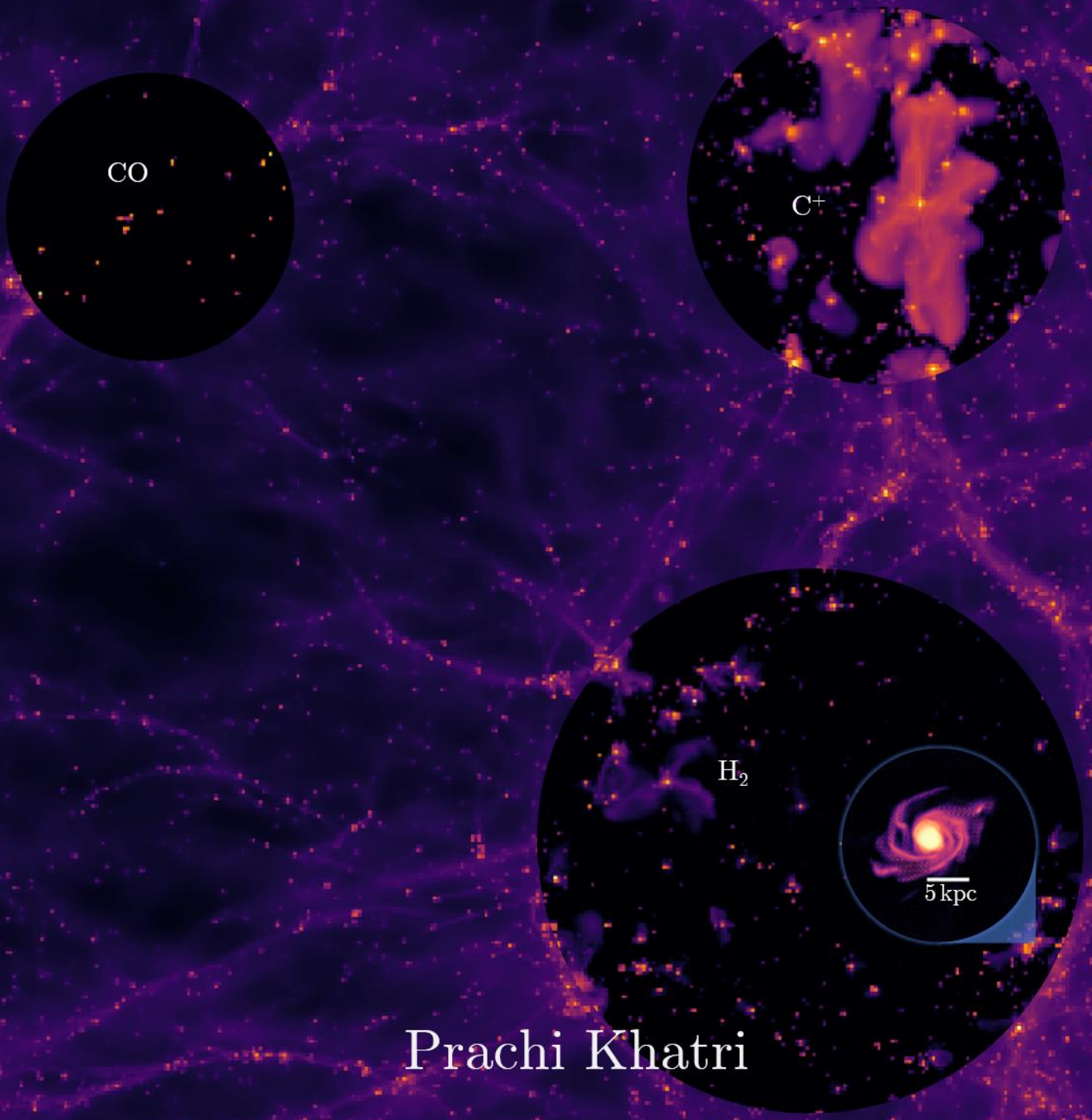
Modelling Molecular Gas and Its Tracers Across Cosmic Time



Prachi Khatri

Modelling Molecular Gas and Its Tracers Across Cosmic Time

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von
Prachi Khatri
aus
Bikaner, India

Bonn, January 2025

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

Gutachter/Betreuer:	Prof. Dr. Cristiano Porciani
Gutachterin:	Prof. Dr. Serena Viti

Tag der Promotion:	04.04.2025
Erscheinungsjahr:	2025

To my parents,
who have given me everything,
and then some more.

Per aspera ad astra

Abstract

Our understanding of the distant Universe and the processes governing galaxy formation and evolution largely stems from observing the light from stars and its interaction with the material surrounding them. However, an essential piece of this picture lies in the role of the interstellar medium (ISM) in shaping star formation within galaxies. Unveiling this aspect requires tracing the fuel for star formation – molecular gas. This thesis explores molecular gas in galaxies across cosmic time using cosmological simulations of galaxy formation.

Simulating the molecular gas content of galaxies is challenging as it requires modelling various physical and chemical processes happening on a wide range of (spatial and temporal) scales. On large scales, galaxy growth is affected by gas accretion from and outflows into the cosmic web. On the other hand, molecular gas chemistry is regulated by conditions on sub-parsec scales, which are beyond the resolving capabilities of large-scale simulations needed to investigate the evolution of the cosmic molecular gas budget. To tackle this multi-scale problem, we have developed a sub-grid model called Hyacinth – *HYdrogen And Carbon chemistry in the INTerstellar medium in Hydro simulations* – that can be embedded into large-scale cosmological simulations to track the non-equilibrium abundances of molecular hydrogen (H_2), and its carbon-based observational proxies, namely, carbon monoxide (CO), atomic carbon (C), and singly-ionised carbon (C^+), on the fly.

We have implemented Hyacinth into the widely-used Ramses code to perform the Marigold simulations, that track the dynamical evolution of H_2 and its tracers within galaxies in a cosmological context. Our simulated cosmic H_2 density is in excellent agreement with current observational constraints. Additionally, we find that low-mass ($M_{H_2} < 10^8\,M_\odot$) galaxies contain nearly half of the cosmic H_2 in the early Universe (i.e., when it was less than one billion year old). However, the sensitivity of current instruments renders these galaxies "invisible", indicating a potential underestimation of the cosmic H_2 density in existing surveys.

In recent years, the fine-structure line of C^+ at 158 microns the [C II] line – has gained significant attention as a (molecular) gas tracer during the epoch when the Universe was less than two billion years old. Being one of the brightest emission lines in galaxies, it offers a unique window into the molecular ISM of distant galaxies, where conventional tracers like CO become observationally expensive. We tested the reliability of this line as a molecular gas tracer using a statistical sample of galaxies at different cosmic epochs from the Marigold simulations. Our analysis reveals that the [C II]-molecular gas correlation

is relatively weak in the first billion years of the Universe but grows in both strength and tightness over time. Moreover, the relation exhibits a clear secondary dependence on the star-formation rate. We further examined the time evolution of the [C II] luminosity function and the cosmic [C II] luminosity density ($\rho_{[\mathrm{C\,II}]}$), and found that faint ($L_{[\mathrm{C\,II}]} < 10^7\,\mathrm{L}_\odot$) galaxies contribute nearly half of the cosmic $\rho_{[\mathrm{C\,II}]}$ in the early Universe. Since these faint galaxies fall below the sensitivity limits of current instruments, detecting them would require alternative observational strategies.

Overall, this thesis highlights the pivotal role of cosmological simulations in interpreting observations and providing crucial insights into the molecular gas reservoir of galaxies, that serves as the fuel for star formation across cosmic time.

List of Publications

The following list indicates the relevant first-author publications that are discussed and presented in this thesis. This thesis culminates in one peer-reviewed and published research publication and one submitted for publication to the Astrophysics & Astronomy journal .

1. **Prachi Khatri**, Cristiano Porciani, Emilio Romano-Díaz, Daniel Seifried, and Alexander Schäbe (2024),
   HYACINTH: HYdrogen And Carbon chemistry in theINTerstellar medium in Hydro simulations, Astronomy & Astrophysics
   DOI: 10.1051/0004-6361/202449640

   Author's Contribution: P.K., C.P., and E.R.D. designed the sub-grid model; A.S. provided useful inputs in the initial development of the chemical network; D.S. provided the data from the SILCC-Zoom simulations that guided the development and testing of HYACINTH; P.K. wrote the code for HYACINTH and interfaced it with RAMSES, performed all tests and comparisons, applied it to a simulated galaxy in post-processing and analysed the results; P.K. wrote the manuscript with helpful feedback from co-authors, particularly C.P.

2. **Prachi Khatri**, Emilio Romano-Díaz, and Cristiano Porciani (2024),
   The [C II] *line emission as an interstellar medium probe in the* MARIGOLD *galaxies*, **Submitted** for publication to Astronomy & Astrophysics
   DOI: 10.1051/0004-6361/202453048

   Author's Contribution: P.K., E.R.D. and C.P. designed the simulation specifications; E.R.D. generated the initial conditions; P.K. performed the simulations, post-processed the snapshots to obtain halo and galaxy catalogs with useful suggestions from E.R.D. and C.P.; P.K. performed the radiative transfer of the simulated galaxies to obtain their [C II] emission, with inputs regarding radiative transfer from C.P.; C.P. significantly contributed to the analysis of the [C II] luminosity function; P.K. analysed all results and wrote the manuscript, with suggestions from C.P. and E.R.D.

Acknowledgements

The journey of a thousand miles begins with a single step.

- Lao Tzu

If I have come this far in my academic journey, it has only been possible with the support of many people who, in their own unique ways, nudged me to take the next step, often when I had no clue where the road led. I hope I manage to thank them all here.

First and foremost, I want to thank my supervisor, Cristiano Porciani, for introducing me to not one, but two amazing research areas, for guiding me throughout my masters and PhD, for your deep insights, and for teaching me how to do research. Thank you for always showing enthusiasm for my work, for teaching me the value of being rigorous, nurturing my writing skills over the years, and gradually helping me to mature on the path to becoming a scientist. Above all, I thank you for always believing in my potential, since the days of theoretical astrophysics till this day, for pushing me to be my best version of scientist-in-prep, even during the phases I had forgotten how to be.

I sincerely thank Emilio Romano-Díaz for his guidance throughout my masters and PhD years, for helping me navigate the challenges of the computational world, and especially for your encouragement and care during the challenging pandemic and post-pandemic years.

I thank Peter Schneider for the superb excitement and energy you brought to the lectures on Cosmology, that always inspired me. I would like to thank Serena Viti for gladly agreeing to be on my PhD committee and being patient with the changes in my timeline. I was lucky to have had the chance to attend a couple of your talks here in Bonn and to chat with you about astrochemistry. In the course of that brief chat, you taught me to reflect on what's really important for my science case, thank you for that. I also extend my thanks to Lucie Flek and Sebastian Neubert for showing enthusiasm for my thesis topic, promptly accepting to be on my PhD committee, and being patient with me. I extend my thanks to Sabine and the AIfA IT team, especially Oliver and Andreas, for helping me throughout my time at AIfA. A special thanks to my undergrad professors, V. Ravishankar and Varsha Banerjee, for pushing me to pursue higher studies.

I am grateful to my friend Kainaat for always being there for me since my very first day in Bonn, helping me navigate the difficulties of living abroad, and inspiring me with your zeal towards your own research. Thanks to Abraham for his friendship and support. You

both have always made me feel at home in Bonn. I am thankful to Aishwarya, my study buddy, my academic companion, my go-to person for all things astronomy. Countless times in my PhD, I have wished that we shared an office like in the good ol' days of masters. Our long phone conversations about everything under the comoving horizon is something I will always cherish.

I would like to thank my PhD friends at AIfA, whose companionship has made this journey so much more joyful. To Yvonne, for guiding me through RAMSES, simulations, slurming, bureaucracy, and more. I will always remember that evening when I was submitting to arXiv for the very first time, and you were there for me when was completely baffled by all the steps. To Anna, for always being there to listen to all I had to say, offering me a cookie, reminding me to take a short break, and to eat. Your excitement and encouragement for HYACINTH and MARIGOLD means a lot to me. To Mandar, for always asking me how I am really doing and lending an ear to all my woes. I am grateful for your friendship and your calm manner even in chaotic times. I have learnt a lot from you during my time in Bonn. To Elena, for being an amazing office buddy and friend, and for cheering me up on difficult days. To Benedetta, for being like a big sister to me, your positive outlook in the midst of challenges has always inspired me.

My deepest gratitude goes to my parents. I cannot express in words how grateful I am for your constant love, care, support and faith in me, for offering me perspective in difficult situations, motivating me, and telling me that I am brave and strong. The truth is, it is *you* who make me strong. Not a day passes without me realizing how awesome you both are. Thank you so much for putting my dreams first, for being my biggest cheerleaders, for supporting me through every step of the journey, for giving me the courage to move away from home in pursuit of my dreams and yet always making sure, no matter what time of the day, home was just a phone call away. Thank you, Bunti, for being the most amazing brother one could ask for, for being excited for my dreams, always rooting for me, for the innumerable times I called you and you fixed me. You are always such a bundle of joy, bringing me laughter, sometimes even at my own expense. Thank you so much for introducing me to poetry and reminding me how beautiful the little things in life are.

If there is one person who has believed in my scientific potential through the ups and downs of my academic journey, even before I had had the chance to do any research, it is Adyant. You picked me up when I had basically given up on my dreams and nudged me to look in a different direction. Your unceasing supply of enthusiasm, emotional strength, and laughter kept me going on. Thank you for reminding me, time and again, to believe in myself and my journey. I can't imagine where I and this thesis would be without your invaluable scientific insights and feedback. I am indeed very lucky to have someone who is more joyous than me in my victories. Special thanks for introducing me to *Modern Family* and for telling me that '*This is all super chill*', for being an awesome cheerleader, and the innumerable times you reminded me that I got this. But most importantly, for keeping me from giving up, time after time. I am glad to have shared this journey with you.

To my grandparents and my uncle Vikas, who have played a bigger part in my journey than they can imagine. Thank you for all the love, positivity, and blessings you have always sent my way.

Contents

CHAPTER 1

Introduction

Molecular gas plays a major role in galaxies near and far, providing the necessary conditions and serving as the primary fuel for star formation. Investigating the build-up of the molecular gas reservoir in galaxies and its cosmic evolution is therefore pivotal to our understanding of the star formation history and galaxy assembly in the Universe. The past two decades have witnessed tremendous growth in our knowledge of molecular gas in an ever-increasing number of galaxies observed at different cosmic epochs. This progress has been fostered by technical advancements in the observational facilities as well as the development of new methods to probe the gas content of galaxies.

In its infancy, this field was restricted to a handful of bright, massive, but rare galaxies. Nowadays, we routinely detect typical galaxies at cosmic times as early as one billion years after the birth of the Universe. The physical conditions within and around these early galaxies (such as radiation fields, chemical composition of the gas, the availability of gas for star formation, galaxy-galaxy interactions, among others) are expected to be different from those in present-day galaxies. As a result, the scaling relations between galaxy observables and intrinsic properties can evolve over time. In this regard, numerical simulations are a useful tool for investigating this evolution. By offering a complementary and broader perspective of the galaxy population at large, they play a pivotal role in guiding the interpretation of observational data, testing the validity of established scaling relations in new regimes, and uncovering potential biases introduced by observational limitations of sensitivity and resolution.

One of the main scientific objectives of this thesis is to model the dynamical evolution of molecular hydrogen (H_2) its carbon-based tracers, namely carbon monoxide (CO), atomic carbon (C), and singly-ionized carbon (C^+), within cosmological simulations of galaxy formation, and study the time evolution of the cosmic molecular gas reservoir. At the same time, these simulations are aimed at providing statistical sample of galaxies

at different epochs that are ideally suited for exploring how the aforementioned scaling relations change with time. In this context, we examine the reliability of an emerging molecular gas tracer – the fine-structure line of C^+ – in the early Universe. The work carried out in this thesis can be divided into three projects:

Project A: *A sub-grid model for hydrogen and carbon chemistry in cosmological simulations*
This project involves developing a sub-grid model that can be embedded into cosmological simulations for calculating the non-equilibrium abundances of H_2, CO, C, and C^+.

Project B: *Molecular gas in galaxies across cosmic time*
The objective of this project is to perform cosmological simulations with the sub-grid model in Project A and investigate the statistical properties of the galaxy population at different cosmic epochs. In particular, we want to answer open questions such as:

- How does the global molecular gas density evolve with time?
- How does the molecular gas fraction vary across galaxies of different masses? How does it evolve over time?

Project C: *The* [C II] *line emission in early galaxies*
This project involves modelling the emission of the fine-structure line of C^+ – the [C II] line – from simulated galaxies. With this study, we want to answer the following science questions:

- How do the physical conditions and galaxy properties affect the brightness of the line emission?
- How does the brightness and abundance of [C II] emitting galaxies evolve over time?
- What is the spatial extent of this emission in galaxies at different cosmic epochs?
- Can this line be used as a probe of the gas content and star formation in galaxies?
- How do the correlations between the line brightness and physical properties change over time? Do they depend on other factors?

These projects are detailed out in Chapters 3, 4, and 5, respectively. Chapter 2 presents an overview of the theoretical concepts and literature relevant for the work carried out in this thesis. Finally, Chapter 6 provides a summary of what we have learned from the three projects and what questions remain to be answered.

In the very beginning,
only H and He were around.
Then stars came along
and spewed heavy elements (aka metals) out.
Slowly and steadily
as cosmic time ticked away,
the galaxies grew larger,
and metals were everywhere.
The metals took the heat of the ISM away
the ISM grew cooler by night and denser by day
Inside these dense regions,
the metal atoms combined
as thus was born dust,
to everyone's surprise.
Though dust formed out of metals,
it remembered its ancestors
And as a tribute promised
to give H_2 some shelter.
So deep within molecular clouds,
H_2 was well-shielded,
the stars kept forming,
and the cycle is ongoing.
Among the various metals,
C, N & O were special
They formed the building blocks
that made life possible.

CHAPTER 2

Theoretical Background

In this chapter, we provide an overview of the theoretical concepts and literature that form the basis of this thesis. In Section 2.1, we briefly review the basic concepts of cosmology and structure formation. Section 2.2 deals with the physics of the interstellar medium. Section 2.4 delves into the details of molecular hydrogen, which is at the core of this thesis. In Section 2.5, we describe the commonly used star formation rate indicators in galaxies. Section 2.6 delves into the star formation history and the evolution of the cosmic molecular gas content. Section 2.3 provides a primer on radiative transfer. Finally, in Section 2.7, we describe simulations of galaxy formation, with a particular focus on modelling hydrogen and carbon chemistry.

2.1. Cosmology and Structure Formation

Matter tells spacetime how to curve, and spacetime tells matter how to move.

-John Wheeler

Modern cosmology is based on General Relativity (GR) and the Cosmological Principle, which states that the Universe is (spatially) isotropic and homogeneous on large scales (a few hundred megaparsecs). The Cosmological Principle requires that the Universe must appear or *look* the same to all observers. GR describes how (matter-)energy and spacetime affect each other via Einstein's field equations[1] (Einstein, 1915):

$$\underbrace{R_{\mu\nu} - \frac{R}{2} g_{\mu\nu}}_{geometry} = \frac{8\pi G}{c^4} \underbrace{T_{\mu\nu}}_{energy} , \tag{2.1}$$

[1] **Note:** Unless otherwise stated, this section draws on the basic concepts of modern cosmology and galaxy formation from Dodelson (2003) and Mo, van den Bosch & White (2010).

where the Ricci tensor $R_{\mu\nu}$ and its trace R, the Ricci scalar, describe the geometry of spacetime, the energy-momentum tensor $T_{\mu\nu}$ encapsulates the various forms of energy such as matter and radiation, G is Newton's gravitational constant and c is the speed of light in vacuum. Einstein's equations relate the geometry of spacetime to the energy content of the Universe.

Observations support that the Universe originated from a much denser and hotter state, commonly referred to as the *Big Bang*, nearly 13.5 billion years ago, and has been expanding since then (e.g., Mo, van den Bosch & White, 2010). In an expanding universe described by general relativity, we can define a set of fundamental observers that move with the expansion of the Universe and do not have any peculiar motion relative to the global expansion. Such observers are called *comoving observers*. All comoving observers agree on a cosmic time t, which quantifies the proper time elapsed since the Big Bang and provides a natural choice of the time coordinate in cosmological models.

The expansion of the Universe is parametrized by a dimensionless scale factor $a(t)$, whose present value is set to 1, i.e., $a = 1$ at the present cosmic time, t_0. If we were to assign spatial coordinates to each position in the Universe at a time t, we would obtain two sets of coordinates – *physical* $\mathbf{r}(t)$ and *comoving* $\mathbf{x}$, where the latter are defined with respect to comoving observers and therefore independent on time. Therefore, by definition, the expansion of the Universe is factored out in comoving coordinates. The two are related as:

$$\mathbf{r}(t) \;=\; a(t)\,\mathbf{x}\,. \tag{2.2}$$

At $t = t_0$, $a = 1$ and we find that the two are identical.

Likewise, the volume elements $\mathrm{d}V_{\mathrm{physical}}$ (in physical coordinates) and $\mathrm{d}V_{\mathrm{comoving}}$ (in comoving coordinates) are related as:

$$\mathrm{d}V_{\mathrm{physical}}(t) \;=\; a^3(t)\,\mathrm{d}V_{\mathrm{comoving}}\,. \tag{2.3}$$

By construction, $\mathrm{d}V_{\mathrm{comoving}}$ is independent of cosmic time.

The time evolution of a is determined by the energy content of the Universe. To quantify how rapidly the Universe is expanding, we define the Hubble rate as

$$H(t) \equiv \frac{1}{a}\,\frac{\mathrm{d}a}{\mathrm{d}t}\,. \tag{2.4}$$

The present-day value of the Hubble rate is called the Hubble constant $H_0 = H(t_0)$.

Einstein's equations for a homogenous and isotropic, expanding Universe, yield the following set of equations, known as the Friedmann equations:

$$H^2(t) \equiv \frac{1}{a^2}\left(\frac{\mathrm{d}a}{\mathrm{d}t}\right)^2 = \frac{8\pi G}{3}\left[\rho(t) - \frac{\kappa c^2}{a^2}\right]\,; \tag{2.5}$$

$$\frac{\mathrm{d}^2 a}{\mathrm{d}t^2} \;=\; -\frac{4\pi G}{3}\left(\rho(t) + \frac{3p(t)}{c^2}\right)\,, \tag{2.6}$$

where it is assumed that the energy content of the Universe can be well described by a perfect fluid,[2] with energy density $\rho(t)$ and pressure $p(t)$ at time t. The first Friedmann

[2] A perfect fluid is isotropic in its rest frame and can be completely characterised by its pressure and density (Carroll, 2004).

equation describes the expansion of the Universe (as parametrized by the scale factor), while the second determines whether this expansion is accelerating or decelerating over time. In Equation (2.5), κ denotes the spatial curvature of the Universe and is defined as

$$\kappa = \frac{1}{c^2} \left(\rho(t_0) - \rho_{\text{crit}}\right) , \tag{2.7}$$

where $\rho(t_0)$ is the present energy density of the Universe and

$$\rho_{\text{crit}} \equiv 3H_0^2/8\pi G \tag{2.8}$$

is the critical energy density required for a spatially flat Universe with Euclidean geometry. If $\rho(t_0) > \rho_{\text{crit}}$, the Universe has a positive spatial curvature (e.g., the surface of a sphere); if $\rho(t_0) < \rho_{\text{crit}}$, the spatial curvature is negative (e.g., a hyperboloid in two dimensions).

In an expanding Universe, the frequency of light emitted by a distant object is altered as it travels to a present-day observer. This happens because two consecutive flashes of light travel slightly different distances between the source and the observer, the observed frequency (ν_{obs}) is smaller than the frequency at which the light was emitted (ν_{em}), and this decrease is defined as the cosmological redshift z of the distant object:

$$1 + z \equiv \frac{\nu_{\text{em}}}{\nu_{\text{obs}}} . \tag{2.9}$$

The scale factor corresponding to the emitting object is related to its redshift as:

$$a = \frac{1}{1+z} . \tag{2.10}$$

Therefore, the cosmological redshift provides a convenient measure of the object's distance from the observer. The exact expression for the distance-redshift relation, however, depends on the energy content of the Universe.

2.1.1. Contents of the Universe

The energy constituents of the Universe and their time evolution are described by a theoretical framework called a cosmological model. Among the various cosmological models, the most widely accepted is the Standard model of cosmology, called the Lambda cold dark matter (ΛCDM) model, that postulates the existence of a yet unknown form of energy called *dark energy* and non-baryonic matter called *dark matter*, that together account for $\sim 95\%$ of the present-day energy density of the Universe. The remaining 5% of the energy budget is contributed by ordinary baryonic matter (also includes electrons that are leptons as per the standard model of particle physics), i.e., atoms that make up stars, planets and everything we see around us. Apart from these, radiation (photons) is also an energy component but its present-day contribution is negligible ($\lesssim 0.01\%$) because the energy density of radiation evolves as $\propto a^{-4}$.

Dark energy is responsible for the present-day accelerated expansion of the Universe (Perlmutter et al., 1999; Riess et al., 1998). In the ΛCDM model, dark energy is in the form of a *cosmological constant* (denoted by Λ), meaning that its energy density remains constant over time. Dark matter interacts solely via gravity and does not emit light, therefore *dark*. In addition, it is *cold* meaning that the velocity dispersion of the dark

matter particles is much smaller than the speed of light since the early stages of the Universe, allowing it to cluster. The energy density of matter (both baryonic and dark) evolves as $\propto a^{-3}$.

The present-day energy densities of the different energy components are often expressed in terms of the critical density $\rho_{\rm crit}$ (Equation 2.8) to obtain the dimensionless parameters:

$$\Omega_i = \frac{\rho_i(t_0)}{\rho_{\rm crit}} \,. \tag{2.11}$$

The density and pressure of the different energy components can be related by the respective equation of state as

$$p(t) = w\,\rho(t)\,c^2 \,, \tag{2.12}$$

where the equation of state parameter w is different for each component: $w = 1/3$ for radiation, $w = 0$ for matter, and $w = -1$ for the cosmological constant.

2.1.2. The Cosmic Microwave Background

In the early Universe, the temperature was sufficiently high to keep electrons and protons from combining to form neutral atoms. Photons were constantly scattering off these free electrons, meaning radiation and baryonic matter were tightly coupled. When the temperature dropped to approximately 3000 K at $z \sim 1100$, electrons and protons combined to form neutral atoms, during a period known as the epoch of recombination. The photons that last scattered off matter during this epoch formed what we observe today as the cosmic microwave background (CMB). The spectrum of the CMB closely resembles a blackbody with a temperature of 2.728 ± 0.004 K (Fixsen et al., 1996). Post-recombination, the baryonic matter in the Universe existed in the form a neutral gas, predominantly hydrogen and helium.

2.1.3. Structure Formation

Even though the flat ΛCDM model assumes spatial isotropy and homogeneity on large scales, we observe structures such as galaxies on smaller scales, which formed due to the growth of primordial density perturbations. The earliest and most compelling evidence of these primordial perturbations comes from observing tiny fluctuations in the CMB. These fluctuations are imprinted in the CMB as temperature anisotropies, which reflect the density fluctuations in the early Universe. The amplitude of a fluctuation can be expressed in the form of a density contrast as:

$$\delta(\vec{x}, t) \equiv \frac{\rho(\vec{x}, t) - \overline{\rho}(t)}{\overline{\rho}(t)} \,, \tag{2.13}$$

where δ denotes the density in excess of $\overline{\rho}(t)$, the average matter density in the Universe at time t. Overdense regions have $\delta > 0$ and underdense regions have $-1 \leq \delta < 0$. Overdensities grow over time by accreting more matter due to their strong gravitational influence on the surrounding matter. Once an overdensity grows sufficiently large ($\delta \sim 1$), its self-gravity counters the expansion of the Universe, causing it to detach itself from the local expansion and gravitationally collapse onto itself. Upon collapse, the collisionless dark matter relaxes violently, forming a dark matter halo, while the gas is shock-heated and,

over time, cools radiatively, leading to the formation of a gravitationally bound object in virial equilibrium.[3] Dark matter halos are the sites of galaxy formation. After virialisation, the average density within the dark matter halo is $\rho_{\rm vir} = \Delta_{\rm vir}\,\overline{\rho}(t)$, where $\Delta_{\rm vir} \approx 200$ in the ΛCDM framework. The virial mass $M_{\rm vir}$ and radius $R_{\rm vir}$ of a halo are related as $M_{\rm vir} = \dfrac{4\pi}{3} R_{\rm vir}^3\,\Delta_{\rm vir}\,\overline{\rho}(t)$.

2.1.4. Galaxy growth

Galaxies grow via smooth accretion of matter from the cosmic web. Depending on the mass of the halo and the cosmic epoch, there are two dominant modes of gas accretion (e.g., Birnboim & Dekel, 2003; Kereš et al., 2005; Dekel & Birnboim, 2006) the *hot mode*, wherein the accreted gas is shock-heated to the halo virial temperature $T_{\rm vir} \propto \frac{GM_{\rm vir}}{R_{\rm vir}}$; and the *cold mode*, wherein the gas is accreted in the form of dense clumps or filaments that can cool quickly (i.e., within a free-fall time[4]) and reach the halo centre.

The hot-accreted gas that cannot cool quickly remains in a gaseous halo. The cold-accreted gas and some of the hot-accreted gas dissipate energy and cool radiatively, then settle into a rotationally supported galactic disk due to the conservation of angular momentum. Dense regions within the disk fragment and collapse further to form stars. Stars inject radiation, energy and momentum into their surrounding medium, enriching it with heavy elements formed through stellar nucleosynthesis within stellar cores. These various forms of matter and energy injected by stars throughout their lifecycle are referred to as stellar feedback.

Stellar feedback can prevent further accretion of gas onto a halo as well as prevent gas cooling, which can (temporarily) halt further star formation. Additionally, feedback-driven outflows can expel gas from galaxies, further depleting the reservoir available for star formation. Stars also emit copious amounts of ionizing radiation (with energy greater than 13.6 eV) and gradually (re-)ionize the intergalactic medium (IGM).

The hydrogen in the Universe underwent a phase transition from predominantly atomic to ionised at $z \gtrsim 6$ in a period known as the Epoch of Reionization (EoR). Reionization affects the growth of galaxies by heating the gas in the IGM, which slows down the cooling and collapse of gas onto a galaxy, impeding further star formation. This suppression of accretion is most pronounced in low-mass ($M_{\rm halo} \lesssim 10^9\,{\rm M}_\odot$) halos, where the halo gravitational potential is not strong enough to retain the heated gas (Gnedin, 2000). Even after the EOR, the IGM is permeated by a background of ionising radiation (Haardt & Madau, 2001), which can heat the gas and inhibit cooling of gas in low-mass halos (see e.g., Barkana & Loeb 2001; Zaroubi 2013 for a detailed review)

In galaxies hosting an accreting black hole, i.e., an active galactic nucleus (AGN), accretion of gas onto the black hole and subsequent AGN feedback can heat up the gas, preventing radiative cooling and further accretion onto the halo. Additionally, AGN feedback can expel gas from the galaxy through powerful outflows and jets, thereby depleting the gas reservoir needed for star formation (Silk & Rees, 1998; Fabian, 2012).

[3] An object is said to be in virial equilibrium when its gravitational potential energy U and kinetic energy K are related as $2K + U = 0$.

[4] The free-fall time of gas is defined as the time it takes for it to collapse under its own gravity in the absence of pressure support.

Apart from smooth accretion, galaxies also grow via collisions with other galaxies, called mergers. The most obvious outcome of a merger is the redistribution of gas and stars among the merging galaxies. In some cases, mergers can cause the gas to lose its angular momentum support, inducing a gas flow to the center(s) of the merging galaxies, thereby triggering intense bursts of star formation and/or AGN activity (Hernquist, 1989)

Thus, the overall growth of galaxies is regulated by gas accretion, gas cooling, star formation, stellar feedback, AGN feedback, reionization, and galaxy interactions.

2.2. The interstellar medium

The material between the stars in a galaxy, commonly known as the interstellar medium (ISM), is highly inhomogeneous and dynamic, and hosts physical and chemical processes occurring in diverse conditions. It encompasses various energy sources and sinks that drive the flow of baryonic matter within and around galaxies. Most importantly, it serves as an interface for matter and energy exchange between stellar and galactic scales.

The ISM comprises mainly hydrogen (H), which makes up approximately 70% by mass and helium (He), which accounts for about 28% by mass. The remaining 2% consists of elements heavier than helium, collectively referred to as *metals*. The abundance of metals, commonly called metallicity, is denoted by Z. Some of these metals are present in solid form, referred to as *dust*. In a Milky-Way-like galaxy, roughly half of the metal mass is locked up in dust i.e., $Z_{\rm d} = 0.01$. Though containing only 1% of the ISM mass in the Milky Way, dust is an important ingredient contributing to the thermal balance in the ISM and acting as a catalyst for molecular chemistry.

The gas in the ISM exists in conditions spanning a wide range of densities and temperatures, from cold and dense ($T \sim 10 - 20$ K, $n \gtrsim 500\,{\rm cm}^{-3}$) molecular clouds to the hot ionised material with temperatures exceeding 10^5 K. Though it is not always possible to split the ISM into distinct phases and more so because all phases are transient, the following crude classification is standard in ISM studies. The conventional way is to split based on the ionisation state of the dominant element - hydrogen. The different phases are distinguished by their characteristic (range of) temperatures and densities and are listed below:

1. **Molecular gas phase**: in this phase, hydrogen is predominantly molecular (in the form of H_2) and is characterised by high gas densities ($n_{\rm H} \gtrsim 100\,{\rm cm}^{-3}$) and low gas temperatures ($T \sim 10 - 20$ K). It can be further split into dense ($n_{\rm H} \gtrsim 10^3\,{\rm cm}^{-3}$) and diffuse ($n_{\rm H} \gtrsim 10^2\,{\rm cm}^{-3}$) molecular gas. The dense part comprises clouds of molecular gas and dust, called molecular clouds, that are opaque to visible light.

2. **Cold neutral medium (CNM)**: comprises largely atomic gas at a temperature close to 100 K and density $n_{\rm H} \sim 30\,{\rm cm}^{-3}$. Together the molecular gas and CNM can account for $\gtrsim 70\%$ of the ISM mass while occupying only $1 - 2\%$ of its volume.

3. **Warm neutral medium (WNM)**: comprises atomic gas that is warmer than the CNM at a temperature of $\sim 6 \times 10^2 - 5 \times 10^3$ K and a density of $n_{\rm H} \sim 0.5 - 0.6\,{\rm cm}^{-3}$.

4. **Warm ionized medium (WIM) and H II regions**: comprises of gas photoionised by the UV radiation from young and massive O stars.[5] The typical temperatures are $\sim 10^4$ K. This phase can be further divided into: a) dense ($n_{e^-} \sim 10^2 - 10^4\,\mathrm{cm}^{-3}$) gas in the immediate vicinity of the star, called an H II region, and b) diffuse ($n_{e^-} \lesssim 10^2\,\mathrm{cm}^{-3}$) extended gas, referred to as the WIM.

5. **Hot ionised medium (HIM)**: comprises low-density gas that has been shock-heated and collisionally-excited by supernova explosions. It has typical temperatures exceeding $10^{5.5}$ K and densities $\sim 0.004\,\mathrm{cm}^{-3}$. This phase has the highest volume filling factor, e.g., in the Milky Way, the HIM occupies $\sim 50\%$ of the disk volume.

Within galaxies, matter can cycle from one phase to another through the various physical processes taking place in the ISM. Radiation from young stars can dissociate H_2 and ionize atomic hydrogen to form H II regions; free electrons and H^+ ions can combine to form neutral atomic hydrogen; and atomic gas can cool down and become dense to facilitate the formation of H_2. Some processes like outflows induced by stellar feedback transport matter back to the IGM, while others like accretion bring in a fresh supply of gas to the ISM. Molecular gas is the phase immediately preceding star formation. Most of this molecular gas exists in molecular clouds that are discussed in depth in Section 2.2.2.

2.2.1. Energy sources in the ISM

The physical state of the ISM is determined by the interaction of gas and dust with the interstellar radiation field (ISRF) and other energy sources (such as turbulence, cosmic rays, etc.). The ISRF at optical and UV wavelengths is determined by stellar radiation while dust continuum emission dominates at infrared wavelengths. The contribution of the cosmic microwave background depends on the redshift of the galaxy. Other energy sources include turbulence, magnetic fields, and cosmic rays.

Starlight

The far-ultraviolet (FUV) radiation (with energy $= 6 - 13.6$ eV and wavelength $\lambda = 912 - 2070$ Å) from young and massive OB stars[6] is responsible for the photoelectric heating of dust grains (Tielens & Hollenbach, 1985) and determining the chemical, ionisation, and thermal state of the star-forming ISM. It further regulates the balance between the different phases present in star-forming regions – H II regions, molecular gas, and CNM. The FUV spectrum of the ISM can be described in terms of the specific energy density $u_\lambda(\lambda)$[7] as a function of the wavelength. The mean energy density can then be obtained by integrating over the FUV band:

$$U_{\mathrm{FUV}} = \frac{1}{2070 - 912} \int_{912\text{Å}}^{2070\text{Å}} u_\lambda(\lambda)\,\mathrm{d}\lambda. \tag{2.14}$$

[5] In the spectral classification of stars, O-type stars are the most massive and luminous, with masses between 16-100 $M_\odot$ and up to 10^6 times more luminous than the Sun.

[6] The term OB stars collectively refers to O-type and B-type stars. O-type stars have masses between 16100 $M_\odot$ and can be up to 10^6 times more luminous than the Sun, while B-type stars have masses of 216 $M_\odot$ and luminosities ranging from 10^2 to 10^5 times that of the Sun.

[7] The specific energy density u_λ is a measure of the energy per unit volume per unit wavelength in a given region.

This mean energy density is often expressed in units of the typical energy density in the solar neighbourhood measured by Habing (1968) and is quantified by the G_0 parameter:

$$G_0 \equiv \frac{U_{\rm FUV}}{U_{\rm FUV,\,Habing}} = \frac{U_{\rm FUV}}{4.6 \times 10^{-17}\,{\rm ergs\,cm^{-3}\,\AA^{-1}}}\,. \tag{2.15}$$

Within the FUV band, a narrower band with energies in $11.2 - 13.6\,{\rm eV}$, called the 'H_2 band' or the Lyman-Werner (LW) band is of particular significance for hydrogen and carbon chemistry (see sections 2.7.1 and 2.7.2). In the solar neighbourhood, the mean energy densities in the LW and FUV bands are related as: $U_{\rm LW}/U_{\rm FUV} \sim 1.1$ (Parravano et al., 2003). The preferred choice of units for expressing the energy density in the LW band is χ which quantifies the energy density at 1000 Å (i.e., at 11.4 eV) with respect to that measured by Habing (1968):

$$\chi \equiv \frac{(\lambda u_\lambda)_{1000\text{Å}}}{(\lambda u_\lambda)_{1000\text{Å, Habing}}} = \frac{(\lambda u_\lambda)_{1000\text{Å}}}{4 \times 10^{-14}\,{\rm ergs\,cm^{-3}}}\,. \tag{2.16}$$

The FUV energy density is sometimes expressed in units of the ISRF estimate by Draine (1978), commonly referred to as the Draine field. It is equivalent to $G_0 = 1.69$ and $\chi = 1.71$.

Infrared emission from dust

UV and optical radiation from stars is absorbed by dust and re-emitted at infrared (IR, $\lambda \sim 1-1000\,\mu$m) wavelengths producing the IR continuum emission from galaxies. Individual dust grains absorb photons with wavelengths smaller than the grain size. As a result, the emission at different wavelengths is dominated by different-sized grains. Broadly speaking, polyaromatic hydrocarbons (PAHs) contribute to short wavelength ($\lambda \sim 3 - 20\,\mu$m) emission, stochastically heated[8] small dust grains dominate at $\lambda \sim 20-60\,\mu$m, while larger grains that are in thermal equilibrium with the ISFR contribute at $\lambda \gtrsim 60\,\mu$m. Consequently, the shape of the dust spectrum depends on not just the dust composition but also on the dust grain-size distribution, as well as on the spectrum of the starlight irradiating the dust.

Cosmic rays

Cosmic rays are an important heating source in the ISM. They are relativistic charged particles[9] (protons, electrons and heavy nuclei) that are thought to be accelerated in supernova remnants. Apart from heating the gas, low-energy ($E \lesssim 0.1\,{\rm GeV}$) cosmic rays are crucial for chemistry as these are the only way of initiating ion-ion chemistry deep within dense molecular clouds that are well-shielded from stellar UV radiation (Dalgarno, 2006; Padovani et al., 2009; Indriolo & McCall, 2013).

[8] Stochastic heating occurs when small dust grains absorb single UV photons, causing a temporary spike in temperature that results in emission, before cooling down and waiting for the next photon.

[9] The term 'cosmic rays' predates the discovery of the proton by Ernest Rutherford in 1917. Cosmic rays were first detected by Victor Hess in 1912 using Balloon experiments, for which he was awarded the Nobel Prize in 1936.

CMB

The CMB (section 2.1) has a blackbody spectrum with a temperature which evolves with redshift as $T_{\rm CMB}(z) = 2.73\,(1+z)$ K. The total energy density U follows from the Stefan-Boltzmann law: $U = 4\,\sigma_{\rm SB}\,T_{\rm CMB}^4\,/\,c$, where $\sigma_{\rm SB}$ is the Stefan-Boltzmann constant. The CMB is an important heating source, especially in high-redshift galaxies where it is believed to establish the temperature floor within molecular clouds.

2.2.2. Molecular clouds

Most of the molecular gas in a galaxy resides in clouds of gas and dust, called molecular clouds, where the high density and dust content provide effective shielding against the ISRF and UV radiation from nearby stars. Molecular clouds come in different sizes and the largest among these are giant molecular clouds (GMCs), with masses of $10^5 - 10^6\,{\rm M}_\odot$ and sizes of a few tens of parsecs. The typical number densities, n, in GMCs are in the range $n \sim 10^2 - 10^3$ cm^{-3}, and the temperatures are $\sim 10 - 20$ K.

Numerical simulations and Milky Way observations have shown that molecular clouds exhibit an intricate density structure that is broadly driven by the interplay between turbulence and gravity, with magnetic fields and radiation pressure also being significant factors. Molecular clouds contain pockets of dense ($n \sim 10^4 - 10^6$ cm^{-3}) regions dispersed within a more diffuse ($n \sim 0.1 - 10$ cm^{-3}) medium. This diversity can be described by a probability distribution function (PDF) of densities. For instance, the mass-weighted PDF gives the probability that an infinitesimal mass element dM has a density in the range $[n, n + {\rm d}n]$.

In an isothermal, supersonically-turbulent molecular cloud, not significantly affected by self-gravity, the mass-weighted PDF is close to a log-normal (Vázquez-Semadeni, 1994; Passot & Vázquez-Semadeni, 1998; Padoan & Nordlund, 2002; McKee & Ostriker, 2007). When self-gravity becomes important, and all or part of the cloud begins to collapse, dense clumps form within the cloud and the PDF develops a power-law tail towards high densities. A figure illustrating the time evolution of the mass-weighted density PDF in an isothermal self-gravitating turbulent medium from the ISM simulations by Kritsuk et al. (2011) is shown in Figure 2.1. Similar results were obtained by Federrath & Klessen (2013) based on high-resolution simulations of turbulent and magnetised molecular clouds.

The dense cores within molecular clouds are crucial for driving physical and chemical processes that depend strongly on gas density. For example, the rate of collisional interactions scales with the square of the density, n^2, indicating that collisions occur 10,000 times more frequently in gas that is 100 times denser. Consequently, any accurate modelling of these processes or their impact on galactic scales must account for these inhomogeneities.

Perhaps the single most important contribution of molecular clouds in shaping galaxies arises from their role as *stellar nurseries*. Observations within the Milky Way and nearby galaxies have also shown that most star formation happens within molecular clouds. Across their different evolutionary stages, stars continuously infuse their parent cloud with ionising radiation, mass, energy, momentum, and metals. Massive stars end their lifecycle in a violent supernova (SN) explosion. SN explosions along with stellar winds and ionising radiation are important sources of turbulence in the ISM. In some cases, the energy injected by star clusters or SN explosions could be sufficient to disrupt the entire cloud.

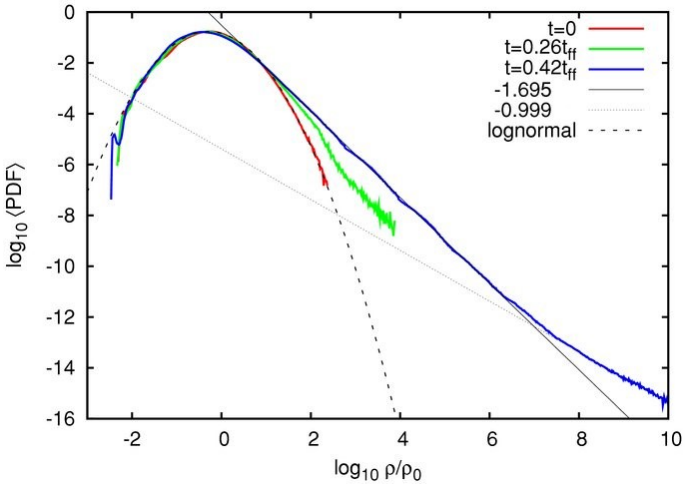


Figure 2.1. – Time evolution of the mass-weighted density PDF in high-resolution simulations of turbulent, magnetised gas *(from Kritsuk et al. 2011)*. The red line shows the initial PDF which is imposed to be a log-normal, shown by the black dashed line. The green and blue curves show the PDF at different times in the simulation. The gray solid and dotted lines represent power-laws with slopes $\sim$ -1.7 and $\sim$ -1, respectively.

As mentioned before, the high-density power-law tail in the density distribution of molecular clouds signifies dense gravitationally collapsed regions or clumps. But what happens when a large amount of energy is injected into this molecular cloud? Do the high-density clumps survive or does everything return to an effectively log-normal distribution? What happens to the molecules? As time passes and the initial effects decay, does the cloud reform? On what timescales?

There is growing evidence in recent years of the fast disruption of molecular clouds by pre-supernova feedback on a $\sim$ 3-5 Myr timescale (see Chevance et al., 2023, for a recent review). While SN feedback begins to act only a few (3-10) million years after star formation, pre-SN feedback in the form of stellar winds, photoionisation, and radiation pressure already starts acting at the onset of star formation. These processes are highly efficient at dispersing the gas within molecular clouds. Ultimately, whether or not the entire cloud and its constituent molecules are destroyed, it is clear that these processes significantly alter the density structure and subsequently the star formation happening within molecular clouds. This is one of the key mechanisms through which stellar feedback regulates subsequent star formation within the molecular cloud.

2.2.3. Photon-dominated regions

Photon-dominated regions (PDRs, historically known as photodissociation regions) represent the interface between H II regions surrounding hot massive OB stars and their parent molecular clouds. PDRs are primarily composed of neutral gas, where the chemistry and heating are regulated by FUV photons. A schematic view of the physical and chemical state across a PDR is shown in Figure 2.2. While the UV radiation from a young star ionizes the gas in its immediate vicinity, the intensity of the radiation decreases as it travels

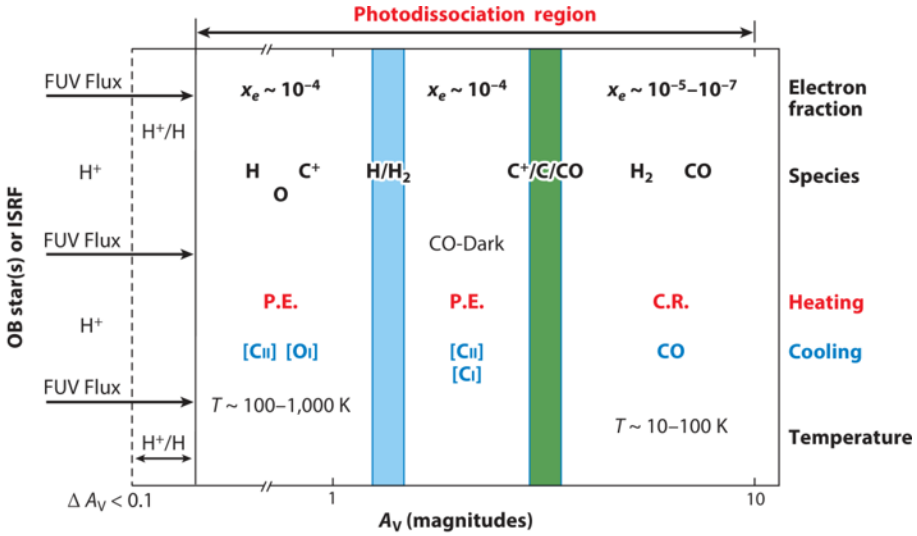


Figure 2.2. – A schematic diagram of a photodissociation region *(Adapted from Wolfire et al. 2022)*.

through a large column of gas, reaching a point where the gas is mostly molecular. This effective shielding against UV radiation is expressed as the visual extinction A_V:

$$A_V = N_{\rm H}\, Z_{\rm d} \,/\, (1.87 \times 10^{21}\,{\rm cm}^{-2})\,, \tag{2.17}$$

where $N_{\rm H}$ is the total column density of hydrogen nuclei, and $Z_{\rm d}$ is the dust abundance relative to the Solar neighbourhood value of 0.01.

PDRs are rich in chemical species ranging from ions, atoms and simple molecules like H_2 and CO to complex organic molecules and polyaromatic hydrocarbons. As of this day, more than 320 molecules have been detected in the ISM or circumstellar shells (the Cologne Database for Molecular Spectroscopy, CDMS Endres et al., 2016). The infrared line and continuum emission from PDRs provide a window into their rich chemistry and physics and are often used to infer the physical conditions within the PDRs. Moreover, these emission processes also give insights into the conditions of star formation (such as density, temperature, and visual extinction). The observation and modelling of PDRs is an active and evolving field of research in astrophysics.

PDR models are employed to solve, often iteratively, for the thermal and chemical state of the ISM and the radiation field as a function of A_V. It is worth noting that despite the simplistic representation of Figure 2.2, PDRs are far from being static and are continuously evolving (see Wolfire et al., 2022, for a recent review).

2.2.4. Star formation and molecular gas

The different processes and components of the ISM affect the star formation within a galaxy, primarily through its gas content. Observationally, the surface density of star formation correlates with the gas[10] surface density as:

$$\Sigma_{\rm SFR} \propto \Sigma_{\rm gas}^{N_{\rm SK}}\,. \tag{2.18}$$

[10]We note that the term 'gas' here refers to the neutral i.e., atomic + molecular, gas.

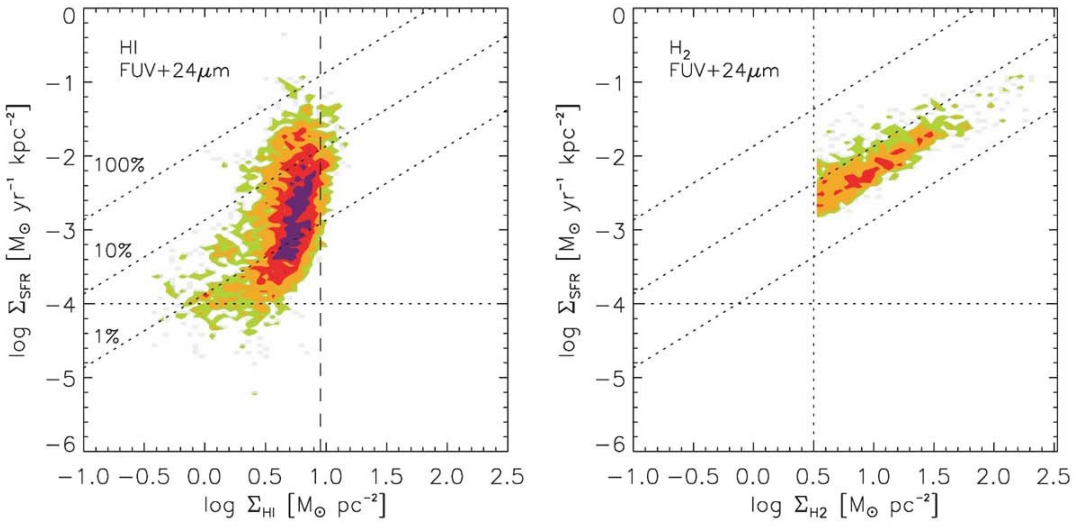


Figure 2.3. – The atomic and molecular Schmidt-Kennicutt relation on sub-kpc scales for nearby galaxies (taken from Bigiel et al., 2008).

This correlation is referred to as the Schmidt-Kennicutt (SK) relation (Schmidt, 1959; Kennicutt, 1998). On galaxy-wide scales in local galaxies, $N_{\rm SK} \sim 1.4 \pm 0.15$ (Kennicutt, 1998; Kennicutt & Evans, 2012a). Several observations have shown that the SK relation holds not only for entire galaxies, but also for (sub-)kpc regions within galaxies (e.g., Wong & Blitz, 2002; Bigiel et al., 2008, 2010; Leroy et al., 2008, 2013; Schruba et al., 2011, etc.), and individual molecular clouds (e.g., Evans et al., 2009; Heiderman et al., 2010; Lada et al., 2010, 2012). It should however be noted that the value of the exponent in Equation 2.18 is sensitive to the scale in consideration. Moreover, while early studies targeted nearby galaxies, the SK relation has now been examined out to $z \sim 6$ (e.g., Bouché et al., 2007; Daddi et al., 2010a,b; Genzel et al., 2010; Saintonge et al., 2013; Tacconi et al., 2013; Freundlich et al., 2013, 2019; Hodge et al., 2015; Béthermin et al., 2023; Vallini et al., 2024).

