was water because, in the Milky Way, it is known to be almost uniquely associated with the most active stages of the star-formation process (van Dishoeck *et al.*, 2021a). Moreover, water emission is routinely observed in the distant Universe, more than 10 billion years ago from us, providing a direct link to galaxies at different evolutionary stages (e.g., van der Werf *et al.*, 2011a; Riechers *et al.*, 2013; Yang *et al.*, 2013; Yang *et al.*, 2016; Jarugula *et al.*, 2019; Apostolovski *et al.*, 2019). Therefore, this thesis presents the model and results of simulating emission from active star formation in galaxies with various star-forming properties. Moreover, with the goal to better understand water emission coming from shocked regions that could be found around forming stars, results of radiative transfer modeling of shock models are presented.

This thesis is structured as follows. Chapter 1 is an introduction to the star formation process and chemistry of star-forming regions. It describes different evolutionary stages and their observational signposts. Then, it gives an overview of the chemistry of star formation, emphasizing water molecules. Since water emission is associated with shocks formed at the earliest stages of star formation, the chapter includes an introduction to the physics and chemistry of interstellar shocks. Then the summary of star formation history in the Universe is presented, highlighting main differences between the local and the distant Universe.

Chapter 2 explains the methods used in this thesis. First, the basics of molecular emission and radiative transfer are outlined. Second, the overview of facilities used for molecular, in particular water, observations is provided. Then, the Water Emission Database is presented, as it was developed as an integral part of this thesis. Lastly, in this chapter, the galaxy-in-a-box model and the shock model are explained.

Chapter 3 is a summary of the articles which are an integral part of this thesis. Also, it describes how this work contributes to the field and presents some future outlooks. The articles themselves are presented in Chapters 4, 5, and 6.

1.1 THE FORMATION OF STARS

Star formation takes place in the dense interstellar medium (ISM) and spans a range of scales. In galaxies ($\sim 10^4$ pc), the predominant stellar nurseries are giant molecular clouds (GMCs; $\sim 10^2$ pc). Molecular clouds exhibit elongated, parsec-scale filamentary structures (see Fig. 1.1; e.g., André *et al.*, 2010; Arzoumanian *et al.*, 2011). In these filaments, smaller scales reveal subparsec dense regions with $n(H_2) \gtrsim 10^4$ cm^{-3}. These structures, known as dense cores or Bok globules, are the parts of the ISM that undergo collapse, leading to the formation of protostars ($\sim 10^{-1}$ pc; Bergin and Tafalla, 2007). Table 1.1 summarizes the typical densities and temperatures of these different ISM environments.

1.1.1 *The collapse of dense cores*

To understand star formation, we first need to understand the collapse of dense cores, which marks the beginning of star formation. It is possible to characterize scales that would form a protostar form within molecular clouds by considering their balance between gravity and thermal pressure.

The widely used criterion when talking about collapsing cores is the Jeans criterion (Jeans, 1902). A core in a molecular cloud is supported by its internal (and, in this case, thermal) pressure. The Jeans criterion tells us if the core's pressure is sufficient to support it against gravitational instability and subsequent collapse.

We can conceptualize the criterion by first considering a parcel of gas with a density ρ, depth $\mathrm{d}z$, and side surface area of $\mathrm{d}A$. If this parcel would be subject to a pressure gradient, i.e., there would be pressure P on one side and slightly increased pressure, $P + \mathrm{d}P$, on the other side, there would be a force, $F_\mathrm{P} = ma = -\mathrm{d}P\mathrm{d}A$, acting on the parcel from the higher to the lower pressure side. Hence, the considered parcel would be subject to an acceleration $a = -\dfrac{1}{\rho}\dfrac{\mathrm{d}P}{\mathrm{d}z}$. We can apply this to a small core within a molecular cloud.

In the parcel scenario, a molecular cloud would be characterized by its density ρ_0, sound speed c_s, and pressure P. Within this molecular cloud, we could then consider a smaller region with radius r, where the density is increased by χ, such that $\rho = (1 + \chi)\rho_0$. In this case, we would get a pressure gradient ∇P inducing the force F_P, just as in the case of the parcel

TABLE 1.1: Comparison between densities and temperatures in the different environments present in the interstellar medium. The density is expressed as hydrogen nuclei per cm^{-3}.

Environment	$n\,(\mathrm{cm}^{-3})$	T(K)	Example
Diffuse clouds	500	50	ζ Ophiuchi
Giant Molecular Clouds	100	15	Orion
Dark Clouds (Complexes)	500	10	Taurus-Auriga
Dark Clouds (Individual)	10^3	10	B1
Dense Cores/Bok Globules	10^4	10	TMC-1

Note: Adapted from Stahler and Palla (2004).

of gas, which expressed as acceleration would be as follows:

$$\frac{F_{\rm P}}{M} \sim \frac{\nabla P}{\rho} \sim \frac{\chi P}{\rho r}. \tag{1.1}$$

In this approximation, we consider an isothermal and uniform gas, hence the pressure can be written as:

$$P = \frac{{\rm k_B} T}{\mu {\rm m_H}} \rho, \tag{1.2}$$

where μ is the mean molecular weight, which accounts for the chemical composition of the gas, $\rm m_H$ is the hydrogen mass in kg, and T is the kinetic temperature of the gas in K. This first part of the equation is, in fact, the speed of sound (squared) of the given medium. Hence, we can rewrite Eq. (1.1) and (1.2) as follows:

$$P = c_{\rm s}^2 \rho \longrightarrow \frac{F_{\rm P}}{M} \sim \frac{\chi c_{\rm s}^2}{r}. \tag{1.3}$$

Then, the gravitational pull inwards would be:

$$\frac{F_{\rm G}}{M} \sim \frac{{\rm G} M \chi}{r^2} \sim {\rm G} \chi \rho_0 r. \tag{1.4}$$

We can use Eq. (1.3) and (1.4) to find the conditions under which gravitation of the cloud will overcome the internal pressure of the core:

$$\begin{aligned} F_{\rm G} \gtrsim F_{\rm P} : {\rm G} \chi \rho_0 r &\gtrsim \frac{\chi c_{\rm s}^2}{r} \\ \Rightarrow r &\gtrsim \sqrt{\frac{c_{\rm s}^2}{G \rho_0}} \equiv \lambda_{\rm J}, \end{aligned} \tag{1.5}$$

where $\lambda_{\rm J}$ is the Jeans length. Hence, if the Jeans length is exceeded, the region will collapse since gravity will take over the pressure. From Eq. (1.5), we can derive the Jeans mass, i.e., defining the critical mass of the region above which it will collapse:

$$M_{\rm J} \equiv \rho_0 \lambda_{\rm J}^3. \tag{1.6}$$

More formal derivations (e.g., Stahler and Palla, 2004) result in commonly known forms of the Jeans length:

$$\lambda_{\rm J} = \sqrt{\frac{\pi c_{\rm s}^2}{G \rho_0}}. \tag{1.7}$$

If we adopt Eq. (1.6) and (1.7) for the typical environment of dense cores, i.e., $n_{\rm H} = 10^4$ cm^{-3} and $T = 10$ K we get $M_{\rm J} = 1.0$ M$_\odot$ and $\lambda_{\rm J} = 0.19$ pc. Hence, assuming typical subparsec lengths and average masses of 10 M$_\odot$, these cores are gravitationally unstable. However, if we embed the core in a medium with external pressure and solve the internal vs. external forces, under the assumption of the hydrostatic equilibrium with a density gradient following the power-law $\rho(r) \propto r^{-2}$, we arrive at the Bonnor-Ebert (BE) solution (Ebert, 1955; Bonnor, 1956). This is the classical star formation description of the maximum mass a core can have to be gravitationally bounded. In this solution the critical core mass, $M_{\rm BE}$, follows:

$$M_{\rm BE} = 1.18 \frac{c_{\rm s}^4}{\sqrt{P_{\rm s} G^3}}, \tag{1.8}$$

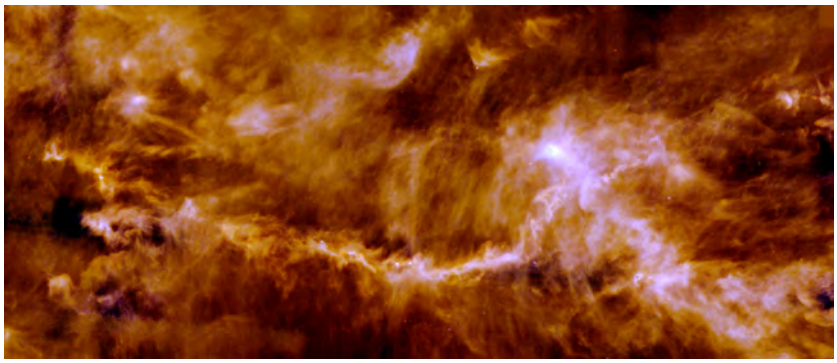


FIGURE 1.1: The star-forming B211/B213 filament in the Taurus molecular cloud ($d \sim 140$ pc; Elias, 1978). Combined observations from *Herschel* at 160 μm (blue), 250 μm (green), and 500 μm (red) show dust glowing within the cloud, revealing the filamentary structure and bright cores, which indicates the ongoing star formation withing the filaments. (Credit: ESA/Herschel/PACS, SPIRE/Gould Belt survey Key Programme/Palmeirim *et al.* (2013))

where P_s, in this case, is the pressure at the surface of the considered sphere. Hence, if the mass of the core exceeds M_{BE}, it begins to collapse in free-fall. The density profile within the collapse radius follows a power-law $\rho(r) \propto r^{-3/2}$. The accretion rate is set by the sound speed, and the infall progresses outwards as time goes by. Hence the infall region expands until it reaches the core boundaries. This is the inside-out collapse (Shu, 1977). Of course, this picture of the prestellar core collapse is an approximation, as the real-life scenario is much more complicated, as has been demonstrated by numerous observations and simulations (e.g., Kuffmeier *et al.*, 2017; Bate, 2018; Zhang *et al.*, 2020). The formation of stars involves more complex physical structures, e.g., rotation of the star-disk system, magnetic fields bending due to this rotation, as well as dynamical interactions and (proto)stellar feedback because stars predominantly form in associations or clusters. Nevertheless, the inside-out collapse serves as a conceptual explanation for what happens, at least in the case of the formation of an individual low-mass star.

1.1.2 *Protostellar evolution*

Once the collapse begins, it leads to a cascade of processes that eventually forms a protostar and possibly a planetary system. The typical classification of protostellar evolution consists of four main stages, as depicted in Figure. 1.2. This classification is based on protostar's spectral energy distribution (SED; Lada, 1987), bolometric temperature (T_{bol}; Myers and Ladd, 1993) and ratio between the submillimeter and bolometric luminosity (L_{submm}/L_{bol}; Dunham *et al.*, 2014). However, this currently holds only for low-mass protostars, since their formation is relatively well understood. There is no established evolutionary sequence for high-mass stars. With an increasing number of observations and numerical simulations, it starts to be clear that high-mass stars do not evolve as singular objects, instead due to lack of observable high-mass prestellar cores, their evolution could be linked to cloud and cluster formation (e.g., Motte *et al.*, 2018; Russeil *et al.*, 2010; Hacar *et al.*, 2018). Hence, in the following, the focus will be on the evolutionary sequence of low-mass protostars.

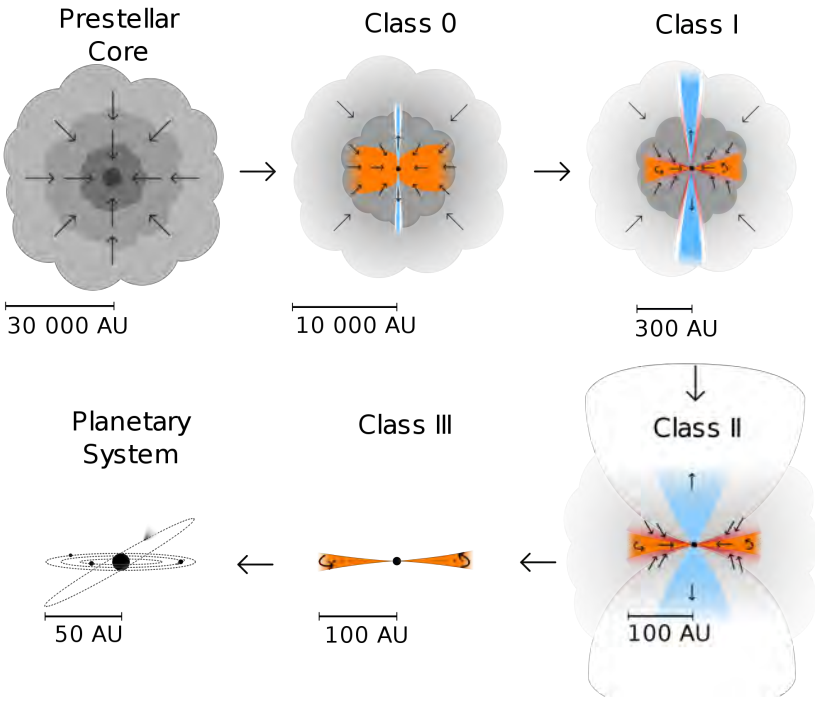


FIGURE 1.2: Stages of low-mass star formation. (Credit: M. V. Persson)

Class 0 and I

In the Class 0 stage, more than 80% of the mass of the protostellar system is located in the envelope (Jørgensen *et al.*, 2009), which typically has a size of a few thousand AU. At this stage, protostellar SEDs peak at submillimeter wavelengths. The Class 0 protostars have $T_{\rm bol} \lesssim 70$ K (Chen *et al.*, 1995) and $L_{\rm submm}/L_{\rm bol} > 0.5\%$ (Dunham *et al.*, 2014). The estimated lifetime of this stage is ~ 50000 years (Kristensen and Dunham, 2018).

Then, in the subsequent stage, the envelopes are less massive, as they get dissipated due to jets and accrete onto the central protostar, and protostars are surrounded by a disk with a radius of 25-500 AU (Williams and Cieza, 2011). The emission of a protostar once it reaches the Class I stage is a mixture of the envelope and disk contributions and the SED peaks in the mid-/far-infrared. These sources have $T_{\rm bol}$ in the range 70 – 650 K (Chen *et al.*, 1995) and their luminosity ratio is $< 0.5\%$ (Dunham *et al.*, 2014). This stage lasts approximately twice long as the Class 0 stage (Kristensen and Dunham, 2018). These two stages of star formation are associated with powerful outflows, which are described in greater detail in further parts of the chapter.

Class II and III

The later stages of star formation are characterized by the dissipation of the protostellar envelope and the flattening of the protoplanetary disk. Similarly, the outflow forces and accretion rates decline. Once a protostar reaches Class II, its SED peaks at a few μm, with a disk component coming from the thermal dust emission at mid-IR to sub-mm wavelengths. The bolometric temperatures of these objects are in the range of 650 – 2800 K (Chen *et al.*, 1995).

The subsequent grain growth and the dissipation of the disk cause the

FIGURE 1.3: The L1527 IRS Class 0 protostar in Taurus ($d \sim 140$ pc; Elias, 1978) driving outflows into the surrounding cloud. The 800 AU scale is indicated in the bottom right corner, and the orientation of the image on the sky is placed in the bottom left corner. These are combined observations from the NIRCam instrument onboard the James Webb Space Telescope from four filters F200W (blue; 1.99μm), F335M (green; 3.35μm), F444W (red; 4.43μm), and F470N (orange; 4.70μm). (Credit: NASA, ESA, CSA, STScI)

disk component of the SED to decrease in intensity, while the placement of the peak itself does not change. The object is now in Class III. There is a dusty but gas-poor disk, and coagulation of the material in the disk surrounding the star may lead to the formation of planetary companions. For this sources $T_{\rm bol}$ exceeds 2800 K (Chen *et al.*, 1995). These two last stages of star formation are also characterized by negligible outflow activity.

Protostellar outflows

One of the most characteristic processes in star formation is the ejection of material through the launch of protostellar outflows and jets (see Fig. 1.3). Broadly speaking, an outflow is protostellar ejecta, which entrains envelope material and moves away from the protostar at supersonic velocities. These outflows are known to be driven by all accreting protostars and have been observed from the 1970s and recognized as shock waves driven at velocities > 100 km/s (see Fig. 1.4; e.g., Snell and Loren, 1977).

One of the properties defining an outflow is its collimation factor, i.e., the length of the outflow divided by its width. In principle, the higher the collimation of the outflow, the younger the protostar (see Fig. 1.5). In the Class 0 stage, when most of the accretion is taking place, protostar drives highly collimated outflows. These outflows entrain and accelerate material from the protostar. At this stage, outflows exhibit extremely high-velocity

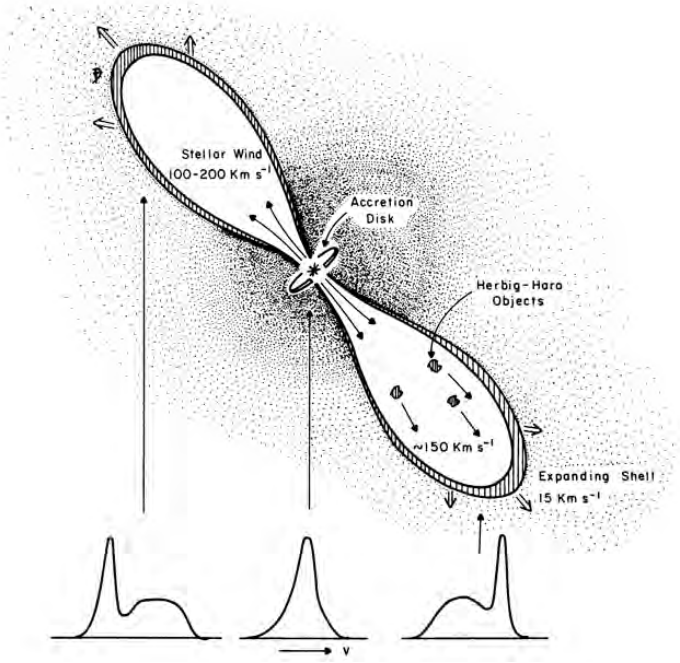


FIGURE 1.4: The first schematic drawing of the wind driven shock model of double-lobed molecular outflows from the L1551 IRS5 protostar, based on observations of CO lines $J = 1 - 0$ and $J = 2 - 1$, presented by Snell and Loren (1977). In this model the protostellar accretion disk, surrounded by the infalling envelope, drives a stellar wind moving at 100-200 km/s, leading to a formation of Herbig-Haro objects, i.e., clumps of ionised gas. This wind pushes the shell into the surrounding medium, which then expands at relatively low velocities of 10-20 km/s. Depending on what is observed, the the observed spectra exhibit different features, e.g., the blue-shifted and red-shifted line-wings coming from the blue-shifted and red-shifted outflow lobes, respectively.

molecular bullets (EHV; velocities > 50 km/s; e.g., Tafalla *et al.*, 2017), a manifestation of pulsed jets driven by a protostar.

As time evolves and the protostar enters the Class I stage, the envelope starts to dissipate in the direction of the outflow and flatten due to rotation, which moves it toward the disk plane. When this happens, the outflows tend to open up, and the collimation factor decreases. The opening angles for Class I object are $\gtrsim 90°$. Also, the EHV molecular bullets are no longer observed at this stage. Then when the protostar gets to Class II, the outflows are poorly collimated, exhibiting high opening angles $\sim 160°$ (e.g., Arce *et al.*, 2007).

However, what stands behind outflows, i.e., what is their driving mechanism, is still a matter of debate. Currently, all models assume that the energy source of an outflow is the rotational energy of a protostar-disk system which transfers the energy to the jet by magnetic fields. These two components require sufficiently large rotation and strong magnetic fields connecting to this rotation. The outflow launching models fall into three main categories: magnetohydrodynamic disk winds (e.g., Pudritz and Norman, 1986; Agra-Amboage *et al.*, 2014), stellar winds (e.g., Shu *et al.*, 1994), and star-disk winds (X-winds; e.g., Shu *et al.*, 1994). All three scenarios predict that the outflow/jet is rotating due to the rotation of the star and/or disk and that the rotation speed of the outflow/jet relates to the launch point rotation speed. This rotation is also now seen with the Atacama Large Millimeter/submillimeter Array (ALMA) and usually indicates a wide-angle

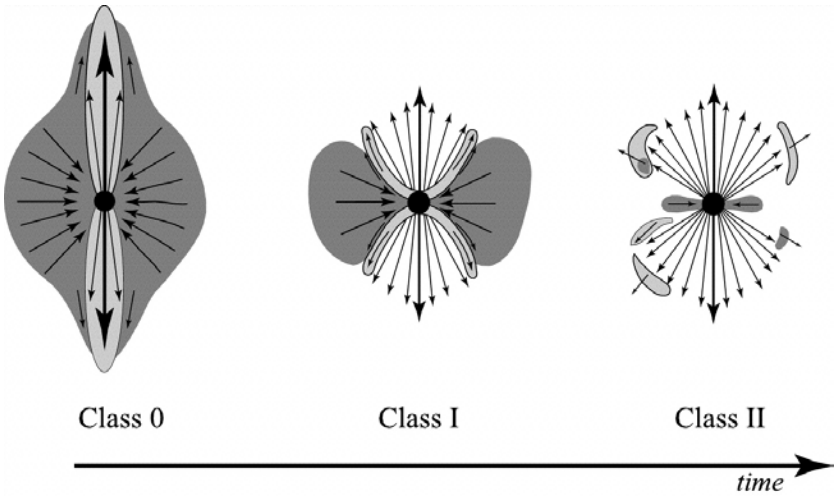


FIGURE 1.5: Schematic evolution of a protostellar outflow and envelope. Dark gray regions correspond to the envelope, while light gray regions to the outflow. Arrows indicate gas motion in the system. Taken from Arce and Sargent (2006).

MHD disk wind as a launch mechanism (e.g., Aso *et al.*, 2015; Bjerkeli *et al.*, 2016; Tabone *et al.*, 2017; de Valon *et al.*, 2020) or the X-wind (Lee *et al.*, 2022). However, even with this growing observational evidence, there is no definite answer to what underlies the process of outflow ejection and where the acceleration of material takes place.

1.1.3 *Star-forming clusters*

The current understanding of star formation provides a framework for an isolated low-mass star formation, as presented in Sect. 1.1.2. However, studies show that most stars form in young massive clusters (see Fig. 1.6; e.g., Lada and Lada, 2003; Lee *et al.*, 2012; Ward *et al.*, 2020), like the Orion Nebula Cluster located at a distance of ~ 414 pc (Palla and Stahler, 1999; Menten *et al.*, 2007). Protostars in these clusters are evolving in a different environment, often being exposed to, e.g., radiation from high-mass (proto)stars and dynamical interactions (e.g., Longmore *et al.*, 2014). These star clusters form due to the gravitational collapse of GMCs (Kennicutt and Evans, 2012) and when they are young, i.e., their age is less than a few Myr, they are deeply connected with the natal molecular clouds (Adamo *et al.*, 2020).

This thesis follows a cluster definition from Lada and Lada (2003), where a cluster is treated as a physically related group of stars with stellar mass volume density large enough to keep the group gravitationally bounded and resistant to tidal disruption by the galaxy and passing interstellar clouds. The cluster must also consist of a sufficient number of members to ensure that its evaporation time (Binney and Tremaine, 1987), i.e., the time it takes for internal stellar encounters to eject all its members, is greater than 10^8 years. In numbers, this definition implies that a stellar cluster is a group of 35 or more physically related stars whose stellar mass density exceeds 1.0 $M_\odot pc^{-3}$.

Commonly when talking about clusters, we refer to globular or open clusters. These clusters are evolved systems with stars that have passed the

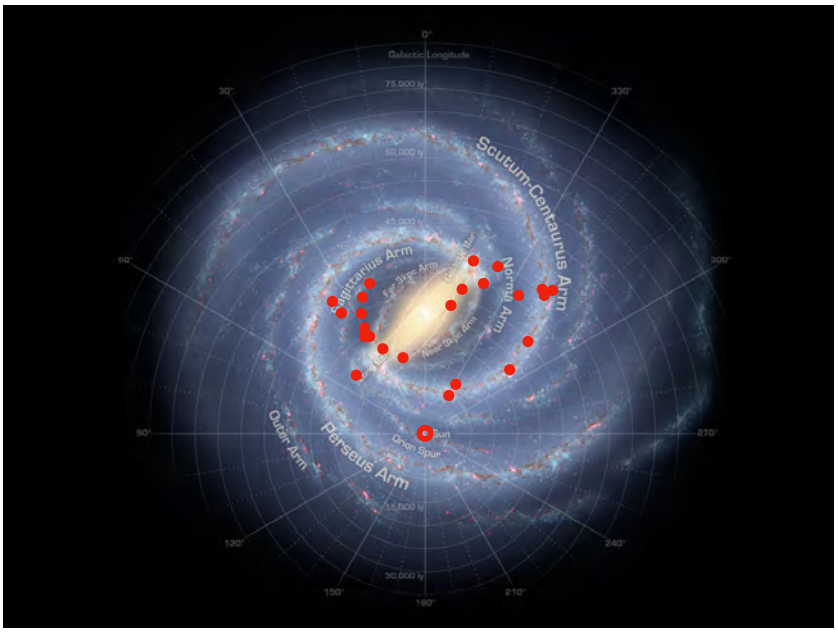


FIGURE 1.6: Cartoon representing the approximate location of 24 GMCs hosting ~ 50% of star formation in the Galaxy (filled circles). The empty circle represents our location in the Milky Way (credit for the background image: NASA/JPL-Caltech/R. Hurt (SSC/Caltech); location of the points based on Lee *et al.*, 2012).

hydrogen-burning point and are easily observable in visible wavelengths. However, in this thesis, the clusters in focus are young and deeply embedded in gas and dust, hosting different types of young stellar objects (YSOs). As mentioned already, these clusters form out of GMCs. Studies indicate this can happen hierarchically, and clusters follow the structure of galactic GMCs (e.g., Kruijssen, 2012; Krumholz and McKee, 2020; Ward *et al.*, 2020). The hierarchical scenario is a subject of this thesis, which focuses on simulating outflow emission from protostars within massive young clusters that formed from a GMC distribution in galaxies. The work on emission from protostars in galactic star-forming clusters is described in Chapters 4 and 5.

1.1.4 *The stellar initial mass function*

One of the most fundamental distributions in astrophysics, especially important for understanding the formation of stars in clusters, is the initial mass function (IMF; e.g., Chabrier, 2003; Offner *et al.*, 2014). This distribution is particularly useful for describing globally the formation of stars in clusters. The classical form of the distribution goes back to 1955 when Salpeter (1955) proposed the so-called original mass function (MF):

$$\xi(\log m) = \frac{\mathrm{d}(N/V)}{\mathrm{d}(\log m)} = \frac{\mathrm{d}(n)}{\mathrm{d}(\log m)}, \tag{1.9}$$

defining the number of stars N in a volume of space V observed at a time t per logarithmic mass interval $\mathrm{d}(\log m)$. The N/V value is the stellar number density n in pc^{-3}. This MF function is a starting point for many variations, and one of them is the IMF.

Following Chabrier (2003), the IMF describes the total number density of starlike objects ever formed per unit log mass. In practice, the IMF provides crucial information about stars formed in a cloud or cluster by providing

their initial mass distribution, i.e., when they would enter the main sequence. In this thesis, the standard form of the IMF is that for young clusters and disks proposed by Chabrier (2003):

$$\begin{aligned} &\text{For } m < 1\ \mathrm{M}_\odot:\ \xi(\log m) = A_{\mathrm{lm}} \exp[-(\log m - \log m_{\mathrm{c}})^2/2\sigma^2] \\ &\text{For } m > 1\ \mathrm{M}_\odot:\ \xi(\log m) = A_{\mathrm{hm}} m^{-1.3\pm0.3}, \end{aligned} \tag{1.10}$$

where the low-mass end normalization constant $A_{\mathrm{lm}} = 0.158^{+0.051}_{-0.046}$, the high-mass end normalization constant $A_{\mathrm{hm}} = 4.4 \times 10^{-2}$, the characteristic mass $m_{\mathrm{c}} = 0.079^{+0.021}_{-0.016}\ \mathrm{M}_\odot$, and the standard deviation $\sigma = 0.69^{+0.05}_{-0.01}$. Usually, the IMF is presented in two forms, i.e., as a number distribution or a mass distribution. In the number distribution the IMF is a function of mass, m, and describes the number of stars in a certain mass bin. While in case of the mass distribution, the IMF is a function of $\log m$ and tells us about the mass of stars in a given mass interval. It is easy to switch between two types of distributions, by adjusting the slope of the function (Scalo, 1986):

$$\xi(\log m) \propto m^x \Rightarrow \xi(m) \propto m^{x-1} \tag{1.11}$$

Hence, knowing the mass of the cloud or a cluster, the IMF lets us predict the final product of the star-formation process. However, there are traps on the way. The universality and the form of the function are a subject of an ongoing debate (e.g. Offner *et al.*, 2014). Besides numerous standard functions (e.g., Salpeter, 1955; Pflamm-Altenburg *et al.*, 2009), some studies propose bottom- or top-heavy IMF forms, i.e., producing more low- or high-mass stars, respectively (e.g., van Dokkum and Conroy, 2010; Chabrier *et al.*, 2014). Other studies propose bottom- or top-light forms, i.e., resulting in a deficit of low- or high-mass stars, respectively (e.g., Figer, 2005; van Dokkum, 2008; Watts *et al.*, 2018). There are also, to name a few, obese, diet Salpeter, and paunchy IMFs (Fardal *et al.*, 2007). These differences in predicted IMFs can be typically attributed to studied galactic types. It is suggested that IMF may evolve with time, such that it is more weighted toward high-mass stars at higher redshift (van Dokkum and Conroy, 2010). However, studies of the most massive ellipticals, thought to be descendants of intensely star-forming galaxies at high redshift, suggest that bottom-heavy IMFs are required with increasing galaxy mass (Smith, 2020). Moreover, it is not clear whether the IMF is a universal or an environment-dependent function (e.g. Davé, 2008; Bastian *et al.*, 2010; Cappellari *et al.*, 2012). Recent studies also propose a temperature-dependent IMF (e.g., Sneppen *et al.*, 2022; Steinhardt *et al.*, 2022). Hence, the IMF takes numerous forms depending on the observations and/or models, and despite sometimes ambiguous definitions, it manifests the lack of consensus on the nature of the star formation process. However, it is clear that any successful model of cloud-scale star formation, must be able to reproduce the IMF. And so it is crucial to have observational constraints on *what* the IMF is in currently star-forming clouds for a range of environmental conditions.

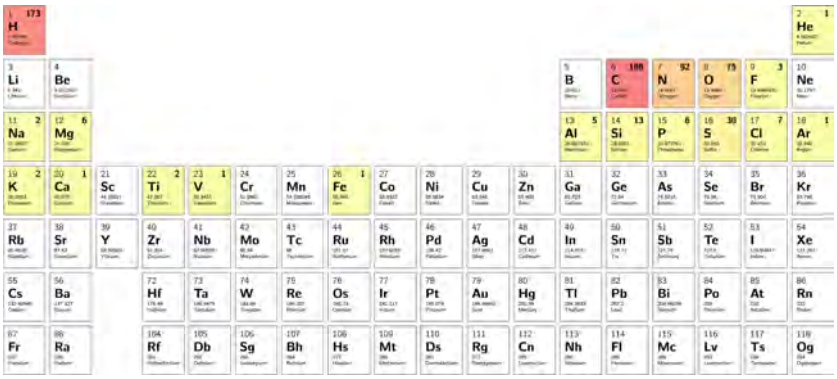


FIGURE 1.7: Periodic table of the elements detected in the Universe (marked in yellow, orange, and red). In the top right cell of each element detected in the ISM there is a number describing its occurrence among species. Taken from McGuire (2022).

1.2 THE CHEMISTRY OF STAR FORMATION

Astrochemistry, i.e., studies of intertwined chemistry and physics of the Universe, covers everything from the circumstellar medium (CSM), interstellar medium (ISM) to the circumgalactic medium (CGM). The molecular census for 2021, that includes observations of the ISM and the CSM, states the presence of 241 molecules composed of 19 different elements (see Fig. 1.7; McGuire, 2022). This is a major progress since the very first molecule, methylidyne (CH), was found in the ISM (Swings and Rosenfeld, 1937). The highest fraction of the identified molecules (36.2%) have been first detected in star-forming regions. Hence, star formation is associated with rich chemistry, which in turn makes it an incredible diagnostic of the evolution of protostars, by tracing, e.g., changes in the density, temperature, and velocity (Jørgensen *et al.*, 2020).

1.2.1 *Heating and cooling in the ISM*

Before going straight into the chemical processes observed in the ISM, and more specifically in the Galactic star-forming regions, it is worth to consider what influences the thermal budget of the ISM. The temperature plays a key role in setting a number of ISM parameters. These include, but are not limited to, chemical reaction rates, the sound speed, the strength of the continuum, and line radiation.

To heat the gas there needs to be an energy transfer into and/or within the medium. One of the ways in which it can be done is through interaction with energetic particles and photons, e.g., cosmic rays (CRs). Cosmic rays are long-known to be an important heating source of the molecular gas, especially that in dense molecular cores, which is shielded from the UV photons (e.g., Goldsmith and Langer, 1978). These highly-energetic particles are associated with massive stars, supernovae remnants (SNRs), and star-forming regions (e.g., Abdo *et al.*, 2010; Armillotta *et al.*, 2022). Since CRs primarily consist of relativistic protons they can ionize atomic and molecular hydrogen, e.g.,:

$$p^+ + H_2 \rightarrow H_2^+ + e^- + p^+. \tag{1.12}$$

In molecular clouds free electrons efficiently heat the gas indirectly through dissociation of H_2 molecules:

$$e^- + H_2 \rightarrow H + H + e^-. \tag{1.13}$$

When H_2 reforms, part of the released binding energy (4.5 eV) goes into heating the gas. However, CRs are mostly important in the densest regions of molecular clouds. Hence, in lower densities what takes over is the interstellar radiation field (IRF), which consist of the cosmic microwave background, dust emission, starlight, and hot plasma. There are three main ways in which it can heat the clouds, i.e, ionisation of atoms and molecules (UV-photons and X-rays), photo-electric effect (ejecting electrons from dust grains with UV-photons), and irradiation of grains (absorption of optical and UV-photons).

The primary cooling process of the ISM is the radiation of the energy. Here, the most important thing to consider are collisions, because cooling typically proceeds by collisional excitation followed by an emission of a photon (radiative de-excitation):

$$\begin{aligned} A + B &\rightarrow A^* + B \\ A^* &\rightarrow A^* + h\nu, \end{aligned} \tag{1.14}$$

where A^* represents a molecule in an excited state. The most important coolant in molecular clouds is CO (e.g., Neufeld *et al.*, 1995), as its low lying levels favor the molecule to be readily (de-)excited. Also, the dust grains can serve as important coolants. Their collisions with atoms and molecules can lead to lattice vibrations, which decay through the emission of infrared (IR) photons.

1.2.2 *Different types of reactions*

Most of the reactions occurring in the star-forming regions are exothermic, i.e., they release energy (Herbst and Klemperer, 1973). The main astrochemical gas-phase reactions can be divided into three groups: bond formation, bond destruction and bond rearrangement, while grain-surface and ice processes include various chemistry processes, freeze-out/adsorption, and ice sublimation/desorption (see Fig. 1.8).

Bond formation

The most effective gas-phase reactions, where bonds form are radiative association and associative detachment, where the bond formation energy is carried away by photons and electrons, respectively:

$$\begin{aligned} A + B &\rightarrow AB + h\nu \\ A^- + B &\rightarrow AB + e^-. \end{aligned} \tag{1.15}$$

For the conditions prevalent in most of the ISM, radiative association is the most important gas-phase bond formation pathway (Herbst and Klemperer, 1973; Öberg and Bergin, 2021). In this type of reaction a collision of two gas-phase species A and B leads to the formation of a bond between the two species and emission of a photon. Similarly, for the associative detachment, where one of the reactants is ionized and an electron is emitted.

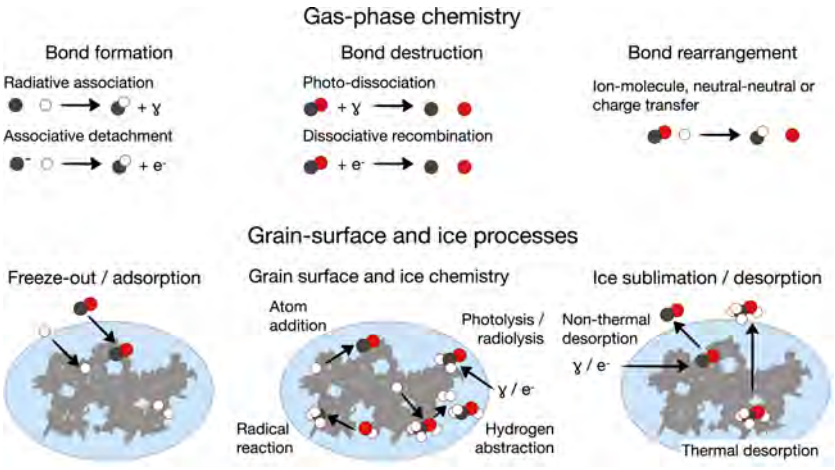


FIGURE 1.8: Gas-phase and grain-surface reactions occurring in the ISM and CSM. Taken from Öberg and Bergin (2021).

Bond destruction

There are three main ways to destroy molecular bonds, i.e., photodissociation (absorption of photons), dissociative recombination (collision with electrons), and collision-induced dissociation:

$$\begin{aligned} AB + h\nu &\rightarrow A + B \\ AB + e^- &\rightarrow A + B \\ AB + C &\rightarrow A + B + C. \end{aligned} \tag{1.16}$$

Photodissociation leads to the breaking of a molecular bond via absorption of a photon. Since typical covalent bond strengths are greater than 5 eV, UV photons are required and photodissociation dominates in parts of the ISM exposed to UV-photons. Hence, it is a dominant process in the vicinity of massive (proto)stars and in photon-dominated regions (PDRs). Other possible destruction pathways are dissociative recombination via collisions with electrons and cosmic rays, as well as collision induced dissociation in dissociative (high-velocity) shocks.

Bond rearrangement

The last type of gas-phase reaction is a bond rearrangement:

$$AB + C \rightarrow A + BC. \tag{1.17}$$

In case of this reaction the net bond formation energy is carried away by A as kinetic energy. Hence, in the bond rearrangement the excess energy does not have to be radiated away. Usually, this reaction involves radicals (atoms, molecules, or ions with at least one unpaired electron) or ions or both.

Therefore, depending on the reactants we divide ion-molecule (typically exothermic) and neutral-neutral (typically endothermic) reactions. The former involves ion-capture by a molecule due to the long-range charge attractions between the two reactants through an induced dipole moment. Ion-molecule reactions tend to dominate gas-phase chemistry in the ISM (Herbst and Klemperer, 1973) and are responsible for the formation of CO, the second most abundant molecule in the Universe (van Dishoeck and Black, 1986). Neutral-neutral reactions typically involves two non-radicals and

typically have high activation barriers. However, if one of the reactants is a radical or an atom, it can lower the activation barrier and such reactions then become important to the formation and destruction of molecules (Smith *et al.*, 2004). Moreover, neutral-neutral reactions lead to formation of water, of the most fundamental species in star-forming regions.

Freeze-out and ice sublimation

At the lowest temperatures of dense clouds, when gas-phase species collide with cold dust grains with temperatures < 30 K, they condense onto grains, except for He and H_2, which do not stick to the grain (Hollenbach and Salpeter, 1971; Crapsi *et al.*, 2007; Pagani *et al.*, 2007). This condensation process is referred to as freeze-out, and cold dust grain surfaces are called icy mantles. Freezing-out of atoms and molecules, along with chemical reactions occurring between them and already condensed species, leads to a build-up of icy mantles.

Molecules locked up in the icy mantles can be released to the gas phase through desorption reactions. When there is a bulk thermal heating of the grain and its temperature is higher than the sublimation temperature of chemical species, they get released into the gas phase. This is what we refer to as the thermal desorption and it dominates in heated regions like protostellar envelopes (e.g., Leger *et al.*, 1985; Fraser *et al.*, 2001). However, there are also local desorption processes induced by UV photons and cosmic rays, that can be direct or indirect.

Local processes do not involve bulk heating of the grain. Instead, energy is deposited locally on the grain and leads to the desorption of one or a few molecules. Examples of such local processes include photodesorption, where a UV photon hits a molecule in the ice mantle, deposits its energy, which is converted to kinetic energy in the molecule. With this additional kinetic energy, the molecule may desorb from the ice. Photodesorption dominates when the impinging UV field is present (e.g., Hollenbach *et al.*, 2009; Fayolle *et al.*, 2011). Cosmic rays may provide local spot heating of the grain, and may provide enough energy to desorb a small number of molecules. When molecules react in the ice, the released bond energy may also be converted to kinetic energy, which can cause the release of one or more molecules, which can be especially important in the UV-shielded regions (e.g., Shen *et al.*, 2004). Finally, molecules may be liberated from the ice mechanically, when the ice is bombarded by, for example, high-velocity H_2 molecules, or H or H_2 atoms. This process is referred to as sputtering, and it is an active process in shocks and outflows (e.g., Kristensen *et al.*, 2010b; Suutarinen *et al.*, 2014).

Grain surface chemistry

The atoms and molecules which freeze-out onto icy mantles are subjected to chemistry, which often leads to the formation of complex organic molecules (COMs), i.e., carbon-bearing molecules with six or more atoms. Moreover, the most abundant molecule in the Universe, H_2, is thought to form efficiently via grain surface chemistry (e.g., Gould and Salpeter, 1963; Hollenbach and Salpeter, 1971; Wakelam *et al.*, 2017).

Even though the conditions prevalent in the CSM and the ISM stand in the way of three-body reactions, what happens on dust grains effectively

is a three-body reaction since the grain acts as a catalyst by adsorbing the bond formation energy (Öberg and Bergin, 2021). Hydrogenation, i.e., H atom addition sets the initial grain-surface chemistry. Hydrogen atoms, due to their small sizes, diffuse on grain surfaces via tunneling (Hasegawa *et al.*, 1992). The H atoms react sequentially with heavier elements and molecules, leading to the formation of more and more complex species, with the most abundant ice species being H_2O (e.g., Ioppolo *et al.*, 2008; Miyauchi *et al.*, 2008). Other reactions involve UV or electron-induced dissociation and hydrogen abstraction. Radicals produced via these reactions can recombine and form more complex molecules (e.g., Garrod *et al.*, 2008; Chuang *et al.*, 2017).

1.2.3 *Shocks*

Shocks are ubiquitous phenomena in the ISM and CGM. In the case of star formation, shocks are born due to protostellar outflows and jets and lead to rich chemical processing of the surrounding medium as they convert kinetic energy into magnetic and thermal energy. Hence, they serve as an excellent diagnostic of the star formation process, especially the earliest stages when protostellar outflows are the strongest due to the interaction between the infalling envelope, winds, and jets launched from the accreting protostar(Bally, 2016).

The formation of a shock is induced by a shock wave moving with velocity v_S through an ambient medium with a local sound speed, c_s, such that $v_S > c_s$. Hence, when the supersonic shock wave collides with the medium, its speed decreases, leading to the subsequent steepening of the wave. When the back of the wave catches up with its front, it breaks, and shock is born. Depending on the nature of the shock wave and the medium, a jump- (J), a continuous- (C), or a CJ-type shock can be born (see Fig. 1.9). For consistency, all explanations in this section are presented in the reference frame of the shock, i.e., the shock front is considered stationary and the pre-shock material streams toward the front, and once the pre-shock encounters the front it decelerates with respect to the front.

Shock types

Broadly speaking, a J-type shock forms if there is no transverse magnetic field or it is very weak. Under unmagnetized conditions, neutrals and ions are fully coupled with each other. Hence, the interaction of the shock front with the surrounding medium slows down both fluids via a combination of viscous stresses and viscous heating. What happens then is that the incoming pre-shock gas quickly jumps from a supersonic to a subsonic regime, with respect to the neutral sound speed, over a distance of about a few mean free paths, resulting in a jump in the temperature, density, and pressure of the gas. In this shock type the maximum kinetic temperature depends on the shock velocity (temperature changes as velocity squared, following the Rankine–Hugoniot relations).

If the transverse magnetic field is strong, a C-type shock emerges. The strongly magnetized nature of the medium leads to the decoupling of neutrals and ions as the Lorentz force decelerates the fluids. Subsequently, neutrals slow down due to the decoupling-induced drag force of the ions. If this

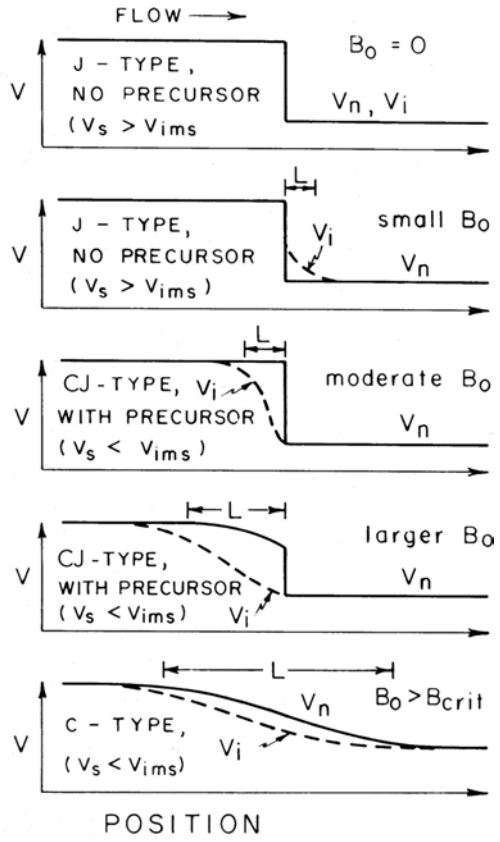


FIGURE 1.9: The evolution of interstellar shocks when shock velocity, v_s, magneto sonic speed in ions, v_{ims}, and the transverse magnetic field, B_0, change. In the case of no or small B_0, there will be no magnetic precursor, and a J-type shock forms. However, when B_0 is large enough for a magnetic precursor to emerge but is too weak to result in a continuous deceleration of neutrals, the structure will be a CJ-type shock. While a C-type shock forms in the presence of strong B_0. Adapted from Draine (1980).

force applies for a longer time, it lowers the strength of the viscous stresses due to the smoothing of the velocity gradient and thermal pressure gradient of neutrals. Hence, the neutrals decelerate continuously via the pressure gradient and the ion-neutral friction. However, if the magnetic field is not strong enough to smooth the velocity gradient, neutrals and ions can recouple, and the C-type shock becomes a J-type shock, raising the CJ-type structure. The difference between the C- and CJ-type shocks can also be explained by the their cooling properties. If the cooling is not efficient, i.e., it cannot keep up with the heating, there will be a discontinuous transition from the supersonic fluid to a subsonic fluid, resulting in a CJ-type shock. However, when the cooling is efficient and the thermal pressure gradient remains strong, a C-type shock forms (Lehmann *et al.*, 2022).

Thus, we can divide shocks can based on their critical velocities and cooling properties (Draine, 1980):

$$\begin{aligned} &\text{J} - \text{type} : V_\text{s} > B/\sqrt{4\pi\rho_\text{i}} \\ &\text{C} - \text{type} : V_\text{s} < B/\sqrt{4\pi\rho_\text{i}} + \text{effcient cooling} \\ &\text{CJ} - \text{type} : V_\text{s} < B/\sqrt{4\pi\rho_\text{i}} + \text{not efficient cooling}, \end{aligned} \tag{1.18}$$

where $B/\sqrt{4\pi\rho_\text{i}}$ is the magnetosonic speed in the charged (ionized) fluid with B being the strength of the transverse magnetic field and ρ_i the ionized

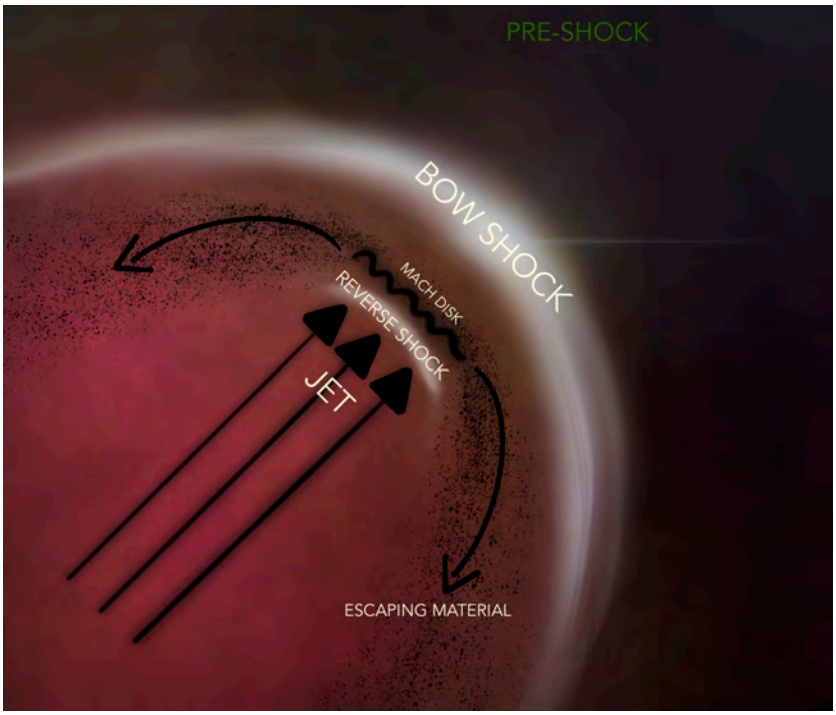


FIGURE 1.10: Cartoon showing the structure of an interstellar shock driven by a protostellar jet as described in Sect. 1.2.3.

fluid density.

Shock structure

The structure of shocks is relatively complex (see Fig. 1.10). Taking the protostellar outflow as an example, the ISM streams toward the front of the outflow, the apex, and the jet moves toward the back of this shock, leading to the formation of a contact discontinuity, where both fluids impact one another. This complex interplay gives rise to numerous structures.

The outermost structure is the shock front from the medium side, called the bow shock or the forward shock. This part of the shock directly interacts with the undisturbed interstellar medium. The shock front seen from the jet side is called a Mach disk. As the jet interacts with the ambient medium and its material slows down, a reverse shock is formed. Moreover, as the two fluids collide, the material escapes sideways from the Mach disk. The material, seeing the ISM streaming toward it, gets deflected and bends backward toward the jet, forming the characteristic shape of the bow shock.

1.2.4 *Water in star forming regions*

Among many molecules found in Galactic star-forming regions, water stands out due to its almost unique association with protostellar outflows (e.g., Mottram *et al.*, 2014; van Dishoeck *et al.*, 2021b). Water can be formed both in the gas phase and on the dust grain surface as is illustrated in Figure 1.11. The three main formation pathways are via ion-neutral reactions, neutral-neutral reactions at high temperatures, and on the dust grains. In star-forming regions, water is sputtered from the icy mantles in shocks. Moreover, the shock chemistry is associated with high temperatures which support its efficient gas-phase synthesis. Hence, when protostellar outflows

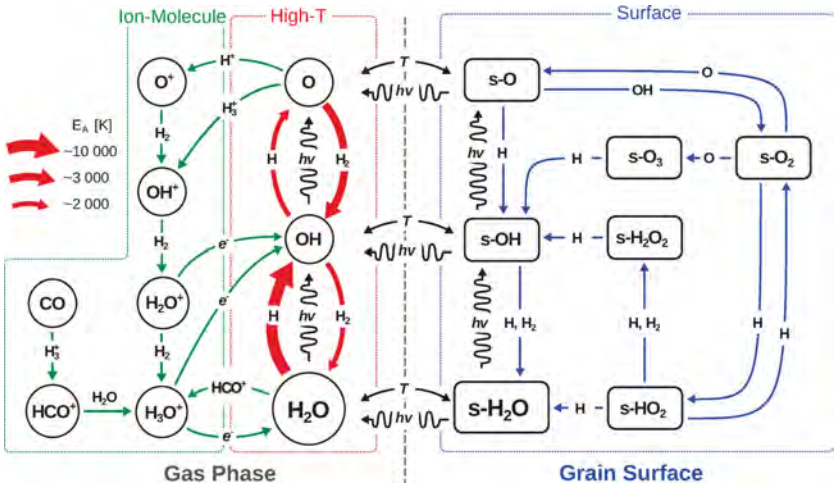


FIGURE 1.11: Water gas-phase and grain surface chemistry. Water can form on dust grain mantles through various diffusive reactions (blue arrows) initiated by the reaction of H with O or O_2. At low gas temperatures (green arrows), water gas-phase formation involves ion-molecule reactions, while at high temperatures (high-T; red arrows), which occur in shocks, water synthesis occurs via neutral-neutral reactions. This model of water chemistry does not include sputtering, which should be represented as additional arrows from the Surface to the High-T region. The figure is taken from van Dishoeck *et al.* (2013).

are launch, there is a jump in the water abundance of many orders of magnitude.

Formation pathways

Water is the most abundant ice species. Its formation on the dust grains happens through various diffusive reactions, as illustrated in Fig. 1.11. These different surface formation pathways have been extensively tested and quantified in laboratories (e.g., Ioppolo *et al.*, 2008; Ioppolo *et al.*, 2010; Lamberts *et al.*, 2013). These surface reactions are typically initiated by the reaction of H with O or O_2. In the latter, the water formation pathway starts by diffusive formation of O_2: $O + O \rightarrow O_2$, as there is no O_2 in the gas phase. Once H and O/O_2 diffuse, they can form the radical OH (for atomic oxygen) or HO_2 radical (for molecular oxygen). Subsequently, OH can get hydrogenated, leading to the formation of water. In the case of HO_2, the hydrogenation needs to happen twice. It first leads to the formation of H_2O_2 and then H_2O along with OH. In this formation route, OH can undergo hydrogenation and form water. However, it is also possible that instead of H_2O_2, the product could be two OH molecules. In this scenario, the two OH molecules could undergo subsequent hydrogenation and form water (van Dishoeck *et al.*, 2013, and references therein).