If we split the gas into atomic hydrogen (H I) and molecular hydrogen (H_2), and inspect separately how each of these relates to $\Sigma_{\rm SFR}$ (as shown in Figure 2.3), an absence of a tight correlation between $\Sigma_{\rm SFR} - \Sigma_{\rm H\,I}$ is immediately evident. In contrast, a strong and nearly linear correlation emerges between $\Sigma_{\rm SFR}$ and $\Sigma_{\rm H_2}$ (see e.g., Wong & Blitz, 2002; Bigiel et al., 2008; Leroy et al., 2008):

$$\Sigma_{\rm SFR} \propto \Sigma_{\rm H_2} \,. \tag{2.19}$$

The existence of this strong correlation has led several astronomers to postulate that molecular gas is the primary fuel for star formation (Krumholz & McKee, 2005; Elmegreen, 2007; Krumholz et al., 2009). On the contrary, another interpretation suggests that the formation of H_2 is a consequence of, rather than a prerequisite for, the conditions conducive to star formation. Molecular hydrogen thrives in dense, cold gas well-shielded from dissociating UV radiation the same conditions that favour the gravitational collapse of gas and subsequent star formation (see Glover & Clark, 2012a, and references therein).

Nevertheless, simulations (Kuhlen et al., 2012b, 2013; Tomassetti et al., 2015) and observations (McKee & Ostriker, 2007; Bigiel et al., 2008) support that most star formation in metal-enriched environments ($Z \gtrsim 10^{-3}\,\mathrm{Z}_{\odot}$) is fuelled by molecular hydrogen. As this thesis does not delve into early metal-poor star formation, in this work, we assume that stars *do* form from molecular gas. This allows us to recast the nearly linear molecular SK relation, given in Equation (2.19), to obtain an analogue connecting the respective volume densities as:

$$\dot{\rho}_{\mathrm{sf}} \;=\; \epsilon_{\mathrm{sf}}\,\frac{\langle \rho_{\mathrm{H_2}} \rangle}{t_{\mathrm{ff}}}\,, \tag{2.20}$$

where ϵ_{sf} is the star formation efficiency i.e., the SFR per unit molecular gas mass and $t_{\mathrm{ff}} = \sqrt{3\,\pi/32\,G_{\mathrm{N}}\,\langle \rho_{\mathrm{gas}} \rangle}$ is the local free-fall time of the gas i.e., the time it takes for the gas to collapse under its own gravity, in the absence of pressure support. The brackets denote an average over a given spatial scale. Observations of Milky-Way molecular clouds have shown that $\epsilon_{\mathrm{sf}} \approx 1 - 2\%$ for over four orders of magnitude in $\rho_{\mathrm{H_2}}$ (Krumholz & Tan, 2007). This general prescription has been widely used as a star-formation recipe in numerical simulations that do not resolve the spatial, temporal, and mass scales relevant for modelling the formation of individual stars.

Equation 2.20 provides an important insight that star formation in a given region or galaxy is governed by the self-gravity of the gas and regulated by a host of processes that counter gravity such as stellar feedback, magnetic fields, and turbulence (reflected in $\epsilon_{\mathrm{sf}} < 100\%$).

Observing the molecular gas content of galaxies alongside their star formation is crucial for understanding the physical conditions and efficiency of star formation within galaxies. Before exploring this further, it is imperative to first describe the observational techniques used for measuring the SFR and molecular gas mass in galaxies across cosmic time. For this, we will first introduce the basic concepts of line emission and radiative transfer that are key to observational astronomy.

2.3. Radiative transfer

Radiative transfer describes how a radiation field changes as it interacts with matter along its path of propagation. The strength of a radiation field can be described in terms of its specific intensity which is defined as the energy passing through a unit area of cross-section per unit time, propagating in a direction $\hat{\boldsymbol{n}}$, per unit frequency interval (around the frequency ν), per unit solid angle $\mathrm{d}\Omega$:

$$I_{\nu}(\nu,\,\hat{\boldsymbol{n}}) = \frac{\text{Energy}}{\text{time}\cdot\text{area}\cdot\mathrm{d}\nu\cdot\mathrm{d}\Omega}\,. \tag{2.21}$$

Alternatively, one can compute the specific energy density at the frequency ν by integrating I_{ν} over the solid angle:

$$u_{\nu} = \frac{\text{Energy}}{\text{Volume}} = \frac{1}{c}\int I_{\nu}\,\mathrm{d}\Omega\,. \tag{2.22}$$

Note that u_{ν} is independent of direction by construction.

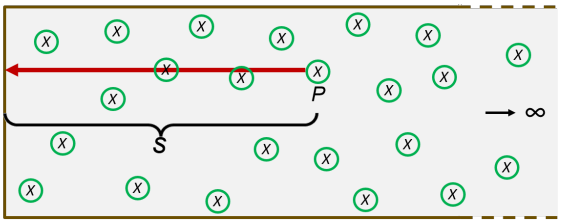


Figure 2.4. – A schematic representation of of semi-infinite slab filled with a species X. The red arrow denotes the path of radiation emitted by a particle of X, located at point P at a depth s into the slab. This simple representation neglects scattering and the change in the direction of radiation. as it travels through the slab.

2.3.1. Line Emission

Consider a two-level system of a species X where the upper and lower energy levels are denoted as u and l, respectively, and the respective number densities are denoted as n_u and n_l. The relative populations of the upper and lower excitation levels are captured by the excitation temperature $T_{\rm ex}$:

$$\frac{n_u}{n_l} = \frac{g_u}{g_l}\, e^{-T_*/T_{\rm ex}}\,, \tag{2.23}$$

where g_u and g_l are the statistical weights of the levels and $E = h\nu_{ul} = k_{\rm B}T_*$ is the energy difference[11] between the levels ($k_{\rm B}$ being the Boltzmann constant and h being the Planck constant). Let u_{ul} be the specific energy density at the transition frequency ν_{ul}. The level populations can change as a result of the following processes:

1. Photon absorption:
$$\frac{{\rm d}n_u}{{\rm d}t} = -\frac{{\rm d}n_l}{{\rm d}t} = B_{lu}\, n_l\, u_{ul}\,.$$

2. Stimulated emission:
$$\frac{{\rm d}n_l}{{\rm d}t} = -\frac{{\rm d}n_u}{{\rm d}t} = B_{ul}\, n_u\, u_{ul}\,.$$

3. Spontaneous emission:
$$\frac{{\rm d}n_l}{{\rm d}t} = -\frac{{\rm d}n_u}{{\rm d}t} = A_{ul}\, n_u\,.$$

4. Collisional excitation:
$$\frac{{\rm d}n_u}{{\rm d}t} = -\frac{{\rm d}n_l}{{\rm d}t} = C_{lu}\, n_l\,.$$

5. Collisional de-excitation:
$$\frac{{\rm d}n_l}{{\rm d}t} = -\frac{{\rm d}n_u}{{\rm d}t} = C_{lu}\, n_u\,.$$

[11] It is common in spectroscopy to report the energy separation between the upper and lower levels of a transition in units of temperature.

Here A_{ul} is the Einstein A coefficient and denotes the rate of spontaneous de-excitation (or spontaneous emission), B_{ul} and B_{lu} are Einstein B coefficients for stimulated de-excitation and excitation, respectively. For a given transition, the three coefficients are related as

$$g_l\, B_{lu} = g_u\, B_{ul}; \qquad B_{ul} = \frac{c^3}{8\pi h \nu_{ul}^3} A_{ul}\,. \tag{2.24}$$

The downward and upward collision rate coefficients, C_{ul} and C_{lu}, are related as

$$\frac{C_{lu}}{C_{ul}} \;=\; \frac{g_u}{g_l}\, \exp(-h\nu_{ul}/k_\mathrm{B}\, T_\mathrm{kin})\,, \tag{2.25}$$

where T_kin is the kinetic temperature of the collision partner.[12] Any given emitting species can have multiple collision partners, each with a different kinetic temperature.

In statistical equilibrium (i.e., $\mathrm{d}n_l/\mathrm{d}t = \mathrm{d}n_u/\mathrm{d}t = 0$), the excitation and de-excitation processes for each level balance each other and we obtain:

$$n_l(B_{lu}u_{ul} + C_{lu}) \;=\; n_u(A_{ul} + B_{ul}u_{ul} + C_{ul})\,. \tag{2.26}$$

Now let us consider a semi-infinite slab of matter filled with the species X as shown in Figure 2.4. The emission originating from a point P at a physical depth s into the slab interacts with other atoms via the aforementioned radiative and collisional processes on its way to the edge of the slab. These interactions alter the intensity of radiation, which can be expressed by the radiative transfer equation:

$$\mathrm{d}I_\nu = -I_\nu\, \kappa_\nu\, \mathrm{d}s + j_\nu\, \mathrm{d}s\,. \tag{2.27}$$

The first term captures the net change due to absorption and stimulated emission, while the second term denotes the enhancement of the radiation field due to spontaneous emission. Here, j_ν is referred to as the emissivity of the material at the frequency ν. Note that the equation neglects scattering.

To obtain the radiation intensity escaping the slab, one must integrate Equation 2.27 across the slab. The emissivity j_ν can vary along the depth of the slab and depends on the level populations of the emitting species X at each position. These level populations are influenced by the local radiation intensity, which in turn is affected by emissions from all other locations within the slab. Therefore, we are faced with a conundrum: solving for the level populations (Equation 2.26) requires knowledge of the radiation density u_{ul}, which is influenced by the level populations themselves. Over the years, several numerical schemes have been developed to solve this system of coupled equations, often employing iterative approaches and/or simplifying assumptions.

The simplest among these assumes a local thermal equilibrium between the excitation temperature of the species of interest and the gas kinetic temperature, i.e., $T_\mathrm{ex} = T_\mathrm{kin}$. This condition is known as Local Thermodynamic Equilibrium (LTE). In LTE, the level populations follow a Boltzmann distribution determined by the local kinetic temperature. LTE is generally applicable at high densities where collisions are frequent enough to bring the level populations in equilibrium with the thermal motion of the collision partner.

[12] The kinetic temperature of a species is the temperature characterising the velocity distribution of the constituent atoms/molecules, that follows a Maxwell-Boltzmann distribution.

Another class of methods relaxes the LTE assumption but nevertheless assumes 'local excitation' which means that the local radiation field determines the excitation in a given region. These solve for statistical equilibrium between the excitation and de-excitation channels.[13] The most widely used among these are described below:

1. Escape Probability formalism: This method estimates the fraction of photons that escape a medium without being absorbed or scattered, based on its optical depth. Depending on the system's geometry (e.g., plane-parallel slab, spherical symmetry, etc.), one can calculate the effective escape fraction of photons from the system. Radiative transfer codes like RADEX (van der Tak et al., 2007), DESPOTIC (Krumholz, 2014) and CLOUDY (Ferland et al., 1998) are based on the escape probability formalism. These codes solve the 3D radiative transfer problem by effectively reducing the system to a 1D geometry, e.g., by assuming a spherically symmetry gas distribution.

2. Large Velocity Gradient (LVG): The LVG approximation (also known as the Sobolev approximation, Sobolev 1957) assumes the existence of large velocity gradients across a region. Initially developed for an expanding spherical shell geometry, this method is also widely used to mimic turbulent motions with moderate velocity gradients. A crucial difference between the escape probability and LVG methods is that while the former reduces a 3D system to an effective 1D system, the latter calculates the angle-averaged velocity gradient based on which an escape probability is estimated.

Finally, the most advanced methods further relax the assumption of local excitation, calculating the level populations at any given location by accounting for the radiation field at all other locations. The most common among these are Monte Carlo methods, which sample the radiation field in a region with photon packets and follow the interaction of each packet with gas and dust along its path. Consequently, these methods are expected to be the most accurate among all described so far, but they can be computationally expensive due to the high number of photon packets and scattering processes involved. Examples of radiative transfer codes based on this approach include RADMC-3D (Dullemond et al., 2012) and SKIRT (Baes et al., 2011).

2.3.2. Units

The specific intensity of a blackbody at temperature T follows from Planck's law:

$$I_\nu \;=\; B_\nu(T) \;=\; \frac{2h\nu^3}{c^2}\,\frac{1}{\exp(h\nu/k_\mathrm{B}T)-1}\,. \tag{2.28}$$

In astronomy, it is often convenient to express the specific intensity I_ν of a source in terms of a brightness temperature T_B, where T_B is simply the temperature of a blackbody with $B_\nu(T_\mathrm{B}) = I_\nu$, which gives:

$$T_\mathrm{B}(\nu) \equiv \frac{h\nu\,/\,k_\mathrm{B}}{\ln\left[1+2h\nu^3\,/\,c^2 I_\nu\right]}\,. \tag{2.29}$$

[13] Note that LTE is a special case of statistical equilibrium, and statistical equilibrium does not always imply LTE. In non-LTE situations, the level populations do not follow a Boltzmann distribution, but the level populations can still be in statistical equilibrium, i.e., $\mathrm{d}n_i/\mathrm{d}t = 0$, if the upward and downward rates are balanced.

The line luminosity $L_{\rm line}$ of a source, i.e., the total energy emitted per unit time in all directions in a given spectral line, can be obtained by integrating the specific intensity I_ν over the solid angle, surface area, and frequency interval, accounting for the line profile. The line profile $\phi_\nu(\nu)$ describes the frequency dependence of $I_\nu(\nu)$ around the central frequency ν_{ul}. It includes contributions from both the uncertainty in the energy of levels u and l (natural broadening[14]) and the thermal motion of the particles of the emitting species (thermal broadening). The luminosity obtained this way is expressed in units of solar luminosity $\mathrm{L}_\odot$.

Alternatively, one can integrate the source brightness temperature $T_{\rm B}(\nu)$ over the surface area of the source and the frequency (again accounting for the line profile). The line luminosity obtained this way is conventionally denoted as $L'_{\rm line}$ and has units $\mathrm{K\,km\,s^{-1}\,pc^2}$. Here the line profile is expressed in units of velocity and is related to $\phi_\nu(\nu)$ as: $\phi_v(v) = \phi_\nu(\nu_{ul}\,(1 - v/c))$ (Draine, 2003).

2.4. Observing molecular gas | Molecular hydrogen and the need for tracers

Molecular hydrogen (H_2) is the dominant component of molecular gas. Among the various types of radiative transitions (electronic, rotational, and vibrational) of the H_2 molecule, rotational transitions have the lowest energy. However, being a symmetric molecule, H_2 lacks a permanent dipole moment and does not exhibit any dipolar rotational transitions. The two lowest quadrupole rotational transitions (one each for ortho and para H_2) have an energy difference corresponding to a temperature $T \gtrsim 500$ K, and are only excited in gas with $T \gtrsim 100$ K. Consequently, H_2 does not emit electromagnetic radiation under the conditions prevalent in molecular clouds that typically have temperatures of $10 - 20$ K.

Studies of molecular gas therefore typically rely on electromagnetic radiation from another species that serves as a proxy or tracer of H_2. A reliable tracer must satisfy the following requirements: a) it must spatially coexist with H_2; and b) it must emit electromagnetic radiation under the typical conditions (i.e., temperature and density) of molecular clouds. There are several tracers for observing the molecular gas in high-redshift galaxies (see Tacconi et al., 2020, for a recent review; also see Carilli & Walter 2013 and Combes 2018). The most common among these are reviewed below.

2.4.1. Carbon monoxide

The second most abundant molecule in the ISM – carbon monoxide (CO) is the most commonly used tracer for H_2. CO has a weak permanent dipole moment ($\mu \sim 0.11$ D) and its lowest rotational transition ($J = 1 \rightarrow 0$) is excited at $T \sim 5.5$ K. Because of the high Einstein A coefficient of this transition, CO emits brightly at radio wavelengths and can be easily detected by ground-based telescopes at $z \lesssim 1$. Therefore, CO is routinely used to trace molecular gas in galaxies (Solomon et al., 1987; Solomon & Barrett, 1991; Solomon & Vanden Bout, 2005).

[14]Natural broadening arises from the Heisenberg uncertainty principle which states that the energy E_i of an energy level i, with a lifetime τ_i, is uncertain by an amount ΔE_i such that $\Delta E_i\,\tau_i \approx h/4\pi$.

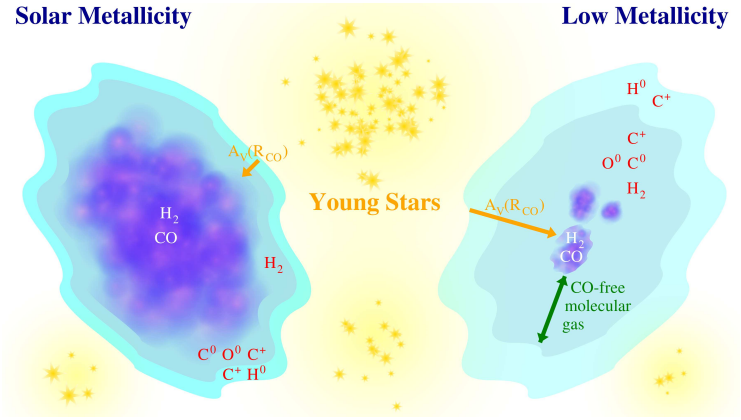


Figure 2.5. – An illustration of the CO-dark molecular gas in low-metallicity environments (from Madden et al., 2020).

Since the CO lines are typically optically thick, the CO mass (and likewise the H_2 mass, M_{H_2}) in a galaxy cannot be directly inferred from the CO luminosity alone. The standard method for estimating M_{H_2} from CO emission therefore relies on a conversion factor α_{CO} that relates the CO $J = 1 \rightarrow 0$ luminosity ($L'_{CO\,(1-0)}$) to M_{H_2} as: $M_{H_2} = \alpha_{CO}\, L'_{CO\,(1-0)}$.

Several observational (Genzel et al., 2012; Schruba et al., 2012; Leroy et al., 2013; Accurso et al., 2017) and numerical (Glover & Mac Low, 2011; Shetty et al., 2011a,b; Feldmann et al., 2012a,b; Clark & Glover, 2015; Narayanan et al., 2012; Narayanan & Krumholz, 2014; Madden et al., 2020) studies have found that α_{CO} varies with metallicity, gas density, and the strength of the UV radiation field. In Milky-Way-like ISM conditions, the typical value is $\alpha_{CO} = 3.6\,M_\odot\,(K\,km\,s^{-1}\,pc^2)^{-1}$ (Bolatto et al., 2013). Ultra luminous infrared galaxies (ULIRGs with infrared luminosities $L_{IR} \gtrsim 10^{12}\,L_\odot$) tend to have lower values of $\alpha_{CO} \sim 0.8\,M_\odot\,(K\,km\,s^{-1}\,pc^2)^{-1}$ (Downes & Solomon, 1998; Papadopoulos et al., 2012; Dunne et al., 2022). On the other hand, a higher value is exhibited by low-metallicity dwarf galaxies (Wolfire et al., 2010; Sandstrom et al., 2013) that are expected to have similar metal enrichment as typical high-redshift galaxies.

At higher redshifts ($z \gtrsim 1$), the use of CO is complicated by the fact that the $J = 1 \rightarrow 0$ transition is not accessible from the ground and observers have to rely on higher-J transitions to obtain an estimate for it. Down-converting the observed CO $J \rightarrow (J-1)$ luminosity to the CO $1 \rightarrow 0$ luminosity requires knowledge of the CO excitation ladder, i.e., the relative occupancy of the different rotational levels of CO, thereby introducing another systematic uncertainty in employing CO as a molecular gas tracer.

Moreover, low-metallicity galaxies are expected to harbour large amounts of molecular gas devoid of CO emission, commonly referred to as 'CO-dark gas' (Wolfire et al., 2008). This happens for two main reasons: a lower metallicity results in a reduced carbon abundance and a lower dust content, that together result in a reduced shielding of CO molecules from dissociating UV photons.

2.4.2. Atomic carbon

The ^3P fine-structure levels ($^3P_0$, $^3P_1$, and $^3P_2$) in the ground electronic state of atomic carbon form a three-level system, where $^3P_1$ and $^3P_2$ are at $T_1 = 23.5$ K and $T_2 = 62.5$

K above the ground level, respectively. This system results in two fine-structure lines – [C I] $^3P_1 \rightarrow {}^3P_0$ or [C I] (1-0)) at 609 μm and [C I] $^3P_2 \rightarrow {}^3P_1$ at 370 μm. Due to the low energy separations, these transitions can be easily excited under typical molecular cloud conditions. Observations (Ikeda et al., 2002) and numerical simulations (Papadopoulos et al., 2004; Bisbas et al., 2015, 2017; Glover et al., 2015) have shown that [C I] emission is fully concomitant with CO emission in molecular clouds. Simulations further suggest that [C I] is a more reliable molecular gas tracer than CO in metal-poor regions (Papadopoulos et al., 2004; Glover & Clark, 2016).

As a result, atomic carbon has been employed as a molecular gas tracer in star-forming galaxies at $z \gtrsim 1$ (Weiß et al., 2003, 2005; Walter et al., 2011; Alaghband-Zadeh et al., 2013; Valentino et al., 2018; Boogaard et al., 2020; Harrington et al., 2021; Henríquez-Brocal et al., 2022). The key advantage of carbon lines over CO lines is that the former are optically thin (Ojha et al., 2001; Ikeda et al., 2002; Weiß et al., 2003), which allows for a direct estimation of the C I mass from the line luminosity, if one assumes an excitation temperature $T_{\rm ex}$ (Weiß et al., 2003):

$$\frac{M_{\rm C\,I}}{\rm M_\odot} = 5.706 \times 10^{-4}\, Q(T_{\rm ex})\, \frac{1}{3}\, \exp(T_1/T_{\rm ex})\, \frac{L'_{\rm [C\,I]\,(1-0)}}{\rm K\,km\,s^{-1}\,pc^2}\,; \tag{2.30}$$

$$\frac{M_{\rm C\,I}}{\rm M_\odot} = 4.566 \times 10^{-4}\, Q(T_{\rm ex})\, \frac{1}{5}\, \exp(T_2/T_{\rm ex})\, \frac{L'_{\rm [C\,I]\,(2-1)}}{\rm K\,km\,s^{-1}\,pc^2}\,, \tag{2.31}$$

where $Q = 1 + 3e^{-T_1/T_{\rm ex}} + 5e^{-T_2/T_{\rm ex}}$ is the partition function of the ^3P system of C I (Weiß et al., 2003, 2005).

If the level populations are assumed to be in LTE, $T_{\rm ex}$ can instead be determined directly from the line luminosity ratio $R = L_{\rm [C\,I]\,(2-1)}/L_{\rm [C\,I]\,(1-0)}$ as (Weiß et al., 2003, 2005):

$$T_{\rm ex} = \frac{38.8\,\rm K}{\ln(2.11/R)}\,. \tag{2.32}$$

Once the atomic carbon mass has been estimated, it can be translated into an H_2 mass by assuming an atomic carbon abundance relative to H_2:

$$X_{\rm C\,I} = [\rm C\,I\,/\,H_2] = \frac{M_{\rm C\,I}}{6 M_{\rm H_2}}\,, \tag{2.33}$$

which represents the primary systematic uncertainty associated with this tracer. Observations at $z \gtrsim 2$ typically adopt a value of 3×10^{-5} (Weiß et al., 2003). Independent calibrations of $X_{\rm C\,I}$ using other molecular gas tracers in $z > 1$ galaxies yield values between $(2-8) \times 10^{-5}$ (e.g., Valentino et al., 2018; Boogaard et al., 2020; Harrington et al., 2021). However, unlike CO, a systematic calibration of $X_{\rm C\,I}$ across the different galaxy types and redshifts is still lacking.

2.4.3. Ionized carbon

The ground electronic state of singly ionized carbon (C^+) features a two-level fine-structure system, consisting of the $^2P_{3/2}$ and $^2P_{1/2}$ levels, separated by an energy difference of 91.25 K. The fine-structure transition $^2P_{3/2} \rightarrow {}^2P_{1/2}$ produces an emission line at

157.74, μm, commonly referred to as the [C II] line. [C II] is an important cooling line in the neutral ISM and has been observed to correlate with the SFR in local and high-redshift galaxies (Stacey et al., 1991, 2010; De Looze et al., 2011, 2014; Herrera-Camus et al., 2015; Carniani et al., 2018; Schaerer et al., 2020), on both global and resolved scales. In recent years, the line has also emerged as a tracer of molecular gas (e.g., Hughes et al. 2017; Zanella et al. 2018; Madden et al. 2020; also see Carilli & Walter 2013 for a review), especially of CO-dark molecular gas (Madden et al., 2020).

Owing to its brightness, contributing ~ 0.1 to a few percent of the total IR luminosity (Stacey et al., 1991), the line is now routinely used as a molecular gas tracer at $z \gtrsim 4$ (Dessauges-Zavadsky et al., 2020; Aravena et al., 2024). At these high redshifts, conventional tracers like CO become observationally expensive, while [C II] remains bright and is conveniently redshifted into a more favourable atmospheric window accessible from the ground, making it a powerful probe of the high-z ISM. The line has been found to correlate well with CO$(1-0)$ emission in $z \gtrsim 2$ star-forming galaxies (Gullberg et al., 2015), further adding to its credibility.

Zanella et al. (2018) compiled a sample of [C II]-detected galaxies from $z \sim 0-5.5$ with independent estimates of the molecular gas mass ($M_{\rm mol}$[15]) and obtained the following best-fit relation for their galaxy sample:

$$\log\left(\frac{L_{\rm [C\,II]}}{\rm L_\odot}\right) = -1.28(\pm0.21) + 0.98(\pm0.02)\log\left(\frac{M_{\rm mol}}{\rm M_\odot}\right), \tag{2.34}$$

with a scatter of ≈ 0.3 dex around the best fit, resulting in a $\alpha_{\rm [C\,II]} \equiv M_{\rm mol}\,/\,L_{\rm [C\,II]}$ of $31^{+30}_{-15}\,\rm M_\odot\,L_\odot^{-1}$. The scarcity of observed galaxies with independent molecular gas mass estimates at $z \gtrsim 4$ has hampered efforts to calibrate this relation and examine its evolution over time.

2.4.4. 850 μm dust emission

Since dust and gas are well-mixed within the ISM, measurements of the infrared dust emission has been proposed as a reliable tracer of the total gas mass (Scoville, 2013; Scoville et al., 2016), which provides an upper limit on the molecular gas mass. If there are observations of the dust spectral energy distribution (SED) in the Rayleigh-Jeans part of the SED (i.e., where $\lambda >> h\,c/\rm k_B\,T_{dust}$), then one can assume that the emission is optically thin and directly obtain the total dust mass from a single broadband measurement around 850 μm. Assuming a dust-to-gas ratio allows one to infer the total gas mass.

However, this method requires assuming a dust temperature, a dust opacity (that determines the shape of the dust SED), and the dust-to-gas ratio. As a result, the combined uncertainty introduced is ~ 0.3 dex which is similar to that while using higher-J CO lines.

2.5. SFR indicators

Emission from galaxies at wavelengths from FUV ($\lambda \sim 1000-2000$ Å) through IR ($\lambda \sim 1-1000\,\mu$m) to radio wavelengths ($\lambda \gtrsim 0.5$ m) is used to infer their SFR. Here we briefly review the most widely used SFR indicators (see Kennicutt & Evans, 2012a; Calzetti, 2013, for a detailed review):

[15] Note that $M_{\rm mol}$ includes a 36% contribution by mass from helium.

1. Hα **emission**: The Hα or Balmer α recombination line of hydrogen at 6564.6 Å is an optical emission line arising from H II regions around young and massive OB stars. The Hα emission is dominated by stars younger than 10 Myr, and therefore traces recent star formation.

2. **UV continuum emission**: The UV continuum radiation comprises of light from stars with ages $\sim 10 - 200$ Myr, with young stars (ages 10-100 Myr) dominating the emission at shorter wavelengths (FUV; $\lambda \sim$ 1220-2000 Å) and evolved stars (ages 100-200 Myr) at longer wavelengths (NUV; $\lambda \sim$ 2000-4000 Å). About 90% of the emission at FUV (NUV) wavelengths comes from stars younger than 100 Myr (200 Myr; see Table 1 of Kennicutt & Evans, 2012a).

3. **IR emission**: UV and optical radiation from stars is absorbed by dust and re-emitted at longer (IR) wavelengths. The total IR luminosity provides a useful, though *crude*, tracer of star formation on $\sim$ 100 Myr timescales (Kennicutt, 1998). It is most commonly employed for nearby galaxies, where complete infrared wavelength coverage can be achieved.

   In galaxies where this is not feasible (e.g., high-redshift galaxies), monochromatic IR indicators, typically centred on specific instrument wavebands such as 24 μm, 70 μm, and 100 μm provide an excellent alternative (see e.g., Calzetti et al., 2010). In general, shorter wavelengths ($\lesssim$ 60 μm) trace the emission from dust heated by massive, short-lived stars, while longer wavelengths ($\sim$ 100-150 μm) are dominated by dust heated by longer-lived, low-mass stars. The 24 μm emission primarily arises from stochastically heated small dust grains in star-forming regions and is directly proportional to the UV photon flux, making it useful for correcting the Hα-based SFR for dust extinction (Calzetti et al., 2005, 2007). The 24 μm emission is sensitive to star formation on 100 Myr timescales (Kennicutt & Evans, 2012a).

4. **Far infrared lines**: FIR emission lines such as [C II] at 157.74 μm, [O I] at 63 μm, and [O III] at 88 μm have been shown to trace star formation in galaxies(De Looze et al., 2014). These lines offer the advantage of being largely unaffected by dust extinction and thus trace the total (obscured+unobscured) star formation. Additionally, they are bright and therefore particularly useful for estimating star formation in high-redshift galaxies. Testing the reliability of these lines as SFR tracers, both in the local Universe and at high redshifts, is a field of continued research (e.g., De Looze et al., 2014; Herrera-Camus et al., 2015; Carniani et al., 2018; Schaerer et al., 2020).

2.6. The history of star formation in the Universe

After reviewing the common observational tracers of H_2 and star formation, we are ready to zoom out of individual galaxies and their ISM and consider the star formation happening on global scales. Mapping out the star formation history of the Universe has been the objective of numerous observational studies over the past three decades. These

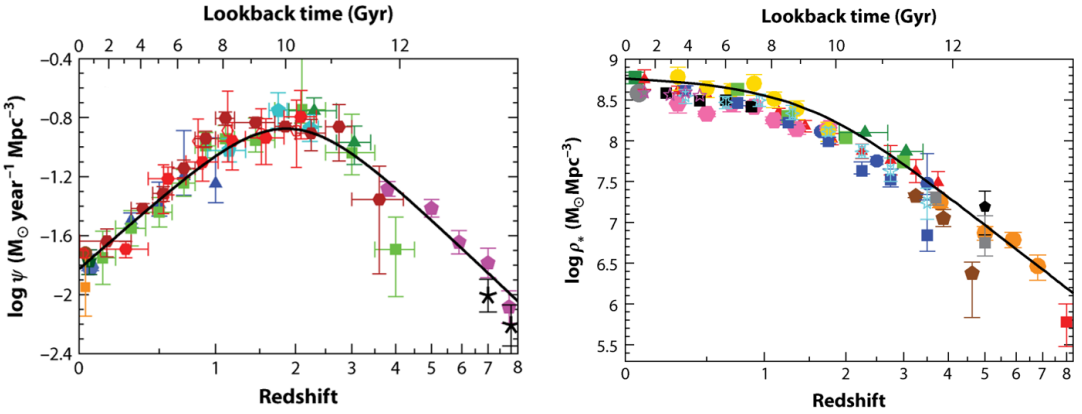


Figure 2.6. – The evolution of the cosmic SFR and stellar mass density as obtained from measurements of the UV and infrared emission in galaxies. The different coloured symbols represent different observations. *Adapted from Figures 9 and 11 of Madau & Dickinson 2014.*

studies have surveyed well-defined regions or fields of the sky with sufficient *depth*[16] to encompass a cosmologically representative volume.

By measuring the stellar light from thousands of galaxies within these fields, such surveys have quantified the star formation rate per unit comoving volume – the star formation rate density (SFRD) – across cosmic time, reaching out to redshift $z \sim 8$ (see Madau & Dickinson, 2014, for a review). Together, they have established a consistent picture where the SFRD rose steadily as $\propto (1+z)^{-2.9}$ from $z \sim 8$ to $z \sim 2$, before declining at a rate of $\propto (1+z)^{2.7}$, decreasing by roughly an order of magnitude between $z = 2$ and $z = 0$ (as shown in Figure 2.6; Madau & Dickinson, 2014). The SFRD peaked at $1.5 \lesssim z \lesssim 2.5$, a period often referred to as the *cosmic noon*.

It is then only natural to ask *what fuelled the higher star formation at* $z \sim 2$ *and why did it decay further?* Was it driven by an evolution of the gas content of galaxies, particularly molecular gas that serves as the fuel for star formation? Or was there an evolution in the efficiency with which galaxies convert their gas into stars? To answer these questions, we need to quantify the amount of molecular gas in large samples of galaxies at different epochs.

Until the mid-2010s, observations of the gas content of galaxies were largely restricted to the low-redshift Universe (Scoville et al., 1995, 1997) and a handful of bright, but rare galaxies at $z \gtrsim 1$ (Tacconi et al. 2010, Daddi et al. 2010a, Daddi et al. 2010b; also see Carilli & Walter 2013, Tacconi et al. 2020, and Förster Schreiber & Wuyts 2020 for a review). The advent of interferometers like the Atacama Large Millimeter Array (ALMA), the Northern Extended Millimeter Array (NOEMA), and the Karl G. Jansky Very Large Array (VLA) has revolutionized the field. The enhanced sensitivity and wide frequency coverage of these instruments have enabled astronomers to study the gas distribution in typical star-forming galaxies that dominate the cosmic star formation budget at any given epoch, but are much fainter (see Hodge & da Cunha, 2020, for a recent review).

[16] Note that here depth refers to the redshift interval Δz across which a given feature of interest (such as line or continuum emission) can be mapped. It is determined by the frequency coverage of a given survey.

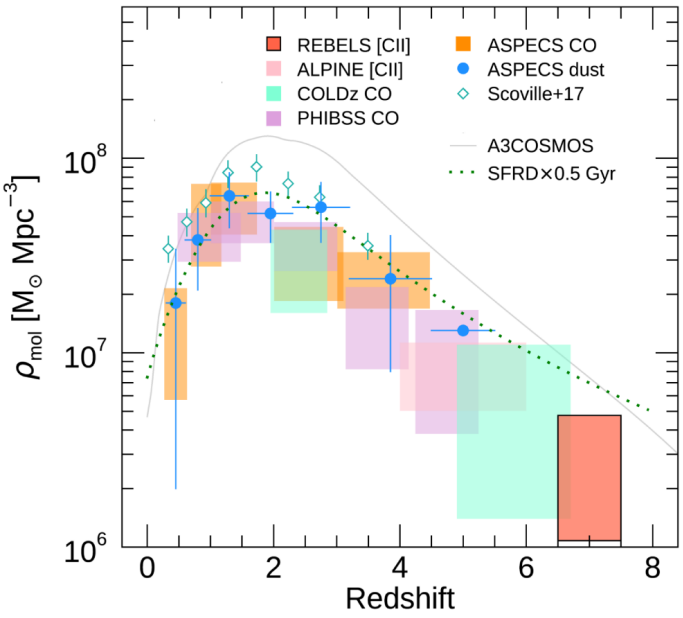


Figure 2.7. – Evolution of the cosmic H_2 density obtained from a variety of targeted (A3COSMOS, PHIBSS, Scoville et al. 2017, ALPINE, REBELS) and blind (COLDz, ASPECS) surveys from $z \sim 7$ down to the present day. The dotted curve represents the cosmic SFR density (left panel of Figure 4.2), scaled by a typical depletion timescale of 0.5 Gyr. (*Adapted from Figure 5 in Aravena et al. 2024.*)

In Section 2.4, we discussed various tracers of molecular gas in galaxies. Over the past decade, numerous studies using one or more of these tracers have been conducted to map the global H_2 content across different cosmic epochs. These studies can be broadly categorized as:

1. **Targeted observations** that focus on estimating the molecular gas mass in pre-selected galaxies that have been identified in optical and infrared wavelengths. Examples include:
   a) studies of CO rotational lines such as the ALMA public archive in the COSMOS deep field (A3COSMOS) survey (Liu et al., 2019) and the Plateau de Bure High-z Blue Sequence Survey (PHIBSS Lenkić et al., 2020);
   b) observations of dust continuum emission such as Scoville et al. (2016, 2017); Magnelli et al. (2020);
   c) surveys targeting the [C II] line such as the ALMA Large Program to Investigate C^+ at Early Times (Faisst et al., 2022, ALPINE;) at $4.4 \leq z \leq 5.9$ and the ALMA Reionization Era Bright Emission Line Survey (ALMA-REBELS; Bouwens et al., 2022) at $6 \leq z \leq 9$.

2. **Blind surveys** that perform a blind search of molecular gas in a given region of the sky. As a result, these can detect all molecular gas sources down to a sensitivity limit. Examples of blind surveys of CO lines and dust continuum include the ALMA Spectroscopic Survey in the HUDF (ASPECS, Walter et al., 2014; Decarli et al.,

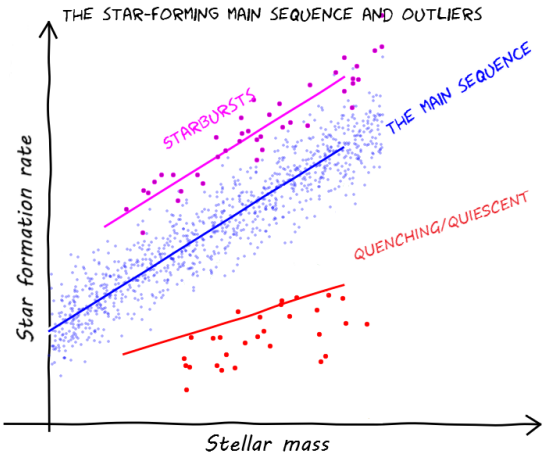


Figure 2.8. – A schematic representation of the main-sequence of star-forming galaxies (Source).

2014, 2016; Walter et al., 2020) and the VLA CO Luminosity Density at High Redshift (COLDz, Riechers et al., 2019).

3. **Intensity mapping surveys** that measure the collective emission from all galaxies within a survey volume without the need to resolve individual galaxies. Intensity mapping can, in principle, detect galaxies fainter than the sensitivity limits of targeted observations. Examples of CO intensity mapping experiments include the CO Power Spectrum Survey-II (COPSS-II; Keating et al., 2016), the Millimeter-wave Intensity Mapping Experiment (mmIME; Keating et al., 2020), and the CO Mapping Array Project (COMAP; Chung et al., 2024).

A compilation of ρ_{H_2} constraints from these observations is shown in the left panel of Figure 2.7. Despite large uncertainties in the various estimates, a clear redshift evolution is evident: the cosmic H_2 density increases over time, reaches a broad peak at $z \sim 1.5$ (Tacconi et al., 2020) and decays towards lower redshifts, in a similar manner as the SFRD.

2.6.1. The Main Sequence of star-forming galaxies

Observations support that the cosmic SFR activity is dominated by star-forming galaxies (SFGs) that grow secularly, forming a tight correlation in the SFR-stellar mass (M_*) plane. This tight correlation is referred to as the "main sequence" (MS; see e.g., Brinchmann et al., 2004; Daddi et al., 2007; Elbaz et al., 2007; Noeske et al., 2007; Speagle et al., 2014; Whitaker et al., 2014; Schreiber et al., 2015; Pearson et al., 2018; Popesso et al., 2023) and has a scatter of 0.2-0.3 dex, that remains largely constant across stellar mass and redshift (Speagle et al., 2014). Observations estimate that $\sim 90\%$ ($\sim 68\%$) of the cosmic star formation from redshift $z \sim 2.5$ ($z \sim 4$) to $z = 0$ takes place in MS galaxies (Rodighiero et al., 2011, 2015; Schreiber et al., 2015). Therefore, these are popular targets for understanding cosmic trends.

The distance from the MS

$$\Delta(\mathrm{MS}) \equiv \log_{10}\,[\mathrm{SFR}\,/\,\mathrm{SFR}(\mathrm{MS},\,M_*,\,z)] \tag{2.35}$$

quantifies the offset of a galaxy from the MS at its redshift. Galaxies with $-0.3 \leq \Delta(\mathrm{MS}) \leq 0.3$ are considered to be MS galaxies.

Galaxies above the MS (i.e., $\Delta(\mathrm{MS}) > 0.3$) exhibit significantly elevated SFRs at a given stellar mass and are known as "starbursts". Because of their intense star formation, these are often much brighter than MS galaxies and were among the firsts to be detected at high redshifts (Frayer et al., 1998, 1999). These also contain copious amounts of dust, which absorbs most of their stellar light and re-radiates it at longer submillimeter wavelengths in the infrared. For this reason, they are also referred to as submillimeter galaxies (Blain et al. 2002; also see Casey et al. 2014 for a recent review). Starbursts contribute roughly 10% of the cosmic SFRD. Galaxies below the main-sequence are referred to as quiescent or quenching, depending on whether or not star formation has completely ceased.

2.7. Simulating galaxy formation

Numerical simulations of galaxy formation model a representative volume of the Universe in a cosmological context, tracking the formation and growth of dark matter halos and the galaxies they host. Simulating galaxy formation is a complex multi-scale problem in astrophysics and involves modelling various physical processes. The large-scale structure of the cosmic web determines where and when halos of a given mass form, which itself is dictated by the location and amplitude of the density peaks in the initial conditions and their subsequent evolution. Gas flows from the cosmic web to the center of halos, where some of this gas cools down to form neutral atomic hydrogen, which further cools down to form H_2. Clouds of molecular gas collapse and fragment to give way to star formation. Throughout their lifecycles, stars inject energy, mass, radiation, and momentum into their surroundings. This process enriches the ISM with metals, many of which act as coolants. The basic physical processes and interactions that are common to all galaxy formation simulations are gravity, hydrodynamics, gas cooling and heating, star formation and stellar feedback, although the exact mode (e.g., thermal, kinetic, radiative) of stellar feedback varies. Some simulations also include black hole growth, AGN feedback, chemical evolution, radiative transfer, magnetic fields and cosmic ray transport. Note that cosmological simulations generally adopt a Newtonian prescription for gravity as it provides an excellent approximation to GR in the limit of low velocities induced by the growth of structures. The expansion of the Universe itself, however, is modelled by following the Friedmann equations (Equations (2.5) and (2.6)).

Dark-matter-only simulations (also known as N-body simulations) model the evolution of the dark-matter density field due to gravitational interactions alone. The density field is approximated by discrete macroscopic dark matter particles[17] that are treated as collisionless. The gravitational force on a given particle is computed by summing over the forces from all other particles which is then used to advance the particle's velocity and subsequently its position. Over time, the particles cluster together, resulting in the growth of the initial overdensities in the field. By following the evolution of dark matter in an expanding Universe, N-body simulations achieve a high accuracy in capturing the gravitational interactions and structure formation of dark matter (see Kuhlen et al., 2012a; Angulo & Hahn, 2022, for a review).

[17]Note that the term 'particle' here does not refer to the fundamental dark-matter particle.

Hydrodynamical simulations, on the other hand, follow the evolution of both dark matter and baryonic components (see Vogelsberger et al. 2020 and Crain & van de Voort 2023 for a review). At the start of the simulation, all baryons are in the form of pristine gas i.e., hydrogen and helium. As the simulated (patch of the) Universe evolves, some of this gas is converted into stars. These simulations then additionally follow the equations of motion of star particles that are treated as collisionless particles. Note that in cosmological simulations, each star particle collectively represents a stellar population.[18] The gas component is treated as an inviscid ideal fluid following Euler equations, whose exact formulation depends on the frame of reference adopted i.e., fluid-frame (Lagrangian framework) or coordinate-frame (Eulerian framework).

For an inviscid ideal fluid with density ρ, pressure P and velocity $\mathbf{v}$, and total energy per unit mass $e = u + \mathbf{v}^2/2$, where u is the internal energy of the fluid, the Euler equations in Eulerian form can be written as

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho\, \mathbf{v}) = 0\,; \quad \frac{\partial (\rho \mathbf{v})}{\partial t} + \nabla \cdot (\rho \mathbf{v} \otimes \mathbf{v}) = -\nabla P + \rho \mathbf{g}\,; \quad \frac{\partial (\rho e)}{\partial t} + \nabla \cdot (\rho e + P)\mathbf{v} = \rho \mathbf{v} \cdot \mathbf{g}\,. \tag{2.36}$$

In the Lagrangian form, these become:

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} = -\rho \nabla \cdot \mathbf{v}\,; \quad \frac{\mathrm{D}\mathbf{v}}{\mathrm{D}t} = -\frac{1}{\rho} \nabla P + \mathbf{g}\,; \quad \frac{\mathrm{D}e}{\mathrm{D}t} = -\frac{1}{\rho} \nabla \cdot (P\mathbf{v}) + \mathbf{v} \cdot \mathbf{g}\,, \tag{2.37}$$

where the Lagrangian or convective derivative $\mathrm{D}/\mathrm{D}t \equiv \partial/\partial t + \mathbf{v} \cdot \nabla$.

These two frames of reference have led to different discretization schemes for solving the Euler equations numerically. As a result, hydro codes can broadly be divided into mesh- or grid-based (Eulerian) and fluid-particle-based (Lagrangian):

1. **Eulerian methods:** These methods discretize the space into a grid of finite-volume cells and follow the motion of the fluid across the grid. Eulerian hydro codes define the fluid properties and their spatial derivatives at the cell level and compute the advection of these properties across the cell boundaries. Owing to the large dynamic range that needs to be simulated, the grid can be adaptively refined to attain a higher spatial resolution in regions of interest (e.g., high-density regions): grid cells that satisfy the refinement criteria are further divided into smaller cells (i.e., cells at a higher refinement level), thereby increasing the spatial resolution. This technique is known as adaptive mesh refinement (AMR). In principle, AMR codes like Ramses (Teyssier, 2002) can achieve arbitrarily high spatial resolution by progressively increasing the levels of refinement in the region of interest. This is a key advantage of mesh-based codes. In practice, this is limited by the high computational cost involved.

2. **Lagrangian methods:** Instead of following the evolution of a finite volume, these methods discretize the fluid into small elements and follow the motion of the individual fluid elements through space. Smooth particle hydrodynamics (SPH; see

[18] A stellar population is a collection of stars that are coeval i.e., born at the same time and therefore, have the same chemical composition or metallicity. The metallicity significantly influences the spectrum of light emitted by the stellar population. The stars have different masses and while it is impossible to directly probe the distribution of these masses, a theoretical concept called the stellar initial mass function (IMF) is used to describe the distribution. The IMF is sensitive to the metallicity and temperature of the star-forming gas.

Springel, 2010a, for a review) is the most widely used Lagrangian method, where the continuum fluid is approximated by discrete particles. Physical properties are computed by smoothing or smearing out the particle properties within a smoothing kernel that has a characteristic length scale called the smoothing length (denoted by h): typically, high-(particle)-density regions have a small h while low-(particle)-density regions have a larger h. Because of their Lagrangian nature, SPH codes like GADGET-2 (Springel, 2005) can naturally follow the mass flow at a high resolution.

Another class of hydro codes combine the natural adaptivity of SPH and the better-resolving capabilities of AMR. For instance, the moving-mesh code Arepo (Springel, 2010b) employs an unstructured mesh defined by the Voronoi tessellation of a set of discrete points that are allowed to move freely (hence the name 'moving-mesh').

Sub-grid recipes: The smallest unit of operation in a grid-based code is a grid cell while that in an SPH code is a gas particle. Therefore, the minimum grid cell size or the smoothing length defines the smallest spatial scale that a simulation can resolve. Similarly, the mass of the dark matter, stellar, or gas particle decides the respective mass resolution. However, quite often, sub-grid models are employed to approximate the processes that occur below the resolution scale of the simulation. These recipes are either theoretically/empirically- motivated or informed by high-resolution simulations. For instance, most galaxy formation simulations do not have sufficient mass or spatial resolution to model the gravitational collapse of gas into stars. Instead, they employ a recipe to convert their gas into stars based on the empirical Schmidt-Kennicutt relation (section 2.2.4).

This thesis focuses on modelling molecular gas chemistry within cosmological simulations of galaxy formation. In the following sections, we will review the numerous advances made in this field over the last two decades, that serve as the foundation for this thesis.

2.7.1. H_2 chemistry in galaxy simulations

H_2 primarily forms on the surface of dust grains. At typical ISM densities, this process is several orders of magnitude more efficient than three-body reactions (Glover, 2003). However, in the extremely metal-poor ($Z \lesssim 10^{-5}$ and therefore dust-poor) environments of the early Universe, three-body reactions are suspected to have been the dominant pathway for H_2 formation (Christensen et al., 2012; Lenoble et al., 2024).

The main destruction channels for H_2 are UV radiation and cosmic rays. Photodissociation is carried out by Lyman-Werner photons with energies in the range 11.2-13.6 eV (i.e., $\lambda = 912 - 1108$ Å), and occurs through a two-step process known as spontaneous radiative dissociation. First, an H_2 molecule absorbs a photon with energy > 11.2 eV and is excited to a higher electronic state. On de-excitation, it either returns to a level in rotational-vibrational continuum of the ground electronic state or (in $\approx 15\%$ of the cases Draine & Bertoldi, 1996), returns to one of the vibrational levels that form a continuum. Since the energy of these levels is higher than the dissociation energy of H_2 (≈ 4.5 eV), the molecule is dissociated once in the vibrational continuum. At temperatures $T \gtrsim 5000$ K, H_2 can also be destroyed by collisions with other species.

Capturing these complex molecular processes in simulations is challenging, as it requires resolving both the local ISM microphysics governing H_2 formation and destruction, as well

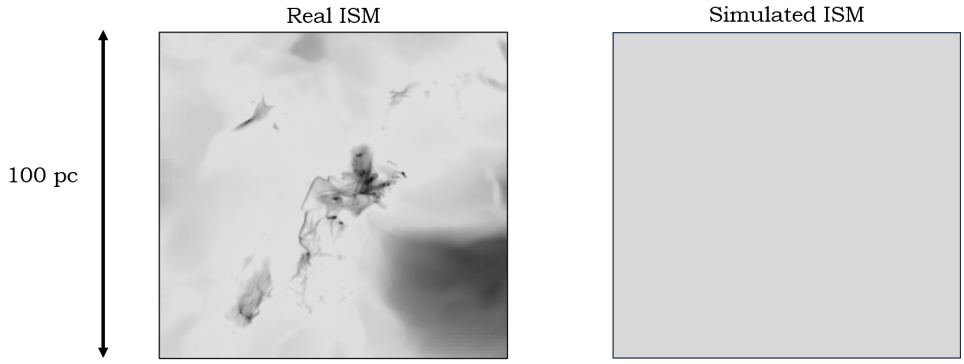


Figure 2.9. – A schematic showing the clumpy ISM missed by the finite resolution of numerical simulations. The left panel is adapted from SILCC-Zoom simulations (Seifried et al., 2017).

as the larger-scale galaxy environment that influences the overall gas content of galaxies. On large scales, galaxies are affected by gas accretion from and outflows into the cosmic web, while on smaller scales, molecular gas chemistry is shaped by local conditions such as metallicity, dust abundance, density structure, turbulence, and radiation field strength. Modelling these small-scale processes from first principles is beyond the resolving capabilities of cosmological simulations needed to investigate the evolution of the cosmic molecular gas budget.

For instance, the left panel of Figure 2.9 presents a 100 pc slice from the SILCC-Zoom simulations (Seifried et al., 2017), which have a spatial resolution of 0.06 pc. In contrast, the right panel illustrates how the same region of the ISM would appear in a low-resolution simulation with a minimum grid-cell size of 100 pc. Such low-resolution simulations assume a uniform distribution of properties like density and temperature across the entire slice, failing to capture the intricate density structure seen on the left. These variations are critical since many chemical reactions, including H_2 formation, are collisional processes whose rates are enhanced in the presence of dense substructures. Ignoring this complexity in low-resolution simulations results in an underestimation of the H_2 formation rate, which necessitates the use of sub-grid prescriptions to account for unresolved density fluctuations.

Pelupessy et al. (2006) were among the firsts to devise such a prescription and incorporate the molecular gas phase in galaxy simulations. They implemented a time-dependent scheme for H_2 formation on dust grains and destruction by UV radiation from young stars, coupled with a sub-grid scheme to approximate gas cells as gas clouds with a radial density profile and that follow the empirical density-size scaling relation (from Larson, 1981; Elmegreen, 1989). An alternative to the sub-grid gas clouds of Pelupessy et al. (2006) was proposed by Gnedin et al. (2009) who enhanced the H_2 formation rate by an effective *'clumping factor'*

$$C = \langle n_{\rm H}^2 \rangle / \langle n_{\rm H} \rangle^2 \,, \tag{2.38}$$

where the brackets denote a spatial average. The clumping factor quantifies the degree of density fluctuations within a given region with mean density $\langle n_{\rm H} \rangle$. The basic premise for using C is: *By artificially boosting the* H_2 *formation rate, one can bypass the need to explicitly resolve the density fluctuations below the resolution scale.* Gnedin et al. (2009) argued based on observations and simulations of turbulent molecular clouds that C mostly

lies in the range 3-10. Some later studies (e.g., Lupi et al., 2018) also implemented a clumping factor varying with the local Mach number. The technique was further extended to SPH simulations by Christensen et al. (2012).

Another class of H_2 models that gained attention early on were two analytical approximations for the H_2 fraction as a function of the gas surface density, UV field strength, and metallicity, proposed in a series of papers by Krumholz et al. There are two varieties of these – 'KMT-EQ' (Krumholz et al., 2009) that gives the equilibrium H_2 fraction in a region with a given gas density and metallicity and is independent of the UV field strength by construction; and 'KMT-UV' (Krumholz, 2013), that calculates the H_2 fraction assuming a local chemical equilibrium between H_2 formation and destruction by UV radiation. Krumholz & Gnedin (2011) compared the KMT-EQ model against the Gnedin et al. (2009) model and found that the former worked reasonably well for gas with metallicity higher than 1% solar ($Z \gtrsim 10^{-2}\,Z_\odot$).