At low gas temperatures, water gas-phase formation involve ion-neutral reactions. These reactions involve dissociative recombination of H_2O^+ and H_3O^+, where the former also forms the radical OH. This formation route is inefficient in dark molecular clouds (Hollenbach *et al.*, 2009), but in PDRs and X-ray-dominated regions (XDRs), this mechanism may be dominant (Meijerink *et al.*, 2011). At gas temperatures greater than $\sim$ 300 K (e.g., Draine *et al.*, 1983; Bergin *et al.*, 1998) water is formed via neutral-neutral reactions:

$$\begin{aligned} H_2 + O &\rightarrow OH + H \\ H_2 + OH &\rightarrow H_2O + H. \end{aligned} \tag{1.19}$$

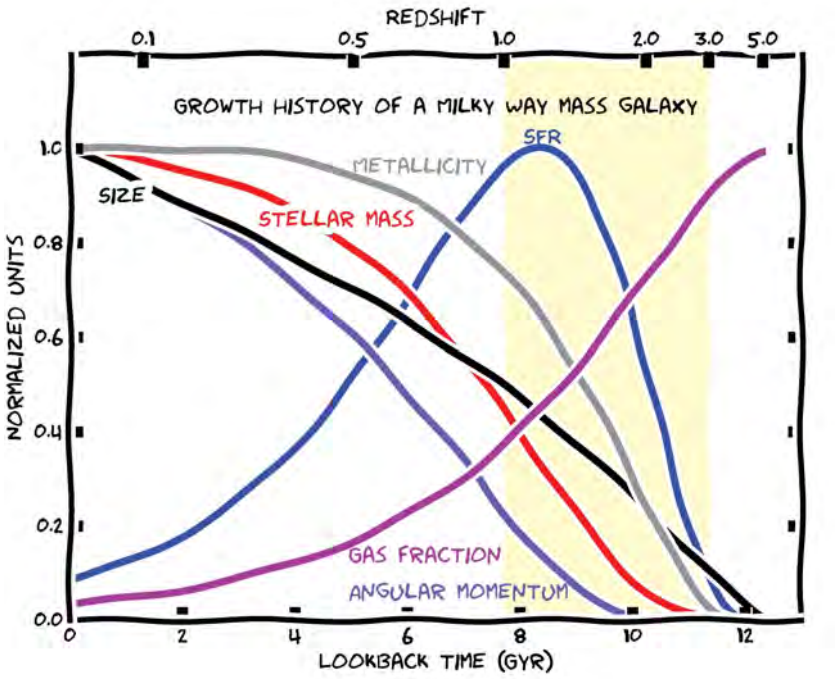


FIGURE 1.12: Plot presenting the evolutionary history of the Milky Way including its gas mass fraction (magenta), star formation rate (blue), gas-phase metallicity (gray), rest-optical effective radius (black), and stellar angular momentum (purple). The yellow shaded region correspond to cosmic noon. Taken from Förster Schreiber and Wuyts (2020).

These neutral-neutral reactions at high temperatures are the most efficient route of the gas-phase synthesis of water and are associated with the interstellar shocks. Hence, in star-forming regions these reactions are indicate tight relation of the water chemistry with the shock chemistry.

Two dominant paths for getting the water into the gas phase are thermal desorption and sputtering. The former one is the most effective and happens if the dust grains have temperatures $\gtrsim$ 100 K (e.g., Fraser *et al.*, 2001). This typically applies to the inner $\sim 10^2$ AU of forming stars (e.g., Bisschop *et al.*, 2007). Sputtering is the dominant process in shocked regions, such as protostellar outflows, but because of its energetic character, it can also lead to the dissociation of water molecules. However, within outflows, temperatures are high enough to make gas-phase synthesis of water effective, hence sustaining the high abundance of water molecules (e.g., Kristensen *et al.*, 2010b; Tafalla *et al.*, 2013; Suutarinen *et al.*, 2014).

1.3 THE UNIVERSALITY OF CHEMISTRY AND STAR FORMATION

Astrochemistry is a way of studying star formation. Emission from molecules can often penetrate through opaque protostellar envelopes and trace the earliest stages of protostellar evolution. We observe a molecular richness toward Galactic star-forming regions, and so they naturally become testbeds for our understanding of astrochemistry (e.g., Herbst and van Dishoeck, 2009; Jørgensen *et al.*, 2020). The same molecules are readily observed toward external galaxies, including the most unique star formation tracers like water and high-J CO (e.g., Hodge and da Cunha, 2020). Hence, the

question is, are these molecules tracing the same physical components across the Universe? Answering this question depends on many additional unknowns. However, the rich chemistry of Galactic star-forming regions implies that it is reasonable to assume that we will also see similar richness across the Universe. Hence, bridging Galactic and extragalactic regimes with astrochemistry could help us assess, e.g., the universality of star formation laws.

1.3.1 *The history of star formation in the Universe*

Studies of cosmic star formation history reveal that most of the star formation in the Universe happened at redshifts of $z \sim 1-3$ (see Fig. 1.13; Madau and Dickinson, 2014; Förster Schreiber and Wuyts, 2020), i.e., approximately 10.5 Gyr - 11.5 Gyr ago (under the assumption of a flat ΛCDM cosmology). This peak epoch of cosmic star formation history is referred to as cosmic noon. Star-forming galaxies (SFGs) observed at cosmic noon are today's massive disk and elliptical galaxies, and scaling relations that involve size, kinematic, and metal and gas content hold out to $z \sim 3$ (e.g., van der Wel *et al.*, 2014; Tacconi *et al.*, 2020).

Distant galaxies (in this thesis, "high-redshift" or "distant" refers to $z > 1$) can help us understand if star formation itself changes with the environment. Moreover, these young galaxies are expected to be "full of Orions" (see Sect. 1.1.3; Rybak *et al.*, 2020). Hence, it should be possible to understand the differences between the Galactic Orion and extragalactic Orions. However, the star formation activity of these galaxies is rather extreme with star formation rates (SFRs) reaching thousands of solar masses per year while the Milky Way currently forms approximately one solar mass per year (Kennicutt and Evans, 2012). This strongly suggests that the environment may be very different.

One of the measurable changes to the star formation process is the star formation efficiency (SFE), which describes the amount of molecular reservoir turned into stellar mass. For Galactic star-formation, SFE values can be as low as 1% (Lada and Lada, 2003), while high-redshift star-forming regions can have SFEs of the order of 30% (e.g., Dessauges-Zavadsky *et al.*, 2019). Moreover, as a natural consequence of the earlier evolutionary stage, cosmic-noon galaxies have clumpy morphologies and higher black hole accretion rates, a larger amount of available molecular reservoir, and lower metallicities (e.g., Conselice, 2014; Förster Schreiber and Wuyts, 2020; Tacconi *et al.*, 2020). These differences would also hold for the Milky Way at earlier evolutionary stages (see Fig. 1.12). Clumpy morphologies could be attributed to the presence of massive star-forming clumps. Studies suggest that they can form through a mechanism similar to that observed in local galaxies, i.e., fragmentation, but which would happen under different ISM conditions (e.g., Adamo *et al.*, 2013; Dessauges-Zavadsky and Adamo, 2018). For example, high-density star formation of UltraLuminous InfraRed Galaxies (ULIRGs; see also Sect. 1.3.2) throughout the Universe is thought to exhibit cosmic-ray energy densities $U_{\mathrm{CR,ULIRG}}$ that are $\sim$ few $\times$ $(10^3 - 10^4) U_{\mathrm{CR,Gal}}$ (Papadopoulos, 2010). Also, the minimum temperatures of the molecular gas in dense cores in extreme starbursts could be $\sim$ 80 K, while in our Galaxy

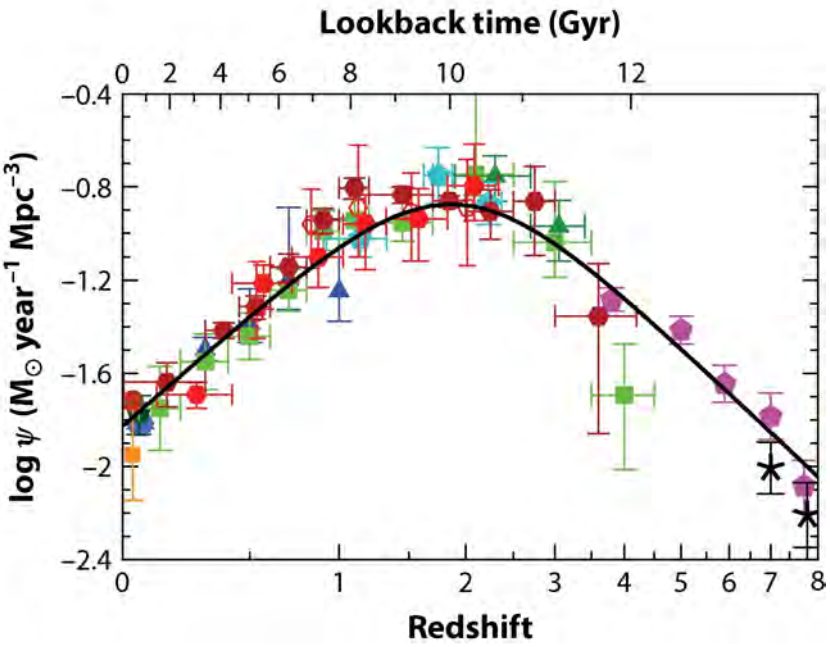


FIGURE 1.13: Cosmic star formation history based on far-UV and infrared rest-frame measurements. Different markers correspond to different determinations of the cosmic star formation rate density. A full description of symbols can be found in Table 1 of Madau and Dickinson (2014), from where the figure was taken.

these are ~ 10 K (Goldsmith, 2001; Papadopoulos, 2010).

1.3.2 *Extragalactic star formation tracers*

Approximately 30% (73 out of 241) of the molecules observed in the Galactic ISM and CSM have also been observed in the extragalactic regime (McGuire, 2022), with the twelve atom C_6H_6 being the biggest molecule observed (Bernard-Salas *et al.*, 2006). Molecular emission from, e.g., CO and H_2O, is observed out to redshift ~ 7 (Jarugula *et al.*, 2021) when the Universe was less than 1 Gyr old.

Due to its high abundance and relatively low critical densities (see Sect. 2.1.1) of the lower-J transitions ($n_{H_2} = 10^2 - 10^3$ cm^{-3}) typically the most easily detectable molecule in high-redshift galaxies is CO (Hodge and da Cunha, 2020). In the Milky Way, CO efficiently forms in the gas phase and its lower-lying transitions can trace entire molecular clouds. It is reasonable to assume that it is also the case elsewhere in the Universe. However, there is no guarantee that the entire cloud material will be turned into stars, as is implied by the star formation efficiency (SFE), describing the amount of molecular gas that ends up in stars. Only the densest parts of the cloud will collapse and form stars, so ideally higher-density tracers are needed ($n_{H_2} > 10^3$ cm^{-3}). At high-redshift readily observed high density tracers are, e.g., HCN, HNC, and HCO^+ (e.g., Solomon *et al.*, 2003; Riechers *et al.*, 2006; Riechers *et al.*, 2011; Oteo *et al.*, 2017). However, their abundances are not accurately known, so it is difficult to turn them into quantitative star-formation tracers.

Water has high critical densities ($n_{H_2} \approx 10^7$ cm^{-3}) and is one of the most abundant molecules in the ISM in the form of ice. Hence, in observations

of high-redshift galaxies, it is typically used as a high-density, rather than a star-formation, tracer (e.g., van der Werf *et al.*, 2011a). It has the strongest molecular emission after CO at submillimeter wavelengths. It has been observed out to redshift ≈ 6.34 (Riechers *et al.*, 2013), with numerous observations toward nearby and distant sub-Luminous InfraRed Galaxies (sub-LIRGs; $L_{\rm IR} < 10^{11}$ L$_\odot$), Luminous InfraRed Galaxies (LIRGs; 10^{11} L$_\odot$ $\leq$ $L_{\rm IR} < 10^{12}$ L$_\odot$), ULIRGs (10^{12} L$_\odot$ $\leq$ $L_{\rm IR} < 10^{13}$ L$_\odot$), HyperLuminous InfraRed Galaxies (HyLIRGs; $L_{\rm IR} \geq 10^{13}$ L$_\odot$), and quasars, both lensed and unlensed (e.g., González-Alfonso *et al.*, 2010; Lis *et al.*, 2011; van der Werf *et al.*, 2011a; Combes *et al.*, 2012; Omont *et al.*, 2013; Riechers *et al.*, 2013; Yang *et al.*, 2013; Yang *et al.*, 2016; Apostolovski *et al.*, 2019; Jarugula *et al.*, 2019), with the first observation being reported already in 1977 (Churchwell *et al.*, 1977). All luminous infrared galaxies, from sub-LIRGs to HyLIRGs are characterized by their FIR luminosities. They are likely very dusty and bulk of their FIR emission is assumed to be generated in starbursts.

At high redshift, water emitting at submillimeter wavelengths (except for the ground-state transitions) is thought to be radiatively excited by the local infrared radiation field (in the 50-200μm range), i.e., through the FIR pumping (e.g., van der Werf *et al.*, 2011a; González-Alfonso *et al.*, 2014). This would imply different excitation conditions than in the Milky Way, where it is primarily excited via collisions in outflows (e.g., Mottram *et al.*, 2014; San José-García *et al.*, 2016). This discrepancy raises many questions about the nature of emission, which is known to trace the earliest stages of star formation in the Galaxy. However, for one of the brightest water transitions, which is also readily observed at high redshift, i.e., the para-H_2O $2_{02} - 1_{11}$ line at 988 GHz (Yang *et al.*, 2013; Yang *et al.*, 2016; van Dishoeck *et al.*, 2021b), the high redshift excitation interpretation has been revisited, suggesting the importance of collisions for observed emission (González-Alfonso *et al.*, 2022).

In the cold, dark Galactic clouds, the water abundance (with respect to H_2) is very low, $\lesssim 10^{-8}$ (e.g., Melnick *et al.*, 2020). Meanwhile, extragalactic water abundances are thought to be $\gtrsim 10^{-7}$ (Liu *et al.*, 2017), which suggests that there must be something else releasing the water into the gas phase. It is possible that this additional mechanism is associated with molecular outflows. Hence, there is a need to join or merge the existing scenarios to understand the origins of water emission across cosmic time and ultimately use emission from this molecule as a quantitative star-formation tracer.

1.4 OPEN QUESTIONS

With observations, we learn more about the tight dependence of how local processes in galaxies influence whole galactic ecosystems. And as star formation plays a pivotal role in the baryon cycle, we need to find ways to couple observations and modeling to decipher how star-forming structures arise from and interact with their natal environment, i.e., the interstellar medium.

Following Sect. 1.2, star-formation is associated with rich chemistry. Observations of this chemistry brought immense progress to our understanding of how stars form and helped us build a large molecular toolbox, which is

now readily used in star-formation-related studies. However, if we want to understand the rules governing star formation, we need to go beyond the Milky Way, where we observe star formation close-up (Sect. 1.3.1). Luckily, the same molecular emission which we observe towards nearby well-resolved low-mass star-forming regions is now routinely observed in regions associated with active star formation at high redshift (Sect. 1.3.2). This poses a question: can we use our toolbox in this distant regime? It seems reasonable to assume that the chemistry should not change when we move from the local to the distant Universe, but is it really true? How can we utilize our astrochemical legacy in the extragalactic regime?

Developing a proxy based on this Galactic toolbox is one of the options. Such a Galactic-based proxy would have to be able to reproduce observed star formation laws while taking into account characteristics of different cosmic epochs (Sect. 1.3.1). However, this poses a question: are these star formation laws universal? Our understanding of the star formation process is focused on individual stars, while most stars form in clusters (Sect. 1.1). Hence, how can we properly resolve star formation and unveil all (proto)stellar populations in clustered environments? The work presented in this thesis was a first attempt to put these puzzles together to construct the proxy for star-formation studies across the Universe.

2

Methods

After describing our current understanding of the star formation process, this chapter describes the observational methods that constitute the present-day astrochemical picture of star formation (Section 2.1 and 2.2). It also explains the modeling tools used in the presented studies, i.e., the galaxy-in-a-box model developed for the purposes of this thesis (Section 2.4) and the Paris-Durham shock code (Section 2.5). Additionally, this chapter contains a description of the Water Emission Database, gathering available water emission data, also developed for the purpose of the work presented in this thesis (Section 2.3).

2.1 MOLECULAR EMISSION

The main ideas utilized in this Ph.D. project emerged based on molecular observations of Galactic star-forming regions and high-redshift galaxies. Hence, it is crucial to understand the origin of what we observe and why it is possible. This section deals with how molecular lines are generated (Sect. 2.1.1), what radiative transfer is and why it is needed (Sect. 2.1.2), giving an example of how water molecules are excited.

2.1.1 *How molecular emission lines are generated*

Energy levels of a molecule are a combination of electronic $E_{\rm elec}$, vibrational $E_{\rm vib}$, and rotational $E_{\rm rot}$, which summed up, give the total energy $E_{\rm tot}$:

$$E_{\rm tot} = E_{\rm elec} + E_{\rm vib} + E_{\rm rot}. \tag{2.1}$$

Different transitions are associated with different frequency/wavelength ranges as can be seen in Figure 2.1. Electronic transitions are a results of electrons being excited from one energy level to a higher energy level. These transitions are typically observed at UV and optical wavelengths. When molecules vibrate, just like the harmonic oscillator, they can go from one vibrational state to another, emitting at near- and mid-infrared wavelengths. Rotational transitions, on the other hand, are associated with rotation of the molecules, i.e., the change of their angular momentum, and emit at submillimeter and millimeter wavelength regime. Different transitions are illustrated in Figure 2.2.

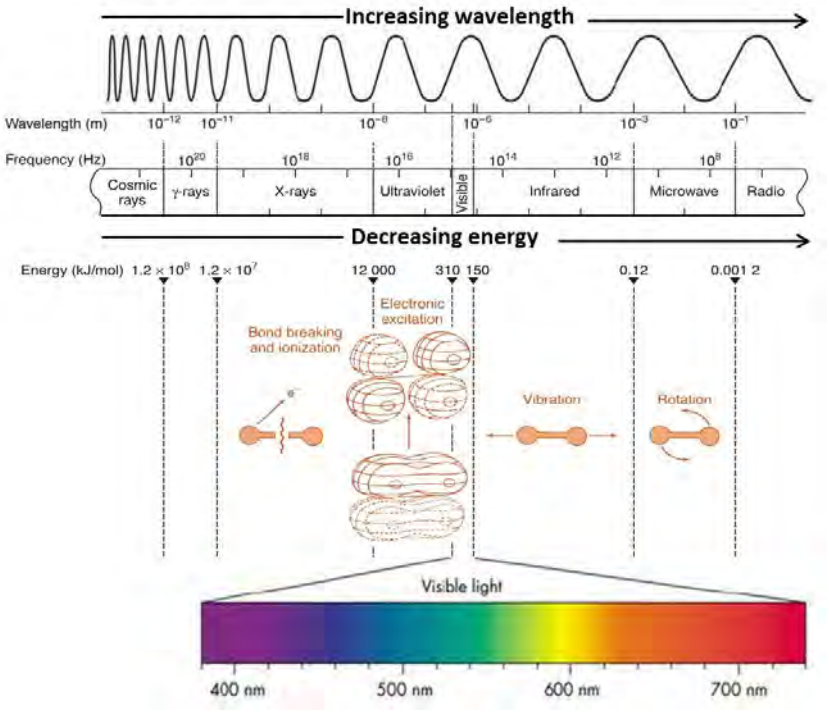


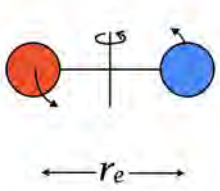
FIGURE 2.1: Schematic representation of expected wavelength, frequency, and energy ranges for electronic, vibrational, and rotational transitions. Electronic transitions are the highest ones on the energy scale, while rotational transitions are the ones requiring the lowest energies. Electronic transitions are expected to emerge at UV and optical wavelengths, vibrational in the infrared regime, and rotational in the microwave.

For a classic rigid rotor, like a diatomic linear molecule, the energy required to populate its rotational levels is:

$$E_{\rm rot}(J) = J(J+1)\left(\frac{h}{2\pi}\right)^2 \frac{1}{2I}, \tag{2.2}$$

where I is the moment of inertia, J is the rotational quantum number, h is Planck's constant. Hence, once we get to high-J levels, the rotational energy goes roughly as J^2. The emitted frequency of a any given transition between $J+1 \rightarrow J$ is proportional to J:

$$\nu = \frac{h(J_{\rm low}+1)}{4\pi^2 I}. \tag{2.3}$$



The moment of inertia is related to the reduced mass of the molecule (μ) and the radial extent of the mass ($r_{\rm e}$), and follows:

$$I = \mu r_{\rm e}^2. \tag{2.4}$$

As illustrated in Figure 2.1, rotational transitions require relatively low energies. Hence, the ISM and star-forming regions exhibit conditions favorable to populate pure rotational levels. Moreover, observations of the pure rotational transitions of molecules are most often (nearly exclusively) conducted with radio astronomy facilities (McGuire, 2022). These transitions are the ones considered in this thesis.

Vibrational and electronic levels require higher energies as can be seen in Figure 2.1. Therefore, observations of emission from vibrational lines typically requires exceptionally warm regions or a radiative pumping mechanism

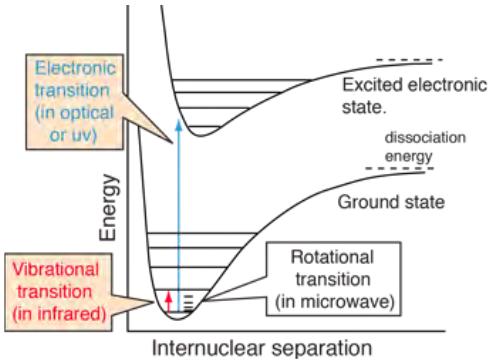


FIGURE 2.2: Schematic representation of energy levels of a molecule with marked electronic, vibrational, and rotational transitions, with indications of expected wavelength ranges.

to populate the levels. Meanwhile, electronic transitions are typically seen only in stellar atmospheres and other extremely energetic environments. However, under these conditions, many molecules can dissociate into their constituent atoms (McGuire, 2022).

2.1.2 *Radiative transfer*

Imagine a two-level system with an upper level energy E_u and a lower level energy E_l. The simplest form of a transition that we can have is a spontaneous transition from the upper to the lower energy level where a photon with a given frequency is emitted and the transition probability for this is given by the Einstein A-coefficient A_ul. Similarly, stimulated emission can occur, and the probability is given by the Einstein B-coefficient B_ul which depends on the intensity of the incoming (stimulating) radiation $I\,B_\mathrm{ul}$. Finally, collisional de-excitation can happen, where collisions induce the transition from the level E_up to E_up, which is described by the collisional rate nC_ul, where n is the density of the collision partner and C_ul is the collisional rate coefficient. Similarly, we may have collisional excitation from the lower energy level to the upper energy level nC_lu, as well as the the absorption of the photon with the right energy denoted by $I\,B_\mathrm{lu}$.

For such a two-level system, the equation of radiative transfer can be expressed through the Einstein A and B coefficients. The energy change involved in each transition depends on the number of molecules in the upper (for emission) and the lower energy level (for excitation), on the corresponding Einstein coefficient, and for stimulated emission and absorption, on the intensity of the incoming radiation. Hence, the energy from all elements can be written as:

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}s} = -\frac{h\nu_0}{c}(n_\mathrm{l}B_\mathrm{lu} - n_\mathrm{u}B_\mathrm{ul})I_\nu\phi(\nu) + \frac{h\nu_0}{4\pi}n_\mathrm{u}A_\mathrm{ul}\phi(\nu), \tag{2.5}$$

where I_ν is the specific intensity of the incoming radiation, $\phi(\nu)$ is the relative probability per unit frequency interval that the photon is emitted, n_u and n_l is the number of molecules in the upper and the lower energy level per unit volume, respectively. The left term involving the Einstein B coefficients takes into account everything that has to do with stimulated emission, hence it accounts also for the absorption, while the right term everything that has to do with spontaneous emission, and they are typically

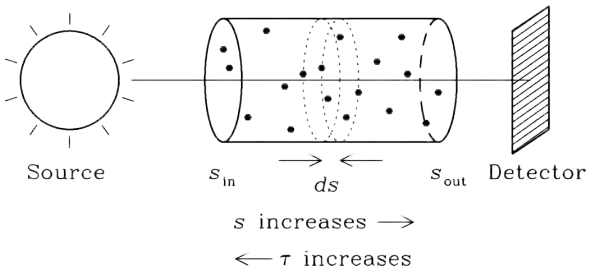


FIGURE 2.3: Schematic representation the light passing a surrounding medium until it reaches the detector. The light enters the medium at s_{in} and propagates along until it reaches the exit point at s_{out}. The optical depth τ increases in the direction of the source. Taken from Condon and Ransom (2016).

denoted as $\alpha_\nu I_\nu$ and j_ν, respectively:

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}s} = -\alpha_\nu I_\nu + j_\nu. \tag{2.6}$$

When light is emitted it passes through the surrounding medium as is illustrated in Figure 2.3. Understanding what will be observed once this happens, i.e., will there be a net gain or net loss in radiation, is the goal of solving the radiative transfer equations. Following Eq. (2.6):

$$\begin{aligned} \mathrm{d}I_{\nu,\mathrm{abs}} &= -\alpha_\nu I_\nu \mathrm{d}s \\ \mathrm{d}I_{\nu,\mathrm{em}} &= j_\nu \mathrm{d}s. \end{aligned} \tag{2.7}$$

From now on α_ν will be expressed as $\rho\kappa_\nu$, where ρ is the density of the medium and κ_ν is the opacity of the medium.

The change in optical depth $\mathrm{d}\tau_\nu$ is defined as $\rho\kappa_\nu \mathrm{d}s$, while a source function describing the incoming emission can be written as $S_\nu = j_\nu/\rho\kappa_\nu$. With that we can derive the typically used form of the radiative transfer equation:

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}\tau_\nu} = -I_\nu + S_\nu. \tag{2.8}$$

For a uniform cloud, we can solve the equation of radiative transfer. In this solution the intensity is equal to the source function modulo the optical depth:

$$I_\nu = S_\nu(1 - \mathrm{e}^{-\tau_\nu}). \tag{2.9}$$

In the Rayleigh–Jeans limit, which is appropriate for low-energy transitions ($h\nu \ll k_\mathrm{B}T$), the intensity depends on the temperature:

$$I_\nu \sim B_\nu(T) = \frac{2kT}{\mathrm{c}^2}\nu^2, \tag{2.10}$$

where $B_\nu(T)$ is the blackbody radiation field at temperature T, which in this case is the brightness temperature T_B. Then, depending on the optical depth we can define the relation between the brightness temperature T_B, i.e., what we see, with the excitation temperature T_ex, i.e., what is being emitted:

$$\begin{aligned} &T_\mathrm{B} = T_\mathrm{ex}(1 - \mathrm{e}^{-\tau_\nu}) \\ &\tau \gg 1 : T_\mathrm{B} = T_\mathrm{ex} \\ &\tau \ll 1 : T_\mathrm{B} = T_\mathrm{ex}\tau_\nu. \end{aligned} \tag{2.11}$$

Hence, for large optical depths, i.e., $\tau \gg 1$, the exponential of $-\tau_\nu$ goes to zero, meaning that in the optically thick case the brightness temperature

will approximate the excitation temperature and we trace the blackbody radiation. Therefore, in the Rayleigh–Jeans limit, for high optical depths we get a handle on the physical temperature. Then, in the optically thin limit, where $\tau \ll 1$, the brightness temperature is equal to the excitation temperature times the optical depth. In this case we get a handle on the column density of the emitting molecule.

Level populations

However, the previous section does not explain the role of collisions, which are important in setting the level populations. When the radiative and collisional de-excitation rates are equal we derive the critical density $n_{\rm crit}$:

$$\begin{aligned} -A_{\rm ul} n_{\rm u} &= -n C_{\rm ul} n_{\rm u} \\ n_{\rm crit} &= \frac{A_{\rm ul}}{C_{\rm ul}}, \end{aligned} \tag{2.12}$$

which differentiate which type of processes dominate in the medium, i.e., if the medium's density is bigger than $n_{\rm crit}$ then the collisional processes dominate. Otherwise, we will have the dominance of spontaneous emission.

Then, assuming a steady-state we can derive the ratio of the upper-level population to the lower-level population:

$$\begin{aligned} \frac{{\rm d}n_{\rm u}}{{\rm d}t} &= n C_{\rm lu} n_{\rm l} - n C_{\rm ul} n_{\rm u} - A_{\rm ul} n_{\rm u} = 0 \\ n_{\rm u}\left(n C_{\rm ul} + A_{\rm ul}\right) &= n C_{\rm lu} n_{\rm l} \\ \frac{n_{\rm u}}{n_{\rm l}} &= \frac{n C_{\rm lu}}{A_{\rm ul}} \frac{1}{1 + \dfrac{n C_{\rm ul}}{A_{\rm ul}}}. \end{aligned} \tag{2.13}$$

The collisional rate coefficients are related to each other such that:

$$C_{\rm lu} = C_{\rm ul} \frac{g_{\rm u}}{g_{\rm l}} \exp\left(\frac{-{\rm h}\nu}{{\rm k}T}\right), \tag{2.14}$$

where $g_{\rm u}$ and $g_{\rm l}$ are state degeneracies of the upper and lower energy level, respectively. We can combine Equations (2.12), (2.13), and (2.14) to derive the relative level populations in the low ($n \ll n_{\rm crit}$) and high ($n \gg n_{\rm crit}$) density limits. Following Eq. (2.12), in the low density limit $n C_{\rm ul} \ll A_{\rm ul}$. Therefore, the level population ratio can be expressed as:

$$\frac{n_{\rm u}}{n_{\rm l}} = \frac{n C_{\rm lu}}{A_{\rm ul}}. \tag{2.15}$$

Thus, the excitation only depends on the density of the collision partner, mostly H_2.

In the high density limit $n C_{\rm ul} \gg A_{\rm ul}$, so $A_{\rm ul}$ will be insignificantly small and:

$$\frac{n_{\rm u}}{n_{\rm l}} = \frac{n C_{\rm lu}}{n C_{\rm ul}} = \frac{C_{\rm lu}}{C_{\rm ul}} \Rightarrow \frac{n_{\rm u}}{n_{\rm l}} = \frac{g_{\rm u}}{g_{\rm l}} \exp\left(\frac{-{\rm h}\nu}{{\rm k}T}\right), \tag{2.16}$$

the level population is given by the Boltzmann distribution and only depend on temperature.

In the optically thin limit and under the assumption of local thermodynamic equilibrium (LTE; corresponding to the collisional regime, where $n \gg n_{\rm crit}$), having measurements of many emission lines from the same

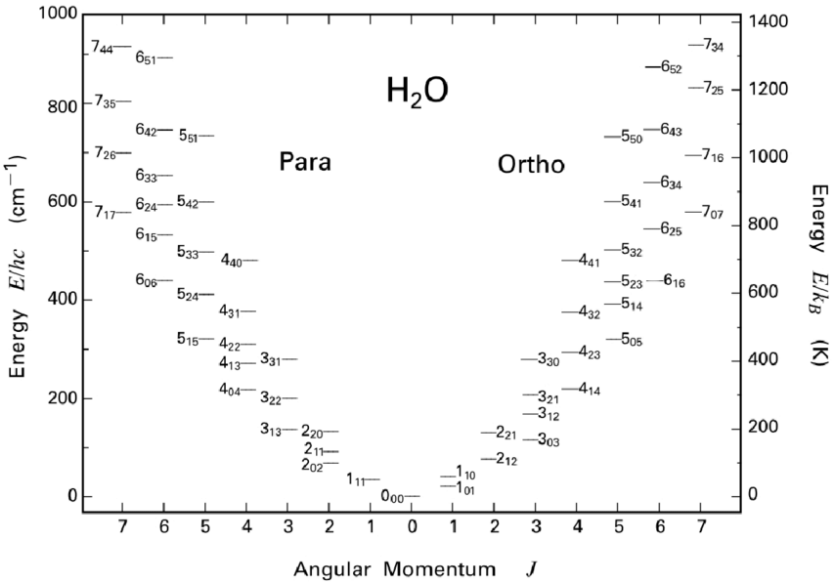


FIGURE 2.4: Para- and ortho-water energy levels structure up to upper level energy temperatures of 1400 K. Taken from Banzatti (2013).

molecule should enable estimation of the excitation temperature and the total column density $N_{\rm tot}$, using the rotational diagram method (e.g., Goldsmith and Langer, 1999). Rotational diagrams present upper level column densities as a function of a upper-level energy and is given by the Boltzmann distribution. A functional form of the fraction of particles can be expressed by:

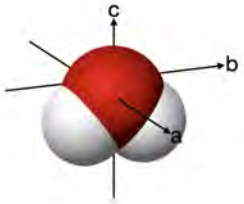
$$\ln\left(\frac{N_{\rm u}}{g_{\rm u}}\right) = -\frac{E_{\rm up}}{T_{\rm ex}} + \ln\left(\frac{N_{\rm tot}}{Q(T_{\rm ex})}\right), \tag{2.17}$$

where $N_{\rm u}$ is the column density of the upper level molecules, $Q(T_{\rm ex})$ is the partition function (the total sum of the upper energy level degeneracies times the exponential of $-E_{\rm up}/kT$). Hence, with a linear fit, $y = aE_{\rm up} + b$, we can derive:

$$\begin{aligned} T_{\rm ex} &= -\frac{1}{a} \\ N_{\rm tot} &= \exp(b)Q(T_{\rm ex}). \end{aligned} \tag{2.18}$$

Solving radiative transfer problems is difficult, as it requires, e.g., to account for the interdependence of the molecular level populations and the local radiation field, requiring iterative solution methods (van der Tak *et al.*, 2007). To overcome these difficulties it is possible to simplify the calculations, e.g., by introducing a geometrically averaged escape probability β, i.e., the probability that a photon will escape the medium from where it was created. One of the widely used approximations is the Sobolev or large velocity gradient (LVG) approximation (e.g., Sobolev, 1960), where the emission at any point in the cloud is completely decoupled from emission at all other points in the cloud. One of the examples of the codes that utilized this approximation is RADEX, a widely used non-LTE radiative transfer code, assuming an isothermal and homogeneous medium without large-scale velocity fields (van der Tak *et al.*, 2007).

Water excitation



Water is an asymmetric rotor, i.e., its principal moments of inertia are different ($I_a \neq I_b \neq I_c$) . Its energy levels are characterized by quantum numbers $J_{K_A K_C}$, where J is the principal rotational quantum number, and K corresponds to the projections of J on the symmetry axes of inertia. Due to its asymmetric nature, water has a large permanent electric dipole moment, which allows for fast dipole transitions with $\Delta J = 0, \pm 1$ and $\Delta K = \pm 1, \pm 3$. Hence, water has an easy-to-populate series of transitions, which are very close in energy. Due to the nuclear spin statistics of its hydrogen atoms, these energy levels are grouped into ortho and para ladders. In ortho-water, the H nuclei have parallel nuclear spins ↑↑ (total spin = 1), and $K_A + K_C$ gives an odd number. The opposite holds for para-water, i.e., spins are antiparallel ↑↓ (total spin = 0) with even $K_A + K_C$ (see Fig. 2.4). These water transitions can be rotational, vibrational, or electronic and observed at far-infrared/submillimeter, infrared, and ultraviolet wavelengths, respectively. Radiative transitions between ortho- and para-water do not exist. Water has high critical densities, e.g., for the ground-state para-line $1_{11} - 0_{00}$ $n_{\rm crit} \approx 10^8\,{\rm cm}^{-3}$, so it requires high densities to be excited.

In past decades, water has been readily observed from space toward Galactic objects. Galactic water observations revealed that outflows dominate water emission in star-forming regions (van Dishoeck *et al.*, 2021b, and references therein). The ground state levels of ortho- and para-water have been found to trace the warm outflowing gas (Mottram *et al.*, 2014), similarly transitions, with $E_{\rm up} \sim 100 - 300$ K. While water transitions with $E_{\rm up} > 300$ K are only populated in high-temperature gas and strong shocks (Herczeg *et al.*, 2012; van Dishoeck *et al.*, 2013). Water, except for the ground state transitions, may also be excited by far-infrared (FIR) pumping, i.e., by absorbing FIR photons leading to excitation to upper-level states from which they cascade toward the lines at which they are observed (González-Alfonso *et al.*, 2014). This is thought to be the dominant mechanism in extragalactic sources. However, in the Galactic sources, water excitation is collisionally dominated, and there are no signs that other processes, such as FIR pumping, play a significant role in the excitation (Mottram *et al.*, 2014; Goicoechea *et al.*, 2015).

2.2 OBSERVATIONAL FACILITIES

Since this thesis focuses on sub-mm/FIR water emission, this section outlines the observational facilities that brought an immense progress into our understanding of its origin in the Universe. It describes single-dish, primarily space-based, observatories in Section 2.2.1 and interferometric arrays in Section 2.2.2.

2.2.1 *Single-dish observatories*

Ground-based observations are limited due to the transparency of the Earth's atmosphere as can be seen in Fig. 2.5. The majority of the infrared light is absorbed by Earth's atmospheric gases, mainly H_2O and CO_2, the optical and radio regimes remain available for observations. Hence, the vast majority

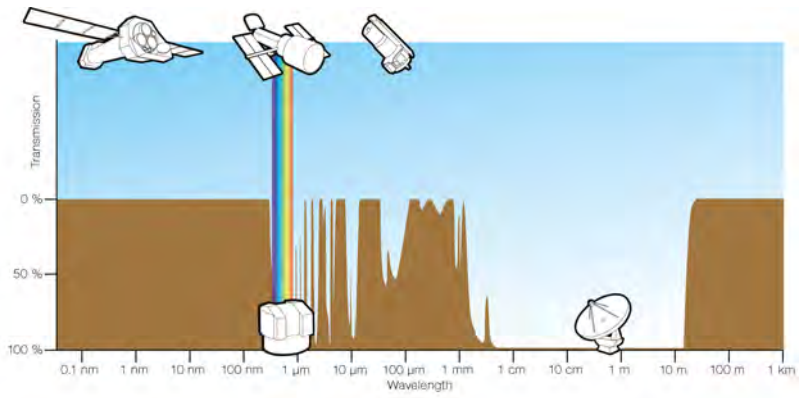


FIGURE 2.5: Cartoon representing the transparency of the Earth's atmosphere. The brown filled curve shows how transparent the atmosphere is at the given wavelength to radiation from space. The major windows are at visible wavelengths (rainbow) and at radio wavelengths from about 1 mm to 10 m. Space telescopes presented in this figure are XMM-Newton, Hubble and the Spitzer Space Telescope, respectively. Credit: ESA/Hubble (F. Granato)

of pure rotational water lines get block by water present in the atmosphere (van Dishoeck *et al.*, 2013).

Galactic observations of water are available through no-longer active space-based observatories. The Infrared Space Observatory (ISO; Kessler *et al.*, 1996) was an infrared space telescope, which covered the 2.5-240 μm range, which operated between 1995 and 1998. It had a modest resolving power $R = \lambda/\Delta\lambda$ ranging from $\sim 200 - 10^4$. Its mid- to far-infrared range provided first insights into rich gas-grain chemistry of water (van Dishoeck, 2004; Cernicharo and Crovisier, 2005). After ISO, the Submilimeter Wave Astronomy Satellite (SWAS; Melnick *et al.*, 2000) launched in 1998 and the Swedish-led satellite Odin (Nordh *et al.*, 2003) launched in 2001, brought progress in the field of water chemistry in space. Both mission were designed to observe the lowest pure rotational $1_{10} - 1_{01}$ ortho-water transition at 557 GHz. Both telescopes had high spectral resolving power with $R > 10^6$ (van Dishoeck *et al.*, 2013). The Spitzer Space Telescope, operative in 2003-2009, observed the $5 - 38\mu$m range, but only with a modest resolving power $R \approx 50 - 600$. It observed highly excited pure rotational lines and solid-state bands of water ice.

However, the greatest progress in water observations were brought with the *Herschel* Space Observatory (Pilbratt *et al.*, 2010), which was launched in 2009 and worked until 2013. With *Herschel* it was possible to observe the $55 - 672\mu$m range. It had three instruments: the heterodyne instrument for the Far-Infrared (HIFI), the Photodetector Array Camera and Spectrometer (PACS), and the Spectral and Photometric Imaging REceiver (SPIRE). From the three instruments, HIFI presented an impressive resolution with R up to 10^7. With its unprecedented sensitivity as well as spatial and spectral resolution, *Herschel* enabled studies of water in the Milky Way. Especially, with the Water In Star-forming regions with *Herschel* (WISH; van Dishoeck *et al.*, 2011; van Dishoeck *et al.*, 2021b) program, which targeted $\sim$ 80 protostellar sources with different luminosities and evolutionary stages.

Typical spectrometer technologies used to obtain high-resolution observations are heterodyne, e.g., HIFI, and grism, e.g., NIRSpec onboard the

FIGURE 2.6: Artistic impressions of the space-based observatories and photos of grand-based sub-mm arrays that have increased our understanding of water chemistry in the Universe. Credits for space observatories: ISO (ESA), ODIN (ESA), Spitzer (NASA/JPL-Caltech), SWAS (NASA), *Herschel* (ESA; background: Hubble Space Telescope, NASA/ ESA/ STScI). Credits for sub-mm arrays: NOEMA (Jérémie Boissier/IRAM/CNRS Photothéque).

James Webb Space Telescope (Greene *et al.*, 2017). Light detection that uses heterodyne spectrometers involves translating the frequency range of the astronomical signal that is being observed to a lower frequency where it is easier to perform the required measurements. This is done by mixing the incoming signal with a very stable monochromatic signal, generated by a local oscillator, and extracting the difference frequency for further processing. This process amplifies the signal, making hetorodynes a widely used spectrometers by nearly all practical radiometers (Condon and Ransom, 2016). Meanwhile, a grism spectrometer is a combination of a diffraction GRating that disperses the observed light and the prISM, which redirects the light back into the camera. In practice it makes grisms multi-object spectrometers, which produce spectra of many objects simultaneously.

2.2.2 *Interferometers*

Single-dish radio telescopes have relatively low angular resolution and pointing accuracy, small field-of-view, and limited sensitivity (Condon and Ransom, 2016). This can be easily tested with the formula for the angular resolution of an antenna θ:

$$\theta \approx \frac{\lambda}{D}, \tag{2.19}$$

where λ is the wavelength at which we want to observe, and D is the diameter of the antenna (Rayleigh, 1879). Hence, assuming that observations of Galactic water emission at FIR/submillimeter wavelengths would be possible from the ground, and we would like to observe water emission at $\sim 303\mu$m (corresponding to the 2_{02}-1_{11} para-transition) with a resolution of $0\farcs1$,

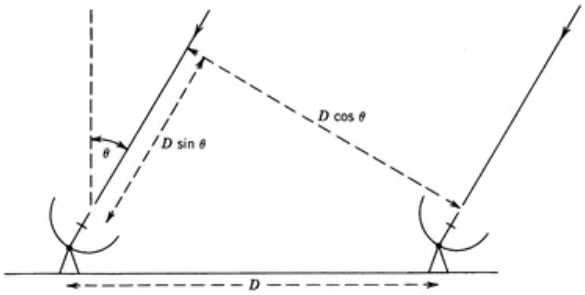


FIGURE 2.7: Schematic geometry of a two-antenna elementary interferometer with a baseline D. Antennas observing the source in the direction θ will receive the signal at different times. In this scenario, the left antenna is the subject of the geometrical time delay.

the single dish antenna diameter would have to have at least ≈ 624 m in diameter.

The limitations of single-dish antennas can be overcome by joining them together into an interferometer, a technique that reaches back to 1890 (Michelson, 1890). The very first interferometer was optical, but then in 1946 Ryle and Vonberg constructed a two-element radio interferometer to investigate cosmic radio emission discovered by Jansky in 1933 (Jansky, 1933; Thompson *et al.*, 2017). Radio-interferometry combines the collected signal from individual antennas to one high-resolution output. Antennas form an array, and the resulting angular resolution of this array is inversely proportional to the largest distance between two antennas:

$$\theta \approx \frac{\lambda}{D_{\max}}, \tag{2.20}$$

The easiest way to demonstrate the basics of interferometry is by considering a two-antenna system. These two antennas simultaneously observe the same field of view. Assuming that the signal does not reach the antennas at the same time and that the source is located in direction θ, the signal will reach one of the antennas with the so-called geometric delay τ_g (see Fig. 2.7, Thompson *et al.*, 2017):

$$\tau_g = \frac{D}{c}\sin\theta, \tag{2.21}$$

where D is the interferometer baseline, i.e., distance between the two antennas, c is the light velocity. The signals received by the antennas reach the correlator, where the signal voltages are multiplied and then time-averaged (this is when they are corrected for τ_g). The output of the correlator called the interferometer response $r(\tau_g)$ is then integrated over time, producing the sky visibility $V(u, v)$, where u and v are the coordinates of the plane describing the position of the baseline, commonly known as the uv plane.

Following the van Cittert-Zernike theorem (van Cittert, 1934; Zernike, 1938), an image of the intensity of the source can be achieved by inverting the Fourier transform of the visibility. Hence, the visibility is defined as the Fourier transform of the specific intensity of the source on the sky $I(x, y)$:

$$V(u, v) = \iint A(x, y)I(x, y)e^{-2\pi i(ux+vy)}\,\mathrm{d}x\mathrm{d}y, \tag{2.22}$$

where $A(x, y)$ is the effective collecting area of the of the antenna. Then, through a Fourier inversion, the modified specific intensity $I'(x, \mathrm{y})$ can be

found:

$$I'(x,y) = A(x,y)I(x,y) = \iint V(u,v)e^{2\pi i(ux+vy)}\,dudv. \tag{2.23}$$

In practice, for interferometers with higher number of antennas, the more the baselines, the more visibilities can retrieved, i.e., more information on the intensity of the source will be obtained. Moreover, longer integration times result in a better coverage of the uv plane due to the rotation of the Earth.

Submilimeter interferometers are readily used to observe water emission in the distant Universe, since its observed frequency gets redshifted and changes as:

$$\nu_{\rm obs} = \nu_{\rm lab}(1+z)^{-1}. \tag{2.24}$$

Hence, e.g., at $z \sim 2$, emission from the $2_{02} - 1_{11}$ line at 988 GHz shifts to 329 GHz. This shift enables ground-based observations of water at cosmic noon and farther.

A prime example of a sub-mm interferometer is the Atacama Large Millimeter/submillimeter Array (ALMA). It consists of a large array of fifty 12-m antennas with baselines up to 16 km and an additional compact array of twelve 7-m and four 12-m antennas. These arrays are all located on the Chajnantor Plateau in northern Chile at 5000 m. Currently, ALMA have eight bands covering frequencies from 84 to 950 GHz. ALMA has uncovered a whole new molecular Universe (e.g., Hodge and da Cunha, 2020) and enabled many new water observations in the distant Universe, including the farthest one (e.g., Jarugula *et al.*, 2021).

Another example of a sub-mm interferometer readily observing water in distant Universe is the Northern Extended Millimeter Array (previously known as the Plateau de Bure Interferometer). Covering the 80 to 370 GHz range, NOEMA consists of an array of 12 individual 15-meter antennas, located on the Plateau de Bure in the French Alps at 2550 meters above sea level. It can achieve a resolving power equivalent to that of a single-dish telescope with a diameter of 1.7 kilometers. With NOEMA numerous galaxies at $z \sim 2-4$ were observed (e.g., Omont *et al.*, 2013; Yang *et al.*, 2013).

2.3 WATER EMISSION DATABASE

Water has been observed towards many source via different observational program. Thus, as a part of this thesis, the Water Emission Database (WED) was created with the goal to gather these observations in one place, from ground-based observatories and the *Herschel* Space Observatory. Currently, WED covers 79 observed water transitions up to the para-H_2O $9_{19} - 8_{08}$ transition at 5280.73 GHz (56.77 μm). However, since its development was dictated by the projects presented in this thesis, WED remains under development. Ultimately, the database will include all observations of water, including the extragalactic ones.

2.3.1 *Observations of Galactic water emission*

Among available water observations, most data comes from Herschel observational programs, e.g., Water In Star-forming regions with *Herschel* (WISH; van Dishoeck *et al.*, 2011); the Dust, Ice, and Gas in Time (DIGIT; Green *et al.*, 2013); and the William Herschel Line Legacy Survey (WILL; Mottram *et al.*, 2017). Using Herschel, these programs managed to cover a whole range of protostellar sources, providing key insights into water chemistry in star-forming regions.

The WISH program predominantly relied on the HIFI and PACS instruments onboard *Herschel*. It was a guaranteed time Key Program, which aimed to use water vapor as physical and chemical diagnostics and follow the water abundance throughout the different phases of star- and planet-formation (van Dishoeck *et al.*, 2021b). WISH observed ~ 80 YSOs. Bolometric luminosities of observed objects $1 L_\odot < L_{\rm bol} < 10^5 L_\odot$, including low-, intermediate-, and high-mass protostars. These protostars were at different evolutionary stages including the stage prior to collapse to the dissipation of envelopes. It observed low- and high-excitation lines of water and its oxygen-related isotopologues, i.e., $H_2^{17}O$ and $H_2^{18}O$, chemically related species O and OH. It also targeted some high-J CO lines and its isotopologues, hydrides, C^+, and HDO (van Dishoeck *et al.*, 2021b).

There have been many subprograms of WISH, which together with the main program are known as WISH+ program. These programs include, among many others, WILL and a Cygnus program targeting intermediate- and high-mass YSOs (PI: S. Bontemps; San José-García, 2015). The Cygnus sample covered 86 sources, from which 8 could be considered as equivalent to Class 0 protostars, while the remaining 78 are more evolved objects (Class I and II type with no radio emission). Available observed fluxes include para-H_2O $1_{11} - 0_{00}$ and $2_{02} - 1_{11}$ lines, and CO isotopologues transitions, i.e, ^{13}CO= 10 – 9 and C^{18}O= 9 – 8 transitions (San José-García, 2015). Meanwhile, WILL covered 49 objects (van Dishoeck *et al.*, 2021b), focusing on low-mass star forming YSOs, covering transitions of H_2O, its isotopologues, as well as CO and its isotopologues.

There have been other programs that also observed water in star-forming regions. These include, e.g., DIGIT observing 30 low-mass sources with PACS, the CHEmical Survey of Star-forming regions (CHESS; Ceccarelli *et al.*, 2010) aimed at understanding the relation between the chemistry with the protostellar properties; Herschel observations of EXtraOrdinary Sources (HEXOS; Bergin *et al.*, 2010) focused on observations of Orion KL, Sgr B2 (N), Sgr B2 (M), Orion S, and the Orion Bar. Each observational program was dedicated to deepening our understanding of different aspects of (water) chemistry.

2.3.2 *Design of the database*

The database is stored and maintained using the MySQL Database Service. The choice of using the Structured Query Language (SQL), which is a programming language designed for accessing, modifying and extracting information from databases. It is commonly used in astronomy, e.g., by the Centre de Données astronomiques de Strasbourg (CDS), where typically

TABLE 2.1: Description of WED table columns

Column	Description
obs_id	Ordinal number of the input
object	Name of the object
obj_type[a]	Emitting object type
ra_2000	RA (J2000)
dec_2000	Dec (J2000)
transition	Observed water transition
freq[b]	Rest frequency of the observed transition
telescope	Name of the telescope used in the observations
instrument[c]	Instrument used in the observations
obs_res	Resolution ($''$)
distance	Distance to the object (pc)
luminosity	Bolometric luminosity ($L_\odot$)
tbol[c]	Bolometric temperature (K)
menv[c]	Envelope mass ($M_\odot$)
vlsr[c]	Velocity (km s^{-1})
flux	Observed water flux
flux_err[c]	Flux error
unit	Unit of the observed flux (K km s^{-1}; W cm^{-2}; W m^{-2}; erg $s^{-1}cm^{-2}$)
ref[d]	Reference to the flux measurement(s)
extra	Other relevant information

Notes: (a) Object types currently in use: YSO - young stellar object, IM - intermediate-mass, LM - low-mass, IR-q - IR-quiet, HM - high-mass, mIR-q - mIR-quiet, HMPO - high-mass protostellar object, HMC - hot molecular core, UCHII - ultra-compact HII region, C0 - Class 0, CI - Class I, CII - Class II, PS - possible pre-stellar core, PDR - photodissociation region. Classification is based on the source papers; (b) All of the frequencies to corresponding transitions are taken from the LAMDA database (Schöier *et al.*, 2005); (c) When available; (d) If more than one flux measurement is available, then the most recent or commonly used one is provided with the references to the remaining ones being stored in this column.

astronomical data sets are published. Its possibilities make it also a favorable language in data science, being used by the biggest companies in the world, e.g., Amazon. In the case of this kind of growing and developing database, it was a natural choice to maintain it locally, in contrary to uploading the data to the CDS, where it could not be altered. However, currently the access to the data is granted through ASCII and CSV files available online on the project website[1]. The ultimate goal for the database is to make an interactive tool for retrieving information on water observations.

[1] katarzynadutkowska.github.io/WED/

The stored values focus on water fluxes and parameters describing source properties, keeping a unified system of parameters important to characterize the emitting sources. This will help to plan for potential future follow-up observations with next FIR probe. Since the idea for the database was a result of the projects presented in this thesis, some of the properties were

motivated by the needs of the galaxy-in-a-box model, e.g, envelope masses. Currently, WED holds the data in tables arranged in 20 columns, which are described in Table 2.1.

Currently, WED is primarily a composite of observations from WISH+ and CHESS, with the highest lying transitions being those with $J_{K_A K_C} = 9_{K_A K_C} - 8_{K_A K_C}$ observed with PACS towards NGC 1333 IRAS 4B (Herczeg *et al.*, 2012). All classification is based on the papers cited in the references column. Moreover, important information are also saved in the final column, e.g., alternative names, reported possibilities of blending, specific component observed if the flux was not reported for the whole source, whether a detection was an upper limit. Transitions are reported in a unified format following the Leiden Atomic and Molecular Database[2] (Schöier *et al.*, 2005). Moreover, all elements in the database are Python friendly, such that they could be easily used for computations and downloaded straight to the code, like is done for the case of the galaxy-in-a-box model. Currently, the official version of the database consists of ~ 600 fully described entries, while the full workbench stands at more than 10^3 entries.