The KMT models have found extensive application in cosmological simulations (Fu et al., 2010; Lagos et al., 2011; Kuhlen et al., 2012b, 2013; Krumholz & Dekel, 2012; Hopkins et al., 2014; Thompson et al., 2014; Lagos et al., 2015; Davé et al., 2016). However, despite their wide use, these equilibrium models do not account for the dynamic nature of the ISM and the long formation timescale for H_2 (Tielens & Hollenbach, 1985). This is particularly relevant for the early Universe, where the dust content of the ISM is lower than in present-day galaxies. This results in an H_2 formation timescale that can be longer than the dynamical timescale on which typical molecular clouds are fully disrupted or their physical conditions are drastically altered.

Therefore, several authors have investigated the effect of non-equilibrium H_2 chemistry on the properties of simulated galaxies. For example, Pelupessy & Papadopoulos (2009) investigated the effect of non-equilibrium H_2 chemistry using galaxy-scale simulations with H_2-based star formation. They found that the majority of the gas (mass) exhibits an H_2 fraction different from the equilibrium prediction. They noted that gravitational collapse and change in ISM conditions happen on much smaller timescales compared to H_2 formation and destruction, resulting in an out-of-equilibrium H_2 abundance. Richings & Schaye (2016) also found a significant effect of non-equilibrium H_2 chemistry on the mass and chemical composition of outflows in their simulated galaxies. Similarly, Pallottini et al. (2017) found a better agreement of their simulated galaxy with the observed Schmidt-Kennicutt relation when using non-equilibrium H_2 chemistry. Similar findings were also reported by Schäbe et al. (2020), and Hu et al. (2021).

These studies demonstrate that assuming steady-state chemistry can give a biased picture of the chemical composition of the ISM as well as the global properties of simulated galaxies. Hydrodynamical simulations with on-the-fly computations of non-equilibrium chemical abundances are rare so far (Dobbs et al., 2008; Pelupessy & Papadopoulos, 2009; Gnedin & Kravtsov, 2010, 2011; Micic et al., 2012; Christensen et al., 2012; Tomassetti et al., 2015; Lupi et al., 2018; Lupi, 2019; Schäbe et al., 2020; Katz et al., 2022; Hu et al., 2023) and often restricted to simulations of individual galaxies.

2.7.2. Carbon chemistry in galaxy simulations

Beyond the challenges outlined in the previous section, modelling carbon chemistry within numerical simulations is further complicated by the sheer volume of the network,

both in terms of the number of species involved and the complexity of the reactions. Several authors have therefore attempted to restrict the chemical networks to retain only the dominant formation and destruction pathways for the species of interest (commonly CO and C^+). The first among these were (Nelson & Langer, 1997, hereafter NL97) who assumed a fast conversion of C^+ directly into CO via the formation of the hydrocarbon radical CH_2^+, thereby effectively bypassing the need to include atomic carbon in the network.

$$C^+ + H_2 \rightarrow CH_2^+ + \gamma\,; \tag{2.39}$$

$$CH_2^+ + O \rightarrow CO + H\,. \tag{2.40}$$

The NL97 network was later modified by (Nelson & Langer, 1999, hereafter NL99) to track the abundance of atomic carbon in addition to CO and C^+, as well as include other channels for CO formation. In total, the NL99 network included 23 chemical reactions. A much more extended network comprising 218 chemical reactions among 32 species was developed by Glover et al. (2010) and they were among the first to include a treatment of hydrogen and carbon chemistry in an ISM simulation. Glover & Clark (2012b) performed a detailed comparison of several chemical networks and found that the time-dependent CO abundance predicted by the Glover et al. (2010) network is similar to (although slightly lower than) that from the NL99 network. A key difference between NL99 and Glover et al. (2010) networks is that the former use pseudo species CH_x and OH_x to collectively refer to a group of molecules, molecular ions, radicals containing only C & H and O & H i.e., OH_x stands for O_2, H_2O, OH, OH^+, H_2O^+, H_2O^+ and CH_x stands for CH, CH_2, CH^+, CH_2^+, on account of their similar reactions and associated reaction rates. Such simplification was also recently adopted by Gong et al. (2017) in their chemical network retaining 50 chemical reactions involving 18 species. Despite the varying levels of complexity of the networks discussed above, all of these are far simpler than the extensive chemical networks employed in one-dimensional PDR models (see Röllig et al., 2007, for an example of commonly used PDR codes in the literature), that provide the standard benchmark to test the robustness of the chemical predictions from simplified networks.

Unfortunately, even the reduced networks are far from ideal for cosmological simulations. For instance, in the ISM simulation of a cubic box of side length 20 pc by Glover et al. (2010), the authors reported that their simplified chemical network consumes $\sim 90\%$ of the total computational time. The overhead can be even more pronounced for large-scale simulations that follow the evolution of hundreds or a few thousands of galaxies, if not more. Solving for chemical abundances on the fly in cosmological simulations therefore requires further simplifications that might lead to a trade-off between accuracy and computational time. Chapter 3 of this thesis deals with these simplifications to obtain a reasonably accurate network for hydrogen and carbon chemistry that can be incorporated within cosmological simulations without a large computational overhead.

CHAPTER 3

HYACINTH: A sub-grid model for hydrogen and carbon chemistry in cosmological simulations

HYACINTH: HYdrogen And Carbon chemistry in the INTerstellar medium in Hydro simulations[1]

Prachi Khatri, Cristiano Porciani, Emilio Romano-Díaz, Daniel Seifried, and Alexander Schäbe

2024 Astronomy & Astrophysics, 688, A194, 21 pp.
(DOI: 10.1051/0004-6361/202449640)

Overview

Modelling the molecular gas content of galaxies is challenging as molecular gas chemistry is regulated by conditions on sub-parsec scales, which are beyond the resolving capabilities of current cosmological simulations needed to investigate the evolution of molecular gas on cosmic scales. To tackle this multi-scale problem, we have developed a sub-grid model called HYACINTH – *HYdrogen And Carbon chemistry in the INTerstellar medium in Hydro simulations* – that circumvents this challenge by a employing physically-motivated prescription for capturing the effects of the 'microscopic' (i.e., unresolved) density structure on the 'macroscopic' (i.e., resolved) chemical abundances in cosmological simulations. In this chapter, we describe the various components of the model and test it against common approaches in the literature. We also present an application of this model as a post-processing tool for hydrodynamical simulations.

[1]The published paper is reproduced in its entirety in Appendix B.

3.1. The components of HYACINTH

The basic framework of HYACINTH comprises:

1. a variable sub-grid density PDF to mimic the effect of gravitational collapse and pre-SN feedback on the density structure within molecular clouds (section 3.1.1);
2. a metallicity-dependent temperature-density relation based on high-resolution[2] molecular cloud simulations (from Hu et al., 2021);
3. a simplified chemical network for hydrogen and carbon chemistry within molecular clouds (section 3.1.2);
4. a metallicity-dependent dust-to-gas mass ratio (from Péroux & Howk, 2020);
5. a variable cosmic ray ionization rate, $\zeta_{\rm H}$. The default mode assumes $\zeta_{\rm H} \propto \chi^2$ (following Kosenko et al., 2021), but other options are available as well;
6. the UV flux in the Lyman Werner bands in Habing units.

3.1.1. The sub-grid density PDF

Most of the molecular gas in the ISM resides within giant molecular clouds that exhibit a highly inhomogeneous density distribution. The density fluctuations within a molecular cloud can be statistically described by a probability distribution function. For instance, a mass-weighted PDF $\mathcal{P}_{\rm M}(n_{\rm H})$ of sub-grid densities $n_{\rm H}$ gives the fraction of the total mass (within a computational volume e.g., a grid cell in a mesh-based simulation) present at densities between $[n_{\rm H}, n_{\rm H} + {\rm d}n_{\rm H}]$. The shape of this density PDF is governed by the interplay between turbulence and self-gravity.

In an isothermal, turbulent medium, not significantly affected by the self-gravity of gas, $\mathcal{P}_{\rm M}$ is expected to take a log-normal shape (see e.g. Vázquez-Semadeni, 1994; Passot & Vázquez-Semadeni, 1998; McKee & Ostriker, 2007, for a review). Conversely, in regions where self-gravity is important, dense clumps form and the PDF develops an extended tail at high densities. In such regions, the PDF remains log-normal at low densities with a power-law tail at high densities.

For a log-normal (hereafter LN) distribution, the mass-weighted PDF is given by

$$\mathcal{P}_{\rm M}(n_{\rm H}) = \frac{1}{\sqrt{2\pi}\sigma n_{\rm H}} \exp\left[-\frac{(\ln n_{\rm H} - \mu)^2}{2\sigma^2}\right] , \tag{3.1}$$

where σ and μ are parameters that decide the width and the location of the peak of the distribution. The parameter μ is related to the mean density $\langle n_{\rm H}\rangle$ in the region as $\mu = \ln\langle n_{\rm H}\rangle + \dfrac{\sigma^2}{2}$. The parameter σ is related to the clumping factor C (equation 2.38) as $\sigma = \sqrt{\ln C}$. Following previous studies (Gnedin et al., 2009; Christensen et al., 2012; Tomassetti et al., 2015), we used a fixed clumping factor of 10.

[2]These simulations have a particle mass resolution of 1 $M_\odot$ and a spatial resolution ∼0.2 pc.

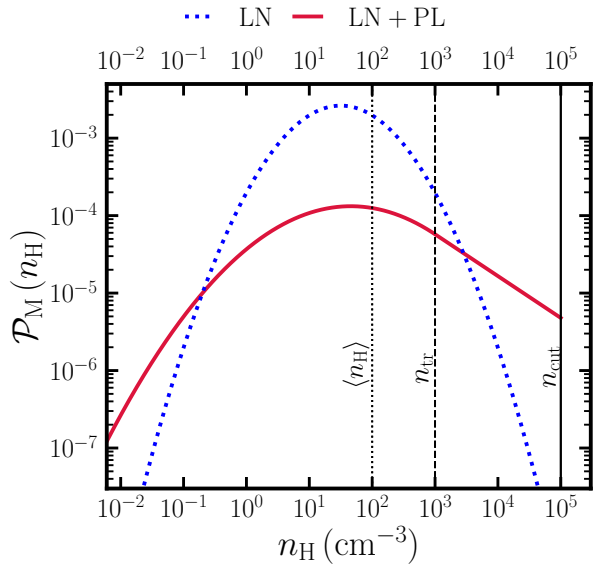


Figure 3.1. – The two mass-weighted probability distribution function (PDF) used in this study. Panel (a) shows the mass-weighted PDFs in sample simulation cells as a function of the sub-grid density $n_{\rm H}$, where $\mathcal{P}_{\rm M}(n_{\rm H})\,{\rm d}n_{\rm H}$ denotes the fraction of the total cell mass present at sub-grid densities in the range $[n_{\rm H}, n_{\rm H} + {\rm d}n_{\rm H}]$. The log-normal (LN; equation 3.1) PDF is shown in blue and the log-normal+power-law (LN+PL, equation 3.2) PDF is shown in red. The sample cells have a mean hydrogen density $\langle n_{\rm H} \rangle = 100\,{\rm cm}^{-3}$ (shown by the dotted black line). For the log-normal+power-law PDF, the transition density $n_{\rm tr}$ and the cut-off density $n_{\rm cut}$ are shown by the dashed and solid black lines, respectively.This figure is a modified version of Figure 1 in Khatri et al. (2024a).

For a log-normal+power-law (hereafter LN+PL) distribution of $n_{\rm H}$, the PDF is given by

$$\mathcal{P}_{\rm M}(n_{\rm H}) = \begin{cases} \dfrac{Q_1}{n_{\rm H}} \exp\left[-\dfrac{(\ln n_{\rm H} - \mu_2)^2}{2\sigma_2^2}\right], & \text{if } n_{\rm H} \leq n_{\rm tr} \\ Q_2 \left(\dfrac{n_{\rm H}}{n_{\rm tr}}\right)^{\alpha}, & \text{if } n_{\rm tr} < n_{\rm H} \leq n_{\rm cut} \\ 0, & \text{if } n_{\rm H} > n_{\rm cut}, \end{cases} \tag{3.2}$$

where $\alpha < 0$ is the slope of the power law and $n_{\rm tr}$ is the density at which the power-law tail begins. The parameters μ_2 and σ_2 characterise the location of the peak and the width of the log-normal part of the PDF. These are calculated, along with constants Q_1 and Q_2, for a given $\langle n_{\rm H} \rangle$, α, and $n_{\rm tr}$ to match the mean density to $\langle n_{\rm H} \rangle$ and ensure the continuity, differentiability, and normalization of the PDF. We use $\alpha = -0.54$ in this study based on the analytical model of a spherically collapsing cloud by Girichidis et al. (2014). This value is also consistent with the range of power-law slopes observed for star-forming clouds (Kainulainen et al., 2009) and in simulations (Kritsuk et al., 2011). We set $n_{\rm tr}$ equal to 10 times the mean density $\langle n_{\rm H} \rangle$ (Kritsuk et al., 2011) and impose a cut-off of $n_{\rm cut} = 1000\,\langle n_{\rm H} \rangle$ above which the PDF is set to zero, to prevent the integral of the PDF from diverging. An example of the two PDFs in sample grid cells with mean density $\langle n_{\rm H} \rangle = 100\,{\rm cm}^{-3}$ is shown in Figure 3.1.

Species	Standard HYACINTH	Extended HYACINTH	Treatment
H_2	✓	✓	solve ODE
CO	✓	✓	solve ODE
C^+	✓	✓	solve ODE
C	✓	✓	conservation of C nuclei
H_3^+	✓	✓	local equilibrium
CH_x	✓	✓	local equilibrium
OH_x	✓	✓	local equilibrium
He^+	✗	✓	solve ODE
HCO^+	✗	✓	solve ODE

Table 3.1. – Table showing the complete list of chemical species and their treatment in standard and extended HYACINTH chemical networks.

3.1.2. The chemical network

The chemical network in HYACINTH is derived from the widely used Nelson & Langer (1999, hereafter NL99) network for hydrogen and carbon chemistry with modifications adopted from the recent work of Gong et al. (2017, hereafter G17). We have limited the number of chemical species and reactions to include only the most dominant formation and destruction channels for H_2, CO, C, and C^+ under the physical conditions prevalent in molecular clouds. Overall our standard network comprises 19 chemical reactions. The included chemical species can be divided into main species and helper species. The distinction lies in how these species are treatment in the network. While we explicitly solve the chemical rate equations for the main species, the helper species are included to assist in accurately predicting the abundance of the main species. We assume "local" (i.e., at each sub-grid density) equilibrium for the helper species.

We test the performance of our standard chemical network against an extended version that includes two additional species, namely He^+ and HCO^+ as these are expected to play an important role in CO chemistry. We include a total of 10 additional reactions involving these species in our extended chemical network. The chemical reactions included in both versions are listed in Table A1 of Khatri et al. 2024a. Table 3.1 shows a list of all species and how these are treated in the two versions of HYACINTH.

We assume that all hydrogen in a cell is either atomic or molecular. This assumption holds very well in molecular-cloud regions, where numerical simulations (e.g., Glover & Mac Low, 2007; Hu et al., 2021) have shown that the temperature $T \lesssim 200\,\mathrm{K}$ and therefore the contribution of ionized hydrogen to the total hydrogen budget is negligible. Thus, the total hydrogen density in a cell can be written as

$$\langle n_{\mathrm{H}} \rangle = \langle n_{\mathrm{HI}} \rangle + 2 \langle n_{\mathrm{H_2}} \rangle \,, \tag{3.3}$$

where $\langle ... \rangle$ denote cell-averaged quantities. The mean H_2 fraction within the cell is defined as

$$f_{\mathrm{H_2}} = \frac{2 \langle n_{\mathrm{H_2}} \rangle}{\langle n_{\mathrm{H}} \rangle} \tag{3.4}$$

In a numerical simulation, it is highly impractical to record the chemical abundances at each sub-grid density within every cell. Therefore, we make a further assumption that in any given cell, hydrogen becomes fully molecular above a critical density $n_{\rm crit,\,H_2}$. This sharp H I $\rightarrow$ H_2 transition arises because of the self-shielding nature of H_2 and is supported by various numerical (e.g., Dobbs et al., 2008; Krumholz et al., 2008, 2009; Gnedin et al., 2009) and observational (e.g., Savage et al., 1977; Tumlinson et al., 2002; Schruba et al., 2011; Shull et al., 2021) studies. As a result, equation (3.3) can be written as

$$\langle n_{\rm H} \rangle = \underbrace{\langle n_{\rm H} \rangle \int_0^{n_{\rm crit,H_2}} \mathcal{P}_{\rm M}\, {\rm d}n_{\rm H}}_{\langle n_{\rm H_I} \rangle} + \underbrace{\langle n_{\rm H} \rangle \int_{n_{\rm crit,H_2}}^{\infty} \mathcal{P}_{\rm M}\, {\rm d}n_{\rm H}}_{2\langle n_{\rm H_2} \rangle} \tag{3.5}$$

Likewise, we assume that all carbon is in ionized form at sub-grid densities below $n_{\rm crit,\,C\,I}$, turns to atomic above this density and becomes fully molecular above sub-grid densities $n_{\rm crit,\,CO}$. All reaction rates are computed at the sub-grid level and the cell-level rates are obtained by integrating over the (sub-grid) density PDF. These calculations are presented in full detail in Appendix A of Khatri et al. (2024a).

3.2. Tests of chemical abundances

Given our simplifications, it is instructive to evaluate the performance of our model against previous approaches in the literature.

3.2.1. Comparison with other chemical networks

First, we perform a 'benchmark' test wherein we compare the chemical predictions from Hyacinth (standard and extended) against other approaches in the literature, namely two (relatively) extensive chemical networks – the NL99 and G17 networks – and the PDR code used in G17 which tracks the abundances of 74 species accounting for 322 chemical reactions and includes a more sophisticated treatment of radiative transfer. This PDR code is derived from Tielens & Hollenbach (1985) and updated by Wolfire et al. (2010), Hollenbach et al. (2012), and Neufeld & Wolfire (2016). For this test, we employ a one-dimensional semi-infinite slab irradiated from one side by a UV field of strength 1 in Draine units (i.e., $\chi = 1$). The slab has a uniform hydrogen number density $n_{\rm H} = 1000\,{\rm cm^{-3}}$ throughout. We show in Figure 3.2 the equilibrium abundances of each species as a function of the visual extinction A_V (equation 2.17). Here we show the case for a CRIR of $1 \times 10^{-17}\,{\rm s^{-1}\,H^{-1}}$. In the manuscript, we also compare for a higher CRIR of $2 \times 10^{-16}\,{\rm s^{-1}\,H^{-1}}$.

Our main findings are:

(a) The A_V-dependence of $f_{\rm H_2}$ from all the different approaches is very similar, indicating that hydrogen chemistry is rather insensitive to the exact treatment of carbon chemistry.

(b) The C $\rightarrow$ CO transition in Hyacinth occurs at a similar A_V as in G17 and the PDR code.

(c) at $A_V \lesssim 0.5$, the C abundance in all other approaches is roughly an order of magnitude higher than that in NL99. This is a consequence of additional destruction

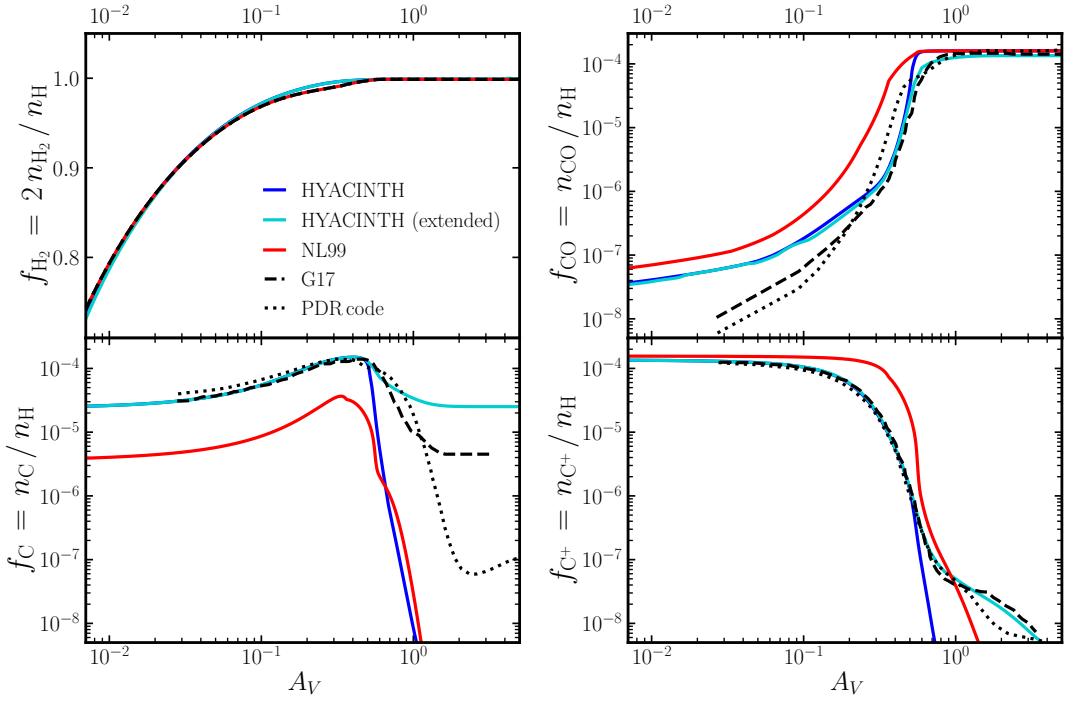


Figure 3.2. – Comparison of the chemical abundances from HYACINTH with those from NL99 and G17 networks. The abundances of H_2, CO, C, and C^+ as a function of the visual extinction A_V in a semi-infinite plane-parallel slab are shown in different panels. The blue, turquoise, and red lines represent the results from HYACINTH, extended HYACINTH, and NL99 respectively. Here extended HYACINTH refers to the HYACINTH network with additional chemical reactions for He^+ and HCO^+ that are not part of standard HYACINTH (see text for more details). The dashed and dotted black lines show, respectively, the abundances from the chemical network and the PDR code in G17. The slab has a uniform hydrogen density $n_H = 1000\,cm^{-3}$, solar metallicity and solar dust abundance, a CRIR of $10^{-17}\,s^{-1}\,H^{-1}$ and is illuminated from one side by a UV field of strength $\chi = 1$. This figure is a modified version of Figure 2 from Khatri et al. (2024a)

channels. In standard HYACINTH these are i) grain-assisted recombination of C^+, ii) an additional outcome for the $C^+ + H_2$ reaction.

(d) In the A_V range where each of the carbon species dominates, i.e., C^+ at $A_V \lesssim 0.1$, C at $0.1 \lesssim A_V \lesssim 1$, and CO at $A_V \gtrsim 1$, the predictions from (standard) HYACINTH agree very well with G17 and the PDR code.

(e) At $A_V \gtrsim 2$ the CO abundance in standard HYACINTH differs from that in extended HYACINTH by 16%, thereby showing that the improvement contributed by 10 additional chemical reactions in extended HYACINTH is not commensurate with the ~ 3.3 higher computational time with respect to standard HYACINTH.[3]

[3]This is computed for the slab test performed here.

3.2.2. Comparison with molecular-cloud simulations

Now we evaluate the performance of the chemical network in conjunction with the sub-grid density PDF. To do so, we compare the full time evolution of the chemical abundances from HYACINTH employed as a sub-grid model within a hydrodynamical simulation against high-resolution molecular-cloud simulations. Specifically, we use two simulations: the SILCC-Zoom simulations (Seifried et al., 2017, 2020) at solar metallicity and the Glover & Mac Low (2011, hereafter GML11) simulations at $Z = 0.1\,Z_\odot$.[4]

For this comparison, we adopt a grid cell size of 25 pc, which is similar to the highest spatial resolution achieved in the cosmological simulations described in Chapter 4. These comparisons are shown in Figures 3 and 4 of Khatri et al. (2024a) and are described in detail in section 3.2 of Khatri et al. (2024a). Overall, we find that our LN+PL shows a better agreement with the SILCC-Zoom simulations while our LN PDF performs better in case of the GML11 simulations. These findings are consistent with the fact that the two simulations exhibit different PDFs – the PDF in GML runs closely matches a log-normal (see their Figure 1) while the SILCC-Zoom simulations have a much more prominent high-density tail as the simulated molecular cloud is gravitational collapsing (see Figure 20 in Walch et al. 2015).

3.3. Application: HYACINTH as a post-processing tool

Now we turn our attention to an immediate application of HYACINTH as a post-processing tool. Although designed as a sub-grid model to be fully integrated within simulations, here we it apply to a pre-simulated galaxy at redshift $z \sim 2.5$ from Tomassetti et al. (2015, hereafter T15). The T15 simulation was performed with a modified version of the AMR code RAMSES and includes a chemistry module for computing the H_2 abundance dynamically within the simulation. The H_2 chemistry in HYACINTH and T15 differ in two ways: (a) HYACINTH includes additional channels for H_2 formation and destruction; (b) unlike T15, HYACINTH uses a metallicity-dependent temperature-density relation and dust-to-gas ratio (see 3.1).

For this application, we only use HYACINTH with the LN PDF (as in T15) for a fair comparison with the dynamically-evolved f_{H_2} from T15. We further compare with two analytical approximations– namely 'KMT-EQ' (Krumholz et al., 2009) and 'KMT-UV' (Krumholz, 2013). In KMT-EQ, f_{H_2} is computed based on the metallicity and total gas column density, independent of the strength of the UV radiation, while KMT-UV accounts for the effect of UV radiation.

Figure 3.3 shows a comparison of the median f_{H_2} (equation 3.4) obtained from different approaches. We find that the f_{H_2} from HYACINTH is sensitive to the assumed cosmic ray ionization rate (CRIR, denoted by ζ_H), particularly at low densities. In Figure 3.3, we show results from three different CRIR laws:

1. *black solid line:* using a fixed CRIR equal to the Milky-Way value of of $3 \times 10^{-17}\,s^{-1}\,H^{-1}$.

[4]Note that neither the SILCC-Zoom nor the GML11 runs account for star formation and stellar feedback.

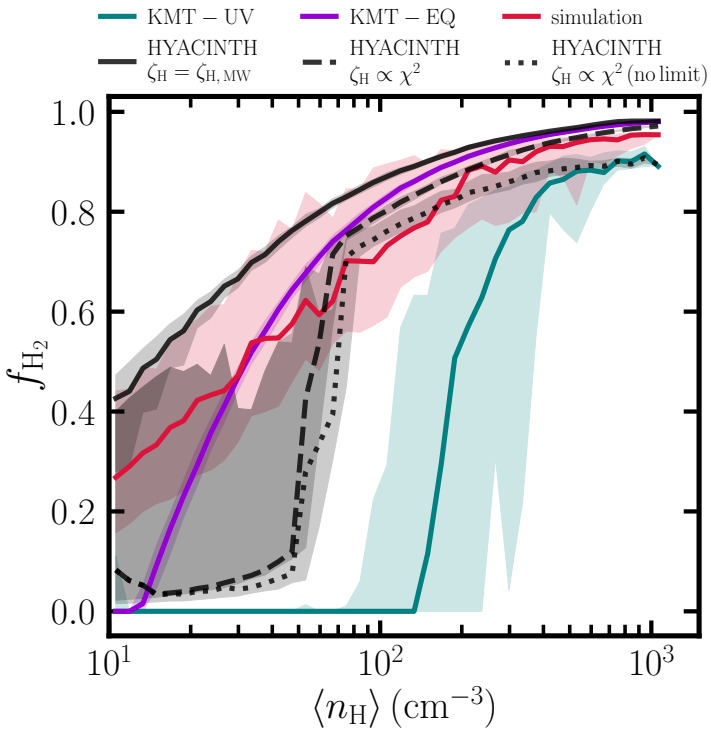


Figure 3.3. – Comparison of the H_2 fraction ($f_{H_2} = 2\langle n_{H_2}\rangle/\langle n_H\rangle$) as a function of $\langle n_H\rangle$ using different approaches. The equilibrium f_{H_2} from post-processing the T15 galaxy with HYACINTH (black) compared with the f_{H_2} from the simulation (red) and two analytical estimates – KMT-EQ (purple) and KMT-UV (green). The lines show the median value in a given $\langle n_H\rangle$ bin while the shaded areas enclose the 16th to 84th percentiles. The solid black line denotes the median f_{H_2} when using a uniform CRIR of $\zeta_H = 3 \times 10^{-17}\,s^{-1}\,H^{-1}$ in HYACINTH. We also show the median f_{H_2} when using $\zeta_H \propto \chi^2$) with (dashed black line) and without (dotted black line) the upper limit of $3\times10^{-14}\,s^{-1}\,H^{-1}$ on the CRIR This figure is a modified version of Figure 6 in Khatri et al. (2024a).

2. *black dashed line:* using a CRIR that scales quadratically with the UV flux ($\zeta_H \propto \chi^2$), with an upper limit on $\zeta_H = 3 \times 10^{-14}\,s^{-1}\,H^{-1}$. This is the default option in HYACINTH.

3. *black dotted line:* same as in (b) but without the upper bound.

The effect of using a quadratic CRIR is the strongest at low densities $\langle n_H\rangle \lesssim 100\,cm^{-3}$. Conversely, removing the upper bound on the CRIR reduces the f_{H_2} at $\langle n_H\rangle \gtrsim 100\,cm^{-3}$ by $\sim$ 10% and has no effect on the lower densities. This is simply because with the quadratic scaling between ζ_H and χ, the χ in the low-density cells is not sufficiently high to exceed the CRIR upper bound of 3×10^{-14}. Nevertheless we note that the bulk ($\gtrsim$ 80%) of the H_2 in this simulated galaxy is present at densities $\langle n_H\rangle \gtrsim 100\,cm^{-3}$. Therefore, these differences would have minimal consequences on the total H_2 mass of galaxies. We enlist in Table 3.2, the masses of the chemical species in the post-processed galaxy for different choices of the CRIR.

Method	$\zeta_{\rm H}$	$\mathbf{M}_{\rm H_2}$ ($10^{10}\,{\rm M}_\odot$)	$\mathbf{M}_{\rm CO}$ ($10^{7}\,{\rm M}_\odot$)	$\mathbf{M}_{\rm C\,I}$ ($10^{7}\,{\rm M}_\odot$)	$\mathbf{M}_{\rm C^+}$ ($10^{7}\,{\rm M}_\odot$)
HYACINTH	$\zeta_{\rm H,\,MW}$	4.70	9.54	0.63	0.62
HYACINTH	default	4.12	9.38	0.66	0.66
HYACINTH	$\propto \chi^2$ (no ceiling)	3.84	6.52	0.90	1.64

Table 3.2. – Total mass of the different chemical species in the post-processed galaxy for different choices of the CRIR. For reference, the (dynamically evolved) $M_{\rm H_2}$ from the simulation is $4.21 \times 10^{10}\,{\rm M}_\odot$ (From Table D2 in Khatri et al. 2024a.)

3.4. Conclusions

In this work, we presented a sub-grid model called HYACINTH, that can be embedded into cosmological simulations for evolving the non-equilibrium abundances of H_2 and its carbon-based tracers, namely CO, C, and C^+. HYACINTH comprises a variable sub-grid density PDF to capture the unresolved density structure in simulations and a simplified chemical network for hydrogen and carbon chemistry. These simplifications were introduced to make the network highly efficient for use in large-scale simulations.

We compared HYACINTH against more sophisticated approaches in the literature for modelling hydrogen and carbon chemistry including two extensive chemical networks (NL99 and G17) and a PDR code, using a one-dimensional semi-infinite slab setup. HYACINTH reproduced the $f_{\rm H_2}$-A_V relation from these methods highlighting that H_2 chemistry is insensitive to the exact treatment of carbon chemistry. Moreover, despite its simplicity and size, HYACINTH captured the $C^+ \rightarrow C$ and $C \rightarrow CO$ transitions very well, as predicted by more complex approaches. A comparison with high-resolution ISM simulations shows reasonable agreement in chemical abundances, further supporting the robustness of HYACINTH.

CHAPTER 4

The MARIGOLD suite: molecular gas in galaxies across cosmic time

Overview

In the previous chapter, we presented a new sub-grid model called HYACINTH for following hydrogen and carbon chemistry on the fly in hydrodynamical simulations. We have embedded HYACINTH into the adaptive mesh refinement (AMR) code RAMSES (Teyssier, 2002). This enables us to study the evolution of the molecular gas content in galaxies across cosmic time. Using our updated version of RAMSES, we have performed a suite of cosmological simulations called the MARIGOLD simulations. This chapter describes the physical and chemical processes included in these simulations and presents the global properties of the simulated galaxies at $z \geq 3$. The primary goal of this chapter is to describe the physics included in these simulations and demonstrate that the simulated galaxy population resembles the observed galaxy population. This sets the stage for further applications of these simulations to make predictions and aid the interpretation of observations.

4.1. Numerical Methods

Our simulation suite comprises two hydrodynamical simulations – a $(25\,\mathrm{Mpc})^3$ comoving volume (denoted as `M25`) and a $(50\,\mathrm{Mpc})^3$ comoving volume (denoted as `M50`). These simulations have different mass and spatial resolutions, chosen to collectively probe a wide dynamical range of galaxy masses. The specifications of these simulations are provided in Table 4.1.

We adopt the flat ΛCDM model with Planck cosmology (Planck Collaboration et al., 2020), i.e., $\Omega_\Lambda = 0.6847$, $\Omega_\mathrm{m} = 0.3153$, $\Omega_\mathrm{b} = 0.0493$, $\sigma_8 = 0.8111$, $n_\mathrm{s} = 0.9649$, and $H_0 = 67.36\mathrm{km\,s^{-1}\,Mpc}$. The simulations are started from uni-grid initial conditions (ICs) set at $z = 99$ generated with the code MUSIC (Hahn & Abel, 2011). The ICs have an initial refinement level $l_\mathrm{ini} = 10$ corresponding to 1024^3 grid cells and an equal number of

Simulation	$L_{\rm box}$ (cMpc)	$N_{\rm DM}$	$\ell_{\rm initial}$	$\ell_{\rm final}$	$\Delta x^{\rm min}$ (pc)	$m_{\rm DM}$ ($\rm M_\odot$)	m_* ($\rm M_\odot$)	$m_{\rm gas}^{\rm ini}$ ($\rm M_\odot$)
`M25`	25	1024^3	10	17	32	5.0×10^5	7.2×10^3	9.3×10^4
`M50`	50	1024^3	10	17	64	4.0×10^6	5.8×10^4	7.4×10^5

Table 4.1. – Specifications of the MARIGOLD simulation suite. From left to right, the columns list: the name of the simulation, the comoving box size, the number of dark-matter (DM) particles, the initial and final refinement levels, the minimum cell size achieved in the simulation in physical units, the DM and stellar particle masses, and the average gas mass per grid cell in the initial conditions.

dark matter particles. We allow for seven additional levels of refinement (i.e. $\ell_{\rm final} = 17$) between the ICs and the final redshift $z = 3$, which results in a maximum spatial resolution $\Delta x^{\rm min}$ of 32 pc and 64 pc, respectively, for the `M25` and `M50` runs. The gas is modelled using an equation of state with polytropic index $\gamma = 5/3$. To avoid spurious fragmentation, we add an artificial pressure support to cells at the highest refinement level as commonly done in mesh-based simulations (e.g., Robertson & Kravtsov, 2008; Kuhlen et al., 2012b). This artificial pressure mimics the pressure support from turbulent motions on scales below the resolution. The simulation volumes have periodic boundary conditions and the dynamical evolution of dark matter, gas, and stars is tracked with a modified version of RAMSES. These modifications are described in the following subsections.

4.1.1. Molecular gas chemistry

At every time step in the simulation, the abundances of H_2, CO, C, and C^+ in each gas cell are evolved with the sub-grid model HYACINTH (Chapter 3). HYACINTH requires six input parameters – the average density of hydrogen nuclei $\langle n_{\rm H} \rangle$ in the cell, the gas-phase metallicity Z, the UV flux in Lyman-Werner (LW) bands in Habing units G_0 (section 2.4), the cell size Δx, the density PDF $\mathcal{P}_{\rm M}$, and the time step Δt. Depending on the state of star formation in a given cell, $\mathcal{P}_{\rm M}$ can take one of the two functional forms presented in section 3.1.1. We describe how this is done in practice in section 4.1.3. HYACINTH then assigns a temperature to each sub-grid density in the PDF based on a metallicity-dependent temperature-density relation from Hu et al. (2021). The chemical rate equations are solved at each sub-grid density and the cell-averaged chemical abundances are obtained by integrating over $\mathcal{P}_{\rm M}$. The initial chemical abundances are chosen such that all hydrogen and carbon are in atomic form.

4.1.2. Metal enrichment by unresolved early star formation

Population III stars are the first generation of stars that form from metal-free gas in the early Universe. These first stars explode as supernovae and release metals into their surrounding gas, enriching the ISM for subsequent star formation. As metal-free gas can only cool via atomic and H_2 cooling, this sets a lower limit on the halo masses where the cooling is efficient and gas can collapse to form stars. This limit is redshift-dependent and at $30 \lesssim z \lesssim 50$ (i.e., when the Universe was $\sim 50-100$ Myr old) corresponds to halo masses in the range of $\sim 10^5 - 10^7\,\rm M_\odot$ (e.g., Tegmark et al., 1997). The DM particle mass in our simulations ($\sim 0.5-4\times10^6\,\rm M_\odot$) is too high to resolve the host haloes of Population III stars

at these redshifts as well as some of the early population II star formation.[1] Therefore, to mimic the metal enrichment of the ISM by these early episodes of star formation, we impose an initial metallicity floor of $10^{-3}\,Z_\odot$ and $10^{-2}\,Z_\odot$ for our `M25` and `M50` runs, respectively. This is a common approach adopted in cosmological simulations that do not resolve the formation of the first stars (see e.g., Kuhlen et al., 2012b; Tomassetti et al., 2015; Pallottini et al., 2017). It is supported by high-resolution simulations that resolve the formation of and metal enrichment by population III and population II stars in the first few hundred Myr of the Universe. (Wise et al., 2012; O'Shea et al., 2015; Smith et al., 2018). Introducing a metallicity floor is particularly crucial for modelling H_2 chemistry as H_2 formation predominantly happens on the surface of dust grains at the densities we resolve in our simulations. Note that we adopt a higher metallicity floor for the `M50` run compared to the `M25` run as the latter has higher mass and spatial resolutions that can resolve the star formation as early as $z \sim 21$, while in the case of `M50`, this can only happen at $z \sim 18$. Despite the deceptively small time (≈ 30 Myr) elapsed between these redshifts, it is well known that supernovae explosions are capable of enriching their surrounding gas to $10^{-3} - 10^{-1.5}\,\mathrm{Z}_\odot$ metallicities, depending on the intensity of early star formation (see e.g., Figures 4 and 6 in Wise et al., 2012, and the discussion therein).

4.1.3. Star formation & stellar feedback

As commonly done in cosmological simulations, the star formation recipe is derived from the Schmidt-Kennicutt relation. We adopt H_2-based star formation wherein the local star formation rate (SFR) in a grid cell is proportional to the molecular gas density:

$$\langle \rho_{\rm SFR} \rangle \;=\; \epsilon_{\rm sf}\,\frac{\langle \rho_{\rm mol} \rangle}{t_{\rm ff}}\,, \tag{4.1}$$

where the brackets indicate a spatial average over the grid cell, $t_{\rm ff} = \sqrt{3\,\pi/32\,G_{\rm N}\,\langle \rho_{\rm gas} \rangle}$ is the local free-fall time of the gas, and $\epsilon_{\rm sf}$ is the star formation efficiency (SFE). The SFE represents the fraction of molecular gas converted into stars within a free-fall time. Based on observations of Milky-Way molecular clouds, Krumholz & Tan (2007) showed that for over four orders of magnitude in molecular gas densities, the local SFE is approximately 1-2 %. As the sizes of our highest refinement cells ($\approx$ 32-128 pc) are comparable to the sizes of molecular clouds (section 2.2.2), we adopt a constant $\epsilon_{\rm sf} = 1\%$ in both simulations. We do not impose a density threshold for star formation: all cells with sufficient H_2 can and do form stars.

At every timestep $\Delta\,t$,[2] the available molecular gas mass in a cell is used to form N equal mass collisionless stellar particles, where N is sampled from a Poisson distribution with mean $\lambda = \langle \rho_{\rm SFR} \rangle\,(\Delta\,x)^3\,\Delta\,t/m_*$, where m_* is the stellar particle mass (see Table 4.1). However, to ensure numerical stability, at any given time, no more than 90% of the molecular gas mass in the cell is allowed to turn into stars. Each stellar particle thus formed represents a simple stellar population sampled following a Kroupa IMF (Kroupa, 2001). The initial metallicity of the stellar particle is set equal to the metallicity of the gas from

[1] Population II stars form from gas enriched by supernovae of population III stars. This gas is metal-poor but not devoid of metals.

[2] Note that Ramses adopts an adaptive timestep scheme wherein $\Delta\,t$ varies with the refinement level of the cell.

which it is formed. The mass of molecular gas that goes into stars is reduced from both the molecular and total gas masses within the cell and the molecular gas fraction $f_{\rm H_2}$ is recomputed.

We include stellar feedback in the form of thermal energy from supernova (SN) and the metal enrichment from stellar evolution based on the STARBURST99 templates with a Kroupa IMF and solar metallicity yields. We model thermal feedback by assuming that each stellar particle explodes as a Type II SN approximately 10 Myr after its birth and releases $E_{\rm SN} = 10^{50}$ erg per unit solar mass of thermal energy into the parent grid cell. Previous studies have found that grid-based numerical simulations do not retain this thermal energy (see e.g. Stinson et al., 2006; Scannapieco et al., 2012; Agertz et al., 2013) and therefore, a common workaround is to adopt a delayed cooling mechanism. In our simulations, the cooling of the gas is delayed for 40 Myr.

Several recent studies have shown that pre-SN feedback in the form of stellar winds, photoionization, and radiation pressure starts acting as soon as stars form and is efficient at dispersing the gas within molecular clouds (see e.g., Chevance et al., 2020). Although we do not explicitly model these processes in our simulations, we account for their effect on the sub-grid density structure of our gas cells by changing the functional form of $\mathcal{P}_{\rm M}$ from a log-normal+power-law (LN+PL) to a log-normal (LN) distribution. The functional forms and a schematic representation of these PDFs can be found in section 3.1.1. For a given mean density, the LN+PL PDF has a higher mass fraction at high densities, representing dense clumps within molecular clouds. As these clumps are dispersed by pre-SN feedback, transitioning from the LN+PL to the LN PDF mimics this dispersal. The formation of a new cloud and subsequent development of high-density clumps are modelled by switching back to the LN+PL PDF after approximately 40 Myr. This timescale accounts for both the molecular cloud dispersal and the assembly of the next cloud. We note that this duration is fixed in our model, while in the real ISM, individual molecular clouds evolve and disperse at different rates.

4.1.4. Heating & Cooling

We adopt the standard prescriptions for cooling and heating in RAMSES that comprise the following cooling processes: collisional ionization, radiative recombination, dielectric recombination, bremsstrahlung, and inverse Compton cooling. Additionally we account for metal line cooling and far-infrared fine-structure line cooling based on the prescription described in Teyssier (2002). The heating processes include radiative and Compton heating. Our simulations also include a uniform cosmic UV background as implemented in Courty & Alimi (2004). We approximate gas shielding against UV heating and ionization by assuming that all cells with hydrogen density above $0.013\,{\rm cm}^{-3}$ are well-sheilded (Tajiri & Umemura, 1998).

4.1.5. Lyman-Werner radiation

The UV radiation in the Lyman-Werner (LW) bands from each stellar particle is modelled based on its age using stellar templates from the STARBURST99 library (Leitherer et al., 1999). This UV radiation is then propagated to nearby cells following the radiative transfer recipe of Tomassetti et al. (2015). Briefly, this recipe calculates the LW UV flux in each gas cell by first calculating the LW luminosity of each stellar particle in the simulation

as a function of its age using the STARBURST99 templates (Leitherer et al., 1999). It is assumed that each stellar particle represents a stellar population where the stellar masses follow the Kroupa IMF. The LW flux (G) in each gas cell is then computed by considering all stellar particles within an octet, i.e., 8 cells at the highest refinement level. The LW flux at the location $\vec{x}$ of a grid cell at time t can then be written as:

$$G(\vec{x}, t) = \frac{\alpha}{G_0} \frac{\Sigma_{i \in \mathrm{oct}} \, L_{\mathrm{LW}}(t - t_{s,i})}{4\pi \, (\Delta \, x)^2 \Delta\lambda \, c} \tag{4.2}$$

where the flux is expressed in Habing units (Habing, 1968, section 2.2.1). The LW luminosity of stellar particle i (in $\mathrm{erg\,s^{-1}}$) at a time $t - t_{s,i}$ after its birth (at $t = t_{s,i}$) is denoted by L_{LW}; α is a geometrical correction factor t correct the flux at the grid cell centre in case of a uniform distribution of stars within the octet.

4.2. Simulations

Using a modified version of RAMSES with physical prescriptions described in the previous section, the simulations are started from the ICs at $z = 99$ and run down to $z = 3$. In this section we describe how galaxies are identified from the simulation snapshots and how we combine the galaxies from the two simulations.

4.2.1. Halo and galaxy identification

We use the Amiga Halo Finder (AHF; Knollmann & Knebe, 2009) to identify haloes and subhaloes in each simulation snapshot. AHF identifies haloes by locating density peaks within the simulation and then iteratively determining the gravitationally bound particles that constitute each peak. Each resulting halo is a spherical region with virial radius R_{vir} and a mean matter density (i.e., including dark matter, gas, and stars) equal to 200 times $\overline{\rho}(t)$, the average matter density in the Universe at the cosmic time t of the snapshot. The virial mass of the halo can be written as $M_{\mathrm{vir}} = \frac{4}{3}\pi R_{\mathrm{vir}}^3 \; 200 \; \overline{\rho}(t)$, where the masses and sizes of haloes are calculated accounting for unbinding. These are referred to as 'main haloes' in the following. The comoving number density of haloes as a function of their mass, commonly referred to as the halo mass function (HMF), is shown in Figure 4.1 for the two simulations at redshifts 5, 4, and 3. We find an excellent agreement between the `M25` and `M50` HMFs down to halo mass of $10^9\,\mathrm{M}_\odot$. Because of the different mass resolution of the two simulations, the `M50` HMF starts to decrease below this mass, while the `M25` HMF continues to grow.

Subhaloes are defined as gravitationally bound objects within main haloes and lying within common isodensity contours of the host halo. We impose that every halo is resolved with at least a 100 particles. Galaxies are defined in terms of their parent halo. For main haloes, the stellar concentration at their centre is referred to as the main or central galaxy.

For each main galaxy, we start with a spherical region of size 0.1 R_{vir} and calculate the stellar half-mass radius $r_{1/2,*}$ (that is, the radius containing half of the stellar mass within $0.1R_{\mathrm{vir}}$). The galaxy is defined in terms of $2r_{1/2,*}$ and all (galaxy-integrated) quantities are measured within this radius. Conversely, the stellar concentration residing at the centre of a subhalo is called a satellite galaxy whose size is defined by the radius corresponding

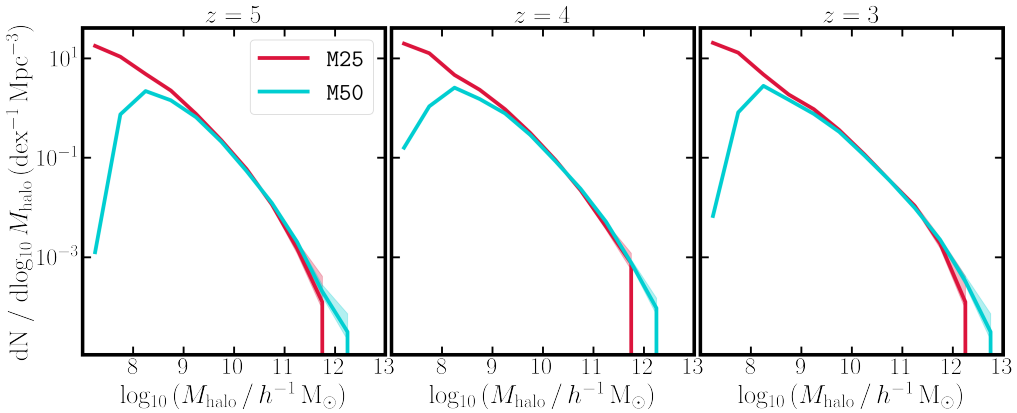


Figure 4.1. – The (main) halo mass functions from the `M25` (red) and `M50` (turquoise) simulations at different redshifts. The shaded areas represent the lower and upper Poisson uncertainties on the number counts in each mass bin at 16% confidence limit from Gehrels (1986).

to the maximum of the subhalo rotation curve, $R_{V_{\rm max}}$ (Klypin et al., 2011; Prada et al., 2012). In other words, $R_{\rm V_{max}}$ sets the boundary of a satellite galaxy.

4.2.2. Dynamical relaxation and galaxy selection

Two-body relaxation is a dynamical phenomenon wherein collisions between the constituent "particles" of a system become frequent enough that the system "evaporates" on a timescale shorter than the age of the Universe. Numerical simulations treat the constituent stellar particles in a system (like a galaxy or dark matter halo) as collisionless particles. However, when the number density of stars in a given region becomes too high, collisions become frequent and one can no longer neglect the effect of these collisions on the redistribution of energy within the system that leads to the eventual evaporation of the system on timescales smaller than the age of the Universe. Since numerical simulations do not account for these effects, they might contain some spurious galaxies/haloes that would not be present in case of a proper treatment of collisional dynamics. We identify such objects in our simulations by evaluating their relaxation time as:

$$t_{\rm relax} = \frac{N}{8\,\ln(N)}\,t_{\rm cross} = \frac{N}{8\,\ln(N)}\left(\frac{R^3}{G\,M}\right)^{1/2}, \tag{4.3}$$

where N is the number of stellar particles enclosed in a sphere of radius $R = 2\,r_{1/2,*}$ and M is the total mass of these stars. We assume that the galaxies are approximately in virial equilibrium such that the typical velocity of the stellar particles can be expressed as

$$\sigma_{\rm v} = \sqrt{\frac{GM}{R}}\,. \tag{4.4}$$

Equation (4.3) yields an order of magnitude approximate value for the relaxation time such that a system of particles can be considered as collisionless if $t_{\rm relax} >> t_{\rm age}$, where $t_{\rm age}$ is the age of the Universe at the galaxy's redshift. In our sample, we consider all galaxies

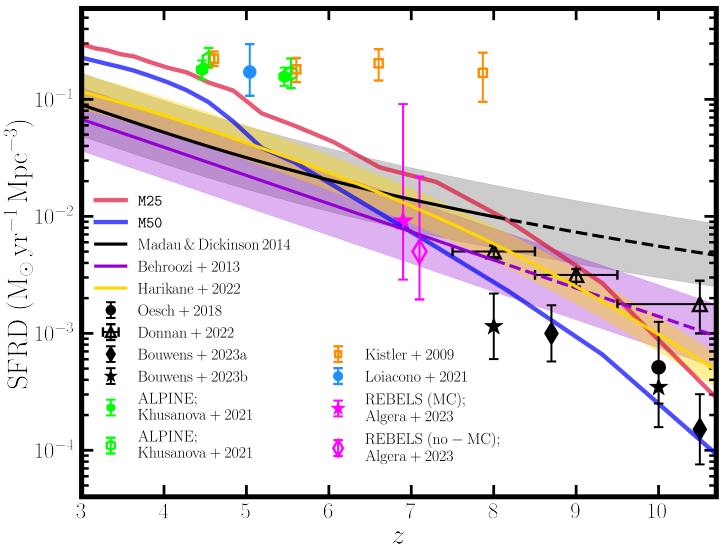


Figure 4.2. – The evolution of the simulated cosmic SFRD from $z \sim 10$ to $z = 3$ in the `M25` (red) and `M50` (blue) simulations. The black, purple, and yellow lines show the best fits from Madau & Dickinson (2014), Behroozi et al. (2013), and Harikane et al. (2022a). The shaded areas denote the uncertainty on the respective fits. Dashed lines indicate extrapolation beyond the redshift range constrained by the measurements. The dust-corrected SFRD measurements based on UV observations from Oesch et al. (2018, black circle), Donnan et al. (2022, black triangle), Bouwens et al. (2023b, black diamond) and Bouwens et al. (2023a, black star) are shown as black symbols with error bars. The SFRD determined by Khusanova et al. (2021) at $z = 4.5$ and 5.5 based on ALPINE galaxies are shown as green symbols, for two different methods of determining the obscured SFR fraction – using stellar mass (open symbols) or FUV magnitudes (filled symbols) as a proxy for the IR luminosity. These symbols have been slightly shifted from $z = 4.5$ and $z = 5.5$ for the sake of clarity. The estimates from REBELS (Algera et al., 2023) are shown as magenta symbols, wherein the 'MC' estimate (filled star) accounts for uncertainty on the stellar mass while deriving the SFRD, while the 'no-MC' estimate (open diamond) follows the standard approach of keeping stellar mass fixed. The constraints from Kistler et al. (2009) using gamma-ray bursts at $z > 4$ is also shown as orange open squares and those from the [C II] luminosity function at $z \sim 5$ from ALPINE (Loiacono et al., 2021) is shown as a light blue circle. Note that the Kistler et al. (2009) estimates are shifted by +0.1 at all redshifts for the sake of clarity.

with $t_{\rm relax} < 10\ t_{\rm age}$ to be affected by two-body relaxation. Then we group our galaxies into logarithmic stellar mass bins of 0.5 dex and discard all galaxies within a bin where more than 10% of the galaxies suffer from two-body relaxation. This results in a stellar mass cut of $10^{7.5}\,{\rm M}_\odot$ at all redshifts for the `M25` simulation and that of $10^{8}\,{\rm M}_\odot$ at $z \geq 5$ and $10^{8.5}\,{\rm M}_\odot$ at lower redshifts for the `M50` simulation.