[2] https://home.strw.leidenuniv.nl/~moldata/

2.4 THE GALAXY-IN-A-BOX MODEL

On galactic scales, star formation predominantly takes place in clusters that form from GMCs. These GMCs form complexes that follow certain spatial distribution in galaxies. Hence, following the bottom-up approach, a model of galactic emission from active star formation can be built starting from these star-forming clusters.

2.4.1 *The cluster-in-a-box model*

The building block of the galaxy-in-a-box model involving the star-forming clusters is based on the cluster-in-a-box model by Kristensen and Bergin (2015). This model simulates young star-forming clusters with the protostellar outflow emission assigned to every embedded low- and intermediate-mass protostar. The cluster in the cluster-in-a-box model can be characterized by spatial, age, mass, and outflow angular distributions.

The radial extent of the cluster is built on a power-law function from Adams *et al.* (2014):

$$R_{\max} = R_0 (N/N_0)^{\alpha_c}, \tag{2.25}$$

where N is the number of stars in the cluster and the power-law slope $\alpha_c = 1/3$. Then, the spatial distribution of stars within the cluster is given by radial probability distribution function (Adams *et al.*, 2006; Adams *et al.*, 2014):

$$\frac{\mathrm{d}P}{\mathrm{d}r} = \frac{3-p}{R_{\max}} \left(\frac{r}{R_{\max}} \right)^{2-p}, \tag{2.26}$$

where p is the power-law index of the density distribution, which in the case of the cluster-in-a-box is set to 1.5. Typically higher-mass stars are located closer to the bottom of the gravitational potential well, i.e., closer to the cluster center (e.g., Hillenbrand and Hartmann, 1998). Therefore, once the spatial distribution is calculated following Eq. (2.25) and (2.26), the cluster-in-a-box re-scales the distance of protostars to the center by $M^{-0.1}$.

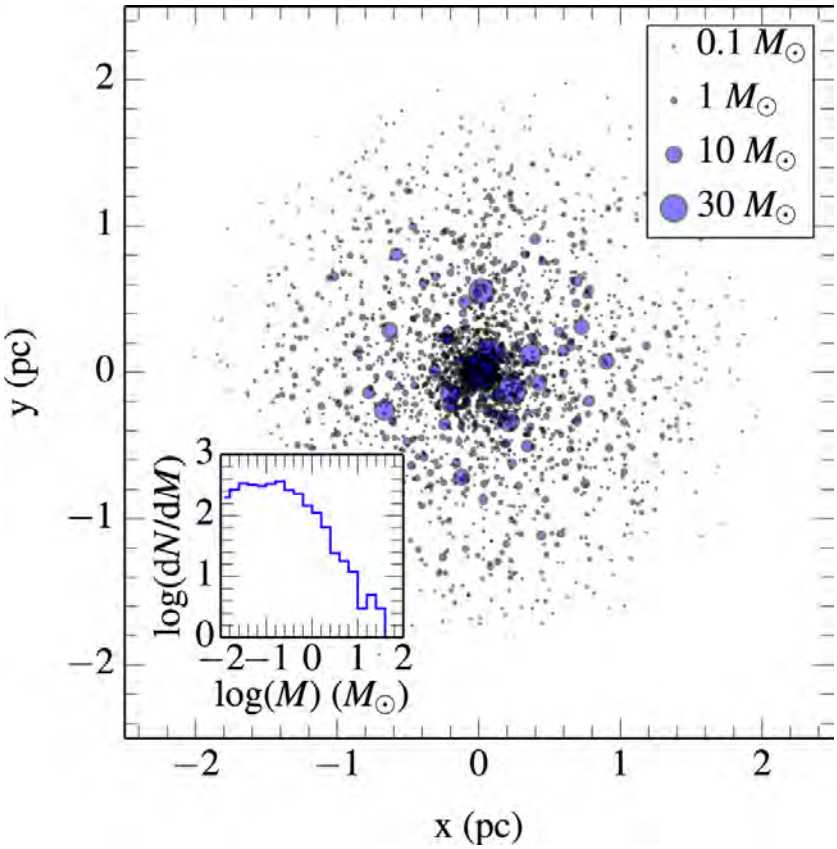


FIGURE 2.8: Spatial distribution of protostars in the cluster-in-a-box model for a cluster containing 3000 stars. The symbol sizes are scaled to the mass of each (proto)star. The inset represent the final mass distribution. Taken from Kristensen and Bergin (2015).

This power-law scaling ensures that the most massive members of the cluster are located within 10%-20% of the maximum cluster radius, whereas the lower-mass members are hardly affected (Kristensen and Bergin, 2015).

The age distribution in the cluster-in-a-box model follows that of the low-mass Perseus star-forming cluster, which has an estimated age of ~ 1 Myr (Evans *et al.*, 2009). The fraction of protostars in the Class 0 was taken from Sadavoy *et al.* (2014), while I, II, and III stages as well as the "flat-spectrum" stage between Class I and II, from Evans *et al.* (2009).

The IMF in the model follows that of Chabrier (2003) for young clusters and disk stars, which is provided in Equation (1.10). The mass distribution is randomly sampled over the mass interval from 0.01 to 100 $M_\odot$. These masses can then be re-scaled to correspond to the protostellar stage. Class 0 cores typically have masses that are three times higher than the final stellar mass (e.g., André *et al.*, 2010, and references therein), while Class I cores have envelope masses 1.5 times as high as the final stellar mass. The masses of protostars in the remaining evolutionary classes are typically close to their final mass (Kristensen and Bergin, 2015). However, as mentioned in Section 1.1.2, Class II and III sources exhibit negligible molecular outflow activity, hence, these are not included in the model.

Then, for the outflow angular distribution, an outflow position angle is randomly set between 0° – 180°. Then, each outflow lobe is given a random distance from the protostar to simulate differences in inclination angles and

dynamical ages, where the maximum simulated separation corresponds to the maximum separation in the observed data, i.e., 2×10^4 AU (Kristensen and Bergin, 2015).

The cluster-in-a-box simulate molecular emission from the methanol $7_0 - 6_0$ A^+ line at 338 GHz, empirically known to be a great outflow tracer (Kristensen *et al.*, 2010a). This emission assignment is based on an observed relation between the observed intensity and envelope masses of protostars, using the re-scaled stellar masses as described above. Once the emission is assigned, the resulting image is convolved with a Gaussian beam to simulate observational-like output.

2.4.2 *Extrapolating clustered star-formation*

Star formation predominantly occurs in clustered environments as outlined in Section 1.1.3. Hence, in the bottom-up approach, the galactic star formation can be built up by adding up many independent, small star-forming regions (Krumholz, 2015), which in this case would be represented by star-forming clusters. Following this, by extrapolating the cluster-in-a-box model, it should be possible to build a distribution of all galactic clusters with ongoing star formation and simulate molecular emission associated with active star formation in a whole galaxy.

The result of the scaling of the cluster-in-a-box model resulted in the development of the galaxy-in-a-box model, which consists of a template (spiral) galaxy with molecular cloud spatial, age, and mass distributions, and of template stellar clusters with assigned outflow emission based on the cluster-in-a-box model (Dutkowska and Kristensen, 2022). This model is the basis of two projects presented in this thesis in Chapter 4 and 5.

As already mentioned, the galaxy-in-a-box model utilizes the cluster-in-a-box model. However, it does not do it straightforwardly, but rather adapts it to best serve the galactic scale modeling. First, for the galaxy-in-a-box water is chosen to be the default molecule for the simulation of galactic emission for the reasons outlined in Section 1.2.4. For the purpose of this model, the emission of the *para*-H_2O $2_{02} - 1_{11}$ transition at 987.927 GHz ($E_{\rm up}$ = 100.8 K) was chosen, as it is among the brightest H_2O transitions observed toward Galactic star-forming regions. Moreover, it is not a ground-state transition, and so it only mildly suffers from self-absorption even toward high-mass objects (van der Tak *et al.*, 2013). Finally, this transition is routinely observed toward extragalactic and even high-z objects (e.g., Yang *et al.*, 2016; Jarugula *et al.*, 2019). Another major change was introduced to the spatial configuration of the cluster model. At a distance of several Mpc and farther, the structure of individual clusters is practically unresolvable. Hence, the spatial component for the galactic model was discarded. Also, the protostellar age distribution was updated with a novel distribution from Kristensen and Dunham (2018).

The architecture of the galaxy-in-a-box model is rooted in three elements of the galactic star-forming environment, i.e., GMCs, clusters, and protostars. First, a spatial and mass distribution of GMCs is generated based on the observational data of other galaxies. The GMCs that will be passed to the next steps of the simulation are chosen randomly from the mass distribution

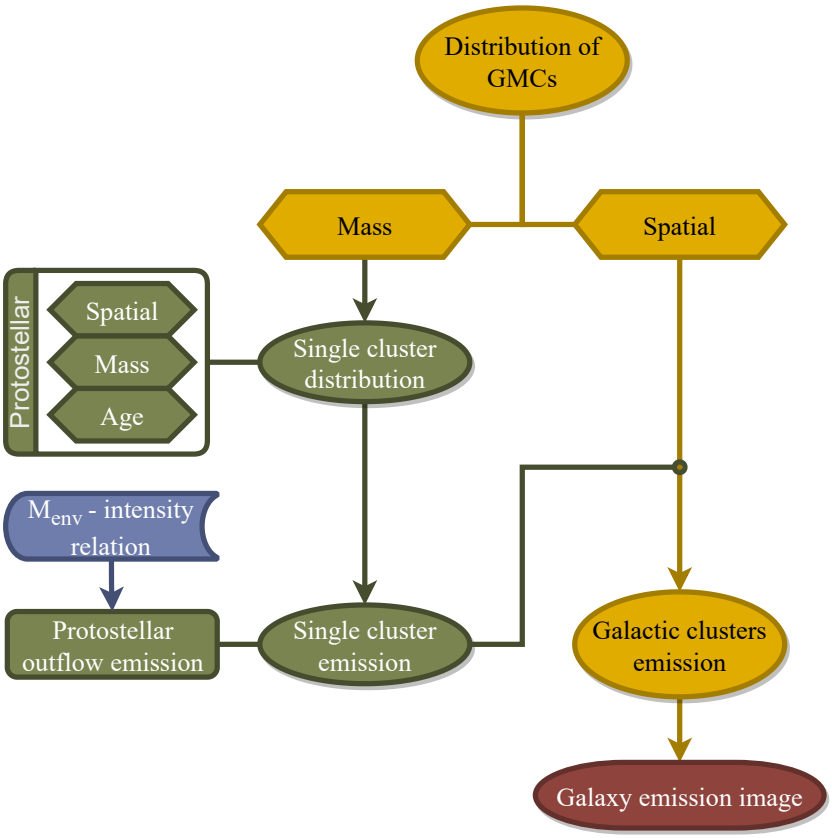


FIGURE 2.9: Schematic flowchart of the galaxy-in-a-box model. Yellow part of the flowchart corresponds to the galactic part of the model, green to the cluster model, red to the final stage, and blue to the external data input from WED. Figure taken from Dutkowska and Kristensen (2022).

until the available galactic molecular reservoir, or the total number of GMCs, is reached. The molecular reservoir can be specified directly or via the far-infrared luminosity, which is recalculated to mass based on the observed galactic virial mass-luminosity relation (Scoville and Good, 1989). Then, each GMC mass acts as an initial cluster mass. Before being passed to the cluster model, each cloud is assigned an age based on its free-fall time, which is then then randomly scaled between being newly formed and completely collapsed. Then, the cluster model returns protostellar mass, age, and spatial distribution. However, as already mentioned, the latter is disregarded in the calculations.

With the complete protostellar distribution, molecular emission is assigned to each Class 0 and Class I protostar. This action results in providing the total expected emission from a given cluster. If there are no introduced constraints on the galactic star formation rate, the model runs until this step is repeated for all GMCs. Otherwise, the model cuts the calculations once the total allowed galactic SFR is reached. Then, the information about total cluster mass (expressed as the total mass of (proto)stellar content) and emission is returned to the galactic spatial grid. After accounting for the sizes of clusters and their location, the raw galactic emission image is convolved with a Gaussian beam. Hence, the model returns statistics on galactic clusters (their number of stars, mass, emission, and star formation rates) and integrated intensity image. However, the level of the details in

the returned statistics can be easily adjusted. Therefore, the exact protostellar distributions of each cluster can be saved. The exact formulas and distributions used in the galaxy-in-a-box model are described in Chapter 4, while the treatment of the star formation rates and the FIR-luminosity–mass relation in Chapter 5.

2.4.3 *Cautionary tales about the model*

The galaxy-in-a-box is a work in progress. Commonly used modeling tools are typically developed in big collaborations over many years, e.g., IllustrisTNG (Weinberger *et al.*, 2017; Pillepich *et al.*, 2018), the Paris-Durham shock code (Flower and Pineau des Forêts, 2013; Godard *et al.*, 2019, `https://ism.obspm.fr`), or Starburst99 (Leitherer *et al.*, 1999). The projects and stage of the code presented in this thesis are at a stage of the proof-of-concept studies, as more sophisticated solutions need to be applied. The general caveats of the model are outlined below, while the possible future extensions are discussed in greater detail in Chapter 3.5.1.

Since the galaxy-in-a-box is aimed to serve as a modeling tool both for observers and theoreticians that could be used to analyze and/or plan observations, it first needs to follow more realistic galactic morphologies to achieve the mock observations level. This includes the distribution of dense molecular gas in galaxies. Currently, the model follows relatively simple recipes for how the GMCs are distributed in galaxies. It may work to a first approximation, but galactic morphologies are far more complicated. Currently, the easiest to set spatial setup is that for spiral galaxies. However, there are more galactic types, and the physical features change as we go back in time.

Certain approximations have also been introduced for the star-forming clusters themselves. First, GMCs are expected to form more than one star-forming cluster (Kennicutt and Evans, 2012). In the presented work, a simple scenario is being followed in which one GMC results in one massive cluster. Moreover, the internal structure of the clusters is rather simplified. Currently, going straight from the IMF, no possible internal interactions are taken into account. Even though this is a top-level model, i.e., it follows observational laws and scalings, the complex nature of clusters could result in changes in molecular emission and abundance. Especially considering the presence of massive (proto)stars.

When calculating star formation rates, the model does not take into account possible star formation histories of a given galaxy. Moreover, it is possible that because of the approximations done on the way, the model may not be correctly assessing the full extent of the ongoing star formation. However, since the distributions applied in the model are randomized, hence it is possible to estimate a possible spread of values expected for a given galaxy and star-forming parameters.

Despite the impressive data collected by *Herschel*, observations of Galactic high-mass young stellar objects are limited due to both their number and sensitivity. Even the most detailed water observations conducted toward the Cygnus X sources (San José-García, 2015) do not recover the total emission from a high-mass star-forming complex. This is due to the spatial resolution

and sensitivity limitations of the HIFI instrument, as the surveys consisted of single-pointing observations. Therefore, it is possible that already at the input data stage, the expected emission can be underestimated. This is further assessed in Chapter 5.

Other caveats come with the differences between the local and the distant Universe, as outlined in Section 1.3.1. For example, the cosmic microwave background (CMB) temperature changes with redshift, and while at $z = 0$ $T_{\rm CMB} = 2.72548 \pm 0.00057$ (Fixsen, 2009), at $z \approx 6$ it can be in the range of 16.4 – 30.2 K (Riechers *et al.*, 2022). Besides, these younger galaxies will exhibit higher gas surface densities (e.g., Carilli and Walter, 2013), ionization rates (Papadopoulos, 2010), there will be more galaxy-galaxy interactions (e.g., Conselice, 2014), and active galactic nuclei (AGN) may contribute to the quenching of star formation (e.g., King and Pounds, 2015). These effects are currently not included in the model.

Lastly, the galaxy-in-a-box model is written in Python and adapted to its $>$ 3.x versions. In its current version, the model utilizes several techniques to improve the performance of the calculations, like, for example, only working with pickle files throughout. However, a major computational downside is that the code is not parallelized. Improving this aspect of the code in the future will increase the available parameter space for calculations.

2.5 THE PARIS-DURHAM SHOCK CODE

The Paris-Durham shock code (Flower and Pineau des Forêts, 2013; Godard *et al.*, 2019, `https://ism.obspm.fr`) calculates a time-dependent dynamical, chemical, and thermal evolution of matter in 1D stationary shocks using the plane-parallel geometry. The code has undergone numerous changes during its more than 35 years history (Flower *et al.*, 1985). Currently its standard chemical network includes $\sim$ 1000 chemical reactions involving $\sim$140 chemical species. The code is designed for for interstellar molecular shocks driven by super-alfvénic turbulence and cloud collisions, jets and outflows from forming stars, slow AGB winds, and old supernova remnant shells. Shocks in the Paris-Durham shock code can be externally irradiated by a UV radiation field.

2.5.1 *The architecture of the shock model*

The computations start with calculating the thermo-chemical pre-shock conditions in *a chemical steady-state model* as illustrated in Figure 2.10 for a given initial proton density ($n_{\rm H} = n({\rm H}) + 2n({\rm H_2})$), radiation field ($G_0$), the cosmic-ray ionization rate ($\zeta_{\rm H_2}$) and PAH abundance ($X({\rm PAH})$), and visual extinction $A_{\rm V}$ of 10^{-9}. This computations are run for long enough that equilibrium is reached. If an irradiated shock is considered, i.e., an isotropic flux of UV photons if $G_0 > 0$, the next step is *a PDR model*. In this PDR model, a tracer particle with an advection speed $\ll 1\,{\rm km\,s^{-1}}$ crosses the PDR front. The chemical and physical conditions are integrated in parallel, which in case of the model used in this thesis, means going from an $A_{\rm V}$ of 10^{-9} to 10^{-1}. Then, the final PDR model results are used as the input parameters of *the shock code*, where the dynamical, chemical, and thermal evolution of

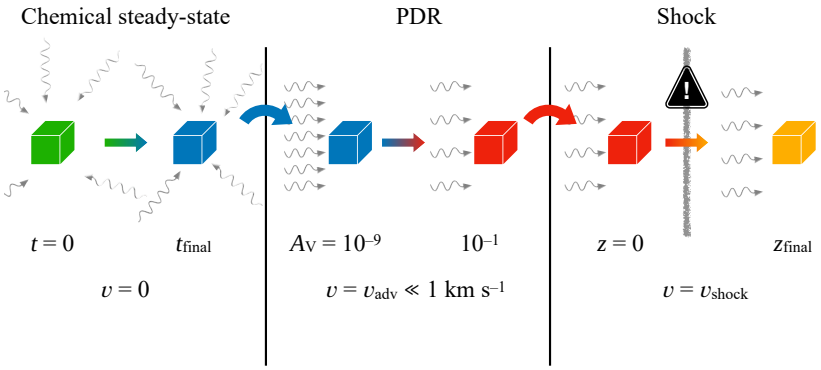


FIGURE 2.10: Schematic representation of the Paris-Durham shock code in the plane-parallel geometry. Computations start with a chemical steady-state model, which calculates the thermal, chemical, and excitation for given initial conditions until it reaches an equilibrium. In case of externally irradiated shocks this output is used as input for a PDR model, integrating over a distance set by visual extinction starting from A_V of 10^{-9} to 10^{-1}. The output of the PDR model is then used as input for the shock model. Taken from Kristensen et al. (in prep).

matter in the shock itself is calculated. *The post-shock layer*, marking the end of the calculations, is reached once the medium is in a steady-state.

Hence, several assumptions are made in the code, e.g., that the calculations are performed for a steady-state ($\partial/\partial t = 0$) dynamical, thermal, and chemical structure of a plane-parallel shock wave, which propagates in a molecular and dusty homogeneous medium with a uniform magnetic field, which is transverse to the propagating shock. All possible shock types for a medium with a transverse magnetic field are available, i.e., the J-type, C-type, C* (a C-type shock, where the neutral fluid becomes subsonic along its trajectory), and CJ-type shocks. However, the differentiation between the C and C* type shocks is for semantic and numerical reasons only (Godard *et al.*, 2019).

The model has been tested and found to produce valid result over the following ranges: the pre-shock proton densities of $10^{-1}\,\mathrm{cm}^{-3} \leq n_\mathrm{H} \leq 10^{8}\,\mathrm{cm}^{-3}$, interstellar radiation field $0 \leq G_0 \leq 10^4$ (here G_0 is the scaling factor of the standard UV radiation field of Mathis *et al.* (1983), the cosmic H_2 ionization rate $10^{-17}\,\mathrm{s}^{-1} \leq \zeta_{H_2} \leq 10^{-15}\,\mathrm{s}^{-1}$. Moreover, the code presented in this thesis does not treat the transfer of Lyman photons, and only considers singly ionized species (this was implemented in version by Lehmann *et al.* (2020)). Since in J-type shocks the maximum kinetic temperature depends on shock velocity (see Sect. 1.2.3), J-shock velocities should be restricted to $V_S \leq 30\,\mathrm{km\,s}^{-1}$, because above this limit temperatures are high enough to get full ionization in the shock the treatment of Lyman photons needs to be included.

2.5.2 *Chemistry in the Paris-Durham shock code*

The shock code accounts for various cooling and heating processes. The cooling mechanisms include, e.g., H_2 line cooling, rotational and vibrational radiative cooling by CO, H_2O, and ^{13}CO, pure rotational cooling by OH and NH_3, collisional dissociation and ionisation. However, while the H_2 cooling

is calculated explicitly, the CO and H_2O cooling uses cooling functions (Kaufman and Neufeld, 1996). The available heating can account for processes like viscous and compressive heating, heating by comic-rays, photoionization and photodissociation associated heating, and collisional de-excitation of H_2 levels. There are also other mechanism that can either heat or cool. These include, e.g., energy heat/loss via gas-grain collisions, exothermic or endothermic reactions.

The standard chemical network of the code includes ~ 1000 chemical reactions involving ~ 140 chemical species, which can form from H, He, C, N, O, Na, Mg, Si, S, and/or Fe. Gas-phase reactions are limited to a maximum of four products and include neutral-neutral reactions, ion-neutral reactions, radiative and dissociative recombination, charge exchange, and ionisation and dissociation (via FUV photons and cosmic-rays). Gas-grain processes include, e.g., sputtering of icy mantles, thermal desorption of ice mantles, FUV-induced and cosmic-ray induced desorption from ice mantles. In the code, H_2 is formed on the dust grains adopting sticking coefficient on water ice given by Hollenbach and McKee (1979). Additionally, its ortho- to para-conversion (Wilgenbus *et al.*, 2000) and collisional dissociation are treated with improved accuracy. Moreover, the Paris-Durham shock code calculates a number of line intensities in the LVG formalism, but not for water. Calculating water line intensities for the shock code models is a subject of the project presented in Chapter 6.

3

This thesis

"Things are only impossible until they're not."
—Star Trek: The Next Generation

This chapter presents the scientific contributions of this thesis. It highlights the main questions that motivated the conducted projects (Sect. 3.1) and the main conclusions (Sect. 3.3). Each scientific project was presented in a separate article, where one is published, one is submitted, and one is in preparation. I summarize these in Sect. 3.2. I will also describe my contribution and that of my co-authors in Sect. 3.4. Building upon these studies, I present the perspectives and future outlook in Sect. 3.5.

3.1 RESEARCH CONTEXT AND DESCRIPTION

The question of how most stars in the Universe form remains open. On one side, it has to do with the fact that the current framework focuses on isolated star formation, but star formation predominantly happens in young massive clusters. These clusters host a whole range of (proto)stellar masses, including high-mass (proto)stars, which dominate in terms of energetic output, and so relatively few high-mass stars can easily outshine all low-mass stars, even though the low-mass population dominates in both total number and mass (Kroupa, 2001). Moreover, clustered environments promote interactions between its members. Here, high-mass stars are known to introduce winds and radiation influencing the neighboring objects and medium. For dense clusters, there is additional confusion when trying to assess their constituent populations. Therefore, the fact that most stars form in clusters introduces challenges that impact our constraints of the initial stellar mass and the core mass functions. These functions are observed to be near-universal, at least in nearby star-forming regions, and any theory of (clustered) star formation must be able to account for these mass distributions.

From the other side most star formation in the Universe happened approximately 10.5 Gyr - 11.5 Gyr ago. At these distances, we will never resolve individual forming stars. We might never even resolve individual clusters. Currently, we can only resolve GMC complexes (e.g., Jarugula *et al.*, 2019). However, if we really want to understand how most stars form, we need to account for this time in cosmic history and to do that, we need to scale what we currently know and understand, and try to take into account the different physical conditions at high-z in these scaling as best as possible.

All stars form in molecular clouds and numerous surveys have provided detailed molecular inventories of such clouds within the Galaxy. Thanks to

state-of-the-art facilities such as the Atacama Large Millimeter/submillimeter Array (ALMA), we are now routinely observing the distant Universe in molecular line emission. As we observe these molecules across the Universe, we start to fill the informational gap between high-z galaxies and the Milky Way. However, the question is: *how?* Trying to answer this questions is the subject of the first, already published, article.

In the second article this question still resonates, but under a different angle. Being the continuation of the first project, the second project focuses on using molecular emission associated with active star formation as a clean tracer of the star formation rates of galaxies. Since star formation lies at the very center of the baryon cycle and plays a pivotal role in shaping galactic ecosystems, it is important to correctly asses this parameter. This is where the star formation rate becomes very important, as it provides a quantitative description of the star-forming properties of a given object by relating the total mass of stars formed in a give time unit. Moreover, the star formation rate is used to establish the cosmic star formation history, which in turn is used to understand and quantify the evolution of galaxies.

Besides these big-picture questions, it is important to understand the origin of the molecular emission itself. More specifically, it is important to understand *under which conditions* we should expect *the most* and *the least* emission. This thesis was very much focused on water, which is a known tracer of active star formation in the Milky Way. This water emission predominantly comes from shocked regions around forming stars. However, there are different shock types and different conditions of the interstellar medium. Therefore, for water it was natural to ask under which shock conditions we should see the most and the least water emission. One way to do it is to by using shock and radiative transfer modeling.

3.2 PUBLICATIONS

This section provides a summary overview of the scientific articles included in Chapters 4, 5, and 6. The articles are:

1. Dutkowska, K. M. & Kristensen, L. E. 2022, A&A, 667, A135. “Water emission tracing active star formation from the Milky Way to high-z galaxies”;

2. Dutkowska, K. M. & Kristensen, L. E. 2022, A&A, submitted. “Star formation rate estimates from water emission”;

3. Dutkowska, K. M., Godard, B., Gusdorf A., & Kristensen, L. E., in preparation. “Water line emission in externally irradiated shocks”.

3.2.1 *Article I: Water emission tracing active star formation from the Milky Way to high-z galaxies*

This article presents a large-scale statistical galactic model, the galaxy-in-a-box model, which simulates molecular emission from young massive clusters. The model is built on observations of well-resolved nearby protostars and clusters. Specifically, the model simulates emission from molecular outflows,

which are known to scale with the mass of protostars, with the goal of serving as a proxy predicting the emission from clustered star formation on galactic scales. This study focused on the para-H_2O $2_{02} - 1_{11}$ line, which is one of the brightest transitions observed toward Galactic star-forming protostars and is now routinely observed toward distant galaxies.

In this article the exact methodology behind the galaxy-in-a-box model is described. It also includes a presentation of the relation between the observed water intensity and the envelope mass, which demonstrates that water is a low-contrast tracer of star formation, with $\int I_\nu \propto M_{\rm env}$. To evaluate model outputs, we conducted a parameter space study. The results of this study were analyzed by deriving flux distributions among simulated star-forming clusters, radial profiles of each galaxy, and total galactic emission. Since each distribution happened to be left-skewed, we used the interquartile range to measure the spread of the data.

From the model results, we concluded that water, due to its low-contrast properties, is highly sensitive to changes in the star formation efficiency. We also concluded that simulated water emission highly depends on cluster ages since younger clusters have more Class 0 and Class I protostars, exhibiting the highest protostellar outflow activity. Finally, the IMF and the molecular cloud mass distributions had negligible impact on the emission. Tentative comparisons of this parameter space study with observations indicated that the possible change in the emission between our Milky Way-like emission with that in other nearby galaxies could have to do, among other things, with different star formation rates.

3.2.2 *Article II: Star formation rate estimates from water emission*

This article is a follow-up to our first study. Here, we aimed to understand the dependence of the predicted para-H_2O $2_{02} - 1_{11}$ line emission on the star formation rates. Hence, we presented an extension to the model which enables deriving and putting constraints on star formation rates from the galaxy-in-a-box model on local and global scales. Moreover, we introduced a new way of defining the number of GMCs using galactic mass or FIR luminosity.

First, we summarized the modifications to the galaxy-in-a-box model and presented the galactic virial mass–FIR luminosity relation based on the data from Scoville and Good (1989). To evaluate the relation between the water emission and star formation rates, we ran more than 10^4 simulations with varying input parameters, focusing on the star formation efficiency, the free-fall time scaling factor, and the initial mass function. To analyze simulation results, we derived the following relations: stellar mass–star formation rate, FIR luminosity–water luminosity, FIR luminosity–star formation rate, water luminosity–star formation rate, and FIR luminosity–water-to-FIR luminosity ratio. We then compared these to literature relations.

Based on the comparisons with literature relations and observations, we concluded that with a straightforward application of the current Galactic observations and star formation properties, we are underestimating the total galactic emission while overestimating the star formation rates, particularly for more starburst-like configurations. Hence, a possible next step would be

to account for additional sources of emission, including supernovae, PDRs, and galactic outflows. However, revisiting the derived water emission for Galactic high-mass star-forming regions might be required, as we might "miss" the bulk of emission already at this stage.

3.2.3 *Article III: Water line emission in externally irradiated shocks*

In the third article, we probe the origin of water emission further. Hence, we took a step back to study water emission in shocked regions, as this is where most of the Galactic water emission originates. In this publication, we presented radiative transfer modeling of results from sophisticated shock models using the Paris-Durham shock code. These shock models covered various conditions of the pre-shock medium and the shock itself.

To better understand how pre-shock and shock conditions influence the emission, we considered pre-shock densities ranging from 10^2 to 10^6 cm^{-3} and UV radiation field strengths with G_0 between 0 and 10^3 for both shock types. For J-type shocks, we changed velocities between 5 and 30 km s^{-1}, while for the C-type shocks between 5 and 20 km s^{-1}. The scaling of the strength of the transverse magnetic field was constant for the C-type shock, while for the J-type shock with velocities of 25 and 30 km s^{-1}, we additionally explored the impact of a stronger magnetic field. We conducted the radiative transfer modeling of these shocks to infer the expected water intensities and excitation. We analyzed the results by comparing derived integrated intensities, specific intensities, excitation diagrams, and fractional abundance of water as a function of time through the shock.

From the analysis of the results, we concluded that water emission is highly dependent on the density of the medium and the velocity of shock waves. More specifically, pre-shock densities can have a dramatic impact on the intensities and excitation of water for both shock types. We have also concluded that J-type shocks are, on average, associated with higher water emission and excitation. However, for velocities $\geq$ 25 km/s, there are clear signs of collisional dissociation of water molecules. In C-type shocks, shock velocity is crucial for the emergence of transitions with $E_{\rm up} \gtrsim 300$ K, which may be observable by the James Webb Space Telescope (JWST).

3.3 OVERALL CONCLUSIONS

Taken together, the main conclusions of this thesis are:

- Water emission, at least from the para-H_2O $2_{02} - 1_{11}$ line, is a low-contrast tracer of active star formation, and hence, it is sensitive to changes in star-formation efficiency. Coming from the earliest stages of star formation, it is also sensitive to cluster ages, where the larger the population of Class 0 and Class I protostars, the higher the emission.

- If the model simulating emission from young massive clusters will only rely on the current Galactic observations and star formation properties, it will underestimate the total galactic emission while overestimating its star formation rate. This is especially true for high-z quasars, ULIRGs, and HyLIRGs.

- The radiative transfer modeling of shock models stressed the importance of the density and velocity on water emission. For C- and J-type shocks, pre-shock densities have a significant impact on the intensities and excitation of water, while shock velocities can promote excitation and emission of water lines, but also collisional dissociation.
- Creating a tool predicting molecular emission from galactic star formation requires accounting for various processes and emitting sources that could contribute to the same type of emission. Some of the emission sources that could impact observed galactic emission are supernovae, PDRs, and galactic outflows. However, the model would also account for different physical properties of the galaxies, including morphologies, distribution of dense gas, and the ISM properties.

3.4 OWN CONTRIBUTIONS VERSUS CONTRIBUTIONS OF COLLABORATORS

In this section I will explain my own contribution to the articles which form the body of this thesis, compared to the contributions from my collaborators.

Article I This project was based on the development of a new astrophysical modeling tool, which I developed, besides the cluster part, which is based on the work of L. E. Kristensen, the Ph.D. supervisor. I have performed the analysis of the results, I wrote the paper on my own and produced all the tables and figures. L. E. Kristensen supervised the development of the model, commented on the paper and the figures, and actively participated in editing the article.

Article II This paper presents a follow-up study of the first article, which included extending the galaxy-in-a-box model. I wrote the star formation rate extension and L. E. Kristensen wrote a new way of defining the number of giant molecular clouds. L. E. Kristensen also provided the observational data of the extragalactic objects used in the analysis of the results. I ran the simulations, performed the analysis, wrote the paper on my own, and produced all the tables and figures. L. E. Kristensen provided comments on the paper and the figures and actively participated in editing the article.

Article III This study involves the large grid of shock models calculated by B. Godard[†] (to be presented in Kristensen et al. in prep.) and a modified version of the LVG model developed by B. Godard and A. Gusdorf[†]. To conduct this project, I attended the Interstellar Shocks School in Les Houches (France) to familiarize myself with the shock code, and then I spent a month at the École Normale Supérieure in Paris to work with B. Godard and A. Gusdorf on modifying the LVG model such that it would be suitable for working with water emission. L.E. Kristensen chose and retrieved the presented shock models. I performed the analysis, wrote the paper, and produced all the tables and figures. L. E. Kristensen provided comments on the paper and the figures and actively participated in editing the article. A. Gusdorf supplied important comments to the article. We are currently fine-tuning the new LVG code version to better reproduce results from previous studies. This

[†]Observatoire de Paris, École Normale Supérieure, Université PSL, Sorbonne Université, CNRS, LERMA

should not change the main conclusions of the paper. However, I will re-run the LVG code and remake the figures once we agree on the final version of the LVG code.

3.5 FUTURE OUTLOOK

In this section I will explore possible pathways for future projects, building on the results and contributions of this thesis. I start by presenting ideas of possible model improvements in Sect. 3.5.1 and observational ideas in Sect. 3.5.2.

3.5.1 *Comparing Milky Way-like star formation with that of nearby galaxies*

Already undertaken studies, including those performed with the galaxy-in-a-box model, demonstrate that before going all the way to the high-redshift regime, it will be necessary to focus on the local Universe and try to match what we see there. Hence, in order to compare the Milky Way-like star formation with that of nearby galaxies, e.g., NGC 1365 (Sandqvist *et al.*, 2021) or Mrk 231 (Omont *et al.*, 2011), the model needs to be tailored to more complex galactic types, morphologies, and astrophysical processes. To achieve this, one would have to rely on archival data from, e.g., ALMA, *Herschel*, the NASA Hubble Space Telescope, and soon the James Webb Space Telescope (JWST) to properly deconstruct morphologies, asses the amount of available molecular material, and understand the spatial extent of the emission. These observations would have to include main shock tracers, i.e., H_2O (*Herschel* for nearby, and ALMA for distant galaxies), high-J CO ($J_{\rm up} \geq 10$; *Herschel* for nearby, and ALMA for distant galaxies), and H_2 (JWST, ground-based observatories), as well as low-density tracers, e.g., low-J CO (ALMA).

Galaxies are known to be complex ecosystems with complex morphologies. Moreover, they can host AGNs and drive galactic-scale outflows. These properties may or may not affect star formation processes in a galaxy, but they will certainly affect the observed molecular emission. In particular, galaxies at high redshift have physical conditions heavily influenced by the nuclear regions of the galaxies, and depending on the galaxy type, these could be a mixture of extreme nuclear starbursts and obscured AGN or even dominated by AGN over the entire electromagnetic spectrum. Therefore, to properly understand molecular excitation conditions of star-forming galaxies and understand the contribution to observed emission from active star formation we need to put all the observational pieces of the puzzle together and fully characterize emission from these objects. Thus, it is crucial to account for different galactic properties in the model. Moreover, to accurately assess the expected emission, the model needs to account for the possible contribution from, e.g., supernovae and PDRs.

There are other, perhaps more fundamental, aspects of the galactic star forming environment, e.g., a whole spectrum of the commonly used initial mass functions and the detailed distribution of dense molecular gas to better match the picture of nearby galaxies. Star-forming regions in galaxies are described by several scaling relations and distributions. On the most

FIGURE 3.1: General overview of possible extensions to the galaxy-in-a-box model to make it tailored for comparing Milky Way-like star formation with that of nearby galaxies. The considered extensions cover the complexity of galactic ecosystems that would definitely influence the observed molecular emission (morphology, galactic-scale outflows, AGN feedback, SNe, and PDRs) and that associated with the galactic star formation environment (distribution of dense molecular gas, various forms of the IMF). Moreover, other shock tracers, like H_2 and high-J CO, should be considered to break observational degeneracies and provide additional constraints on the emission. The presented overview also includes a potential future extension that would account for various ISM conditions, influencing both the emission and star formation. Accounting for changing ISM conditions is particularly important when considering distant galaxies, but also for the proper treatment of central regions of galaxies. All extensions marked with solid lines are further described in Section 3.5.1, while the extensive comments to the extension marked with the dashed line can be found in Section 1.3 and 2.4.3.

fundamental level, star formation is described by the initial mass function. However, the outcome of this function depends on the mass of molecular clouds in which stars form, which on galactic scales also follow various mass functions. Moreover, these molecular clouds follow different spatial distributions in galaxies, usually complex and dependent on their galactic location (Kennicutt and Evans, 2012). Accounting for these star-formation recipes is crucial to ensure that the model is suited for comparing the Milky Way-like star formation with that of nearby (and ultimately farther) galaxies. Luckily, the model has the advantage that its spatial configuration can be freely parameterized. Therefore, a synergy with cosmological magnetohydrodynamical simulations of galaxy formation, e.g., the IllustrisTNG project, could ultimately result in mock observations of molecular emission from active star formation across the Universe.

Besides water emission, high-J CO transitions ($J_{\rm up} \gtrsim 10$) could serve as complementary shock tracers, helping to understand the shock conditions and breaking possible degeneracies, as they have similar excitation conditions as water (e.g., Kristensen *et al.*, 2017). As argued in Chapter 1 (give precise reference), CO is less sensitive to chemical effects and so its high-J transitions may be a "cleaner" and more robust tracer of protostellar outflows.

In the JWST era, there is another extremely important shock tracer to consider, i.e., H_2. By being the most abundant molecule in the Universe with widely spaced energy levels, H_2 easily lights up in regions where energy is deposited. This deposition can either be in the form of UV radiation (i.e., in PDRs associated with high-mass stars) or mechanical (e.g., in the form of shocks from forming stars or supernova explosions). The results from running the large shock grid by Kristensen et al., in prep, showed

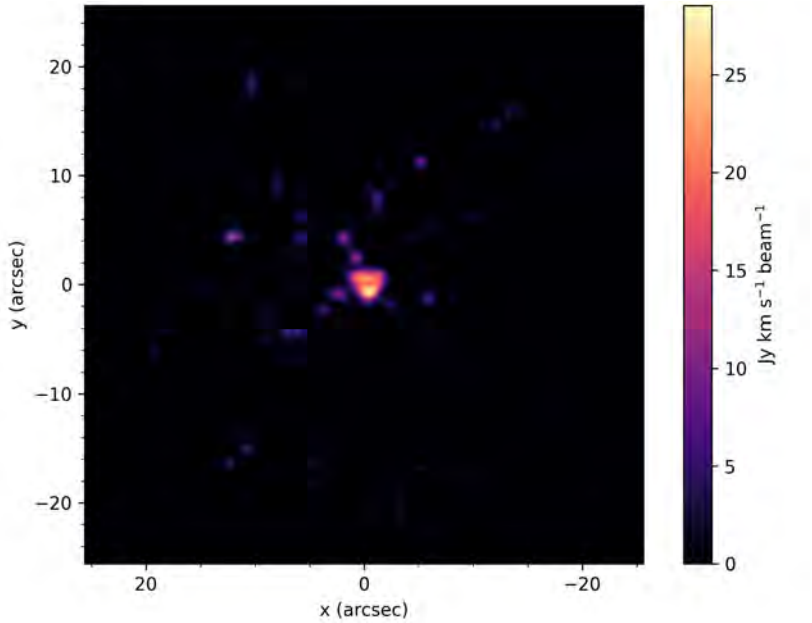


FIGURE 3.2: Generated integrated intensity map of methanol emission at 338.4 GHz, convolved with 0.83″ beam. The cluster configuration was chosen to resemble G10.47+0.03. The color scale is shown on the right.

that H_2 emission is particularly bright at higher densities ($> 10^4$ cm^{-3}) and higher magnetic field strengths, with the 1-0 S(3) and the 1-0 Q(1) lines being typically the brightest. These lines are located at 1.96 and 2.40 μm, respectively, falling within JWST's NIRSpec range. Furthermore, the excitation is set by the temperature of the shock, which in turn scales as the velocity squared, and at very high densities ($\geq 10^6$ cm^{-3}), H_2 is no longer the dominant coolant, but H_2O instead takes over. In this hot gas, H_2O has a number of highly excited transitions that are readily visible also to the JWST. Thus, observing H_2O emission together with H_2 will provide strong constraints on particularly the velocity and density structure of these shocks, along with the strength of the magnetic field. Being able to constrain these key parameters will be of crucial importance for understanding the physical conditions of these interstellar shock waves. Thus, observing H_2O emission together with H_2 would provide strong constraints, in particular on the velocity and density structure of these shocks, along with the strength of the magnetic field.

3.5.2 *Observations of a high-mass star forming region*

The research presented in this thesis could be continued, e.g., with observations, which would spatially resolve emission from a Galactic high-mass star-forming region to improve our understanding of the contribution of different components to the emission. This way, the building block of the galaxy-in-a-box model could be benchmarked against observations. It was the subject of a Submillimeter Array (SMA) proposal in the 2021A semester (PI: K. Dutkowska, 2021) in which we proposed to observe a high-mass star-forming region G10.47+0.03. The proposal was successful and B-ranked. However, due to multiple antenna failures and the COVID lockdown, it wasn't possible to set the array in the configuration required for the observations,

and in effect they could not happen. We have not resubmitted this proposal.

The idea for these SMA observations was to disentangle emission from clustered star formation into its constituent parts: the contribution from low- and high-mass protostars and the feedback from the high-mass protostars on their surroundings. We intended to achieve this disentanglement by utilizing the capabilities of the cluster-in-a-box model. Originally, this cluster model relies on a relationship between emission from methanol at 338 GHz and envelope mass, built on observations of low- and intermediate-mass protostars in NGC 1333 and Serpens Main. The model was benchmarked against a single high-mass embedded cluster IRAS 16547-4247 and successfully disentangled its emission (Kristensen and Bergin, 2015). However, by relying on a single high-mass observation, it is hard to evaluate if it was just luck with the source choice or if the model can reproduce the morphology of embedded clusters in general. Thus, we proposed to study the emission from a high-mass star-forming region G10.47+0.03 to double the number of high-mass clusters where emission from the various cluster components are identified (our predicted emission map can be seen in Fig. 3.2). The chosen region is located at $d \sim 8.6$ kpc and has a luminosity of $8.1{\times}10^5$ $L_\odot$ (van der Tak *et al.*, 2019), and it hosts a hot molecular core (e.g., Gensheimer *et al.*, 1996; Cesaroni *et al.*, 2010). Moreover, by being relatively distant, insights from this source would be relevant for the galaxy-in-a-box model. Other SMA observations of G10.47+0.03 (Rolffs *et al.*, 2011) proved that this region exhibits a great range of molecular species, like OCS, HC_3N, and CH_3OH. The targeted methanol 7_k–6_k branch at 338 GHz would also cover other molecules useful for studying the chemistry of high-mass star-forming regions, in particular, the hot-core chemistry (Jørgensen *et al.*, 2020). CH_3OH is chemically similar to water: it primarily forms on dust grains, and it is efficiently released in shocks due to sputtering of the ice mantles. Besides, it is easily observable from the ground by existing facilities. The main difference to water, is that while both molecules are easily destroyed in shocks through collisional dissociation, water can readily reform in the gas phase while methanol cannot. Furthermore, methanol emission is observed to also originate in molecular hot cores, even in single-dish observations, whereas water is not. Thus, by targeting a high-mass star-forming region at high angular resolution, the emission would be spatially disentangled in a way that is not currently possible for water emission.

3.5.3 *Water emission from high-z galaxies*

Another natural continuation of this thesis would be ALMA observations of water emission at high redshift. Even though numerous studies report water emission toward $\sim$ 20 high redshift galaxies (e.g., Omont *et al.*, 2011; Yang *et al.*, 2013; Jarugula *et al.*, 2019), no systematic surveys have been undertaken for targeting multiple lines at uniform sensitivity. Undertaking such a systematic survey was the subject of an ALMA proposal in Cycle 8 (PI: K. Dutkowska, 2021) in which we proposed to observe water emission toward a diverse sample of 17 high-z galaxies. Unfortunately, the proposal was not accepted. We decided not to resubmit the proposal in Cycle 9.

The idea for this ALMA proposal was related to trying to understand

if stars form in the same manner as in the present-day Milky Way, or not. To trace star formation in external galaxies, studies typically rely on observations at either optical/UV lines (e.g., Ly-α) or dust continuum emission at far-infrared wavelengths, which are then calibrated to star-formation rates. However, an alternative exists, which has not been used in the past: molecular line emission. By observing directly the gas in which the stars form, we bypass a number of proxies, such as uncertain dust properties, and gain direct access to the bulk mass. So far, molecular emission observed toward the high-z star-forming regions resembles emission from our own Galaxy. Thus, we proposed to systematically calibrate molecular emission as a unique tracer of star formation across cosmic time by linking galactic-scale molecular emission to the molecular emission sites associated with star formation in our own Galaxy. These observations would also further serve to pave the way for follow-up observations, targeting full molecular inventories in these distant sources.

We focus this survey on H_2O transitions that are likely unaffected by FIR pumping, but instead collisionally excited (González-Alfonso *et al.*, 2014). San José-García *et al.* (2016) found that there is a direct proportionality between the far-infrared luminosity, $L_{\rm FIR}$, of these Galactic star-forming regions and their H_2O line luminosity. Interestingly, extragalactic sources perfectly matches this correlation, even HyLIRGs. The origin of this tight correlation has been interpreted as water emission being caused by the FIR radiation field (González-Alfonso *et al.*, 2014); however, an alternative explanation exists, namely that the emission from lower-excited transitions is caused by active star formation. If that is the case, then the scaling can be used to calibrate the star-formation rate from the Galactic sources, all the way to high z. In order to further test this hypothesis, we proposed to target several lines of water, together with high-J CO lines, that are unaffected by FIR pumping, i.e., para-H_2O $2_{02}-1_{11}$ and ortho-H_2O $3_{12}-3_{03}$ lines, at 987.9267 and 1097.36505 GHz, respectively, and CO 10−9 or 9−8 transitions, at 1151.985452 and 1036.912393 GHz, respectively. This combination of bright lines would allow us to calibrate star-formation rates independently from dust observations.

For this study we chose a sample of 17 submillimetre galaxies (SMGs) (for details see Tab.3.1) consisting of ULIRGs and HyLIRGs. All of the chosen galaxies are weakly to strongly lensed. Gravitational lensing of the high-z galaxies can be of crucial importance for observational studies. As the lensed background source receives a boost in apparent flux by a magnification factor μ, the study of emission that otherwise would be too faint to detect becomes possible. Here, we specifically chose the galaxies where a lens model is either available or there is a basis to infer the probable magnification factor. Hence, by ensuring that the chosen sources have available μ values, we would retrieve the actual emission output while taking full advantage of gravitational lensing.

TABLE 3.1: *Source coordinates and properties. L_{IR} refers to intrinsic luminosity, μ is the magnification factor.*

Source	RA	Dec	z_{spec}	$L_{IR}(L_\odot)$	μ
NAv1.56	13:44:29.518	+30:30:34.050	2.3010	9.8×10^{12}	9.79
HXMM-01	02:20:16.603	−06:01:43.200	2.3079	1.75×10^{13}	1.30
HXMM-02	02:18:30.607	−05:31:31.440	3.3903	5.14×10^{12}	5.33
SPT2349-50	23:49:42.200	−50:53:30.900	2.8759	1.3×10^{13}	2.15
SPT2232-61	22:32:51.140	−61:14:44.180	2.8936	7.0×10^{12}	5.5
SPT0314-44	03:14:28.330	−44:52:22.180	2.9345	1.5×10^{13}	5.5
NAv1.195	13:26:30.216	+33:44:07.600	2.9510	1.83×10^{13}	4.1
SPT0551-50	05:51:39.420	−50:58:02.000	3.1642	8.0×10^{12}	5.5
SPT0226-45	02:26:49.460	−45:15:39.000	3.2330	8.0×10^{12}	5.5
G12.v2.30	11:46:37.900	−00:11:32.000	3.2592	1.64×10^{13}	9.5
SPT2129-57	21:29:12.530	−57:01:54.300	3.2601	8.0×10^{12}	5.5
SPT0652-55	06:52:07.240	−55:16:00.100	3.3466	1.3×10^{13}	5.5
SPT0529-54[a]	05:29:03.090	−54:36:40.200	3.3689	2.5×10^{12}	13.23
SPT0516-59	05:16:37.980	−59:20:32.100	3.4039	5.0×10^{12}	5.5
SPT0112-55	01:12:09.030	−55:16:36.000	3.4430	1.9×10^{12}	5.5
SPT0027-50	00:27:06.540	−50:07:19.800	3.4436	1.5×10^{13}	5.49
NCv1.143	12:56:32.544	+23:36:27.630	3.5650	1.14×10^{13}	11.3

[a] Already observed in H_2O 2_{02}–1_{11} by Jarugula *et al.* (2019) and this transition was excluded from observations.

3.6 SUMMARY AND CONCLUSION

This chapter summarized the work done as a part of this Ph.D. project. This work consisted of developing a novel astrophysical modeling tool (Sect. 3.2.1 and 3.2.2) and theoretical studies of water chemistry in externally irradiated shocks (Sect. 3.2.3). From the main results of this thesis, we concluded that a straightforward application of the current Galactic observations and star formation properties is not enough to properly reproduce the emission observed toward distant galaxies. However, they also showed that the assumption that water emission should trace active star formation across cosmic epochs, or at least trace it to some degree, is reasonable. After accounting for more processes (including other emission sources) and differences in the star-forming and ISM properties, we should get closer to a more realistic scenario and much better reproduce the emission.

I also presented different ways to develop the scientific work further. It seems natural that the next step should involve extending and improving the current version of the galaxy-in-a-box model (Sect. 3.5.1). By accounting for the complexity of galactic ecosystems influencing the observed molecular emission (morphology, galactic-scale outflows, AGN feedback, SNe, and PDRs) and the properties of the galactic star formation environment (distribution of dense molecular gas, various forms of the IMF), it should be possible to build a version of the model that would connect the Galactic and extragalactic regime and help us get closer to unveiling the cosmic properties of star formation. However, this extension should also consider various ISM

conditions, as these would impact both the emission and star formation. Accounting for changing ISM conditions would be particularly important for distant galaxies and the proper treatment of central regions of galaxies. Moreover, to be more accurate other shock tracers, like H_2 and high-J CO. These additional traces would help to break observational degeneracies and provide additional constraints on the emission.

Another avenue would be to focus on observations first and get a better grasp of the molecular emission within star-forming clusters and distant galaxies. One possibility would be to conduct observations similar to that from the SMA proposal (Sect. 3.5.2). With such observations, we could disentangle emission from clustered star formation into its constituent parts: the contribution from low- and high-mass protostars and the feedback from the high-mass protostars on their surroundings. When it comes to the distant regime, a systematic survey like the one proposed for ALMA observations (Sect. 3.5.3) could help us calibrate molecular emission as a unique tracer of star formation between galaxies. Moreover, aiming for a diverse list of sources (e.g., with and without AGNs) would allow for assessing the contribution of non-star-forming emission sources to the emission, which would ultimately help us better estimate the feedback from star formation. All in all, we are taking the first steps toward a more realistic model for interpreting molecular emission from distant galaxies, and not just treating them as single "blobs" of emission.

4

Water emission tracing active star formation from the Milky Way to high-z galaxies

K. M. Dutkowska[1], L. E. Kristensen[1]

[1] Niels Bohr Institute & Centre for Star and Planet Formation, University of Copenhagen, Øster Voldgade 5−7, 1350 Copenhagen K., Denmark

Published in Astronomy & Astrophyics, vol. 667, article no. A135, Nov 2022

Abstract

Context. The question of how most stars in the Universe form remains open. While star formation predominantly takes place in young massive clusters, the current framework focuses on isolated star formation. This poses a problem when trying to constrain the initial stellar mass and the core mass functions, both in the local and distant Universe.

Aims. One way to access the bulk of protostellar activity within star-forming clusters is to trace signposts of active star formation with emission from molecular outflows. These outflows are bright (e.g., in water emission), which is observable throughout cosmological times, providing a direct observational link between nearby and distant galaxies. We propose to utilize the in-depth knowledge of local star formation as seen with molecular tracers, such as water, to explore the nature of star formation in the Universe.

Methods. We present a large-scale statistical galactic model of emission from galactic active star-forming regions. Our model is built on observations of well-resolved nearby clusters. By simulating emission from molecular outflows, which is known to scale with mass, we create a proxy that can be used to predict the emission from clustered star formation on galactic scales. In particular, the para-H_2O $2_{02} - 1_{11}$ line is well suited for this purpose as it is one of the brightest transitions observed toward Galactic star-forming regions and is now

routinely observed toward distant galaxies.
Results. We evaluated the impact of the most important global star formation parameters (i.e., initial stellar mass function, molecular cloud mass distribution, star formation efficiency, and free-fall time efficiency) on simulation results. We observe that for emission from the para-H_2O $2_{02} - 1_{11}$ line, the initial mass function and molecular cloud mass distribution have a negligible impact on the emission, both locally and globally, whereas the opposite holds for star formation efficiency and free-fall time efficiency. Moreover, this water transition proves to be a low-contrast tracer of star formation, with $\int I_\nu \propto M_{\rm env}$.
Conclusions. The fine-tuning of the model and adaptation to morphologies of distant galaxies should result in realistic predictions of observed molecular emission and make the galaxy-in-a-box model a tool for analyzing and better understanding star formation throughout cosmological times.

4.1 INTRODUCTION

Water is one of the key molecules tracing active and current star formation (SF); in the Milky Way water emission is almost uniformly associated with molecular outflows from protostars (van Dishoeck *et al.*, 2021b). These outflows arise at the earliest stages of star formation, when the protostar is in its main accretion phase and the interaction between the infalling envelope, winds, and jets launched from the protostar is particularly strong (Bally, 2016). When this happens, water predominantly locked up as ice on dust grains is released from the icy grain mantles into the gas phase, causing a jump in the abundance of many orders of magnitude. At the same time, the physical conditions are conducive to water being readily excited into rotational states, and the de-excitation leads to subsequent cooling (Suutarinen *et al.*, 2014). Therefore, whenever star formation occurs, these outflows light up in water emission.

Water emission is also observed toward high-redshift galaxies (e.g., Yang *et al.*, 2016; Jarugula *et al.*, 2019; Stanley *et al.*, 2021). The origin of this emission is interpreted as the molecular clouds from which stars form, and not the protostellar outflows. This interpretation is primarily grounded in a very tight correlation between the far-infrared luminosity ($L_{\rm FIR}$) and water line luminosity ($L_{\rm H_2O}$), where $L_{\rm FIR}$ is thought to trace dust (e.g., González-Alfonso *et al.*, 2008; González-Alfonso *et al.*, 2014; Omont *et al.*, 2013), which indicates that $L_{\rm FIR}$ indirectly traces molecular clouds, and the excitation of water molecules is expected to be caused by the FIR radiation field through radiative pumping.

Two dominant mechanisms contribute to returning the water ice to the gas phase. The first, and the most effective, is thermal desorption if the temperature of the dust grains rises above $\sim$ 100 K (e.g., Fraser *et al.*, 2001). Such high temperatures are typically found within the inner $\sim 10^2$ AU of forming stars (e.g., Bisschop *et al.*, 2007). The second is sputtering of ice from the dust grains when neutral species or ions with sufficient kinetic energy (predominantly H_2, H, and He) collide with the ice mantle. Due to its highly energetic character, sputtering can cause the dissociation of

water molecules. However, the high temperatures within outflows make the gas-phase synthesis of water effective enough to sustain the high abundance of water molecules (Suutarinen *et al.*, 2014). Finally, water may also be directly synthesized in the gas from ion-neutral reactions. In dark molecular clouds, this path is inefficient (Hollenbach *et al.*, 2009), but in photon and X-ray-dominated regions (PDRs and XDRs) where the ionization fraction is high, this mechanism may be dominant (Meijerink *et al.*, 2011).

Observations of emission from the ground state levels of ortho- and para-water (e.g., the ortho-H_2O $1_{10} - 1_{01}$ line at 557 GHz) are known to trace the warm outflowing gas (Mottram *et al.*, 2014), as do the mid-excited transitions, with $E_{up} \sim 100 - 300$ K, like the para–H_2O $2_{02} - 1_{11}$ line at 988 GHz. Subsequently, highly excited water transitions with $E_{up} > 300$ K, such as the ortho-H_2O $5_{23} - 5_{14}$ line at 1411 GHz, are only populated in high-temperature gas and strong shocks (van Dishoeck *et al.*, 2013). Water, except for the ground state transitions, may also be excited by pumping to higher-excited levels by FIR photons (González-Alfonso *et al.*, 2014). However, in the Galactic outflows where water excitation is collisionally dominated, there are no signs that other processes, such as FIR pumping, play a significant role in the excitation (Mottram *et al.*, 2014). It raises the question of whether water behaves differently at high redshift.

With the great progress in astrochemistry in the past years, particularly thanks to the observational programs carried out with the *Herschel* Space Observatory (active 2009 – 2013) and the Atacama Large Millimeter/submillimeter Array (ALMA), we are now routinely observing the distant Universe in molecular line emission (Hodge and da Cunha, 2020). Numerous surveys have provided detailed chemical inventories of star-forming regions within the Galaxy (for a recent review, see Jørgensen *et al.*, 2020), and as we observe the same molecules across the Universe (McGuire, 2022), we can now start to fill the informational gap between high-redshift galaxies and the Milky Way and start comparing the observational results between these regimes.

One of the questions we can answer is how molecular line emission can be used to quantitatively trace active star formation. Most stars form in clusters (Lada and Lada, 2003). In clusters all ranges of stellar masses are present and relatively few main-sequence high-mass stars can easily outshine the entire low-mass population. Moreover, the younger the protostar, the deeper it is embedded in gas and dust. Therefore, we need to use reliable tracers of active star formation that are common and bright enough to be easily observed. One of the best tracers in our Galaxy, also observed in the distant Universe, is water; the emission is particularly bright in the deeply embedded phase, when the protostars drive molecular outflows (e.g., Bally, 2016).

In this work we present a model that can be used to compare observations from different galaxies with the emission that could arise from active star-forming regions. In the model we simulate the emission from molecular outflows, one of the key signposts of active and current star formation, that would arise from protostars within star-forming clusters. These star-forming clusters are then incorporated into a large-scale galactic model, which contains a range of molecular clouds in which the stars form. In this

study we focus on simulating water emission at 988 GHz (the $J_{KaKc} = 2_{02} - 1_{11}$ line), which is particularly bright in Galactic star-forming regions and has been observed toward many high-redshift galaxies (e.g., van Dishoeck *et al.*, 2021b; van der Tak *et al.*, 2013), but the model is set up such that it can ingest and predict any type of outflow emission.

This paper is organized as follows. Section 4.2 describes our galactic model in detail and provides the methods used to obtain the results. Subsequently, in Section 4.3 we present the results of a parameter space study of the model, which we then discuss in Section 4.4 and present some future prospects. Finally, we present our conclusions in Section 4.5.

4.2 MODEL

On galactic scales, stars predominantly form in giant molecular clouds (GMCs). These GMCs form complexes that follow a certain spatial distribution in galaxies, as will be outlined below. Hence, to build a model of galactic emission from active star-forming regions, we broke this distribution into its constituent parts. We used an existing cluster model (Sect. 4.2.1) as a starting point and adapted it into a cloud model. We subsequently used this cloud model as the building blocks for the galaxy-in-a-box model (see Sect. 4.2.2). Finally, we built the observational template used for emission assignment in the form of a database where we gathered the available water data from ground-based observations and the *Herschel* Space Observatory (Sect. 4.2.3). The model is outlined in Fig. 4.1 with the different modules highlighted.

4.2.1 *Cluster-in-a-box model*

Most stars form in clusters, especially in high-mass clusters (Lada and Lada, 2003). These clusters harbor protostars covering the whole range of stellar masses. However, at the time of formation they are also deeply embedded in their natal clouds, and so it is impossible to access the initial main-sequence stellar populations forming within these clusters directly. Moreover, massive stars dominate cluster emission, making the low-mass population hard to access observationally. An alternative is to probe this population with outflow emission. Studies show that there is a proportionality between this emission and protostellar envelope mass (e.g., Bontemps *et al.*, 1996; Skretas and Kristensen, 2022). Kristensen and Bergin (2015) used this link to construct the cluster-in-a-box model[1], simulating methanol emission from low-mass outflows in embedded star-forming clusters.

[1]The codes required for running the model are publicly available at https://github.com/egstrom/cluster-in-a-box
doi: 10.5281/zenodo.13184

The cluster model consists of a template cluster and molecular emission assigned to each protostar in the cluster. The spatial distribution of protostars in the template cluster is based on the model by Adams *et al.* (2014), where the radial extent of the cluster can be described by the power-law function $R_{\rm max} = R_0(N/N_0)^{\alpha_{\rm c}}$, where N is the number of stars in the cluster and the power-law slope $\alpha_{\rm c} = 1/3$. The age distribution of protostars in the stages Class 0, Class I, "flat-spectrum", Class II, and Class III follows that of the Perseus low-mass star-forming cluster (Evans *et al.*, 2009; Sadavoy *et al.*, 2014). The model applies the Chabrier initial mass function (IMF)

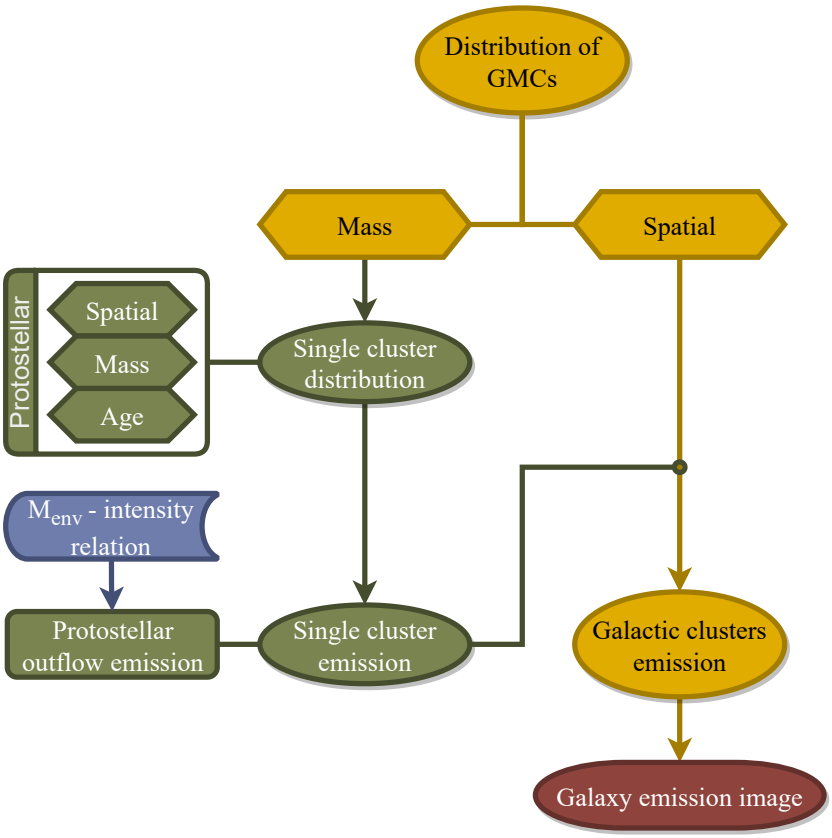


FIGURE 4.1: Schematic flowchart of the galaxy-in-a-box model. The computation starts with generating the spatial (Sect. 4.2.2) and mass (Sect. 4.2.2) distributions of GMCs in the simulated galaxy. The GMC mass distribution serves as the input to the module generating the protostellar spatial, mass, and age distributions within individual star-forming clusters. Here, each randomly chosen GMC mass is an initial mass of the cluster. Having calculated these distributions, the model uses them to assign molecular outflow emission to each protostar within the cluster, based on the envelope mass–outflow intensity relation calculated using the Water Emission Database (Sect. 4.2.3). After repeating these calculations for all GMCs, the emission and mass of star-forming clusters are returned to the galaxy-in-a-box model. Subsequently, the model merges the spatial distribution of the initial GMCs with water emission emerging from the corresponding star-forming clusters. Once the model returns the expected emission from the galaxy, this raw galactic emission grid is convolved with a Gaussian beam, producing the integrated intensity image of the galaxy. In this flowchart yellow corresponds to the galactic part of the model (Sect. 4.2.2), green to the cluster model (Sect. 4.2.1), red to the last stage of the model, and blue to the external data input.

(Chabrier, 2003) for young clusters and disks. The outflow position angles are chosen randomly from 0° to 180°, as is the distance from the protostar to the outflow lobe with the maximum separation equal to 2×10^4 AU. The molecular outflow emission is assigned based on a scaling relation of the observed outflow emission from single low-mass protostars in the nearby low-mass star-forming regions NGC 1333 and Serpens Main and their modeled envelope masses. However, the emission is assigned only to Class 0 and I protostars, because flat-spectrum, Class II, and Class III objects only produce negligible molecular outflows (Arce *et al.*, 2007). The cluster-in-a-box model focuses on the $7_0 - 6_0$ A$^+$ methanol line at 338.409 GHz.

The cluster model did not include the contribution from high-mass sources, neither in the form of their outflows nor their hot cores. Nevertheless, a proof-of-concept study showed that the model reproduces the extended emission from a high-mass star-forming region to within a factor of two without tweaking the input parameters, suggesting that low-mass out-

flows account for ~50% of the total cluster emission. These results indicate that a toy model of this kind can be used to constrain parameters of star-forming clusters and decipher the contribution from their components (i.e., molecular outflows and hot cores), and to reproduce their morphologies.

4.2.2 *Galaxy-in-a-box*

New telescope facilities, particularly ALMA, are now routinely observing molecular emission at high redshift (e.g., out to $z \gtrsim 6$, Strandet *et al.*, 2017). One possibility for understanding the origin of this emission is to use Galactic star-forming clusters as templates of emission. This approach would consist first in scaling Galactic observations to cover entire galaxies, and then in comparing these scalings with actual observations of local galaxies. Next, the scalings would be extrapolated to the high-redshift regime ($z \gtrsim 1$), where they can be compared to observations. Practically, the approach would consist of first creating a cluster model (Sect. 4.2.1), then populating a galaxy with these model clusters, thereby going from a cluster-in-a-box model to a galaxy-in-a-box model. This model consists of a template (spiral) galaxy with molecular cloud spatial, age, and mass distributions, and of template stellar clusters with assigned outflow emission based on the cluster-in-a-box model. In this manner, emission from an entire galaxy can be simulated, with the advantage that the model only depends on a few input parameters.

Our knowledge of astrochemistry and star formation primarily comes from observations of the Milky Way (e.g., Herbst and van Dishoeck, 2009). Thus, when first going to the extragalactic regime, the goal is to use the knowledge from the Milky Way together with a similar galaxy that could provide the pivotal information on its spatial structure. Furthermore, the galaxy should be nearby, well-studied, and ideally face-on, such that line-of-sight effects are minimized. One example of such a galaxy is the grand-design spiral Whirlpool Galaxy, M51. In addition to the spiral structure, M51 has an apparent size of 24 kpc (Jarrett *et al.*, 2003), which is roughly comparable to the estimated size of the Galactic disk $\gtrsim$ 30 kpc (Bland-Hawthorn and Gerhard, 2016). It is nearby ($D \sim 7.6$ Mpc; Ciardullo *et al.*, 2002) and almost face-on ($i \sim 22°$; Colombo *et al.*, 2014), making it an object of numerous studies, for example the Plateau de Bure Interferometer Arcsecond Whirlpool Survey (PAWS; Schinnerer *et al.*, 2013). Therefore, in the following, we base the template galaxy against observational data from M51.

For the galaxy-in-a-box we picked water as a default molecule for the simulation of galactic emission. The reason is that from the 30% of the molecular species observed in the Milky Way that were also detected in external galaxies (McGuire, 2022), water stands out as a ubiquitous star formation tracer in the Milky Way with emission dominated by molecular outflows and is readily observed toward high-z galaxies (e.g., Yang *et al.*, 2016; Yang *et al.*, 2017; Jarugula *et al.*, 2019; van Dishoeck *et al.*, 2021b). For the purpose of this work, we focused on the emission of the para-H_2O $2_{02} - 1_{11}$ line at 987.927 GHz.

In addition to the change in the molecular species used to obtain the mass-intensity relation, the cluster model underwent a few upgrades while

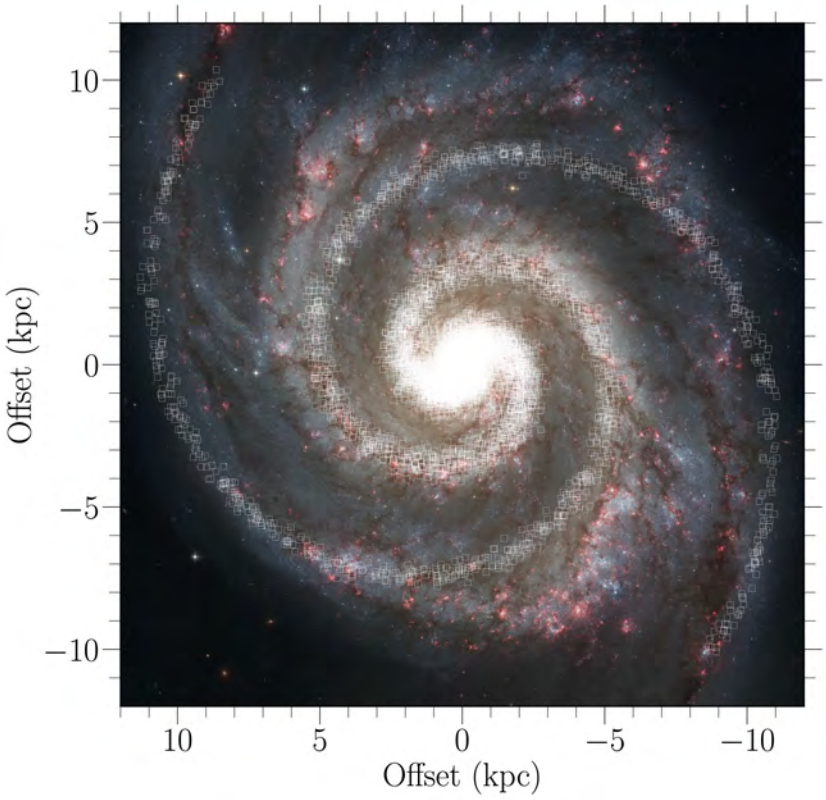


FIGURE 4.2: Two-part spatial configuration used in the galaxy-in-a-box model mapped onto the image of M51 from NASA's Hubble Space Telescope (credit: NASA, ESA, S. Beckwith (STScI) and the Hubble Heritage Team (STScI/AURA)). The M51 image was scaled to fit within the spatial size settings used in the model. The white squares represent the location of stellar clusters along the spiral arms.

being adapted to the galactic model. One of the major changes is the spatial configuration defined in the cluster model. At a distance of ⩾7.6 Mpc the structure of individual clusters is practically unresolvable (1″ corresponds to ∼ 40 pc). Therefore, the spatial component for the galactic model was discarded. Moreover, we used a novel distribution of protostellar ages following Kristensen and Dunham (2018). We describe all of the relevant changes and upgrades motivated by scaling up the cluster model in greater detail in the following paragraphs. First we describe the spatial distribution applied in the galaxy model (Sect. 4.2.2), then we define the molecular cloud mass distribution (Sect. 4.2.2), and from there we go to the age distribution (Sect. 4.2.2).

Spatial distribution

The spatial distribution of GMCs, in which young clusters form, in the galaxy-in-a-box model follows Ringermacher and Mead (2009):

$$r(\phi) = \frac{A}{\log\left(B \tan \frac{\phi}{2N_S}\right)}. \tag{4.1}$$

Here A is a scale parameter for the entire structure, while B and N_S determine the spiral pitch. This formula assumes that all galaxies have "bars" hidden within a bulge. Increasing the N value results in tighter winding, and increasing B in greater arm sweep and smaller bars and/or bulge. To emulate

M51 we adopted the following values: $A = 8.0$, $B = 1.0$, and $N_S = 8.26$. To obtain long spiral arms that wrap around each other, we chose an angle coverage ϕ of 500 degrees. We also introduced a direct scaling parameter $S = 1.5$ to shift spiral arms closer together, toward the galaxy center, without altering their spatial setups. This is especially useful to simulate a central bulge within a galaxy. The parameter is designed to be added at the end of Eq. 4.1. The values were chosen to fit a galaxy with a $\sim$ 23 kpc diameter, which is roughly equivalent to the estimates of the M51 spatial size (e.g., Jarrett *et al.*, 2003). Figure 4.2 illustrates the quality of our fit.

We built our radial distribution of stellar clusters by utilizing an exponential decline of stellar surface density $\Sigma_{\rm star}$ with radius R as

$$\Sigma_{\rm star} = \exp(-R/h_R), \tag{4.2}$$

where h_R is a characteristic scale-length. Here the exponential radial distribution corresponds to a probability density function for the location of stellar clusters along the spiral arms, which are then randomly located according to this function. We follow Casasola *et al.* (2017) and use the value $h_R = 2.38$ pc in this study.

The density distribution of stars in M51 resembles a skewed normal distribution (Scheepmaker *et al.*, 2009). Therefore, the model initially assigns a given stellar cluster a randomly generated location along the spiral arm, and then a random position along the cross section of the spiral arm given by the skewed normal distribution. Studies show that the gas and dust density in galaxies typically decrease as a function of the radius from the center (e.g., Bianchi, 2007; Hunt *et al.*, 2015). Along with the stationary density wave predicting an age gradient across the arms, this decrease implies that star formation activity preferentially occurs in a narrowing band of the spiral arms. To simulate this effect, the standard deviation associated with the skewed normal distribution is scaled as a function of the distance from the center:

$$\sigma = (2 + 0.5r)^{-1} \, . \tag{4.3}$$

This σ value was arbitrarily chosen based on a qualitative good fit with observations of star-forming regions in M51 (Koda *et al.*, 2011).

Molecular cloud mass distribution

In the galaxy-in-a-box model, the initial number of GMCs is specified and then each GMC is randomly assigned a mass following the molecular cloud mass distribution. The latter is described by the molecular cloud mass probability density function (PDF):

$$\frac{{\rm d}N}{{\rm d}M} \propto M^{\alpha} \, . \tag{4.4}$$

We adopt a value of the slope, $\alpha = -1.64$ following Roman-Duval *et al.* (2010). This value is in a good agreement with other Galactic studies of the GMCs, clouds, and clumps (e.g., Solomon *et al.*, 1987; Urquhart *et al.*, 2014). However, this power-law slope was derived for molecular clouds with masses between 10^5 $M_\odot$–10^6 $M_\odot$. Therefore, we assume that lower masses follow a similar slope and so we can use this α value for our study, where

we utilize this relation for the mass range 10^4 $M_\odot$–10^6 $M_\odot$. Estimates of extragalactic α show that this value is probably not constant among galaxies, and there have been reports of variations reaching $\alpha \sim -3.0$ and estimates of the average $\alpha \sim -2.0$ (e.g., Rosolowsky, 2005; Guszejnov *et al.*, 2018; Mok *et al.*, 2020). We evaluate the impact of different α values on the model in Sect. 4.3.1.

Subsequently, we use the mass distribution obtained with Eq. 4.4 to calculate the size of each molecular cloud. Here we follow the recent estimate of the mass–size relation for Galactic GMCs from Lada and Dame (2020):

$$R = 3.3 \times 10^{-3} \text{ pc} \left(\frac{M}{M_\odot}\right)^{0.51}. \tag{4.5}$$

To account for the fact that not all of the molecular cloud mass is converted to stellar mass, we assign a star formation efficiency, $\varepsilon_{\rm SF}$, to determine the total mass of the stellar population from the molecular cloud mass. In the model we apply $\varepsilon_{\rm SF} \sim 10\%$ for embedded clusters following Lada *et al.* (2010).

Age distribution

The characteristic timescale associated with star-forming regions is the free-fall timescale $t_{\rm ff}$,

$$t_{\rm ff} = \sqrt{\frac{3\pi}{32G\rho}}, \tag{4.6}$$

where ρ is the density of the cluster calculated as the total mass of the progenitor molecular cloud divided by the volume of the cloud. The free-fall time reflects the time required for a medium with negligible pressure support to gravitationally collapse. Here, we utilize this timescale to determine a lifetime of the clusters. However, not all of the molecular reservoir will undergo gravitational collapse. Recent studies find that $\varepsilon_{\rm SF}$ per $t_{\rm ff}$ remains constant among different molecular clouds (e.g., Pokhrel *et al.*, 2021). To account for this inefficiency and its influence on the efficiency of $t_{\rm ff}$, we impose a scaling factor $\tau_{\rm ff}^{\rm sc}$. In this study we set the standard value of this factor to be 1. We also assume a constant free-fall time for the entire cluster.

To assign a random age to the cluster we scale $t_{\rm ff}$ with the chosen $\tau_{\rm ff}^{\rm sc}$, and subsequently choose random values ranging between 0 (newly formed) and 1 (completely collapsed). The assigned ages are used to calculate the star formation rate, given by

$$\lambda_{\rm SF} = \frac{N(t)}{t}, \tag{4.7}$$

where $N(t)$ is the number of stars at time t, which is the current age of the cluster calculated from the free-fall time. Here, we make an assumption that $\lambda_{\rm SF}$ is constant for the entire cluster.

To assign the ages to protostars and determine their distributions within clusters, we follow Kristensen and Dunham (2018) and adopt a novel age distribution module. We start with the assumption that protostellar evolution is sequential, it begins at Class 0 and then goes through Class I, flat-spectrum, Class II, and ends at Class III. Then, with the constant star formation rate and

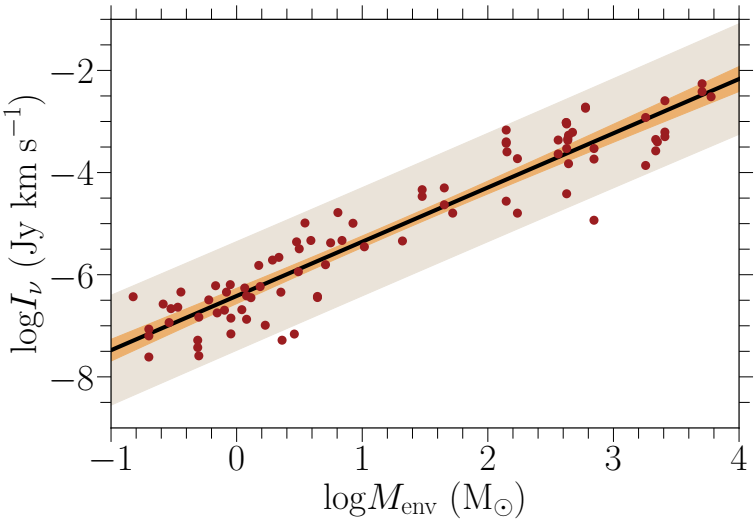


FIGURE 4.3: Water emission at 998 GHz vs. envelope mass ($M_{\rm env}$) for objects from WED used in the simulations. Colors as in Fig. 4.5.

protostellar half-lives, sequential decay is applied. This decay, associated with protostars going through the evolutionary stages, is characterized by the decay constant $\lambda_{\rm D}$, where D represents the protostellar class. Values of $\lambda_{\rm D}$ for each evolutionary stage are estimated based on the observations of seven Galactic clouds (for further details, see Kristensen and Dunham, 2018). With this, we calculate the fractional population of stars in each evolutionary class for all galactic clusters.

4.2.3 *Water Emission Database*

Our model relies on archival water observations. Thus, as a part of this project, we created the Water Emission Database (WED). The main goal of creating this database is to gather in one place all of the available water data, from both ground-based observatories and the *Herschel* Space Observatory, and to make it publicly available. This way the data serves the scientific community. The database is stored and maintained using the MySQL Database Service. However, access to the data is granted through regularly updated ASCII and CSV files available online and is independent of the database driver for safety measures.

Data from many Galactic surveys and observational projects are included in WED, for example Water In Star-forming regions with *Herschel* (WISH; van Dishoeck *et al.*, 2011); the William Herschel Line Legacy Survey (WILL; Mottram *et al.*, 2017); and Dust, Ice, and Gas in Time (DIGIT; Green *et al.*, 2013). Ultimately, the database will also include extragalactic observations of water emission. The values that we store are particularly useful for this study. For example, we focused on water fluxes and parameters describing source properties. This means that we not only store the values from specific studies, but we also keep a unified system of parameters important to characterize the sources. Currently, WED covers 79 observed water transitions up to the para-H_2O $9_{19} - 8_{08}$ transition at 5280.73 GHz (56.77 μm). Emitting sources

TABLE 4.1: Overview of the most important global parameters in the galaxy-in-a-box model.

Category	Parameter	Description	Standard value	Ref.
Star formation	x	power-law slope for the high-mass end of IMF, i.e., for stars with masses $> 1M_{\odot}$; 0 $\rightarrow$ standard, $x = -2.3$; 1 $\rightarrow$ top-heavy, $x = -1.3$; 2 $\rightarrow$ bottom-heavy, $x = -3.3$	0	1
	$\varepsilon_{\rm SF}$	star formation efficiency	10%	2
	$\tau_{\rm ff}^{\rm sc}$	free-fall time scaling factor	1	...
	α	power-law slope of molecular cloud mass distribution	-1.64	3
	$N_{\rm CL}$	number of simulated clusters	10^4	...
	$M_{\rm GMC}$	minimum and maximum masses of progenitor giant molecular clouds	$10^4 M_{\odot} \leqslant M_{\rm GMC} \leqslant 10^6 M_{\odot}$	...
Morphology	A	galactic scaling factor	8.0	...
	B	galactic arms sweep	1.0	...
	$N_{\rm S}$	spiral winding number	8.26	...
	ϕ	angular coverage of spiral arms	500°	...
	R_g	galactocentric radius	12 kpc	4
	h_R	characteristic scale-length	2.38 pc	5
Observational	D	distance to the galaxy	7.6 Mpc	6
	θ	beam size convolved with galaxy image	$2\farcs55$	...
	$p_{\rm size}$	pixel size	$0\farcs51$	...
	dim	image size in pixels	1280 x 1280	...

References: (1) Chabrier (2003); (2) Lada and Lada (2003); (3) Roman-Duval *et al.* (2010); (4) Ciardullo *et al.* (2002); (5) Casasola *et al.* (2017); (6) Jarrett *et al.* (2003).

TABLE 4.2: Description of WED table columns

Column	Description
obs_id	Ordinal number of the input
object	Name of the object
obj_type[a]	Emitting object type
ra_2000	RA (J2000)
dec_2000	Dec (J2000)
transition	Observed water transition
freq[b]	Rest frequency of the observed transition
telescope	Name of the telescope used in the observations
instrument[c]	Instrument used in the observations
obs_res	Resolution (″)
distance	Distance to the object (pc)
luminosity	Bolometric luminosity ($L_\odot$)
tbol[c]	Bolometric temperature (K)
menv[c]	Envelope mass ($M_\odot$)
vlsr[c]	Velocity (km s^{-1})
flux	Observed water flux
flux_err[c]	Flux error
unit	Unit of the observed flux (K km s^{-1}; W cm^{-2}; W m^{-2}; erg $s^{-1}cm^{-2}$)
ref[d]	Reference to the flux measurement(s)
extra	Other relevant information

Notes: (a) Object types currently in use: YSO - young stellar object, IM - intermediate-mass, LM - low-mass, IR-q - IR-quiet, HM - high-mass, mIR-q - mIR-quiet, HMPO - high-mass protostellar object, HMC - hot molecular core, UCHII - ultra-compact HII region, C0 - Class 0, CI - Class I, CII - Class II, PS - possible pre-stellar core, PDR - photodissociation region. Classification is based on the source papers; (b) All of the frequencies to corresponding transitions are taken from the LAMDA database (Schöier *et al.*, 2005); (c) When available; (d) If more than one flux measurement is available, then the most recent or commonly used one is provided with the references to the remaining ones being stored in this column.

at these transitions include the whole range of Galactic protostellar sources, with the majority of low-mass protostars.

The database holds the data in tables arranged in 20 columns (see Table 4.2) and shares them in the form of CSV and ASCII files available online on the project website[2]. All of the files that are available for download are fully described and updated whenever there is a change in the database. The galaxy-in-a-box model downloads the data directly from the website, which makes access to the model completely independent from the restricted MySQL server.

[2] https://katarzynadutkowska.github.io/WED/

For the purpose of this work, we use a very particular subset of WED. We chose the data for the para-H_2O $2_{02}-1_{11}$ line at 987.927 GHz. This water line is among the brightest H_2O transitions observed toward Galactic star-forming regions. Furthermore, it is not a ground-state transition, and so it only mildly suffers from self-absorption even toward high-mass objects (van der Tak *et al.*,

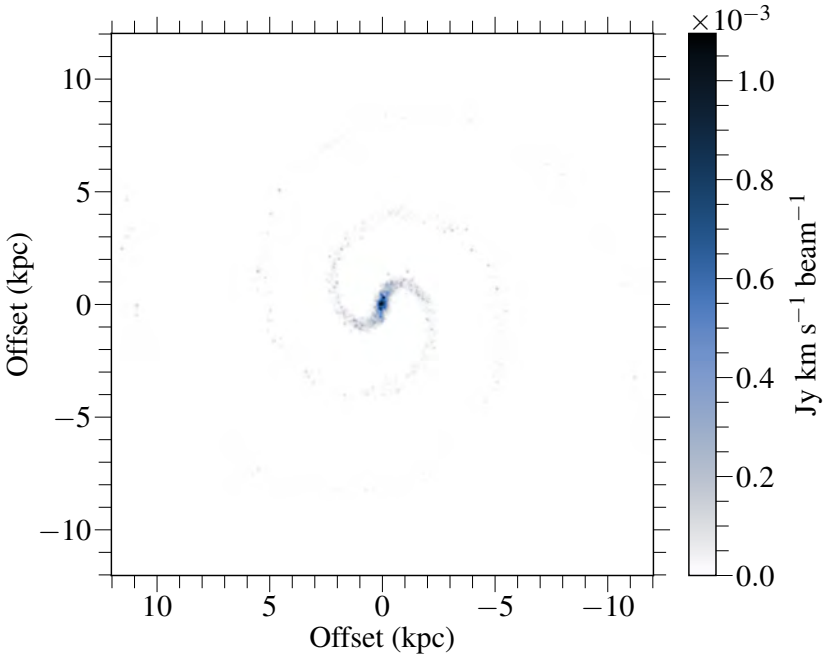


FIGURE 4.4: Example integrated intensity map of the template galaxy with the standard setup.

2013). Finally, this transition is routinely observed toward extragalactic and even high-z objects (e.g., Yang *et al.*, 2016; Yang *et al.*, 2017; Jarugula *et al.*, 2019). The data available in WED for this particular line cover the whole range of sources, and therefore gives a broad overview of water emission. San José-García *et al.* (2016) identified an intensity–envelope mass relation for this line, $\log L_{\mathrm{H_2O}} = (-2.91 \pm 0.10) + (1.19 \pm 0.05) \cdot \log M_{\mathrm{env}}$, which we also observe for the data used in this study (see Fig. 4.3).

As mentioned, the emission assignment utilizes the relationship between the line intensity and envelope mass. At first, Class 0 and Class I objects are assigned with a stellar mass sampled from the IMF. Then we subsequently convert the stellar masses to envelope masses by assuming the envelope mass corresponds to 3× and 1.5× stellar mass for Class 0 and I protostars, respectively (e.g., André *et al.* (2010), and for a more in-depth discussion Offner *et al.* (2014)). Following this, by using the intensity–envelope mass relation, we assign outflow emission to these deeply embedded protostars. We build this relation for the para-H_2O $2_{02} - 1_{11}$ line data from the WISH and WILL samples. The observed intensities are distance-normalized to get a distance-independent measurement. To assess the goodness of fit of the correlation in our regression model, we examined its R-squared value, which in this case corresponds to 89%, indicating a strong relationship between envelope mass and intensity. We derived the correlation as

$$\log I_\nu \ (\mathrm{Jy\ km\ s^{-1}}) = -6.42 \pm 0.08 + (1.06 \pm 0.04) \cdot \log M_{\mathrm{env}}(\mathrm{M_\odot}) \ , \quad (4.8)$$

where the intensity is normalized to the distance of M51 (i.e., 7.6 Mpc). From the above correlation we see that there is a near-proportionality between I_ν and M_{env}, $I_\nu \propto M_{\mathrm{env}}$.

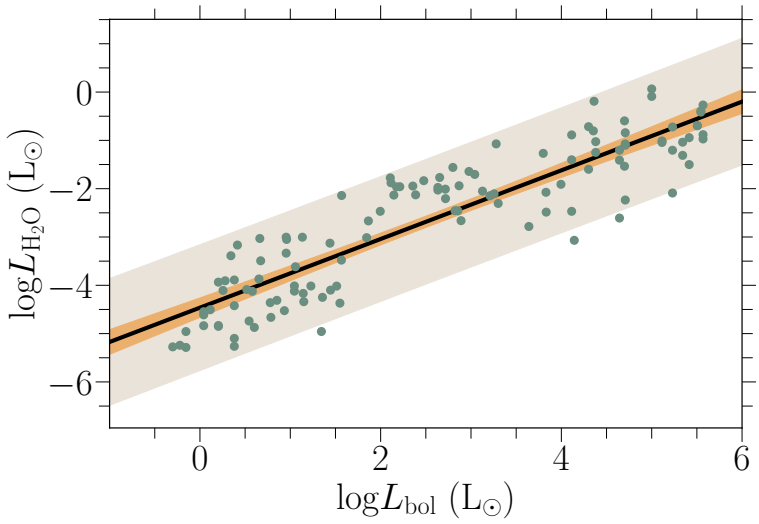


FIGURE 4.5: Water line luminosity at 988 GHz vs. bolometric luminosity for objects from WED used in the simulations, with inclusion of a few additional sources that were excluded from the simulations due to lack of M_{env} data (San José-García, 2015). The solid black line shows the best-fit proportionality, the orange filled region corresponds to the 95% confidence region of the correlation, and the shaded red region represents the region that contains 95% of the measurements.

4.3 RESULTS

With the default galactic and star formation parameters described in Sect. 4.2.2 and gathered in Table 4.1, we obtain an integrated intensity map of the desired molecular emission, as well as mass, total emitted emission, and number of stars of each star-forming cluster within the simulated galaxy. An example integrated intensity map for the model with default parameters is presented in Fig. 4.4. With the chosen spatial setup, most of the emission comes from the innermost parts of the galaxy where the bulge is located, and here individual clusters are not resolved with the applied beam size of 2.″55 (see Table 4.1). The farther from the bulge, the lower the emission and the easier it is to resolve clusters within spiral arms, although the surface brightness of course also decreases.

To explore the impact of the global star formation parameters on the expected emission from clusters in a simulated galaxy as well as the galaxy itself, we conducted a parameter-space study. The changes in parameters were set with respect to the standard model configuration (Table 4.1). We focused on the variations caused by the most important global SF-related parameters: (i) α, describing the slope of molecular cloud mass distribution; (ii) ε_{SF}, the star formation efficiency; (iii) τ_{ff}^{sc}, the free-fall scaling parameter; and (iv) the power-law slope for the high-mass end of IMF. For each change in parameters, we run ten simulations to derive the average of the predicted emission, while for the standard setup we decided on 30 model runs to lower the variations in the derived values. The choice of running ten simulations was motivated by cutting down on the computational time, and it is enough to show the variability in the model outcomes. We explored the cumulative impact of these parameters on the total galactic emission, radial profiles of

TABLE 4.3: Simulation results for different molecular cloud mass distributions

α	$\tilde{I}$ [Jy km s^{-1}]	$I_{\rm tot}$ [Jy km s^{-1}]	IQR [Jy km s^{-1}]
-1.50	1.91×10^{-3}	7.04×10^{1}	5.91×10^{-3}
-1.64	1.86×10^{-3}	7.02×10^{1}	5.85×10^{-3}
-2.00	1.72×10^{-3}	6.61×10^{1}	5.40×10^{-3}
-2.90	1.51×10^{-3}	5.91×10^{1}	4.68×10^{-3}

Notes: Results from running ten simulations per model configuration: α - power-law slope of the molecular cloud mass distribution, ($\tilde{I}$) - median flux, ($I_{\rm tot}$) - total galactic emission, IQR - midspread

the emission maps, and distributions of emitted flux by the galactic clusters. As we show below, these seem to be consistently skewed. Therefore, we chose median values as a measure of central tendency and explored the spread of these distributions with the interquartile range method (IQR or midspread), providing information on the middle 50% of values with the median being in the center of the range.

4.3.1 *Molecular cloud mass distributions*

The standard value of α is set to -1.64 (Roman-Duval *et al.*, 2010). Different studies report a spread in α depending on the studied regions (e.g., Solomon *et al.*, 1987; Rosolowsky, 2005; Mok *et al.*, 2020), and following these studies we explore the change in expected emission for $\alpha = -1.5$, -2, and -2.9. The highest α follows the steepest index reported by Rosolowsky (2005). To investigate this impact we compared the distributions of flux emitted by the clusters and radial profiles of galactic emission.

We observe no apparent variations in the expected emission caused by the change in α. It is true both for the flux distributions and for the mean radial profiles (Fig. 4.6). However, looking at the values obtained for the molecular cloud mass distribution (see Table 4.3) we see a clear trend, indicating that with increasing α, the median flux, the total galactic emission, and the interquartile range increase. This result is consistent with the nature of the corresponding mass distributions as the steeper the slope, the more emission comes from low-mass clusters, which in turn lowers the total observed emission.

4.3.2 *Initial mass function*

In the model we adopted three types of IMF based on the Chabrier (2003) IMF form for young clusters and disk stars. By introducing changes in the slope of the high-mass end of the IMF (x), which applies to stars with $M_\star > 1 M_\odot$, we defined bottom-heavy and top-heavy forms. With the standard value of $x = -2.3$, the slope for the bottom-heavy IMF is defined as $x - 1$, while for the top-heavy it is $x + 1$. This is a purely empirical parameterization, although it is in reasonable agreement with studies reporting x values for bottom-heavy and top-heavy IMF forms (for a recent review, see Smith, 2020).

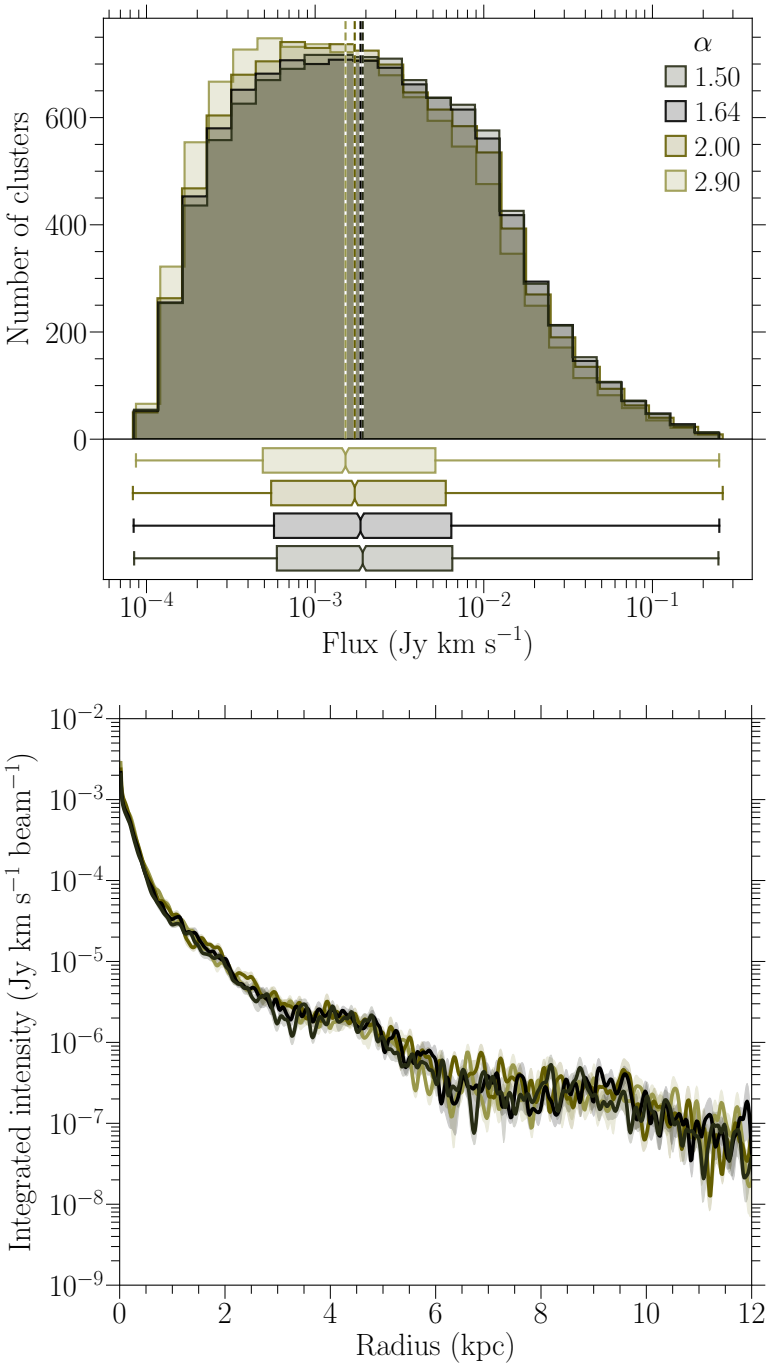


FIGURE 4.6: Results of multiple model runs with varying power-law slopes α of molecular cloud mass distributions. **Top.** Distributions of cluster emission derived for simulations with different power-law slopes α of molecular cloud mass distributions. The vertical dashed lines correspond to the median flux of each distribution. In the bottom box plot the interquartile ranges are presented. The notches in the boxes indicate the 95% confidence intervals around the median. The whiskers spread to the beginning (0%) and the end (100%) of the distributions. These are the mean distributions from a series of ten simulations for each varying parameter. **Bottom.** Radial profiles of emission from the galaxies of the corresponding α values. The radial profiles were calculated from the center of the galaxy all the way to its outskirts. The solid lines correspond to the mean profiles derived from ten simulations, while the shaded regions represent the spread of the mean values based on their standard deviations.

There is no apparent difference in the examined values for any of the

TABLE 4.4: Simulation results for different IMF configurations

IMF	$\tilde{I}$ [Jy km s^{-1}]	I_{tot} [Jy km s^{-1}]	IQR [Jy km s^{-1}]
top-heavy	2.51×10^{-3}	8.57×10^{1}	7.87×10^{-3}
standard	1.86×10^{-3}	7.02×10^{1}	5.85×10^{-3}
bottom-heavy	1.81×10^{-3}	6.90×10^{1}	5.73×10^{-3}

Notes: Results from running ten simulations per model configuration: IMF - form of the initial mass function, ($\tilde{I}$) - median flux, (I_{tot}) - total galactic emission, IQR - midspread

TABLE 4.5: Simulation results for different star formation efficiencies

$\varepsilon_{\mathrm{SF}}$	$\tilde{I}$ [Jy km s^{-1}]	I_{tot} [Jy km s^{-1}]	IQR [Jy km s^{-1}]
1%	1.79×10^{-4}	6.96	5.78×10^{-4}
3%	5.45×10^{-4}	2.10×10^{1}	1.76×10^{-3}
10%	1.86×10^{-3}	7.02×10^{1}	5.85×10^{-3}
30%	5.43×10^{-3}	2.11×10^{2}	1.75×10^{-2}

Notes: Results from running ten simulations per model configuration: $\varepsilon_{\mathrm{SF}}$ - star formation efficiency, ($\tilde{I}$) - median flux, (I_{tot}) - total galactic emission, IQR - midspread

IMF types (see Table 4.4), although it is clear that our top-heavy IMF model tends to produce slightly more emission than the bottom-heavy model. We discuss this further in Sect. 4.4. The lack of dominance of any IMF type is also true for the mean radial profiles of galaxies depicted in Fig. 4.7. Here, we see that neither around the inner part of spiral arms nor around their outer parts do any of the considered IMF types take over the emission, and the radial profiles are indistinguishable.

4.3.3 *Star formation efficiencies*

We probed the impact of $\varepsilon_{\mathrm{SF}}$ on emission outputs by varying its values from 1% to 30%. The outputs vary strongly between different $\varepsilon_{\mathrm{SF}}$ values with a clear trend of increasing flux with $\varepsilon_{\mathrm{SF}}$, as seen in Fig. 4.8. The difference between the highest and the lowest values roughly corresponds to one order of magnitude for all of the considered values. Moreover, we see that the shape of the distribution does not vary significantly across different $\varepsilon_{\mathrm{SF}}$ values; instead, higher $\varepsilon_{\mathrm{SF}}$ merely translates distributions to higher flux values. This way, for the lowest $\varepsilon_{\mathrm{SF}}$ = 1% we derived the total galactic emission of 6.96 Jy km s^{-1}, while one order of magnitude higher $\varepsilon_{\mathrm{SF}}$ = 10% results in an increase in the same parameter of approximately one order of magnitude, giving 7.02×10^{1} Jy km s^{-1}. In addition to the total galactic emission I_{tot}, this trend holds for the median fluxes $\tilde{I}$, and for the midspreads, and it is clear that the multiplication of $\varepsilon_{\mathrm{SF}}$ on average corresponds to the same multiplication of flux (see Table 4.5).

From mean radial profiles (see Fig. 4.8) it is also clear that the increase

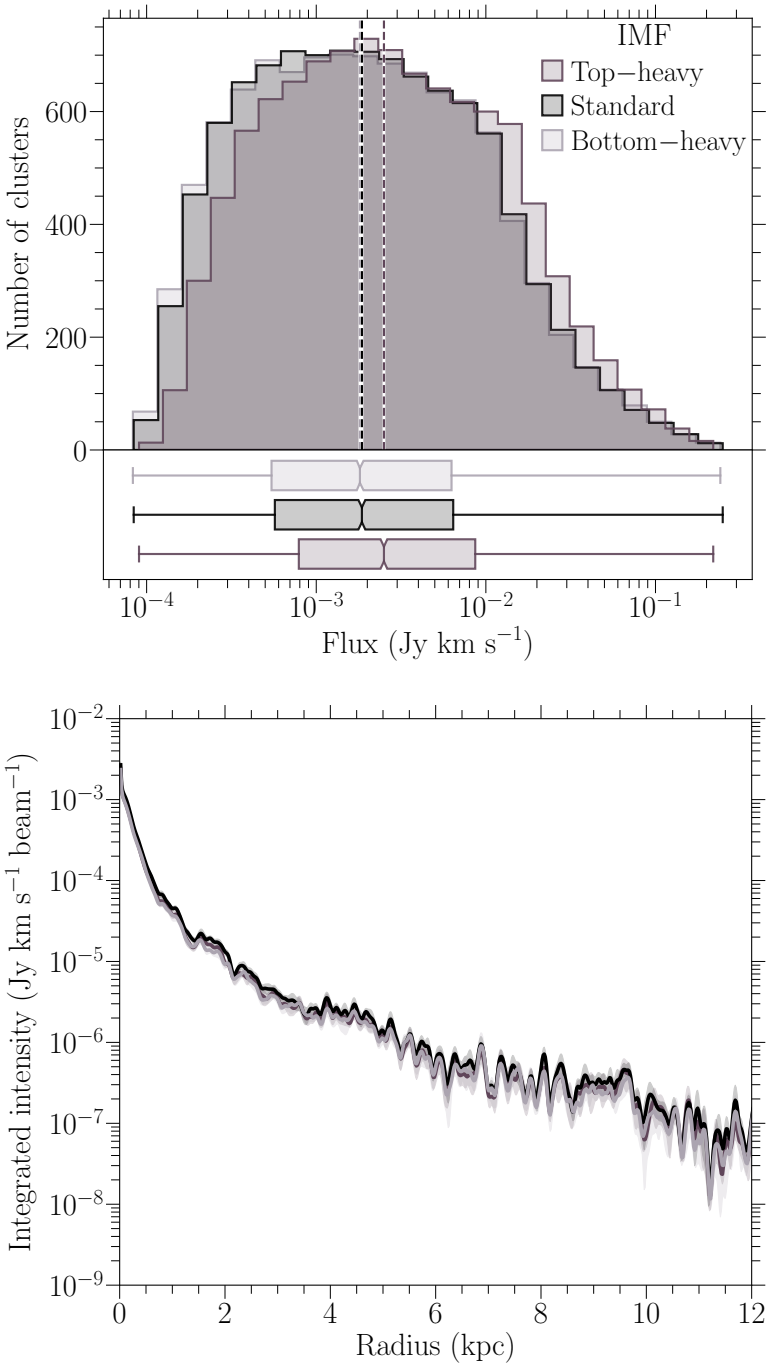


FIGURE 4.7: As in Fig. 4.6, but for different IMF forms.

in the ε_{SF} value results in a subsequent increase in average emission from the galaxy. Here the highest differences in intensities are also around one order of magnitude. Therefore, the higher the ε_{SF}, the more emission comes from spiral arms at different points of the radius. In addition, for ε_{SF} = 1% and ε_{SF} = 3%, the drop in emission in the outermost parts of the galaxy results in higher variations and a more significant drop in the observed emission.