4.3. Cosmic trends

As discussed in Section 2.6, the time evolution of the cosmic star formation rate density (SFRD) and cosmic H_2 density are of key importance in the study of galaxy formation and evolution. Here, we examine their evolution in the MARIGOLD simulations.

4.3.1. The cosmic star formation rate density

The evolution of the cosmic SFR density (SFRD) is the most prominent tool to study galaxy evolution and an important benchmark for all cosmological galaxy formation models. Figure 4.2 shows the evolution of the simulated SFRD compared with observational estimates. These include the best fit from Madau & Dickinson (2014) based on a compilation of SFR measurements from UV and IR observations at $0 \leq z \leq 8$; the scaling from Behroozi et al. (2013) following a comprehensive analysis of galaxy stellar mass function in the same redshift range; and the estimate from Harikane et al. (2022a) based on UV measurements at $z \gtrsim 8$ as well as literature data at lower redshifts. Remarkably, the Harikane et al. (2022a) SFRD differs significantly from the extrapolated Behroozi et al. (2013) and Madau & Dickinson (2014) best fits at $z \geq 8$, suggesting a steeper decline in the SFRD at high redshifts. We also show individual UV-based estimates at specific redshift intervals from Oesch et al. (2018), Donnan et al. (2022) ,Bouwens et al. (2023b), and Bouwens et al. (2023a).

The simulated SFRD includes the contribution from all stars within the simulation volume. The shape of the simulated SFRD is consistent with the Madau & Dickinson (2014), Behroozi et al. (2013), and Harikane et al. (2022a) estimates and is well within the uncertainty of these relations in the redshift range $z \gtrsim 5$. At lower redshifts, however, our simulated SFRD is consistently higher than the Madau & Dickinson (2014) estimate. At $z = 3$, the SFRD in the `M50` (`M25`) run is $\approx 1.3\sigma$ ($\approx 1.7\sigma$) higher than Madau & Dickinson (2014).

Nevertheless, we find a good agreement with recent estimates from Khusanova et al. (2021) based on ALPINE galaxies at $4.5 \leq z \leq 5.9$. Till date, this is the largest sample of galaxies at these redshifts with multiwavelength observations of galaxies targeting both their UV and IR emission, thereby providing an estimate of the total (dust-obscured + unobscured) star formation in main-sequence galaxies at these redshifts. The authors further note that at $z \sim 4.5$ and $z \sim 5.5$, the obscured star formation accounts for $68^{+18}_{-25}\%$ and $61^{+20}_{-25}\%$, respectively, of the total star formation at these redshifts. Their results indicate that the original compilation by Madau & Dickinson (2014) underestimates the correction due to dust-obscured SFRD at $z \gtrsim 4$ (also see Rowan-Robinson et al., 2016; Gruppioni et al., 2020).

At $z \gtrsim 8$, our simulated SFRD shows a steeper decline compared to the to the extrapolated Behroozi et al. (2013) and Madau & Dickinson (2014) relations, and similar to the Harikane et al. (2022a) relation. Overall, the SFRD predicted from our simulations shows a general agreement with observational estimates at specific redshifts. It is worth noting that the latter exhibit large uncertainties at $z \geq 4$, primarily because of the difficulty involved in estimating the dust-obscured SFR and the small sample of galaxies observed at these high redshifts.

At $5 \leq z \leq 9$, our simulated SFRD[3] is well described by:

$$\rho_{\rm SFR} \propto 10^{-0.43\,(1+z)}\,, \tag{4.5}$$

while at $z < 5$, we obtain:

$$\rho_{\rm SFR} \propto 10^{-0.36\,(1+z)}\,. \tag{4.6}$$

[3]This scaling is obtaining using the SFRD(z) from the `M50` simulation because of its eight times larger volume.

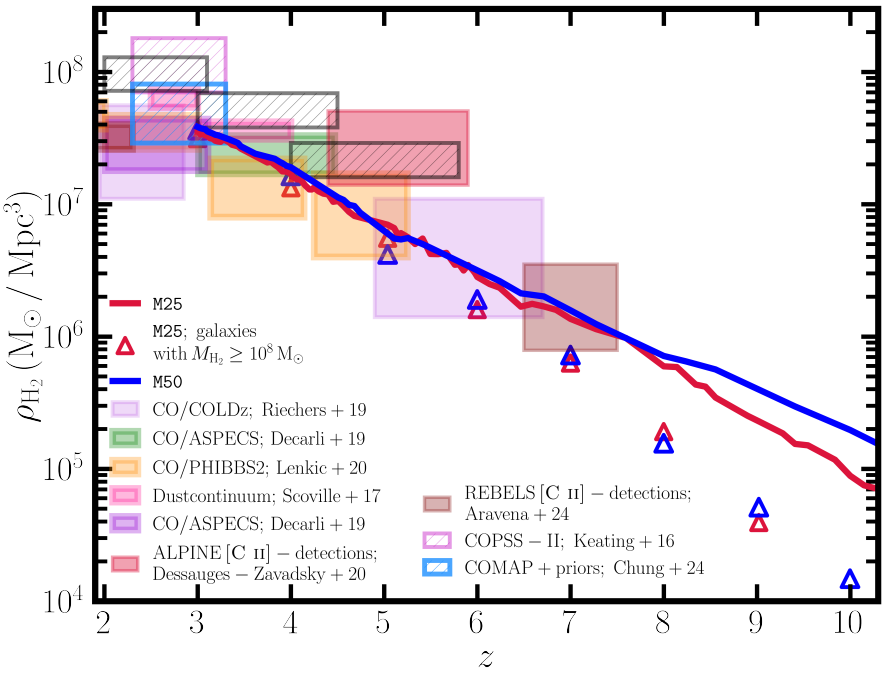


Figure 4.3. – Evolution of the cosmic density of H_2 in the MARIGOLD simulations versus current observational estimates. The solid lines in red (M25) and blue (M50)are for the entire simulation volumes while the open triangles in respective colours show the contribution from galaxies with $M_{H_2} > 10^8\,M_\odot$. The different observational constraints are from Scoville et al. (2017); Decarli et al. (2019); Riechers et al. (2019); Lenkić et al. (2020); Keating et al. (2020); Dessauges-Zavadsky et al. (2020); Aravena et al. (2024).

4.3.2. The cosmic H_2 density

Mapping the time evolution of the cosmic H_2 density provides a way to understand the evolution of the cosmic SFR density. Figure 4.3 shows the simulated cosmic H_2 density ρ_{H_2} at $3 \leq z \leq 10$ and a comparison with current observational constraints (shown by coloured rectangles) using the different observational probes of the H_2 content of galaxies (see section 2.4 for an overview of the different H_2 tracers). These include: (a) measurements of the CO luminosity function for different rotational transitions of CO from blind surveys such as ASPECS Decarli et al. (2019); Walter et al. (2020) and COLDz (Riechers et al., 2019) as well as targeted surveys such as PHIBSS2 (Lenkić et al., 2020); (b) measurements of the dust continuum in galaxies in the COSMOS field with ALMA (Scoville et al., 2017; Magnelli et al., 2020); (c) [C II]-based molecular gas mass estimates from ALPINE at $4.4 \lesssim z \lesssim 5.9$ (Dessauges-Zavadsky et al., 2020) and REBELS at $6.5 \lesssim z \lesssim 7.5$ (Aravena et al., 2024); and (d) constraints from the CO Line Intensity Mapping experiments – mmIME (Keating et al., 2020) at $1 \lesssim z \lesssim 5$, COPSS-II (Keating et al., 2016) and COMAP (Chung et al., 2024) at $2.2 \lesssim z \lesssim 3.2$.

Firstly, we see large uncertainties in individual measurements as well as offset between different estimates. These are due to the systematic uncertainties in the conversion of observed luminosities to H_2 masses, e.g., the CO-to-H_2 conversion factor, the dust-to-gas ratio etc. as well as the small sample size in most of the observations. In contrast, the

simulated ρ_{H_2} is calculated by summing up all the H_2 in the simulation and therefore represents the actual H_2 masses without any dependence on tracers. We find that our simulated ρ_{H_2} agrees very well with most of the observational constraints.

It is worth noting that despite their different resolutions, the ρ_{H_2} from the two simulations coincide with each other for most of the redshift range shown here. Overall, we predict a steady decline of ρ_{H_2} towards high redshifts. In the redshift range $3 \leq z \leq 10$, this is well described by a power-law:

$$\rho_{H_2} \propto 10^{-0.35\,(1+z)}\,. \tag{4.7}$$

As a result the cosmic H_2 density increases by factor of ~ 25 between $z = 7$ and $z = 3$, while at $z = 10$ it decreases to $\approx 7 - 12\%$ of its $z = 7$ value.

Additionally, we quantify the contribution of massive ($M_{H_2} \geq 10^8\,M_\odot$) galaxies to the cosmic H_2 budget. This is denoted by open triangles in Figure 4.3. At $z = 3$, these galaxies contribute $\approx 91\%$[4] of the cosmic ρ_{H_2}. However, this fraction decreases to $\approx 46\%$ by $z = 7$ and at $z = 10$, it is $< 10\%$. This demonstrates that at $z \gtrsim 7$, the cosmic H_2 budget is dominated by low-mass galaxies that are not probed by current surveys like REBELS (see e.g., Figure 1 of Aravena et al., 2024, also see Figure 4.4).

Comparing Equation 4.7 with the $z-$evolution of the SFRD presented in Section 4.3.1, we find that it closely matches the decline in our predicted SFRD at $z < 5$ (Equation 4.6), but is shallower than the decline of the SFRD at $z \geq 5$ (Equation 4.5). This indicates that at $z \lesssim 5$, the SFRD is predominantly regulated by the molecular gas supply, but at earlier cosmic epochs, additional factors also influence the SFRD.

4.3.3. The H_2 mass function

Figure 4.4 shows the simulated H_2 mass function at $z = 4$ and $z = 3$ compared with estimates from ASPECS. The latter are based on measurements of the CO luminosity function for the $J = 4 \rightarrow 3$ transition in the redshift range $3.0 < z < 4.5$ (median $z = 3.8$) and for the $J = 3 \rightarrow 2$ transition in the redshift range $2.0 < z < 3.1$ (median $z = 2.61$). To obtain the CO $J = 1 \rightarrow 0$ luminosity function from these, the authors assume thermalised level populations for the different rotational levels of CO.[5] They estimate the H_2 mass function from the CO $J = 1 \rightarrow 0$ luminosity function by adopting a MW-like conversion factor $\alpha_{CO} = 3.6\,M_\odot\,(K\,km\,s^{-1}\,pc^2)^{-1}$.

We include the upper and lower Poisson uncertainties on the number counts in each mass bin, calculated at a 16% confidence limit following Gehrels (1986). Due to the small size of our simulation volume, we do not have any objects with $M_{H_2} \geq 10^{11}\,M_\odot$ at the redshifts shown. Therefore, for higher masses, we show the (16% confidence) Poisson upper limit on zero counts. At $z = 3$, our highest mass bin overlaps with the lowest mass bin from ASPECS and we find a good agreement between the two. At $z = 4$, however, our prediction is lower than ASPECS. At $M_{H_2} \lesssim 10^8\,M_\odot$, the mass function from our M50 simulation tends to flatten out because of its lower resolution, while the M25 mass function continues to rise steadily.

[4] Because of its eight times larger volume, we report the fractions for the `M50` simulation, but obtain similar results for `M25` at $z \lesssim 7$.

[5] For thermalised level populations, the luminosity of the $1 \rightarrow 0$ transition is related to the luminosity of the $J \rightarrow J-1$ as $r_{J1} = L'(\mathrm{CO}\ J \rightarrow J-1)\,/\,L'(\mathrm{CO}\ 1 \rightarrow 0) = J^2$ for all J.

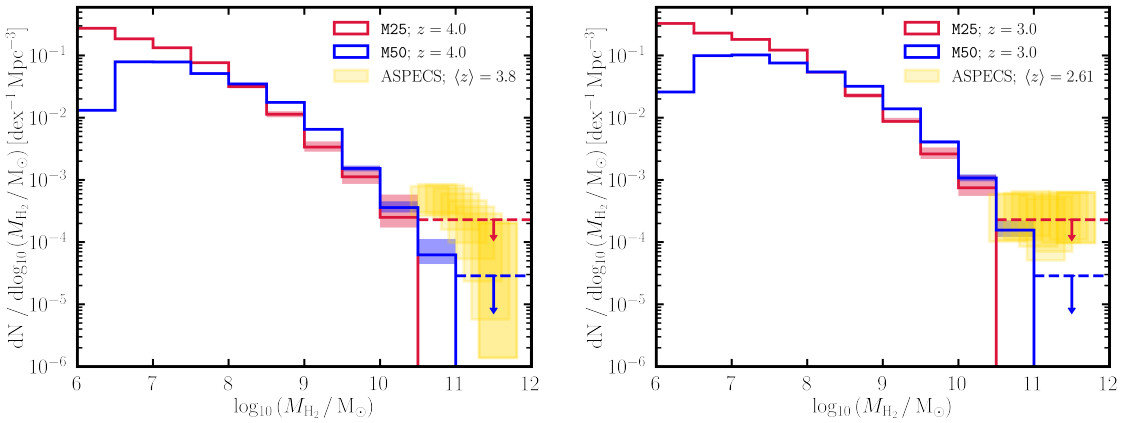


Figure 4.4. – The simulated H_2 mass functions at redshifts $z = 4.0$ (left) and $z = 3.0$ (right) compared with observational constraints in the redshift ranges $3.0 < z < 4.5$ (median $z = 3.8$) and $2.0 < z < 3.1$ (median $z = 2.61$), respectively from ASPECS measurements of the CO luminosity function (Decarli et al., 2019). For the simulated mass functions, the shaded area in each mass bin encloses the lower and upper Poisson uncertainties at 16% confidence limit. The horizontal dashed lines show the upper 16% confidence limit on zero counts.

4.4. Statistical properties of the MARIGOLD galaxies

After examining the evolution of the cosmic SFRD, the molecular gas budget, and the H_2 mass function, we now turn our attention to the statistical properties of our galaxies. To do so, we first combine the galaxy samples from the two simulations – `M25` and `M50` – to form the basis for our statistical analysis of the relationships among various galaxy properties.

4.4.1. Resolution effects and galaxy sample

The different spatial and mass resolutions of `M25` and `M50` can potentially impact the total mass and baryonic content of the simulated galaxies. As shown in Figure 4.1, the halo mass functions from the two simulations agree remarkably well down to a halo mass of $10^9\,h^{-1}\,\mathrm{M}_\odot$. Therefore, to investigate potential resolution effects on the baryonic content, we examine the masses of the different baryonic components as a function of halo mass in the two simulations. These comparisons at $z = 3$ are shown in Figure 4.5. Here, we focus exclusively on central galaxies from the two simulations, as the convergence of the halo mass function is robustly established for main halos but not for substructures.

We find that the galaxies from the two simulations show an excellent agreement in the total baryonic mass vs. halo mass plane. Here the baryonic mass comprises both the stellar and total gas mass. However, there is a clear difference in how this baryonic mass is distributed between the stellar and gas content of galaxies from the two simulations. At a given halo mass, the `M25` simulation tends to have a higher stellar mass at the expense of a lower gas mass. Nevertheless, the distributions from the two simulations show a good overlap. Likewise, the star formation rates (SFR) in the `M25` simulation tend to be on the higher side of the SFR distribution of `M50` galaxies at a given halo mass. In contrast, the

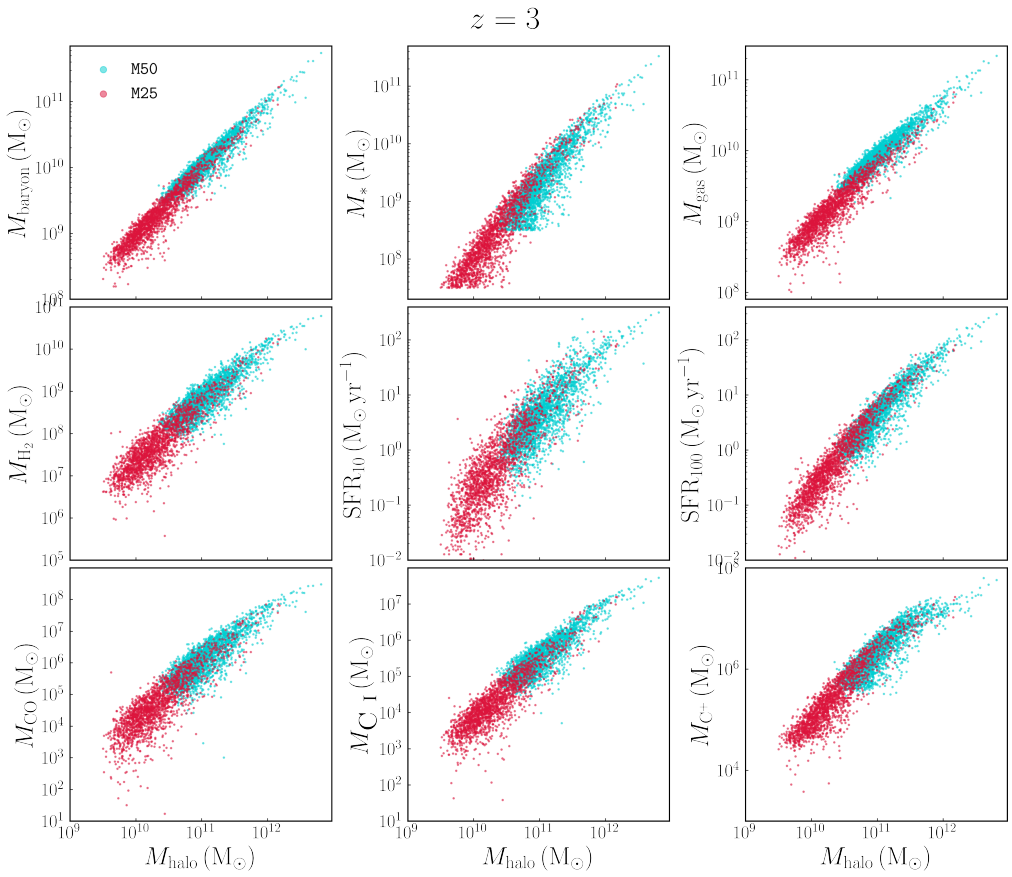


Figure 4.5. – The different baryonic properties of central galaxies from the two simulations –M25 (red) and M50 (turquoise) as a function of the halo mass at $z = 3$. The first panel shows the total baryonic mass which is obtained from the sum of the stellar and total gas mass within a given galaxy.

mass of H_2, CO, C, and C^+ seems to be rather agnostic to the resolution of the simulation and shows excellent agreement between the two simulations at all halo masses. We obtain qualitatively similar results for $z = 5$ and $z = 4$ as well (these are shown in Appendix A). This test validates our choice to combine the galaxies from the two simulations and analyse their global properties together.

We further inspect the stellar mass functions (SMF) of the central galaxies from these simulations. These are shown for $z = 5$, 4, and 3 in Figure 4.6. We find that for all redshifts, at $M_* \geq 10^{9.5}\,\rm M_\odot$, the SMF from the two simulations show a good agreement. At lower M_*, the M50 simulation shows a flattening of the SMF on account of its lower resolution, while the M25 SMF continues to grow. Therefore, in the rest of this chapter, we only consider galaxies from the M25 simulation in the stellar mass range $M_* < 10^{9.5}\,\rm M_\odot$, while at $M_* \geq 10^{9.5}\,\rm M_\odot$, we consider both M25 and M50. In the following, we refer to this combined sample as 'simulated galaxies' or 'MARIGOLD galaxies'.

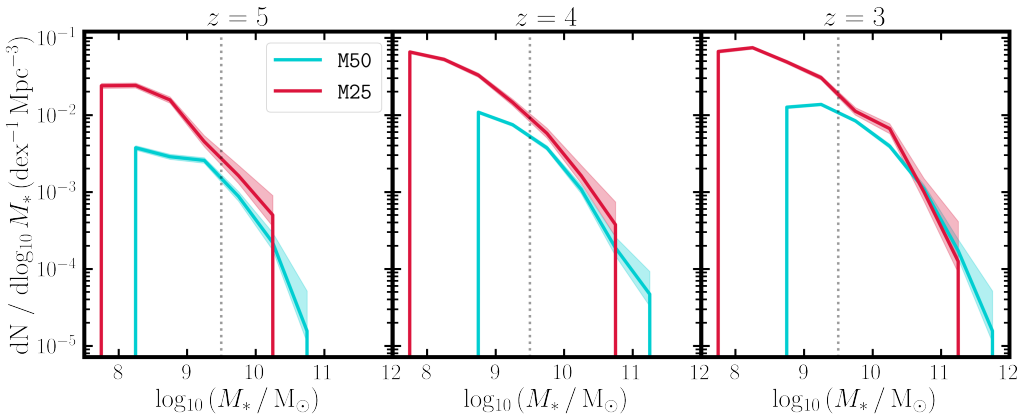


Figure 4.6. – Stellar mass functions for the central galaxies in the `M25` (red) and `M50` (turquoise) simulations at different redshifts. The shaded areas represent the lower and upper Poisson uncertainties on number counts at 16% confidence limit from Gehrels (1986). The dotted line shows the stellar mass cut of $10^{9.5}\,\mathrm{M}_\odot$ above which the mass functions from the two simulations show a good agreement.

4.4.2. The main sequence of star-forming galaxies

The cosmic star formation at any given epoch is dominated by galaxies that show a tight correlation in the SFR-M_* plane, known as the main sequence (MS) of star-forming galaxies (see Section 2.6.1). Here we examine the distribution of our simulated galaxies in the SFR-M_* plane and compare with observational estimates of the MS. Given that the MS relations in the literature use various SFR indicators that are sensitive to the SFR on different timescales (see section 2.5 for a detailed discussion), we compute the SFRs of our simulated galaxies on two timescales – 10 Myr (SFR_{10}) and 100 Myr (SFR_{100}). We show the median SFR in M_* bins along with the interquartile range in Figure 4.7. The observed relations are from Speagle et al. (2014), Shapley et al. (2023), and Clarke et al. (2024) for SFR averaged over 10 Myr and Schreiber et al. (2015), Popesso et al. (2023), and Clarke et al. (2024) for SFR averaged over 100 Myr. It is worth noting that the Shapley et al. (2023) MS at $2.7 < z < 6.5$ has a similar slope as the Speagle et al. (2014) MS but with an offset of -0.34 dex. The authors argue that the MS parametrization of Speagle et al. (2014) adopts a solar metallicity conversion factor between the Hα luminosity and the SFR (Hα) which gives a biased estimate of the SFR of high-redshift galaxies that predominantly have sub-solar metallicities.

At all redshifts, our medians show an excellent agreement with the observational estimates. At the high-mass end, the SFR_{100} in our galaxies tend to be slightly lower than the Schreiber et al. (2015) and Popesso et al. (2023) relations. At these masses, our galaxies agree better with the Clarke et al. (2024) MS who find a slightly lower (~ 0.3 dex) normalization compared to Schreiber et al. (2015). At all M_*, SFR_{10} exhibits a higher scatter compared to SFR_{100}. This is not surprising since star formation in galaxies fluctuates on short timescales and these fluctuations are smoothed out when averaging over a longer timescale. Similar findings were also reported recently by Clarke et al. (2024) from a comparison of the intrinsic scatter in the star-forming main sequence using different SFR

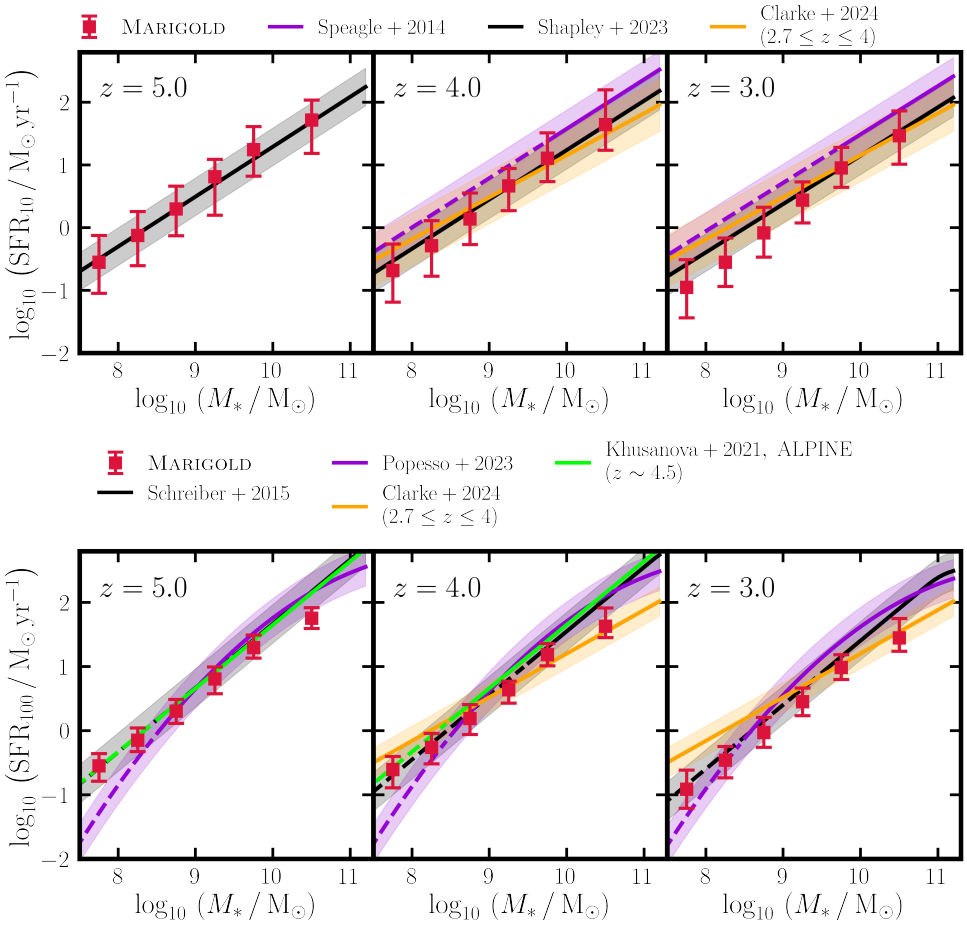


Figure 4.7. – The median SFR in M_* bins for our simulated galaxies at different redshifts compared with MS relations from observations based on Hα observations (Speagle et al., 2014; Shapley et al., 2023; Clarke et al., 2024) and UV observations (Schreiber et al., 2015; Popesso et al., 2023; Clarke et al., 2024). As these observations are sensitive to star formation on different timescales, we show the SFR averaged over 10 Myr *(top panels)* and 100 Myr *(bottom panels)*.

tracers. A detailed discussion on the scatter around the main sequence can be found in Matthee et al. (2019).

4.4.3. The mass-metallicity relation

The metal content of a galaxy is a useful probe of the complex interplay between gas accretion, stellar nucleosynthesis and feedback, and gas mixing within the ISM. Observations have reported a strong correlation between the gas-phase metallicity ($Z_{\rm gas}$) and the stellar mass of galaxies, known as the mass-metallicity relation (MZR; Tremonti et al., 2004). Figure 4.8 shows the MZR for our simulated galaxies at different redshifts. The $Z_{\rm gas}$ of a galaxy is computed as a gas-mass-weighted average over all cells within the galaxy and is

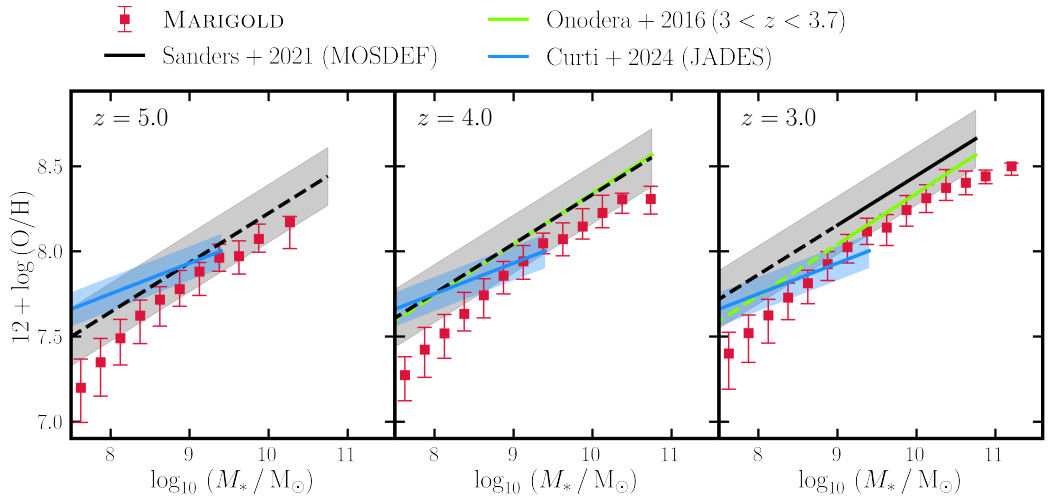


Figure 4.8. – The gas-metallicity expressed as 12 + log (O/H) vs. stellar mass correlation in our simulated galaxies at different redshifts. The squares represent the median 12 + log (O/H) in different stellar-mass bins and the error bars denote the interquartile rang. The black line shows an extrapolation of the $z = 2.3$ MZR from Sanders et al. (2021) to the respective redshift, while the shaded area denotes the 1σ scatter around the MZR. The green line shows the best-fit MZR for a sample of UV-selected galaxies form Onodera et al. (2016). For clarity, here we only show the scatter around the best-fit relation for the Sanders et al. (2021) relation. The Onodera et al. (2016) relation has a similar scatter. For both relations, the solid line shows the extent of stellar masses constrained by their observations while the dashed line denotes the extrapolation to lower masses. The blue line shows the best-fit MZR (at the respective redshift) derived by Curti et al. (2024) based on a sample of galaxies at $3 \leq z < 6$, with the shaded area representing the 1σ scatter.

normalized to the solar value of 0.02[6] (Karakas, 2010). The $Z_{\rm gas}$ values are converted to the oxygen abundance 12 + log (O/H) following: 12 + log (O/H) = $\log_{10}(Z_{\rm gas}\,/\,{\rm Z}_\odot)$+8.69. The best fit to the MZR for a sample of 41 UV-selected galaxies at $3.0 \lesssim z \lesssim 3.7$ with Keck/MOSFIRE spectroscopy (Onodera et al., 2016) is shown for comparison. Sanders et al. (2021) used a sample of ~ 300 and ~ 150 galaxies at $z = 2.3$ and $z = 3.3$ from the MOSFIRE Deep Evolution Field (MOSDEF) survey to study the redshift-evolution of the MZR. Here we use the MZR they found for their highest redshift sample ($z = 2.3$) and extrapolate it to the redshifts shown in the figure to compare with our predicted MZR. Recently, Curti et al. (2024) analysed a sample of low-mass ($6.4 \lesssim \log(M_*/{\rm M}_\odot) \lesssim 9.4$) galaxies at $3 < z < 10$ from the JWST Advanced Deep Extragalactic Survey (JADES), thereby complementing previous MZR estimates that primarily targeted more massive galaxies.

Overall our median metallicities tend to be on the lower side of the Sanders et al. (2021) MZR (although consistent within the scatter). At $z = 3$, our galaxies show an excellent agreement with Onodera et al. (2016) at $M_* \gtrsim 10^{9.5}\,{\rm M}_\odot$ and with Curti et al. (2024) for lower masses. At $z = 4$ and 5, our medians fall slightly below the Curti et al. (2024) estimates.

[6]Note that the observational MZRs we compare to assume an absolute solar metallicity value of 0.014. For a fair comparison, we increment our values by $\log_{10}(0.02/0.014)$.

We observe a weak redshift-evolution in the MZR of ≈ 0.13 dex per Gyr, independent of the stellar mass. At $z = 3$ and 4, we see signs of a flattening of the MZR at of $M_* \sim 10^{10.5}\,\mathrm{M}_\odot$ in both simulations, which is not apparent at $z = 5$. A flattening of the MZR has previously been reported in other simulations (e.g., Zahid et al., 2014; Torrey et al., 2019) as well as observations at low (Tremonti et al., 2004; Mannucci et al., 2010) and high (Sanders et al., 2021) redshifts. In their $z \sim 2.3$ and $z \sim 3.3$ samples Sanders et al. (2021) found that the metallicity in the highest mass bins ($M_* \sim 10^{10.5}\,\mathrm{M}_\odot$) falls below the respective best-fit MZRs (although still within 2σ of the best-fit), which they perceive as a hint of the flattening of the MZR above those masses. However, they caution an apparent flattening could also result from a selection effect against more metal- rich (and therefore dusty) galaxies. A possible physical explanation for this flattening is a change in the fundamental physical processes related to metal production in galaxies and the flow of metals (alongside baryons) in and out of galaxies (Tremonti et al., 2004; Zahid et al., 2014). We note that in our simulation, this could also be an effect of the small number of galaxies at these masses. Therefore, verifying a flattening of the MZR would require a larger simulation volume.

4.4.4. The stellar size-mass relation

In this section, we compare the stellar size-mass relation of our simulated galaxies with observational estimates. The latter generally determine the half-light radius from the optical or UV emission from galaxies. Inferring the half-mass radius of a galaxy starting from an estimate of its half-light radius is non-trivial and depends on the emission wavelength (van der Wel et al., 2014; Szomoru et al., 2013). Using rest-frame optical and near-IR observations of a sample of 177 massive ($M_* > 10^{10.7}\,\mathrm{M}_\odot$) galaxies at $0.5 < z < 2.5$, Szomoru et al. (2013) compared the half-mass radius ($r_{1/2,*}$) and the $g-$band half-light radius (r_g) of galaxies. They found a stronger mass concentration compared to the optical light concentration, i.e., $r_{1/2,*} < r_\mathrm{g}$ and on average, $r_{1/2,*} = 10^{-0.1} r_\mathrm{g}$ for their galaxies. They further found no significant redshift evolution of this ratio. In the following, we assume that this relation holds out to $z = 5$ and scale the observed half-light radii of galaxies by a conversion factor of $10^{-0.1}$ to obtain their corresponding half-mass radii for comparison with our simulated galaxies. In general, the half-light radius of a galaxy varies with the wavelength and it is expected that $r_\mathrm{UV} > r_\mathrm{optical}$ (van der Wel et al., 2014). Therefore, we expect a higher ratio between the observed UV half-light radii and the corresponding half-mass radii. However, because of a lack of such a calibration between r_UV and $r_{1/2,*}$, we apply the same scaling factor for UV observations as well.

Figure 4.9 shows the median half-mass radius of galaxies in a given stellar mass bin.[7] For comparison, we show the $r_{1/2,*}$ obtained from rest-frame UV observations of galaxies from Shibuya et al. (2015) at the respective redshifts as well as $r_{1/2,*}$ for massive SFGs and quiescent galaxies from Straatman et al. (2015). We also include the size-mass relation from van der Wel et al. (2014) based on photometric and spectroscopic measurements of $\sim$ 30000 galaxies at $0 < z < 3$ (in black). The authors found no strong evidence for the redshift evolution of the slope of the relation, while for the intercept, they found that

[7] Note that here the stellar mass is the one calculated within 0.1 R_vir and $r_{1/2,*}$ contains half of this mass. In all other analysis, we refer to all quantities calculated with the galaxy radius defined as twice the half-mass radius ($2\,r_{1/2,*}$).

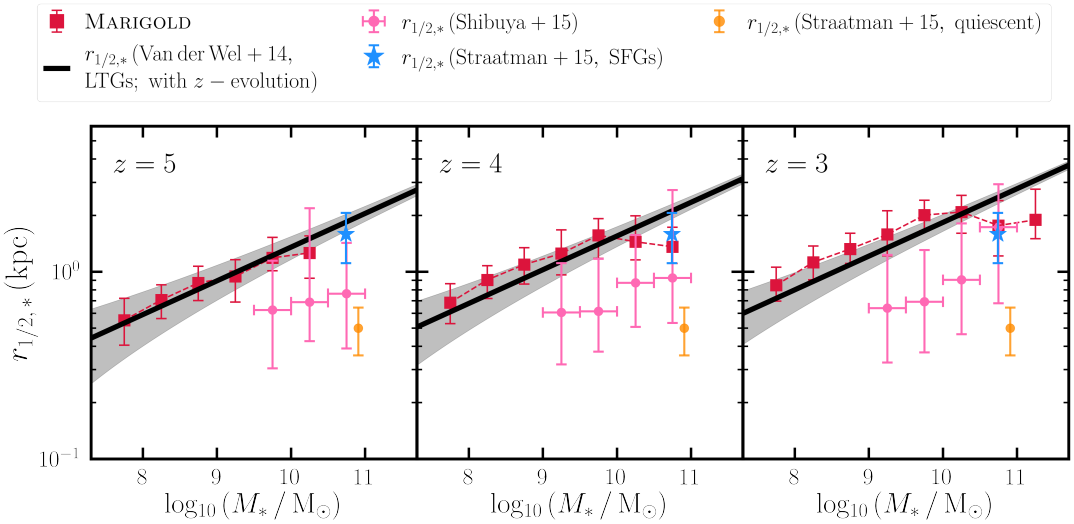


Figure 4.9. – The stellar half-mass radii of simulated galaxies at redshifts $3 \leq z \leq 5$ as a function of their stellar mass. The median size of galaxies in a given mass bins are shown by red squares and the error bars represent the interquartile range. The size-mass relation from van der Wel et al. (2014) with the expected redshift-evolution from their study is shown in black with the shaded area representing the observed scatter of 0.19 dex. The median $r_{1/2,*}$ of galaxies at the respective redshifts from Shibuya et al. (2015) are shown in pink. The median sizes of massive SFGs (light blue star) and quiescent galaxies (orange circle) from ZFOURGE (FourStar Galaxy Evolution Survey) at $3.4 \leq z \leq 4.2$ are taken from Straatman et al. (2015). The error bars denote the interquartile range of the respective observational estimates. For all observations, we converted their estimated half-light radii to the corresponding half-mass radii using the scaling $r_{1/2,*} = 10^{-0.1}\, r_{1/2,\,\mathrm{light}}$ found by Szomoru et al. (2013, see text for details).

the sizes of late-type galaxies decrease with redshift as $r \propto (1+z)^{-0.75}$. Their size-mass relation for $2.5 < z < 3$ is shown in the right panel of Figure 4.9. To compare with our galaxies at $z = 4$ and 5, we adopt their redshift scaling and the corresponding relations are shown as a solid black line in the respective panels.

In general, the $r_{1/2,*}$ of our galaxies show a good agreement with observations. For both simulations, we see that the galaxy sizes increase with stellar mass and flatten out at $M_* \gtrsim 10^{10}\,\mathrm{M}_\odot$. At a given stellar mass, the galaxy sizes increase over time. We further find that our highest mass ($M_* \sim 10^{10.5}\,\mathrm{M}_\odot$) galaxies tend to be fairly compact, with sizes $\lesssim 2$ kpc in agreement with the median size reported by Straatman et al. (2015) for SFGs with similar masses (light blue star with error bar).

4.4.5. Molecular gas fraction

The presence of neutral atomic hydrogen (H I) in galaxies is inferred through the hyperfine transition of hydrogen that arises from a spin flip of the electron with respect to the proton. The transition results in an emission line at 21 cm. The lack of H I 21 cm observations in individual galaxies at high redshifts ($z \gtrsim 0.5$) has led to several assumptions

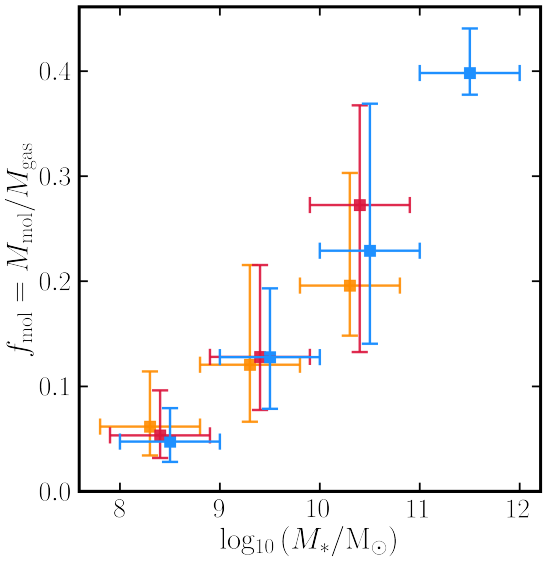


Figure 4.10. – Redshift evolution of the molecular fraction $f_{\rm mol} = M_{\rm mol}/M_{\rm gas}$ in simulated galaxies as a function of their stellar mass. The squares denote the median in a stellar mass bin while the error bars denote the 16-84 percentile range. The points at different redshifts are slightly shifted along the $x-$ axis for clarity.

regarding the H_2/H I mass ratio. Particularly, at $z \sim 2-3$, it is commonly assumed that the gas content in $M_* \gtrsim 5 \times 10^{10}\,{\rm M}_\odot$ galaxies is dominated by H_2 (Tacconi et al., 2018).

We test this assumption by computing the molecular gas to total gas fraction:

$$f_{\rm mol} \equiv \frac{M_{\rm mol}}{M_{\rm gas}} = \frac{1.36\,M_{\rm H_2}}{M_{\rm gas}}\,. \tag{4.8}$$

The parameter $f_{\rm mol}$ denotes the fraction of total gas within the galaxy present in molecular form. Figure 4.10 shows the median $f_{\rm mol}$ as a function of M_* at different redshifts in the MARIGOLD simulations.

Across redshift, $f_{\rm mol}$ increases with M_*, albeit with a significant scatter at all stellar masses. At a given stellar mass, however, $f_{\rm mol}$ does not evolve with redshift, implying that galaxies with a given stellar mass retain a fixed fraction of their gas content in molecular form. Most importantly, $f_{\rm mol}$ is always less than 1 for all stellar masses, indicating a significant contribution of atomic and ionised gas to the total gas budget in our galaxies. Similar findings have been reported in previous numerical simulations (Popping et al., 2014; Lagos et al., 2015; Davé et al., 2020).

Similar findings are beginning to emerge from observational studies probing high-redshift galaxies across multiple wavelengths. Recently, Catán et al. (2024) estimated the molecular gas fraction by comparing molecular gas mass estimates derived from mid-J ($J = 3, 4, 5$) CO lines and dust continuum emission in a stack of 13 main-sequence galaxies at $z \sim 3$ with $M_* \in [10^{9.5}\,{\rm M}_\odot, 10^{10.7}\,{\rm M}_\odot]$. Assuming that the dust continuum emission traces the total (atomic + molecular) gas mass in these galaxies, while CO emission provides a reliable estimate of the molecular gas mass, they calculated $f_{\rm mol}$ as the ratio of the two mass

estimates. Their results indicate that molecular gas accounts for only $\sim 9\%$ of the total gas mass in these galaxies, with the rest being atomic.

4.5. Conclusions

In this chapter, we described the physical and chemical processes included in the MARIGOLD simulations and examined the properties of the simulated galaxies in the context of the established scaling relations in the literature as well as the cosmic evolution of the H_2 density and the SFRD. Our key findings can be summarized as:

1. The cosmic SFRD in our simulations evolves as $\rho_{\rm SFR} \propto 10^{-0.36\,(1+z)}$ at $z < 5$ and as $\rho_{\rm SFR} \propto 10^{-0.43\,(1+z)}$ at $5 \leq z \leq 10$. This results in a steeper decline at $z \geq 8$ compared to the extrapolation of the best fit from Madau & Dickinson (2014), in agreement with recent observational estimates of the SFRD from Harikane et al. (2022b).

2. The cosmic H_2 density in our simulations agrees remarkably well with observations and shows a 25-fold increase between redshifts 7 and 3. In the redshift regime $z > 7$, currently unconstrained by observations, we find that $\rho_{\rm H_2}$ decreases by an order of magnitude from $z = 7$ to $z = 10$. Overall, at $3 \leq z \leq 10$, we report a steady power-law evolution: $\rho_{\rm H_2} \propto 10^{-0.35\,(1+z)}$.

3. The $z-$evolution of the simulated SFRD declines mirrors that of $\rho_{\rm H_2}$ at $z < 5$, supporting the idea that the evolution of the SFRD is primarily driven by the availability of molecular gas. However, at $z \geq 5$, we observe a steeper decline in the SFRD compared to $\rho_{\rm H_2}$, suggesting that the SFRD is regulated not only by the supply of molecular gas but also by additional factors.

4. We find that low-mass ($M_{\rm H_2} < 10^8\,{\rm M}_\odot$) galaxies contribute significantly to the cosmic H_2 budget at $z \gtrsim 7$, with their contribution reaching $\gtrsim 50\%$ at $z = 7$ and $\gtrsim 90\%$ at $z = 10$.

5. Our simulated galaxies align well with established scaling relations from the literature between various galaxy properties and stellar mass (M_*), including the main sequence of star-forming galaxies (SFR-M_*), the mass-metallicity relation ($Z_{\rm gas}$-M_*), and the stellar size-mass relation (stellar half-mass radius-M_*).

The remarkable agreement of the MARIGOLD simulations with observed cosmic trends, the H_2 mass function, and the scaling relations between galaxy properties lends credibility to our approach and underscores the reliability of the simulated galaxy population. The large dynamical range of galaxy masses collectively probed by the `M25` and `M50` simulations enables us to validate our predictions against observations in the regime where observational data exist (primarily $M_* \gtrsim 10^9\,{\rm M}_\odot$), while simultaneously offering insights into the properties of low-mass galaxies.

CHAPTER 5

The [C II] line emission as an ISM probe in the MARIGOLD galaxies

The [C II] line emission as an ISM probe in the MARIGOLD galaxies[1]

Prachi Khatri, Emilio Romano-Díaz, and Cristiano Porciani
Submitted for publication to Astronomy & Astrophysics
arXiv:2411.09755 (DOI: https://doi.org/10.48550/arXiv.2411.09755)

Overview

The [C II] line is one of the fine-structure transition of singly-ionised carbon, C^+. It is among the brightest emission lines at infrared wavelengths and accounts for $\sim 0.1 - 1\%$ of the total infrared luminosity in star-forming galaxies. The luminosity of the [C II] line correlates strongly with the star formation rate (SFR; Stacey et al., 2010; De Looze et al., 2014; Carniani et al., 2018; Schaerer et al., 2020) and the molecular gas mass ($M_{\rm mol}$; Hughes et al., 2017; Madden et al., 2020; Zanella et al., 2018) in galaxies near and far. In this work, we evaluate the utility of this line as a tracer of the SFR and $M_{\rm mol}$ at $z \geq 3$ from a numerical perspective. For this, we use the galaxies from the MARIGOLD simulations (described in Chapter 4). In particular, we want to answer the following open questions:

1. Does the [C II]-SFR relation evolve with redshift?
2. Does the [C II]-$M_{\rm mol}$ relation depend on other galaxy properties?
3. How does the number of galaxies emitting a given [C II] luminosity change over time?
4. How is the emission distributed in and around galaxies?

[1]The version of this paper submitted for publication to A&A is reproduced in its entirety in Appendix C.

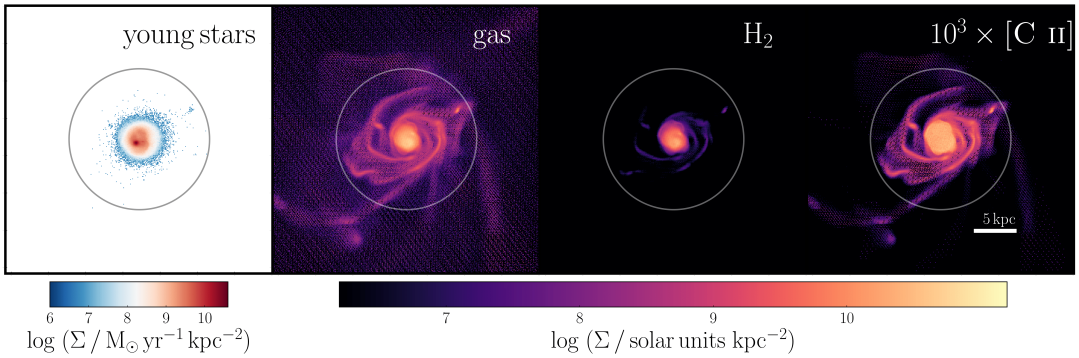


Figure 5.1. – Face-on view of a simulated galaxy at $z = 3$. From left to right, the columns show the surface density of young stars (with ages $\leq$ 200 Myr), total gas surface density, H_2 surface density, and [C II] surface brightness. In each panel the circle indicates 0.1 times the virial radius of the parent DM halo. The second panel refers to the total gas in the ISM, including H_2. (This is similar to Figure 1 in Khatri et al. (2024b), where we present a galaxy at a different redshift).

5.1. Modelling [C II] emission

We calculate the [C II] emission from the MARIGOLD galaxies after the simulation has been performed, i.e., in post-processing. An example of the [C II] emission in a simulated galaxy at $z = 3$ is shown in the rightmost panel of Figure 5.1. For reference, we also show the distribution of young stars (ages $\leq$ 200 Myr), total gas, and molecular hydrogen within the galaxy. The figure presents a 2D face-on projection of the galaxy, with the angular momentum axis of the stellar particles pointing out of the paper towards the reader.

5.2. The [C II] luminosity function

The [C II] luminosity function (LF) $\phi(L_{[\mathrm{C\,II}]})$ depicts the number density n of galaxies emitting within a given $L_{[\mathrm{C\,II}]}$ interval such that[2]

$$\phi(L_{[\mathrm{C\,II}]}) \equiv \mathrm{d}n/\mathrm{d}\log L_{[\mathrm{C\,II}]}\,. \tag{5.1}$$

In this section, we examine how the [C II] LF evolves with redshift.

As a reminder, the MARIGOLD suite comprises two simulations: `M25` and `M50`. The former has eight times better mass resolution and half the minimum cell size compared to the latter. Conversely, `M50` has an eight times larger simulation volume at a lower resolution. This allows `M25` to better resolve low-mass galaxies (see Table 4.1), while `M50` can simulate rare and bright galaxies. We aim to utilize these simulations with different spatial and mass resolutions to jointly constrain the [C II] LF over a broad range of [C II] luminosities. To achieve this, we first compare the separate LFs from the two simulations. Above a threshold luminosity L_{thr}, the LFs show small differences compatible with sample variance, but substantial systematic differences for $L_{[\mathrm{C\,II}]} < L_{\mathrm{thr}}$. The threshold luminosity is approximately $10^{5.5}\,\mathrm{L}_\odot$ for $5 \leq z \leq 7$ and $10^8\,\mathrm{L}_\odot$ for $3 \leq z < 5$.

[2]Note that here and throughout this chapter, we use 'log' to denote $\log_{10}$; for the natural logarithm $\log_e$, we use 'ln' instead.

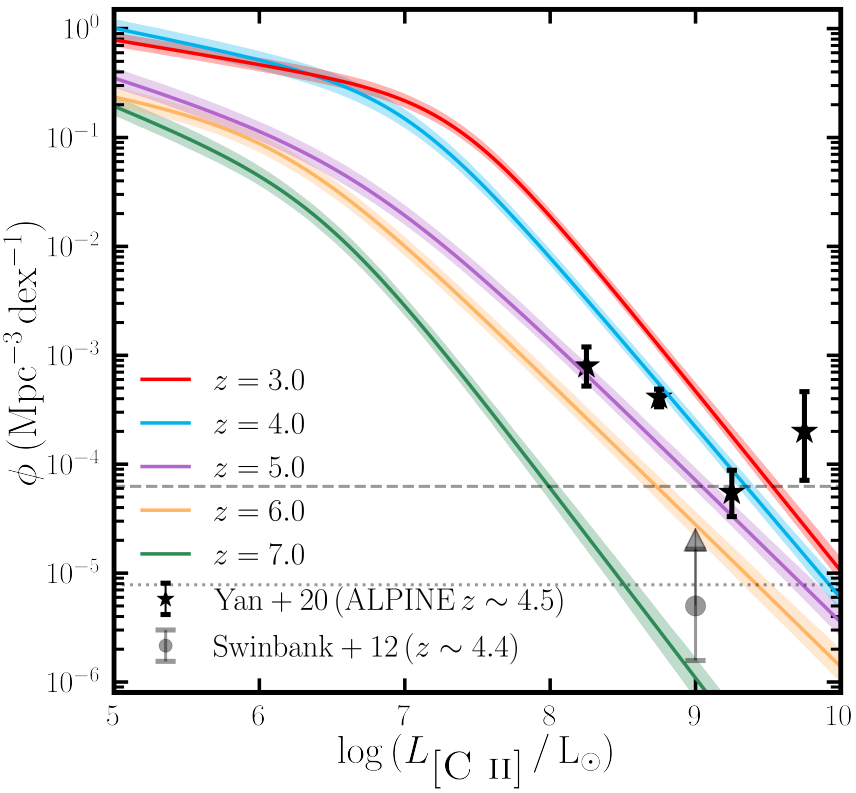


Figure 5.2. – The simulated [C II] LF compared with observational estimates. The coloured lines represent the best-fit DPL – Equation (5.3) – to the simulated LF and the shaded area represents the central 68% credibility range obtained using the MCMC chains. Black stars represent the observational estimates at $z \sim 4.5$ from the ALPINE survey (Yan et al., 2020) and the grey arrow shows the lower limit from Swinbank et al. (2012) based on observations of two galaxies at $z \sim 4.4$. The dashed and dotted horizontal lines represent a number count of 1 per dex in the entire simulation volume of `M25` and `M50`, respectively.

We fit two different functional forms to the [C II] LF at each redshift – a Schechter function and a double power-law (DPL):

$$\phi(L_{[\mathrm{C\,II}]}) = \ln(10)\,\phi_*\left(\frac{L_{[\mathrm{C\,II}]}}{L_*}\right)^{\alpha+1}\exp\left(-\frac{L_{[\mathrm{C\,II}]}}{L_*}\right), \tag{5.2}$$

$$\phi(L_{[\mathrm{C\,II}]}) = \ln(10)\,\phi_*\left[\left(\frac{L_{[\mathrm{C\,II}]}}{L_*}\right)^{-(\alpha+1)} + \left(\frac{L_{[\mathrm{C\,II}]}}{L_*}\right)^{-(\beta+1)}\right]^{-1}, \tag{5.3}$$

where ϕ_* is the normalization of the LF, L_* is the turnover luminosity where the LF changes from a powerlaw to an exponential (in case of the Schechter function) or changes slope (in case of a DPL), $\alpha < 0$ is the faint-end slope, and $\beta < 0$ is the bright-end slope, such that $|\beta| \geq |\alpha|$. The fit to the simulated LFs is perfomed using a Bayesian approach for sampling the parameter space using the Markov-chain Monte-Carlo (MCMC) method implemented within the python package `emcee` (Foreman-Mackey et al., 2013).

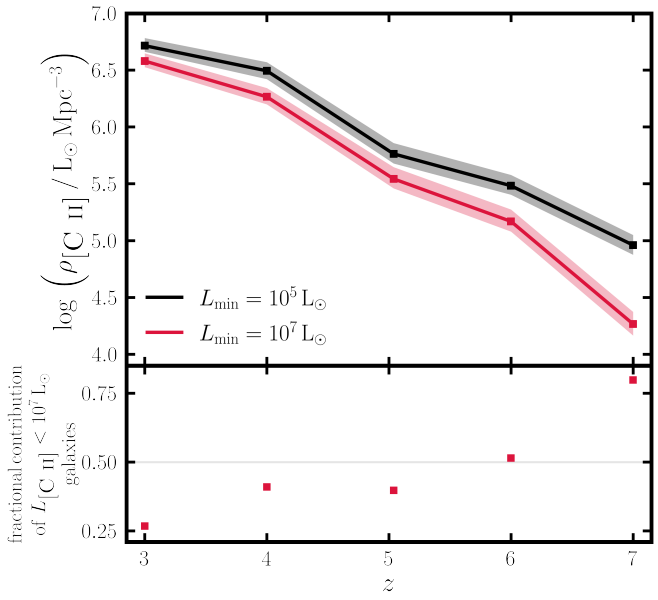


Figure 5.3. – Contribution of faint ($L_{[\mathrm{C\,II}]} < 10^7\,\mathrm{L}_\odot$) galaxies to the total luminosity density $\rho_{[\mathrm{C\,II}]}$. The $\rho_{[\mathrm{C\,II}]}$ at each redshift is obtained by integrating the [C II] LF shown in Figure 5.2 down to a L_{min}. We consider two values of L_{min} – $10^7\,\mathrm{L}_\odot$, that is typical of surveys like ALPINE and REBELS, and a lower value of $10^5\,\mathrm{L}_\odot$, which corresponds to the lowest luminosity we probe in the MARIGOLD simulations. The $\rho_{[\mathrm{C\,II}]}$ with these values of L_{min} are shown by the black and red curves in the top panel. The bottom panel shows one minus the ratio of the two and quantifies the fractional contribution of $L_{[\mathrm{C\,II}]} < 10^7\,\mathrm{L}_\odot$ emitters to the total $\rho_{[\mathrm{C\,II}]}$.

To account for the sample variance of the relatively small computational volumes in the fitting procedure, we follow an approach similar to that presented in Trenti & Stiavelli (2008). Briefly, we fit the LFs from the two simulations with the same functional form but allow them to differ in their normalization. These separate normalizations differ from the 'cosmic' normalization ϕ_* by a sample variance term such that $\phi_{*,j} = \phi_* + \Delta_j$, where Δ_j is the correction due to sample variance in the j^{th} simulation. We impose Gaussian priors on the two Δ_j such that $\Delta_j \sim \mathcal{N}(0, (\sigma_{v,j}/\ln 10)^2$, where $\sigma^2_{v,j}$ is the sample variance of the overdensity within the respective simulation volume, estimated following Somerville et al. (2004) (considering the halo mass bin that gives the dominant contribution to the counts of emitters around L_*). We adopt broad flat priors on the other parameters.

Finally, we employ the deviance information criterion (DIC; Spiegelhalter et al., 2002) to investigate which functional form provides a better fit to our simulated data based on the MCMC chains. The DIC rules out the Schechter function by a large margin at all redshifts.

The resulting DPL fits at different redshifts are shown in Figure 5.2. Our predictions are in excellent agreement with current observational estimates from the ALPINE survey (Yan et al., 2020) and the upper limit from Swinbank et al. (2012). We find a flattening of the faint-end slope (α) at late times ($z \lesssim 4$) and an increase in the turnover luminosity (L_*) over time. Moreover, our LFs predict an $\approx$ 600-fold increase in the abundance of $L_{[\mathrm{C\,II}]} \sim 10^9\,\mathrm{L}_\odot$ galaxies from $z = 7$ to $z = 3$.

Now we inspect the evolution of the [C II] luminosity density in the simulation. For this we integrate the [C II] luminosity function shown in Figure 5.2 down to $\log(L_{\rm min} / {\rm L}_\odot) = 5$. This is shown by a black line in Figure 5.3. Additionally, we obtain another estimate of $\rho_{\rm [C\,II]}$ by integrating the [C II] LF down to $\log(L_{\rm min} / {\rm L}_\odot) = 7$ (shown as a red line in the figure). Subtracting the ratio of the two from 1 gives the contribution of faint[3] galaxies to the cosmic $\rho_{\rm [C\,II]}$ at different redshifts. This is shown in the bottom panel of the figure. We find that at $z = 3$, faint galaxies contribute approximately a quarter of the total luminosity. This fraction steadily increases with redshift at $z \gtrsim 6$, more than half of [C II] emission arises from these faint emitters.

Similar to the comparison of LFs described at the beginning of this section, we also compare the [C II] LFs for central and satellites galaxies separately. We find that for $L_{\rm [C\,II]} \geq 10^5\,{\rm L}_\odot$, the LFs of centrals from the two simulations are in good agreement, while the satellites LFs differ substantially. Therefore, in the following sections, we only consider central galaxies for deriving the scaling relations between $L_{\rm [C\,II]}$ and galaxy properties.

5.3. The [C II]-SFR relation

The strong correlation between the [C II] line luminosity and the SFR can be exploited to obtain an estimate of the SFR in distant galaxies. This is particularly useful as the [C II] line is not affected by dust-obscuration and as a result provides a straightforward SFR tracer. To do so, we must first understand how the [C II]-SFR relation evolves with redshift. This relation for MARIGOLD galaxies is shown for three different redshifts in Figure 5.4.

At each redshift, we fit a relation of the form $\log\left(L_{\rm [C\,II]}/{\rm L}_\odot\right) = a \log\left({\rm SFR}_{200}/{\rm M}_\odot\,{\rm yr}^{-1}\right) + b$ (where $\{a, b\} \in \mathbb{R}^2$) to our simulated galaxies using an ordinary least squares (OLS) linear regression. The distribution of galaxies in the $L_{\rm [C\,II]}$-SFR plane at each redshift is shown as purple hexagons in Fig. 5.4, along with the OLS best-fit as a solid red line. We compare with empirical relations from De Looze et al. (2014) based on a sample of $0.5 \leq z \leq 6.6$ galaxies and Schaerer et al. (2020) based on galaxies from the ALPINE survey at $4.4. \lesssim z \lesssim 5.9$. Several individual galaxies are also shown as coloured scatter points. Overall, we find a good agreement of the simulated [C II]-SFR relation with observations. We find a mild evolution in the slope of the [C II]-SFR relation (~ 0.15 dex). In contrast, we note a strong evolution in the intercept that increases by a factor of approximately 3 from $z = 7$ to $z = 3$. As a result, our [C II]-SFR relation at $z = 3$ is in excellent agreement with the De Looze et al. (2014) relation based on a sample of $0.5 \leq z \leq 5.5$ galaxies; while at higher redshifts, we observe a clear vertical offset.

We also examine the spatially-resolved [C II]-SFR relation at $z = 4$. For this, we obtain the [C II] surface brightness and SFR surface density maps of each central galaxy. We then apply a 2D Gaussian smoothing with the beam sizes matching those in Posses et al. (2024), which represents one of the highest resolution achieved in current high-z observations. At $z = 4$, their beam size corresponds to a full width at half-maximum of ~ 1.4 kpc and ~ 1.2 kpc, respectively, for [C II] and SFR. To examine possible variations in the galaxy

[3]Here, we define galaxies as faint based on a luminosity threshold of 10^7 **based** on the sensitivity of current surveys at $z \gtrsim 4$ such as ALPINE and REBELS

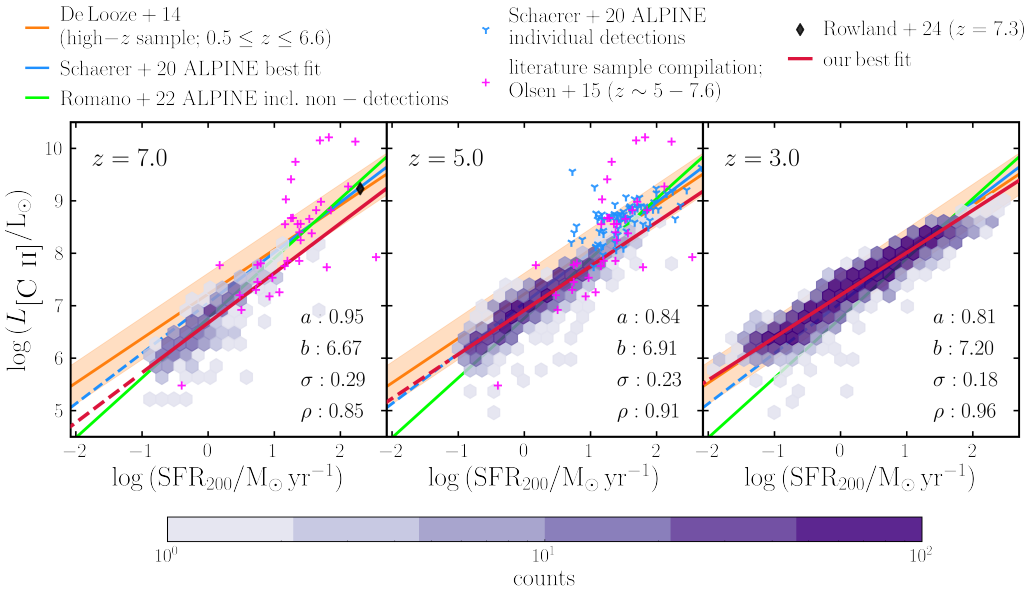


Figure 5.4. – The [C II]−SFR relation from the MARIGOLD simulations compared with observations. Each panel shows a different redshift and the simulated galaxy population is represented as purple hexbins, with the colour indicating the galaxy counts per bin. The red line shows the best-fit: $\log(L_{[\mathrm{C\,II}]}/\mathrm{L}_\odot) = a\log(\mathrm{SFR}_{200}/\mathrm{M}_\odot\,\mathrm{yr}^{-1}) + b$. The fit parameters are reported in each panel, along with the scatter (σ) around the best-fit and the Spearman's rank correlation coefficient (ρ). For comparison we include the best-fit relations from De Looze et al. (2014) for their high−z ($0.5 < z < 6.6$) sample (orange line, with the scatter shown as a shaded area) and Schaerer et al. (2020) for the ALPINE survey (blue line). The individual ALPINE galaxies (at $4.5 \lesssim z \lesssim 5.9$), the literature sample (at $5 \lesssim z \lesssim 7.6$) taken from Olsen et al. (2015), and REBELS-25 from the REBELS survey Rowland et al. (2024) are shown with blue, pink, and black symbols, respectively. This figure is a modified version of Figure 5 in Khatri et al. (2024b).[4]

populations, we split our galaxy sample into three stellar mass (M_*) bins – $10^8\,\mathrm{M}_\odot \leq M_* < 10^9\,\mathrm{M}_\odot$, $10^9\,\mathrm{M}_\odot \leq M_* < 10^{10}\,\mathrm{M}_\odot$, and $M_* \geq 10^{10}\,\mathrm{M}_\odot$. We find an excellent agreement of our median $\Sigma_{[\mathrm{C\,II}]}$ at a given Σ_{SFR} with the $\Sigma_{[\mathrm{C\,II}]} - \Sigma_{\mathrm{SFR}}$ relation for ALPINE galaxies for $\Sigma_{\mathrm{SFR}} \lesssim 1\,\mathrm{M}_\odot\,\mathrm{yr}^{-1}\,\mathrm{kpc}^{-2}$. Towards higher Σ_{SFR}, however, we find a flattening of $\Sigma_{[\mathrm{C\,II}]}$, which is driven by the increased abundance of CO at the expense of C^+ at high surface densities in our galaxies.

5.4. [C II] as a molecular gas tracer

In recent years, the [C II] line has garnered significant attention as a molecular gas tracer (Zanella et al., 2018; Dessauges-Zavadsky et al., 2020; Aravena et al., 2024), particularly owing to its brightness at $z \gtrsim 4$, where conventional tracers like CO rotational lines become observationally expensive. However, this relation and its redshift evolution have not been extensively investigated in numerical simulations. We examine this using the MARIGOLD galaxies in Figure 5.5. Here, again we fit a linear relation at each redshift: $\log(L_{[\mathrm{C\,II}]}) = a\log(M_{\mathrm{mol}}) + b$. In contrast to the [C II]-SFR relation, we find a stronger redshift evolution in the [C II]-M_{mol} relation. In particular, the correlation is relatively

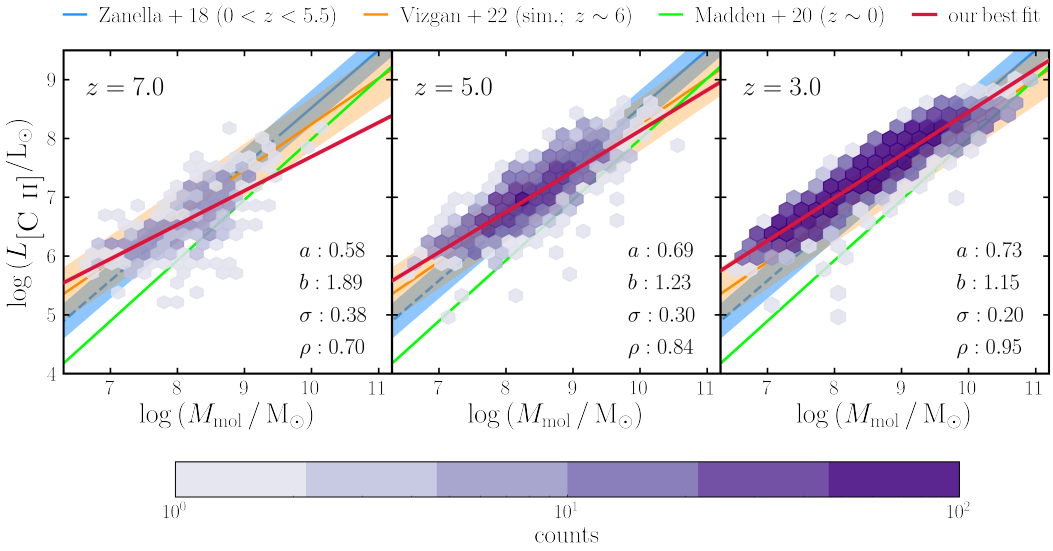


Figure 5.5. – The [C II] – $M_{\rm mol}$ relation from our simulations compared with observations. The simulated galaxies are represented by purple hexbins, where the colour indicates the number of galaxies in each bin. The solid red line is the best-fit: $\log(L_{\rm [C\,II]}/{\rm L}_\odot) = a\log(M_{\rm mol}/{\rm M}_\odot\,{\rm yr}^{-1}) + b$. The fit parameters are reported in each panel, along with the scatter (σ) around the best-fit. Also included is the Spearman's rank correlation coefficient (ρ). The best fit to the observed galaxy sample at $0 \leq z \leq 5.5$ by Zanella et al. (2018) is shown in blue and the fit to the $z \sim 0$ dwarf galaxies (Madden et al., 2020) is shown in lime. The relation from Simba simulations at $z = 6$ (Vizgan et al., 2022) is shown in orange. This figure is a modified version of Figure 7 in Khatri et al. (2024b).

weak at $z \gtrsim 5$ and strengthens over time (as evident from the value of the Spearman correlation coefficient ρ reported in each panel). Moreover, the slope steepens over time. At all redshifts, our best-fit relation is shallower than the relation from Zanella et al. (2018) based on a sample of massive galaxies at $0 \lesssim z \lesssim 5.5$. Consequently, only our high-mass ($M_{\rm mol} \gtrsim 10^9\,{\rm M}_\odot$) galaxies follow their relation, while our low-mass galaxies exhibit higher $L_{\rm [C\,II]}$ than expected from extrapolation of their relation. The apparent discrepancy may arise from the fact that the Zanella et al. (2018) sample is restricted to bright ($L_{\rm [C\,II]} \geq 10^{8.5}\,{\rm L}_\odot$) galaxies at $z \geq 2$.

We further investigate how the conversion factor $\alpha_{\rm [C\,II]} \equiv M_{\rm mol}/L_{\rm [C\,II]}$ depends on other galaxy properties and evolves over time. Figure 5.6 shows $\alpha_{\rm [C\,II]}$ as a function of the gas-phase metallicity 12 + log (O/H) and the SFR averaged over the last 5 Myr ($\rm SFR_5$). For the sake of clarity, we show the median and interquartile range in bins of 12 + log (O/H) and $\rm SFR_5$. We observe that $\alpha_{\rm [C\,II]}$ increases with $\rm SFR_5$ at all redshifts. Similar to Zanella et al. (2018), we find a weak correlation with 12 + log (O/H) at all redshifts and $\alpha_{\rm [C\,II]}$ values spanning nearly two orders of magnitude. The mean $\alpha_{\rm [C\,II]}$ of $31^{+31}_{-15}\,{\rm M}_\odot\,/\,{\rm L}_\odot$ obtained by Zanella et al. (2018) is shown as a gray dotted line and a shaded band in Figure 5.6.

Based on these findings, we devise a prescription for inferring $M_{\rm mol}$ from a measurement of the [C II] luminosity while accounting for secondary dependences in the [C II]-$M_{\rm mol}$ relation. For this, we perform a principal component analysis (PCA) in the 5D parameter space: [$M_{\rm mol}$, $L_{\rm [C\,II]}$, $\rm SFR_5$, $\rm SFR_{200}$ and 12 + log(O/H)]. As PCA identifies dominant

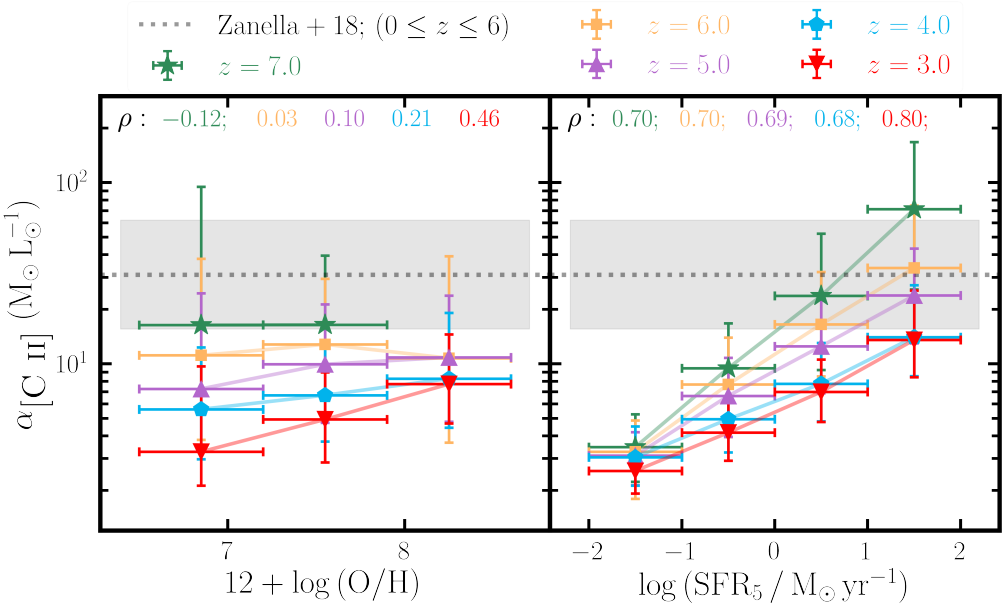


Figure 5.6. – The conversion factor $\alpha_{[\mathrm{C\,II}]}$ in our simulated galaxies as a function of gas metallicity (left) and the SFR averaged over the last 5 Myr (right) at different redshifts. The coloured symbols represent the median $\alpha_{[\mathrm{C\,II}]}$ in each bin, while the error bars represent the 16-84 percentiles. The Spearman's rank correlation coefficient between the variables on the y and x axes are shown in each panel as well. The grey dotted line denotes the mean $\alpha_{[\mathrm{C\,II}]}$ from Zanella et al. (2018) and the shaded band represents the corresponding scatter. This is a modified version of Figure 8 in Khatri et al. (2024b).

patterns and correlations between the parameters, it is often employed to reduce the dimensionality in the data while approximately preserving variance (see e.g., Mannucci et al., 2010; Lara-López et al., 2010; Hunt et al., 2012; Bothwell et al., 2016, for astrophysical applications).

Here, we use PCA to identify the principal component that contributes the least to the total variance. By setting this component to zero, we can obtain a linear scaling between the aforementioned parameters. The results of this analysis at different redshifts are reported in Table 5.1. We report the standard deviation of the difference between the true and predicted M_{mol}. For reference, we also compute this quantity for the linear $M_{\mathrm{mol}} - L_{[\mathrm{C\,II}]}$ relation[5] as well. At all redshifts, we find that the PCA-based relation performs significantly better than the two-variable $M_{\mathrm{mol}} - L_{[\mathrm{C\,II}]}$ relation, such that the former shows an ≈ 2.3 times lower standard deviation.

Even accounting for typical measurement uncertainties in the SFR and metallicity in observations,[6] this relation reliably predicts the true M_{mol} within a factor of 2 for 95% of the galaxies at $3 \leq z \leq 6$. Given that the SFR and metallicity estimates are now becoming

[5]Note that this is different from the best-fit $L_{[\mathrm{C\,II}]} - M_{\mathrm{mol}}$ relation as the dependent and independent variables are swapped.

[6]Here, we adopt an uncertainty of 0.1 dex for $L_{[\mathrm{C\,II}]}$ and 0.24 dex for SFR_5 and SFR_{200}, both based on ALPINE galaxies (Béthermin et al., 2020), and that of 0.05 dex for $12 + \log(\mathrm{O/H})$ (Sanders et al., 2015).

z	a	b	c	d	e	σ (dex)			
						PCA	$M_{\rm mol}$-$L_{\rm [C\,II]}$	PCA*	$M_{\rm mol}$-$L_{\rm [C\,II]}$*
7	0.50	0.79	−0.16	+0.04	4.39	0.20	0.45	0.28	0.47
6	0.42	0.73	−0.02	−0.03	5.46	0.17	0.41	0.24	0.42
5	0.45	0.65	−0.00	−0.03	5.27	0.13	0.36	0.2	0.37
4	0.47	0.59	+0.01	+0.09	4.11	0.13	0.30	0.2	0.32
3	0.57	0.59	−0.06	+0.13	3.12	0.11	0.25	0.19	0.28

Table 5.1. – Coefficients for the PCA-based prescription for estimating the molecular gas mass at different redshifts using the equation: $\log(M_{\rm mol}/{\rm M}_\odot) = a\log(L_{\rm [C\,II]}/{\rm L}_\odot) + b\log({\rm SFR}_5/{\rm M}_\odot\,{\rm yr}^{-1}) + c\log({\rm SFR}_{200}/{\rm M}_\odot\,{\rm yr}^{-1}) + d\,[12+\log({\rm O/H})] + e$. The seventh and eighth columns enlist the standard deviation of the difference between the true and predicted $M_{\rm mol}$ when using the PCA-based and the best fit $M_{\rm mol}$-$L_{\rm [C\,II]}$ relations, respectively, while the last two (denoted by a *) enlist the same when accounting for typical observational uncertainties (see text). Table adapted from Table 5 in Khatri et al. (2024b).

available for a growing sample of high-z galaxies with the James Webb Space Telescope and are expected to increase in the future, this relation can provide more reliable estimates of $M_{\rm mol}$ in [C II]-detected galaxies.

5.5. Extended [C II] emission

Now we examine the relative extent of the [C II] emission and the star formation activity in MARIGOLD galaxies. For this we first consider a 25 kpc sphere centred on a given galaxy and obtain 2D projections of this sphere along three orthogonal lines of sight (these are taken to be the coordinate axes of the simulation box). Then from each projection, we compute the cumulative [C II] surface brightness profile and SFR surface density profiles and calculate the radius containing 90% of the emission and star formation – $r_{90,\,\rm [C\,II]}$ and $r_{90,\,\rm SFR}$ (see the last three panels of Figure 5.7 for an illustration). For the [C II] profile, we additionally compute $r_{70,\,\rm [C\,II]}$ i.e., the radius containing 70% of the [C II] emission. From these values we calculate two parameters: (a) the parameter $\mathcal{R} \equiv r_{90,\,\rm [C\,II]}\,/\,r_{90,\,\rm SFR}$ that quantifies the relative extent of the [C II] emission compared to SF; (b) the multicomponent extent parameter, $\mathcal{E} \equiv r_{90,\,\rm [C\,II]}\,/\,r_{70,\,\rm [C\,II]}$ that measures the spread in the [C II] emission. Additionally, from the 3D distribution of the [C II] emission, we compute the fraction of [C II] luminosity arising from outside the central galaxy (see the left panel of Figure 5.7 for an illustration). We refer to this as the satellite contribution parameter, $\mathcal{S}$.

To understand how the parameter $\mathcal{R}$ is affected by the other two parameters – $\mathcal{E}$ and $\mathcal{S}$, we show in Figure 5.8 the $r_{90,\,\rm [C\,II]}$ versus $r_{90,\,\rm SFR}$ for $z = 5$ galaxies. In this plane, the diagonal line corresponds to $\mathcal{R} = 1$. Each point represents one of the three projections of a galaxy and is colour-coded by the $\mathcal{E}$ value for this projection. The shape of the symbol denotes whether the satellite contribution $\mathcal{S}$ is greater than 10%. We notice that most of the $\mathcal{S} < 0.1$ points (denoted by a x) occupy the region around the diagonal while the pluses ($\mathcal{S} \geq 0.1$) preferentially occupy the region above the $\mathcal{R} = 2$ line. Moreover, the points with low $\mathcal{E}$ mostly occupy the region around the diagonal line while higher $\mathcal{E}$ points

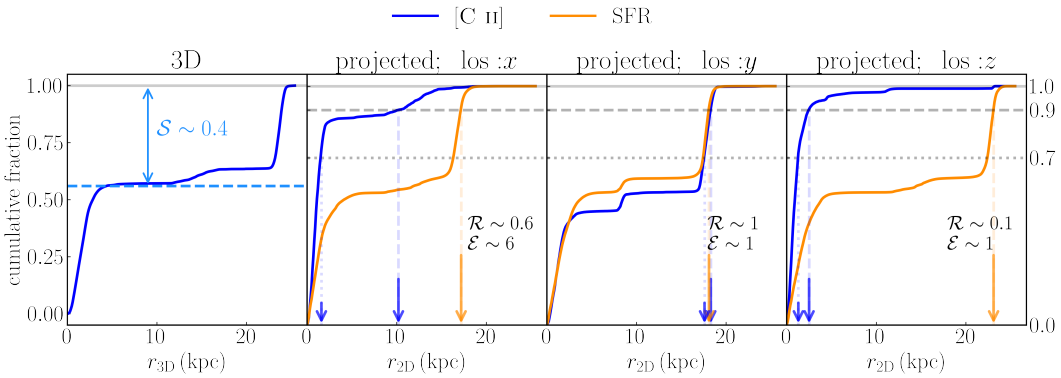


Figure 5.7. – An example illustrating the calculation of the $\mathcal{S}$, $\mathcal{R}$, and $\mathcal{E}$ parameters in a simulated galaxy. The left panel shows the the cumulative profile for the 3D distribution of [C II], with the dashed line marks the contribution of the central galaxy to the total [C II] emission (as evident from the flattening of the cumulative profile). The remaining fraction (denoted by $\mathcal{S}$) represents the contribution of satellites. The other panels show cumulative profile constructed from the [C II] surface brightness (blue) and SFR surface density (orange). For the profiles obtained from projections, the value of the $\mathcal{R}$ and $\mathcal{E}$ parameters are indicated in each panel. In all but the leftmost panel, the dotted, dashed and solid horizontal lines denote cumulative fractions of 70%, 90%, and 100%, respectively. The small and large blue arrows mark $r_{70,\,[\mathrm{C\,II}]}$ and $r_{90,\,[\mathrm{C\,II}]}$, respectively and the the orange arrow denotes $r_{90,\,\mathrm{SFR}}$. The parameter $\mathcal{R} \equiv r_{90,\,[\mathrm{C\,II}]}/r_{90,\mathrm{SFR}}$ is calculated from the ratio of the r values denoted by the large blue and orange arrows, while the parameter $\mathcal{E} \equiv r_{90,\,[\mathrm{C\,II}]}/r_{70,\,[\mathrm{C\,II}]}$ from the ratio of the large and small blue arrows. In a real observation, we can only observe a galaxy from a single line of sight. However, as we can see from Figure 5.7, the inferred extent of the [C II] emission relative to the SFR (quantified by the $\mathcal{R}$ parameter) is sensitive to the line of sight.

reside above the $\mathcal{R} = 2$ line. Overall, we find that systems with $\mathcal{R} \geq 2$ (i.e., where the [C II] emission is at least twice as extended as star formation), have preferentially higher values of the parameters $\mathcal{E}$ and $\mathcal{S}$ compared to galaxies with $\mathcal{R} < 2$. The median $\mathcal{E}$ values for the two subsamples are 3.99 and 1.5, respectively, and the median $\mathcal{S}$ values are 0.17 and 0.03.

5.6. Conclusions

The [C II] line is a powerful probe of galaxy properties such as the SFR and the molecular gas mass in galaxies. In this work, we examined several aspects of the [C II] line at $3 \leq z \leq 7$ using the MARIGOLD simulations. Our key findings can be summarized as:

1. The [C II] LF shows a strong z-evolution, with the number density of $L_{[\mathrm{C\,II}]} \sim 10^9\,\mathrm{L}_\odot$ emitters increases by 600 times in the above redshift range. The faint-end slope α tends to flatten at late times ($z \leq 4$).

2. The slope of [C II]-SFR relation shows a mild evolution with redshift, while the intercept increases by 0.5 dex between $z = 7$ and $z = 3$.

3. The strength of the [C II]-M_{mol} correlation increases over time.

4. At all redshifts, the conversion factor $\alpha_{[\mathrm{C\,II}]}$ increases with the star formation rate averaged over the last five Myr.

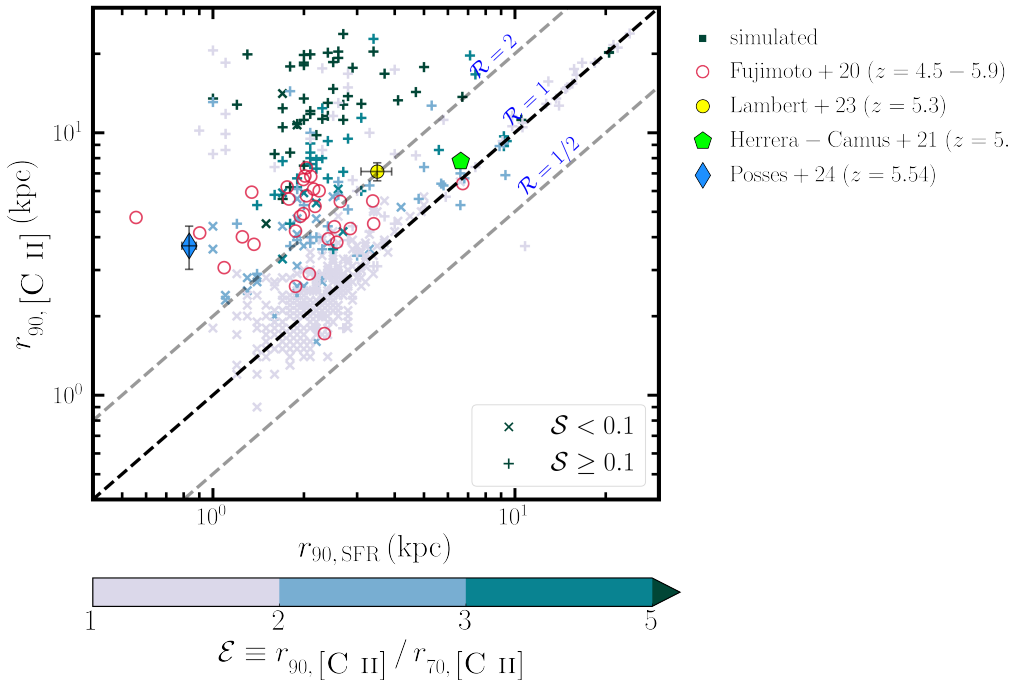


Figure 5.8. – Comparison of $r_{90,\,[\mathrm{C\ II}]}$ and $r_{90,\,\mathrm{SFR}}$ for simulated galaxies at $z = 5$ (left). The galaxies are colour-coded by their multicomponent extent parameter $\mathcal{E}$ defined as the ratio of the r_{90} and r_{70} values of the [C II] surface brightness profile. The shape of the symbol reflects the $\mathcal{S}$ parameter that quantifies the satellite contribution to the total [C II] emission (see text for details). We use crosses (pluses) to denote galaxies with $< 10\%$ ($\geq 10\%$) satellite contribution. The $r_{90,\,[\mathrm{C\ II}]}$ versus $r_{90,\,\mathrm{SFR}}$ values of observed galaxies are shown as red open circles (Fujimoto et al., 2020), a yellow plus (Lambert et al., 2023), a green pentagon (Herrera-Camus et al., 2021), and a blue diamond (Posses et al., 2024). The black dashed line indicates a 1:1 relation, while the top and bottom grey dashed lines indicate 2:1 and 1:2 relations, respectively. Note that the error bars are not shown for Fujimoto et al. (2020) galaxies for the sake of clarity.

5. Our 5-variable PCA-based prescription performs better at predicting the true molecular gas mass compared to the two-variable $M_{\mathrm{mol}} - L_{[\mathrm{C\,II}]}$ relation, improving the precision of molecular gas mass estimates by a factor of ~ 2.3.

6. Galaxies with a more extended [C II] emission than their star formation activity tend to have a higher contribution of satellite galaxies to the total [C II] luminosity as well as a more spread out distribution of the [C II] emission.

CHAPTER 6

Conclusions

6.1. Summary

Observations of the molecular ISM in distant galaxies have opened up an exciting avenue, enabling astronomers to measure the molecular gas content – the fuel for star formation – in an ever-increasing number of typical galaxies, out to the earliest epochs, when the Universe was barely a billion years old (just 7% of its present age). However, with new data come new challenges. Current methods for probing molecular gas in galaxies have been rigorously calibrated in local galaxies (Bolatto et al., 2013; Scoville et al., 2016; Hughes et al., 2017), but their reliability at higher redshifts remains uncertain. With the frontier of molecular ISM studies now being at $z \gtrsim 4$, there is an urgent need to test these calibrations. Unfortunately, the small sample sizes of current observations pose significant challenges. This thesis addressed this issue from a numerical perspective. A schematic representation of the work carried out in this thesis is shown in Figure 6.1.

In **Chapter 3**, we developed a sub-grid model called Hyacinth that can be embedded into cosmological simulations to track the time-dependent abundances of H_2, CO, C and C^+ within galaxies. At the heart of this model is a variable sub-grid density PDF and a temperature-density relation calibrated on high-resolution molecular-cloud simulations. The model captures the effect of microscopic (i.e., unresolved) density fluctuations on the macroscopic (i.e., on resolved scales) chemical abundances.

The model has been benchmarked against two widely used, extensive chemical networks from the literature, a more comprehensive PDR code, and high-resolution molecular cloud simulations with different metallicities. We applied Hyacinth in post-processing to a pre-simulated galaxy, comparing its predicted chemical abundances to observational data from both nearby and high-redshift galaxies. Overall, the chemical abundances from Hyacinth show a remarkable agreement with these observations.

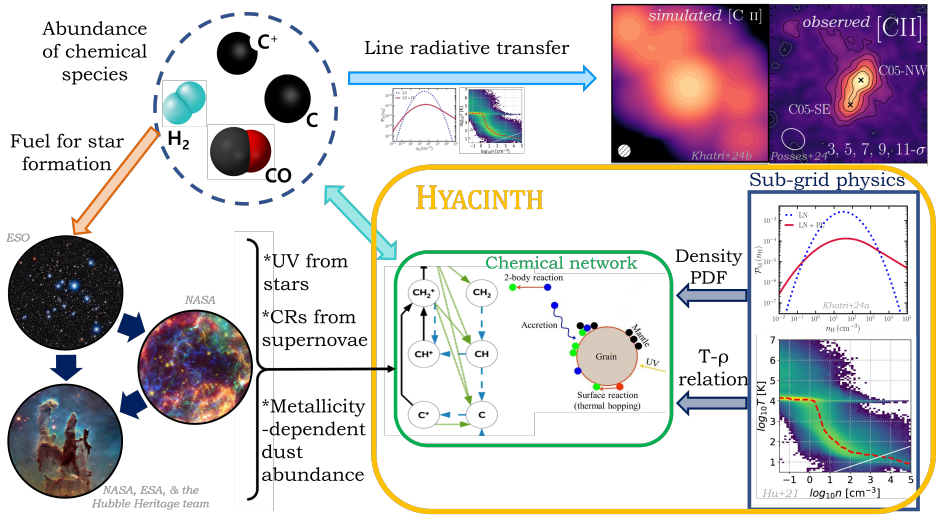


Figure 6.1. – A schematic representation of the work carried out in this thesis.

In essence, by incorporating insights from high-resolution simulations of molecular clouds, HYACINTH represents a significant step towards bridging the gap between ISM-scale and galaxy-scale simulations.

In **Chapter 4**, we incorporated HYACINTH into the adaptive mesh refinement code RAMSES to perform hydrodynamical simulations of galaxy formation in a cosmological context – the MARIGOLD simulations. These simulations predict that at $z < 5$, the evolution of the cosmic star formation rate density (SFRD) mirrors that of the H_2 density, while at $z \geq 5$, the former exhibits a steeper decline. Additionally, we quantified the contribution of low-mass ($M_{H_2} < 10^8\,M_\odot$) galaxies to the cosmic H_2 budget and found that these galaxies contain nearly half of the cosmic H_2 at $z \gtrsim 6$. However, the current sensitivity of instruments like ALMA renders these galaxies "invisible", indicating a potential underestimation of the cosmic H_2 density in current surveys.

We further examined how the molecular gas fraction (i.e., the fraction of the total gas mass in molecular form) f_{mol} varies across galaxies of different stellar masses and evolves over time. We found that on galaxy-wide scales, f_{mol} remains less than 50% at stellar masses $M_* \leq 10^{11}\,M_\odot$ and $z \geq 3$, thereby refuting the common assumption in observational studies of the gas being fully molecular, particularly for massive ($M_* \gtrsim 5 \times 10^{10}\,M_\odot$) galaxies at $z \gtrsim 0.4$ (Tacconi et al., 2018).

In **Chapter 5**, we investigated the reliability of the [C II] fine-structure line as a molecular gas tracer in galaxies across $3 \leq z \leq 7$. Our results indicate that the $L_{[C\,II]} - M_{mol}$ relation is relatively weak at $z \gtrsim 5$ but strengthens over time. This likely reflects the early stages of galaxy evolution at high redshifts, where galaxies are still building their molecular gas reservoirs while simultaneously enriching their interstellar medium with metals. We propose that it takes time for a stable and tight relationship to emerge between $L_{[C\,II]}$ and M_{mol}. In a follow-up study, we would test this hypothesis by tracking the evolution of the [C II] emission and molecular gas content in individual galaxies over time.

Across redshifts, we found a strong positive correlation between the [C II]-to-$M_{\rm mol}$ conversion factor, $\alpha_{\rm [C\,II]}$, and the star-formation rate (SFR) of galaxies. At high SFRs ($\gtrsim 10\,{\rm M}_\odot\,{\rm yr}^{-1}$), the $\alpha_{\rm [C\,II]}$ values for MARIGOLD galaxies agree well with the widely-used calibration from Zanella et al. (2018), derived for a sample of $0 \leq z \leq 5.5$ galaxies with SFR $\geq 35\,{\rm M}_\odot\,{\rm yr}^{-1}$. However, deviations emerge at lower SFRs, which implies that using a constant $\alpha_{\rm [C\,II]}$ (calibrated on the high-SFR galaxies alone) would overestimate the molecular gas mass in low-SFR galaxies. This highlights the importance of accounting for variations in $\alpha_{\rm [C\,II]}$ across different galaxy populations. Furthermore, we demonstrated that incorporating secondary dependencies in the [C II]-$M_{\rm mol}$ relation improves molecular gas mass predictions by a factor of 2.3 across all redshifts.

Finally, zooming out of individual galaxies, we examined how the [C II] luminosity function evolves over time. We found that at $z = 3$, faint galaxies with $L_{\rm [C\,II]} < 10^7\,{\rm L}_\odot$[1] contribute roughly a quarter of the total [C II] luminosity. By $z \gtrsim 6$, this fraction rises significantly, with faint galaxies producing over half of the cosmic [C II] emission. This emphasizes the crucial role of faint galaxies in shaping the [C II] luminosity function at high redshifts.

6.2. Outlook

6.2.1. Scope for improvement

While the focus of this thesis was to enable hydrogen and carbon chemistry calculations on the fly in large-scale numerical simulations, the simulations can greatly benefit from improving the implementations of the other physical processes, of which, we review four key improvements: **A)** *Variable metal yields:* HYACINTH currently assumes constant metal yields (e.g., a fixed fraction of the metals is assumed to be present in elemental carbon), which could be improved by incorporating more detailed chemical enrichment histories where the fraction varies with the metallicity and/or the age of the Universe. **B)** *Improved radiative transfer modelling*: Allowing UV radiation to propagate from a given stellar particle in a spherically-symmetric manner (i.e., as $\propto r^{-2}$) instead of propagating in an octet as is presently done. **C)** *Coupling cooling to chemistry*: the numerical implementation of cooling can be improved by including molecular line and [C II] line cooling based on the non-equilibrium abundances. **D)** *Follow dust growth and evolution*: HYACINTH assumes a metallicity-dependent dust-to-gas ratio calibrated on observations of $0 < z < 5$ galaxies, but a more accurate treatment of dust in the simulations would improve the accuracy as dust plays a crucial role in H_2 chemistry. In addition, this would also allow for the production of dust continuum maps post-processing and the testing of the reliability of dust continuum emission as a molecular gas tracer.

6.2.2. Future studies

Zooming in on the high-z galaxies: In this thesis, we explored the [C II]-$M_{\rm mol}$ relation on galaxy-wide scales, investigating its dependence on various galaxy properties. However, the high spatial resolution of the MARIGOLD simulations enables us to extend this analysis to resolved scales within individual galaxies. As a preliminary example, Figure 6.2 shows

[1]The threshold of $10^7\,{\rm L}_\odot$ corresponds to the typical sensitivity limit of current surveys at $z \gtrsim 4$, such as ALPINE and REBELS.

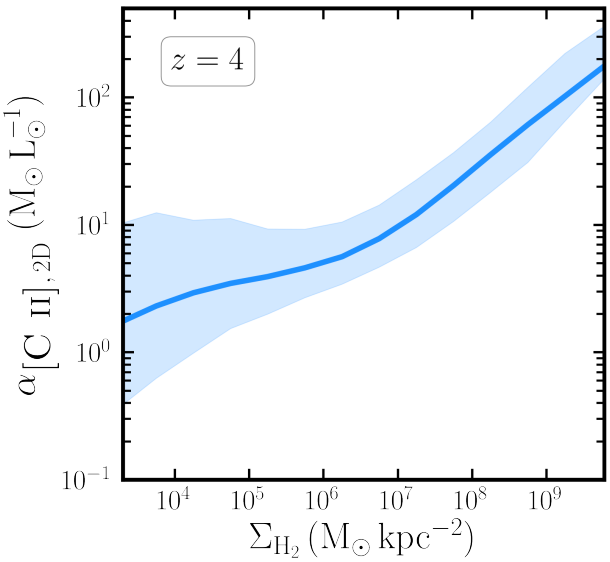


Figure 6.2. – The spatially-resolved $\alpha_{[\mathrm{C\,II}],\,2\mathrm{D}} \equiv 1.36\,\Sigma_{\mathrm{H_2}}/\Sigma_{[\mathrm{C\,II}]}$ versus the $\mathrm{H_2}$ surface density in MARIGOLD galaxies at $z = 4$ versus. The solid line shows the median $\alpha_{[\mathrm{C\,II}],\,2\mathrm{D}}$ in a given $\Sigma_{\mathrm{H_2}}$ while the shaded area represents the interquartile range. The [C II] surface brightness and $\mathrm{H_2}$ surface densities are obtained from a face-on projection of a cube centred on the galaxy and of side length equal to twice the size of the galaxy (same as in Figure 6 of Khatri et al. 2024b).

the median $\alpha_{[\mathrm{C\,II}],\,2\mathrm{D}}$ as a function of the $\mathrm{H_2}$ surface density across the simulated galaxy population at $z = 4$. For this analysis, we first obtain the [C II] surface brightness and $\mathrm{H_2}$ surface density maps for each galaxy from the face-on projection of a cube centered on the galaxy, with a side length twice the size of the galaxy. To match the resolution of current high-redshift [C II] observations capable of resolving kpc-scale regions within $z \gtrsim 4$ galaxies (e.g., Posses et al., 2024), we apply a 2D Gaussian smoothing to the simulated maps. For this, we adopt a beam size (in terms of the full-width at half-maximum, FWHM) of 0.17 arcsec (same as in Posses et al., 2024), which at $z = 4$, corresponds to ~ 1.2 kpc.

From the figure, we see that $\alpha_{[\mathrm{C\,II}],\,2\mathrm{D}}$ gradually increases with $\Sigma_{\mathrm{H_2}}$ and shows almost two orders of magnitude variation. This shows that the increase in the [C II] surface density towards higher densities is not commensurate with the rise in the $\mathrm{H_2}$ surface density.

Kiloparsec-scale stellar clumps are ubiquitously observed in high-redshift ($z \gtrsim 2$) galaxies (Förster Schreiber et al., 2011; Wuyts et al., 2012). Spatially-resolved observations of the gas and star formation in these galaxies are crucial for understanding their internal structure and assembly history. To fully exploit these observations, however, it is crucial to understand how the scaling relations between observables (such as infrared line luminosities) and ISM properties (such as gas mass, star formation rate) behave on spatially resolved scales. Figure 6.3 illustrates how the [C II]-to-$\mathrm{H_2}$ conversion factor varies on kpc scales across a pair of interacting galaxies from the MARIGOLD simulations at $z = 4$ and underscores the potential risk of applying empirical relations calibrated on galaxy-wide scales to resolved observations.

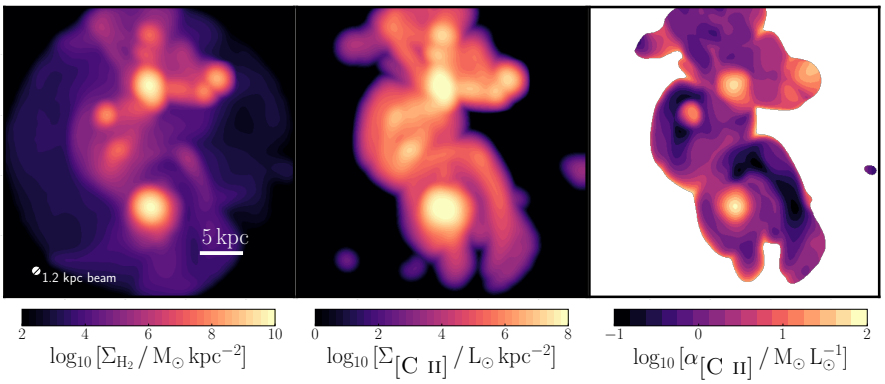


Figure 6.3. – Figure illustrating the variation of the [C II]-to-H_2 conversion factor (right panel) on $\sim$ kpc scales in a pair of interacting galaxies from the MARIGOLD simulations at $z = 4$.

[C II] deficit: Both the [C II] line and the FIR dust continuum emission are important cooling channels in the ISM. The [C II]/FIR[2] ratio quantifies the relative contribution of [C II] line cooling compared to dust cooling. Early observations of nearby (ultra) luminous infrared galaxies[3] (ULIRGS/LIRGS; Malhotra et al., 1997, 2001; Luhman et al., 1998, 2003; Graciá-Carpio et al., 2011) found that the [C II]/FIR ratio decreases with increasing total infrared luminosity ($L_{\rm IR}$). Normal star-forming galaxies typically have a [C II]/FIR $\gtrsim 4 \times 10^{-3} - 10^{-2}$, while for LIRGs/ULIRGs, it is $\lesssim 10^{-3}$ and can be as low as few times 10^{-4} (Graciá-Carpio et al., 2011). This phenomenon is observed at low and high redshifts alike and is commonly referred to as the "[C II] line deficit" or simply the "[C II] deficit". The origin of this deficit remains an open question in the field and possible explanations include saturation of the [C II] line, optical depth effects, a high ionization parameter, or a reduced neutral gas fraction, among others.

Using the MARIGOLD simulations, we can explore the cause of the [C II] deficit by calculating the FIR emission from these galaxies. As a preliminary step, we approximate the FIR emission as detailed below and examine the [C II]/FIR ratio in these galaxies to gain an initial insight into this problem.

To obtain the $L_{\rm IR}$ of our galaxies, we adopt the following scaling relation between the total IR luminosity and the star formation rate from Hao et al. (2011); Murphy et al. (2011):

$$\log_{10}\left(\frac{L_{\rm IR}}{\rm erg\,s^{-1}}\right) = \log_{10}\left(\frac{\rm SFR}{\rm M_\odot\,yr^{-1}}\right) + 43.41\,, \tag{6.1}$$

where the SFR is an average over the last 100 Myr (see e.g., Table A.1 in De Looze et al., 2014). The FIR luminosity is obtained from $L_{\rm IR}$ following (Calzetti et al., 2000, also see Herrera-Camus et al. 2018):

$$L_{\rm FIR} = \frac{1.0}{1.75}\,L_{\rm IR}\,. \tag{6.2}$$

[2]Here FIR represents the total FIR luminosity in the wavelength range $42.5 - 122.4\,\mu$m (Helou et al., 1988).

[3]LIRGs have $10^{11}\,{\rm L_\odot} \lesssim L_{\rm IR} \lesssim 10^{12}\,{\rm L_\odot}$, while ULIRGs have $L_{\rm IR} \gtrsim 10^{12}\,{\rm L_\odot}$.

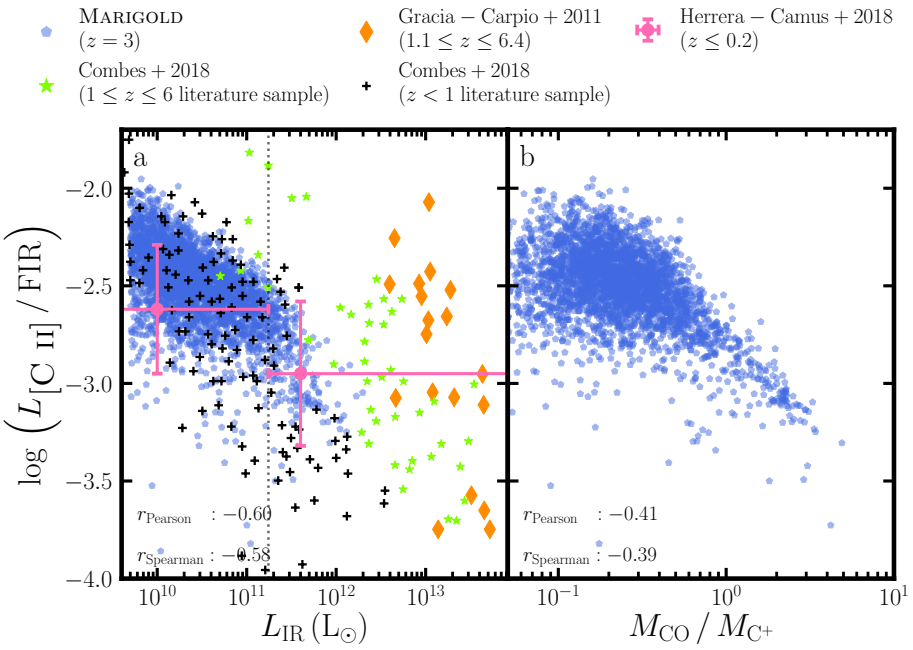


Figure 6.4. – The [C II]/FIR ratio of our simulated galaxies at $z = 3$ as a function of the total IR luminosity (panel a) and the ratio of CO to C^+ mass of the galaxy (panel b). The orange diamonds are from the $1 \leq z \leq 6.6$ sample of Graciá-Carpio et al. (2011), the black pluses and green stars are the low-redshift ($z < 1$) and high-redshift ($1 \leq z \leq 6$) data compiled by Combes (2018), and the pink data points with error bars denote the median in different bins from the sample of $z \leq 0.2$ main-sequence galaxies from Herrera-Camus et al. (2018).