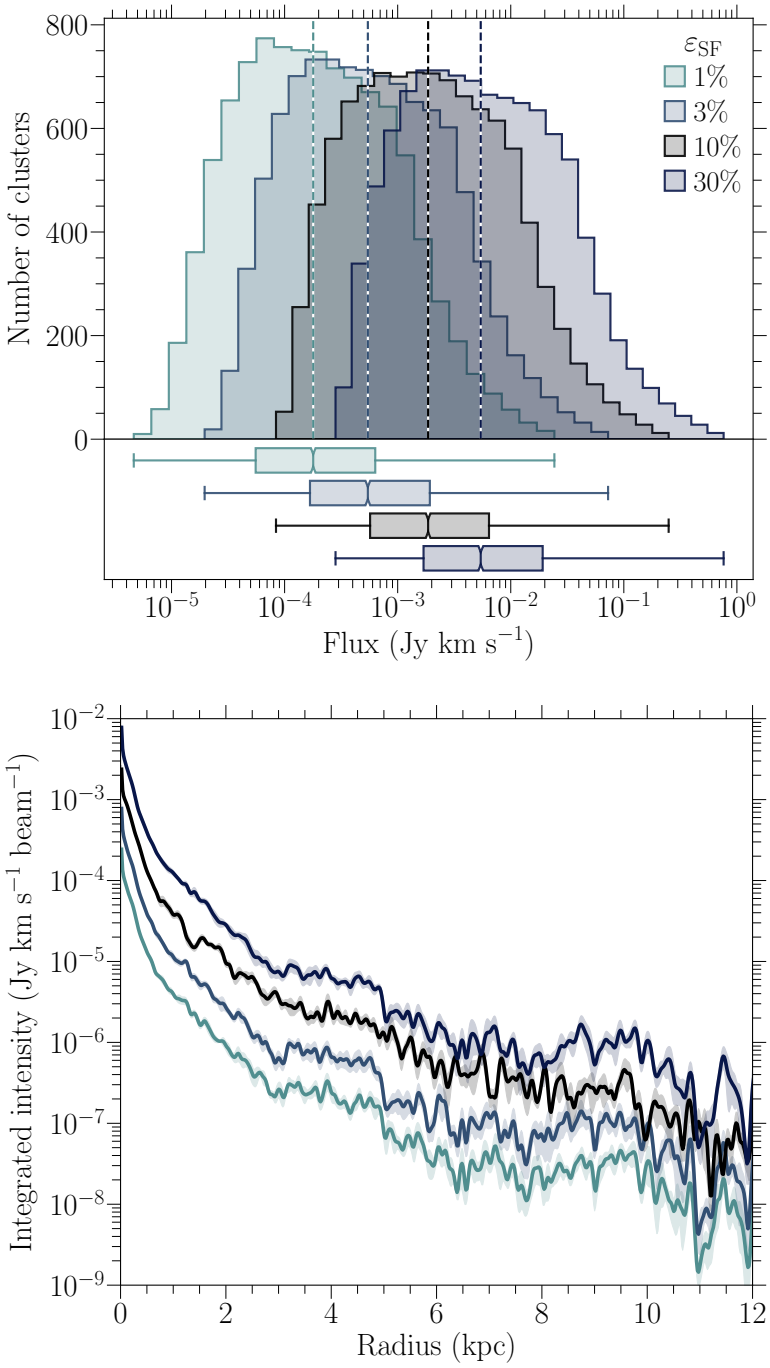


FIGURE 4.8: As in Fig. 4.6, but for the varying $\varepsilon_{\rm SF}$.

4.3.4 *Free-fall-time scaling*

We studied the impact of the free-fall time in the form of $\tau_{\rm ff}^{\rm sc}$ by adopting values ranging from $\tau_{\rm ff}^{\rm sc} = 0.5$ to $\tau_{\rm ff}^{\rm sc} = 5.0$. The scaling factor introduced in this study represents how many free-fall times it takes to form most of the stellar population in a single cluster and relates to the free-fall time efficiency as $\epsilon_{\rm ff} = 0.9 \frac{M_*}{M_{\rm tot}} \frac{t_{\rm ff}}{t_{\rm form}} = 0.9 \frac{M_*}{M_{\rm tot} \tau_{\rm ff}^{\rm sc}}$, following Da Rio *et al.* (2014), where they estimated the time required to form 90% of stars in the cluster.

TABLE 4.6: Simulation results for different free-fall time scaling factors

$\tau_{\rm ff}^{\rm sc}$	$\tilde{I}$ [Jy km s^{-1}]	$I_{\rm tot}$ [Jy km s^{-1}]	IQR [Jy km s^{-1}]
0.5	3.36×10^{-3}	1.12×10^{2}	1.03×10^{-2}
1.0	1.86×10^{-3}	7.02×10^{1}	5.85×10^{-3}
3.0	6.06×10^{-4}	2.94×10^{1}	2.00×10^{-3}
5.0	3.64×10^{-4}	1.91×10^{1}	1.21×10^{-3}

Notes: Results from running ten simulations per model configuration: $\tau_{\rm ff}^{\rm sc}$ - free-fall time scaling factor, ($\tilde{I}$) - median flux, ($I_{\rm tot}$) - total galactic emission, IQR - midspread

Therefore, with this choice of the $\tau_{\rm ff}^{\rm sc}$ values, we evaluate the impact of the free-fall time efficiencies spreading over one order of magnitude, between $\epsilon_{\rm ff} \sim 0.01 - 0.1$.

We observe a very distinct relation between emitted flux and $\tau_{\rm ff}^{\rm sc}$ values, namely that with decreasing $\tau_{\rm ff}^{\rm sc}$ the observed total flux increases. Moreover, decreasing $\tau_{\rm ff}^{\rm sc}$ is associated with condensation of flux distributions, which get both narrower and flatter, and are shifted toward higher flux values (see Fig. 4.9). The lowest $\tau_{\rm ff}^{\rm sc}$ results in a median flux value that is one order of magnitude higher than that derived for the highest $\tau_{\rm ff}^{\rm sc}$ (see Table 4.6). In addition, the beginnings of each distribution are shifted by one order of magnitude from $\sim 10^{-5}$ to $\sim 10^{-4}$ Jy km s^{-1} for the highest and lowest $\tau_{\rm ff}^{\rm sc}$, respectively.

From the radially averaged flux from galaxies with different $\tau_{\rm ff}^{\rm sc}$ we see a trend similar to that for varying $\varepsilon_{\rm SF}$ values. The flux profile from different model outcomes divides into distinguishable pairs for $\tau_{\rm ff}^{\rm sc} \leq 1$ and $\tau_{\rm ff}^{\rm sc} > 1$, although the differences stop being prominent at the galactic outskirts where the flux is the weakest. Here, in particular, the profiles for $\tau_{\rm ff}^{\rm sc} = 3$ and 5 get blended and cause major fluctuations of more than two orders of magnitude in the observed flux.

4.3.5 *Total galaxy emission*

We calculated the integrated galactic emission for model outcomes with varying parameters (Fig. 4.10). The total integrated flux $I_{\rm tot}$ was calculated from the mean flux distributions, and for the standard setup it is equal to 7.02×10^{1} Jy km s^{-1}.

From Fig. 4.10 we see that only two $I_{\rm tot}$-values significantly exceed the default model outcome. The highest value of $I_{\rm tot}$ is observed for simulations with $\varepsilon_{\rm SF} = 30\%$ and is equal to $I_{\rm tot} = 2.11 \times 10^{2}$ Jy km s^{-1}. The second highest value comes from the setup with $\tau_{\rm ff}^{\rm sc} = 0.5$ with 1.12×10^{2} Jy km s^{-1}. For the varying α the highest total emission is derived for $\alpha = -1.5$ and falls almost at the same level as the output from the standard model. A similar thing happens for the top-heavy IMF, which exceeds the default $I_{\rm tot}$ by 1.56×10^{1} Jy km s^{-1}.

The most visible changes are for the outputs that fall below the standard threshold. Here we observe that the lowest total emission output is derived

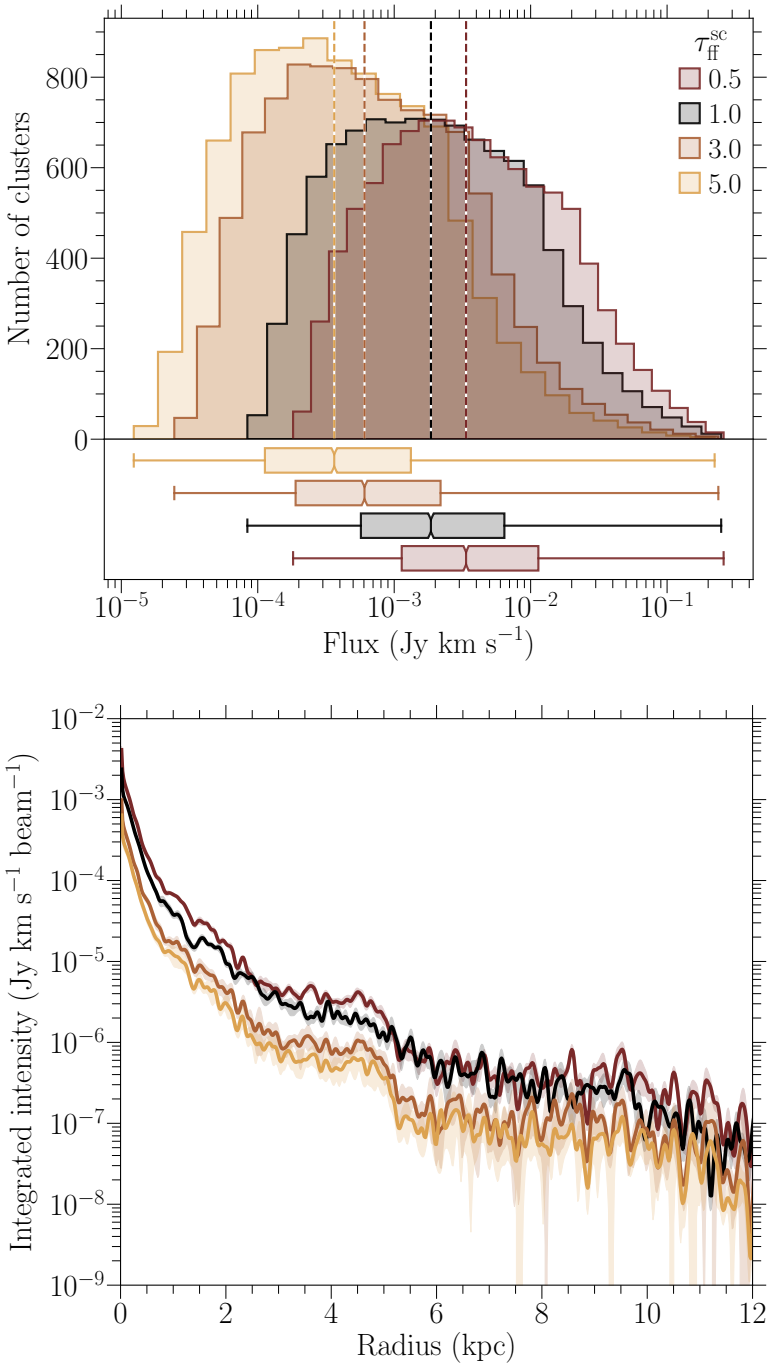


FIGURE 4.9: As in Fig. 4.6, but for galaxies with different $\tau_{\rm ff}^{\rm sc}$.

for the setup with the lowest $\varepsilon_{\rm SF}$ resulting in a one order of magnitude drop in $I_{\rm tot}$ = 6.96 Jy km s^{-1}. Subsequently, the second lowest value is a result of setting $\tau_{\rm ff}^{\rm sc}$ to 5.0 with $I_{\rm tot} = 1.91 \times 10^1$ Jy km s^{-1}. However, the second lowest value of $\varepsilon_{\rm SF}$ results in a very similar result with $I_{\rm tot} = 2.10 \times 10^1$ Jy km s^{-1}. Therefore, these two parameters have the biggest impact on emission and show the highest spread in derived $I_{\rm tot}$ values, while the lowest impact is observed for changes introduced to the molecular cloud mass distribution with the α index.

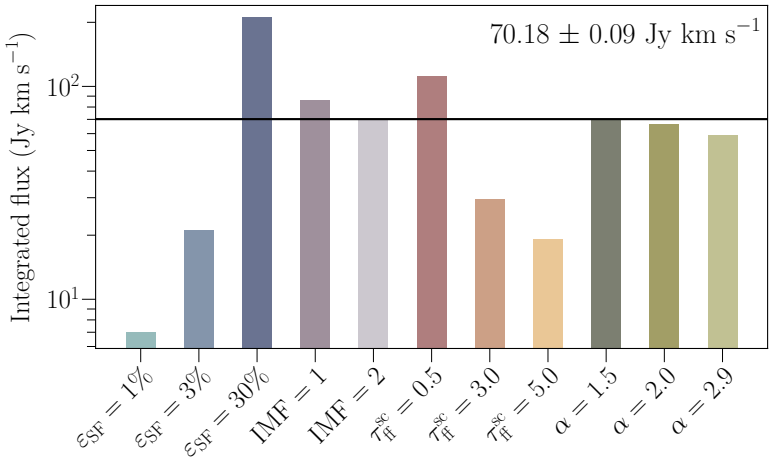


FIGURE 4.10: Total galactic emissions derived from all of the clusters averaged over ten simulations for each setup. The dashed black horizontal line corresponds to the standard setup described in Table 4.1.

4.4 DISCUSSION

In the following we discuss model outcomes and their possible explanations. We also evaluate the impact of different star formation parameters and compare the joint effect of the most influential ones. Moreover, we focus on addressing the question of what other star formation-associated processes, not incorporated into the current version of the galaxy-in-a-box model, could influence the results. Finally, we explore the implications of this proof-of-concept study for observations.

4.4.1 *Varying molecular cloud mass distributions*

Molecular cloud mass distributions are usually described by a single power-law function (Eq. 4.4). Some studies (e.g., McKee and Williams, 1997; Rosolowsky, 2005) propose truncated power-law distributions. However, when the truncated distribution applies, the cut-off point usually lies outside the mass range considered in this study (i.e., for $M_{GMC} > 10^6 M_\odot$). The mass distribution can be expressed either in differential form, as in this work, or in cumulative form with $\alpha > -1$ (Heyer and Dame, 2015). Many Galactic surveys report $\alpha > -2$ (e.g., Solomon *et al.*, 1987; Roman-Duval *et al.*, 2010; Urquhart *et al.*, 2014), while even steeper distributions are found in the outer parts of the Milky Way and in extragalactic regions, with $-2.9 < \alpha \leqslant -2$ (e.g., Rosolowsky, 2005; Guszejnov *et al.*, 2018; Mok *et al.*, 2020). The α index indicates whether most of the mass is contained in high-mass ($\alpha > -2$) or low-mass clouds ($\alpha < -2$).

We evaluated the impact of α on the predicted emission. It appears that steeper distributions result in lower medians and lower total fluxes (see Figs. 4.6 & 4.10). For the standard setup with $\alpha = -1.64$, we see a clear difference when comparing these outcomes to $\alpha = -2.9$. For these, the median values differ by 3.44×10^{-4} Jy km s^{-1}, where the IQRs are narrower by 1.17×10^{-3} Jy km s^{-1} for $\alpha = -2.9$. This small yet potentially observable

level of discrepancy means that the model could distinguish the molecular cloud distributions for slopes with a difference on the order of ~1.

This effect of lowered values with increasing steepness of the mass distribution is somewhat expected. Steeper distributions result in a greater number of molecular clouds with lower masses and produce smaller star-forming clusters. The greater number of low-mass clusters in turn emit less and thus lower total galactic emission, and this is what we see in Fig. 4.6.

Comparing the impact of molecular cloud mass distribution and IMF, as these two seem to have the smallest impact on the predicted emission, we see that the standard and bottom-heavy IMFs result in median fluxes similar to molecular cloud mass distributions with $\alpha \geqslant -2$. However, the most bottom-heavy form of the molecular cloud mass distribution stands out, similarly to the top-heavy IMF. Therefore, when conducting observational comparisons to model outputs, putting constraints on the slope of α, at least for its most varying values, or IMF shape, may be required to fine-tune the model and obtain better agreement with the observations.

4.4.2 *IMF constraints*

The parameterization of the IMF varies between studies, where the used format and high-mass cutoff differs between objects and redshifts (e.g., Chabrier, 2003; Hoversten and Glazebrook, 2008; van Dokkum, 2008; Smith, 2020); the standard form is as follows: $N(M)\mathrm{d}M \propto M^{-2.35}$ (Salpeter, 1955). For more bottom-heavy IMF parameterizations, more low-mass stars are formed, while more top-heavy distributions lead to the presence of more high-mass stars.

In this study we followed a widely used form of the IMF, the Chabrier IMF (Chabrier, 2003), and adjusted it so it roughly represents the three main functional forms: standard, bottom-heavy, and top-heavy. As the building blocks of our model are molecular clouds from which individual star-forming clusters form, the IMF directly influenced the stellar mass distribution of each cluster and emission component. By studying variations on these local building blocks and large galactic scales, we see no significant variations imposed by the different IMF forms. However, for the standard IMF we see that the top-heavy distribution results in a slight increase in emission, while the opposite happens after adopting the bottom-heavy distribution. This result is expected. On the one hand, low-mass protostars dominate star formation in total mass and number (Kroupa, 2001). The size of this population is increased or decreased for the bottom- and top-heavy IMFs, respectively. On the other hand, high-mass protostars are far more energetic than low-mass ones. Moreover, with $\int I_\nu \propto {M_\mathrm{env}}^1$ water is a low-contrast mass tracer. Hence, the more massive the envelope, the higher the emission.

When comparing results obtained for different IMF forms, we also see that the total flux obtained for the bottom-heavy IMF is very similar to that derived for the standard IMF. These two are also very similar when we consider their flux distributions and radial profiles as seen in Fig. 4.7. The same for their IQRs. Therefore, the model cannot distinguish these from one another. The top-heavy IMF, on the other hand, seems to differ when it comes to the IQR and the range spanned by the flux distribution. However,

the variation is in the range of $5.73 - 7.87 \times 10^{-3}$ Jy km s^{-1} for IQR and only $1.81 - 2.51 \times 10^{-3}$ Jy km s^{-1} for $\tilde{I}$. Nevertheless, this is the only IMF form that could be necessary to fine-tune the model when comparing it with observations.

Looking at the total flux plot in Fig. 4.10 we see that the output for the standard and bottom-heavy IMFs is comparable to other outputs derived for molecular cloud mass distributions for which α was set to -1.5 and 2.0. The only difference between these setups can be seen in the shapes of their radial profiles; however, this may be not significant enough to distinguish these distributions from each other.

4.4.3 *Effect of star formation efficiency*

The star formation efficiency describes the amount of molecular gas that ends up in stars. The increase in $\varepsilon_{\rm SF}$ directly translates to an increase in the number of (proto)stars, which results in more emission from clusters. Different values of $\varepsilon_{\rm SF}$ are reported toward different types of sources across cosmic time, varying from 3% in nearby molecular clouds to 30% in Galactic embedded clusters (Lada and Lada, 2003) and extragalactic GMCs (Dessauges-Zavadsky *et al.*, 2019). In this work, the impact of $\varepsilon_{\rm SF} > 30\%$ is not evaluated, as $\varepsilon_{\rm SF}$ is closely related to the gas depletion time, and with higher $\varepsilon_{\rm SF}$ molecular gas is used at a higher rate and is sustained for a shorter time.

Analyzing the impact of $\varepsilon_{\rm SF}$ on the expected emission locally and on a galactic scale, we observe a clear and systematic increase in emission with increasing $\varepsilon_{\rm SF}$. The observed increase in emission is roughly proportional to the increase in $\varepsilon_{\rm SF}$. There is a shift of the flux distributions as seen in Fig. 4.8. The IQRs follow the same trend, and are in the range $\sim 6 \times 10^{-4} - 2.0 \times 10^{-2}$ Jy km s^{-1}. This suggests that the model can be used to distinguish different values of $\varepsilon_{\rm SF}$, at least when no other parameter varies.

Distributions drawn from model outputs with varying $\varepsilon_{\rm SF}$ show significant variations when considering all of the analysis, which is also true for the impact of $\tau_{\rm ff}^{\rm sc}$. However, these two parameters significantly differ when it comes to the shape of the flux distributions and radial profiles. Therefore, it should be possible to evaluate which parameter could explain the observed galactic emission.

4.4.4 *Influence of the free-fall time scaling factor*

The last considered parameter is the free-fall time scaling factor $\tau_{\rm ff}^{\rm sc}$. Here we arbitrarily chose all of the values to explore how altering the ages of clusters could affect the expected emission. With $\tau_{\rm ff}^{\rm sc} < 1$ we effectively lower the ages of protostars within the cluster, and therefore increase the contribution from Classes 0 and I. Therefore, with lower $\tau_{\rm ff}^{\rm sc}$ values we would expect higher emission both globally and locally.

From the flux distributions and radial profiles in Fig. 4.9 we see that there is indeed an increase in flux with the decrease in $\tau_{\rm ff}^{\rm sc}$. Moreover, all of the distributions tend to flatten with this decrease. We also observe a peculiar shape of the distribution derived for the smallest $\tau_{\rm ff}^{\rm sc}$. The possible explanation for this peculiar shape is that such a huge change in free-fall

time results in constraints on the age distribution of clusters within galaxies. It is also the distribution with the higher median, which indicates a greater number of Class 0 sources within clusters, which produce more molecular emission from outflows.

Following Kristensen and Dunham (2018), the fraction of Class 0/I cores decreases with the age of the cloud and reaches a steady state at $\sim$ 0.5 Myr. Therefore, as the scaling of the free-fall time increases, especially when $\tau_{\rm ff}^{\rm sc} \geqslant 1$, clusters more accurately represent the dynamics of stellar formation. This in turn results in a greater range of flux distributions and lower median fluxes, as the fraction of Class 0/I cores decreases.

The outcome for $\tau_{\rm ff}^{\rm sc} = 5.0$ is similar to that for $\varepsilon_{\rm SF} = 3\%$, when considering the cumulative galactic flux shown in Fig. 4.10. Nevertheless, the difference between these outputs is potentially observable, especially that $\tau_{\rm ff}^{\rm sc} = 5.0$ gives a flatter flux distribution. Therefore, the model could distinguish the emission for these global parameters.

4.4.5 *Interplay of the most influential parameters*

The most influential parameters in the model are $\tau_{\rm ff}^{\rm sc}$ and $\varepsilon_{\rm SF}$. Thus, to understand and explore the combined effect of these parameters on simulated galaxies we ran the model for all of the possible combinations of the values considered in this study. Then we evaluated the outcomes of these simulations by calculating the distributions of cluster fluxes and their corresponding midspreads (see Fig. 4.11) and galactic radial profiles (Fig. 4.12). Moreover, we color-coded the results of each simulation based on the integrated intensities of the flux distribution. The heat map with corresponding integrated fluxes is presented in Appendix 4.A.

The distribution of fluxes changes according to what we observed when studying the impact of $\tau_{\rm ff}^{\rm sc}$ and $\varepsilon_{\rm SF}$ separately, namely that median flux and integrated intensity within galaxies increases with increasing $\varepsilon_{\rm SF}$ and decreasing $\tau_{\rm ff}^{\rm sc}$. Interestingly, $\varepsilon_{\rm SF}$ seems to mainly influence the median flux by shifting the distribution toward higher flux values proportionally to the magnitude of the increase. In addition, the shift is not associated with any significant changes in the shape of the distributions. On the other hand, $\tau_{\rm ff}^{\rm sc}$ increases the median fluxes, but does not shift the whole distribution. With the decrease in $\tau_{\rm ff}^{\rm sc}$ the distributions flatten and, based on the midspreads, the high-flux tail seems to catch up with the rest of the distribution. Subsequently, there is a decrease in the spread of observed flux values. The lower-flux part of the distribution shifts toward higher flux values, but it does not affect the highest flux values.

The changes observed on galactic scales also reveal complex outcomes of the interplay of these parameters. Here we observe that $\varepsilon_{\rm SF}$ basically scales the radial profiles up and increases the level of each emission point, especially in the inner galaxy where most of the clusters reside. It also influences the visibility of the spiral arm bumps in the radial profiles. Surprisingly, these bumps are more prominent with the increase in the free-fall time scaling factor. However, this change is also associated with the increased radial profile variability.

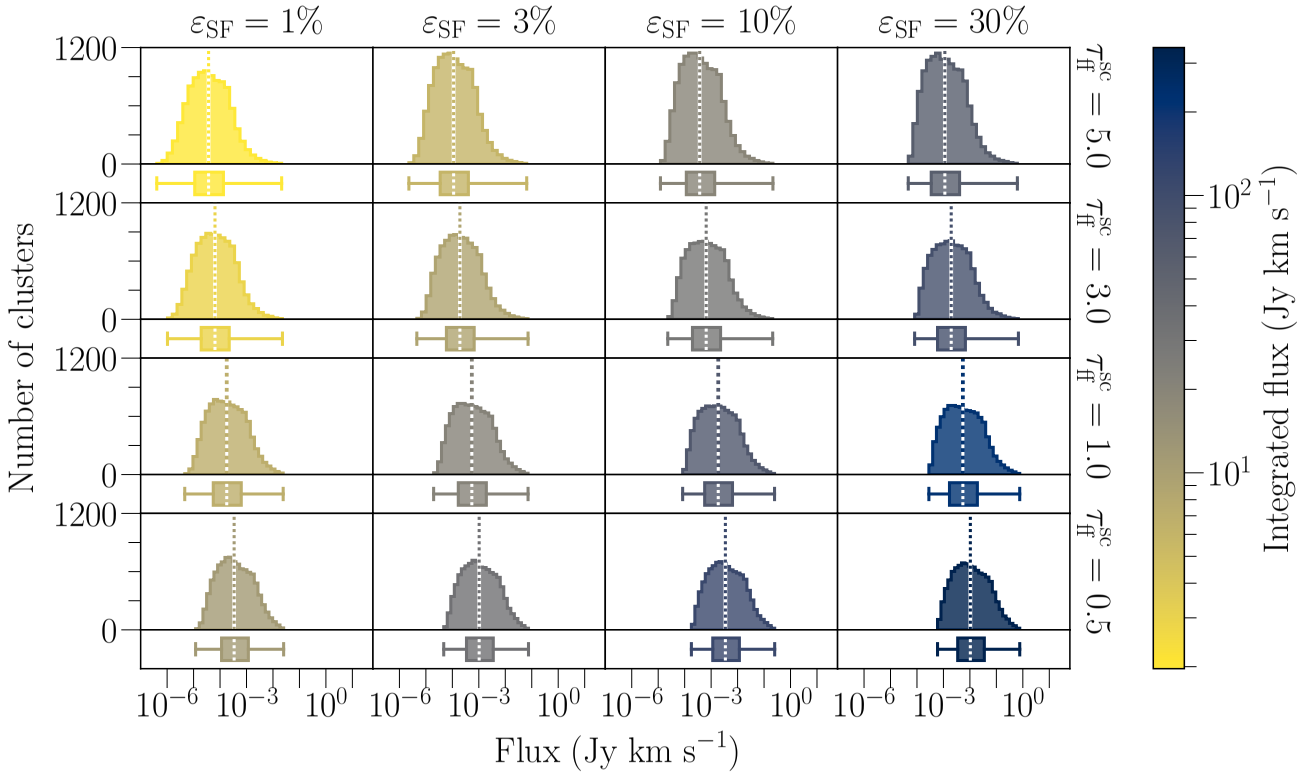


FIGURE 4.11: Distributions of cluster emission derived for simulations with varying free-fall time scaling factor τ_{ff}^{sc} and star formation efficiency ε_{SF}. The vertical dashed lines correspond to the median flux of each distribution and a measure of central tendency in form of IQR is presented in the bottom box plots with whiskers spreading from the beginning (0%) to the end (100%) of each distribution. These are the mean distributions from a series of ten simulations for each varying pair of τ_{ff}^{sc} and ε_{SF}. The y- and x-axes have the same ranges for all rows and columns. The color-coding is based on the integrated fluxes of each distribution. The exact values of these fluxes are available in Fig. 4.A.1.

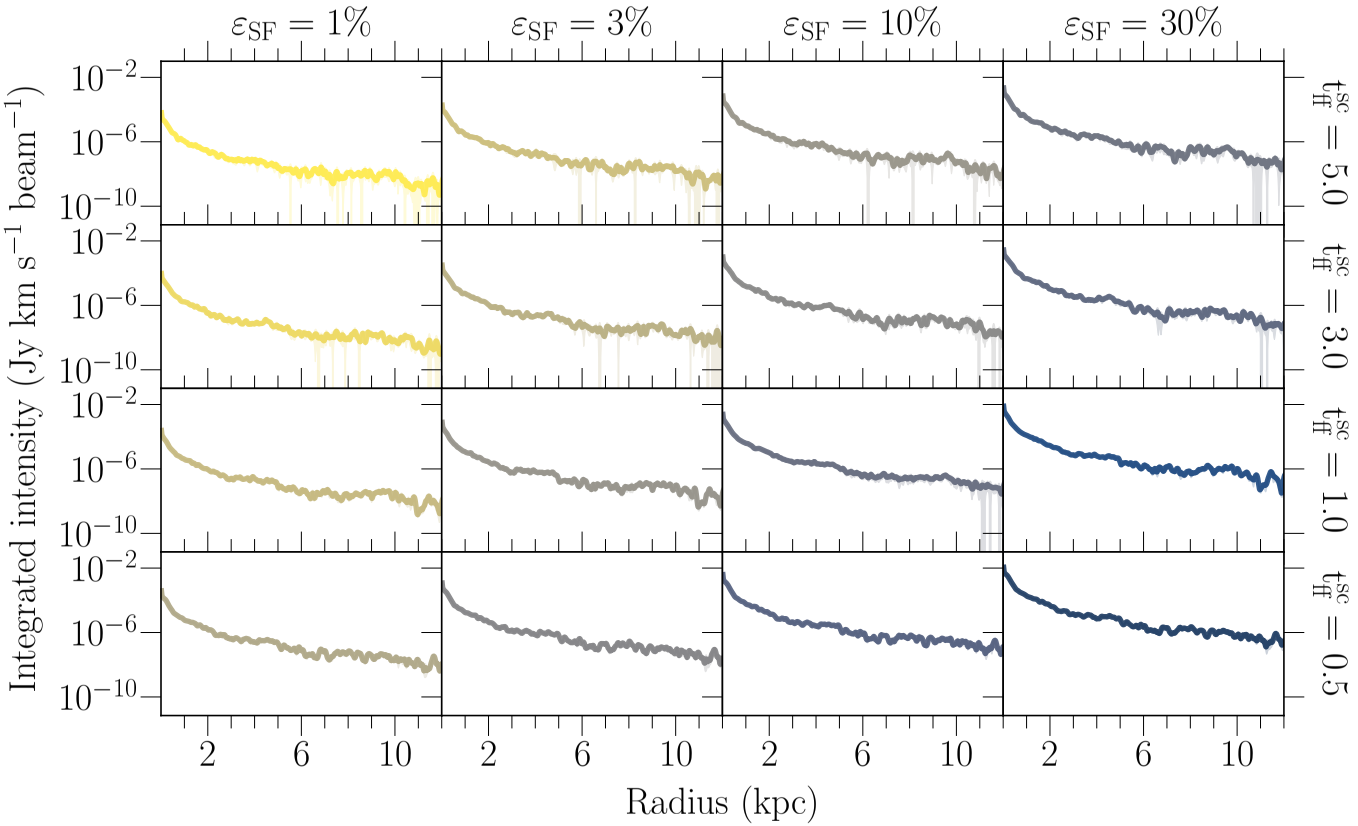


FIGURE 4.12: Radial profiles of emission from the same set of simulations and color-coding as in Fig. 4.11. The solid lines correspond to the mean radial profile, while the shaded regions represent the variability of each profile, based on the standard deviation of each profile.

By looking at the simulations obtained for all of the available combinations, we see that the impact of each parameter is different, and the only common characteristic is a strong influence on the observed emission. From the flux distributions we find that for spatially resolved observations the possible value of each parameter can be estimated because these parameters introduce very distinct features to the shape and properties of each distribution, while in the case of unresolved observations these values can be determined from the features seen in the galactic radial profiles. Therefore, our model can be used to reveal these global star formation parameters or at least indicate which one has the prevalence in increased or decreased star formation activity in a galaxy.

4.4.6 *Other effects*

Other factors could play a role in water excitation. These include the local and global effects of star formation processes and galactic evolution and structure.

The warm ($\gtrsim$ 100K) and dense ($\gtrsim 10^6 \mathrm{cm}^{-3}$) inner regions of protostars, known as hot cores, exhibit conditions that support the presence of a rich chemistry (Herbst and van Dishoeck, 2009). Under these conditions all water ice should sublimate, and the observed gaseous water abundances should match the expected water ice abundances. However, the observations do not follow these expectations, showing that most of the observed gaseous water is universally associated with warm outflowing and shocked gas, with a negligible contribution from hot cores (van Dishoeck *et al.*, 2021b). The low surface brightness of the hot cores along with the high dust opacity at 1 THz obscuring the hot cores makes them practically invisible in a *Herschel* beam (Visser *et al.*, 2013; Herpin *et al.*, 2012).

On larger scales the question arises of the emission from molecular clouds themselves. Here, water vapor is generally not detected in the Milky Way (e.g., Dionatos *et al.*, 2020). The only noteworthy exception are the regions exposed to enhanced UV radiation, the so-called photon dominated regions with one narrow emission component (Bjerkeli *et al.*, 2012). However, the overall molecular cloud contribution to the observed water emission is insignificant for the results of this study, particularly for the higher-excited 2_{02}–1_{11} transition. Taking a global view of galaxies, active galactic nuclei (AGNs) could play a role in increasing or decreasing water emission, both locally and globally. Studies report quenching of star formation in AGN host galaxies, which would lower the number of protostars and thus outflows (e.g., Fabian, 2012; King and Pounds, 2015; van Dishoeck *et al.*, 2021b, and references therein). Moreover, AGNs can produce conditions favoring molecular excitation or dissociation if the radiation becomes too strong. The exact influence of the AGN feedback on water excitation is not well understood, but it appears that the presence of AGNs has little impact on the excitation of the water line considered in this study, the para-H_2O $2_{02} - 1_{11}$ line at 987.927 GHz. Specifically, Jarugula *et al.* (2019) spatially resolved H_2O emission in this transition toward the Cloverleaf quasar, which is a strongly lensed AGN, at a resolution of 1 kpc, but found no trend with distance to the actual AGN. Thus, considering AGN feedback would likely

have a negligible effect on the results of this study.

4.4.7 *Implications for observations*

Verification of the model can only be obtained by benchmarking its outcomes against observations. Ideally, these observations should spatially resolve individual star-forming clusters. This way, the cluster flux distribution is compared with a simulated galaxy. To come down to $\sim$ 10 pc scales and spatially resolve the whole range of molecular cloud sizes, the resolution should be on the order of 0.″3 at 7.6 Mpc.

The results presented from our proof-of-concept study are for a resolution of 2.″55, which at 7.6 Mpc corresponds to $\sim$ 70 pc. This resolution is comparable to the resolution at which M51 was observed as part of the PAWS program (Schinnerer *et al.*, 2013), and where individual GMCs are resolved. Therefore, smaller clouds are unresolved in the galactic image. However, only a handful of high-redshift star-forming galaxies are spatially resolved in H_2O emission, although at $\sim$ 1 kpc scales (Jarugula *et al.*, 2019). Most observations do not resolve emission, and comparisons would have to be made based on the total fluxes or water line luminosities rather than on radial profiles or the shape of cluster flux distributions. With this assumption, we can make a tentative comparison of water line luminosities observed toward nearby and distant galaxies with the values derived in this study.

The average total flux of $\sim$ 70 Jy km s^{-1}, corresponding to $\sim$ 1300 $L_\odot$, derived for the simulated galaxies in this study remains approximately one order of magnitude below the luminosity derived for the nearby starburst M82 (Yang *et al.*, 2013), where the star formation rate (SFR) is approximately one order of magnitude higher (e.g., de Grijs *et al.*, 2001) than the Milky Way or M51. The observed luminosities toward several LIRGs (Luminous InfraRed Galaxies) and ULIRGs (Ultra-Luminous InfraRed Galaxies) at larger distances (Yang *et al.*, 2013) or high-z starbursts at $z \sim 2-4$ (e.g., van der Werf *et al.*, 2011b; Omont *et al.*, 2011; Omont *et al.*, 2013; Yang *et al.*, 2016; Jarugula *et al.*, 2019) remain up to $\sim 2-4$ orders of magnitude higher. However, this difference is expected and consistent with the increasing SFRs of these galaxies, especially when considering the high-z galaxies where SFRs often exceed $\sim$ 1000 $M_\odot$/yr, which naturally boosts star formation, and hence the emission coming from the protostellar outflows. However, more comparisons are needed to fully assess the differences between the model and high-redshift observations, but this is beyond the scope of this paper.

There are several ways in which to interpret the difference between the model outcomes and the observations of high-z galaxies. First of all, our template galaxy resembles the nearby M51 galaxy. We chose this particular galaxy instead of a well-studied high-redshift galaxy because we wanted to start with an object with a known molecular cloud distribution (e.g., Hughes *et al.*, 2013; Colombo *et al.*, 2014) as this is one of the building blocks in our model. Second, our results are for a standard IMF (Lada and Lada, 2003); there are indications that IMFs toward high-z galaxies are significantly more top-heavy than those we tested here, which would serve to further boost emission from the high-mass stars. However, this in turn implies that we

are dealing with a different spatial setup, galactic size, and internal galactic environment. This size difference is very prominent, as spatially resolved high-redshift galaxies have radii in the range of 0.95 – 2.24 kpc (Jarugula *et al.*, 2019), while M51 has a radius of ∼ 12 kpc.

On the other hand, there is reasonable agreement between the model results and observations of galaxies that lie closer to M51. Sandqvist *et al.* (2021) reported water flux measurements from the *Herschel* SPIRE observations toward the NGC 1365 galaxy, lying at a distance of 18.6 Mpc (Madore *et al.*, 1998). The observed flux corresponds to 3081.9 Jy km s^{-1}, which falls on the higher end of the fluxes derived for the model results when distance-corrected, and corrected for the SFR in NGC 1365 being approximately one order of magnitude higher than that of the Milky Way and M51. For a nearby starburst, Mrk 231 at a distance of ∼ 200 Mpc (van der Werf *et al.*, 2010), Omont *et al.* (2011) reports a flux of 718 Jy km s^{-1}, which is distance- and SFR-corrected, and also falls at the high end of the simulated fluxes.

It is clear that both the star formation efficiency and the free-fall scaling parameter can affect the H_2O flux dramatically (e.g., Fig. 4.11). A single integrated H_2O flux is not going to constrain either parameter, and additional constraints are needed. To constrain the star formation efficiency, for example, the total number of stars formed combined with the amount of molecular material available should be observed. The former is best constrained through an in-depth look into stellar masses in galaxies, both nearby and at high-redshift. One way to do this is through near- and mid-IR observations, where the James Webb Space Telescope (JWST) will provide a great advance, especially for the high-redshift regime. The molecular material available can be probed either through low-J CO emission or dust emission. Although there are known problems with both tracers (e.g., Pitts and Barnes, 2021), they are the best tracers at the moment for the total gas mass. Thus, with the combination of JWST observations of the stellar mass and, for example, ALMA observations of the total gas mass, the star formation efficiency can be independently constrained.

Another factor to consider is the detailed comparisons of spatially resolved observations with model results, where it would be possible to evaluate which sets of star formation parameters can reproduce the galactic emission. Here, for example, by analyzing the flux distribution of the observed emission (similar to Fig. 4.11), it would be possible to put constraints on these parameters and pinpoint their most probable values.

4.5 CONCLUSIONS

We created a galactic model of emission that could arise from active galactic star-forming regions. In this paper we demonstrated the main principles behind the galaxy-in-a-box model and explored how it can serve as a tool to study and better understand star formation activity in galaxies even at high redshift. For a template galaxy set to resemble the grand-design spiral Whirlpool Galaxy M51, we evaluated the impact of important global star formation parameters on model results. We conducted this parameter space

study for the para-H_2O $2_{02} - 1_{11}$ line at 987.927 GHz. The main results are as follows:

- Emission from the para-H_2O $2_{02} - 1_{11}$ line is a low-contrast tracer of active star formation with $\int I_\nu \propto M_{\rm env}$;
- The initial mass function along with molecular cloud mass distribution have little impact on predicted water emission;
- An increase and decrease in star formation efficiency $\varepsilon_{\rm SF}$ will respectively increase and decrease the predicted emission, both locally and globally;
- With the decrease in free-fall time scaling factor $\tau_{\rm ff}^{\rm sc}$ we observe a corresponding increase in galactic emission and flattening of star-forming flux distribution, which indicates increasing populations of Class 0 and Class I protostars;
- At the moment, further constraints are needed to break model degeneracies; these additional constraints include JWST observations combined with low-J CO observations, and resolved observations of H_2O emission.

A tentative comparison of model outcomes with observational data for high-redshift galaxies yields realistic results and opens new paths to improve the model, so it can become a reliable proxy to reveal star formation in galaxies throughout cosmological times. In the near future we plan to introduce the possibility of turning on or off AGN feedback and to conduct detailed comparisons of model results with observations of local and distant LIRGs, ULIRGs, and HyLiRGs (Hyper-Luminous InfraRed Galaxies). Furthermore, since our model is not designed specifically for water molecules, we intend to explore the results for other unique outflow tracers, like high-J CO ($J \geq 10$). It will be important to constrain which global star formation parameters that have not impacted our results for water emission will behave differently for other molecular tracers.

ACKNOWLEDGEMENTS

The research of KMD and LEK is supported by a research grant (19127) from VILLUM FONDEN. This article has been supported by the Polish National Agency for Academic Exchange under Grant No. PPI/APM/2018/1/00036/U/001.

Appendix

4.a INTEGRATED FLUX VALUES OF FLUX DISTRIBUTIONS

In order to evaluate the interplay between the most influential parameters, $\tau_{\mathrm{ff}}^{\mathrm{sc}}$ and $\varepsilon_{\mathrm{SF}}$, we color-coded the results from each set of simulations based on the integrated flux values calculated from the cluster flux distributions. We also created a corresponding flux value guide in the form of a heat map presented in Fig. 4.A.1.

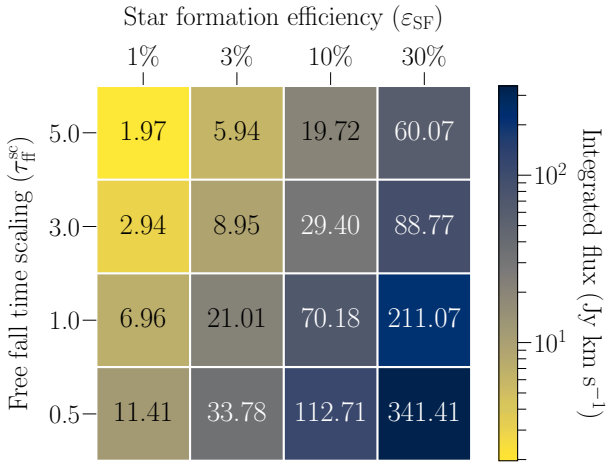


FIGURE 4.A.1: Heat map with integrated flux values calculated from the flux distributions presented in Fig. 4.11.

5

Star formation rate estimates from water emission

K. M. Dutkowska[1], L. E. Kristensen[1]

[1] Niels Bohr Institute, University of Copenhagen, Øster Voldgade 5–7, 1350 Copenhagen K., Denmark

Submitted to Astronomy & Astrophyics

Abstract

Context. The star formation rate quantitatively describes the star formation process in galaxies throughout cosmic history. Current ways to calibrate this rate usually do not employ observational methods accounting for the low-mass end of stellar populations as their signatures are too weak.

Aims. Accessing the bulk of protostellar activity within galactic star-forming regions can be achieved by tracing signposts of ongoing star formation. One such signpost are molecular outflows, which are particularly strong at the earliest stages of star formation. These outflows are bright in molecular emission, which is readily observable. We propose to utilize the protostellar outflow emission and use it as a tracer of the star formation rate.

Methods. In this work we introduce a novel version of the galaxy-in-a-box model, which can be used to relate molecular emission from star formation in galaxies with the star formation rate. We measure the predicted para-water emission at 988 GHz (which is particularly bright in outflows) and corresponding star formation rates for galaxies with $L_{\rm FIR} = 10^8 - 10^{11} {\rm L}_\odot$ in a distance-independent manner, and compare them with expectations from observations.

Results. We evaluated the derived results by varying star-forming parameters, i.e., the star formation efficiency, the free-fall time scaling factor, and the initial mass function. We observe that for the chosen water transition, relying on the current Galactic observations and

star formation properties, we are underestimating the total galactic emission, while overestimating the star formation rates, particularly for more starburst-like configurations.

Conclusions. The current version of the galaxy-in-a-box model only accounts for a limited number of processes and configurations, i.e., it focuses on ongoing star formation in massive young clusters in a spiral galaxy. Thus, the inferred results, underestimating the emission and overestimating the star formation rate, are not surprising: known sources of emission are not included in the model. To improve the results, the next version of the model needs to include a more detailed treatment of the entire galactic ecosystem and other processes that would contribute to the emission. Thus, the galaxy-in-a-box model is a promising step toward unveiling the star-forming properties of galaxies across cosmic time.

5.1 INTRODUCTION

Star formation lies at the very center of the baryon cycle and plays a pivotal role in shaping galactic ecosystems. There are different measures of this process, which all help to understand and characterize its behavior through cosmic history. The star formation rate (SFR) is one of these measures, as it provides a quantitative description of the star-forming properties of a given object by relating the total mass of stars formed in a give time unit, i.e., $M_*/\Delta t$. The star formation rate is used to establish the cosmic star formation history (e.g., Lilly *et al.*, 2013; Madau and Dickinson, 2014), which in turn is used to understand and quantify the evolution of galaxies.

The key epoch of cosmic star formation history ("cosmic noon"), when star formation peaked, marks a critical stage during the evolution of today's galaxy population (e.g., Shapley, 2011; Madau and Dickinson, 2014; Förster Schreiber and Wuyts, 2020). Cosmic noon galaxies, lying at redshifts of 2–3, exhibit star formation rates extremely different from those observed in the local Universe, reaching $> 1000\,M_\odot\,yr^{-1}$, while the Milky Way is forming stars at a rate of $\sim 1\,M_\odot\,yr^{-1}$ (e.g., Kennicutt and Evans, 2012).

There a various ways of deriving SFRs in galaxies from nebular line, UV, infrared, radio, and X-ray emission (Madau and Dickinson, 2014). These methods all assume that there is a scaling between the luminosity in a given band and the SFR. However, the observed emission is usually dominated by high-mass stars, which easily outshine low-mass stars due to their energetic output, and so an initial mass function is applied to correct for low-mass stars, which is where most of the mass is. In the local Universe the SFR is readily traced and calibrated with Hα, Hβ, [O II], and [O III] emission (e.g., Kennicutt, 1998; Tresse *et al.*, 2002; Kewley *et al.*, 2004; Salim *et al.*, 2007; Villa-Vélez *et al.*, 2021). However, in the past 20 years, with advances in astrochemistry we have additional ways to trace star formation, even in its most embedded stages, and thus tracing the episodes of current star formation in galaxies (e.g., Herbst and van Dishoeck, 2009; Jørgensen *et al.*, 2020).

Molecular emission from protostars in not yet a common SFR tracer. Nevertheless, this emission has the potential to trace even low-mass pop-

ulations directly. At the earliest stages, the forming star itself is deeply embedded in gas and dust and thus completely obscured. Thus, the key is to trace signposts of these early stages that are not obscured. One of these signposts are outflows, which are launched from protostars in their main accretion phase, when the interaction between the infalling envelope, winds, and jets launched from the protostar are particularly strong (Bally, 2016). These outflows are launched from close to the protostar, but quickly punch their way through to the surrounding molecular cloud, where they are not obscured (Bachiller *et al.*, 1990). In our Galaxy, one of the best tracers of this protostellar component is water (van Dishoeck *et al.*, 2021b), which is predominantly locked up as ice on dust grains, but is released from the grain mantles into the gas phase, causing a jump in the abundance of many orders of magnitude. At the same time, the physical conditions are conducive to water being readily excited into rotational states (e.g., Suutarinen *et al.*, 2014).

Water emission is also observed toward high-redshift galaxies (e.g., Yang *et al.*, 2013; Yang *et al.*, 2016; Jarugula *et al.*, 2019), where it was also calibrated to serve as a SFR tracer (Jarugula *et al.*, 2019). However, at high-redshift water is thought to trace dusty molecular clouds illuminated by either massive stars or a central galactic nucleus, and thus the excitation is assumed to be via far-infrared (FIR) pumping (e.g., González-Alfonso *et al.*, 2008; González-Alfonso *et al.*, 2014). However, toward the Galactic sources, which were extensively observed with the *Herschel* Space Observatory (e.g., the WISH survey; van Dishoeck *et al.*, 2011; van Dishoeck *et al.*, 2021b), water emission is almost uniquely associated with outflows, where its excitation is collisionally dominated and other processes, such as FIR pumping, have a negligible contribution to the excitation (Mottram *et al.*, 2014; Goicoechea *et al.*, 2015).

With the goal of tracing active star formation in galaxies with molecular emission from protostars, Dutkowska and Kristensen (2022) created a galactic model, the so-called galaxy-in-box model, simulating emission from star-forming regions. Utilizing the state-of-the-art understanding of Galactic star formation and astrochemical observations of Galactic protostars, the galaxy-in-a-box model simulates emission from young clusters in a chosen galaxy, at the same time providing details into the statistics of the star formation process. The default molecular emission is that from water at 988 GHz ($J_{Ka,Kc} = 2_{02} - 1_{11}$), which is readily observed even at high redshifts, where its emission is thought to be dominated by the FIR pumping dominated regions as outlined above.

In this work, we present an extension to the galaxy-in-a-box model, which allows us to derive SFRs from simulated galaxies and their individual star-forming clusters, as well as put constraints on local and global SFRs. We focus on water emission at 988 GHz, and simulate emission for galaxies with $L_{\rm FIR} = 10^8 - 10^{11} \mathrm{L}_\odot$ for varying star formation parameters.

This paper is organized as follows. Section 5.2 describes all of the changes introduced to the galaxy-in-a-box model. Subsequently, in Section 5.3 we present the results of this study and test them against observations and literature, which we then discuss in Section 5.4. Finally, we present our conclusions in Section 5.5.

5.2 METHODS

In this study we explore the relation between the SFR, L_{H_2O} and L_{FIR} using the galaxy-in-a-box model (Dutkowska and Kristensen, 2022, for an overview of the model see Appendix 5.A). This is a novel, top-level astrophysical modeling tool, that simulates emission from young clusters in a galaxy, at the same time providing detailed insights into the constituents of the star formation process and derived parameters. The model relies on relatively few input parameters, giving the user great flexibility to define global and local galactic parameters.

For deriving the SFR and relating L_{FIR} to the virial mass of galaxies, we implemented a number of upgrades to the model, which we describe in Sect. 5.2.1. In Sect. 5.2.2 we describe the choice of parameters for the simulated galaxies.

5.2.1 *Changes to the galaxy-in-a-box model*

For the purposes of this study, we introduced the star formation rate as an input and output parameter in the galaxy-in-a-box model. We only used the output SFRs. However, in the following, we will describe the full extent of the new SFR feature. The star formation rate tell us how much material is turned into the stars per unit time. With that in mind, we defined the star formation rate for a cloud in a galaxy as

$$\begin{aligned} \mathrm{SFR}_{\mathrm{cloud}} &= N_* \left(\frac{\langle M_* \rangle}{\mathrm{M}_\odot}\right) \left(\frac{t_{\mathrm{cloud}}}{\mathrm{yr}}\right)^{-1} \\ &= N_* \left(\frac{\langle M_* \rangle}{\mathrm{M}_\odot}\right) \left(\frac{\tau_{\mathrm{ff}}^{\mathrm{sc}} t_{\mathrm{ff}}}{\mathrm{yr}}\right)^{-1}, \end{aligned} \tag{5.1}$$

where N_* is the number of formed protostars, $\langle M_* \rangle$ is the average protostellar mass, t_{cloud} is the age of the cloud, $\tau_{\mathrm{ff}}^{\mathrm{sc}}$ is the unitless free-fall scaling factor (Dutkowska and Kristensen, 2022), and t_{ff} is the free-fall time of the cloud. In the case of the galaxy-in-a-box model the age is randomized, i.e., it randomly scales the ages such that they range from newly formed to completely collapsed. The global star-forming rate of the entire galaxy, SFR, is thus the sum of the individual rates for each cloud.

In the model, we assume that each cloud goes on to form one cluster; in nature, clouds may go on to form several generations of clusters, however, for the purposes of this study where we consider global star formation, that is not relevant. With this implementation of the star formation rate, we introduced a possibility to constrain the star formation rate also at the cloud or cluster level. The cluster module can now be run with a fixed star formation rate, where the age of the cluster will be adjusted through the free-fall time scaling factor, which can be easily derived from Eq. (5.1):

$$\tau_{\mathrm{ff}}^{\mathrm{sc}} = N_* \left(\frac{\langle M_* \rangle}{\mathrm{M}_\odot}\right) \left(\frac{t_{\mathrm{ff,random}}}{\mathrm{Myr}}\right) \left(\frac{\mathrm{SFR}_{\mathrm{cloud}}}{\mathrm{M}_\odot\,\mathrm{Myr}^{-1}}\right). \tag{5.2}$$

In this equation, t_{ff} is already randomized ($t_{\mathrm{ff,random}}$) to avoid poor SFR adjustment due to age assignment that takes place later in the model. However, Eq. (5.2) is not used in this study.

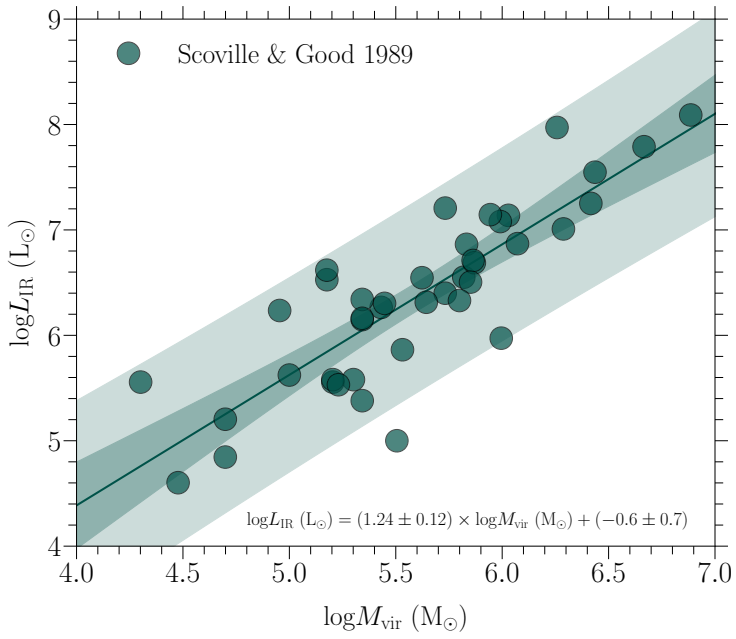


FIGURE 5.2.1: Correlation between log $M_{\rm vir}$ and log $L_{\rm FIR}$ following Scoville and Good (1989). The solid straight line represents the best-fit power-law to the data points, the darker shaded region corresponds to the 95% confidence region of the correlation, and the lighter shaded region represents the region that contains 95% of the measurements.