Previous studies (e.g., Calzetti et al., 2010; Kennicutt & Evans, 2012b; Madden et al., 2013) have shown that the total IR emission is an excellent proxy of the star formation rate in galaxies with $L_{\mathrm{IR}} \gtrsim 5 \times 10^9\,\mathrm{L}_\odot$. Therefore we apply Equation (6.1) to all simulated galaxies with $L_{\mathrm{IR}} \gtrsim 5 \times 10^9\,\mathrm{L}_\odot$ or equivalently $\mathrm{SFR} \gtrsim 0.75\,\mathrm{M}_\odot\,\mathrm{yr}^{-1}$.

The left panel of Figure 6.4 shows the [C II]/FIR ratio versus L_{IR} in our galaxies at $z = 3$. We also show the observed high-redshift ($1.1 \leq z \leq 6.4$) galaxy compilation Graciá-Carpio et al. (2011),[4] and the literature sample compiled by Combes (2018) at $z < 1$ and $z \geq 1$. We also include the mean $L_{[\mathrm{C\,II}]}/L_{\mathrm{FIR}}$ ratio in different L_{IR} bins for $z < 0.2$ galaxies from Herrera-Camus et al. (2018). Our simulated galaxies show a similar decline in [C II]/FIR towards high L_{IR} as reported in observations.

To investigate the cause of this decline, in the right panel of Figure 6.4, we plot the [C II]/FIR ratio versus the CO/C^+ mass ratio in our simulated galaxies. We find that [C II]/FIR is nearly constant albeit with a large scatter for $M_{\mathrm{CO}}/M_{\mathrm{C^+}} \lesssim 0.4$ while a gradual decline is observed for higher $M_{\mathrm{CO}}/M_{\mathrm{C^+}}$ values. Our findings indicate that galaxies with high infrared luminosities have a higher fraction of their total carbon in the form of CO. Thus, we predict that the [C II] deficit in LIRGs and ULIRGs results from the physical conditions prevalent in these galaxies, particularly high densities and metallicities, that

[4] Graciá-Carpio et al. (2011) report the L_{FIR} for these galaxies from which we obtain L_{IR} using Eq. (6.2).

favour the conversion of C^+ into CO. This preliminary analysis looks promising and the cause of [C II] deficit can be explored further.

Multi-line studies of the ISM: While this thesis focused on testing the reliability of [C II] as a molecular gas tracer, CO rotational lines – particularly the $J = 1 \rightarrow 0$ transition – remain one of the most commonly used tracers of molecular gas, especially at $z \lesssim 3$ (see Section 2.4.1). The variation of the CO$(1-0)$-to-H_2 conversion factor, $\alpha_{\rm CO}$, with physical conditions such as metallicity, and stellar surface density, among others, has been well-studied in the Milky Way and nearby galaxies (Bolatto et al., 2013). However, at higher redshifts, doing so becomes challenging due to the limited number of galaxies observed in CO, the inaccessibility of the CO $J = 1 \rightarrow 0$ transition, and the poor constraints on CO excitation conditions.

To provide a theoretical insight on these, we can perform the radiative transfer of CO lines in the MARIGOLD galaxies. By building templates of the CO Spectral Line Energy Distribution (SLED) for a statistically significant sample of simulated galaxies with varied ISM conditions, we can investigate the key galaxy properties influencing the excitation of the CO rotational levels at different cosmic epochs.

Moreover, low-metallicity dwarf galaxies are known to contain large amounts of CO-dark molecular gas (Wolfire et al., 2010; Madden et al., 2020) Building upon the work carried out in this thesis on the [C II] line as a molecular gas tracer, we can explore how CO and [C II] lines complement each other, potentially providing a more comprehensive view of the high-redshift ISM.

The Road goes ever on and on
Down from the door where it began.
Now far ahead the Road has gone,
And I must follow, if I can,
Pursuing it with eager feet,
Until it joins some larger way
Where many paths and errands meet.
And whither then? I cannot say.

- J.R.R. Tolkien

Bibliography

Accurso G., et al., 2017, MNRAS, 470, 4750

Agertz O., Kravtsov A. V., Leitner S. N., Gnedin N. Y., 2013, ApJ, 770, 25

Alaghband-Zadeh S., et al., 2013, MNRAS, 435, 1493

Algera H. S. B., et al., 2023, MNRAS, 518, 6142

Angulo R. E., Hahn O., 2022, Living Reviews in Computational Astrophysics, 8, 1

Aravena M., et al., 2024, A&A, 682, A24

Baes M., Verstappen J., De Looze I., Fritz J., Saftly W., Vidal Pérez E., Stalevski M., Valcke S., 2011, ApJS, 196, 22

Barkana R., Loeb A., 2001, Phys. Rep., 349, 125

Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57

Béthermin M., et al., 2020, A&A, 643, A2

Béthermin M., et al., 2023, A&A, 680, L8

Bigiel F., Leroy A., Walter F., Brinks E., de Blok W. J. G., Madore B., Thornley M. D., 2008, AJ, 136, 2846

Bigiel F., Leroy A., Walter F., Blitz L., Brinks E., de Blok W. J. G., Madore B., 2010, AJ, 140, 1194

Birnboim Y., Dekel A., 2003, MNRAS, 345, 349

Bisbas T. G., Papadopoulos P. P., Viti S., 2015, ApJ, 803, 37

Bisbas T. G., van Dishoeck E. F., Papadopoulos P. P., Szűcs L., Bialy S., Zhang Z.-Y., 2017, ApJ, 839, 90

Blain A. W., Smail I., Ivison R. J., Kneib J. P., Frayer D. T., 2002, Phys. Rep., 369, 111

Bolatto A. D., Wolfire M., Leroy A. K., 2013, ARA&A, 51, 207

Boogaard L. A., et al., 2020, ApJ, 902, 109

Bothwell M. S., Maiolino R., Cicone C., Peng Y., Wagg J., 2016, A&A, 595, A48

Bouché N., et al., 2007, ApJ, 671, 303

Bouwens R. J., et al., 2022, ApJ, 931, 160

Bouwens R., Illingworth G., Oesch P., Stefanon M., Naidu R., van Leeuwen I., Magee D., 2023a, MNRAS, 523, 1009

Bouwens R. J., et al., 2023b, MNRAS, 523, 1036

Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, MNRAS, 351, 1151

Calzetti D., 2013, in Falcón-Barroso J., Knapen J. H., eds, , Secular Evolution of Galaxies. p. 419, doi:10.48550/arXiv.1208.2997

Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682

Calzetti D., et al., 2005, ApJ, 633, 871

Calzetti D., et al., 2007, ApJ, 666, 870

Calzetti D., et al., 2010, ApJ, 714, 1256

Carilli C. L., Walter F., 2013, ARA&A, 51, 105

Carniani S., Maiolino R., Smit R., Amorín R., 2018, ApJ, 854, L7

Carroll S. M., 2004, Spacetime and Geometry. Cambridge University Press

Casey C. M., Narayanan D., Cooray A., 2014, Phys. Rep., 541, 45

Catán V., et al., 2024, arXiv e-prints, p. arXiv:2408.07849

Chevance M., et al., 2020, MNRAS, 493, 2872

Chevance M., Krumholz M. R., McLeod A. F., Ostriker E. C., Rosolowsky E. W., Sternberg A., 2023, in Inutsuka S., Aikawa Y., Muto T., Tomida K., Tamura M., eds, Astronomical Society of the Pacific Conference Series Vol. 534, Protostars and Planets VII. p. 1 (arXiv:2203.09570), doi:10.48550/arXiv.2203.09570

Christensen C., Quinn T., Governato F., Stilp A., Shen S., Wadsley J., 2012, MNRAS, 425, 3058

Chung D. T., et al., 2024, A&A, 691, A337

Clark P. C., Glover S. C. O., 2015, MNRAS, 452, 2057

Clarke L., Shapley A. E., Sanders R. L., Topping M. W., Brammer G. B., Bento T., Reddy N. A., Kehoe E., 2024, arXiv e-prints, p. arXiv:2406.05178

Combes F., 2018, A&A Rev., 26, 5

Courty S., Alimi J. M., 2004, A&A, 416, 875

Crain R. A., van de Voort F., 2023, ARA&A, 61, 473

Curti M., et al., 2024, A&A, 684, A75

Daddi E., et al., 2007, ApJ, 670, 156

Daddi E., et al., 2010a, ApJ, 713, 686

Daddi E., et al., 2010b, ApJ, 714, L118

Dalgarno A., 2006, Proceedings of the National Academy of Science, 103, 12269

Davé R., Thompson R., Hopkins P. F., 2016, MNRAS, 462, 3265

Davé R., Crain R. A., Stevens A. R. H., Narayanan D., Saintonge A., Catinella B., Cortese L., 2020, MNRAS, 497, 146

De Looze I., Baes M., Bendo G. J., Cortese L., Fritz J., 2011, MNRAS, 416, 2712

De Looze I., et al., 2014, A&A, 568, A62

Decarli R., et al., 2014, ApJ, 782, 78

Decarli R., et al., 2016, ApJ, 833, 70

Decarli R., et al., 2019, ApJ, 882, 138

Dekel A., Birnboim Y., 2006, MNRAS, 368, 2

Dessauges-Zavadsky M., et al., 2020, A&A, 643, A5

Dobbs C. L., Glover S. C. O., Clark P. C., Klessen R. S., 2008, MNRAS, 389, 1097

Dodelson S., 2003, Modern Cosmology

Donnan C. T., et al., 2022, arXiv e-prints, p. arXiv:2207.12356

Downes D., Solomon P. M., 1998, ApJ, 507, 615

Draine B. T., 1978, ApJS, 36, 595

Draine B. T., 2003, ApJ, 598, 1017

Draine B. T., Bertoldi F., 1996, ApJ, 468, 269

Dullemond C. P., Juhasz A., Pohl A., Sereshti F., Shetty R., Peters T., Commercon B., Flock M., 2012, RADMC-3D: A multi-purpose radiative transfer tool, Astrophysics Source Code Library, record ascl:1202.015

Dunne L., Maddox S. J., Papadopoulos P. P., Ivison R. J., Gomez H. L., 2022, MNRAS, 517, 962

Einstein A., 1915, Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften, pp 844–847

Elbaz D., et al., 2007, A&A, 468, 33

Elmegreen B. G., 1989, ApJ, 338, 178

Elmegreen B. G., 2007, ApJ, 668, 1064

Endres C. P., Schlemmer S., Schilke P., Stutzki J., Müller H. S. P., 2016, Journal of Molecular Spectroscopy, 327, 95

Evans Neal J. I., et al., 2009, ApJS, 181, 321

Fabian A. C., 2012, ARA&A, 50, 455

Faisst A. L., et al., 2022, Universe, 8, 314

Federrath C., Klessen R. S., 2013, ApJ, 763, 51

Feldmann R., Gnedin N. Y., Kravtsov A. V., 2012a, ApJ, 747, 124

Feldmann R., Gnedin N. Y., Kravtsov A. V., 2012b, ApJ, 758, 127

Ferland G. J., Korista K. T., Verner D. A., Ferguson J. W., Kingdon J. B., Verner E. M., 1998, PASP, 110, 761

Fixsen D. J., Cheng E. S., Gales J. M., Mather J. C., Shafer R. A., Wright E. L., 1996, ApJ, 473, 576

Foreman-Mackey D., et al., 2013, emcee: The MCMC Hammer, Astrophysics Source Code Library, record ascl:1303.002

Förster Schreiber N. M., Wuyts S., 2020, ARA&A, 58, 661

Förster Schreiber N. M., et al., 2011, ApJ, 739, 45

Frayer D. T., Ivison R. J., Scoville N. Z., Yun M., Evans A. S., Smail I., Blain A. W., Kneib J. P., 1998, ApJ, 506, L7

Frayer D. T., et al., 1999, ApJ, 514, L13

Freundlich J., et al., 2013, A&A, 553, A130

Freundlich J., et al., 2019, A&A, 622, A105

Fu J., Guo Q., Kauffmann G., Krumholz M. R., 2010, MNRAS, 409, 515

Fujimoto S., et al., 2020, ApJ, 900, 1

Gehrels N., 1986, ApJ, 303, 336

Genzel R., et al., 2010, MNRAS, 407, 2091

Genzel R., et al., 2012, ApJ, 746, 69

Girichidis P., Konstandin L., Whitworth A. P., Klessen R. S., 2014, ApJ, 781, 91

Glover S. C. O., 2003, ApJ, 584, 331

Glover S. C. O., Clark P. C., 2012a, MNRAS, 421, 9

Glover S. C. O., Clark P. C., 2012b, MNRAS, 421, 116

Glover S. C. O., Clark P. C., 2016, MNRAS, 456, 3596

Glover S. C. O., Mac Low M.-M., 2007, ApJ, 659, 1317

Glover S. C. O., Mac Low M. M., 2011, MNRAS, 412, 337

Glover S. C. O., Federrath C., Mac Low M. M., Klessen R. S., 2010, MNRAS, 404, 2

Glover S. C. O., Clark P. C., Micic M., Molina F., 2015, MNRAS, 448, 1607

Gnedin N. Y., 2000, ApJ, 542, 535

Gnedin N. Y., Kravtsov A. V., 2010, ApJ, 714, 287

Gnedin N. Y., Kravtsov A. V., 2011, ApJ, 728, 88

Gnedin N. Y., Tassis K., Kravtsov A. V., 2009, ApJ, 697, 55

Gong M., Ostriker E. C., Wolfire M. G., 2017, ApJ, 843, 38

Graciá-Carpio J., et al., 2011, ApJ, 728, L7

Gruppioni C., et al., 2020, A&A, 643, A8

Gullberg B., et al., 2015, MNRAS, 449, 2883

Haardt F., Madau P., 2001, in Neumann D. M., Tran J. T. V., eds, Clusters of Galaxies and the High Redshift Universe Observed in X-rays. p. 64 (arXiv:astro-ph/0106018), doi:10.48550/arXiv.astro-ph/0106018

Habing H. J., 1968, Bull. Astron. Inst. Netherlands, 19, 421

Hahn O., Abel T., 2011, MNRAS, 415, 2101

Hao C.-N., Kennicutt R. C., Johnson B. D., Calzetti D., Dale D. A., Moustakas J., 2011, ApJ, 741, 124

Harikane Y., et al., 2022a, ApJS, 259, 20

Harikane Y., et al., 2022b, ApJ, 929, 1

Harrington K. C., et al., 2021, ApJ, 908, 95

Heiderman A., Evans Neal J. I., Allen L. E., Huard T., Heyer M., 2010, ApJ, 723, 1019

Helou G., Khan I. R., Malek L., Boehmer L., 1988, ApJS, 68, 151

Henríquez-Brocal K., et al., 2022, A&A, 657, L15

Hernquist L., 1989, Nature, 340, 687

Herrera-Camus R., et al., 2015, ApJ, 800, 1

Herrera-Camus R., et al., 2018, ApJ, 861, 95

Herrera-Camus R., et al., 2021, A&A, 649, A31

Hodge J. A., da Cunha E., 2020, Royal Society Open Science, 7, 200556

Hodge J. A., Riechers D., Decarli R., Walter F., Carilli C. L., Daddi E., Dannerbauer H., 2015, ApJ, 798, L18

Hollenbach D., Kaufman M. J., Neufeld D., Wolfire M., Goicoechea J. R., 2012, ApJ, 754, 105

Hopkins P. F., Kereš D., Oñorbe J., Faucher-Giguère C.-A., Quataert E., Murray N., Bullock J. S., 2014, MNRAS, 445, 581

Hu C.-Y., Sternberg A., van Dishoeck E. F., 2021, ApJ, 920, 44

Hu C.-Y., Sternberg A., van Dishoeck E. F., 2023, ApJ, 952, 140

Hughes T. M., et al., 2017, A&A, 602, A49

Hunt L., et al., 2012, MNRAS, 427, 906

Ikeda M., Oka T., Tatematsu K., Sekimoto Y., Yamamoto S., 2002, ApJS, 139, 467

Indriolo N., McCall B. J., 2013, Chemical Society Reviews, 42, 7763

Kainulainen J., Beuther H., Henning T., Plume R., 2009, A&A, 508, L35

Karakas A. I., 2010, MNRAS, 403, 1413

Katz H., et al., 2022, MNRAS, 510, 5603

Keating G. K., Marrone D. P., Bower G. C., Leitch E., Carlstrom J. E., DeBoer D. R., 2016, ApJ, 830, 34

Keating G. K., Marrone D. P., Bower G. C., Keenan R. P., 2020, ApJ, 901, 141

Kennicutt Robert C. J., 1998, ApJ, 498, 541

Kennicutt R. C., Evans N. J., 2012a, ARA&A, 50, 531

Kennicutt R. C., Evans N. J., 2012b, ARA&A, 50, 531

Kereš D., Katz N., Weinberg D. H., Davé R., 2005, MNRAS, 363, 2

Khatri P., Porciani C., Romano-Díaz E., Seifried D., Schäbe A., 2024a, arXiv e-prints, p. arXiv:2402.11023

Khatri P., Romano-Díaz E., Porciani C., 2024b, arXiv e-prints, p. arXiv:2411.09755

Khusanova Y., et al., 2021, A&A, 649, A152

Kistler M. D., Yüksel H., Beacom J. F., Hopkins A. M., Wyithe J. S. B., 2009, ApJ, 705, L104

Klypin A. A., Trujillo-Gomez S., Primack J., 2011, ApJ, 740, 102

Knollmann S. R., Knebe A., 2009, ApJS, 182, 608

Kosenko D. N., Balashev S. A., Noterdaeme P., Krogager J. K., Srianand R., Ledoux C., 2021, MNRAS, 505, 3810

Kritsuk A. G., Norman M. L., Wagner R., 2011, ApJ, 727, L20

Kroupa P., 2001, MNRAS, 322, 231

Krumholz M. R., 2013, MNRAS, 436, 2747

Krumholz M. R., 2014, MNRAS, 437, 1662

Krumholz M. R., Dekel A., 2012, ApJ, 753, 16

Krumholz M. R., Gnedin N. Y., 2011, ApJ, 729, 36

Krumholz M. R., McKee C. F., 2005, ApJ, 630, 250

Krumholz M. R., Tan J. C., 2007, ApJ, 654, 304

Krumholz M. R., McKee C. F., Tumlinson J., 2008, ApJ, 689, 865

Krumholz M. R., McKee C. F., Tumlinson J., 2009, ApJ, 693, 216

Kuhlen M., Vogelsberger M., Angulo R., 2012a, Physics of the Dark Universe, 1, 50

Kuhlen M., Krumholz M. R., Madau P., Smith B. D., Wise J., 2012b, ApJ, 749, 36

Kuhlen M., Madau P., Krumholz M. R., 2013, ApJ, 776, 34

Lada C. J., Lombardi M., Alves J. F., 2010, ApJ, 724, 687

Lada C. J., Forbrich J., Lombardi M., Alves J. F., 2012, ApJ, 745, 190

Lagos C. D. P., Baugh C. M., Lacey C. G., Benson A. J., Kim H.-S., Power C., 2011, MNRAS, 418, 1649

Lagos C. d. P., et al., 2015, MNRAS, 452, 3815

Lambert T. S., et al., 2023, MNRAS, 518, 3183

Lara-López M. A., et al., 2010, A&A, 521, L53

Larson R. B., 1981, MNRAS, 194, 809

Leitherer C., et al., 1999, ApJS, 123, 3

Lenkić L., et al., 2020, AJ, 159, 190

Lenoble R., Commerçon B., Rosdahl J., 2024, arXiv e-prints, p. arXiv:2401.16821

Leroy A. K., Walter F., Brinks E., Bigiel F., de Blok W. J. G., Madore B., Thornley M. D., 2008, AJ, 136, 2782

Leroy A. K., et al., 2013, AJ, 146, 19

Liu D., et al., 2019, ApJ, 887, 235

Loiacono F., et al., 2021, A&A, 646, A76

Luhman M. L., et al., 1998, ApJ, 504, L11

Luhman M. L., Satyapal S., Fischer J., Wolfire M. G., Sturm E., Dudley C. C., Lutz D., Genzel R., 2003, ApJ, 594, 758

Lupi A., 2019, MNRAS, 484, 1687

Lupi A., Bovino S., Capelo P. R., Volonteri M., Silk J., 2018, MNRAS, 474, 2884

Madau P., Dickinson M., 2014, ARA&A, 52, 415

Madden S. C., et al., 2013, PASP, 125, 600

Madden S. C., et al., 2020, A&A, 643, A141

Magnelli B., et al., 2020, ApJ, 892, 66

Malhotra S., et al., 1997, ApJ, 491, L27

Malhotra S., et al., 2001, ApJ, 561, 766

Mannucci F., Cresci G., Maiolino R., Marconi A., Gnerucci A., 2010, MNRAS, 408, 2115

Matthee J., et al., 2019, ApJ, 881, 124

McKee C. F., Ostriker E. C., 2007, ARA&A, 45, 565

Micic M., Glover S. C. O., Federrath C., Klessen R. S., 2012, MNRAS, 421, 2531

Mo H., van den Bosch F. C., White S., 2010, Galaxy Formation and Evolution

Murphy E. J., et al., 2011, ApJ, 737, 67

Narayanan D., Krumholz M. R., 2014, MNRAS, 442, 1411

Narayanan D., Krumholz M. R., Ostriker E. C., Hernquist L., 2012, MNRAS, 421, 3127

Nelson R. P., Langer W. D., 1997, ApJ, 482, 796

Nelson R. P., Langer W. D., 1999, ApJ, 524, 923

Neufeld D. A., Wolfire M. G., 2016, ApJ, 826, 183

Noeske K. G., et al., 2007, ApJ, 660, L43

O'Shea B. W., Wise J. H., Xu H., Norman M. L., 2015, ApJ, 807, L12

Oesch P. A., Bouwens R. J., Illingworth G. D., Labbé I., Stefanon M., 2018, ApJ, 855, 105

Ojha R., et al., 2001, ApJ, 548, 253

Olsen K. P., Greve T. R., Narayanan D., Thompson R., Toft S., Brinch C., 2015, ApJ, 814, 76

Onodera M., et al., 2016, ApJ, 822, 42

Padoan P., Nordlund Å., 2002, ApJ, 576, 870

Padovani M., Galli D., Glassgold A. E., 2009, A&A, 501, 619

Pallottini A., Ferrara A., Bovino S., Vallini L., Gallerani S., Maiolino R., Salvadori S., 2017, MNRAS, 471, 4128

Papadopoulos P. P., Thi W. F., Viti S., 2004, MNRAS, 351, 147

Papadopoulos P. P., van der Werf P., Xilouris E., Isaak K. G., Gao Y., 2012, ApJ, 751, 10

Parravano A., Hollenbach D. J., McKee C. F., 2003, ApJ, 584, 797

Passot T., Vázquez-Semadeni E., 1998, Phys. Rev. E, 58, 4501

Pearson W. J., et al., 2018, A&A, 615, A146

Pelupessy F. I., Papadopoulos P. P., 2009, ApJ, 707, 954

Pelupessy F. I., Papadopoulos P. P., van der Werf P., 2006, ApJ, 645, 1024

Perlmutter S., et al., 1999, ApJ, 517, 565

Péroux C., Howk J. C., 2020, ARA&A, 58, 363

Planck Collaboration et al., 2020, A&A, 641, A6

Popesso P., et al., 2023, MNRAS, 519, 1526

Popping G., Somerville R. S., Trager S. C., 2014, MNRAS, 442, 2398

Posses A., et al., 2024, arXiv e-prints, p. arXiv:2403.03379

Prada F., Klypin A. A., Cuesta A. J., Betancort-Rijo J. E., Primack J., 2012, MNRAS, 423, 3018

Richings A. J., Schaye J., 2016, MNRAS, 458, 270

Riechers D. A., et al., 2019, ApJ, 872, 7

Riess A. G., et al., 1998, AJ, 116, 1009

Robertson B. E., Kravtsov A. V., 2008, ApJ, 680, 1083

Rodighiero G., et al., 2011, ApJ, 739, L40

Rodighiero G., et al., 2015, ApJ, 800, L10

Röllig M., et al., 2007, A&A, 467, 187

Rowan-Robinson M., et al., 2016, MNRAS, 461, 1100

Rowland L. E., et al., 2024, MNRAS, 535, 2068

Saintonge A., et al., 2013, ApJ, 778, 2

Sanders R. L., et al., 2015, ApJ, 799, 138

Sanders R. L., et al., 2021, ApJ, 914, 19

Sandstrom K. M., et al., 2013, ApJ, 777, 5

Savage B. D., Bohlin R. C., Drake J. F., Budich W., 1977, ApJ, 216, 291

Scannapieco C., et al., 2012, MNRAS, 423, 1726

Schäbe A., Romano-Díaz E., Porciani C., Ludlow A. D., Tomassetti M., 2020, MNRAS, 497, 5008

Schaerer D., et al., 2020, A&A, 643, A3

Schmidt M., 1959, ApJ, 129, 243

Schreiber C., et al., 2015, A&A, 575, A74

Schruba A., et al., 2011, AJ, 142, 37

Schruba A., et al., 2012, AJ, 143, 138

Scoville N. Z., 2013, in Falcón-Barroso J., Knapen J. H., eds, , Secular Evolution of Galaxies. p. 491, doi:10.48550/arXiv.1210.6990

Scoville N. Z., Yun M. S., Brown R. L., Vanden Bout P. A., 1995, ApJ, 449, L109

Scoville N. Z., Yun M. S., Windhorst R. A., Keel W. C., Armus L., 1997, ApJ, 485, L21

Scoville N., et al., 2016, ApJ, 820, 83

Scoville N., et al., 2017, ApJ, 837, 150

Seifried D., et al., 2017, MNRAS, 472, 4797

Seifried D., Haid S., Walch S., Borchert E. M. A., Bisbas T. G., 2020, MNRAS, 492, 1465

Shapley A. E., Sanders R. L., Reddy N. A., Topping M. W., Brammer G. B., 2023, ApJ, 954, 157

Shetty R., Glover S. C., Dullemond C. P., Klessen R. S., 2011a, MNRAS, 412, 1686

Shetty R., Glover S. C., Dullemond C. P., Ostriker E. C., Harris A. I., Klessen R. S., 2011b, MNRAS, 415, 3253

Shibuya T., Ouchi M., Harikane Y., 2015, ApJS, 219, 15

Shull J. M., Danforth C. W., Anderson K. L., 2021, ApJ, 911, 55

Silk J., Rees M. J., 1998, A&A, 331, L1

Smith B. D., Regan J. A., Downes T. P., Norman M. L., O'Shea B. W., Wise J. H., 2018, MNRAS, 480, 3762

Sobolev V. V., 1957, Soviet Ast., 1, 678

Solomon P. M., Barrett J. W., 1991, in Combes F., Casoli F., eds, Vol. 146, Dynamics of Galaxies and Their Molecular Cloud Distributions. p. 235

Solomon P. M., Vanden Bout P. A., 2005, ARA&A, 43, 677

Solomon P. M., Rivolo A. R., Barrett J., Yahil A., 1987, ApJ, 319, 730

Somerville R. S., Lee K., Ferguson H. C., Gardner J. P., Moustakas L. A., Giavalisco M., 2004, ApJ, 600, L171

Speagle J. S., Steinhardt C. L., Capak P. L., Silverman J. D., 2014, ApJS, 214, 15

Spiegelhalter D. J., Best N. G., Carlin B. P., Van Der Linde A., 2002, Journal of the Royal Statistical Society Series B: Statistical Methodology, 64, 583

Springel V., 2005, MNRAS, 364, 1105

Springel V., 2010a, ARA&A, 48, 391

Springel V., 2010b, MNRAS, 401, 791

Stacey G. J., Geis N., Genzel R., Lugten J. B., Poglitsch A., Sternberg A., Townes C. H., 1991, ApJ, 373, 423

Stacey G. J., Hailey-Dunsheath S., Ferkinhoff C., Nikola T., Parshley S. C., Benford D. J., Staguhn J. G., Fiolet N., 2010, ApJ, 724, 957

Stinson G., Seth A., Katz N., Wadsley J., Governato F., Quinn T., 2006, MNRAS, 373, 1074

Straatman C. M. S., et al., 2015, ApJ, 808, L29

Swinbank A. M., et al., 2012, MNRAS, 427, 1066

Szomoru D., Franx M., van Dokkum P. G., Trenti M., Illingworth G. D., Labbé I., Oesch P., 2013, ApJ, 763, 73

Tacconi L. J., et al., 2010, Nature, 463, 781

Tacconi L. J., et al., 2013, ApJ, 768, 74

Tacconi L. J., et al., 2018, ApJ, 853, 179

Tacconi L. J., Genzel R., Sternberg A., 2020, ARA&A, 58, 157

Tajiri Y., Umemura M., 1998, ApJ, 502, 59

Tegmark M., Silk J., Rees M. J., Blanchard A., Abel T., Palla F., 1997, ApJ, 474, 1

Teyssier R., 2002, A&A, 385, 337

Thompson R., Nagamine K., Jaacks J., Choi J.-H., 2014, ApJ, 780, 145

Tielens A. G. G. M., Hollenbach D., 1985, ApJ, 291, 722

Tomassetti M., Porciani C., Romano-Díaz E., Ludlow A. D., 2015, MNRAS, 446, 3330

Torrey P., et al., 2019, MNRAS, 484, 5587

Tremonti C. A., et al., 2004, ApJ, 613, 898

Trenti M., Stiavelli M., 2008, ApJ, 676, 767

Tumlinson J., et al., 2002, ApJ, 566, 857

Valentino F., et al., 2018, ApJ, 869, 27

Vallini L., et al., 2024, MNRAS, 527, 10

Vázquez-Semadeni E., 1994, ApJ, 423, 681

Vizgan D., et al., 2022, ApJ, 929, 92

Vogelsberger M., Marinacci F., Torrey P., Puchwein E., 2020, Nature Reviews Physics, 2, 42

Walch S., et al., 2015, MNRAS, 454, 238

Walter F., Weiß A., Downes D., Decarli R., Henkel C., 2011, ApJ, 730, 18

Walter F., et al., 2014, ApJ, 782, 79

Walter F., et al., 2020, ApJ, 902, 111

Weiß A., Henkel C., Downes D., Walter F., 2003, A&A, 409, L41

Weiß A., Downes D., Henkel C., Walter F., 2005, A&A, 429, L25

Whitaker K. E., et al., 2014, ApJ, 795, 104

Wise J. H., Abel T., Turk M. J., Norman M. L., Smith B. D., 2012, MNRAS, 427, 311

Wolfire M. G., Tielens A. G. G. M., Hollenbach D., Kaufman M. J., 2008, ApJ, 680, 384

Wolfire M. G., Hollenbach D., McKee C. F., 2010, ApJ, 716, 1191

Wolfire M. G., Vallini L., Chevance M., 2022, ARA&A, 60, 247

Wong T., Blitz L., 2002, ApJ, 569, 157

Wuyts S., et al., 2012, ApJ, 753, 114

Yan L., et al., 2020, ApJ, 905, 147

Zahid H. J., Dima G. I., Kudritzki R.-P., Kewley L. J., Geller M. J., Hwang H. S., Silverman J. D., Kashino D., 2014, ApJ, 791, 130

Zanella A., et al., 2018, MNRAS, 481, 1976

Zaroubi S., 2013, in Wiklind T., Mobasher B., Bromm V., eds, Astrophysics and Space Science Library Vol. 396, The First Galaxies. p. 45 (arXiv:1206.0267), doi:10.1007/978-3-642-32362-1_2

van der Tak F. F. S., Black J. H., Schöier F. L., Jansen D. J., van Dishoeck E. F., 2007, A&A, 468, 627

van der Wel A., et al., 2014, ApJ, 788, 28

APPENDIX A

Supplementary figures

Figures A.1 and A.2 show the different baryonic properties as a function of halo mass ($M_{\rm halo}$) in the MARIGOLD simulations – `M25` and `M50` at redshifts 4 and 5, respectively.

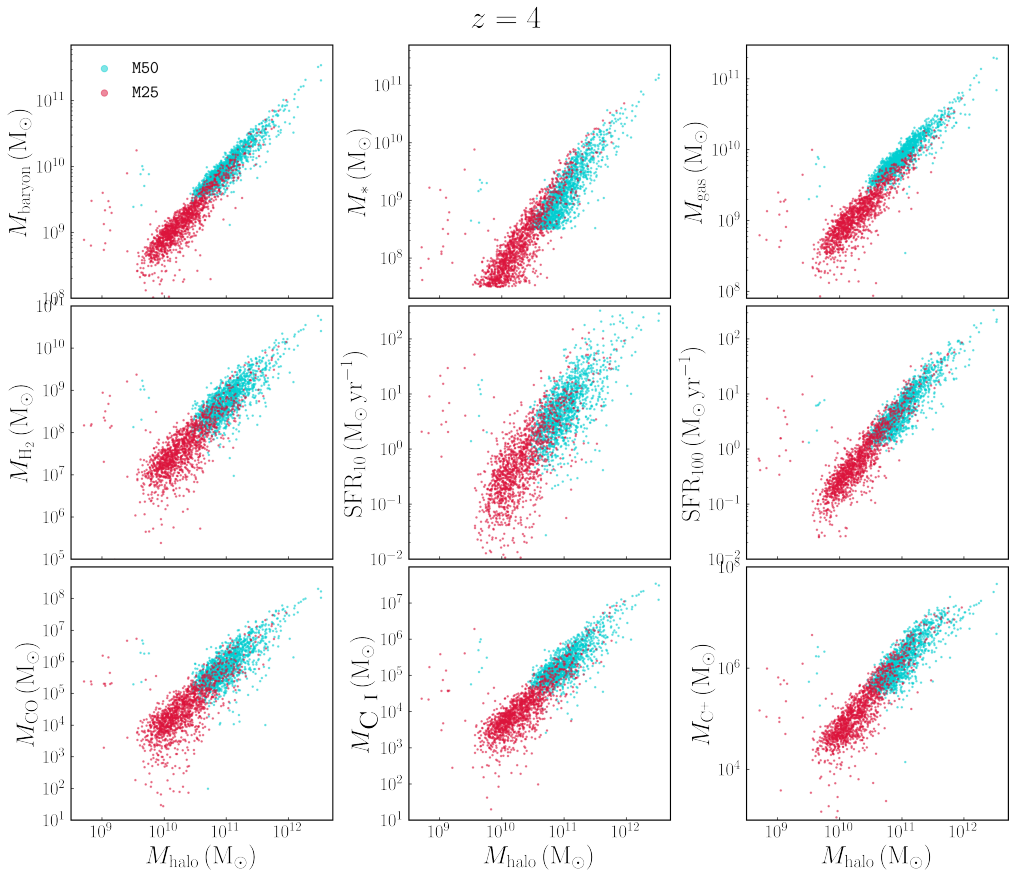


Figure A.1. – Same as Figure 4.5 but at $z = 4$.

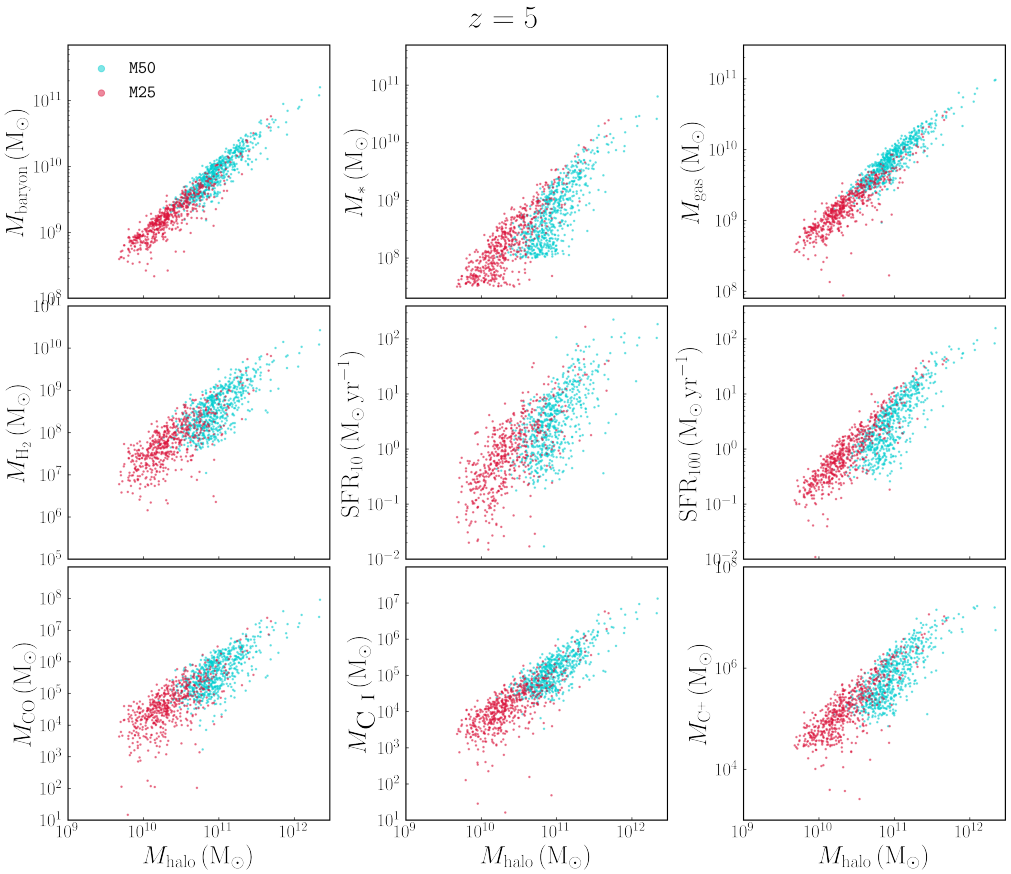


Figure A.2. – Same as Figure 4.5 but at $z = 5$.

APPENDIX B

HYACINTH: HYdrogen And Carbon chemistry in the INTerstellar medium in Hydro simulations

The paper *Khatri et al. A&A (2024), 688, A194* has been published in A&A and is reproduced below in its original form.

A&A, 688, A194 (2024)
https://doi.org/10.1051/0004-6361/202449640

Astronomy & Astrophysics

HYACINTH: HYdrogen And Carbon chemistry in the INTerstellar medium in Hydro simulations

Prachi Khatri[1], Cristiano Porciani[1,2,3,4], Emilio Romano-Díaz[1], Daniel Seifried[5], and Alexander Schäbe[6]

[1] Argelander Institute für Astronomie, Auf dem Hügel 71, 53121 Bonn, Germany
e-mail: pkhatri@astro.uni-bonn.de
[2] SISSA, International School for Advanced Studies, Via Bonomea 265, 34136 Trieste, TS, Italy
[3] Dipartimento di Fisica – Sezione di Astronomia, Università di Trieste, Via Tiepolo 11, 34131 Trieste, Italy
[4] IFPU, Institute for Fundamental Physics of the Universe, via Beirut 2, 34151 Trieste, Italy
[5] Universität zu Köln, I. Physikalisches Institut, Zülpicher Str 77, 50937 Köln, Germany
[6] TÜV NORD EnSys GmbH & Co. KG, Am TÜV 1, 30519 Hannover, Germany

Received 16 February 2024 / Accepted 13 June 2024

ABSTRACT

Aims. We present a new sub-grid model, HYACINTH – HYdrogen And Carbon chemistry in the INTerstellar medium in Hydro simulations – for computing the non-equilibrium abundances of H_2 and its carbon-based tracers, namely CO, C, and C^+, in cosmological simulations of galaxy formation.

Methods. The model accounts for the unresolved density structure in simulations using a variable probability distribution function of sub-grid densities and a temperature-density relation. Included is a simplified chemical network that has been tailored for hydrogen and carbon chemistry within molecular clouds and easily integrated into large-scale simulations with minimal computational overhead. As an example, we applied HYACINTH to a simulated galaxy at redshift $z \sim 2.5$ in post-processing and compared the resulting abundances with observations.

Results. The chemical predictions from HYACINTH are in reasonable agreement with high-resolution molecular-cloud simulations at different metallicities. By post-processing a galaxy simulation with HYACINTH, we reproduced the H I – H_2 transition as a function of the hydrogen column density N_H for both Milky-Way-like and Large-Magellanic-Cloud-like conditions. We also matched the N_{CO} versus N_{H_2} relation inferred from absorption measurements towards Milky-Way molecular clouds, although most of our post-processed regions occupy the same region as (optically) dark molecular clouds in the N_{CO}–N_{H_2} plane. Column density maps reveal that CO is concentrated in the peaks of the H_2 distribution, while atomic carbon more broadly traces the bulk of H_2 in our post-processed galaxy. Based on both the column density maps and the surface density profiles of the different gas species in the post-processed galaxy, we find that C^+ maintains a substantially high surface density out to ∼10 kpc as opposed to other components that exhibit a higher central concentration. This is similar to the extended [C II] emission found in some recent observations at high redshifts.

Key words. methods: numerical – ISM: abundances – ISM: molecules – galaxies: formation – galaxies: high-redshift – galaxies: ISM

1. Introduction

Molecular gas plays a major role in the interstellar medium (ISM) of galaxies, providing the necessary conditions and likely serving as the primary fuel for star formation. The cosmic molecular gas density in the Universe, as inferred from blind surveys such as the VLA CO Luminosity Density at High Redshift[1] (COLDz, Riechers et al. 2019) and the ALMA Spectroscopic Survey in the HUDF[2] (ASPECS, Decarli et al. 2019; Walter et al. 2020), increases by roughly an order of magnitude between redshifts $z \sim 6$ and $z \sim 2$. This is accompanied by a similar trend in the cosmic star formation rate density that reaches its peak value at $z \sim 3$–2, a period known as 'cosmic noon' (see Madau & Dickinson 2014; Förster Schreiber & Wuyts 2020, for a review).

Investigating the build-up of the molecular gas reservoir in galaxies and its cosmic evolution is therefore pivotal to our understanding of the star formation history and galaxy assembly in the Universe. Molecular hydrogen (H_2) is the dominant molecular gas component in galaxies. However, because of the lack of a permanent dipole moment and high excitation temperatures ($T \gtrsim 500$ K) for its ro-vibrational transitions, H_2 does not emit light under typical conditions of molecular clouds ($T \lesssim 100$ K). Therefore, the presence and mass of H_2 are routinely inferred via emission from tracers such as dust, CO, and atomic fine-structure lines.

The low-J rotational transitions of CO are the most commonly used tracers of molecular gas in galaxies (e.g. Dickman et al. 1986; Solomon & Barrett 1991; Downes & Solomon 1998; Solomon & Vanden Bout 2005; Tacconi et al. 2006, 2008, 2010; Daddi et al. 2010a,b; Genzel et al. 2010; Bolatto et al. 2013). The CO-to-H_2 conversion factor α_{CO} captures the relation between the observed CO $J = 1 \rightarrow 0$ luminosity of a galaxy and the underlying molecular gas mass. Alternatively, for spatially resolved observations within the Milky Way (MW) and nearby galaxies, X_{CO} relates the CO intensity (W_{CO}) to the H_2 column density (N_{H_2}) along the line of sight. The variation of X_{CO} with physical conditions, such as metallicity, stellar surface density, galactocentric distance, among others, has been extensively investigated in the MW and nearby galaxies (see Bolatto et al. 2013, for a review). At higher redshifts, however, the limited

[1] http://coldz.astro.cornell.edu/
[2] https://aspecs.info/

number of galaxies observed in CO poses a challenge to studying the dependence of α_{CO} on other galaxy properties. This is further complicated by the fact that the CO $J = 1 \rightarrow 0$ transition is not accessible at high redshifts and observers have to rely on higher-J transitions to obtain an estimate for it. This requires knowledge about the CO excitation ladder, thereby introducing another systematic uncertainty in employing CO as a molecular gas tracer. Moreover, in some galaxies (e.g. low-metallicity dwarf galaxies), a large amount of H_2 is not traced by CO emission and is referred to as CO-dark molecular gas (Wolfire et al. 2010; Madden et al. 2020).

Other tracers of molecular gas such as the [C II] fine-structure line suffer from similar systematic effects. The nature and origin of this line are highly debated as it can arise from multiple ISM phases and not all of the [C II] emission of a galaxy is associated with the molecular gas phase. The presence of a [C II] halo in some galaxies extending 2–3 times farther than their rest-frame UV emission (see e.g. Fujimoto et al. 2020) hints at an extended C^+ reservoir devoid of molecular gas. Additionally, the decrease in the [C II]/far-infrared (FIR) luminosity ratio with increasing FIR luminosity, known as the '[C II] deficit', further complicates the use of this line as a reliable tracer of molecular gas across galaxies.

Atomic carbon has been proposed as another reliable tracer of molecular gas. The fine structure lines of atomic carbon are expected to trace the bulk of the molecular gas in galaxies (Papadopoulos et al. 2004; Tomassetti et al. 2014; Glover & Clark 2016) and have been used both in the local Universe as well as at high redshifts (e.g. Gerin & Phillips 2000; Ikeda et al. 2002; Weiß et al. 2003, 2005; Walter et al. 2011; Valentino et al. 2018; Henríquez-Brocal et al. 2022). However, the use of these lines requires assumptions about the relative abundance of atomic carbon and molecular hydrogen.

Cosmological simulations are a useful tool for investigating the reliability of molecular gas tracers under different ISM conditions and galaxy environments. However, simulating the molecular gas content of galaxies is challenging as it requires modelling the various physical and chemical processes happening on a wide range of (spatial and temporal) scales. On the one hand, it is necessary to simulate galaxies in realistic environments as their ISM is affected by outflows and gas accretion from the cosmic web. On the other hand, molecular-cloud chemistry is regulated by conditions on sub-parsec scales. Early attempts at modelling H_2 assumed a local chemical equilibrium between H_2 formation and destruction (Krumholz et al. 2008, 2009; McKee & Krumholz 2010) and these models have found extensive application in cosmological simulations (Kuhlen et al. 2012, 2013; Hopkins et al. 2014; Thompson et al. 2014; Lagos et al. 2015; Davé et al. 2016). Despite their wide use, these equilibrium models do not account for the dynamic nature of the ISM and the long formation timescale for H_2 (Tielens & Hollenbach 1985). For example, Pelupessy & Papadopoulos (2009), Tomassetti et al. (2015), Richings & Schaye (2016), Pallottini et al. (2017), Schäbe et al. (2020), and Hu et al. (2021) provide a detailed discussion on the effect of non-equilibrium H_2 chemistry on the integrated properties of simulated galaxies. However, cosmological hydrodynamical simulations with on-the-fly computations of the non-equilibrium chemical abundances are rare (e.g. Dobbs et al. 2008; Christensen et al. 2012; Tomassetti et al. 2015; Semenov et al. 2018; Lupi et al. 2018; Lupi 2019; Schäbe et al. 2020; Katz et al. 2022; Hu et al. 2023) and often restricted to individual galaxies. Some of these studies only include a non-equilibrium chemical network for H I and H_2 but not the tracers (Lupi & Bovino 2020; Lupi et al. 2020).

Despite tremendous improvement in the spatial resolution of cosmological simulations over the last decade, the state-of-the-art today is still far from resolving the clumpy ISM in molecular clouds. A clumpy ISM would allow for pockets of high-density gas where H_2 formation would be enhanced. This enhancement is missed by simulations as the density is uniform below their resolution scale. Gnedin et al. (2009) accounted for these unresolved densities by enhancing the H_2 formation rate by an effective clumping factor C assuming a log-normal density distribution for the gas. This technique was later tested by Micic et al. (2012) in their numerical study on the effect of the nature of turbulence on H_2 formation. They found that using a clumping factor systematically overpredicts the H_2 formation rate in regions with a high molecular fraction ($f_{H_2} \gtrsim 0.5$). Christensen et al. (2012) adapted this method in their smooth particle hydrodynamics (SPH) simulations to model the non-equilibrium abundance of H_2 in a cosmological simulation of a dwarf galaxy. However, similar to Micic et al. (2012), they cautioned that this approach works well for dwarf galaxies where fully molecular gas is rare but would need further modifications at high molecular fractions. Moreover, the finite resolution of simulations implies that the temperature tracked in simulations is an average over the resolution element, similar to any other quantity followed explicitly. This average fails to capture the true heterogeneous temperature distribution that is essential for determining the rates of chemical reactions taking place in the ISM.

To overcome these limitations, Tomassetti et al. (2015, hereafter T15) developed a sub-grid model that accounts for the unresolved density structure in simulations by assuming a log-normal probability distribution function (PDF) of sub-grid densities. They assigned a temperature to each sub-grid density in the PDF using a temperature-density relation from high-resolution simulations of molecular clouds (Glover & Mac Low 2007b). They obtained the H_2 abundance in each resolution element by evolving their chemical network at the sub-grid level and integrating over the density PDF. They found good agreement between different gas and stellar properties of their simulated galaxy at $z = 2$ and observations of high-redshift galaxies.

In this paper, we improve upon the work of T15 and introduce a new sub-grid model, HYACINTH – HYdrogen And Carbon chemistry in the INTerstellar medium in Hydro simulations – for on-the-fly computation of the non-equilibrium abundances of H_2 and associated carbon tracers (CO, C, and C^+) within cosmological simulations of galaxy formation. The paper is organised as follows: in Sect. 2, we describe the components of HYACINTH and how it can be incorporated into cosmological simulations. A comparison of our chemical network with two more complex chemical networks and a photon-dominated region code is presented in Sect. 3.1. We further compare the chemical evolution from HYACINTH against high-resolution simulations of molecular clouds – the SILCC-Zoom simulations (Seifried et al. 2017, 2020) and the Glover & Mac Low (2011) simulations in Sect. 3.2. Although HYACINTH is primarily designed to be embedded as a sub-grid model in cosmological simulations, as an immediate application, we applied it to a galaxy simulation (from T15) in post-processing and directly compared it with observations related to the abundances of H_2, CO, C, and C^+ in nearby and high-redshift galaxies. These are discussed in Sect. 4. In Sect. 5, we elaborate on the caveats involved in using HYACINTH as a post-processing tool and compare this with other approaches

in the literature. A summary of our findings is presented in Sect. 6.

2. Methods

At the core of HYACINTH is a PDF of sub-grid densities and a simplified chemical network. The PDF accounts for the unresolved density structure in simulations, statistically incorporating its impact on the chemistry at resolved scales. The chemical network for hydrogen and carbon chemistry is designed to be efficient and scalable for integration into large-scale galaxy formation simulations. Additionally, HYACINTH can be used as a post-processing tool as well; see Sect. 4 for a sample application. In the following, we describe the technical specifications of these two components.

2.1. The sub-grid density PDF

The density structure of molecular clouds is governed by the interplay between turbulence and self-gravity. The PDF of densities is an important statistical property that describes this structure. For instance, the mass-weighted PDF gives the probability that an infinitesimal mass element $\mathrm{d}M$ has a density in the range $[\rho, \rho + \mathrm{d}\rho]$. This distribution is expected to take a log-normal shape in an isothermal, turbulent medium, not significantly affected by the self-gravity of gas (see e.g. Vázquez-Semadeni 1994; Passot & Vázquez-Semadeni 1998; McKee & Ostriker 2007, for a review). Using near-infrared dust extinction mapping of nearby molecular clouds, Kainulainen et al. (2009) found that the column density PDF in quiescent clouds is very well described by a log-normal. However, in star-forming clouds, they found large deviations from a log-normal: the PDF in these clouds has a log-normal shape only at low column densities and resembles a power law at high column densities. A similar picture is supported by numerical simulations of turbulent and self-gravitating gas (e.g. Nordlund & Padoan 1999; Klessen 2000; Glover & Mac Low 2007b; Ballesteros-Paredes et al. 2011; Ward et al. 2014). On scales where the turbulence is supersonic and self-gravity is unimportant, the gas follows a log-normal distribution. In regions where gas is collapsing under its own gravity, the density structure is log-normal at low densities and develops a power-law tail at high densities. Motivated by these findings, we have modified the original PDF of T15 and use a log-normal PDF for turbulence-dominated regions and a log-normal+power-law PDF for gravity-dominated regions.

The mass-weighted log-normal PDF of sub-grid densities n_H is given by (T15)

$$\mathcal{P}_\mathrm{M}(n_\mathrm{H}) = \frac{1}{\sqrt{2\pi}\sigma n_\mathrm{H}} \exp\left[-\frac{(\ln n_\mathrm{H} - \mu)^2}{2\sigma^2}\right], \tag{1}$$

where σ and μ are parameters that decide the width and the location of the peak of the distribution. It is often convenient to introduce the clumping factor

$$C = \langle n_\mathrm{H}^2 \rangle / \langle n_\mathrm{H} \rangle^2 \tag{2}$$

that captures the inhomogeneity and the degree of clumpiness in the medium. For a log-normal PDF, it is related to the parameter σ above as $C = \mathrm{e}^{\sigma^2}$. We use a constant clumping factor of 10, which has previously been shown to reproduce the observed H_2 fractions in nearby galaxies (see e.g. Gnedin et al. 2009; Christensen et al. 2012, T15). We note, however, that it is possible to adopt a more sophisticated approach by varying C with the local Mach number $\mathcal{M}$ or the 3D velocity dispersion (see e.g. Lupi et al. 2018). The parameter μ is related to the mean density $\langle n_\mathrm{H} \rangle$ in the region as $\mu = \ln \langle n_\mathrm{H} \rangle + \frac{\sigma^2}{2}$.