On a global scale, i.e., when introducing constraints on the total galactic SFR, the new version of the galaxy-in-a-box model monitors the total SFR of the given galaxy and computations stop, when the specified SFR is reached. The allowed deviation from the specified SFR is ±10%. There may be situations where the galaxy-in-a-box model will not converge: these situations are unphysical, and an example would be a very low-mass galaxy with a very high SFR. There needs to be enough gas that can be turned into stars at the desired rate.

One of the changes to the galaxy-in-a-box model that was essential for this study was to set the number of clusters as limited by the total molecular gas reservoir, rather than setting it as a fixed number in the input file. The way we infer the number of molecular clouds, from which star forming clusters form, is by putting an upper mass limit for the total mass of the molecular clouds, which are randomly generated using the molecular cloud mass distribution. When the limit is reached, the clouds are no longer passed to the cluster part of the calculations (see Fig. 1 of Dutkowska and Kristensen, 2022). This way we ensure that the mass of clouds does not exceed the available molecular reservoir.

Lastly, the mass properties of the galaxy can be now set by defining the far-infrared luminosity of the galaxy. Following Scoville and Good (1989) the model derives the mass of the molecular reservoir through the observed $M_{\rm vir} - L_{\rm IR}$ relation (see Fig. 5.2.1). The viral mass of the galaxy can be expressed as:

$$\frac{M_{\rm vir}}{\rm M_\odot} = 10^{0.5\pm0.6}\left(\frac{L_{\rm IR}}{\rm L_\odot}\right)^{0.81\pm0.08}, \tag{5.3}$$

TABLE 5.2.1: Parameters considered in this study

			Galactic type ($\log(L_{\rm FIR}/L_\odot)$)			
			8 – 8.9	9 – 9.9	10 – 10.9	11
Parameter	$\varepsilon_{\rm SF}$	1%	x	x	x	x
		10%	x	x	x	x
		30%	x	x		x
	$\tau_{\rm ff}^{\rm sc}$	1	x	x	x	x
		5	x	x	x	x
	IMF	*s*	x	x	x	x
		t-h	x	x	x	x
		b-h	x	x	x	x

Notes: Star-forming and galactic parameters considered in this study. Red filling corresponds to parameter space used for the high-z correlation test, while x refers to those considered together with the standard correlation. For the latter, the only omitted galactic type was that with $L_{\rm FIR} = 10^{10} L_\odot$ for all combinations involving $\varepsilon_{\rm SF} = 30\%$. In the IMF parameters, *s* refers to standard, *t-h* to top-heavy, and *b-h* to bottom-heavy.

where $L_{\rm IR}$ is the total far-infrared luminosity of the cloud. With Eq. (5.3) we can simulate $L_{\rm H_2O}$ for galaxies with different $L_{\rm IR}$, including typical galaxy-types observed with H_2O emission, i.e., sub-Luminous InfraRed Galaxies (sub-LIRGs; $L_{\rm IR} < 10^{11}$ $L_\odot$), Luminous InfraRed Galaxies (LIRGs; 10^{11} $L_\odot \leq L_{\rm IR} < 10^{12}$ $L_\odot$), UltraLuminous InfraRed Galaxies (ULIRGs; 10^{12} $L_\odot \leq L_{\rm IR} < 10^{13}$ $L_\odot$), and HyperLuminous InfraRed Galaxies (HyLIRGs; $L_{\rm IR} \geq 10^{13}$ $L_\odot$). In this study, we are interested in relative values for derived luminosities and SFRs, and as such we make an assumption that $L_{\rm FIR}$ is a proxy for $L_{\rm IR}$, and we will use these luminosities interchangeably.

5.2.2 *Considered parameters*

The goal of this study is to explore what SFRs we will derive with the galaxy-in-a-box model and how they relate to derived luminosities. To achieve this goal we decided to use the template galaxy from Dutkowska and Kristensen (2022) with emission from the para-H_2O $2_{02} - 1_{11}$ line at 987.927 GHz, and tweak the star formation efficiency, the free-fall time scaling factor as well as the initial mass function. The exact range of probed parameters is described in Table 5.2.1.

For the galactic masses, or in this case luminosities, we decided to probe galaxies with $L_{\rm FIR} = 10^8 - 10^{11}$ $L_\odot$, where for the range $10^8 - 10^{10}$ $L_\odot$ we continued with a increment corresponding to the given order of magnitude (i.e., $10^8, 2\times10^8, 3\times10^8$, etc.), we stopped at 10^{11} $L_\odot$. The reason behind this choice was that, we wanted to probe the chosen regime relatively uniformly. Moreover, the lower limit was dictated by low galactic mass (10^8 $L_\odot$ corresponds to $\sim 10^7$ $M_\odot$), while the upper one was dictated by the limitations of the computational power. As will be shown below, the inferred SFRs can readily be extrapolated to even higher luminosities.

In the model, we use the relation between mass and H_2O line luminosity obtained only from Galactic sources to estimate the amount of emission

generated by protostars. As a sanity check, we can include the high-z observations in this correlation, as shown in Fig. 5.3.1. Including the high-z measurements shift the correlation slightly, such that low-mass protostars would be assigned less emission, and vice versa for high-mass protostars. Thus, if these high-z sources are included, this would have implications for the assumed IMF. For getting luminosity distances for high-z objects, we use a Planck 2018 flat ΛCDM cosmology with $H_0 = 67.7\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$ and $\Omega_M = 0.310$ as implemented in the ASTROPY package (The Astropy Collaboration *et al.*, 2018).

5.3 RESULTS

Extracting SFRs together with $L_{\mathrm{H_2O}}$, while at the same time defining galaxies rather by their luminosities than masses directly, is a way to confront expectations about the star formation process as seen in the Milky Way with the literature, while testing the galaxy-in-a-box model. Therefore, in this proof-of-concept study we ran a number of simulations spanning a range of parameters representing different galactic and star formation properties (see Tab. 5.2.1).

As mentioned in Sect. 5.2.1, we used two different mass-line luminosity correlations: in the first, we only use the Galactic data points, and in the second we include the high-z data points. We excluded certain parameters from the high-z test, because they were either computationally heavy or unnecessary for testing the impact of the high-z extrapolation (for further discussion, see Sect. 5.4). For galaxies with $L_{\mathrm{FIR}} = 10^8 - 9 \times 10^8\ \mathrm{L_\odot}$ we run 40 simulations for each setup, while for other luminosity ranges we run 20 simulations per setup. The increased number of simulations for this specific galactic type was dictated by higher SFR variations, as the molecular reservoir is relatively low, which is reflected in larger variations in the number of formed stars. We also excluded calculations for galaxies with $L_{\mathrm{FIR}} = 10^{10} - 9 \times 10^{10}\ \mathrm{L_\odot}$ that would have $\varepsilon_{\mathrm{SF}} = 30\%$, because they were the most computationally heavy, and including them would not affect any conclusions of this study. In total we ran 15200 simulations, including 12240 main runs and 2960 runs for the high-z test.

Uncertainties for the simulations results are calculated as a standard-deviation from the mean value, derived for all runs with the same set of parameters. The best-fits were obtained using linear regression, while accounting for the spread in the y-direction. If the spread is not shown, it means that the size is smaller than the marker or the line size. When recalculating fluxes to luminosities, we naturally account for the propagation of uncertainties.

We describe the literature sample chosen for this study in Sect. 5.3.1. We then present the results through derived $L_{\mathrm{FIR}} - L_{\mathrm{H_2O}}$ (Sect. 5.3.3), L_{FIR} – SFR (Sect. 5.3.4), and $L_{\mathrm{H_2O}}$ – SFR (Sect. 5.3.5) relations, which we compare to those provided in the literature.

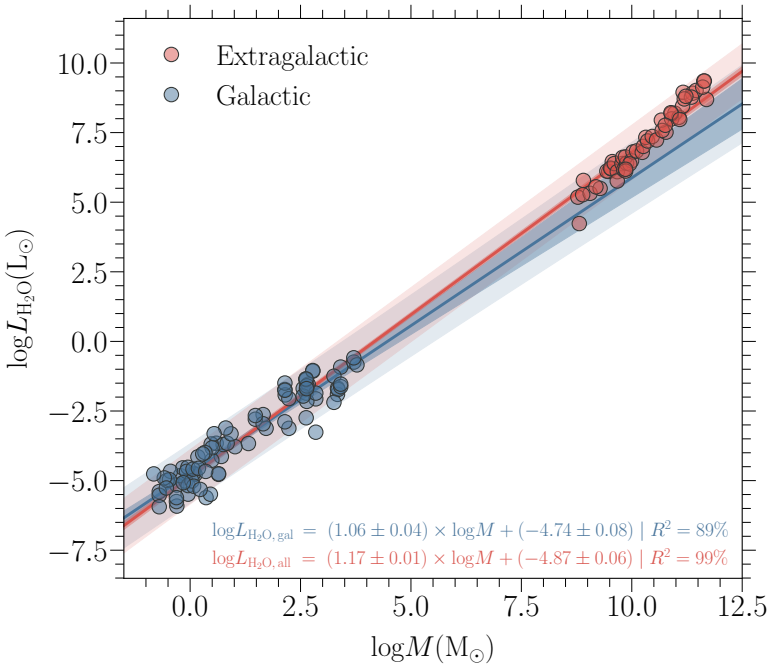


FIGURE 5.3.1: Two types of correlations and observational samples used in this study. The blue solid line corresponds to the best-fit to the Galactic with the data points taken from the Water Emission Database (Dutkowska and Kristensen, 2022), while the red solid line represent the best-fit that also includes the extragalactic sample consisting of nearby sub-LIRGs, LIRGs and quasars, as well as high-z quasars, ULIRGs, and HyLIRGs (for details see Sect. 5.3.1). Markers correspond to the observations from each sample. Shading follows that from Fig. 5.2.1.

5.3.1 *Literature sample*

As a default source of Galactic observations we use data from the Water Emission Database (Dutkowska and Kristensen, 2022) for the para-H_2O $2_{02} - 1_{11}$ line at 987.927 GHz, which consists of Galactic low-, intermediate-, and high-mass protostars observed as a part of the Water In Star-forming regions with Herschel (WISH; van Dishoeck *et al.*, 2011) and the William Herschel Line Legacy Survey (WILL; Mottram *et al.*, 2017).

The sample of extragalactic sources, used in the high-z test, were taken directly from van der Werf *et al.* (2011a), Combes *et al.* (2012), Omont *et al.* (2013), Riechers *et al.* (2013), Yang *et al.* (2013), Yang *et al.* (2016), Apostolovski *et al.* (2019), and Jarugula *et al.* (2019). This sample includes both nearby sub-LIRGs, LIRGs and quasars, as well as high-z quasars, ULIRGS and HyLIRGs, with the farthest one being the HyLIRG, HFLS3, at z = 6.337 (D_L = 62834.75 Mpc; for more details see Riechers *et al.*, 2013). A detailed description of the sample and exact values used in this study can be found in Kristensen et al. subm.

5.3.2 *M_* vs SFR*

We evaluated the derived SFRs by exploring their relation with the total stellar mass of the corresponding galaxies. From Fig. 5.3.2 we see that we are overestimating the SFRs, when looking at functions derived by, e.g. Salmon *et al.* (2015) for the main-sequence galaxies and Rinaldi *et al.* (2022)

for the starbursts.

With the chosen set of properties, galaxies with $M_* < 10^{6.5}$ $M_\odot$ seem to lie close to the main-sequence estimates from Salmon *et al.* (2015), at least in their lower limits. However, going to cases where the combination of considered parameters resulted in increase of star formation rates, especially galaxies with $M_* > 10^{6.5}$ $M_\odot$ we start overestimating SFRs by at least one order of magnitude when compared to the literature (Rinaldi *et al.*, 2022).

We also observe two distinct populations that appear to be dictated by the value of the free-fall time scaling factor. For $\tau_{\rm ff}^{\rm sc} = 1$, we let the efficiency of the free-fall time depend only on the density of the progenitor molecular cloud, while by introducing $\tau_{\rm ff}^{\rm sc} = 5$ we prolong the time required to form most of the stellar population resulting in a more diverse range of protostellar ages. From Fig. 5.3.2 we see how decreasing of the free-fall time scaling factor influences the derived star formation rate. Considering the relatively low efficiency of the star formation process, the lower-SFR branch is likely to be more consistent with the nature of the star formation process. We discuss it further in Sect. 5.4.4.

5.3.3 $L_{\rm FIR} - L_{\rm H_2O}$ *correlation*

To compare the predicted fluxes with observations, we have first converted them to luminosities using the following expression:

$$\frac{L_{\rm line}}{\rm L_\odot} = 99.04 \left(\frac{I}{\rm Jy\,km\,s^{-1}}\right) \left(\frac{\lambda_0}{\mu\rm m}\right)^{-1} \left(\frac{D_{\rm L}}{\rm Mpc}\right)^2, \tag{5.4}$$

where I is the total intensity in Jy km s^{-1}, λ_0 the wavelength in microns (303.4557 μm for the para-H_2O $2_{02}-1_{11}$ line) and $D_{\rm L}$ the luminosity distance of the source in Mpc. By converting fluxes we can quantitatively compare our results with observations, as we are no longer distance dependent.

Using linear regression, we derived best-fit lines to the following expression:

$$\log_{10}\left(L_{\rm H_2O}/\rm L_\odot\right) = a \times \log_{10}\left(L_{\rm FIR}/\rm L_\odot\right) + b. \tag{5.5}$$

In Tab. 5.3.1 we provide all of the derived slopes and intercepts. In the following we focus on the two setups exhibiting the highest and lowest water emission. These are the models with $\varepsilon_{\rm SF}$=30%, IMF = top-heavy, $\tau_{\rm ff}^{\rm sc}$=1, and $\varepsilon_{\rm SF}$=1%, IMF = bottom-heavy, $\tau_{\rm ff}^{\rm sc}$=5, respectively. For the least emitting case we derive $a = 0.809 \pm 0.003$ and $b = -7.269 \pm 0.029$, while for the most emitting case we derive $a = 0.809 \pm 0.001$ and $b = -5.135 \pm 0.012$. In both cases $R^2 = 99.9\%$. For all of the simulations the slope stays roughly constant with $a \approx 0.81$, hence the span of luminosities is described by the intercept falling in the range of -7.269 to -5.135. We can derive the general relation for water line luminosity depending on the intercept value:

$$L_{\rm H_2O}/\rm L_\odot = 10^b \left(L_{\rm FIR}/\rm L_\odot\right)^{0.81}. \tag{5.6}$$

From Fig. 5.3.3 we see that we deviate from extragalactic observations by a factor of a few to ~ 2 orders of magnitude. We observe that the expectations built on the extragalactic sample, taken from Jarugula *et al.* (2019), where

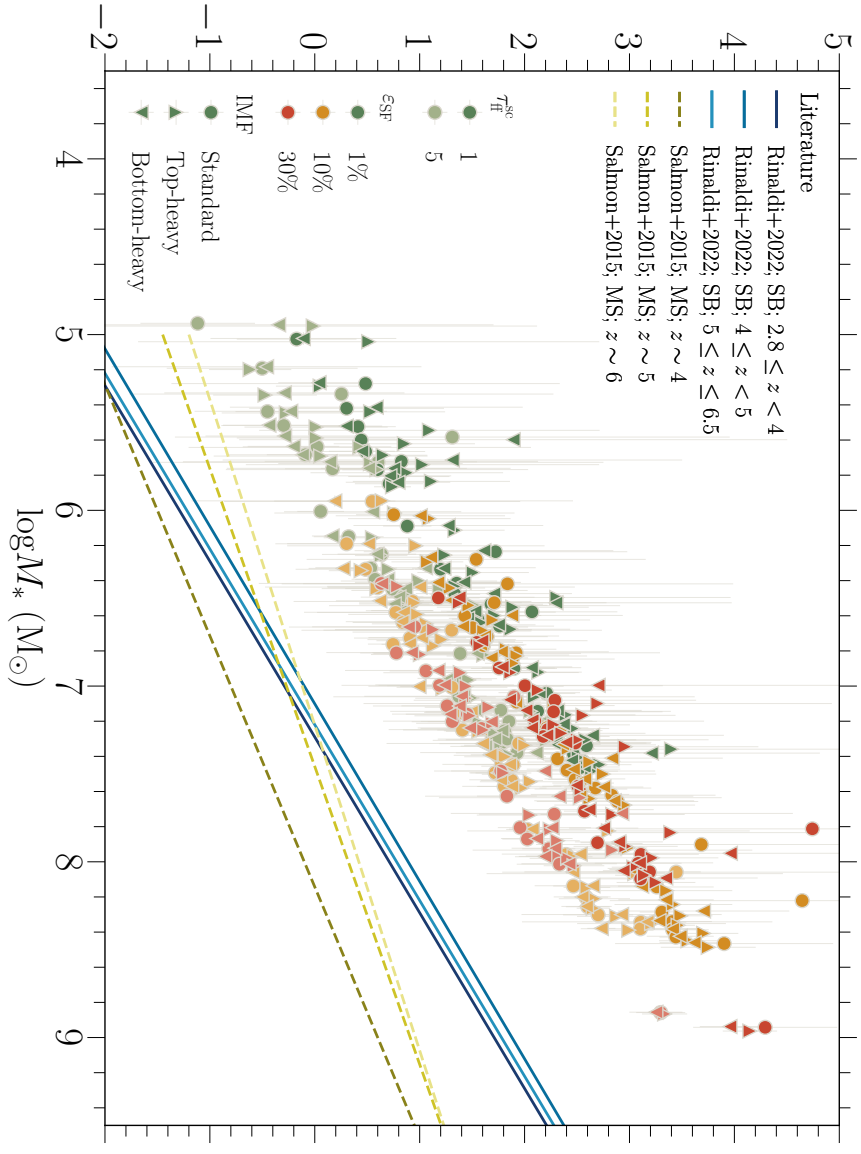


FIGURE 5.3.2: Star formation rate as a function of stellar mass of each galaxy. Full color markers represent results, where the free-fall time scaling factor was set to 1, while their lighter versions correspond to $\tau_{\rm ff}^{\rm sc}$ of 5. Circles represent setups with the standard IMF (Chabrier, 2003), triangles pointing upwards its top-heavy version, while pointing downwards the bottom-heavy one. Different colors of the markers refer to different star formation efficiencies, where green, orange and red means $\varepsilon_{\rm SF}$ of 1%, 10% and 30%, respectively. Solid lines represent best-fit lines from Rinaldi *et al.* (2022) to their starburst (SB) population, while dashed ones best-fits to the main-sequence galaxies from Salmon *et al.* (2015).

Galactic type			Model results $\log_{10}Y = a \cdot \log_{10}X + b$	
$\varepsilon_{\rm SF}$	$\tau_{\rm ff}^{\rm sc}$	IMF	$Y = \frac{SFR}{\rm M_\odot\,yr^{-1}},\ X = \frac{L_{\rm FIR}}{\rm L_\odot}$	$Y = \frac{L_{\rm H_2O}}{\rm L_\odot},\ X = \frac{L_{\rm FIR}}{\rm L_\odot}$
1%	1	s	$a = 0.908 \pm 0.021,\ b = -7.462 \pm 0.218;\ R^2 = 0.941$	$a = 0.809 \pm 0.001,\ b = -6.685 \pm 0.013$
1%	5	s	$a = 0.945 \pm 0.033,\ b = -8.586 \pm 0.329;\ R^2 = 0.772$	$a = 0.806 \pm 0.002,\ b = -7.202 \pm 0.026$
1%	1	t-h	$a = 0.923 \pm 0.030,\ b = -7.523 \pm 0.304;\ R^2 = 0.827$	$a = 0.808 \pm 0.001,\ b = -6.606 \pm 0.010$
1%	5	t-h	$a = 0.841 \pm 0.028,\ b = -7.424 \pm 0.294;\ R^2 = 0.849$	$a = 0.805 \pm 0.002,\ b = -7.111 \pm 0.026$
1%	1	b-h	$a = 0.903 \pm 0.030,\ b = -7.410 \pm 0.298;\ R^2 = 0.765$	$a = 0.809 \pm 0.001,\ b = -6.707 \pm 0.010$
1%	5	b-h	$a = 0.935 \pm 0.035,\ b = -8.499 \pm 0.354;\ R^2 = 0.879$	$a = 0.810 \pm 0.003,\ b = -7.269 \pm 0.029$
10%	1	s	$a = 0.900 \pm 0.021,\ b = -6.389 \pm 0.214;\ R^2 = 0.832$	$a = 0.808 \pm 0.001,\ b = -5.675 \pm 0.014$
10%	5	s	$a = 0.998 \pm 0.029,\ b = -8.160 \pm 0.287;\ R^2 = 0.870$	$a = 0.812 \pm 0.003,\ b = -6.261 \pm 0.028$
10%	1	t-h	$a = 0.908 \pm 0.019,\ b = -6.348 \pm 0.189;\ R^2 = 0.982$	$a = 0.811 \pm 0.001,\ b = -5.632 \pm 0.010$
10%	5	t-h	$a = 0.925 \pm 0.027,\ b = -7.293 \pm 0.272;\ R^2 = 0.908$	$a = 0.810 \pm 0.002,\ b = -6.157 \pm 0.019$
10%	1	b-h	$a = 0.892 \pm 0.024,\ b = -6.310 \pm 0.238;\ R^2 = 0.956$	$a = 0.807 \pm 0.002,\ b = -5.690 \pm 0.017$
10%	5	b-h	$a = 0.934 \pm 0.024,\ b = -7.521 \pm 0.247;\ R^2 = 0.948$	$a = 0.812 \pm 0.003,\ b = -6.295 \pm 0.028$
30%	1	s	$a = 0.918 \pm 0.066,\ b = -6.079 \pm 0.624;\ R^2 = 0.658$	$a = 0.807 \pm 0.001,\ b = -5.182 \pm 0.013$
30%	5	s	$a = 0.941 \pm 0.031,\ b = -7.083 \pm 0.297;\ R^2 = 0.892$	$a = 0.813 \pm 0.002,\ b = -5.795 \pm 0.025$
30%	1	t-h	$a = 0.899 \pm 0.034,\ b = -5.751 \pm 0.330;\ R^2 = 0.832$	$a = 0.809 \pm 0.001,\ b = -5.135 \pm 0.012$
30%	5	t-h	$a = 0.865 \pm 0.025,\ b = -6.167 \pm 0.240;\ R^2 = 0.923$	$a = 0.805 \pm 0.003,\ b = -5.620 \pm 0.028$
30%	1	b-h	$a = 0.870 \pm 0.032,\ b = -5.611 \pm 0.310;\ R^2 = 0.789$	$a = 0.810 \pm 0.002,\ b = -5.244 \pm 0.016$
30%	5	b-h	$a = 0.873 \pm 0.038,\ b = -6.354 \pm 0.371;\ R^2 = 0.861$	$a = 0.808 \pm 0.003,\ b = -5.771 \pm 0.032$
1%	1	s	—	$a = 0.809 \pm 0.001,\ b = -5.668 \pm 0.008$
1%	1	t-h	—	$a = 0.808 \pm 0.001,\ b = -6.655 \pm 0.012$
10%	1	s	—	$a = 0.811 \pm 0.002,\ b = -5.806 \pm 0.016$
10%	1	t-h	—	$a = 0.810 \pm 0.001,\ b = -5.677 \pm 0.010$

Notes: Results from running simulations with all considered parameters combinations (for more details see Tab. 5.2.1). In the IMF parameters, *s* refers to standard, *t-h* to top-heavy, and *b-h* to bottom-heavy. For details on the $\tau_{\rm ff}^{\rm sc}$ and its relation with the free-fall time efficiency we refer the reader to Dutkowska and Kristensen (2022), where extensively discussed the impact of this factor. The four setups in the bottom of the table divided by the horizontal line refer to our high-z test (see. Sect. 5.4.2). Since the test had impact only on the water emission, we did not provide the SFR – $L_{\rm FIR}$ relations as these are the same as for the corresponding Galactic setups. The correlation between the $L_{\rm H_2O} - L_{\rm FIR}$ had consistent $R^2 \approx 0.999$, hence we do not provide the R^2 values for this relation in the table.

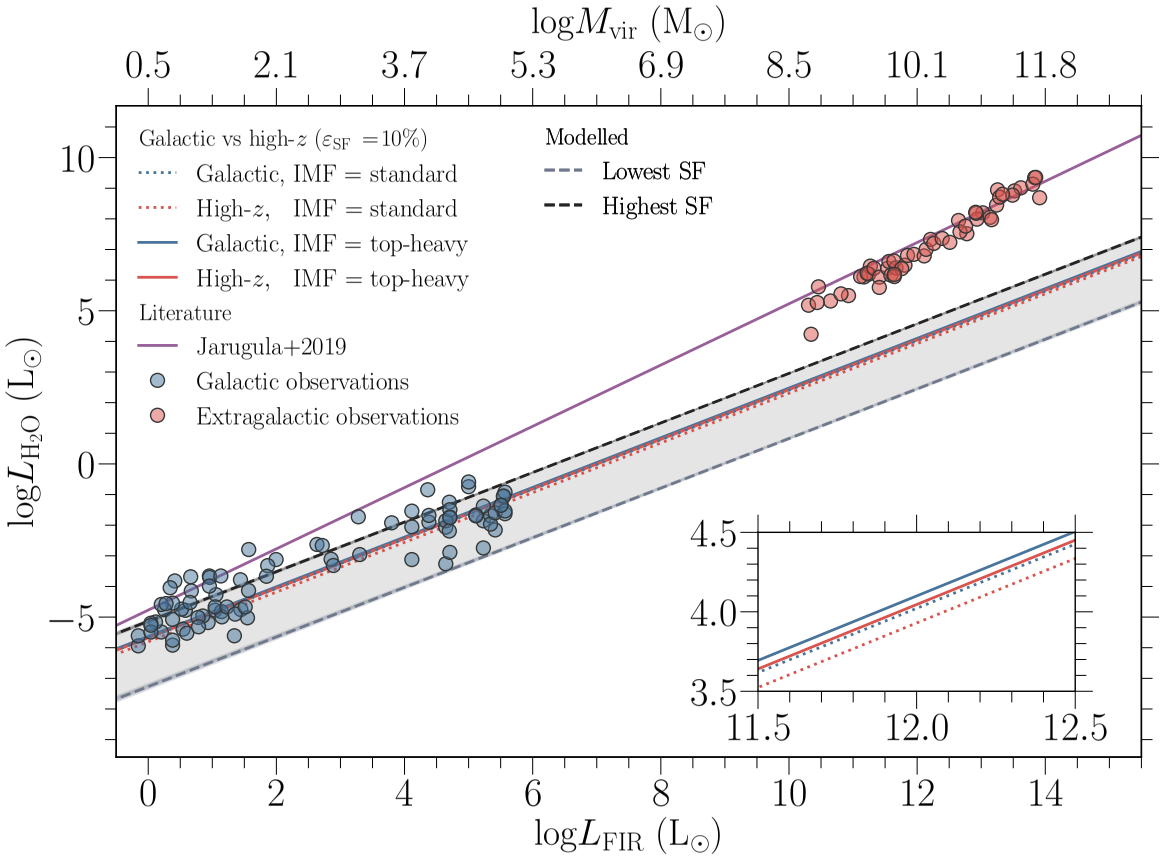


FIGURE 5.3.3: Simulated water line luminosities as a function of far-infrared luminosities. The dashed black line represent the most emitting galaxy in our simulations ($\varepsilon_{\rm SF}$=30%, IMF = top-heavy and $\tau_{\rm ff}^{\rm sc}$=1), while the dashed gray line corresponds to the least star-forming one ($\varepsilon_{\rm SF}$=1%, IMF = bottom-heavy and $\tau_{\rm ff}^{\rm sc}$=5). The shaded gray area between these two lines refer to the probed parameters space and all possible outcomes considered in this study would fall in that regime. Solid blue and red lines refer to the results derived for setups with the top-heavy IMF form for the Galactic and extragalactic $M_{\rm env} - I$ relations, respectively. Dotted lines show the result for these two correlations, when the standard IMF was applied. In both, i.e., the standard and the top-heavy cases, the free-fall time scaling factor was set to 1. Circles refer to observational samples (for more details we refer the reader to Sect. 5.3.1), while the purple line represent the expected relation from Jarugula *et al.* (2019).

$L_{H_2O}/L_{FIR} = 1.69^{+0.79}_{-0.54} \times 10^{-5}$ (we explore this more extensively in Sect. 5.4.1) are especially far-off from our expectations for the brightest high-z galaxies. We discuss this further in Sect. 5.3.3 and we explore the possible impact of the inclusion of high-z starbursts to the correlation between the envelope mass and intensity ($M_{env} - I$ relation), which is the base of emission assignment in the galaxy-in-a-box model, and whether it could explain the observed differences in Sect. 5.4.2.

5.3.4 L_{FIR} – *SFR correlation*

To evaluate derived SFRs further, we explored their relation with corresponding far-infrared luminosities (Fig. 5.3.4). We clearly see that the derived SFRs create different populations that depend on the star formation and the free-fall time scaling factor. Again, we are clearly overestimating the SFRs. However, relations in the literature, e.g., Kennicutt and Evans (2012), Casey *et al.* (2014), fall into our lower prediction regime, meaning that at least for the star forming galaxies with lower star formation activity (with respect to the standard setup in the galaxy-in-a-box model), we are roughly recovering the expected star formation process.

The span of the SFRs derived in this study depends strongly on the efficiency of the process. The discrepancy between the literature values and our simulations can be as high as two orders of magnitude. We focused and derived relations analog to Eq. (5.5) for the least and most emitting setups, as well as the standard model setup from the galaxy-in-a-box model. We provide all of the derived relations in Tab. 5.3.1. Here, we do not derive almost identical slopes, as we did for $L_{FIR} - L_{H_2O}$. For the most extreme cases in L_{FIR}-SFRs relation, we derive slopes of 0.94 ± 0.04 and 0.90 ± 0.03, which agree within the uncertainties, while the derived intercepts (here, the intercept refers to the term b in Eq. (5.5), which is further used as showed in Eq. (5.6)) are equal to -8.50 ± 0.35 and -5.75 ± 0.33, respectively. We discuss the apparent excess in SFRs further in Sect. 5.3.4.

5.3.5 L_{H_2O} – *SFR correlation*

The last explored dependence was that of L_{H_2O} and corresponding SFRs. We see from Fig. 5.4.1 that all of the derived SFRs fall into the same population, which is expected, considering the fact that the greater the luminosity, the more star-forming and massive the corresponding galaxy. By fitting all of the derived points to Eq. (5.5), we get a slope of 1.11 ± 0.01 and intercept -0.083 ± 0.018, indicating a near-proportionality between the SFR and water luminosity.

However, Fig. 5.4.1 suggests that we are systematically overestimating SFRs by ~ 4 orders of magnitude with respect to what have been found by Jarugula *et al.* (2019), where $\mathrm{SFR}\left(\mathrm{M_\odot yr^{-1}}\right) = 7.35^{+5.74}_{-3.22} \times 10^{-6} L_{H_2O}\,(\mathrm{L_\odot})$. If extrapolating their relation to Galactic star-forming regions, we would underestimate SFRs by orders of magnitude (Kristensen et al. subm.). We will discuss this discrepancy in Sect. 5.4.5.

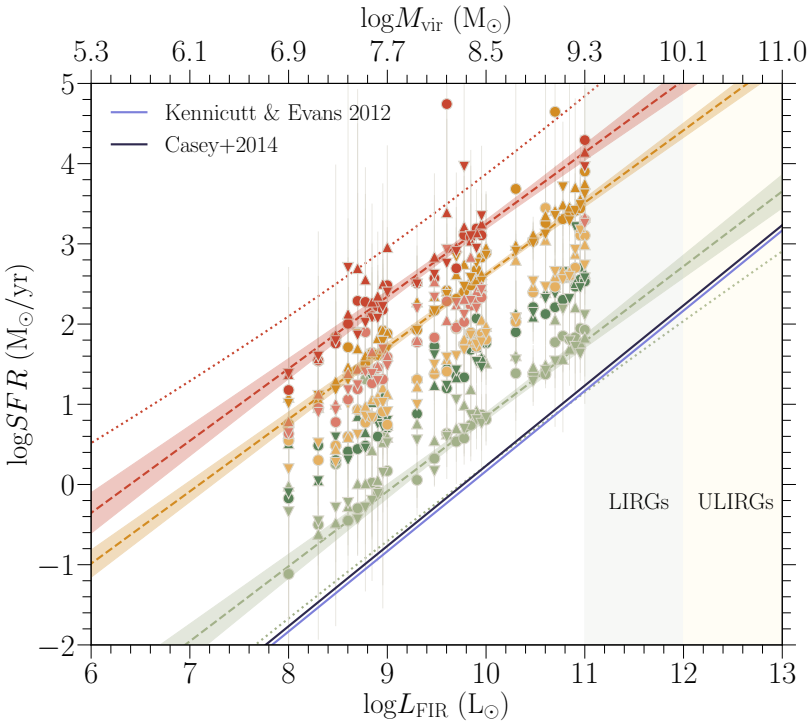


FIGURE 5.3.4: Star formation rate as a function of far-infrared luminosity. Colors and markers as in Fig. 5.3.2. Dotted lines refer to the upper prediction band for the most-star forming setup and the lower prediction band for the least star-forming setup. Shading of the best-fit lines correspond to the the 95% confidence region of each correlation. Solid lines represent the literature estimates.

5.4 DISCUSSION

In the following, we will discuss derived SFRs and water luminosities. We will also evaluate how considered star formation parameters could affect the results and compare our results with the literature. Moreover, we will discuss what other physical processes not considered in this study could impact the derived values and explore other possible influences.

5.4.1 *Insights from $L_{\rm H_2O}/L_{\rm FIR}$ ratios*

The ratio of water luminosity and corresponding infrared luminosity could be used to understand the the source of the observed water emission (this is shown in Fig. 5.4.2). This in turn can help us understand whether water behaves differently in different galactic regions and galactic types. With this in mind, we calculated the ratios derived from the galaxy-in-a-box model and compared them with our galactic and extragalactic samples.

The derived values $10^{-8} < L_{\rm H_2O}/L_{\rm FIR} < 10^{-6}$ fall below those from all objects considered in the extragalactic sample, but coincide with the galactic sample at its high-mass/high-luminosity end (see Fig. 5.3.3). We know from galactic observations (e.g., van Dishoeck *et al.*, 2021b) that water emission from young stellar objects predominantly comes from the shocked material in outflows. Therefore, a natural assumption would be that the galactic sample should be consistent in terms of the calculated ratios. Instead what we see is that low- to intermediate-mass protostars exhibit roughly the same

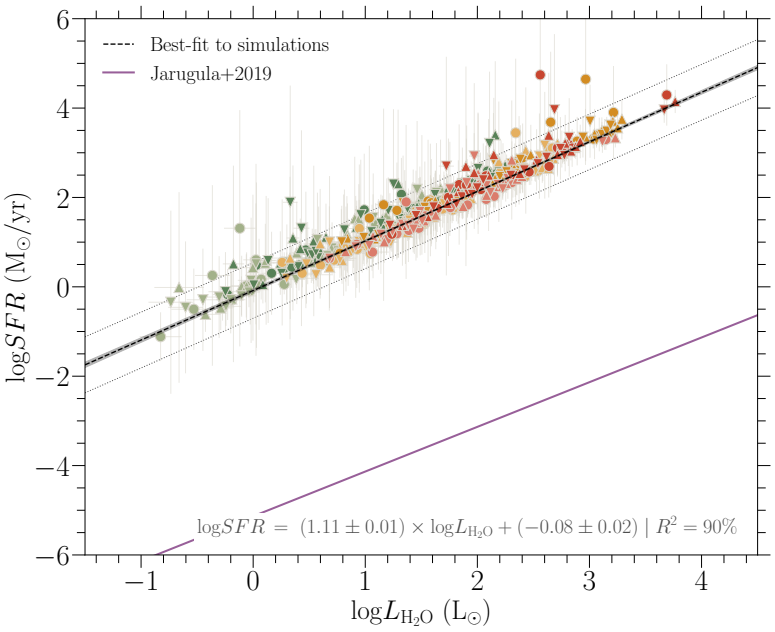


FIGURE 5.4.1: Star formation rate as a function of water line luminosity. Colors and markers as in Fig. 5.3.2. The shaded region corresponds to the 95% confidence region of the correlation, while dotted lines indicate where 95% of the measurements should fall. The solid purple line represent the expected relation from Jarugula *et al.* (2019).

ratios as the extragalactic sample, and we see a clear drop for the most luminous end of the galactic objects.

Available water observations of Galactic high-mass young stellar objects are limited due to both their number and sensitivity. One of the most detailed studies was conducted with a survey towards the Cygnus X star-forming region (PI: Bontemps; San José-García, 2015). Cygnus-X is one of the nearest massive star-forming complexes ($D \sim 1.3$–1.4 kpc, e.g., Rygl *et al.*, 2012). However, even these observations do not recover the total emission that would come from a high-mass star forming complex, due to the spatial resolution and sensitivity limitations of the HIFI instrument on the *Herschel* Space Observatory. This survey, one of the most complete surveys, only consists of single-pointing observations. Thus, new instruments are needed to fully estimate the amount of H_2O emission coming from a forming Galactic cluster.

To take another approach, we estimate the amount of H_2O emission from the nearby W3 high-mass star forming region. The distance is 2 kpc and the age is 2 Myr (Bik *et al.*, 2012). We used a mass of $4{\times}10^5$ $M_\odot$ for the entire cluster (Rivera-Ingraham *et al.*, 2013), corresponding to a total luminosity of $2{\times}10^6$ $L_\odot$ using Eq. 5.3. To estimate the missing emission from all protostars, we ran a model for just one cluster instead of an entire galaxy. The cluster model predicts a total line intensity of 120 K km s^{-1}, which may be compared to the observed value of the high-mass protostar W3-IRS5 of 21.9 K km s^{-1} (van der Tak *et al.*, 2013), which has a luminosity of 10^5 $L_\odot$, or 5% of that of the cluster. The simulated value is highly sensitive to the adopted age of the cluster, and an age of, for example, 1 Myr would result in a predicted intensity of 250 K km s^{-1}. This implies that for an

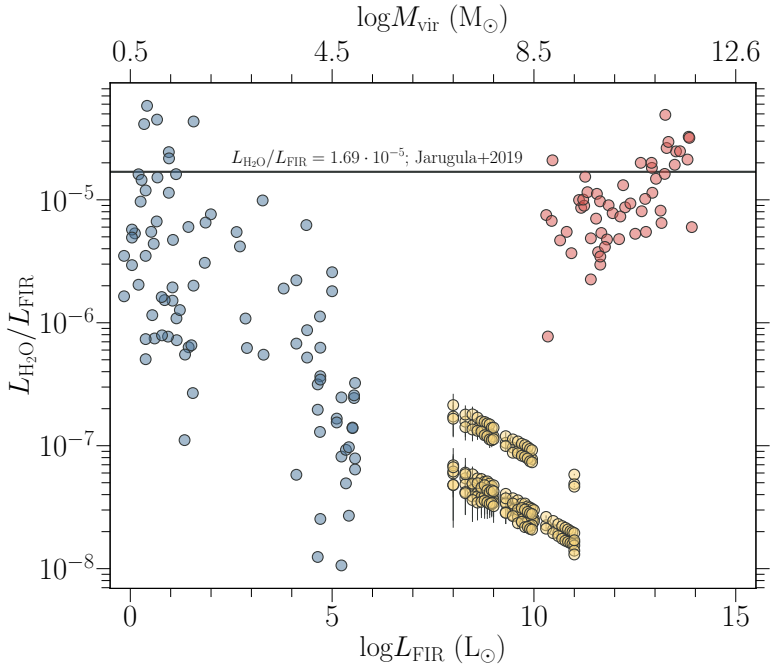


FIGURE 5.4.2: $L_{\rm H_2O}/L_{\rm FIR}$ as a function of $L_{\rm FIR}$. Blue and red points refer to Galactic and extragalactic observations, respectively. Yellow points refer to our simulations with star formation efficiencies of 10% and 30%.

individual cluster, we need to know the age accurately to within 10%, which is not currently possible. It is reasonably possible that the amount of water emission we are missing is between a factor of 6 or 12. Without being able to map the entire cluster in water emission, we will not know exactly how much.

5.4.2 *High-z test*

Knowing that the relation between water emission and far-infrared luminosity spans over many orders of magnitude starting from the low-mass protostars to high-z HyLIRGs, we probed the influence of the extragalactic observations on the $M_{\rm env} - I$ relation, and explore how this extrapolated form of the formula would impact derived intensities.

In Fig. 5.3.1 we see that by including the extragalactic observations we will effectively lower the contribution from the low-mass end of the correlation and that it will only positively impact the high-mass protostars. On the other hand, the purely Galactic correlation lowers the emission from the high-mass protostars. Therefore, considering that we are underestimating water emission, we focused only on the standard and top-heavy IMF forms. The reason is that the standard IMF is already dominated by the low-mass end of the distribution, and we also know from Dutkowska and Kristensen (2022) that the emission derived for the bottom-heavy IMF is practically indistinguishable from the standard one. At the same time the top-heavy IMF would increase the emission even for the normal form of the correlation, and the inclusion of the extragalactic sources increases the slope by $\sim$ 10% (see Fig. 5.3.1).

The results of the test indicate that inclusion of the extragalactic sources on average results in lowered emission, and that the difference with the results with the purely Galactic correlation starts to diminish for higher galactic masses and higher star formation efficiencies. This effect is not surprising as the star formation process is dominated in both total mass and number by low-mass protostars, while in terms of total bolometric luminosity the high-mass stars dominate the picture completely (e.g., Kroupa, 2002). Therefore, the inclusion of the extragalactic sources, lowering the emission from the low-mass protostars, naturally lowers the water emission derived from the simulated galaxies, as this is the main star-forming component if we consider Milky Way-like star formation. However, for the high-z starbursts with high star formation efficiencies and seemingly top- or even extremely top-heavy IMFs, this extrapolation could make a difference, when simulating star formation and its emission. Nevertheless, we will not further investigate this, as this is beyond the scope of this paper.

5.4.3 *SFR estimates*

From the results derived in this study, we are consistently overestimating star formation rates for given galactic types. However, when considering the assumptions behind the model and the fact that in the current version of the model simulating current star formation rates, without correcting for star formation histories and existing populations, the overestimation stops being that prominent.

The galaxy-in-a-box model was created as a tool for simulating emission from active and current star formation in galaxies. Therefore, even though the model accounts for dynamical differentiation of (proto)stellar ages, the model does not account for already existing, older stellar populations, that normally would contribute to observations from which the rates is calculated. Moreover, as noticed from the Fig. 5.3.2 and 5.3.4, the results lie close to the literature estimates, if we assume low star formation activity. A calibration of the SFRs of galaxies depends on their current star formation activity. If the bulk of galaxies are observed during a period of low star formation, we would naturally fall on the lower SFR side. Also, there are many factors influencing star formation activity in galaxies, which are not taken into account in the current version of the galaxy-in-a-box model.

Another important aspect is that when calibrating SFRs from $L_{\rm FIR}$, one has to make assumptions such as the initial mass function, the star formation history, which is a source of additional the uncertainty of the final estimation of the star formation rate. Moreover, $L_{\rm FIR}$ is likely to underestimate the star formation rate in young clusters (Gutermuth *et al.*, 2011) by up to an order of magnitude, and these are the main objects of interest in this study. If this is the case, our SFR estimates are roughly consistent with expectations.

Lastly the galaxy-in-a-box model accounts for all stellar products, i.e., from brown dwarfs to high-mass stars. Therefore, it is not subject to observational limitations and the apparent overestimating can be an effect of accounting for all, including normally unobservable objects, as illustrated in the W3 example above. The scenario we are considering can resemble a bit more the high-z situation, where galaxies are filled with active star-

forming regions and described as "full of Orions" (Rybak *et al.*, 2020). In this case, having relatively young star-forming regions, we trace only active and current star formation without accounting for higher differentiation of ages and stellar populations.

5.4.4 *Impact of the star formation parameters*

In this study we explored simulations for different galaxy types, and as such went through a broad parameter space (see Sect. 5.2.2 and Tab. 5.2.1). As in the first the galaxy-in-a-box study (Dutkowska and Kristensen, 2022) we observe no strong effect of the IMF, even if we included nearby sub-LIRGs, LIRGs and quasars, as well as high-z quasars, ULIRGS and HyLIRGs to the correlation that is used to assign molecular emission to protostars. This is expected as the extrapolation to the high-z regime changes the slope of the correlation only by ~10%.

We observe a strong impact of the star formation efficiency and the free-fall time scaling factor, both for the derived emission and star formation rates. This is of no surprise as both parameters impact the stellar population of each cluster. The free-fall time scaling factor will effectively lower the ages of the clouds and thus increase the emission, while the star formation efficiency regulates how much of the molecular reservoir will be turned into stars, hence increasing the number of stars.

One new studied input parameter was the mass of the galaxy, as derived from Eq. (5.3). Clearly, the more massive the galaxy, the more emission we derive from the model. However, this parameter has its own uncertainty, which would be important especially when considering the predicted water emission. The relation between the mass and luminosity was also derived for young stellar objects by Pitts *et al.* (2022), where:

$$\log\left(M_{\rm env}/{\rm M}_\odot\right) = 0.30^{+0.07}_{-0.06} + 0.79^{+0.01}_{-0.02}\log\left(L_{\rm bol}/{\rm L}_\odot\right). \tag{5.7}$$

Although this expression was inferred for individual protostellar envelopes, it clearly agrees with Eq. 5.3 within the uncertainty. Here, we make the assumption that L_{bol} represents L_{FIR} as young protostars are deeply embedded in gas and dust, and L_{bol} will be dominated by the contribution from L_{FIR}. Hence, if the relation between mass and luminosity is more universal, underestimating or overestimating can respectively underestimate or overestimate the available molecular reservoir.

5.4.5 *Comparison with observations*

When comparing the derived values with observations we clearly see that we are underestimating the water emission by at least one to two orders of magnitude (see Fig. 5.3.3) and overestimating the star formation rates from a factor a few to two orders of magnitude (see Fig. 5.3.4 and 5.3.2).

We discuss the possible explanations of difference in SFRs in Sect. 5.4.3 extensively, and here we will focus solely on the difference between our estimate and that of Jarugula *et al.* (2019). The SFR calibration of Jarugula *et al.* (2019) utilizes the L_{FIR} - SFR relation from Kennicutt and Evans (2012):

$$SFR\ ({\rm M}_\odot\,{\rm yr}^{-1}) = 1.47\times10^{-10}L_{\rm IR}\ ({\rm L}_\odot), \tag{5.8}$$

which, as mentioned in Sect. 5.4.3, is subject to various uncertainties. This is especially important, when considering the initial mass function in the high-z ULIRGs and HyLIRGs as found in many studies (e.g., Zhang *et al.*, 2018), adding uncertainty to the calibration. Moreover if we were to apply the calibration from Jarugula *et al.* (2019), we would heavily underestimate SFRs towards well-studied, resolved Galactic clouds, where the relation inferred for water emission and luminosity is $\approx$ 3000 times higher than that of Jarugula *et al.* (2019) (further discussion in Kristensen et al. 2022, under review).

Focusing on the water emission, there are a few factors that could contribute to the observed difference and we discussed some of them in Sect. 5.4.1. Additionally, one of the reasons for not recovering the emission is that we convert 100% of the galactic mass to an emitting source. There is a number of parameters standing in the way, with the star formation efficiency being the most obvious one. Moreover, currently we consider emission only from Class O and Class I protostars. Therefore, when considering emitting components that constitute only a small percentage of a whole galaxy, we are naturally going to loose a certain amount of emission.

In galaxies there are more emitting components than just protostars. These include, e.g., photodissociation regions, galactic outflows and supernovae. Even though their contribution likely will be lower than that from star formation, it is essential to fully reproduce the emission. As such, the inclusion of those additional emitting components is a part of planned future improvements.

5.5 CONCLUSIONS

We extended the galaxy-in-a-box model to relate the predicted molecular emission from forming stars with star formation rates. In this paper we demonstrated the introduced extension and evaluated the derived results for galaxies with $L_{\rm FIR} = 10^8 - 10^{11} {\rm L}_\odot$ and various levels of star formation activity. We complemented the star formation rate study by extracting predicted emission for the para-H_2O $2_{02} - 1_{11}$ line at 987.927 GHz. The main results are as follows:

- The star formation efficiency and the free fall time efficiency have a strong impact on the star formation rate and emission, whereas the opposite holds for the initial mass function;
- For the most extreme star-forming cases, the galaxy-in-a-box model overestimates the star formation rates by up to two orders of magnitude. However, this difference could be lowered depending on how much the current calibrations using $L_{\rm FIR}$ as a star formation tracer underestimate the actual star formation rate values;
- The model underestimates the water emission especially for the high-z quasars, ULIRGs, and HyLIRGs by up to two orders of magnitude;
- For the moment the model needs to account for additional sources of emission, including supernovae, photodissociation regions and galactic outflows. Moreover, we need to revisit the derived water emission for

Galactic high-mass star forming regions, as we might miss the bulk of emission.

Our estimates deviate from observations and literature. However, the apparent differences are consistent with expectations in the sense that known sources of emission are not included in the model, and thus the galaxy-in-a-box model is a promising step toward shedding light on the star-forming properties of galaxies across cosmic time. In the near future, we plan to introduce a number of extensions that will account for other sources and processes that could contribute to the emission. The planned extensions included accounting for galactic outflows, both AGN and starburst driven, shocks from supernovae, and emission from photodissociation regions. Moreover, we are introducing H_2 and high-J CO emission, which is going to be especially important in the JWST era.

To properly account for water emission in our own Galaxy, in the future we will need a new far-infrared probe with the sensitivity of JWST. Such a probe is the planned PRIMA[1] mission. Only then we can fully recover the emission from star-forming clusters in Galaxy, and properly estimate the contribution from protostars in all stellar mass ranges.

[1] https://prima.ipac.caltech.edu

ACKNOWLEDGEMENTS

The research of KMD and LEK is supported by a research grant (19127) from VILLUM FONDEN.

Appendix

5.a OVERVIEW OF THE GALAXY-IN-A-BOX MODEL

The galaxy-in-a-box model, which was used to derive the result presented in this study, is described in detail in the paper by Dutkowska and Kristensen (2022). Below we provide a general description.

The architecture of the galaxy-in-a-box model is rooted in three elements of the galactic star-forming environment, i.e., GMCs, star-forming clusters, and protostars. The model starts the calculations by generating a spatial and mass distribution of GMCs based on the observational data of other galaxies. The GMCs that will be passed to the next steps of the simulation are chosen randomly from the mass distribution, and each GMC mass acts as an initial cluster mass, i.e., one GMC will form one cluster in the model. Before the GMC mass is passed to the cluster module (based on the cluster-in-a-box model by Kristensen and Bergin (2015)), each cloud is assigned an age based on its free-fall time, which is then randomly scaled between being newly formed and completely collapsed. This affects the number of deeply embedded protostars (Class 0 and I protostars) driving outflows in a given cloud. In the next step of the calculations, the cluster module returns protostellar mass, age, and spatial distribution. However, the latter is disregarded in the galactic-scale calculations.

With the complete protostellar distribution, molecular emission is assigned to each Class 0 and Class I protostar. This results in the total expected molecular outflow emission from a given cluster. The model repeats these calculations for all chosen GMCs. Then, the information about the total cluster mass (expressed as the total mass of (proto)stellar content) and emission is returned to the galactic spatial grid. After accounting for the sizes of clusters and their location, the raw galactic emission image is convolved with a Gaussian beam. Hence, the model returns statistics on galactic clusters (their number of stars, mass, and emission) and integrated intensity image. However, the level of the details in the returned statistics can be easily adjusted, such that the exact protostellar distributions of each cluster can be stored.

6

Water line emission in externally irradiated shocks

K. M. Dutkowska[1], B. Godard[2], A. Gusdorf[2], L. E. Kristensen[1]

[1] Niels Bohr Institute, University of Copenhagen, Øster Voldgade 5−7, 1350 Copenhagen K., Denmark

[2] Laboratoire de Physique de l'École Normale Supérieure, ENS, Université PSL, CNRS, Sorbonne Université, Université Paris Cité, 75005 Paris, France
Observatoire de Paris, PSL University, Sorbonne Université, LERMA, 75014 Paris, France

In preparation

Abstract

Context. Molecular shocks play a significant role in the physics and chemistry of star-forming regions and for the protostars themselves, e.g., by injecting momentum into the surrounding cloud. Understanding their effect on the interstellar medium is necessary for our interpretation of the molecular emission associated with the star formation process.

Aims. In this work, we explored the impact of various externally irradiated stationary molecular shocks on the gaseous water abundance and resulting water emission. We analyzed how different properties of the pre-shock medium impact the temperature of the neutral fluid and water abundance, excitation diagrams, as well as specific and integrated intensities.

Methods. To study the impact of externally irradiated shocks on the water abundance and its line emission, we conducted radiative transfer modeling of models from the Paris-Durham shock code. The chosen models varied in the strength of the external UV radiation field, shock velocities, and pre-shock densities.

Results. We find that pre-shock densities and shock velocities can have a significant positive impact on intensities of water for both

shock types, while the increasing UV radiation has a negative impact on lines with $E_{up} \leq 600$ K. Moreover, J-type shocks are on average associated with higher water emission and excitation than their C-type counterparts with the same pre-shock density and shock velocity.
Conclusions. Our radiative transfer solutions provide insights into the nature of water emission and excitation in shocked regions, where the bulk of Galactic water emission originates. Our solutions were derived for externally irradiated shock, i.e., we assumed the presence of an UV radiation field. Hence, they can be particularly important for understanding the nature of molecular excitation in the presence of high-mass stars, especially for lower-excited lines.

6.1 INTRODUCTION

Shocks are a ubiquitous phenomena in the Universe. They are driven into the surrounding environment by, for example, supernovae (SNe), cloud collisions, protostellar outflows, and galactic outflows. With that they are pivotal for understanding the properties of the interstellar and the circumgalactic medium (ISM and CGM, respectively).

A shock is a hydrodynamical surprise (Chernoff, 1987). Its formation is induced by a shock wave moving with velocity v_S through an ambient medium with a local sound speed, c_s, such that $v_S > c_s$. Hence, when the supersonic shock wave collides with the medium, its speed decreases, leading to the subsequent steepening of the wave. When the back of the wave catches up with its front, it breaks, and a shock is born.

Shocks convert kinetic energy into magnetic and thermal energy, which results in a range of consequences for the medium it interacts with. The changes are long lived (compression, acceleration) or short-lived (heating). In the ISM, the consequence of these effects is a modification of the chemistry, both in the gas-phase and in the processing of the grain material. Moreover, depending on the nature of the shock wave and the nature of the medium, the resulting shocks differ in properties. The two dominant shock types are jump (J-type) shocks and continuous (C-type) shocks. The main difference between these two shock types is set by the balance between the strength of the transverse magnetic field in the medium and the ionization fraction: the magnetic field has the potential to act as a reservoir of energy in a C-type shock, and the ionization fraction controls the recoupling of the ionized and neutral particles in a J-type shock.

In the lack of or in the presence of a very weak transverse magnetic field ions and neutrals are fully coupled with each other. When the shock front interacts with the medium, both fluids are decelerated through the combined actions of viscous stresses and the thermal pressure gradient induced by viscous heating, as the medium moving toward the shock front causes friction. In the reference frame of the shock, the in-coming pre-shock gas quickly jumps from a supersonic to subsonic regime, with respect to the neutral sound speed, over a distance of about a few mean free paths, and we get a jump in the temperature, density, and velocity of the gas. Hence, this type of shock is called a jump shock.

When the magnetic field is strong and if the ionization fraction is low enough, the Lorentz force decelerates the ions, leading to their decoupling from the neutral species. This decoupling induces a drag force, which slows down the neutrals. When the drag force applies for a longer time, there is a smoothing of the neutral velocity gradient and thermal pressure gradient, which reduces the strength of viscous stresses. This in turn leads to a continuous slow down of the neutrals through the pressure gradient and the ion-neutral friction. This is what we call a continuous shock.

Shocks serve as an excellent diagnostic of the ISM, as they carry information about its dynamics. By interpreting their effects on the chemistry of the dust and gas we can study their history, pre-shock conditions and driving sources. Moreover, by studying their energetic and chemical impacts, we can better understand the role they have on star formation. When studying star-forming regions one of the best diagnostics of the earliest stages of the star formation process is protostellar outflows, which arise at the earliest stages of star formation when the interaction between the infalling envelope, winds, and jets launched from the protostar is particularly strong (Bally, 2016). Hence, outflows are tightly related to the physics and chemistry of the shocks, where the driving source is the central protostar. From observations of Galactic star-forming regions we know that a molecule that is almost uniquely associated with outflows is water. In outflows, water, which is predominantly locked up as ice on dust grains, is released from the icy grain mantles into the gas phase through sputtering. It can also be efficiently formed directly through gas-phase reactions. Thus, in outflows the gas-phase abundance of water experiences a jump of many orders of magnitude. At the same time, the physical conditions are conducive to water being readily excited into rotational states, and the de-excitation leads to subsequent cooling of the gas (Tafalla *et al.*, 2013; Suutarinen *et al.*, 2014).

A shock structure itself is more complicated than the picture one can take from the overall definition. Taking protostellar outflows as an example of shocks, if we put ourselves in a shock frame, what happens is that the interstellar medium streams towards us from the front and at the same time the jet moves towards us from the back. Hence, two fluids streaming towards us will create a contact discontinuity, where they will impact one another. Several structures arise during this process. The most prominent structure is the bow shock, i.e., the shock front from the medium side. The bow shock is often called a forward shock, being the furthest structure from the disturbance source and interacting directly with the undisturbed interstellar medium. Then, from the jet side we have a mach disk, being an immediate interaction of a jet with the ambient medium. The mach disk is in fact a shock front from the jet side, and is associated with the presence of the so-called reverse shock, which forms as the result of slowing down the material of the incoming jet. Keeping ourselves in the shock frame, when the two incoming fluids collide, i.e., the interstellar medium and the jet, the material escapes sideways from the mach disk. This material still sees the interstellar medium streaming towards it and so it gets deflected and bends backwards towards the jet, which results for instance in the characteristic shape of a bow-shock in a protostellar outflow.

In this study we explored the impact of various J- and C-type shocks on

gaseous water abundance and its emission, calculated with the Paris-Durham shock code and coupled to it the radiative transfer model (Lehmann *et al.*, 2022). It is convenient to put ourselves in the reference frame of the shock, therefore, we will use it throughout the paper. This paper is organized as follows. Section 6.2 we describe methods used in this study, i.e., the shock code and the radiative transfer model. We also describe the changes to the latter model, which we introduced to enable calculations for water. Then, in Section 6.3 we present the models, which we used in this study and the derived results. Finally, in Section 6.4 we summarize our results.

6.2 METHODS

The Paris-Durham shock code is extensively described in Godard *et al.* (2019) and references therein. Hence, in this study, we will focus only on the general description of the model. Moreover, the models presented in this study are taken from the large-grid of shock models calculated by Kristensen et al. (to be subm.).

From the modeling perspective we can divide shocks to non-irradiated and irradiated. The difference is rooted primarily in excluding or including of the illumination by the UV photons emitted by an external source of radiation, e.g., high-mass stars, located in the pre-shock or by the shock itself, respectively. The role of the externally irradiated models is especially important for star-forming regions, where photons from protostars are expected to play a role in the physics and chemistry of the regions. Water emission from non-irradiated shocks, using shock models from the Paris-Durham shock code, was described by, e.g., Flower and Pineau des Forêts (2010), Flower and Pineau des Forêts (2013), Gusdorf *et al.* (2011), and Gusdorf *et al.* (2015). In this study we explore externally irradiated shocks.

6.2.1 *The Paris-Durham shock code*

The shock models presented in this study where derived using the Paris-Durham shock code (Flower and Pineau des Forêts, 2003; Godard *et al.*, 2019, https://ism.obspm.fr). This code calculates a time-dependent dynamical, chemical, and thermal evolution of matter in 1D stationary shocks using the plane-parallel geometry. The code accounts for various cooling and heating processes including, e.g., fluid interactions, cosmic ray ionization, and collisional (de)excitation and radiative emission. The chemical network of the code includes $\sim$ 1000 chemical reactions involving $\sim$ 140 chemical species, e.g., neutral-neutral and ion-neutral reactions, ionization and dissociation by electrons from cosmic ray ionization, and photoionization. The code is also used to evaluate the initial abundances in chemical equilibrium, that are then passed as initial conditions of the model.

A medium moves toward a shock wave at speed $-V_S$ in the direction perpendicular to the transverse magnetic field and the shock wave is irradiated by an isotropic flux of UV photons if $G_0 > 0$. The computations starts in the pre-shock medium, which marks the time and distance for the computations of the shock, and the model enters the shock layer. Then, the post-shock layer is marked by the medium reaching a steady-state. In terms

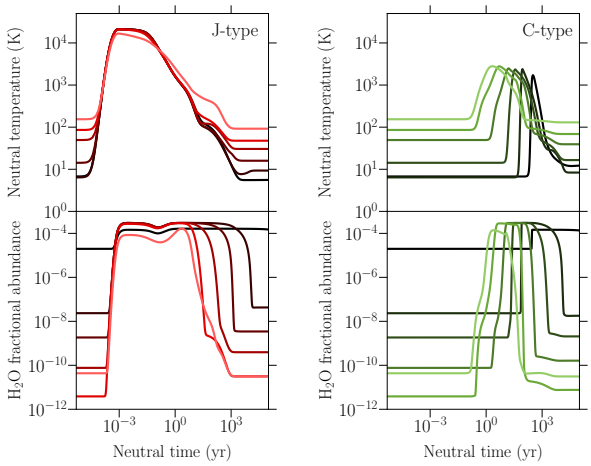


FIGURE 6.2.1: Neutral temperature (top) and fractional abundance of H_2O (bottom) as a function of time for J- and C-type shocks (with $n_H = 10^4$ cm^{-3}, $V_S = 20$ km s^{-1}, and $b = 1$) where UV radiation field was considered. The intensity of the radiation field, G_0, varied between 0 – 1000. The lighter the color, the higher the irradiation, i.e., the G_0 value.

of a shock critical velocity, if $V_s > B/\sqrt{4\pi\rho_i}$, i.e., the magneto sonic speed in the charged (ionized) fluid, a J-type shock will develop, but if $V_s < B/\sqrt{4\pi\rho_i}$ a C-type shock will form (Draine, 1980).

6.2.2 *Radiative transfer*

In order to model the radiative transfer through the interstellar medium we need to solve the radiative transfer equations. In the case of the Paris-Durham shock code, the designated radiative transfer code uses the large velocity gradient (LVG) approximation, in which the emission at any point in the cloud is completely decoupled from emission at all other points in the cloud, i.e., slabs which are radiatively uncorrelated from one another. This is likely a good approximation, as the different layers of the shock move at different velocities, thus Doppler-shifting any emission from the different layers with respect to each other.

We have tested the radiative transfer solution for water against the solution presented by Gusdorf *et al.* (2011), and also adapt the escape probability described by:

$$\beta_\perp = \frac{1 - e^{-\tau_\perp}}{\tau_\perp}, \tag{6.1}$$

where $\tau_\perp$ is the LVG optical depth in the direction of shock propagation, perpendicular to the shock front.

6.2.3 *Collisional partners*

Collisional excitation is highly important in the hot gas in shock waves. The main collisional partners are atomic and molecular hydrogen, helium, and electrons. Here, we are not considering collisions with electrons, since the available H_2O–e^- collisional rates cover only 18 water lines (Faure *et al.*, 2004), and we did not want to extrapolate the rates in temperatures nor levels. For H_2O–H we follow Daniel *et al.* (2015). For collisions between

para-H_2O with para- and ortho-H_2, as well as ortho-H_2O with ortho-H_2 we use collisional rates from Daniel *et al.* (2011), while for ortho-H_2O with para-H_2 from Dubernet *et al.* (2009). The molecular data for water is taken from the LAMDA database (Schöier *et al.*, 2005).

6.3 ANALYSIS

The Paris-Durham shock code provides physical (local density, temperature, velocity) and chemical (abundances) parameters, from which the LVG code computes the level populations, specific and integrated intensities by solving the equations of statistical equilibrium. We benchmarked our radiative transfer solution against that of Gusdorf *et al.* (2011) using their shock model and collisional data, and found to be in the 89% agreement with respect to their integrated intensities. In this study we focused on 30 models for varying initial parameters, consisting of 17 models of J-type shocks and 13 of C-type shocks. When it comes to water excitation, we arbitrarily divided water lines into 3 groups: low-excited water lines with $E_{up} < 300$ K, mid-excited lines with 300 K $\leq E_{up} \leq$ 600 K, and highly-excited lines with $E_{up} > 600$ K. This choice was dictated for the presentation purposes and we will present the results for modeled water emission following this classification.

6.3.1 *Considered parameters for shock models*

We explored variations of the following input parameters: the pre-shock density, strength of the transverse magnetic field and UV radiation field. The pre-shock density in the Paris-Durham shock code is defined as $n_H = n(H) + 2n(H_2)$. The transverse magnetic field $B(\mu G)$ is set by the magnetic parameter b and depends on the pre-shock density as $b\sqrt{n_H(cm^{-3})}$. Presented results are for the cosmic-ray ionization rate (ζ) of 10^{-17} s^{-1}, visual extinction (A_v) of 10^{-1}, and PAH abundance (X(PAH)) of 10^{-6}. In the model the standard Interstellar Radiation Field (ISRF) is applied (Mathis *et al.*, 1983) and is modified by the radiation field scaling factor G_0.

In this study we considered pre-shock densities ranging from 10^2 to 10^6 cm^{-3} and UV radiation field with G_0 between 0 and 10^3 for both shock types. For J-type shocks we varied velocities between 5 and 30 km s^{-1}, while for the C-type between 5 and 20 km s^{-1}. In both cases there was a 5 km s^{-1} increment between subsequent velocities. The magnetic parameter was constant for the C-type shock and set to 1. For the J-type shock we applied $b = 0.1$, but for shock velocities of 25 and 30 km s^{-1}, we additionally explored the impact of $b = 1$. We observed that the increased transverse magnetic field in J-type shocks has a negative impact on integrated intensities ($\sim$ one order of magnitude drop for all transitions; see Fig. 6.C.1), specific intensities (the same effect as in integrated intensities, additionally highly-excited water lines start to diminish; see Fig. 6.D.1) and water excitation diagrams ($\sim$ 5 orders of magnitude drop for transitions with $E_{up} \geq 300K$; see Fig. 6.B.1). However, we will not discuss the impact of the magnetic field on J-type shocks further, as this is beyond the scope of this paper.

6.3.2 *Fractional abundance of water*

The shock chemistry is especially important for interpreting the results of expected emission from chosen molecules. Since in this work we are considering externally irradiated shocks, it is important to understand how the presence of a UV field influences the abundance of water. Hence, we have compared the fractional abundance of water as a function of time and the corresponding temperature profiles of the neutral fluid (see Fig. 6.2.1). Here, the fractional abundance is expressed by the ratio of local water density to proton density $n(H_2O)/n_H$.

In a presence of the UV radiation field we see that the stronger the radiation field the lower the pre-shock water abundance. For considered parameter space, the drop in the abundance can be up to 7 orders of magnitude with respect to the absence of the UV field. This effect is of no surprise and showcases photodissociation of water molecules by UV photons. Nevertheless, even if pre-shock water gets dissociated in the presence of the UV field, in shocks the local abundance jumps to the same level as if there is no impinging UV field. This is associated with water reformation in shocks, through gas phase synthesis. For temperatures $\gtrsim 300$ K the following water formation pathway starts to become efficient (Bergin *et al.*, 1998):

$$\begin{aligned} H_2 + O &\rightleftarrows OH + H \\ H_2 + OH &\rightleftarrows H_2O + H. \end{aligned} \tag{6.2}$$

At high gas temperatures ($\gtrsim 2 \times 10^4$ K) H_2 is collisionally dissociated to atomic H. When this occurs, the back reactions in Eq. (6.2) will dominate, and H_2O also dissociates collisionally. Hence, in the presence of the UV radiation field, the water abundance is set by a balance of three reactions: photodissociation, collisional dissociation and gas-phase synthesis. This balance is set by G_0, V_S, and n_H. Moreover, from Fig. 6.2.1 we see that for the highest G_0 there is a drop in the observed water abundance. Even though photodissociation by UV photons is negligible inside the shock, it seems that for $G_0 = 10^3$ the effect becomes significant.

This particular grid was run with the specific goal of exploring H_2 emission. As such, all gas-grain and grain-grain processes were turned off, with the exception of H_2 formation on grains. The implication for this study is that water remains in the gas phase, and does not freeze out. Similarly, there is no water ice on the grains in the first place, and so there is no sputtering of ices in the shocks. Freeze-out likely only plays a role far downstream from the shock front, where the temperature, density, and time are sufficient for water to freeze out. As for sputtering, this only occurs at shock velocities greater than some threshold ($\sim$ 3–10 km s^{-1}; e.g., Tielens *et al.*, 1994; Suutarinen *et al.*, 2014). Thus, for the lowest-velocity shocks, the abundance of water may be over-estimated.

6.3.3 *UV radiation field*

We compared integrated intensities and water spectrum where the strength of the radiation field, G_0, varied between 0 and 10^3 in Habing units (see Fig. 6.4.1 and Fig. 6.4.2).

The impinging UV radiation field results in a drop of ~ 1 and ~ 6 orders of magnitude for the lower excited water lines in integrated intensities for the J- and C-type shocks, respectively (see Fig. 6.4.1). For the mid-excited lines the integrated intensities fall by ~ one order of magnitude for both shock types, while for highly-excited lines there is a negligible impact for the J-type shocks and for the C-type shock there is a ~ one order of magnitude emission drop at the lower end of the highly-excited lines. In excitation diagrams (see Fig. 6.4.3) we see a similar effect, where the increasing G_0 lowers the low-excited lines by ~ 10 orders of magnitude for both shock types.

However, the changes for the C-type shocks seem to be more prominent than for J-type shocks, which can bee see in the local abundance and temperature profiles, illustrating that as a consequence of magnetization, the shock width increases (Draine, 1980). Moreover, for $G_0 = 10^2$, the increased coupling of the ions and neutrals, make C-type shock a C*-type shock, i.e., a C-type shock, where the neutral fluid becomes subsonic along its trajectory. While for $G_0 = 10^3$ what we actually get is a CJ-type shock, where, due to recoupling of neutrals and ions, the C-type shock becomes a J-type shock. However, in this study the J component only has a negligible effect on the medium and for practical purposes we will refer to it as a C-type shock.

From the water spectra we see that the mid- and highly excited lines start to emerge when we increase the strength of the UV radiation field (see Fig. 6.4.2), while the peak intensities of lower-excited lines drop up to ~ one order of magnitude. Hence, the effect of the UV radiation field has a negative impact for transitions with $E_{\rm up} \leq 600$ K.

6.3.4 *Shock velocity*

The input kinetic energy flux in shocks varies as $1/2\rho V_S^3$, and part of this energy is radiated away through, e.g., water emission. The clearly depends strongly on the velocity, and we have looked at velocities ranging from 5 to 30 km s^{-1} for J-type shocks and 5 to 20 km s^{-1} for C-type shocks.

In J-type shocks the temperature is set by shock velocities (temperature changes as velocity squared, following the Rankine–Hugoniot relations), which is directly reflected in the local abundance profile (see Fig. 6.3.1). For velocities $\geq 25 - 30$ km s^{-1} we see signposts of collisional dissociation of water, when the critical temperature for H_2 dissociation, 2×10^4 K, is crossed. For the 30 km s^{-1} shock the drop in abundance is extreme, as it reaches ~ 10 orders of magnitude. This effect is expected as shocks with velocities $\gtrsim$ 25 km s^{-1} are dissociative. However, once the temperature decreases, we see the effects of water reformation, and the abundance returns to 10^{-4} level. Since dissociation appears in a rather thin layer of the shock, its effect on integrated quantities might not be that significant.

For higher velocities, mid- and highly-excited lines emerge for both J- and C- type shocks (see Fig. 6.4.2). For C-type shocks this effect is also important for total intensities of water lines (see Fig. 6.4.1), where highly-excited lines jump by ~ 10 orders of magnitude in intensity, which is also true for water excitation diagrams (see Fig. 6.4.3).

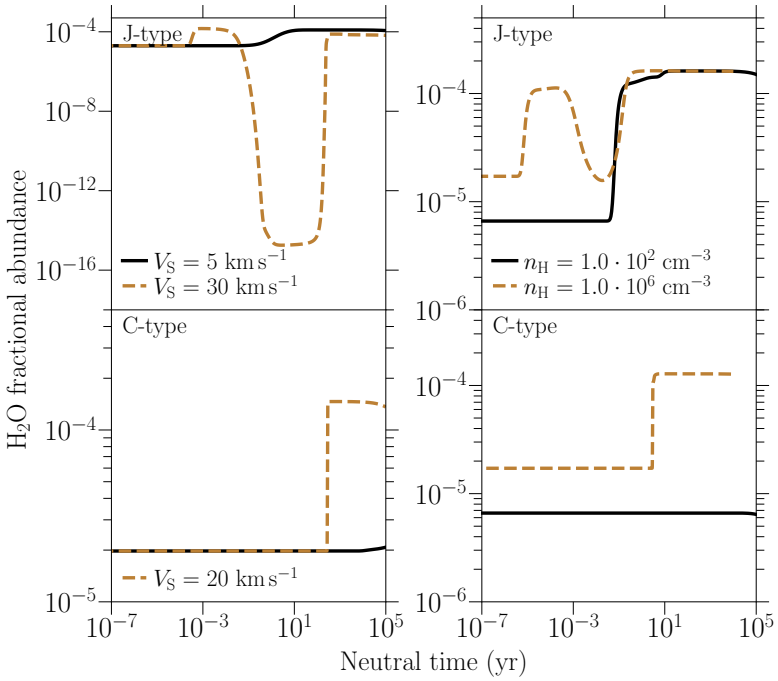


FIGURE 6.3.1: Fractional abundance of H_2O as a function of time for J- and C-type shocks for varying shock parameters (for profiles with varying G_0 see Fig. 6.2.1). Default parameters as in Fig. 6.4.1. Corresponding neutral temperature profiles are presented in Appendix 6.A.

6.3.5 *Pre-shock density*

Higher densities of the pre-shock gas result in a greater number of collisional partners. As mentioned in Sect. 6.3.4 water is mostly excited by collisions. However, even though higher pre-shock densities should result in more collisions and as such more efficient excitation, they also lead to narrower shocks, due to more efficient cooling, which results in lower column densities of water, and possibly lower emission. Moreover, due to its high dipole moment, the critical densities of water are high, e.g., for para-H_2O line at 988 GHz at 300 K (for collisions with para-H_2), $n_{crit} = 8.56 \times 10^7$ cm^{-3}. Hence, high densities are needed to excite H_2O. In this study we considered $n_H = 10^2 - 10^6$ cm^{-3}.

In J-type shocks density impacts the excitation diagrams for all transitions (see Fig. 6.4.3). On the low-excited end the change is of the order of ~ 5 orders of magnitude between the lowest and the highest density, while for the highly-excited end the change can be as high as ~ 15 orders of magnitude. In C-type shocks the change for the highly-excited transitions reaches ~ 75 orders of magnitude. This clearly shows that pre-shock densities, and not shock widths, dominate water emission.

The substantial change experienced in C-type shocks is also very prominent in the water spectra (see Fig. 6.4.2) and integrated intensities (see Fig. 6.4.1). For the lowest density, the change in integrated intensity of low-excited lines is equal to ~ 6 orders of magnitude, while for the highly-excited ones it is ~ 30 orders of magnitude. In J-type shocks the change is lower, yet significant with an increase of ~ 3 orders of magnitude for the low-excited lines and ~ 6 orders of magnitude for the highly-excited ones.

Water spectra reveal that higher densities translate to more brighter lines. For the C-type shock, the change is very prominent, as one low-excited line is replaced by a forest of lines. When it comes to the change in the intensity of the spectrum, the change is by at least ~ 4 orders of magnitude for the J-type shock and ~ 5 orders of magnitude for the C-type shock.

However, the drastic changes for C-type shock values can be explained by the fact that at the lowest density the effects of the shock are not strong. Hence, there is no jump in temperature and the abundance change is negligible. Under this conditions water excitation should not be expected and we see that being reflected in excitation diagrams and intensities.

6.4 CONCLUSIONS

In star forming regions the bulk of water emission is associated with shocked regions around protostars. In this study we have presented results of the radiative transfer modeling of water emission and excitation in J- and C-type shocks from the Paris-Durham shock code. We varied the initial conditions ($n_{\rm H} = 10^2 - 10^6$ cm^{-3}, $G_0 = 0 - 1000$, and $b = 0.1 - 1.0$) and shock velocities ($V_{\rm S} = 5 - 30$ km s^{-1}). The main conclusions are as follows:

- J-type shocks are on average associated with higher water emission and excitation than their C-type counterparts with the same pre-shock density and shock velocity;
- an external UV radiation field has a negative impact on low- and mid-excited lines;
- In J-type shocks the temperature is set by the shock velocity, and for velocities ≥ 25 km s^{-1} the local abundance profiles reveal clear signs of the collisional dissociation of water molecules. In C-type shocks the effect of velocity is especially important for the emergence of mid- and highly-excited transitions;
- Pre-shock densities can have significant impact on intensities and excitation diagrams of water for both shock types. For C-type shocks this effect can be as high as 75 orders of magnitude for the highly-excited lines.

Our radiative-transfer solutions provide insights into the nature of water emission and excitation in shocked regions. These conditions can be applied to Galactic shock regions and help us better understand the observed emission. Moreover, since our solutions assume the presence of an external UV radiation field, they can be especially important for understanding the nature of molecular excitation in a presence of high-mass stars. For future modeling of water emission, the impact of cosmic ray ionization and far-infrared pumping should be evaluated. Cosmic ray ionization affects the ionization and, indirectly, the physical and chemical conditions of the shock. Meanwhile, far-infrared pumping influences the excitation of water. This could be important for assessing molecular excitation in the extragalactic regime, where these two processes are thought to play an important role in water excitation.

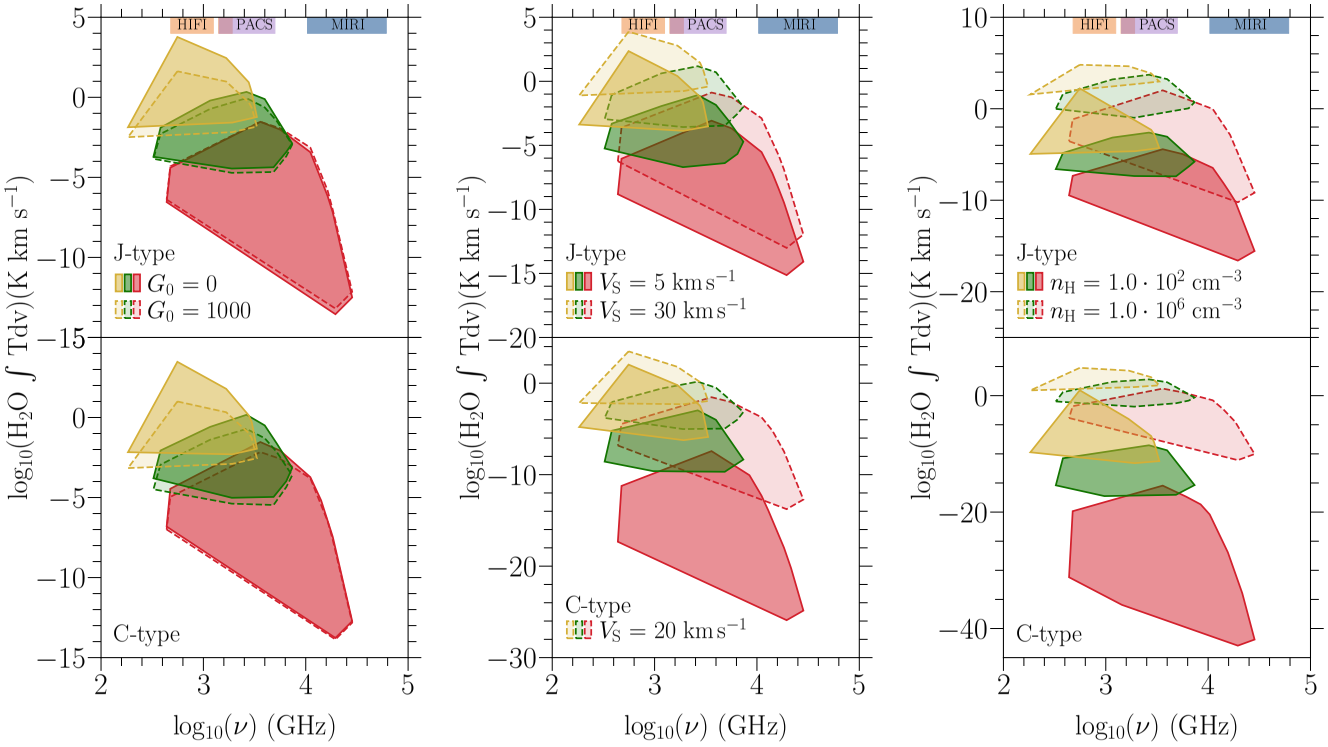


FIGURE 6.4.1: Total integrated intensity of H_2O for J- and C-type shocks for varying initial and shock parameters. The full set of default parameters is $n_H = 10^4$ cm^{-3},$V_S = 20$ km s^{-1}, and $G_0 = 0$. Yellow convex hulls represent low-excited water lines, green mid-excited lines, and red highly-excited ones (see the beginning of Sect. 6.3 for definitions of these lines). At the top of each panel the *Herschel* HIFI and PACS, and the JWST's MIRI ranges are marked. Examples of delineated integrated intensities plots can be seen in Appendix 6.C.

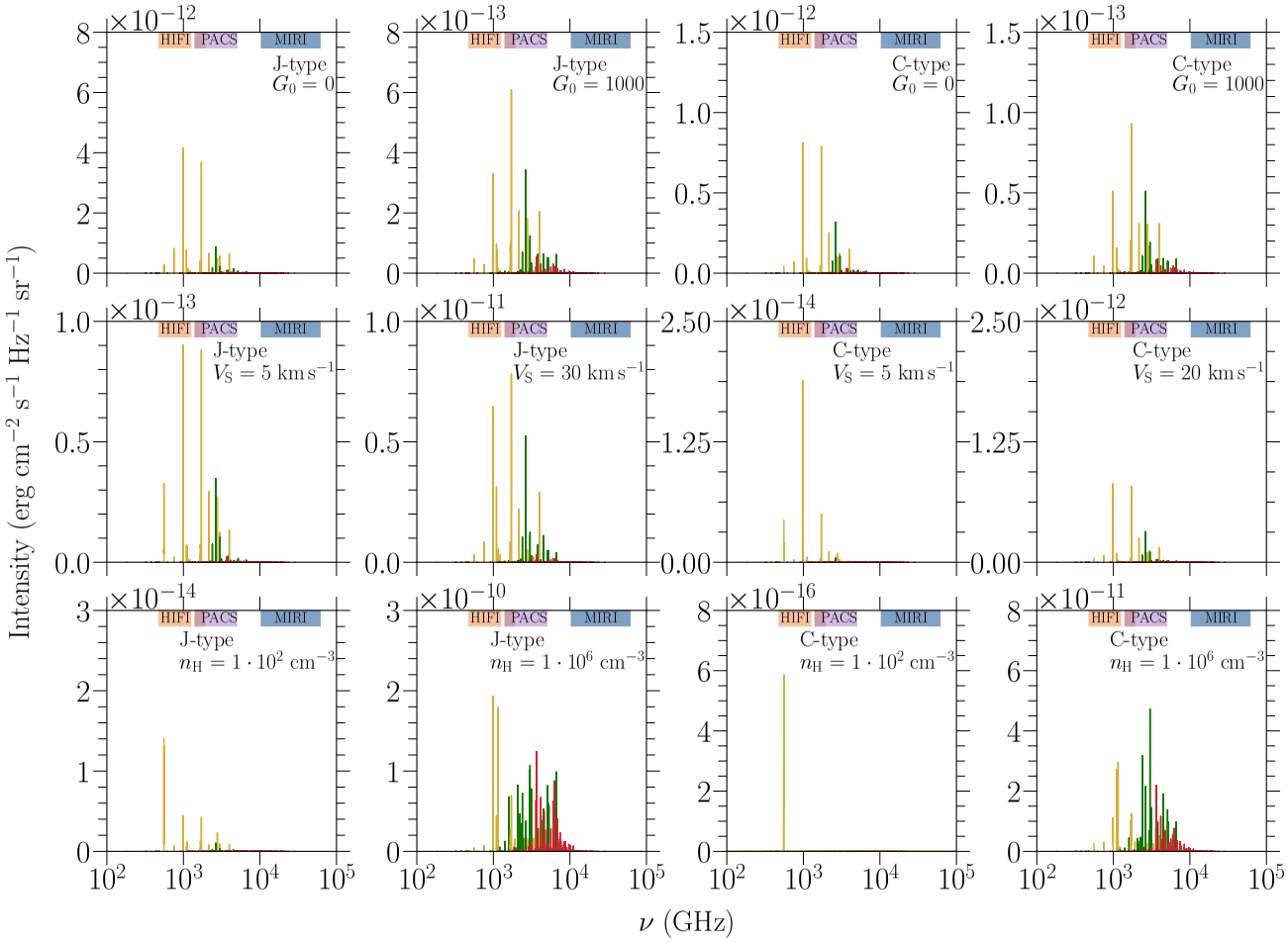


FIGURE 6.4.2: H_2O specific intensities for J- and C-type shocks for varying initial and shock parameters. Default parameters, colors and frequency ranges in Fig. 6.4.1.

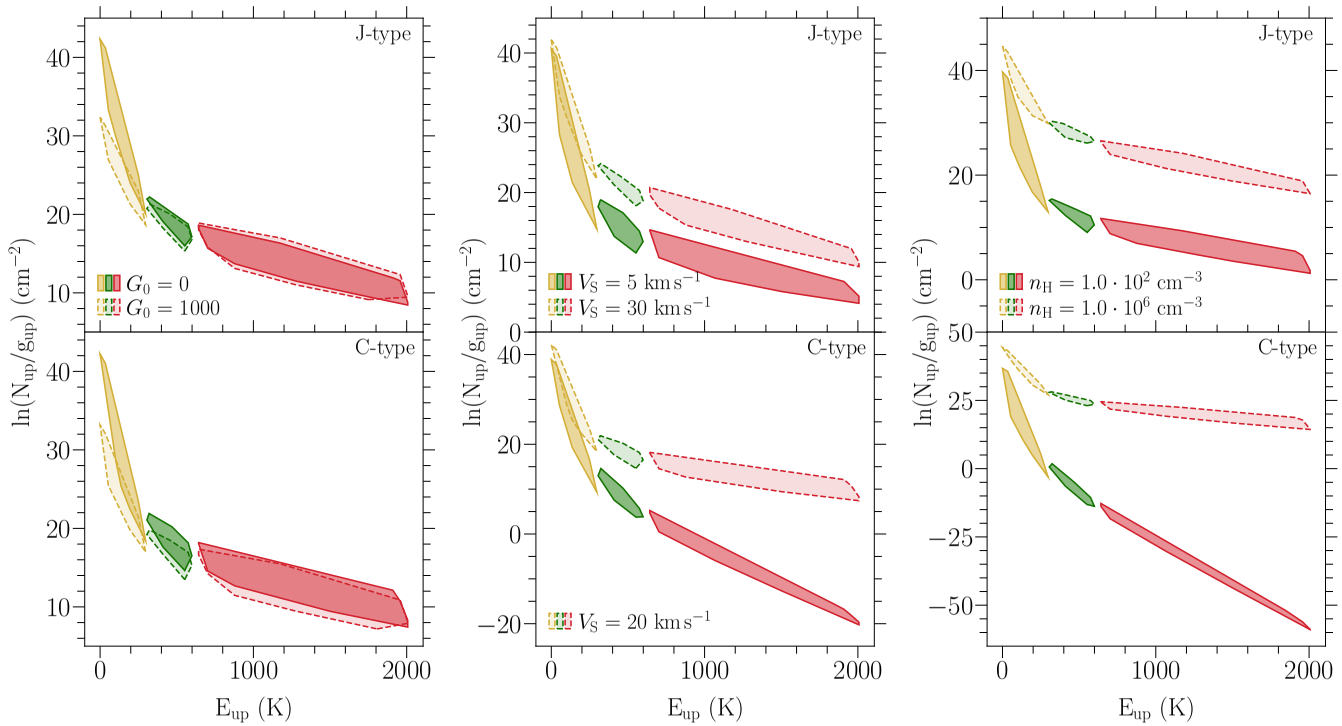


FIGURE 6.4.3: H_2O excitation diagrams for J- and C-type shocks for varying initial and shock parameters. Default parameters and colors as in Fig. 6.4.1. Examples of delineated excitation diagrams can be seen in Appendix 6.B.

Appendix

6.a FRACTIONAL WATER ABUNDANCE AND NEUTRAL TEMPERATURE

In this section we present local abundance profiles of H_2O and neutral temperature profiles as a function of time for J- and C-type shocks with varying shock velocities (Fig. 6.A.1 and 6.A.3, respectively), as well as for the J-type shock with varying magnetic parameter, b (Fig. 6.A.2).

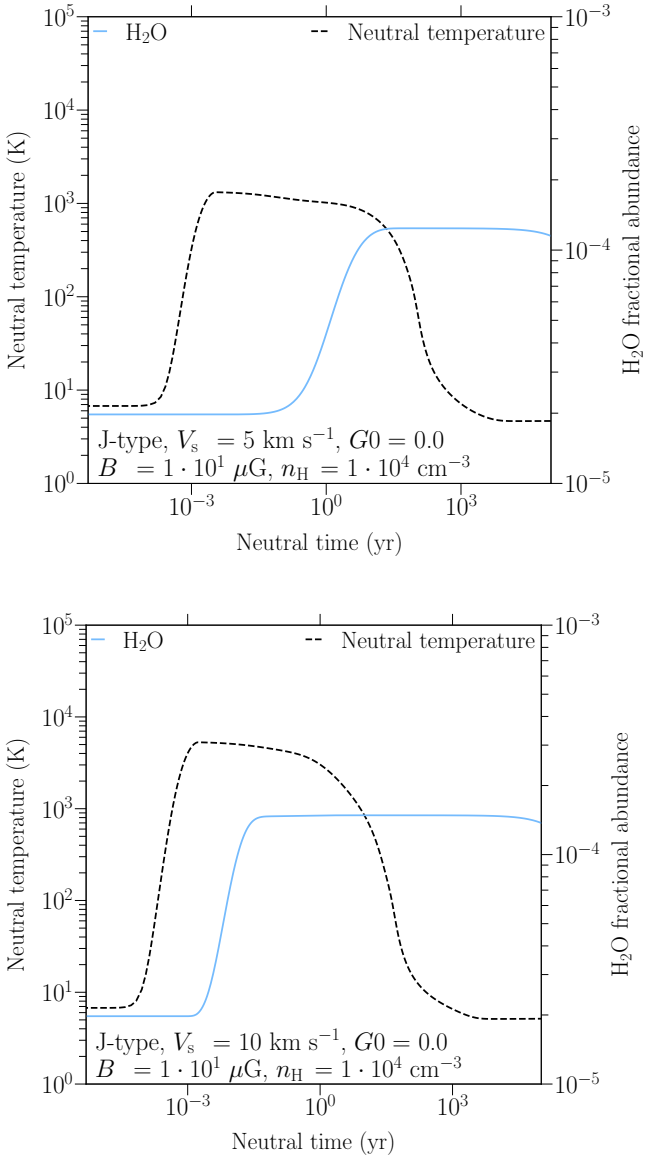


FIGURE 6.A.1: Fractional abundance of H_2O and neutral temperature as a function of time for J-type shocks (with $n_H = 10^4$ cm^{-3} and $b = 0.1$) for shock velocities, V_S, varying between 5 – 30 km s^{-1}, where the UV pumping was not considered.

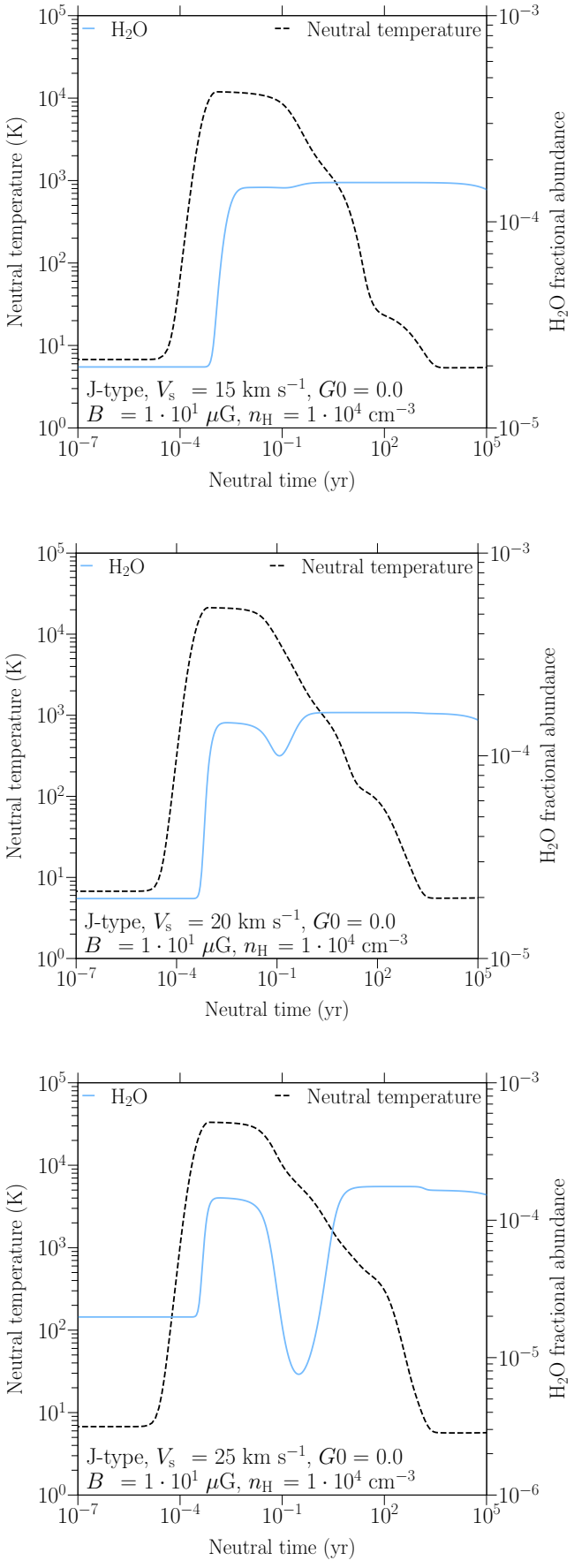


FIGURE 6.A.1: Fractional abundance of H_2O and neutral temperature as a function of time for J-type shocks (with $n_H = 10^4$ cm^{-3} and $b = 0.1$) for shock velocities, V_S, varying between 5 – 30 km s^{-1}, where the UV pumping was not considered.

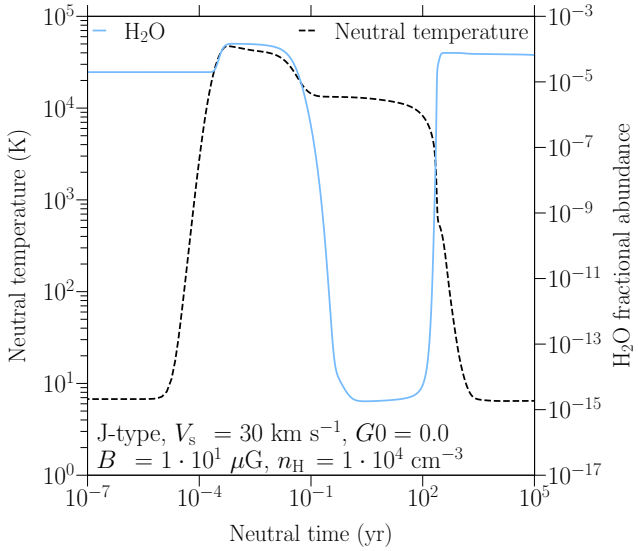


FIGURE 6.A.1: Fractional abundance of H_2O and neutral temperature as a function of time for J-type shocks (with $n_H = 10^4$ cm^{-3} and $b = 0.1$) for shock velocities, V_S, varying between 5 – 30 km s^{-1}, where the UV pumping was not considered.

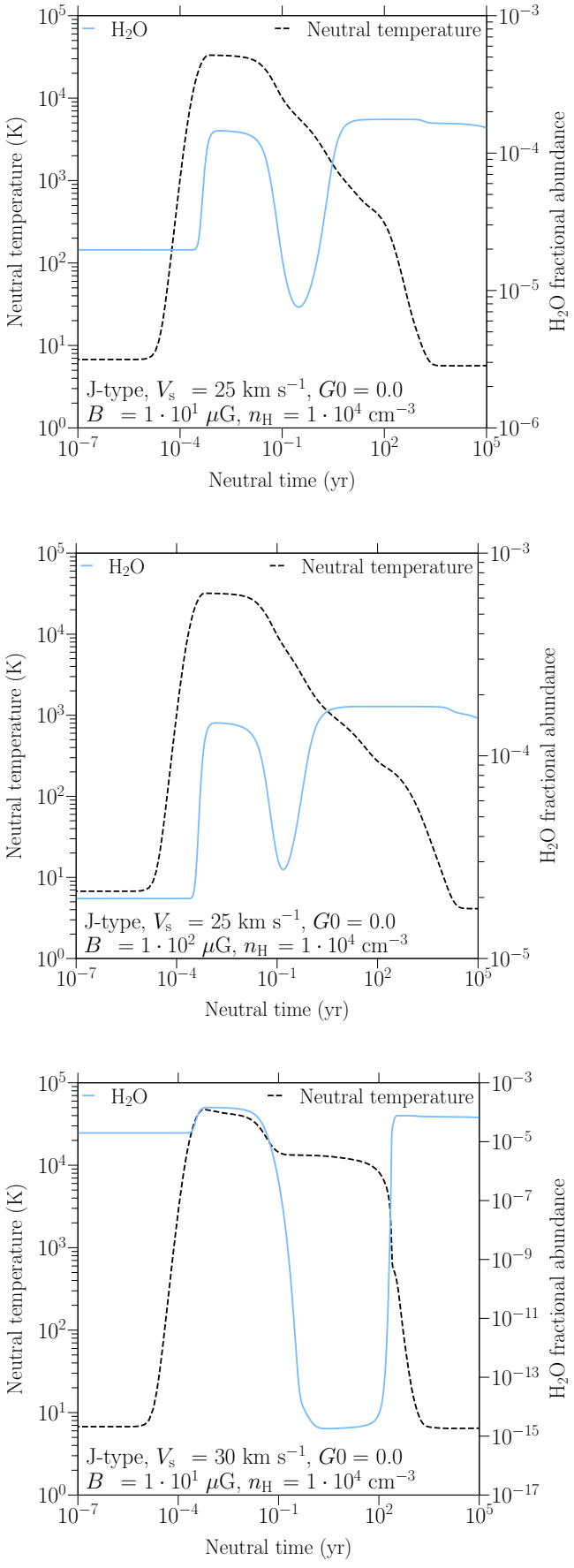


FIGURE 6.A.2: Fractional abundance of H_2O and neutral temperature as a function of time for J-type shocks (n_H = 10^4 cm^{-3} and in a lack of UV pumping) with velocities V_S = 25 km s^{-1} and V_S = 30 km s^{-1}, where magnetic parameter, b, varied between 0.1 and 1, hence, increasing the strength of the transverse magnetic field, B_0, by an order of magnitude.

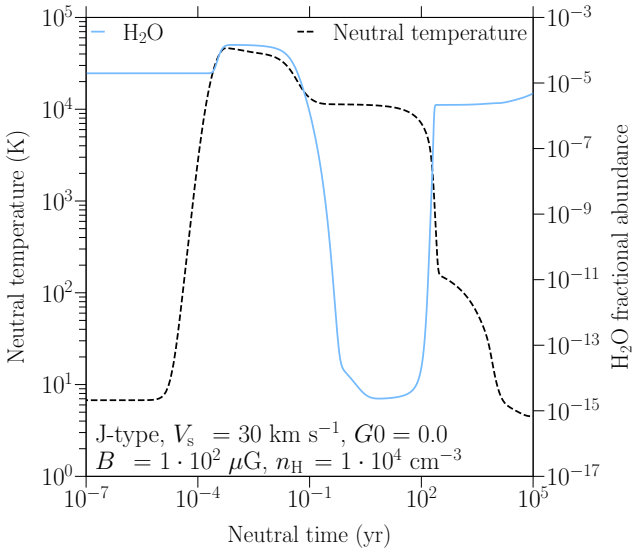


FIGURE 6.A.2: Fractional abundance of H_2O and neutral temperature as a function of time for J-type shocks ($n_H = 10^4$ cm^{-3} and in a lack of UV pumping) with velocities $V_S = 25$ km s^{-1} and $V_S = 30$ km s^{-1}, where magnetic parameter, b, varied between 0.1 and 1, hence, increasing the strength of the transverse magnetic field, B_0, by an order of magnitude.

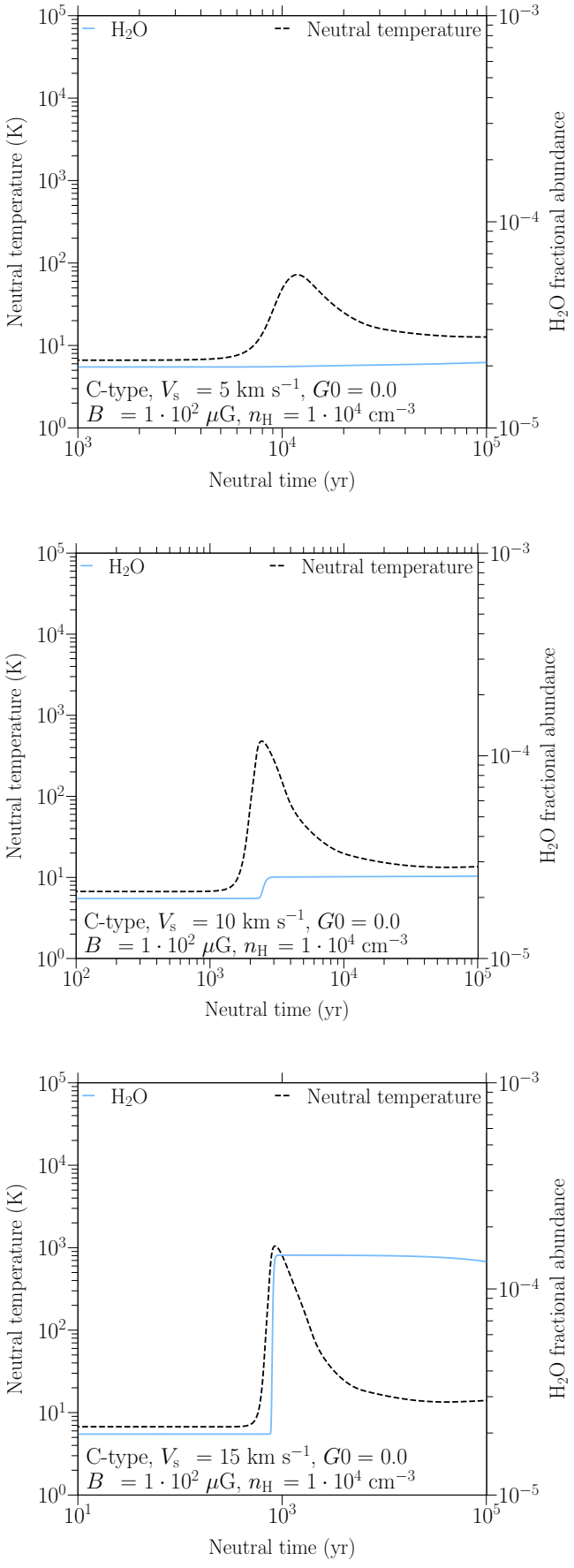


FIGURE 6.A.3: Fractional abundance of H_2O and neutral temperature as a function of time for C-type shocks (with $n_H = 10^4$ cm^{-3} and $b = 1$) for shock velocities, V_S, varying between 5 – 20 km s^{-1}, where the UV pumping was not considered.

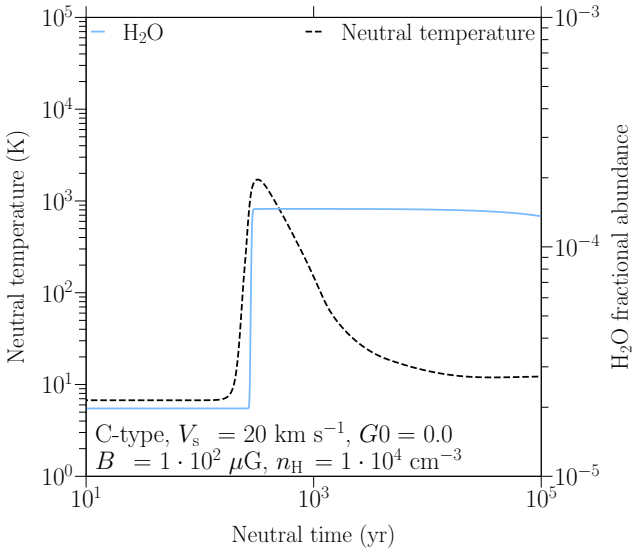


FIGURE 6.A.3: Fractional abundance of H_2O and neutral temperature as a function of time for C-type shocks (with $n_H = 10^4$ cm^{-3} and $b = 1$) for shock velocities, V_S, varying between $5 - 20$ km s^{-1}, where the UV pumping was not considered.

6.b EXCITATION DIAGRAMS

Here, we present derived water excitation diagrams for J-type shocks with varying magnetic parameter, b (Fig. 6.B.1).

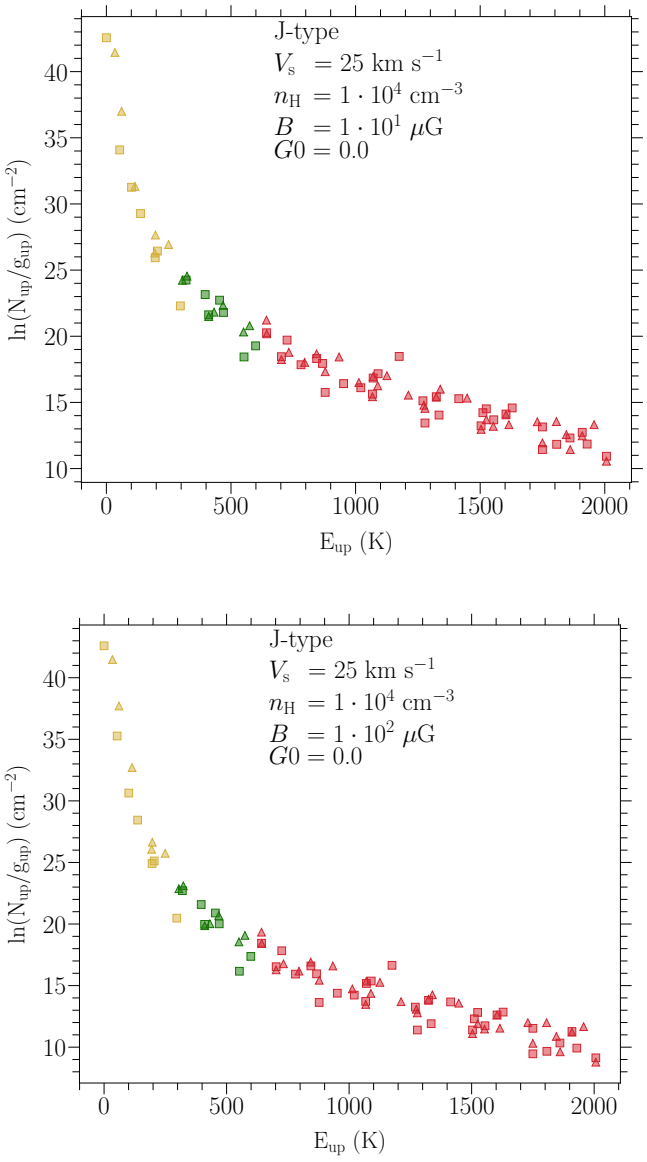


FIGURE 6.B.1: H_2O excitation diagrams for J-type shocks ($n_H = 10^4$ cm^{-3} and in a lack of UV pumping) with velocities $V_S = 25$ km s^{-1} and $V_S = 30$ km s^{-1}, where magnetic parameter, b, varied between 0.1 and 1, hence, increasing the strength of the transverse magnetic field, B_0, by an order of magnitude. Yellow points represent low-excited water lines, green mid-excited lines, and red highly-excited ones. Squares correspond to para-H_2 lines, while triangles to ortho-H_2O lines.