For a log-normal+power-law distribution of n_H, the PDF takes the form

$$\mathcal{P}_\mathrm{M}(n_\mathrm{H}) = \begin{cases} \dfrac{Q_1}{n_\mathrm{H}} \exp\left[-\dfrac{(\ln n_\mathrm{H} - \mu_2)^2}{2\sigma_2^2}\right], & \text{if } n_\mathrm{H} \le n_\mathrm{tr} \\ Q_2 \left(\dfrac{n_\mathrm{H}}{n_\mathrm{tr}}\right)^\alpha, & \text{if } n_\mathrm{tr} < n_\mathrm{H} \le n_\mathrm{cut} \\ 0, & \text{if } n_\mathrm{H} > n_\mathrm{cut}, \end{cases} \tag{3}$$

where $\alpha < 0$ is the slope of the power law and n_tr is the density at which the power-law tail begins. The parameters μ_2 and σ_2 characterise the location of the peak and the width of the log-normal part of the PDF. These are calculated, along with constants Q_1 and Q_2, for a given $\langle n_\mathrm{H} \rangle$, α, and n_tr to match the mean density to $\langle n_\mathrm{H} \rangle$ and ensure the continuity, differentiability, and normalisation of the PDF. Numerical simulations of self-gravitating molecular clouds (e.g. Kritsuk et al. 2011; Federrath & Klessen 2013) have shown that a power law with an index $\alpha > -1$ provides a good fit to the density distribution in regions undergoing gravitational collapse. Girichidis et al. (2014) developed an analytical model for the time evolution of the slope of the power-law tail in a spherically collapsing cloud. They found that irrespective of the initial state of the cloud, after one free-fall time, the power-law tail has a universal slope of −0.54. Their value is in good agreement with the range of values observed for star-forming clouds in Kainulainen et al. (2009) and those found by Kritsuk et al. (2011) in simulations[3]. Therefore, we adopt $\alpha = -0.54$. Furthermore, we set n_tr equal to 10 times the mean density $\langle n_\mathrm{H} \rangle$ (Kritsuk et al. 2011) and, in order to prevent the integral of the PDF from diverging, we impose a cut-off of $n_\mathrm{cut} = 1000 \langle n_\mathrm{H} \rangle$, above which the PDF is set to zero.

For both PDFs, we do not vary the parameters with the spatial resolution as long as it is larger than the typical scales of density fluctuations (i.e. the scales at which clumping takes place) in molecular clouds. For a detailed discussion on the variability of the clumping factor, we refer the interested readers to Micic et al. (2012) and Schäbe et al. (2020).

Figure 1 shows the distribution of sub-grid densities n_H in two sample simulation cells with different PDFs, each with $\langle n_\mathrm{H} \rangle = 100\,\mathrm{cm}^{-3}$. The corresponding cumulative distribution functions (CDF) are plotted in the bottom panel. We can see that the log-normal+power-law distribution spans a broader range of densities as compared to the log-normal. This information is captured by the clumping factor C, which for the log-normal+power-law PDF used here is ∼300 compared to $C = 10$ for the log-normal.

Effect of stellar feedback on the PDF. The internal structure and lifecycle of molecular clouds are affected by star formation and stellar feedback. While supernova (SN) feedback begins to act only a few million years (3–10) after star formation, pre-SN feedback in the form of stellar winds, photoionisation, and radiation pressure, particularly the latter in dense regions, can already start to act once stars form within molecular clouds.

[3] Several studies (e.g. Kritsuk et al. 2011) use a volume-weighted density PDF. The power-law index α_V of the volume-weighted PDF is related to the power-law index α_M of the mass-weighted PDF as $\alpha_\mathrm{V} + 1 = \alpha_\mathrm{M}$. Unless otherwise stated, we use α to refer to α_M.

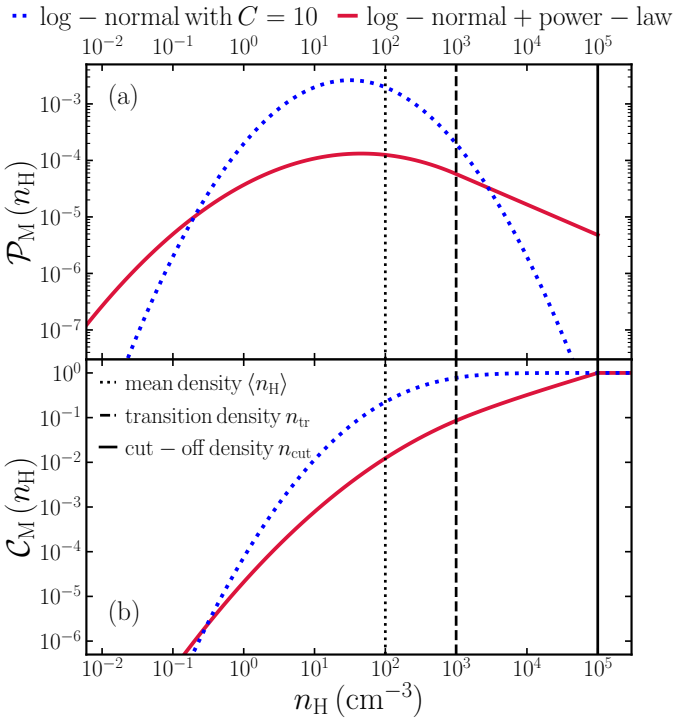


Fig. 1. Probability distribution functions (PDFs) and the corresponding cumulative distribution functions (CDFs) used in this study. Panel a shows the mass-weighted PDFs in sample simulation cells as a function of the sub-grid density $n_{\rm H}$, where $\mathcal{P}_{\rm M}(n_{\rm H})\,{\rm d}n_{\rm H}$ denotes the fraction of the total cell mass present at sub-grid densities in the range $[n_{\rm H}, n_{\rm H} + {\rm d}n_{\rm H}]$. The log-normal (Eq. (1)) and the log-normal+power-law (Eq. (3)) PDFs are shown in blue and red, respectively. The sample cells have a mean hydrogen density $\langle n_{\rm H}\rangle = 100\,{\rm cm}^{-3}$ (shown by the dotted black line). For the log-normal+power-law PDF, the transition density $n_{\rm tr}$ and the cut-off density $n_{\rm cut}$ are shown by the dashed and solid black lines, respectively. Panel b shows the corresponding CDFs, that is, $C_{\rm M}(n_{\rm H})$ denotes the fraction of the total cell mass present at sub-grid densities below $n_{\rm H}$. For the log-normal+power-law PDF shown here, the power-law tail encloses ∼91% of the total cell mass.

Recent observations of molecular clouds in nearby galaxies (Hollyhead et al. 2015; Hannon et al. 2019; Kruijssen et al. 2019; Chevance et al. 2020) have found that these effects efficiently disperse the gas within molecular clouds.

A typical molecular-cloud region would cycle through episodes of star formation followed by stellar feedback. We use a log-normal+power-law PDF for gravitationally collapsing regions before star formation. At the onset of star formation in this region, we transition to a log-normal PDF to capture the combined effects of pre-SN and SN feedback. After a period of 40 Myr (Oey & Clarke 1997; also see Chevance et al. 2023 for a recent review of molecular cloud lifetimes), we switch back to the log-normal+power-law PDF. This 40 Myr timescale includes the molecular cloud dispersal time plus the time it takes for the assembly of the next cloud, but still prior to collapse. We note that the exact value of this timescale time will vary as a function of the local density and might be overestimated for some regions.

We note that our model does not explicitly capture the collapse of gas and star formation but only approximates their effect on chemistry by modifying the PDF from a log-normal to a log-normal+power-law form. Hence, these PDFs are a tool to mimic the effects of the 'microscopic' (i.e. unresolved) density structure on the 'macroscopic' (resolved) chemistry at different stages in the lifecycle of a molecular cloud.

2.2. The chemical network

We use a simplified version of the Nelson & Langer (1999, hereafter NL99) chemical network with some modifications (described in the following subsections) from the recent work of Gong et al. (2017, hereafter G17). Our simplifications reduce the number of chemical species and reactions that we follow to retain only the dominant formation and destruction channels for H_2, CO, C, and C^+ under the physical conditions prevalent in molecular clouds (see Table A.1 in Appendix A). This simplifies the NL99 network for easier integration into cosmological simulations, allowing on-the-fly computation of chemical abundances without a significant increase in the computational overhead of these simulations.

Numerical simulations of molecular clouds (see e.g. Glover & Mac Low 2007b; Hu et al. 2021) show that at the densities relevant for H_2 formation in molecular clouds, the temperatures are mostly $\lesssim$200 K. At these temperatures, the contribution of ionised hydrogen (H^+) to the total hydrogen mass is expected to be negligible. Therefore, we make a further simplifying assumption that all hydrogen is either atomic or molecular. Hence, in practice, we solve the system of rate equations for only three species, namely H_2, CO, and C^+. The electron abundance follows from change conservation (i.e. $f_{e^-} = f_{C^+}$). We compute the abundances of H and C based on the conservation of hydrogen and carbon nuclei, respectively. We assume that the total gas-phase elemental abundances of carbon and oxygen are proportional to the gas metallicity Z, i.e. $f_{\rm C,tot} = 1.41\times10^{-4}\,(Z/{\rm Z}_\odot)$ and $f_{\rm O,tot} = 3.16\times10^{-4}\,(Z/{\rm Z}_\odot)$ (Savage & Sembach 1996).

2.2.1. H_2 chemistry

We consider the following formation and destruction channels for H_2:

1. H_2 formation on dust grains
   $H + H + {\rm grain} \rightarrow H_2 + {\rm grain}$;
2. radiative recombination of H_3^+ with an electron
   $H_3^+ + e^- \rightarrow H_2 + H$;
3. collisional reactions of H_3^+ with atomic carbon and oxygen
   $H_3^+ + C \rightarrow CH_x + H_2$,
   $H_3^+ + O \rightarrow OH_x + H_2$;
4. photodissociation of H_2
   $H_2 + \gamma \rightarrow 2H$;
5. ionisation of H_2 by cosmic rays
   $H_2 + {\rm CR} \rightarrow H_2^+ + e^-$;
6. collisional reaction of H_2 with C^+
   $H_2 + C^+ + e^- \rightarrow CH_x + H$,
   $H_2 + C^+ + e^- \rightarrow C + 2H$.

The rate of H_2 formation on dust grains depends on the dust abundance. We adopt a metallicity-dependent dust-to-gas mass ratio (DTG) based on observational measurements and theoretical predictions of the DTG as a function of the gas-phase metallicity in galaxies at redshifts $0 < z \lesssim 5$ (Péroux & Howk 2020

and Popping & Péroux 2022). The DTG in our model is given as:

$$\log_{10}(\mathrm{DTG}) = 1.3\log_{10}(\mathrm{Z}/\mathrm{Z}_{\odot}) - 2.02, \quad (4)$$

where $Z/Z_{\odot}$ is the gas-phase metallicity in solar units. We use $Z_{\odot} = 0.02$ (Karakas 2010). The above non-linear dependence of the DTG on gas metallicity reflects a variable dust-to-metals (DTM) ratio (i.e. the fraction of metals locked up in dust grains). Based on a compilation of absorption-line studies of high-redshift objects, Popping & Péroux (2022) found that the DTG does not show significant evolution for $0 < z < 5$. Hence, we further assume that the same relation holds at all redshifts.

For the collisional reaction between H_2 and C^+, NL99 only considers the first outcome. However, it has been shown by Wakelam et al. (2010) that the reaction between C^+ and H_2 gives CH_x + H only 70% of the times. In the remaining 30% of the cases, C + 2H are formed instead. This has important consequences for the relative abundances of C and CO. The CH_x formed in the first outcome can react with an O atom to form CO or could photodissociate into C and H atoms, whereas, the second outcome acts as an additional formation channel for C in our network as well as in G17. The dissociation of H_2 is carried out by photons in two narrow bands of energies in the range 11.2–13.6 eV (λ = 912–1108 Å), called Lyman-Werner photons. We do not explicitly include three-body interactions in our network as these are inefficient at most ISM densities resolved in cosmological simulations. However, three-body reactions are the main mechanism for H_2 formation at high redshifts $z \gtrsim 12$ (see e.g. Christensen et al. 2012; Lenoble et al. 2024).

2.2.2. H_3^+ chemistry

The ionisation of H_2 by cosmic rays produces the H_2^+ ion that quickly reacts with an H atom to form H_3^+. H_3^+ can be destroyed by reactions with e^-, C, and O. Because of its high reactivity (Oka 2006), we assume a local (i.e. at each sub-grid density) equilibrium between its formation and destruction channels (see Eqs. (A.4)–(A.5) in Appendix A). We note that NL99 also include the reaction of H_3^+ with CO, which is excluded from our chemical network to limit the number of chemical species and reactions. Another difference with respect to NL99 is that, following G17, we consider two outcomes for the recombination of H_3^+ with e^-: a) H_2 + H b) 3 H. Of these, only the first one is included in NL99.

2.2.3. CO chemistry

The formation of CO proceeds via the reaction of $CH_x(OH_x)$ with O (C) which is formed by the reaction of H_3^+ with C (O) (see reactions (3)–(6) in Table A.1). CH_x can additionally be formed by the collisional reaction between C^+ and H_2. The destruction channels for CO are dissociation into C and O by UV photons in the 912–1100 Å band and cosmic rays. The rate of photodissociation drops off exponentially with the increasing column density of H_2, CO, and dust as a result of the shielding effect of these species on the impinging UV radiation deep inside molecular clouds (see Sect. 2.2.7 for the expression of the shielding functions).

2.2.4. Grain-assisted recombination of C^+

Following Glover & Clark (2012, hereafter GC12) and G17, we include grain-assisted recombination of C^+ in our network in addition to its radiative recombination. This is the main channel for C^+ recombination at solar metallicity (G17), although its importance at sub-solar metallicities would depend on the relative amount of dust to gas in the ISM (the dust-to-gas ratio, see Sect. 2.2.1). Moreover, for several ions, including C^+, the recombination rate on dust grains is often higher than the direct radiative recombination rate, especially in star-forming regions where dust, particularly in the form of polycyclic aromatic hydrocarbons (PAHs) is highly prevalent and frequently collides with ions leading to their recombinations. In a previous study, GC12 stressed that this reaction is particularly effective when the ratio of the UV field strength to the mean density is very small. As a result of including this additional destruction (formation) channel for C^+ (C), we expect the C/C^+ ratio predicted by our network (as well as in G17) to be significantly higher than predicted by NL99 (see Sect. 3.1).

2.2.5. Cosmic rays

Cosmic rays (CRs) with energies $\lesssim$0.1 GeV play a critical role in initiating ion-ion chemistry deep inside dense molecular-cloud regions that are well-shielded from UV radiation (see e.g. Padovani et al. 2009). These reactions become particularly relevant at high redshifts where the cosmic ray ionisation rate per H atom (CRIR, ζ_{H}) is expected to be higher than the canonical MW value of $3 \times 10^{-17}\,\mathrm{s}^{-1}$ because of higher star formation rates (see e.g. Muller et al. 2016; Indriolo et al. 2018, for CRIR estimates at high redshifts). We include the ionisation of H_2, the ionisation of C, and the dissociation of CO by CRs in our network while NL99 only includes the first one.

Based on absorption studies of HD in H_2-bearing damped Lyman-α systems, Kosenko et al. (2021) found that the CRIR scales quadratically with the UV field intensity relative to the Draine field[4] (χ, Draine 1978). Therefore, as a default choice, we adopt the following relation between ζ_{H} and χ:

$$\frac{\zeta_{\mathrm{H}}}{\zeta_{\mathrm{H,MW}}} = \left(\frac{\chi}{\chi_{\mathrm{MW}}}\right)^2, \quad (5)$$

but also consider alternative options in Sect. 4 and Appendix D. We use $\zeta_{\mathrm{H,MW}} = 3 \times 10^{-17}\,\mathrm{s}^{-1}$ and $\chi_{\mathrm{MW}} = 1.0$ for MW. We also impose upper and a lower bounds on the CRIR of $3 \times 10^{-14}\,\mathrm{s}^{-1}$ and $10^{-18}\,\mathrm{s}^{-1}$, respectively, to avoid unreasonable CRIRs. The effect of this upper limit is investigated in Sect. 4 and Appendix D. We note that the χ in Eq. (5) is measured in the FUV band (λ = 912–2070 Å), while HYACINTH requires the UV flux in the Lyman–Werner band (λ = 912–1080 Å) in Habing units as an input. In the solar neighbourhood, the mean energy densities in the two bands are related as: $U_{\mathrm{LW}} / U_{\mathrm{FUV}} \sim 1.1$ (Parravano et al. 2003).

2.2.6. The temperature–density relation

As some chemical reactions in our network have temperature-dependent rate coefficients (see Table A.1), we need to associate a temperature with each sub-grid density in the PDF. For this, we use a metallicity-dependent temperature-density relation obtained from simulations of the ISM (Hu et al. 2021).

[4] Often the UV field strength is expressed in terms of the Habing Field (Habing 1968) and the parameter G_0 captures the ratio between the energy density of a given UV field and the energy density measured in the solar neighbourhood by Habing (1968). The Draine field has $G_0 = 1.7$.

These simulations self-consistently include time-dependent H_2 chemistry and cooling, star formation, and feedback from photoionisation and supernovae. They adopt a linear relationship between the CRIR and the UV intensity (actually both quantities scale with the star-formation-rate surface density) which differs from our quadratic scaling and thus leads to a small inconsistency in our assumptions for dense, well-shielded gas, where CRs are an important heating source. In Sect. 4, we compare the chemical abundances obtained when using the $\zeta_H - \chi$ relation from Hu et al. (2021) against those obtained from Eq. (5).

2.2.7. Shielding functions

Dense, optically thick gas can shield itself against penetrating UV radiation because of the high column densities of H_2, CO, C, and dust. We account for this by modifying the reaction rates of photoionisation and photodissociation reactions by appropriate shielding functions. The (self-) shielding of H_2 is given by (Draine & Bertoldi 1996):

$$f_{s,H_2}(N_{H_2}) = \frac{0.965}{(1+x/b_5)^2} + \frac{0.035}{(1+x)^{0.5}} \exp\left[-8.5\times10^{-4}(1+x)^{0.5}\right], \quad (6)$$

where $x = \frac{N_{H_2}}{5\times10^{14}\,\mathrm{cm^{-2}}}$, N_{H_2} is the column density of H_2 and b_5 is the velocity dispersion of gas in $\mathrm{km\,s^{-1}}$. Following Sternberg et al. (2014) and T15, we use a constant $b_5 = 2$ throughout. However, as noted in G17, the H_2 fraction f_{H_2} is insensitive to the value of b_5 for $f_{H_2} \gtrsim 0.1$.

The CO shielding function $f_{s,CO}(N_{CO}, N_{H_2})$ accounts for both CO self-shielding and the shielding of CO by H_2 and is calculated by interpolating over N_{CO} and N_{H_2} from Table 5 in Visser et al. (2009).

The shielding function for C is given by (Tielens & Hollenbach 1985)

$$f_{s,C}(N_C, N_{H_2}) = \exp(-\tau_C)\, f_{s,C}(H_2), \quad (7)$$

where

$$\tau_C = 1.6\times10^{-17}\frac{N_C}{\mathrm{cm^{-2}}}, \quad f_{s,C}(H_2) = \frac{\exp(-r_{H_2})}{1+r_{H_2}}, \quad r_{H_2} = 2.8\times10^{-22}\frac{N_{H_2}}{\mathrm{cm^{-2}}}, \quad (8)$$

where N_C and N_{H_2} are the column densities of atomic carbon and H_2, respectively.

In addition to self-shielding and H_2−shielding, all species are also shielded against UV radiation by dust grains; the relevant shielding function is given by

$$f_{dust} = \exp(-\gamma A_V), \quad (9)$$

where γ is different for each species and listed in Table A.1 in Appendix A. The visual extinction A_V is related to the total column density of hydrogen nuclei $N_H = N_{H\,I} + 2N_{H_2}$ along the line of sight as

$$A_V = \frac{N_H Z_d}{1.87\times10^{21}\,\mathrm{cm^{-2}}}. \quad (10)$$

In a simulation, recording the H_2 fraction for every sub-grid density within a grid cell at each timestep is computationally expensive and impractical. Therefore, we follow the approach of T15 for distributing the available H_2 to the different sub-grid densities. We assume a sharp transition from atomic to molecular hydrogen at the sub-grid density $n_H = n_{crit,H_2}$, that is $f_{H_2}(n_H < n_{crit,H_2}) = 0$ and $f_{H_2}(n_H \geq n_{crit,H_2}) = 1$. This sharp transition has been observed in various numerical studies (Glover & Mac Low 2007a; Dobbs et al. 2008; Krumholz et al. 2008, 2009; Gnedin et al. 2009), and occurs at densities where H_2 becomes self-shielding (Dobbs et al. 2014). Similarly, we assume that carbon transitions from ionic to atomic form above $n_{crit,C\,I}$ and becomes fully molecular at $n_{crit,CO}$. In a given region, these critical densities depend on the density of total hydrogen and the abundance of the different species involved in the transition (see Appendix C for calculation of n_{crit,H_2} for the H I – H_2 transition).

In practice, our sub-grid model requires six input parameters – the average density of hydrogen nuclei $\langle n_H \rangle$, the gas-phase metallicity Z, the UV flux in Lyman-Werner bands in Habing units G_0, the characteristic length scale Δx of a resolution element (e.g. cell size or smoothing length), the density PDF $\mathcal{P}_M$, and the time Δt over which the chemical abundances are to be evolved. When embedded as a sub-grid model in a simulation, these parameters can be obtained directly from the simulation or calculated in post-processing (e.g. the UV flux). Depending on the value of Z, the temperature-density relation assigns a (sub-grid) temperature to each sub-grid density in the PDF. The chemical network then solves the rate equations for each sub-grid density and integrates over the PDF to obtain the average abundances of H_2, CO, C, and C^+ with respect to the total hydrogen within the region.

3. Comparison of chemical abundances

In this section, we compare different aspects of HYACINTH to previous approaches in the literature. First, we focus on the chemical network and contrast its predictions to those from the NL99 and G17 networks. Subsequently, we compare our full implementation of the sub-grid model (including the density PDF) with the output of high-resolution simulations of individual molecular clouds (Seifried et al. 2017, 2020; Glover & Mac Low 2011).

3.1. Comparison with NL99 and G17

In this section, we compare our chemical network predictions to Figs. 2b and 3b in G17. The setup involves a one-dimensional semi-infinite slab with a uniform hydrogen density $n_H = 1000\,\mathrm{cm^{-3}}$, a solar metallicity ($Z = Z_\odot$), a solar dust-to-gas ratio of 0.01, and fixed gas and dust temperatures of 20 K and 10 K, respectively. The slab is irradiated from one side by a UV field of strength 1 in Draine units (i.e. $\chi = 1$). We test for two different values of the CRIR (same as in G17): $1\times10^{-17}\,\mathrm{s^{-1}\,H^{-1}}$ (left column of Fig. 2) and $2\times10^{-16}\,\mathrm{s^{-1}\,H^{-1}}$ (right column of Fig. 2). For this comparison, we assume the same elemental abundances for carbon and oxygen as in G17, that is, $f_{C,tot} = 1.6\times10^{-4}\,(Z/Z_\odot)$ and $f_{O,tot} = 3.2\times10^{-4}\,(Z/Z_\odot)$, throughout Sect. 3.1.

For a fair comparison with the G17 results, we calculate the shielding to the incident radiation field using their approximation for an isotropic radiation field (as described in section 2.3 of G17 and originally used by Wolfire et al. 2010). Briefly, it approximates an isotropic radiation field with a unidirectional

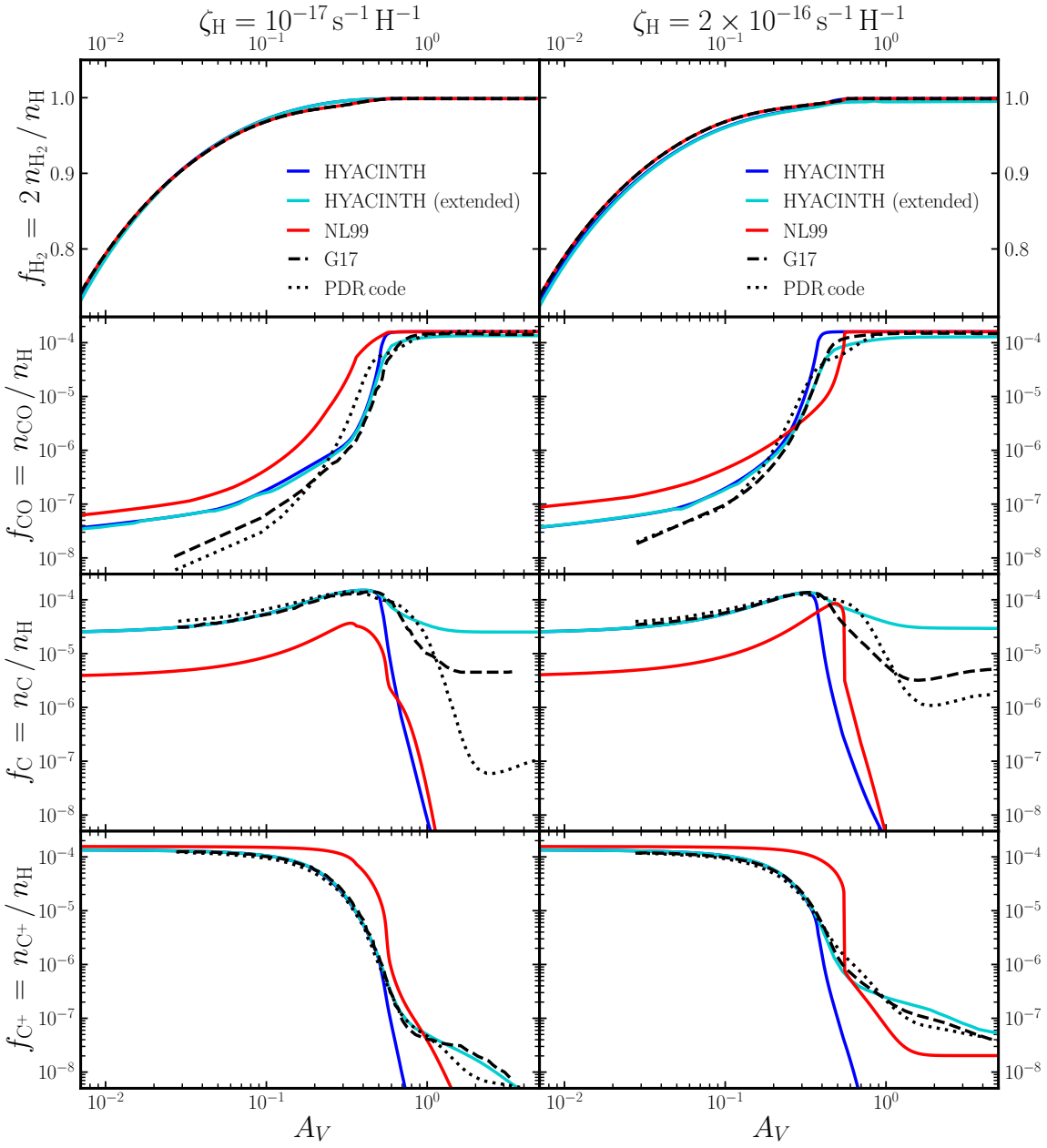


Fig. 2. Comparison of the chemical network in HYACINTH with NL99 and G17 networks. The abundances of H_2, CO, C, and C^+ as a function of the visual extinction A_V in a semi-infinite plane-parallel slab are shown in different panels. The left column shows the abundances for a CRIR (ζ_H) of $10^{-17}\,s^{-1}\,H^{-1}$ while the right one has $\zeta_H = 2\times10^{-16}\,s^{-1}\,H^{-1}$. The blue, turquoise, and red lines represent the results from HYACINTH, extended HYACINTH, and NL99 respectively. Here extended HYACINTH refers to the HYACINTH network with additional chemical reactions for He^+ and HCO^+ that are not part of standard HYACINTH (see text for more details). The dashed and dotted black lines show, respectively, the abundances from the chemical network and the PDR code in G17. The slab has a uniform hydrogen density $n_H = 1000\,cm^{-3}$, solar metallicity and solar dust abundance, and is illuminated from one side by a UV field of strength $\chi = 1$.

field incident at an angle of 60° with the normal to the slab. Thus, for an incident radiation field of strength χ, the effective radiation field at a (perpendicular) depth L into the slab can be expressed as

$$\chi_{\rm eff} = (\chi/2)\, f_{\rm shield}(2L),$$

where $f_{\rm shield}(L)$ represents the shielding function at a depth L into the slab and is different for each chemical species (see Sect. 2.2.7). For this uniform-density slab, the column density of hydrogen, N_H, at depth L can be written as $N_H(L) = n_H\,L$. The slab is divided into 1000 layers with N_H values spaced logarithmically in the range $N_H = 10^{17}\,cm^{-2}$–$10^{22}\,cm^{-2}$ or equivalently, visual extinction A_V (Eq. (10)) in the range 5.35×10^{-5}–5.35. For each layer, the NL99 and HYACINTH networks are evolved until equilibrium.

To put our results in context, we compare them to the output of more complex models. First, we consider the extended HYACINTH chemical network[5] that includes the non-equilibrium treatment of two additional species, namely, He^+ and HCO^+. These species, particularly He^+, serve as the main destruction agents for CO in dense, shielded regions.

[5] The extra chemical reactions treated in extended HYACINTH are listed at the end of Table A.1 (reactions (20)–(29)).

Moreover, we present the output of the photon-dominated region (PDR) code used in G17 which tracks the abundances of 74 species accounting for 322 chemical reactions and includes a more sophisticated treatment of radiative transfer. This PDR code is derived from Tielens & Hollenbach (1985) and updated by Wolfire et al. (2010), Hollenbach et al. (2012), and Neufeld & Wolfire (2016).

Figure 2 shows the equilibrium abundances of the chemical species as a function of A_V. The f_{H_2} versus A_V from our network is in excellent agreement with that from NL99, G17 and the PDR code. In contrast, there are noticeable differences in the $C^+ \rightarrow C$ and $C \rightarrow CO$ transitions in the different networks. For both values of the CRIR, at low A_V, the C^+ abundance in HYACINTH (both standard and extended), G17 and the PDR code is slightly lower than NL99, accompanied by a higher C abundance for these networks with respect to NL99. In HYACINTH, this shift results primarily from the inclusion of grain-assisted recombination of C^+ (see Sect. 2.2.4) and an additional outcome for the $C^+ + H_2$ reaction (see Sect. 2.2.1). Together, these lead to almost an order of magnitude increase in the abundance of atomic carbon at $A_V < 0.5$ in HYACINTH compared to NL99. GC12 observed a similar trend when comparing their networks with and without this reaction. Previous studies have reported that the chemical networks employed in PDR codes tend to produce an elevated atomic carbon abundance with respect to the atomic carbon abundance measured in MW (interstellar) clouds (Sofia et al. 2004; Sheffer et al. 2008; Wolfire et al. 2008; Burgh et al. 2010). This has been long-standing problem for several chemical networks (see e.g. Burgh et al. 2010; Liszt 2011; Gong et al. 2017).

For $\zeta_H = 10^{-17}\,s^{-1}\,H^{-1}$, the $C \rightarrow CO$ transition in HYACINTH (both standard and extended), G17 and the PDR code occurs at a slightly higher A_V than in NL99. Conversely, for $\zeta_H = 2 \times 10^{-16}\,s^{-1}\,H^{-1}$, this transition occurs at a slightly lower A_V in all other approaches compared to NL99. However, this does not have any practical implications for the modelling of CO chemistry in galaxy simulations as we expect the bulk of the CO mass in a molecular cloud or a galaxy to be present at high A_V ($\gtrsim$2).

A noticeable difference between our network and G17 is that at high A_V ($\gtrsim$1), all carbon in our network is in the form of CO, while G17 predicts ~3–10% of the carbon to be in atomic form. The PDR code predicts that $\lesssim$3% of the total carbon is in atomic form at $A_V \gtrsim 1$ for both values of the CRIR. In contrast, extended HYACINTH predicts an even higher abundance of atomic carbon than G17 at $A_V \gtrsim 1$ and consequently a lower CO abundance. Therefore, it is evident that the varying complexities of the different approaches result in significant differences in the atomic carbon abundance at $A_V \geq 1$. However, since the bulk of the atomic carbon mass in all networks is present at intermediate A_V ($0.1 \lesssim A_V \lesssim 1$), the contribution of the atomic carbon in G17 and the PDR code at $A_V \geq 1$ to the total atomic carbon mass is $\lesssim$2% and is therefore, not significant.

Overall, both the standard and extended HYACINTH networks show a good agreement with NL99, G17, and the PDR code in the respective A_V range where each of the carbon species dominates. For H_2, the agreement is excellent for all A_V, showing that hydrogen chemistry is not sensitive to the exact treatment of carbon chemistry. There are, however, noticeable differences between standard and extended HYACINTH for the carbon-based species. For instance, there is a significant difference between the abundances of C and CO at $A_V \gtrsim 0.5$. Extended HYACINTH shows a smoother $C \rightarrow CO$ transition as compared to standard HYACINTH. This transition closely matches that in G17. Nevertheless, at $A_V \gtrsim 2$, the CO abundance in extended HYACINTH is only marginally different from that in the standard one – roughly 16% (20%) less for $\zeta_H = 10^{-17}\,s^{-1}\,H^{-1}$ ($2 \times 10^{-16}\,s^{-1}\,H^{-1}$). This shows that standard HYACINTH provides robust CO abundances with respect to extended HYACINTH, while requiring approximately 3.3 times less computational time. Therefore, in what follows, we only consider the standard HYACINTH network.

3.2. Comparison with molecular-cloud simulations

Now we shift our focus to assessing the performance of the chemical network in conjunction with the sub-grid density PDF and compare against high-resolution simulations of individual molecular clouds – the SILCC-Zoom simulations (Seifried et al. 2017, 2020) with solar metallicity ($Z = Z_\odot$, Sect. 3.2.1) and those from Glover & Mac Low (2011, hereafter GML11) with $Z = 0.1\,Z_\odot$ (Sect. 3.2.2). We note that neither the SILCC-Zoom nor the GML11 runs account for star formation and stellar feedback.

3.2.1. SILCC-Zoom simulations

SILCC-Zoom are adaptive mesh refinement (AMR) simulations that follow the formation of two molecular clouds extracted from the parent SILCC simulations (Walch et al. 2015) with the zoom-in technique. They reach a minimum cell size of 0.06 pc in the densest regions. The original SILCC and SILCC-Zoom simulations were run with a chemical network based on Glover & Mac Low (2007a) and Glover & Mac Low (2007b) for hydrogen chemistry and Nelson & Langer (1997) for CO chemistry. However, the SILCC-Zoom simulations used in this comparison (Seifried et al. 2020) were performed with the NL99 network for CO chemistry.

Our goal is to compare the chemical evolution from HYACINTH (employed as a sub-grid model within a hydro simulation) with that from SILCC-Zoom. For this, we build a low-resolution copy of one of the SILCC-Zoom clouds (MC1-HD) by running a simulation with our modified version of the RAMSES code (Teyssier 2002) that uses HYACINTH for evolving the time-dependent chemistry. We consider a cubic box with a side length of 175 pc split in cells with a size of 25 pc (comparable to the spatial resolution we aim to achieve in our future cosmological simulations). We set the initial conditions (gas properties and chemical abundances) by coarse graining the SILCC-Zoom snapshot after a simulated time of 1 Myr, when all the mesh refinements have been performed, so as to exclude any possible numerical effects of variable spatial resolution on chemical abundances. We then follow the evolution of the system for 3 Myr, that is the full duration of the SILCC-Zoom runs. Beyond self-gravity, our simulation accounts for an external gravitational potential as described in sections 3.1 and 3.2 of Walch et al. (2015). The DTG is set to 0.01 and the elemental abundances of carbon and oxygen are set to $f_{C,tot} = 1.41 \times 10^{-4}$ and $f_{O,tot} = 3.16 \times 10^{-4}$. The cells are irradiated with an ISRF of strength $G_0 = 1.7$ in Habing units and a CRIR of $\zeta_H = 1.3 \times 10^{-17}\,s^{-1}\,H^{-1}$.

Before proceeding, two comments are in order. First, the lack of stellar feedback in the SILCC-Zoom simulations leads to the formation of very high-density clumps. Consequently, the density PDFs within the 25 pc cells vary over time and deviate significantly from the analytical forms we adopt in HYACINTH (see also Fig. 20 in Walch et al. 2015 and Fig. 3 in Buck et al. 2022). For instance, the clumping factor in these cells spans a wide range of values and is often much higher than the values of

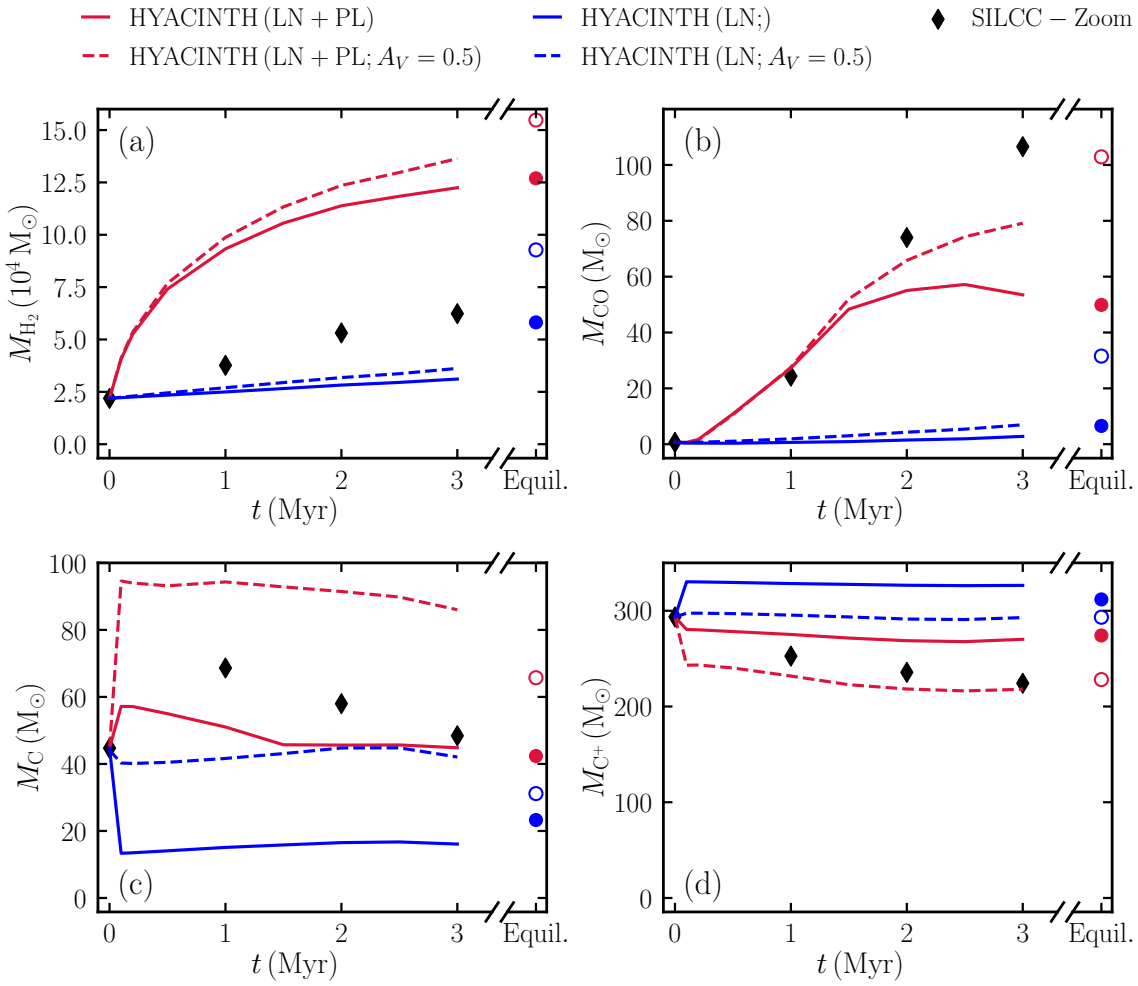


Fig. 3. Time evolution of the total mass of different chemical species contained in a $(175\,\mathrm{pc})^3$ region encompassing the molecular cloud MC1-HD from the SILCC-Zoom simulations (black diamonds). The curves show the corresponding results from four low-resolution hydrodynamic simulations that use HYACINTH as a sub-grid model for the chemical evolution. The colour coding distinguishes runs obtained assuming either the LN PDF (blue) or the LN+PL PDF (red). The line style indicates the strength of the assumed uniform UV ISRF: $G_0 = 1.7$ in Habing units with no attenuation (solid) or with an effective visual extinction of $A_V = 0.5$ (dashed). The circles on the right-hand-side of the panels show the masses derived by assuming that the final snapshot of each simulation is in chemical equilibrium. Filled and empty symbols refer to the zero attenuation and $A_V = 0.5$ runs, respectively.

10 and 300 that we adopt for the log-normal (hereafter LN) and log-normal+power-law (hereafter LN+PL) PDFs, respectively. Second, our simulation does not include radiative transfer of UV radiation. Consequently, the local ISRF is not attenuated due to the shielding from surrounding gas and dust as done in the SILCC-Zoom simulations. To quantify the impact of this effect, we perform a second simulation adopting a uniform effective visual extinction $A_V = 0.5$ in the entire simulation volume.

Figure 3 shows the time evolution of the masses of different chemical species in our original ($A_V = 0$, solid lines) and UV-attenuated ($A_V = 0.5$, dashed lines) simulations compared to SILCC-Zoom (black diamonds). The shape of the density PDF determines how quickly the molecules are produced. The formation timescales for H_2 and CO are comparable to (for the LN+PL PDF) or longer than (for the LN PDF) the duration of the simulations which explains why we find that their masses are growing throughout the simulation. The LN PDF consistently underpredicts M_{H_2} with respect to SILCC-Zoom by a factor of ∼2 and produces little CO within the simulation time. This is consistent with the fact that SILCC-Zoom has a much more prominent high-density tail. On the other hand, the LN+PL PDF forms H_2 at a rate that is nearly twice as high as SILCC-Zoom and nicely matches the evolution of M_{CO} for the first 1.5 Myr, beyond which it saturates at about 50% and 75% of the final value in SILCC-Zoom in the runs with $A_V = 0$ and 0.5, respectively. To give an idea of the long-term evolution of the different species, we also compute their masses at chemical equilibrium for the final snapshot of the simulation (these are indicated with filled ($A_V = 0$) and empty ($A_V = 0.5$) circles on the right-hand side of the plots). These values never differ substantially from the simulation output at 3 Myr.

Overall, despite the differences in the exact treatment of various physical processes and the chemical network in SILCC-Zoom and our simulations, the predicted masses of the chemical species are in reasonable agreement, in particular when the LN+PL PDF is used. We stress that the goal here is not to achieve a perfect match since, as we mentioned, the sub-grid density PDF in MC1-HD is quite different from the analytical models used within HYACINTH, which are meant as averages over an ensemble of turbulent molecular clouds in the presence of stellar feedback. Finally, it is worth remembering that, when using HYACINTH as a sub-grid model in cosmological simulations,

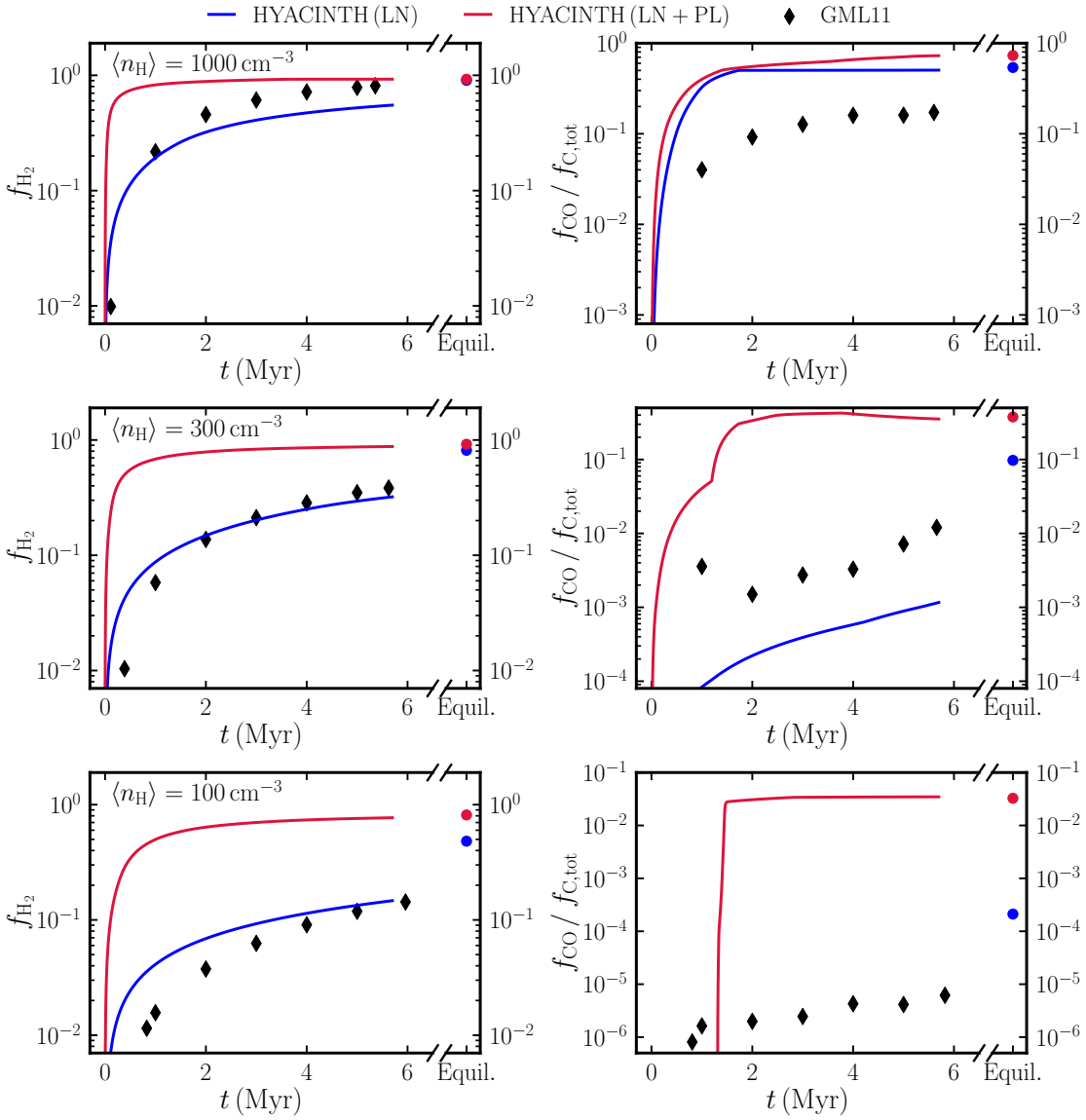


Fig. 4. Comparison of the chemical abundances from HYACINTH against those from the $Z = 0.1\,Z_\odot$ runs from GML11 as a function of time. For $\langle n_H \rangle = 100\,cm^{-3}$, very little CO is formed when using the LN PDF (not visible in the bottom-right panel). The circles on the right-hand side denote the equilibrium abundances.

we switch between the LN and LN+PL PDFs to mimic the combined effects of self-gravity in a turbulent ISM and stellar feedback as explained in Sect. 2.1. The switch takes place on much longer timescales than those investigated in this section and would likely lead to somewhat different results.

3.2.2. GML11 simulations

The GML11 simulations track the thermal and chemical evolution of magnetised and supersonically turbulent gas with typical conditions found in molecular clouds. The computational domain is a cube of side length $L = 20\,pc$ divided into a fixed number of cells. We present results for the runs with a grid-cell size of 0.156 pc which generate a closely LN density PDF well converged with the exception of the high-density tail. The GML11 simulations consider three different mean densities (100, 300, and 1000 cm^{-3}) and use a chemical network (consisting of 218 reactions between 32 species) adapted to model hydrogen, carbon, and oxygen chemistry in molecular clouds. A detailed comparison of the performance in modelling CO chemistry of this more extended network and the simpler one given in NL99 is presented in GC12.

As the boxlength of the GML11 simulations is similar to the cell size we aim to achieve in our future cosmological runs, we model the entire computational volume of GML11 as a single domain in HYACINTH. The DTG is set to $0.01\,(Z\,/\,Z_\odot)$ and the elemental abundances of carbon and oxygen are set to $f_{C,tot} = 1.41 \times 10^{-4}(Z/Z_\odot)$ and $f_{O,tot} = 3.16 \times 10^{-4}(Z/Z_\odot)$ (same as in GML11). All hydrogen is initially in atomic form while all carbon is in ionised form. The regions are irradiated with an interstellar radiation field (ISRF) of strength $G_0 = 1.7$ in Habing units and a cosmic ray ionisation rate of $\zeta_H = 10^{-17}\,s^{-1}\,H^{-1}$. We compute the chemical evolution for 5.7 Myr (the time interval covered by the GML11 runs) and, in Fig. 4, compare the H_2 and CO abundances with those from GML11[6] (as presented in their Fig. 5). For $\langle n_H \rangle = 1000\,cm^{-3}$, both LN and LN+PL give

[6] The abundances of C and C^+ are not presented in GML11.

a reasonable agreement with GML11, while for the lower densities, the f_{H_2} from LN closely matches the f_{H_2} from GML11, but the one from LN+PL is consistently higher. A similar behaviour is observed for CO, although LN predicts roughly an order of magnitude lower f_{CO} at $\langle n_H \rangle = 300\,cm^{-3}$. In contrast, LN+PL gives a significantly higher f_{CO}. At $\langle n_H \rangle = 100\,cm^{-3}$, hardly any CO forms in GML11 and in our LN run, while the f_{CO} from LN+PL comprises a small fraction of the total carbon abundance (∼5%).

As in Fig. 3, on the right hand side of each panel, we show the abundances that would be obtained at chemical equilibrium. For the LN+PL PDF, these are similar to the results obtained after 5.7 Myr, indicating that the molecules form with a characteristic timescale of a few million years. On the contrary, the equilibrium abundances obtained with the LN PDF are significantly higher than those obtained after 5.7 Myr, reflecting a much slower formation rate for the molecules due to the less prominent high-density tail in the PDF.

Overall, HYACINTH with the LN PDF gives H_2 abundances in excellent agreement with the GML11 simulations. The situation is more complex for CO, as the formation timescale of the molecules is very sensitive to the high-density tail of the sub-grid PDF (as we noticed already in Fig. 3) and some fine tuning would be needed to get a good match between the different models. For $\langle n_H \rangle = 100$ and $300\ cm^{-3}$, our results with the LN PDF overestimate the formation time compared to GML11 while those with the LN+PL PDF underestimate it. At $\langle n_H \rangle = 1000\ cm^{-3}$, CO is always produced at a slightly faster rate in HYACINTH than in the GML11 simulations. This simple test suggests that, at such densities and on such time intervals, HYACINTH overestimates the CO abundance by a factor of ∼2 with respect to GML11. We note, however, that, in a less idealised set-up that accounts for star formation and stellar feedback, the high-density regions would be exposed to much more intense UV radiation than assumed in this test (see, e.g. Fig. 6) and this would surely affect the resulting CO abundance.

4. Sample application

In this section, we apply HYACINTH to a simulated galaxy to compare our model predictions with observations of H_2, CO, C, and C^+ abundances in nearby and high-redshift galaxies. Although our primary goal is to integrate HYACINTH as a sub-grid component within simulations, here we post-process an existing simulation from T15 at $z \sim 2.5$ and compute the equilibrium abundances. The galaxy in T15 was simulated with a modified version of the AMR code RAMSES (Teyssier 2002) including a sub-grid model for H_2 chemistry and a method to propagate the UV radiation in the Lyman-Werner (LW) bands to nearby cells. H_2 formation takes place on the surface of dust grains and H_2 is destroyed by LW photons. In contrast, our chemical network includes additional channels for H_2 formation and destruction (see Sect. 2.2.1 and Table A.1 for a complete list). Moreover, we adopt a different DTG than T15 ($\log_{10}(DTG_{T15}) = \log_{10}(Z/Z_\odot) - 2.0$; see Sect. 2.2.1 for our DTG).

An important caveat when applying HYACINTH in post-processing involves the CRIR. In this scenario, the UV field χ from the simulation is used as an input to compute the equilibrium abundances, meaning that it is assumed that χ stays constant until equilibrium is reached. Because of the quadratic relationship between the CRIR and χ in HYACINTH (Sect. 2.2.5), cells in the simulation that have recently undergone star formation will have an extended period of high CRIR. Thus, $\zeta_H \propto \chi^2$ is not the ideal CRIR for a post-processing application[7].Therefore, we adopt a fixed CRIR of $3 \times 10^{-17}\ s^{-1}\ H^{-1}$. Lastly, as applying the model in post-processing does not allow us to switch between the two PDFs, here we only use the log-normal (as in T15).

Figure 5 shows the column-density maps for the species obtained via post-processing as well as the total hydrogen and H_2 column densities from the simulation. Firstly, we see a remarkable agreement between the N_{H_2} maps from the simulation and post-processing, indicating that the differences in our computation of H_2 chemistry compared to T15, namely the additional chemical reactions and the equilibrium treatment do not significantly impact the H_2 column density. This also indicates that the H_2 in the simulated galaxy is close to equilibrium. From the distribution of the three carbon species, we see that CO is the dominant carbon component in the highest N_{H_2} ($\gtrsim 10^{22}\ cm^{-2}$) regions and is hardly present at $N_{H_2} \lesssim 10^{21}\ cm^{-2}$, where N_{CO} drops below $10^{15}\ cm^{-2}$. The distribution of atomic carbon (C) shows a great similarity to that of H_2 both in its extent and the location of peaks. C^+ is found even more extensively throughout the galaxy, resembling the spread of the total hydrogen (H_{tot}). This is because carbon's ionisation energy (11.6 eV) is slightly lower than that of hydrogen, allowing C^+ to exist in all ISM phases. We further discuss this in the context of some recent observations in Sect. 4.2.4. Another noteworthy feature is that while the density peaks in CO and C coincide with those in H_2, the local fluctuations in C^+ are milder. As H_2 assists in shielding the former two from UV radiation, their abundances are enhanced in high N_{H_2} regions. On the other hand, C^+ can exist both in the atomic and molecular phases of the ISM; therefore, we do not see a strong enhancement in N_{C^+} with increasing N_{H_2}.