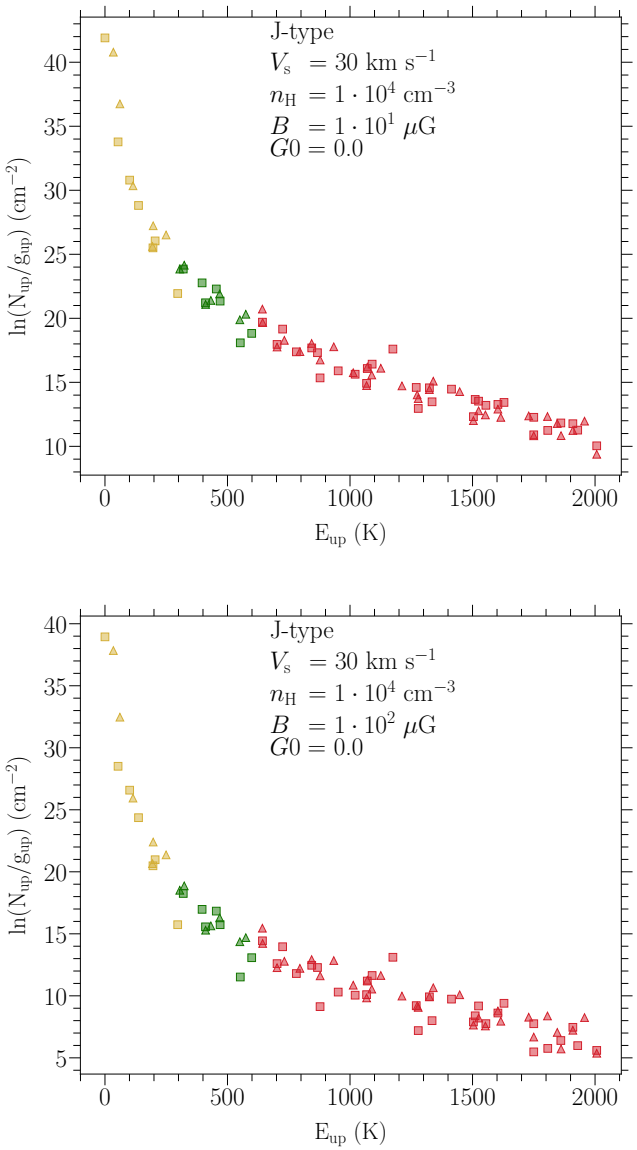


FIGURE 6.B.1: H_2O excitation diagrams for J-type shocks ($n_H = 10^4$ cm^{-3} and in a lack of UV pumping) with velocities $V_S = 25$ km s^{-1} and $V_S = 30$ km s^{-1}, where magnetic parameter, b, varied between 0.1 and 1, hence, increasing the strength of the transverse magnetic field, B_0, by an order of magnitude. Yellow points represent low-excited water lines, green mid-excited lines, and red highly-excited ones. Squares correspond to para-H_2 lines, while triangles to ortho-H_2O lines.

6.c INTEGRATED WATER INTENSITIES

This section showcases integrated intensities of water for J-type shocks with varying magnetic parameter, b (Fig. 6.C.1).

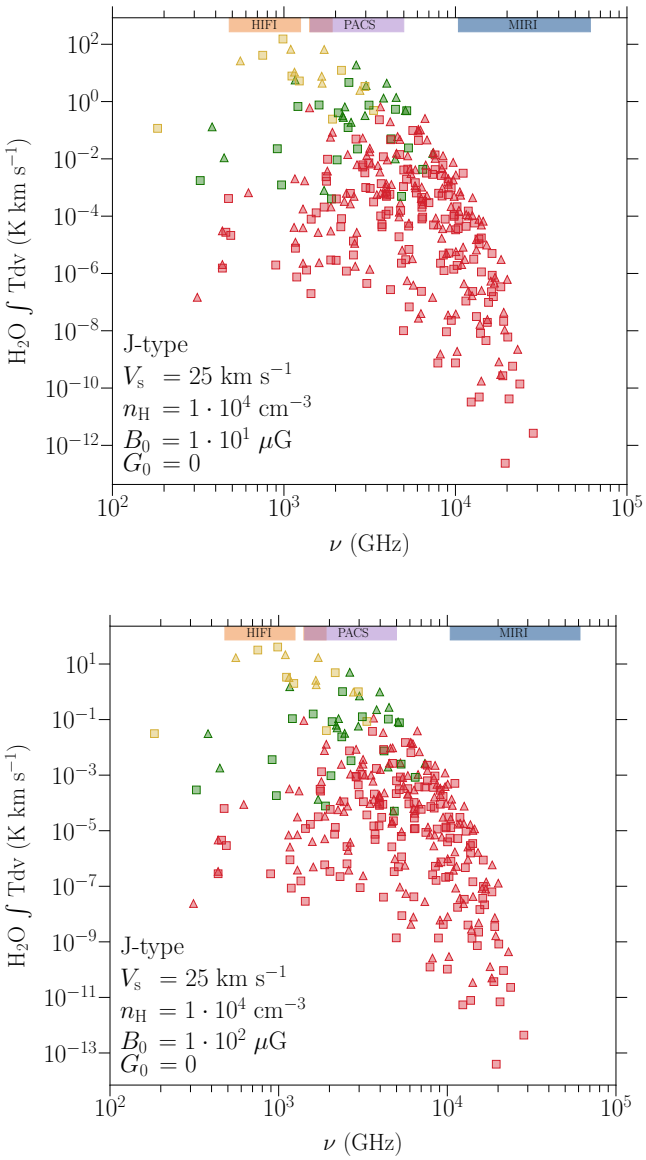


FIGURE 6.C.1: Integrated intensities of H_2O for J-type shocks ($n_H = 10^4$ cm^{-3} and in a lack of UV pumping) with velocities $V_S = 25$ km s^{-1} and $V_S = 30$ km s^{-1}, where magnetic parameter, b, varied between 0.1 and 1, hence, increasing the strength of the transverse magnetic field, B_0, by an order of magnitude. Yellow points represent low-excited water lines, green mid-excited lines, and red highly-excited ones. Squares correspond to para-H_2 lines, while triangles to ortho-H_2O lines. At the top of each plot the *Herschel's* HIFI and PACS, and the JWST's MIRI ranges are marked.

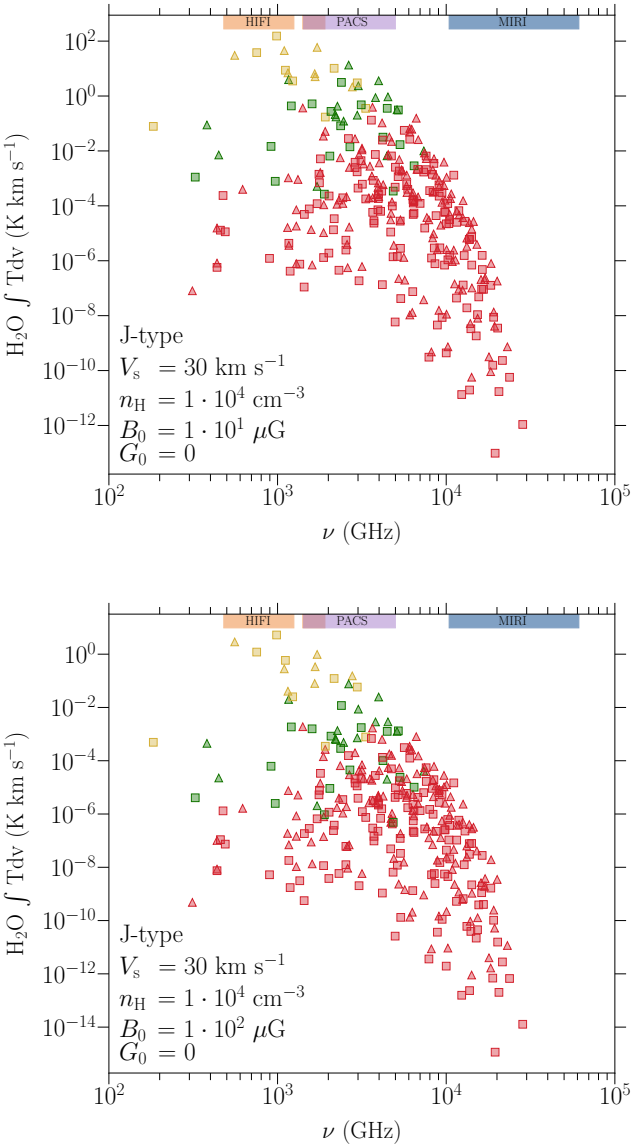


FIGURE 6.C.1: Integrated intensities of H_2O for J-type shocks ($n_H = 10^4$ cm^{-3} and in a lack of UV pumping) with velocities $V_S = 25$ km s^{-1} and $V_S = 30$ km s^{-1}, where magnetic parameter, b, varied between 0.1 and 1, hence, increasing the strength of the transverse magnetic field, B_0, by an order of magnitude. Yellow points represent low-excited water lines, green mid-excited lines, and red highly-excited ones. Squares correspond to para-H_2 lines, while triangles to ortho-H_2O lines. At the top of each plot the *Herschel's* HIFI and PACS, and the JWST's MIRI ranges are marked.

6.d WATER SPECTRA

Below, we present water spectra for J-type shocks with varying magnetic parameter, b (Fig. 6.D.1).

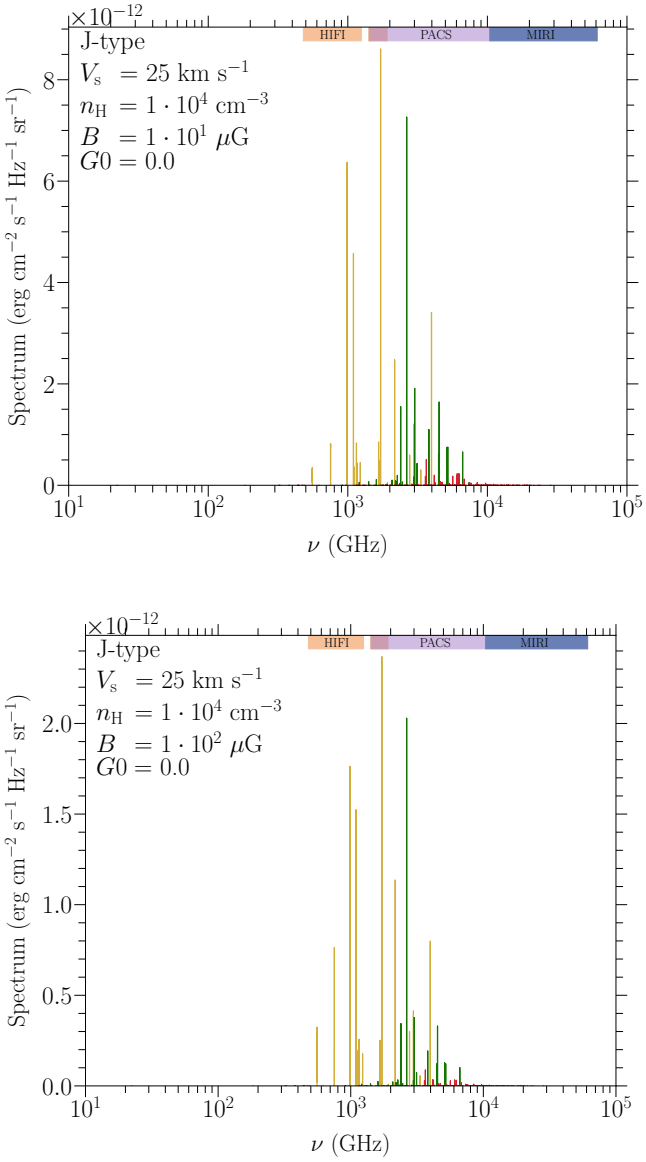


FIGURE 6.D.1: H_2O specific intensities for J-type shocks ($n_H = 10^4$ cm^{-3} and in a lack of UV pumping) with velocities V_S = 25 km s^{-1} and V_S = 30 km s^{-1}, where magnetic parameter, b, varied between 0.1 and 1, hence, increasing the strength of the transverse magnetic field, B_0, by an order of magnitude. Yellow lines represent low-excited water lines, green mid-excited lines, and red highly-excited ones. At the top of each plot the *Herschel's* HIFI and PACS, and the JWST's MIRI ranges are marked.

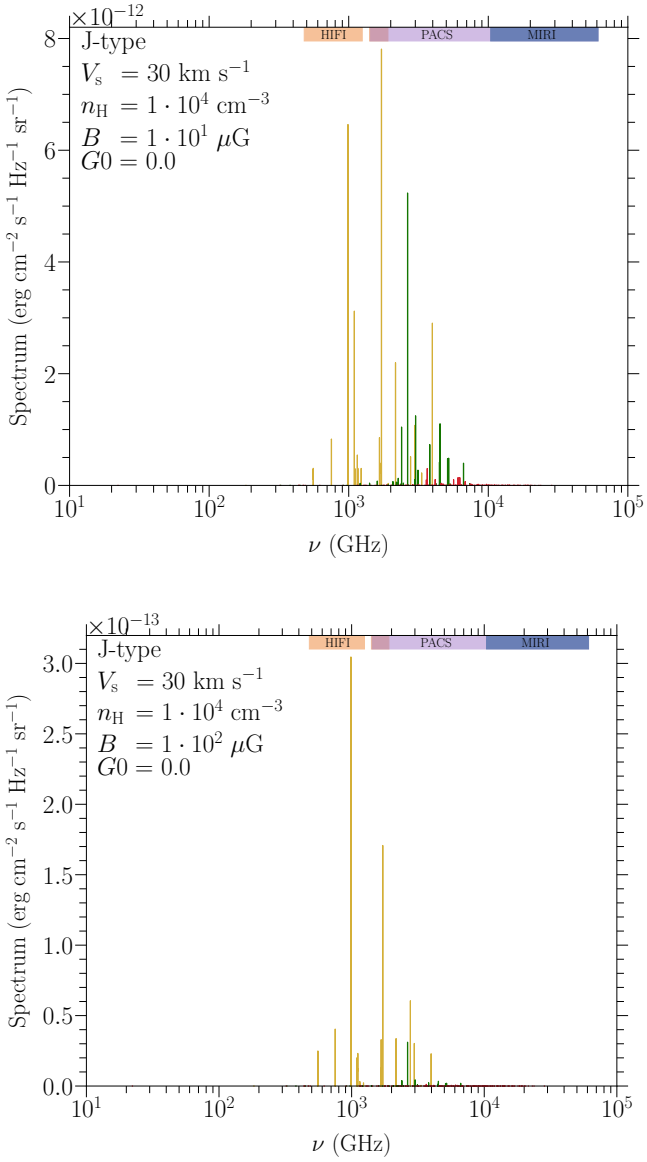


FIGURE 6.D.1: H_2O specific intensities for J-type shocks ($n_H = 10^4$ cm^{-3} and in a lack of UV pumping) with velocities $V_S = 25$ km s^{-1} and $V_S = 30$ km s^{-1}, where magnetic parameter, b, varied between 0.1 and 1, hence, increasing the strength of the transverse magnetic field, B_0, by an order of magnitude. Yellow lines represent low-excited water lines, green mid-excited lines, and red highly-excited ones. At the top of each plot the *Herschel's* HIFI and PACS, and the JWST's MIRI ranges are marked.

Bibliography

1. A. A. Abdo *et al.*, *A&A* **512**, A7, arXiv: 1001.3298 [astro-ph.HE] (Mar. 2010).
2. A. Adamo *et al.*, *ApJ* **766**, 105, arXiv: 1302.2149 [astro-ph.CO] (Apr. 2013).
3. A. Adamo *et al.*, *Space Sci. Rev.* **216**, 69, arXiv: 2005.06188 [astro-ph.GA] (June 2020).
4. F. C. Adams, *ARA&A* **48**, 47–85, arXiv: 1001.5444 [astro-ph.SR] (Sept. 2010).
5. F. C. Adams, M. Fatuzzo, L. Holden, *ApJ* **789**, 86, arXiv: 1405.5142 [astro-ph.SR] (July 2014).
6. F. C. Adams *et al.*, *ApJ* **641**, 504–525, arXiv: astro-ph/0512330 [astro-ph] (Apr. 2006).
7. V. Agra-Amboage *et al.*, *A&A* **564**, A11, arXiv: 1402.1160 [astro-ph.SR] (Apr. 2014).
8. P. André *et al.*, *A&A* **518**, L102, arXiv: 1005.2618 [astro-ph.GA] (July 2010).
9. Y. Apostolovski *et al.*, *A&A* **628**, A23, arXiv: 1905.12738 [astro-ph.GA] (Aug. 2019).
10. H. G. Arce *et al.*, presented at the Protostars and Planets V, ed. by B. Reipurth, D. Jewitt, K. Keil, p. 245, arXiv: astro-ph/0603071 [astro-ph].
11. H. G. Arce, A. I. Sargent, *ApJ* **646**, 1070–1085, arXiv: astro-ph/0605139 [astro-ph] (Aug. 2006).
12. L. Armillotta, E. C. Ostriker, Y.-F. Jiang, *ApJ* **929**, 170, arXiv: 2203.11949 [astro-ph.GA] (Apr. 2022).
13. D. Arzoumanian *et al.*, *A&A* **529**, L6, arXiv: 1103.0201 [astro-ph.GA] (May 2011).
14. Y. Aso *et al.*, *ApJ* **812**, 27, arXiv: 1508.07013 [astro-ph.SR] (Oct. 2015).
15. R. Bachiller *et al.*, *A&A* **231**, 174–186 (May 1990).
16. J. Bally, *ARA&A* **54**, 491–528 (Sept. 2016).
17. A. Banzatti, PhD thesis, Eidgenossische Technische Hochschule, Zurich, Switzerland, Mar. 2013.
18. N. Bastian, K. R. Covey, M. R. Meyer, *ARA&A* **48**, 339–389, arXiv: 1001.2965 [astro-ph.GA] (Sept. 2010).
19. M. R. Bate, *MNRAS* **475**, 5618–5658, arXiv: 1801.07721 [astro-ph.SR] (Apr. 2018).
20. E. A. Bergin *et al.*, *A&A* **521**, L20, arXiv: 1007.2172 [astro-ph.GA] (Oct. 2010).
21. E. A. Bergin, G. J. Melnick, D. A. Neufeld, *ApJ* **499**, 777–792, arXiv: astro-ph/9803330 [astro-ph] (May 1998).
22. E. A. Bergin, M. Tafalla, *ARA&A* **45**, 339–396, arXiv: 0705.3765 [astro-ph] (Sept. 2007).
23. J. Bernard-Salas *et al.*, *ApJ* **652**, L29–L32, arXiv: astro-ph/0609299 [astro-ph] (Nov. 2006).
24. S. Bianchi, *A&A* **471**, 765–773, arXiv: 0705.1471 [astro-ph] (Sept. 2007).
25. A. Bik *et al.*, *ApJ* **744**, 87, arXiv: 1109.3467 [astro-ph.GA] (Jan. 2012).
26. J. Binney, S. Tremaine, *Galactic dynamics*.
27. S. E. Bisschop *et al.*, *A&A* **465**, 913–929, arXiv: astro-ph/0702066 [astro-ph] (Apr. 2007).
28. P. Bjerkeli *et al.*, *A&A* **546**, A29, arXiv: 1209.0294 [astro-ph.GA] (Oct. 2012).
29. P. Bjerkeli *et al.*, *Nature* **540**, 406–409, arXiv: 1612.05148 [astro-ph.SR] (Dec. 2016).
30. J. Bland-Hawthorn, O. Gerhard, *ARA&A* **54**, 529–596, arXiv: 1602.07702 [astro-ph.GA] (Sept. 2016).
31. W. B. Bonnor, *ZAp* **39**, 143 (1956).
32. S. Bontemps *et al.*, *A&A* **311**, 858–872 (July 1996).
33. M. Cappellari *et al.*, *Nature* **484**, 485–488, arXiv: 1202.3308 [astro-ph.CO] (Apr. 2012).
34. C. L. Carilli, F. Walter, *ARA&A* **51**, 105–161, arXiv: 1301.0371 [astro-ph.CO] (Aug. 2013).
35. V. Casasola *et al.*, *A&A* **605**, A18, arXiv: 1706.05351 [astro-ph.GA] (Sept. 2017).
36. C. M. Casey, D. Narayanan, A. Cooray, *Phys. Rep.* **541**, 45–161, arXiv: 1402.1456 [astro-ph.CO] (Aug. 2014).
37. C. Ceccarelli *et al.*, *A&A* **521**, L22 (Oct. 2010).
38. J. Cernicharo, J. Crovisier, *Space Sci. Rev.* **119**, 29–69 (Aug. 2005).
39. R. Cesaroni *et al.*, *A&A* **509**, A50 (Jan. 2010).
40. G. Chabrier, *PASP* **115**, 763–795, arXiv: astro-ph/0304382 [astro-ph] (July 2003).

41. G. Chabrier, P. Hennebelle, S. Charlot, *ApJ* **796**, 75, arXiv: 1409.8466 [astro-ph.GA] (Dec. 2014).
42. H. Chen *et al.*, *ApJ* **445**, 377 (May 1995).
43. D. F. Chernoff, *ApJ* **312**, 143 (Jan. 1987).
44. K. J. Chuang *et al.*, *MNRAS* **467**, 2552–2565, arXiv: 1705.07680 [astro-ph.GA] (May 2017).
45. E. Churchwell *et al.*, *A&A* **54**, 969–971 (Feb. 1977).
46. R. Ciardullo *et al.*, *ApJ* **577**, 31–50, arXiv: astro-ph/0206177 [astro-ph] (Sept. 2002).
47. D. Colombo *et al.*, *ApJ* **784**, 4, arXiv: 1401.3759 [astro-ph.GA] (Mar. 2014).
48. F. Combes *et al.*, *A&A* **538**, L4, arXiv: 1201.2908 [astro-ph.CO] (Feb. 2012).
49. J. J. Condon, S. M. Ransom, *Essential Radio Astronomy*.
50. C. J. Conselice, *ARA&A* **52**, 291–337, arXiv: 1403.2783 [astro-ph.GA] (Aug. 2014).
51. A. Crapsi *et al.*, *A&A* **470**, 221–230, arXiv: 0705.0471 [astro-ph] (July 2007).
52. N. Da Rio, J. C. Tan, K. Jaehnig, *ApJ* **795**, 55, arXiv: 1409.2503 [astro-ph.GA] (Nov. 2014).
53. F. Daniel, M. L. Dubernet, A. Grosjean, *A&A* **536**, A76 (Dec. 2011).
54. F. Daniel *et al.*, *MNRAS* **446**, 2312–2316, arXiv: 1410.8021 [astro-ph.GA] (Jan. 2015).
55. R. Davé, *MNRAS* **385**, 147–160, arXiv: 0710.0381 [astro-ph] (Mar. 2008).
56. M. Dessauges-Zavadsky, A. Adamo, *MNRAS* **479**, L118–L122, arXiv: 1806.07922 [astro-ph.GA] (Sept. 2018).
57. M. Dessauges-Zavadsky *et al.*, *Nature Astronomy* **3**, 1115–1121, arXiv: 1909.08010 [astro-ph.GA] (Sept. 2019).
58. O. Dionatos *et al.*, *A&A* **641**, A36, arXiv: 2006.01087 [astro-ph.SR] (Sept. 2020).
59. B. T. Draine, *ApJ* **241**, 1021–1038 (Nov. 1980).
60. B. T. Draine, W. G. Roberge, A. Dalgarno, *ApJ* **264**, 485–507 (Jan. 1983).
61. M. L. Dubernet *et al.*, *A&A* **497**, 911–925 (Apr. 2009).
62. M. M. Dunham *et al.*, in *Protostars and Planets VI* (Beuther, Henrik et al., 2014), p. 195.
63. K. M. Dutkowska, L. E. Kristensen, *A&A* **667**, A135, arXiv: 2206.01753 [astro-ph.GA] (Nov. 2022).
64. R. Ebert, *ZAp* **37**, 217 (1955).
65. J. H. Elias, *ApJ* **224**, 857–872 (Sept. 1978).
66. I. Evans Neal J. *et al.*, *ApJS* **181**, 321–350, arXiv: 0811.1059 [astro-ph] (Apr. 2009).
67. A. C. Fabian, *ARA&A* **50**, 455–489, arXiv: 1204.4114 [astro-ph.CO] (Sept. 2012).
68. M. A. Fardal *et al.*, *MNRAS* **379**, 985–1002, arXiv: astro-ph/0604534 [astro-ph] (Aug. 2007).
69. A. Faure, J. D. Gorfinkiel, J. Tennyson, *MNRAS* **347**, 323–333 (Jan. 2004).
70. E. C. Fayolle *et al.*, *ApJ* **739**, L36, arXiv: 1109.0281 [astro-ph.GA] (Oct. 2011).
71. D. F. Figer, *Nature* **434**, 192–194, arXiv: astro-ph/0503193 [astro-ph] (Mar. 2005).
72. D. J. Fixsen, *ApJ* **707**, 916–920, arXiv: 0911.1955 [astro-ph.CO] (Dec. 2009).
73. D. R. Flower, G. Pineau des Forêts, *MNRAS* **343**, 390–400 (Aug. 2003).
74. D. R. Flower, G. Pineau des Forêts, *MNRAS* **406**, 1745–1758 (Aug. 2010).
75. D. R. Flower, G. Pineau des Forêts, *MNRAS* **436**, 2143–2150 (Dec. 2013).
76. D. R. Flower, G. Pineau des Forets, T. W. Hartquist, *MNRAS* **216**, 775–794 (Oct. 1985).
77. N. M. Förster Schreiber, S. Wuyts, *ARA&A* **58**, 661–725, arXiv: 2010.10171 [astro-ph.GA] (Aug. 2020).
78. H. J. Fraser *et al.*, *MNRAS* **327**, 1165–1172, arXiv: astro-ph/0107487 [astro-ph] (Nov. 2001).
79. R. T. Garrod, S. L. Widicus Weaver, E. Herbst, *ApJ* **682**, 283–302, arXiv: 0803.1214 [astro-ph] (July 2008).
80. P. D. Gensheimer, R. Mauersberger, T. L. Wilson, *A&A* **314**, 281–294 (Oct. 1996).
81. B. Godard *et al.*, *A&A* **622**, A100, arXiv: 1901.04273 [astro-ph.GA] (Feb. 2019).
82. J. R. Goicoechea *et al.*, *ApJ* **799**, 102, arXiv: 1411.2930 [astro-ph.GA] (Jan. 2015).
83. P. F. Goldsmith, W. D. Langer, *ApJ* **222**, 881–895 (June 1978).
84. P. F. Goldsmith, *ApJ* **557**, 736–746 (Aug. 2001).
85. P. F. Goldsmith, W. D. Langer, *ApJ* **517**, 209–225 (May 1999).

86. E. González-Alfonso *et al.*, *A&A* **518**, L43, arXiv: 1005.3642 [astro-ph.CO] (July 2010).
87. E. González-Alfonso *et al.*, *A&A* **567**, A91, arXiv: 1406.5110 [astro-ph.GA] (July 2014).
88. E. González-Alfonso *et al.*, *ApJ* **675**, 303–315, arXiv: 0711.2281 [astro-ph] (Mar. 2008).
89. E. González-Alfonso *et al.*, *A&A* **666**, L3, arXiv: 2209.06444 [astro-ph.GA] (Oct. 2022).
90. R. J. Gould, E. E. Salpeter, *ApJ* **138**, 393 (Aug. 1963).
91. J. D. Green *et al.*, *ApJ* **770**, 123, arXiv: 1304.7389 [astro-ph.GA] (June 2013).
92. T. P. Greene *et al.*, *Journal of Astronomical Telescopes, Instruments, and Systems* **3**, 035001 (July 2017).
93. A. Gusdorf *et al.*, *A&A* **532**, A53, arXiv: 1108.3692 [astro-ph.GA] (Aug. 2011).
94. A. Gusdorf *et al.*, *A&A* **575**, A98, arXiv: 1502.00488 [astro-ph.GA] (Mar. 2015).
95. D. Guszejnov, P. F. Hopkins, M. Y. Grudić, *MNRAS* **477**, 5139–5149, arXiv: 1707.05799 [astro-ph.GA] (July 2018).
96. R. A. Gutermuth *et al.*, *ApJ* **739**, 84, arXiv: 1107.0966 [astro-ph.SR] (Oct. 2011).
97. A. Hacar *et al.*, *A&A* **610**, A77, arXiv: 1801.01500 [astro-ph.GA] (Mar. 2018).
98. T. I. Hasegawa, E. Herbst, C. M. Leung, *ApJS* **82**, 167 (Sept. 1992).
99. E. Herbst, W. Klemperer, *ApJ* **185**, 505–534 (Oct. 1973).
100. E. Herbst, E. F. van Dishoeck, *ARA&A* **47**, 427–480 (Sept. 2009).
101. G. J. Herczeg *et al.*, *A&A* **540**, A84, arXiv: 1111.0774 [astro-ph.GA] (Apr. 2012).
102. F. Herpin *et al.*, *A&A* **542**, A76, arXiv: 1204.0397 [astro-ph.GA] (June 2012).
103. M. Heyer, T. M. Dame, *ARA&A* **53**, 583–629 (Aug. 2015).
104. L. A. Hillenbrand, L. W. Hartmann, *ApJ* **492**, 540–553 (Jan. 1998).
105. J. A. Hodge, E. da Cunha, *Royal Society Open Science* **7**, 200556, arXiv: 2004.00934 [astro-ph.GA] (Dec. 2020).
106. D. Hollenbach, C. F. McKee, *ApJS* **41**, 555–592 (Nov. 1979).
107. D. Hollenbach, E. E. Salpeter, *ApJ* **163**, 155 (Jan. 1971).
108. D. Hollenbach *et al.*, *ApJ* **690**, 1497–1521, arXiv: 0809.1642 [astro-ph] (Jan. 2009).
109. E. A. Hoversten, K. Glazebrook, *ApJ* **675**, 163–187, arXiv: 0711.1309 [astro-ph] (Mar. 2008).
110. A. Hughes *et al.*, *ApJ* **779**, 46, arXiv: 1309.3453 [astro-ph.GA] (Dec. 2013).
111. L. K. Hunt *et al.*, *A&A* **576**, A33, arXiv: 1409.5916 [astro-ph.GA] (Apr. 2015).
112. S. Ioppolo *et al.*, *ApJ* **686**, 1474–1479, arXiv: 0807.0129 [astro-ph] (Oct. 2008).
113. S. Ioppolo *et al.*, *Physical Chemistry Chemical Physics (Incorporating Faraday Transactions)* **12**, 12065, arXiv: 1009.5272 [astro-ph.SR] (Jan. 2010).
114. K. G. Jansky, *Nature* **132**, 66 (July 1933).
115. T. H. Jarrett *et al.*, *AJ* **125**, 525–554 (Feb. 2003).
116. S. Jarugula *et al.*, *ApJ* **880**, 92, arXiv: 1906.05469 [astro-ph.GA] (Aug. 2019).
117. S. Jarugula *et al.*, *ApJ* **921**, 97, arXiv: 2108.11319 [astro-ph.GA] (Nov. 2021).
118. J. H. Jeans, *Philosophical Transactions of the Royal Society of London Series A* **199**, 1–53 (Jan. 1902).
119. J. K. Jørgensen *et al.*, *A&A* **507**, 861–879, arXiv: 0909.3386 [astro-ph.SR] (Nov. 2009).
120. J. K. Jørgensen, A. Belloche, R. T. Garrod, *ARA&A* **58**, 727–778, arXiv: 2006.07071 [astro-ph.SR] (Aug. 2020).
121. M. J. Kaufman, D. A. Neufeld, *ApJ* **456**, 611 (Jan. 1996).
122. J. Kennicutt Robert C., *ApJ* **498**, 541–552, arXiv: astro-ph/9712213 [astro-ph] (May 1998).
123. R. C. Kennicutt, N. J. Evans, *ARA&A* **50**, 531–608, arXiv: 1204.3552 [astro-ph.GA] (Sept. 2012).
124. M. F. Kessler *et al.*, *A&A* **315**, L27–L31 (Nov. 1996).
125. L. J. Kewley, M. J. Geller, R. A. Jansen, *AJ* **127**, 2002–2030, arXiv: astro-ph/0401172 [astro-ph] (Apr. 2004).
126. A. King, K. Pounds, *ARA&A* **53**, 115–154, arXiv: 1503.05206 [astro-ph.GA] (Aug. 2015).
127. J. Koda *et al.*, *ApJS* **193**, 19, arXiv: 1102.3060 [astro-ph.IM] (Mar. 2011).
128. L. E. Kristensen, M. M. Dunham, *A&A* **618**, A158, arXiv: 1807.11262 [astro-ph.SR] (Oct. 2018).
129. L. E. Kristensen *et al.*, *A&A* **516**, A57, arXiv: 1004.2217 [astro-ph.SR] (June 2010).

130. L. E. Kristensen *et al.*, *A&A* **521**, L30, arXiv: 1007.3031 [astro-ph.SR] (Oct. 2010).
131. L. E. Kristensen *et al.*, *A&A* **605**, A93, arXiv: 1705.10269 [astro-ph.GA] (Sept. 2017).
132. L. E. Kristensen, E. A. Bergin, *ApJ* **807**, L25, arXiv: 1506.04160 [astro-ph.SR] (July 2015).
133. P. Kroupa, *MNRAS* **322**, 231–246, arXiv: astro-ph/0009005 [astro-ph] (Apr. 2001).
134. P. Kroupa, *Science* **295**, 82–91, arXiv: astro-ph/0201098 [astro-ph] (Jan. 2002).
135. J. M. D. Kruijssen, *MNRAS* **426**, 3008–3040, arXiv: 1208.2963 [astro-ph.CO] (Nov. 2012).
136. M. R. Krumholz, *arXiv e-prints*, arXiv:1511.03457, arXiv: 1511.03457 [astro-ph.GA] (Nov. 2015).
137. M. R. Krumholz, C. F. McKee, *MNRAS* **494**, 624–641, arXiv: 1909.01565 [astro-ph.GA] (May 2020).
138. M. Kuffmeier, T. Haugbølle, Å. Nordlund, *ApJ* **846**, 7, arXiv: 1611.10360 [astro-ph.SR] (Sept. 2017).
139. C. J. Lada, presented at the Star Forming Regions, ed. by M. Peimbert, J. Jugaku, vol. 115, p. 1.
140. C. J. Lada, T. M. Dame, *ApJ* **898**, 3, arXiv: 2006.08632 [astro-ph.GA] (July 2020).
141. C. J. Lada, E. A. Lada, *ARA&A* **41**, 57–115, arXiv: astro-ph/0301540 [astro-ph] (Jan. 2003).
142. C. J. Lada, M. Lombardi, J. F. Alves, *ApJ* **724**, 687–693, arXiv: 1009.2985 [astro-ph.GA] (Nov. 2010).
143. T. Lamberts *et al.*, *Physical Chemistry Chemical Physics (Incorporating Faraday Transactions)* **15**, 8287 (Jan. 2013).
144. C.-F. Lee *et al.*, *ApJ* **927**, L27, arXiv: 2203.00961 [astro-ph.GA] (Mar. 2022).
145. E. J. Lee, N. Murray, M. Rahman, *ApJ* **752**, 146, arXiv: 1204.4190 [astro-ph.GA] (June 2012).
146. A. Leger, M. Jura, A. Omont, *A&A* **144**, 147–160 (Mar. 1985).
147. A. Lehmann *et al.*, *A&A* **643**, A101, arXiv: 2010.01042 [astro-ph.GA] (Nov. 2020).
148. A. Lehmann *et al.*, *A&A* **658**, A165, arXiv: 2111.14089 [astro-ph.GA] (Feb. 2022).
149. C. Leitherer *et al.*, *ApJS* **123**, 3–40, arXiv: astro-ph/9902334 [astro-ph] (July 1999).
150. S. J. Lilly *et al.*, *ApJ* **772**, 119, arXiv: 1303.5059 [astro-ph.CO] (Aug. 2013).
151. D. C. Lis *et al.*, *ApJ* **738**, L6, arXiv: 1106.4784 [astro-ph.CO] (Sept. 2011).
152. L. Liu *et al.*, *ApJ* **846**, 5, arXiv: 1707.04914 [astro-ph.GA] (Sept. 2017).
153. S. N. Longmore *et al.*, presented at the Protostars and Planets VI, ed. by H. Beuther *et al.*, p. 291, arXiv: 1401.4175 [astro-ph.GA].
154. P. Madau, M. Dickinson, *ARA&A* **52**, 415–486, arXiv: 1403.0007 [astro-ph.CO] (Aug. 2014).
155. B. F. Madore *et al.*, *Nature* **395**, 47–50 (Sept. 1998).
156. J. S. Mathis, P. G. Mezger, N. Panagia, *A&A* **128**, 212–229 (Nov. 1983).
157. B. A. McGuire, *ApJS* **259**, 30, arXiv: 2109.13848 [astro-ph.GA] (Apr. 2022).
158. C. F. McKee, J. P. Williams, *ApJ* **476**, 144–165 (Feb. 1997).
159. R. Meijerink *et al.*, *A&A* **525**, A119, arXiv: 1010.0788 [astro-ph.CO] (Jan. 2011).
160. G. J. Melnick *et al.*, *ApJ* **539**, L77–L85 (Aug. 2000).
161. G. J. Melnick *et al.*, *ApJ* **892**, 22, arXiv: 2003.04897 [astro-ph.GA] (Mar. 2020).
162. K. M. Menten *et al.*, *A&A* **474**, 515–520, arXiv: 0709.0485 [astro-ph] (Nov. 2007).
163. A. A. Michelson, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **30**, 1–21, eprint: https://doi.org/10.1080/14786449008619983, (https://doi.org/10.1080/14786449008619983) (1890).
164. N. Miyauchi *et al.*, *Chemical Physics Letters* **456**, 27–30, arXiv: 0805.0055 [astro-ph] (Apr. 2008).
165. A. Mok, R. Chandar, S. M. Fall, *ApJ* **893**, 135, arXiv: 2004.02698 [astro-ph.GA] (Apr. 2020).
166. F. Motte, S. Bontemps, F. Louvet, *ARA&A* **56**, 41–82, arXiv: 1706.00118 [astro-ph.GA] (Sept. 2018).
167. J. C. Mottram *et al.*, *A&A* **572**, A21, arXiv: 1409.5704 [astro-ph.SR] (Dec. 2014).
168. J. C. Mottram *et al.*, *A&A* **600**, A99, arXiv: 1701.04647 [astro-ph.SR] (Apr. 2017).
169. P. C. Myers, E. F. Ladd, *ApJ* **413**, L47 (Aug. 1993).
170. D. A. Neufeld, S. Lepp, G. J. Melnick, *ApJS* **100**, 132 (Sept. 1995).
171. H. L. Nordh *et al.*, *A&A* **402**, L21–L25 (May 2003).
172. K. I. Öberg, E. A. Bergin, *Phys. Rep.* **893**, 1–48, arXiv: 2010.03529 [astro-ph.EP] (Jan. 2021).

173. S. S. R. Offner *et al.*, presented at the Protostars and Planets VI, ed. by H. Beuther *et al.*, p. 53, arXiv: 1312.5326 [astro-ph.SR].
174. A. Omont *et al.*, *A&A* **530**, L3, arXiv: 1107.4979 [astro-ph.CO] (June 2011).
175. A. Omont *et al.*, *A&A* **551**, A115, arXiv: 1301.6618 [astro-ph.CO] (Mar. 2013).
176. I. Oteo *et al.*, *ApJ* **850**, 170, arXiv: 1701.05901 [astro-ph.GA] (Dec. 2017).
177. L. Pagani *et al.*, *A&A* **467**, 179–186, arXiv: astro-ph/0701823 [astro-ph] (May 2007).
178. F. Palla, S. W. Stahler, *ApJ* **525**, 772–783 (Nov. 1999).
179. P. Palmeirim *et al.*, *A&A* **550**, A38, arXiv: 1211.6360 [astro-ph.SR] (Feb. 2013).
180. P. P. Papadopoulos, *ApJ* **720**, 226–232, arXiv: 1009.1134 [astro-ph.CO] (Sept. 2010).
181. J. Pflamm-Altenburg, C. Weidner, P. Kroupa, *MNRAS* **395**, 394–400, arXiv: 0901.4335 [astro-ph.GA] (May 2009).
182. G. L. Pilbratt *et al.*, *A&A* **518**, L1, arXiv: 1005.5331 [astro-ph.IM] (July 2010).
183. A. Pillepich *et al.*, *MNRAS* **473**, 4077–4106, arXiv: 1703.02970 [astro-ph.GA] (Jan. 2018).
184. R. L. Pitts *et al.*, *A&A* **657**, A70, arXiv: 2110.11132 [astro-ph.GA] (Jan. 2022).
185. R. L. Pitts, P. J. Barnes, *ApJS* **256**, 3, arXiv: 2106.02047 [astro-ph.GA] (Sept. 2021).
186. R. Pokhrel *et al.*, *ApJ* **912**, L19, arXiv: 2104.04551 [astro-ph.GA] (May 2021).
187. R. E. Pudritz, C. A. Norman, *ApJ* **301**, 571 (Feb. 1986).
188. L. Rayleigh, *Proceedings of the Royal Society of London Series I* **29**, 71–97 (Jan. 1879).
189. D. A. Riechers *et al.*, *ApJ* **645**, L13–L16, arXiv: astro-ph/0605437 [astro-ph] (July 2006).
190. D. A. Riechers *et al.*, *ApJ* **726**, 50, arXiv: 1011.0991 [astro-ph.CO] (Jan. 2011).
191. D. A. Riechers *et al.*, *Nature* **496**, 329–333, arXiv: 1304.4256 [astro-ph.CO] (Apr. 2013).
192. D. A. Riechers *et al.*, *Nature* **602**, 58–62, arXiv: 2202.00693 [astro-ph.GA] (Feb. 2022).
193. P. Rinaldi *et al.*, *ApJ* **930**, 128, arXiv: 2112.03935 [astro-ph.GA] (May 2022).
194. H. I. Ringermacher, L. R. Mead, *MNRAS* **397**, 164–171, arXiv: 0908.0892 [astro-ph.GA] (July 2009).
195. A. Rivera-Ingraham *et al.*, *ApJ* **766**, 85, arXiv: 1301.3805 [astro-ph.GA] (Apr. 2013).
196. R. Rolffs *et al.*, *A&A* **536**, A33, arXiv: 1109.3995 [astro-ph.GA] (Dec. 2011).
197. J. Roman-Duval *et al.*, *ApJ* **723**, 492–507, arXiv: 1010.2798 [astro-ph.GA] (Nov. 2010).
198. E. Rosolowsky, *PASP* **117**, 1403–1410, arXiv: astro-ph/0508679 [astro-ph] (Dec. 2005).
199. D. Russeil *et al.*, *A&A* **515**, A55 (June 2010).
200. M. Rybak *et al.*, *MNRAS* **494**, 5542–5567, arXiv: 1912.12538 [astro-ph.GA] (June 2020).
201. K. L. J. Rygl *et al.*, *A&A* **539**, A79, arXiv: 1111.7023 [astro-ph.GA] (Mar. 2012).
202. S. I. Sadavoy *et al.*, *ApJ* **787**, L18, arXiv: 1404.7142 [astro-ph.SR] (June 2014).
203. S. Salim *et al.*, *ApJS* **173**, 267–292, arXiv: 0704.3611 [astro-ph] (Dec. 2007).
204. B. Salmon *et al.*, *ApJ* **799**, 183, arXiv: 1407.6012 [astro-ph.GA] (Feb. 2015).
205. E. E. Salpeter, *ApJ* **121**, 161 (Jan. 1955).
206. I. San José-García *et al.*, *A&A* **585**, A103, arXiv: 1511.00357 [astro-ph.SR] (Jan. 2016).
207. I. San José-García, PhD thesis, University of Leiden, Netherlands, June 2015.
208. A. Sandqvist *et al.*, *A&A* **647**, A86, arXiv: 2101.03876 [astro-ph.GA] (Mar. 2021).
209. J. M. Scalo, *Fund. Cosmic Phys.* **11**, 1–278 (May 1986).
210. R. A. Scheepmaker *et al.*, *A&A* **494**, 81–93, arXiv: 0812.1417 [astro-ph] (Jan. 2009).
211. E. Schinnerer *et al.*, *ApJ* **779**, 42, arXiv: 1304.1801 [astro-ph.CO] (Dec. 2013).
212. F. L. Schöier *et al.*, *A&A* **432**, 369–379, arXiv: astro-ph/0411110 [astro-ph] (Mar. 2005).
213. N. Z. Scoville, J. C. Good, *ApJ* **339**, 149 (Apr. 1989).
214. A. E. Shapley, *ARA&A* **49**, 525–580, arXiv: 1107.5060 [astro-ph.CO] (Sept. 2011).
215. C. J. Shen *et al.*, *A&A* **415**, 203–215 (Feb. 2004).
216. F. H. Shu, *ApJ* **214**, 488–497 (June 1977).

217. F. Shu *et al.*, *ApJ* **429**, 781 (July 1994).
218. I. M. Skretas, L. E. Kristensen, *A&A* **660**, A39, arXiv: 2202.04676 [astro-ph.GA] (Apr. 2022).
219. I. W. M. Smith, E. Herbst, Q. Chang, *MNRAS* **350**, 323–330 (May 2004).
220. R. J. Smith, *ARA&A* **58**, 577–615 (Aug. 2020).
221. R. L. Snell, R. B. Loren, *ApJ* **211**, 122–127 (Jan. 1977).
222. A. Sneppen *et al.*, *ApJ* **931**, 57, arXiv: 2205.11536 [astro-ph.GA] (May 2022).
223. V. V. Sobolev, *Moving Envelopes of Stars*.
224. P. M. Solomon *et al.*, *ApJ* **319**, 730 (Aug. 1987).
225. P. Solomon *et al.*, *Nature* **426**, 636–638, arXiv: astro-ph/0312436 [astro-ph] (Dec. 2003).
226. S. W. Stahler, F. Palla, *The Formation of Stars*.
227. F. Stanley *et al.*, *A&A* **646**, A178, arXiv: 2011.09991 [astro-ph.GA] (Feb. 2021).
228. C. L. Steinhardt *et al.*, *ApJ* **931**, 58, arXiv: 2205.14161 [astro-ph.GA] (May 2022).
229. M. L. Strandet *et al.*, *ApJ* **842**, L15, arXiv: 1705.07912 [astro-ph.GA] (June 2017).
230. A. N. Suutarinen *et al.*, *MNRAS* **440**, 1844–1855, arXiv: 1402.7214 [astro-ph.GA] (May 2014).
231. P. Swings, L. Rosenfeld, *ApJ* **86**, 483–486 (Nov. 1937).
232. B. Tabone *et al.*, *A&A* **607**, L6, arXiv: 1710.01401 [astro-ph.SR] (Nov. 2017).
233. L. J. Tacconi, R. Genzel, A. Sternberg, *ARA&A* **58**, 157–203, arXiv: 2003.06245 [astro-ph.GA] (Aug. 2020).
234. M. Tafalla *et al.*, *A&A* **551**, A116, arXiv: 1301.5322 [astro-ph.GA] (Mar. 2013).
235. M. Tafalla *et al.*, *A&A* **597**, A119, arXiv: 1610.01614 [astro-ph.GA] (Jan. 2017).
236. The Astropy Collaboration *et al.*, *AJ* **156**, 123, arXiv: 1801.02634 [astro-ph.IM] (Sept. 2018).
237. A. R. Thompson, J. M. Moran, J. Swenson George W., *Interferometry and Synthesis in Radio Astronomy, 3rd Edition*.
238. A. G. G. M. Tielens *et al.*, *ApJ* **431**, 321 (Aug. 1994).
239. L. Tresse *et al.*, *MNRAS* **337**, 369–383, arXiv: astro-ph/0111390 [astro-ph] (Nov. 2002).
240. J. S. Urquhart *et al.*, *MNRAS* **443**, 1555–1586, arXiv: 1406.5078 [astro-ph.GA] (Sept. 2014).
241. J. A. Villa-Vélez *et al.*, *A&A* **654**, A153, arXiv: 2108.13321 [astro-ph.GA] (Oct. 2021).
242. R. Visser *et al.*, *ApJ* **769**, 19, arXiv: 1303.6177 [astro-ph.SR] (May 2013).
243. V. Wakelam *et al.*, *Molecular Astrophysics* **9**, 1–36, arXiv: 1711.10568 [astro-ph.GA] (Dec. 2017).
244. J. L. Ward, J. M. D. Kruijssen, H.-W. Rix, *MNRAS* **495**, 663–685, arXiv: 1910.06974 [astro-ph.SR] (June 2020).
245. A. B. Watts *et al.*, *MNRAS* **477**, 5554–5567, arXiv: 1804.07072 [astro-ph.GA] (July 2018).
246. R. Weinberger *et al.*, *MNRAS* **465**, 3291–3308, arXiv: 1607.03486 [astro-ph.GA] (Mar. 2017).
247. D. Wilgenbus *et al.*, *A&A* **356**, 1010–1022 (Apr. 2000).
248. J. P. Williams, L. A. Cieza, *ARA&A* **49**, 67–117, arXiv: 1103.0556 [astro-ph.GA] (Sept. 2011).
249. C. Yang *et al.*, *A&A* **595**, A80, arXiv: 1607.06220 [astro-ph.GA] (Nov. 2016).
250. C. Yang *et al.*, *A&A* **608**, A144, arXiv: 1709.04740 [astro-ph.GA] (Dec. 2017).
251. C. Yang *et al.*, *ApJ* **771**, L24, arXiv: 1305.6351 [astro-ph.GA] (July 2013).
252. F. Zernike, *Physica* **5**, 785–795 (Aug. 1938).
253. C. Zhang *et al.*, *MNRAS* **497**, 793–808, arXiv: 2006.13410 [astro-ph.GA] (Sept. 2020).
254. Z.-Y. Zhang *et al.*, *Nature* **558**, 260–263, arXiv: 1806.01280 [astro-ph.GA] (June 2018).
255. R. de Grijs, R. W. O'Connell, I. Gallagher John S., *AJ* **121**, 768–792, arXiv: astro-ph/0010046 [astro-ph] (Feb. 2001).
256. A. de Valon *et al.*, *A&A* **634**, L12, arXiv: 2001.09776 [astro-ph.SR] (Feb. 2020).
257. P. H. van Cittert, *Physica* **1**, 201–210 (Jan. 1934).
258. E. F. van Dishoeck, J. H. Black, *ApJS* **62**, 109 (Sept. 1986).
259. E. F. van Dishoeck *et al.*, *PASP* **123**, 138, arXiv: 1012.4570 [astro-ph.GA] (Feb. 2011).
260. E. F. van Dishoeck *et al.*, *A&A* **648**, A24, arXiv: 2102.02225 [astro-ph.GA] (Apr. 2021).

261. E. F. van Dishoeck *et al.*, *A&A* **648**, A24, arXiv: 2102.02225 [astro-ph.GA] (Apr. 2021).

262. E. F. van Dishoeck, *ARA&A* **42**, 119–167, arXiv: astro-ph/0403061 [astro-ph] (Sept. 2004).

263. E. F. van Dishoeck, E. Herbst, D. A. Neufeld, *Chemical Reviews* **113**, 9043–9085, arXiv: 1312.4684 [astro-ph.GA] (Dec. 2013).

264. P. G. van Dokkum, *ApJ* **674**, 29–50, arXiv: 0710.0875 [astro-ph] (Feb. 2008).

265. P. G. van Dokkum, C. Conroy, *Nature* **468**, 940–942, arXiv: 1009.5992 [astro-ph.CO] (Dec. 2010).

266. F. F. S. van der Tak *et al.*, *A&A* **468**, 627–635, arXiv: 0704.0155 [astro-ph] (June 2007).

267. F. F. S. van der Tak *et al.*, *A&A* **554**, A83, arXiv: 1304.2949 [astro-ph.GA] (June 2013).

268. F. F. S. van der Tak *et al.*, *A&A* **625**, A103, arXiv: 1903.11305 [astro-ph.GA] (May 2019).

269. A. van der Wel *et al.*, *ApJ* **788**, 28, arXiv: 1404.2844 [astro-ph.GA] (June 2014).

270. P. P. van der Werf *et al.*, *A&A* **518**, L42, arXiv: 1005.2877 [astro-ph.GA] (July 2010).

271. P. P. van der Werf *et al.*, *ApJ* **741**, L38, arXiv: 1106.4825 [astro-ph.CO] (Nov. 2011).

272. P. P. van der Werf *et al.*, *ApJ* **741**, L38, arXiv: 1106.4825 [astro-ph.CO] (Nov. 2011).