4.1. Comparison of H_2 abundance

A comparison of the post-processed H_2 fraction, $f_{H_2} = 2\langle n_{H_2} \rangle / \langle n_H \rangle$, with that directly obtained from the simulation is presented in Fig. 6. We stress that because of the differences in T15 and our chemical network and because we solve for equilibrium, we do not expect to exactly reproduce the dynamically evolved H_2 abundance from the simulation in every grid cell, but only obtain values similar to those in the simulated galaxy. We also calculate f_{H_2} in each cell using two analytical prescriptions, namely 'KMT-EQ' (Krumholz et al. 2009) and 'KMT-UV' (Krumholz 2013). In the KMT-EQ relation, f_{H_2} is determined by the gas column density and metallicity, and is independent of the strength of the ISRF, G_0, by construction. The KMT-UV relation accounts for the effect of UV radiation on f_{H_2} and is sensitive to the ratio $G_0/\langle n_H \rangle$. In the left panel, we plot the median f_{H_2} as a function of $\langle n_H \rangle$ for the different approaches. Firstly, we see that at $\langle n_H \rangle \gtrsim 100\,cm^{-3}$, the median f_{H_2} from post-processing with a uniform ζ_H (solid black line) agrees very well with the simulation and the KMT-EQ prediction. At lower densities, however, the three show some differences. By $\langle n_H \rangle = 10\,cm^{-3}$, the KMT-EQ prediction gradually decreases to 0, while the simulation has a $f_{H_2} \sim 0.2$. The post-processed f_{H_2} shows a similar trend at low densities but is consistently higher than the simulation. In contrast with all other approaches, KMT-UV predicts a median $f_{H_2} = 0$ at $\langle n_H \rangle \lesssim 100\,cm^{-3}$. This median is dominated by the cells that have a very high $G_0/\langle n_H \rangle$ value and as such the KMT-UV prescription gives $f_{H_2} = 0$ for

[7] It is better suited for dynamically evolving chemistry within simulations where the UV field strength χ would vary over time.

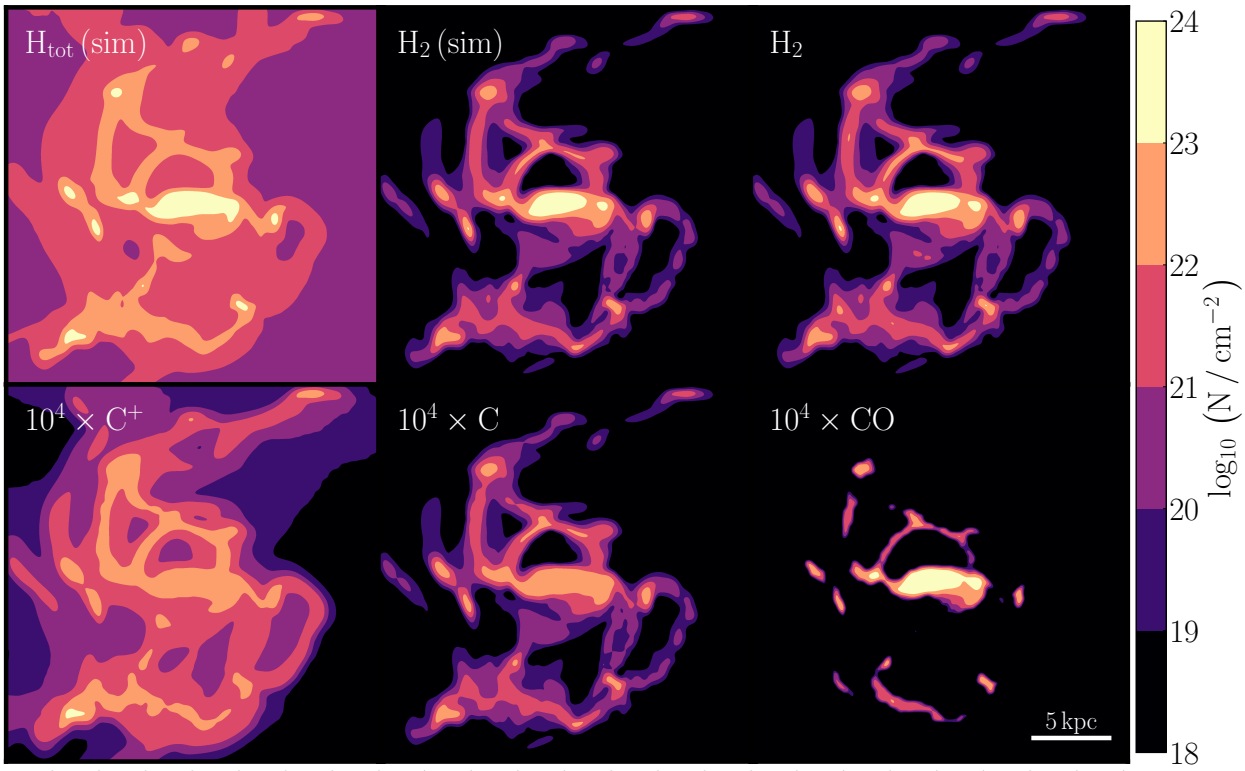


Fig. 5. Column density maps of different species in the simulated galaxy after post-processing. The total hydrogen and H_2 from the simulation are shown in the first two panels. The remaining panels show the column density of H_2, C^+ , C, and CO obtained via post-processing. The H_2 from post-processing (top row, third panel) is remarkably similar to the H_2 from the simulation (top row, second panel). CO dominates the carbon budget in the highest N_{H_2} regions and is essentially absent when $N_{H_2} \lesssim 10^{21}$ cm^{-2}. Atomic carbon traces the full extent of the H_2 distribution, while C^+ extends farther out to regions devoid of a significant amount of H_2.

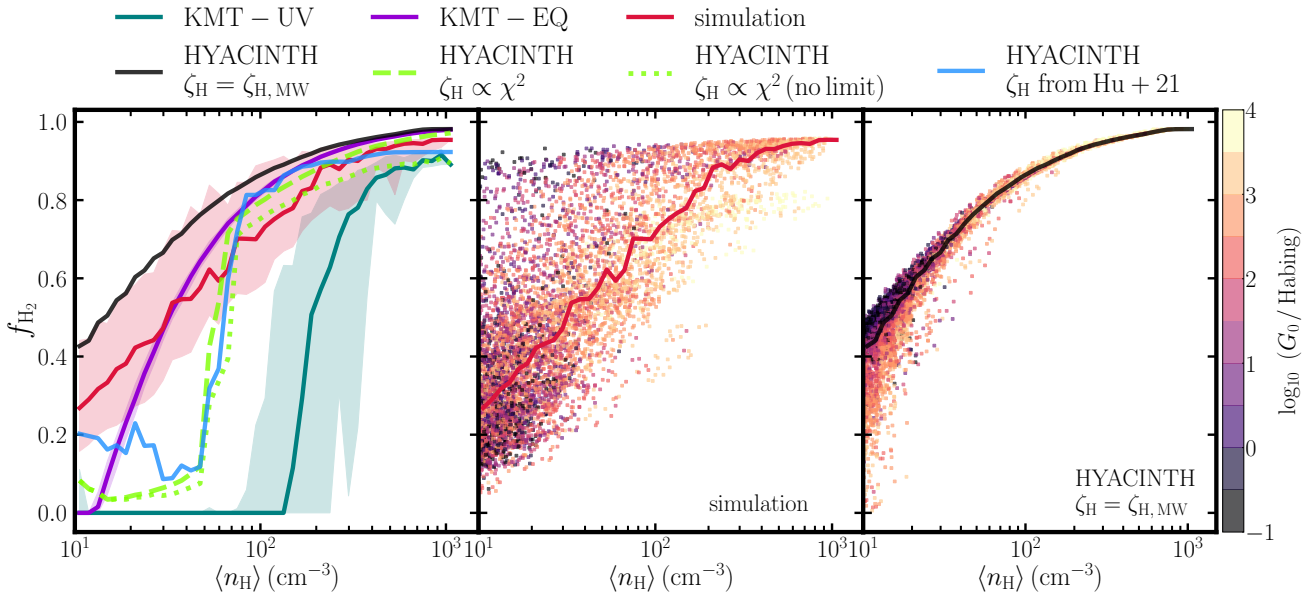


Fig. 6. Comparison of the H_2 fraction ($f_{H_2} = 2\langle n_{H_2} \rangle / \langle n_H \rangle$) as a function of $\langle n_H \rangle$ using different approaches. Left: the equilibrium f_{H_2} from post-processing the T15 galaxy with HYACINTH (black, green, and blue) compared with the f_{H_2} from the simulation (red) and two analytical estimates – KMT-EQ (blue) and KMT-UV (turquoise). The lines show the median value in a given $\langle n_H \rangle$ bin while the shaded area encloses the 16th to 84th percentiles. The solid black line denotes the median f_{H_2} when using a uniform CRIR of $\zeta_H = 3 \times 10^{-17}$ s^{-1} H^{-1} for HYACINTH. We also show the median f_{H_2} obtained from post-processing when using the CRIR from Sect. 2.2.5 (i.e. $\zeta_H \propto \chi^2$) with (dashed green line) and without (dotted green line) the upper limit on the CRIR as well as (solid blue line) the $\zeta_H - \chi$ relation from Hu et al. (2021, i.e., $\zeta_H \propto \chi$, normalised to $\zeta_{H,MW} = 10^{-16}$ s^{-1} H^{-1}). Middle: the f_{H_2} from the T15 simulation colour-coded by the strength of the UV field in the LW bands in Habing units, G_0 (from the simulation), in each grid cell. The red line is the same as in the left panel. Right: the post-processed f_{H_2} (using a uniform ζ_H) for each grid cell within the simulated galaxy colour-coded by G_0. The black line is the same as in the left panel.

all these cells. At $\langle n_H \rangle \gtrsim 100\,\mathrm{cm}^{-3}$, the f_{H_2} increases gradually and gives similar results as the other methods.

Additionally, to demonstrate the impact of using a variable ζ_H in post-processing, we show the f_{H_2} obtained for three different $\zeta_H - \chi$ relations: (i) $\zeta_H \propto \chi^2$, but with an upper limit of $3 \times 10^{-14}\,\mathrm{s}^{-1}\,\mathrm{H}^{-1}$ on ζ_H (dashed green curve); (ii) $\zeta_H \propto \chi^2$ without any upper limit (dotted green curve); (iii) the $\zeta_H - \chi$ relation from Hu et al. (2021, solid blue curve). In all three cases, the post-processed f_{H_2} exhibits a sharp drop at $\langle n_H \rangle \sim 70\,\mathrm{cm}^{-3}$ and shows a great resemblance to the KMT-UV prediction. We see that removing the upper bound of $3 \times 10^{-14}\,\mathrm{s}^{-1}\,\mathrm{H}^{-1}$ on the CRIR in our $\zeta_H \propto \chi^2$ relation (dotted green line) results in a decrease in the median f_{H_2} by $\approx 10\%$ at $\langle n_H \rangle \geq 100\,\mathrm{cm}^{-3}$.

In the middle and right panels of Fig. 6, we show, respectively, the dynamical (i.e. from the simulation) and the equilibrium f_{H_2} (with a uniform ζ_H) for each grid cell colour-coded by G_0, the strength of the UV field in the LW band in the cell. We see that, at a given $\langle n_H \rangle$, the equilibrium f_{H_2} predicted by our model is sensitive to the value of G_0 (similar to KMT-UV), with a higher G_0 resulting in a lower f_{H_2}. In contrast, the simulated f_{H_2} does not show a clear trend with G_0, because the UV field varies throughout the formation history of H_2 in any given region. Overall, we see that the different approaches for H_2 chemistry result in varying predictions for f_{H_2}. In particular, the abundance computed dynamically in the simulation differs from the equilibrium calculations and shows a larger scatter at all densities. Similar plots when using a variable ζ_H are shown in Fig. D.2.

4.2. Comparison with observations

4.2.1. The HI - H_2 transition

We compare the H I – H_2 transition in the post-processed galaxy (with a uniform ζ_H) with the observed one obtained from the absorption spectra of distant quasars and nearby stars in the MW and Large Magellanic Cloud (LMC). For this, we compute the column densities of H_2 and total hydrogen within a position-position-velocity data cube of 156 pc × 156 pc × 40 km s^{-1}[8]. For a total hydrogen column density $N_H = N_{HI} + 2N_{H_2}$, the H_2 fraction can be defined as $f_{H_2} = 2N_{H_2}/N_H$. The results of this comparison are shown in Fig. 7. In the left panel, we show the regions with MW-like conditions, that is $0.6 < Z/Z_\odot < 1.4$ and $0.6 < G_0/1.7 < 1.4$ and in the right panel, those with LMC-like conditions, that is $0.18 < Z/Z_\odot < 0.42$ and $6 < G_0/1.7 < 14$. The observed data are taken from the Copernicus survey (Savage et al. 1977), the HERACLES survey of 30 nearby galaxies (Schruba et al. 2011), and the compilation of several observations of OB stars in the Galactic disc with the Far Ultraviolet Spectrographic Explorer (FUSE) by Shull et al. (2021). The LMC data are from Tumlinson et al. (2002) using a FUSE survey. We see that the H I – H_2 transition from post-processing agrees very well with the observed relation, particularly for MW-like conditions (left panel). As evident from both the post-processed and observed data, the transition is sensitive to the metallicity and the strength of the ISRF: it shifts to higher N_H values for LMC-like conditions as compared to MW-like conditions, since the former has 10 times stronger ISRF than the MW but only about a third of the metals in the MW.

[8] 156 pc is the spatial resolution of the simulation at the redshift of post-processing. The velocity dispersion obtained from the different absorption measurements ranges from 4–35 km s^{-1}; here we adopt 40 km s^{-1} as a safe upper limit for these values.

4.2.2. The relationship between N_{H_2} and N_{CO}

Figure 8 shows the relation between the ratio of CO to H_2 column densities (N_{CO}/N_{H_2}) and the H_2 column density N_{H_2} in the post-processed galaxy for three different CRIRs. For comparison, we include column densities obtained from UV absorption measurements along sight lines towards diffuse and (optically) dark molecular clouds in the MW from Sheffer et al. (2008). The N_{CO}/N_{H_2} values for a compilation of dark-cloud observations by Federman et al. (1990) are also shown. In these observations, dark clouds are defined as those with visual extinction $A_V \gtrsim 5$, while those with $A_V \lesssim 5$ are defined as diffuse. As these are MW observations, we only plot the post-processed regions with $0.6 < Z/Z_\odot < 1.4$. Moreover, Fig. 7 of Sheffer et al. (2008) shows that the ISRF values (expressed as $\chi = G_0/1.7$) for these observations range from ∼0.5 to ∼10. Therefore, for a fair comparison, we only select the post-processed regions with $G_0/1.7$ in the range [0.35, 13], allowing for a scatter of 30%.

At H_2 column densities above $10^{21}\,\mathrm{cm}^{-2}$, our post-processed points (for all CRIRs) span the same region in the N_{CO}/N_{H_2} versus N_{H_2} plane as Federman et al. (1990), albeit with significantly less scatter. At lower $N_{H_2} \sim 10^{21}\,\mathrm{cm}^{-2}$, the N_{CO}/N_{H_2} ratio shows a sharp drop when using a uniform $\zeta_H = \zeta_{H,MW}$, in contrast to the gradual decline seen in the observed data at $N_{H_2} \sim 5 \times 10^{20}$–$10^{21}\,\mathrm{cm}^{-2}$. Conversely, the N_{CO}/N_{H_2} ratio shows a gradual decline similar to the observed data when using a variable ζ_H (CRIR). Nevertheless, both our data and observations exhibit a large scatter in the N_{CO}/N_{H_2} values at $N_{H_2} \in [10^{20}, 10^{21}]\,\mathrm{cm}^{-2}$. Moreover our values are in a good agreement with the observed values at these H_2 column densities. Such large variations in N_{CO}/N_{H_2} values were also reported previously by Smith et al. (2014) in their simulation of a MW-like galaxy. At $N_{H_2} \lesssim 10^{20}$, our data is sparse and some of our sight lines have a factor ∼2 higher N_{CO} compared to those reported by Crenny & Federman (2004) and Sheffer et al. (2008). Nevertheless, the majority of our data remains consistent with the observations shown here. Finally, it is worth noting that the regions within our post-processed galaxy with MW-like metallicity predominantly resemble dark clouds while only a few inhabit the region where $N_{CO} < 10^{16}\mathrm{cm}^{-2}$ and $N_{H_2} < 10^{21}\mathrm{cm}^{-2}$.

4.2.3. The abundance of atomic carbon

The fine structure lines of atomic carbon C I (corresponding to the $^3P_2 - ^3P_1$ and $^3P_1 - ^3P_0$ transitions) are often used to infer the H_2 masses of galaxies – both in the local Universe as well as at high redshifts (e.g. Gerin & Phillips 2000; Ikeda et al. 2002; Weiß et al. 2003, 2005; Walter et al. 2011; Valentino et al. 2018; Henríquez-Brocal et al. 2022). This method requires assuming an atomic carbon abundance relative to H_2. Several observations have tried to measure this abundance using other independent estimates of the H_2 mass of a galaxy such as dust emission in the infrared or rotational lines of CO. Here we compare the atomic carbon abundance relative to H_2 for our post-processed galaxy with those found in observations. Our galaxy (post-processed with a uniform ζ_H) has $M_{CI} = 6.29 \times 10^6\,M_\odot$ and $M_{H_2} = 4.70 \times 10^{10}\,M_\odot$ which implies a galaxy-integrated neutral carbon abundance relative to H_2, $X_{CI} = \dfrac{M_{CI}}{6\,M_{H_2}}$ of 2.23×10^{-5} or equivalently $\log_{10} X_{CI} = -4.65$. The M_{CI} of our galaxy is similar to that obtained by Tomassetti et al. (2014) using a different approach. They use a X_{CI} of 3×10^{-5} from Alaghband-Zadeh et al. (2013) and obtained $M_{CI} = 7.92 \times 10^6\,M_\odot$ for their simulated galaxy at $z = 2$.

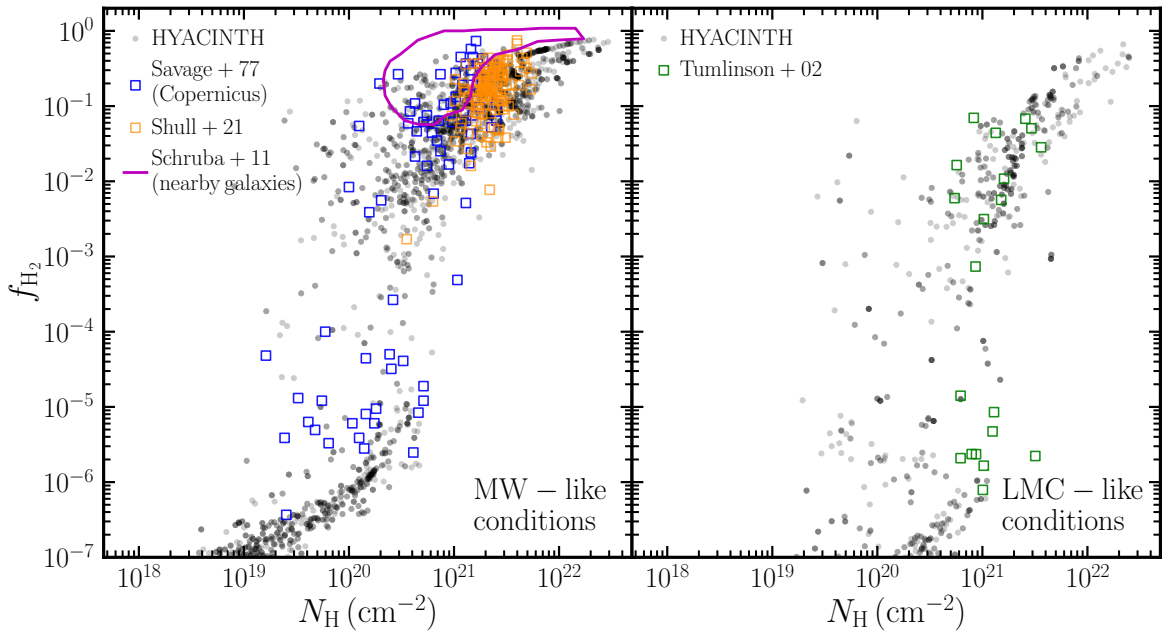


Fig. 7. H_2 fraction as a function of the total column density of hydrogen $N_H = N_{H\,I} + 2N_{H_2}$, where $f_{H_2} = 2N_{H_2}/N_H$. Left: the HYACINTH-post-processed regions with MW-like conditions are shown in black. The observed data are from Savage et al. (1977) (blue open squares), Shull et al. (2021) (orange open squares), Schruba et al. (2011) (magenta dots). For clarity, here we only show a contour enclosing all data points for Schruba et al. (2011). Right: the HYACINTH-post-processed regions with LMC-like conditions (black) compared to the observed data from Tumlinson et al. (2002) (green open squares). For HYACINTH data, we use a uniform $\zeta_H = \zeta_{H,MW}$ and the column densities are calculated in a position–position–velocity cube of 156 pc 156 pc × 40 km s^{-1}, where 156 pc is the spatial resolution of the AMR grid in the simulation at the redshift of post-processing and 40 km s^{-1} is an upper limit on the observed velocity dispersion in the various absorption measurements used in this comparison.

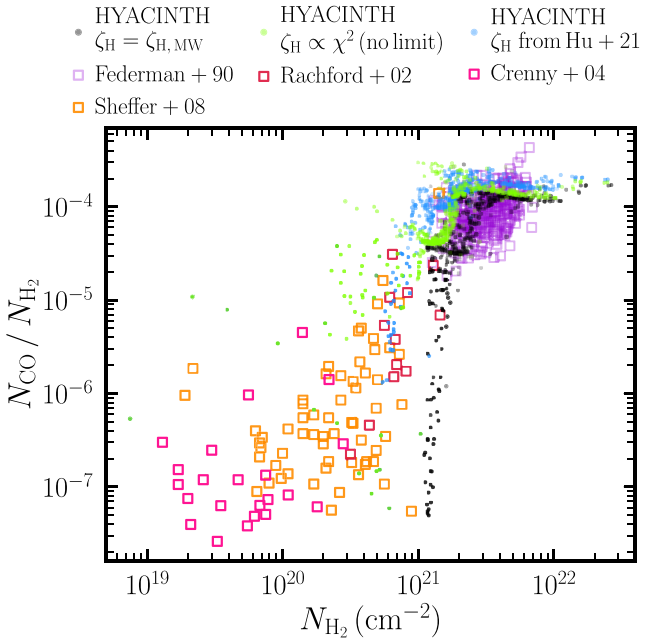


Fig. 8. Ratio of the CO-to-H_2 column density N_{CO}/N_{H_2} versus N_{H_2} for the post-processed galaxy, compared with observations of (optically) diffuse (Rachford et al. 2002; Crenny & Federman 2004; Sheffer et al. 2008) and dark (Federman et al. 1990) molecular clouds in the MW. The black, green, and blue points are from the galaxy post-processed with HYACINTH but using different CRIRs (as indicated in the legend). In each case, the points are selected to have a gas-metallicity $Z \in [0.7, 1.3]$ and $\chi \in [0.35, 1.3]$ to match the respective values in the observed data. The column densities are calculated in a position-position-velocity cube of 156 pc 156 pc × 5 km s^{-1}, to match the velocity dispersion in the absorption measurements.

Walter et al. (2011) carried out a survey of C I emission in $z > 2$ sub-millimetre galaxies (SMGs) and quasar host galaxies using the IRAM Plateau de Bure interferometer and the IRAM 30 m telescope. They obtain a mean $\log_{10} X_{C\,I}$ of $-4.08^{+0.16}_{-0.23}$ for their sample of 10 galaxies. Valentino et al. (2018) carried out a survey of [C I] ($^3P_1 - {}^3P_0$) in far-infrared-selected main-sequence (MS) galaxies in the COSMOS field at $z \sim 1.2$ with the Atacama Large Millimeter Array (ALMA). They found a mean $\log_{10} X_{C\,I}$ of -4.7 ± 0.1 and -4.8 ± 0.2, respectively, for M_{H_2} estimates based on dust and CO measurements for their sample of 12 galaxies. For SMGs at $z \gtrsim 2.5$, they obtain a value of -4.3 ± 0.2 and -4.2 ± 0.1 for $M_{H_2,dust}$ and $M_{H_2,CO}$, respectively. For a sample of 6 MS galaxies at $z = 1$–3 observed as part of the ASPECS, Boogaard et al. (2020) estimated $X_{C\,I} = (1.9 \pm 0.5) \times 10^{-5}$ (i.e. $\log_{10} X_{C\,I} = -4.72^{+0.10}_{-0.13}$). Using a sample of 21 lensed starburst galaxies with [C I] detection at $z \sim 1.3$–3.5, Harrington et al. (2021) found an $X_{C\,I}$ of $(6.82 \pm 3.04) \times 10^{-5}$ (i.e. $\log_{10} X_{C\,I} = -4.17^{+0.16}_{-0.26}$). A slightly lower value of ~ -4.5 is often used in observations of galaxies at $z \gtrsim 2$ (e.g. Weiß et al. 2003). Thus, broadly speaking, the $X_{C\,I}$ in our post-processed galaxy is consistent with those obtained for star-forming galaxies at $z \gtrsim 2$. In particular, our $\log_{10} X_{C\,I} = -4.65$ is similar to those in MS galaxies, but lower than the values reported for starbursts and SMGs.

4.2.4. The extended C$^+$ distribution

The [C II] fine-structure line at 158 µm is one of the brightest emission lines in star-forming galaxies and an important coolant of the ISM. The morphology, extent, and kinematics of this line give crucial insights into the different physical processes in galaxies. Several recent [C II] observations with ALMA (e.g. Nesvadba et al. 2016; Carniani et al. 2018; Fujimoto et al. 2019,

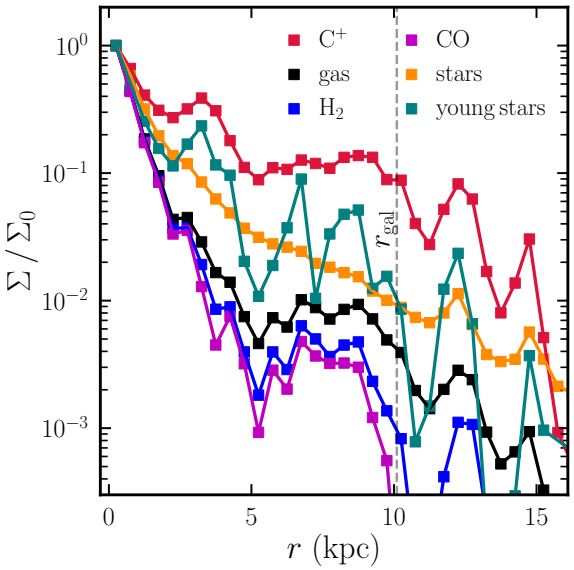


Fig. 9. Surface density profiles of gas (black), H_2 (blue), CO (magenta), C^+ (red), stars (orange), and young stars (teal, with ages less than 10 Myr) within the HYACINTH-post-processed galaxy. For each component, the profiles are normalised by their central value Σ_0. For each radial bin, we calculate the surface density from the face-on projection from a cylinder with height of ±5 kpc. Among all components, C^+ and young stars show the slowest decline as we move away from the centre of the galaxy, while H_2 and CO are more concentrated (see also Fig. 5). The dashed grey line denotes the size of the galaxy as defined in T15.

2020; Herrera-Camus et al. 2021; Lambert et al. 2023) have detected the presence of an extended C^+ reservoir in galaxies out to $z \sim 7$. These observations find that the effective radius of the [C II]-emitting region is ≈ 2–3 times larger than that of the rest-frame UV-emitting region. Possible explanations for this extended emission include metal-rich outflows, mergers, connection to Lyα halo, and contribution from satellite galaxies, without any clear consensus so far regarding its source (Fujimoto et al. 2019, 2020). What causes this extended [C II] emission remains an open question in the field of galaxy formation.

Figure 9 shows the surface density profiles (normalised to one at the centre) for all gas, H_2, CO, C^+, young stars (age ≤ 10 Myr) and all stars in our post-processed galaxy (with a uniform ζ_H) observed face on. The H_2 and CO profiles show a steep decline such that both Σ_{H_2} and Σ_{CO} decrease by more than an order of magnitude from the centre out to 2 kpc. The stellar surface density profile shows a similar although less steep decline. In contrast, the C^+ profile exhibits a gradual decline, maintaining a substantially high surface density at distances extending up to ∼10 kpc, beyond which it shows a steep decline. In this regard, our (normalised) C^+ profile is similar to the stacked one observed in Fujimoto et al. (2019) and used to infer the presence of a C^+ 'halo' (which might be an inappropriate name as, in our simulated galaxy, the extended profile arises from the gaseous disc and not from a spheroidal distribution). Young stars also exhibit a relatively slower decline albeit with strong fluctuations. This reflects that UV radiation is essential for ionising carbon. Moreover, this is consistent with the observed correlation between the [C II] line luminosity and the star formation rate (De Looze et al. 2011, 2014; Herrera-Camus et al. 2015; Schaerer et al. 2020). Also in Fig. 5, we saw that the C^+ distribution strongly mirrors the total hydrogen distribution which is dominated by atomic and ionised hydrogen at larger distances from the centre. Since we apply HYACINTH in post-processing, all C^+ that is farther than the other components is formed in situ and not transported by any dynamical processes such as outflows.

5. Discussion

It is well known that H_2 formation can be greatly enhanced by the clumpiness of the ISM on scales typically below the resolution scale of present-day cosmological simulations. Our model accounts for this unresolved density structure by assuming a sub-grid density PDF that varies with the state of star formation in each gas element. This variable PDF is designed to mimic the effect of pre-supernova feedback on the density structure of molecular clouds. In this regard, our approach is alternative to Lupi et al. (2018) who assume a log-normal PDF with the clumping factor related to the local Mach number. While Lupi et al. (2018) always assume a log-normal PDF with a continuously varying clumping factor, our approach has a binary nature, where the PDF switches between two distinct shapes – log-normal and log-normal+power-law, each with a constant clumping factor.

In Sect. 3.2, we demonstrated that the chemical abundances from HYACINTH are in a reasonable agreement with high-resolution molecular-cloud simulations. This shows that HYACINTH works reasonably well for modelling molecular hydrogen and carbon chemistry. However, our goal is not to replace extensive chemical networks that would be more accurate for modelling chemistry in ISM-scale simulations, but rather the model is tailored to follow hydrogen and carbon chemistry in large cosmological simulations that have a resolution of 20–200 pc.

In Sect. 4, we have shown a post-processing application for a galaxy simulation (at $z \sim 2.5$) with a spatial resolution of 156 pc. The resulting equilibrium H_2 abundance are in agreement with the commonly adopted KMT-EQ relation and the H_2 column density map shows a great similarity to the one obtained directly from the simulation (accounting for the non-equilibrium H_2 abundance). This shows that our approach gives reasonable results on larger scales (∼100–200 pc) that barely resolve individual molecular clouds.

An important caveat about using HYACINTH is that the formation of H_2 in HYACINTH relies on the presence of dust, similar to most other approaches in the literature (e.g. KMT). Therefore, it cannot be applied to gas with arbitrarily low metallicities ($Z \lesssim 10^{-3}\,Z_\odot$). As a result it cannot be applied to pristine gas such as in simulations of population III stars and early galaxy formation ($z \gtrsim 15$, see e.g. Hirano et al. 2015; Lenoble et al. 2024). Accurately modelling hydrogen chemistry in such (proto) galaxies requires accounting for three-body reactions and more importantly a high spatial resolution to resolve the densities where these reactions are efficient ($n_H \gtrsim 10^8\,cm^{-3}$). One workaround for cosmological simulations that do not resolve these early objects is to impose a metallicity floor that mimics the metal enrichment by the first stars (i.e. the Population III stars; see e.g. Kuhlen et al. 2012, 2013; Tomassetti et al. 2015; Pallottini et al. 2022).

6. Summary

We present a new sub-grid model, HYACINTH, that can be embedded into cosmological simulations for calculating the non-equilibrium abundances of H_2 and its carbon-based tracers, namely CO, C, and C^+. HYACINTH comprises a variable sub-grid density PDF to capture the unresolved density structure in

simulations and a simplified chemical network for hydrogen and carbon chemistry. These simplifications were introduced to make the network highly efficient for use in large-scale simulations. Additionally, a metallicity-dependent temperature-density relation, based on high-resolution simulations of the star-forming ISM, was used to assign a temperature to each sub-grid density.

We compared HYACINTH against more sophisticated approaches in the literature for modelling hydrogen and carbon chemistry including two extensive chemical networks (NL99 and G17) and a PDR code, using a one-dimensional semi-infinite slab setup (Fig. 2). HYACINTH reproduced the f_{H_2}-A_V relation from these methods highlighting that H_2 chemistry is insensitive to the exact treatment of carbon chemistry. Moreover, despite its simplicity and size, HYACINTH captured the $C^+ \rightarrow C$ and $C \rightarrow CO$ transitions very well, as predicted by more complex approaches (see Sect. 3.1 for a detailed discussion).

We compared the predictions from HYACINTH with high-resolution molecular-cloud simulations – the SILCC-Zoom simulations at $Z = Z_\odot$ and the GML11 simulations at $Z = 0.1\,Z_\odot$. We find reasonable agreement between the chemical abundances from HYACINTH and those from the simulations. We further find that our log-normal+power-law PDF shows better agreement with the SILCC-Zoom simulations while the log-normal PDF performs better for the GML11 runs, in alignment with the different density structures shown by the two simulations.

Finally, we applied HYACINTH to a simulated galaxy at $z \sim 2.5$ from T15 in post-processing and compared it directly with observations. For regions in the post-processed galaxy with MW-like conditions, the H I-to-H_2 transition in the f_{H_2}-N_H plane aligns very well with that from observations. The observed f_{H_2} and N_H were obtained from measurements of the absorption spectra of quasars and nearby stars in sightlines towards MW molecular clouds. The same is also true for LMC-like regions (Fig. 7). Additionally, the values of N_{CO}/N_{H_2} versus N_{H_2} in MW-like regions are consistent with observations of MW molecular clouds (Fig. 8). It is worth noting that most of our post-processed regions resemble the observed (optically) dark molecular clouds in the $N_{CO}/N_{H_2} - N_{H_2}$ plane. Furthermore, the relative abundance of atomic carbon to molecular hydrogen ($X_{C\,I}$) in our post-processed galaxy consistently matches the $X_{C\,I}$ values found in star-forming galaxies at redshifts $z \gtrsim 1$. Based on surface density profiles of the different baryonic components within the post-processed galaxy, we find an excess of C^+ at large distances from the galaxy centre, similar to those found in observations (e.g. Fujimoto et al. 2019).

In a forthcoming paper (paper II; Khatri et al., in prep.), we will present a suite of cosmological simulations using HYACINTH to model the non-equilibrium abundances of H_2 and its tracers in high-redshift ($z \gtrsim 2$) galaxies. This will open up the possibility of addressing fundamental questions such as what the contribution is of low-mass galaxies to the global H_2 budget at high z, what regulates the molecular gas fraction in galaxies, and for what physical conditions, environments, and galaxies are CO, C, and C^+ all reliable tracers of molecular gas.

Acknowledgements. The authors thank an anonymous reviewer and Pascal Oesch for suggestions and comments that helped improving the paper. They gratefully acknowledge the Collaborative Research Center 1601 (SFB 1601 sub-projects C5, B1, B4) funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 500700252. Early stages of this work were carried out within the Collaborative Research Centre 956 (SFB 956 sub-projects C4, C5, C6), funded by the DFG – 184018867. PK is a part of the International Max Planck Research School in Astronomy and Astrophysics, the Bonn Cologne Graduate School, and a guest at the Max Planck Institute for Radio Astronomy in Bonn. DS acknowledges funding from the programme "Profilbildung 2020", an initiative of the Ministry of Culture and Science of the State of North Rhine-Westphalia. CP is grateful to SISSA, the University of Trieste, and IFPU, where part of this work was carried out, for hospitality and support.

References

Alaghband-Zadeh, S., Chapman, S. C., Swinbank, A. M., et al. 2013, MNRAS, 435, 1493
Anicich, V. G., & Huntress, W. T., J. 1986, ApJS, 62, 553
Badnell, N. R. 2006, ApJS, 167, 334
Badnell, N. R., O'Mullane, M. G., Summers, H. P., et al. 2003, A&A, 406, 1151
Ballesteros-Paredes, J., Vázquez-Semadeni, E., Gazol, A., et al. 2011, MNRAS, 416, 1436
Barlow, S. E. 1984, PhD thesis, University of Colorado, Boulder, USA
Bolatto, A. D., Wolfire, M., & Leroy, A. K. 2013, ARA&A, 51, 207
Boogaard, L. A., van der Werf, P., Weiss, A., et al. 2020, ApJ, 902, 109
Buck, T., Pfrommer, C., Girichidis, P., & Corobean, B. 2022, MNRAS, 513, 1414
Burgh, E. B., France, K., & Jenkins, E. B. 2010, ApJ, 708, 334
Carniani, S., Maiolino, R., Smit, R., & Amorín, R. 2018, ApJ, 854, L7
Chevance, M., Kruijssen, J. M. D., Hygate, A. P. S., et al. 2020, MNRAS, 493, 2872
Chevance, M., Krumholz, M. R., McLeod, A. F., et al. 2023, in ASP Conf. Ser., 534, Protostars and Planets VII, eds. S. Inutsuka, Y. Aikawa, T. Muto, K. Tomida, & M. Tamura, 1
Christensen, C., Quinn, T., Governato, F., et al. 2012, MNRAS, 425, 3058
Crenny, T., & Federman, S. R. 2004, ApJ, 605, 278
Daddi, E., Bournaud, F., Walter, F., et al. 2010a, ApJ, 713, 686
Daddi, E., Elbaz, D., Walter, F., et al. 2010b, ApJ, 714, L118
Davé, R., Thompson, R., & Hopkins, P. F. 2016, MNRAS, 462, 3265
Decarli, R., Walter, F., Gónzalez-López, J., et al. 2019, ApJ, 882, 138
De Looze, I., Baes, M., Bendo, G. J., Cortese, L., & Fritz, J. 2011, MNRAS, 416, 2712
De Looze, I., Cormier, D., Lebouteiller, V., et al. 2014, A&A, 568, A62
de Ruette, N., Miller, K. A., O'Connor, A. P., et al. 2016, ApJ, 816, 31
Dickman, R. L., Snell, R. L., & Schloerb, F. P. 1986, ApJ, 309, 326
Dobbs, C. L., Glover, S. C. O., Clark, P. C., & Klessen, R. S. 2008, MNRAS, 389, 1097
Dobbs, C. L., Krumholz, M. R., Ballesteros-Paredes, J., et al. 2014, in Protostars and Planets VI, eds. H. Beuther, R. S. Klessen, C. P. Dullemond, & T. Henning, 3
Downes, D., & Solomon, P. M. 1998, ApJ, 507, 615
Draine, B. T. 1978, ApJS, 36, 595
Draine, B. T. 2003, ApJ, 598, 1017
Draine, B. T., & Bertoldi, F. 1996, ApJ, 468, 269
Federman, S. R., Huntress, W. T., J., & Prasad, S. S. 1990, ApJ, 354, 504
Federrath, C., & Klessen, R. S. 2013, ApJ, 763, 51
Förster Schreiber, N. M., & Wuyts, S. 2020, ARA&A, 58, 661
Fujimoto, S., Ouchi, M., Ferrara, A., et al. 2019, ApJ, 887, 107
Fujimoto, S., Silverman, J. D., Bethermin, M., et al. 2020, ApJ, 900, 1
Genzel, R., Tacconi, L. J., Gracia-Carpio, J., et al. 2010, MNRAS, 407, 2091
Geppert, W. D., Thomas, R. D., Ehlerding, A., et al. 2005, in J. Phys. Conf. Ser., 4, Journal of Physics Conference Series, 26
Gerin, M., & Phillips, T. G. 2000, ApJ, 537, 644
Girichidis, P., Konstandin, L., Whitworth, A. P., & Klessen, R. S. 2014, ApJ, 781, 91
Glassgold, A. E., & Langer, W. D. 1974, ApJ, 193, 73
Glover, S. C. O., & Clark, P. C. 2016, MNRAS, 456, 3596
Glover, S. C. O., & Mac Low, M.-M. 2007a, ApJS, 169, 239
Glover, S. C. O., & Mac Low, M.-M. 2007b, ApJ, 659, 1317
Glover, S. C. O., & Mac Low, M. M. 2011, MNRAS, 412, 337
Glover, S. C. O., & Clark, P. C. 2012, MNRAS, 421, 116
Glover, S. C. O., Federrath, C., Mac Low, M. M., & Klessen, R. S. 2010, MNRAS, 404, 2
Gnedin, N. Y., Tassis, K., & Kravtsov, A. V. 2009, ApJ, 697, 55
Gong, M., Ostriker, E. C., & Wolfire, M. G. 2017, ApJ, 843, 38
Habing, H. J. 1968, Bull. Astron. Inst. Netherlands, 19, 421
Hannon, S., Lee, J. C., Whitmore, B. C., et al. 2019, MNRAS, 490, 4648
Harrington, K. C., Weiss, A., Yun, M. S., et al. 2021, ApJ, 908, 95
Heays, A. N., Bosman, A. D., & van Dishoeck, E. F. 2017, A&A, 602, A105
Henríquez-Brocal, K., Herrera-Camus, R., Tacconi, L., et al. 2022, A&A, 657, L15
Herrera-Camus, R., Bolatto, A. D., Wolfire, M. G., et al. 2015, ApJ, 800, 1
Herrera-Camus, R., Förster Schreiber, N., Genzel, R., et al. 2021, A&A, 649, A31
Hirano, S., Hosokawa, T., Yoshida, N., Omukai, K., & Yorke, H. W. 2015, MNRAS, 448, 568

Hollenbach, D., Kaufman, M. J., Neufeld, D., Wolfire, M., & Goicoechea, J. R. 2012, ApJ, 754, 105
Hollyhead, K., Bastian, N., Adamo, A., et al. 2015, MNRAS, 449, 1106
Hopkins, P. F., Kereš, D., Oñorbe, J., et al. 2014, MNRAS, 445, 581
Hu, C.-Y., Sternberg, A., & van Dishoeck, E. F. 2021, ApJ, 920, 44
Hu, C.-Y., Sternberg, A., & van Dishoeck, E. F. 2023, ApJ, 952, 140
Hummer, D. G., & Storey, P. J. 1998, MNRAS, 297, 1073
Ikeda, M., Oka, T., Tatematsu, K., Sekimoto, Y., & Yamamoto, S. 2002, ApJS, 139, 467
Indriolo, N., Bergin, E. A., Falgarone, E., et al. 2018, ApJ, 865, 127
Kainulainen, J., Beuther, H., Henning, T., & Plume, R. 2009, A&A, 508, L35
Karakas, A. I. 2010, MNRAS, 403, 1413
Katz, H., Rosdahl, J., Kimm, T., et al. 2022, MNRAS, 510, 5603
Kim, J. K., Theard, L. P., & Huntress, W. T., J. 1975, Chem. Phys. Lett., 32, 610
Klessen, R. S. 2000, ApJ, 535, 869
Kosenko, D. N., Balashev, S. A., Noterdaeme, P., et al. 2021, MNRAS, 505, 3810
Kritsuk, A. G., Norman, M. L., & Wagner, R. 2011, ApJ, 727, L20
Kruijssen, J. M. D., Schruba, A., Chevance, M., et al. 2019, Nature, 569, 519
Krumholz, M. R. 2013, MNRAS, 436, 2747
Krumholz, M. R., McKee, C. F., & Tumlinson, J. 2008, ApJ, 689, 865
Krumholz, M. R., McKee, C. F., & Tumlinson, J. 2009, ApJ, 693, 216
Kuhlen, M., Krumholz, M. R., Madau, P., Smith, B. D., & Wise, J. 2012, ApJ, 749, 36
Kuhlen, M., Madau, P., & Krumholz, M. R. 2013, ApJ, 776, 34
Lagos, C. d. P., Crain, R. A., Schaye, J., et al. 2015, MNRAS, 452, 3815
Lambert, T. S., Posses, A., Aravena, M., et al. 2023, MNRAS, 518, 3183
Lenoble, R., Commerçon, B., & Rosdahl, J. 2024, A&A, 685, A7
Le Teuff, Y. H., Millar, T. J., & Markwick, A. J. 2000, A&AS, 146, 157
Liszt, H. 2003, A&A, 398, 621
Liszt, H. S. 2011, A&A, 527, A45
Lupi, A. 2019, MNRAS, 484, 1687
Lupi, A., & Bovino, S. 2020, MNRAS, 492, 2818
Lupi, A., Bovino, S., Capelo, P. R., Volonteri, M., & Silk, J. 2018, MNRAS, 474, 2884
Lupi, A., Pallottini, A., Ferrara, A., et al. 2020, MNRAS, 496, 5160
Madau, P., & Dickinson, M. 2014, ARA&A, 52, 415
Madden, S. C., Cormier, D., Hony, S., et al. 2020, A&A, 643, A141
McCall, B. J., Huneycutt, A. J., Saykally, R. J., et al. 2004, Phys. Rev. A, 70, 052716
McElroy, D., Walsh, C., Markwick, A. J., et al. 2013, A&A, 550, A36
McKee, C. F., & Krumholz, M. R. 2010, ApJ, 709, 308
McKee, C. F., & Ostriker, E. C. 2007, ARA&A, 45, 565
Micic, M., Glover, S. C. O., Federrath, C., & Klessen, R. S. 2012, MNRAS, 421, 2531
Muller, S., Müller, H. S. P., Black, J. H., et al. 2016, A&A, 595, A128
Nelson, R. P., & Langer, W. D. 1997, ApJ, 482, 796
Nelson, R. P., & Langer, W. D. 1999, ApJ, 524, 923
Nesvadba, N., Kneissl, R., Cañameras, R., et al. 2016, A&A, 593, A2
Neufeld, D. A., & Wolfire, M. G. 2016, ApJ, 826, 183
Nordlund, Å. K., & Padoan, P. 1999, in Interstellar Turbulence, eds. J. Franco, & A. Carraminana, 218
Oey, M. S., & Clarke, C. J. 1997, MNRAS, 289, 570
Oka, T. 2006, PNAS, 103, 12235
Padovani, M., Galli, D., & Glassgold, A. E. 2009, A&A, 501, 619
Pallottini, A., Ferrara, A., Bovino, S., et al. 2017, MNRAS, 471, 4128
Pallottini, A., Ferrara, A., Gallerani, S., et al. 2022, MNRAS, 513, 5621
Papadopoulos, P. P., Thi, W. F., & Viti, S. 2004, MNRAS, 351, 147
Parravano, A., Hollenbach, D. J., & McKee, C. F. 2003, ApJ, 584, 797
Passot, T., & Vázquez-Semadeni, E. 1998, Phys. Rev. E, 58, 4501
Pelupessy, F. I., & Papadopoulos, P. P. 2009, ApJ, 707, 954
Péroux, C., & Howk, J. C. 2020, ARA&A, 58, 363
Petzold, L. R. 1982, Description of DASSL: a differential/algebraic system solver (Sandia National Labs., Livermore, CA USA)
Popping, G., & Péroux, C. 2022, MNRAS, 513, 1531
Rachford, B. L., Snow, T. P., Tumlinson, J., et al. 2002, ApJ, 577, 221
Richings, A. J. ,& Schaye, J. 2016, MNRAS, 458, 270
Riechers, D. A., Pavesi, R., Sharon, C. E., et al. 2019, ApJ, 872, 7
Savage, B. D., & Sembach, K. R. 1996, ARA&A, 34, 279
Savage, B. D., Bohlin, R. C., Drake, J. F., & Budich, W. 1977, ApJ, 216, 291
Schäbe, A., Romano-Díaz, E., Porciani, C., Ludlow, A. D., & Tomassetti, M. 2020, MNRAS, 497, 5008
Schaerer, D., Ginolfi, M., Béthermin, M., et al. 2020, A&A, 643, A3
Schauer, M. M., Jefferts, S. R., Barlow, S. E., & Dunn, G. H. 1989, J. Chem. Phys., 91, 4593
Schruba, A., Leroy, A. K., Walter, F., et al. 2011, AJ, 142, 37
Seifried, D., Walch, S., Girichidis, P., et al. 2017, MNRAS, 472, 4797
Seifried, D., Haid, S., Walch, S., Borchert, E. M. A., & Bisbas, T. G. 2020, MNRAS, 492, 1465
Semenov, V. A., Kravtsov, A. V., & Gnedin, N. Y. 2018, ApJ, 861, 4
Sheffer, Y., Rogers, M., Federman, S. R., et al. 2008, ApJ, 687, 1075
Shull, J. M., Danforth, C. W., & Anderson, K. L. 2021, ApJ, 911, 55
Smith, R. J., Glover, S. C. O., Clark, P. C., Klessen, R. S., & Springel, V. 2014, MNRAS, 441, 1628
Sofia, U. J., Lauroesch, J. T., Meyer, D. M., & Cartledge, S. I. B. 2004, ApJ, 605, 272
Solomon, P. M., & Barrett, J. W. 1991, in Dynamics of Galaxies and Their Molecular Cloud Distributions, 146, eds. F. Combes, & F. Casoli, 235
Solomon, P. M., & Vanden Bout, P. A. 2005, ARA&A, 43, 677
Sternberg, A., Le Petit, F., Roueff, E., & Le Bourlot, J. 2014, ApJ, 790, 10
Tacconi, L. J., Neri, R., Chapman, S. C., et al. 2006, ApJ, 640, 228
Tacconi, L. J., Genzel, R., Smail, I., et al. 2008, ApJ, 680, 246
Tacconi, L. J., Genzel, R., Neri, R., et al. 2010, Nature, 463, 781
Teyssier, R. 2002, A&A, 385, 337
Thompson, R., Nagamine, K., Jaacks, J., & Choi, J.-H. 2014, ApJ, 780, 145
Tielens, A. G. G. M., & Hollenbach, D. 1985, ApJ, 291, 722
Tomassetti, M., Porciani, C., Romano-Diaz, E., Ludlow, A. D., & Papadopoulos, P. P. 2014, MNRAS, 445, L124
Tomassetti, M., Porciani, C., Romano-Díaz, E., & Ludlow, A. D. 2015, MNRAS, 446, 3330
Tumlinson, J., Shull, J. M., Rachford, B. L., et al. 2002, ApJ, 566, 857
Valentino, F., Magdis, G. E., Daddi, E., et al. 2018, ApJ, 869, 27
Vázquez-Semadeni, E. 1994, ApJ, 423, 681
Vissapragada, S., Buzard, C. F., Miller, K. A., et al. 2016, ApJ, 832, 31
Visser, R., van Dishoeck, E. F., & Black, J. H. 2009, A&A, 503, 323
Wakelam, V., Smith, I. W. M., Herbst, E., et al. 2010, Space Sci. Rev., 156, 13
Wakelam, V., Herbst, E., Loison, J. C., et al. 2012, ApJS, 199, 21
Walch, S., Girichidis, P., Naab, T., et al. 2015, MNRAS, 454, 238
Walter, F., Weiß, A., Downes, D., Decarli, R., & Henkel, C. 2011, ApJ, 730, 18
Walter, F., Carilli, C., Neeleman, M., et al. 2020, ApJ, 902, 111
Ward, R. L., Wadsley, J., & Sills, A. 2014, MNRAS, 445, 1575
Weingartner, J. C., & Draine, B. T. 2001, ApJ, 563, 842
Weiß, A., Henkel, C., Downes, D., & Walter, F. 2003, A&A, 409, L41
Weiß, A., Downes, D., Henkel, C., & Walter, F. 2005, A&A, 429, L25
Wolfire, M. G., Tielens, A. G. G. M., Hollenbach, D., & Kaufman, M. J. 2008, ApJ, 680, 384
Wolfire, M. G., Hollenbach, D., & McKee, C. F. 2010, ApJ, 716, 1191
Woodall, J., Agúndez, M., Markwick-Kemper, A. J., & Millar, T. J. 2007, A&A, 466, 1197
Yang, C., Omont, A., Martín, S., et al. 2023, A&A, 680, A95
Zanchet, A., Bussery-Honvault, B., Jorfi, M., & Honvault, P. 2009, Phys. Chem. Chem. Phys., 11, 6182

Appendix A: Reactions in the chemical network

Table A.1 gives a list of all chemical reactions in HYACINTH. For each of these, we use updated reaction rates from UMIST (McElroy et al. 2013) and KIDA (Wakelam et al. 2012) databases.

Table A.1. Reactions in our chemical network along with the rate coefficients.

	Reaction	Type	Rate coefficient	Reference
1	$C^+ + H_2 \rightarrow CH_x + H$	H_2 destruction, CH_x formation	$2.31 \times 10^{-13}\, T^{-1.3}\, \exp(-23/T)$	1,19
2	$C^+ + H_2 \rightarrow C + 2H$	H_2 destruction, C formation	$0.99 \times 10^{-13}\, T^{-1.3}\, \exp(-23/T)$	1
3	$H_3^+ + C \rightarrow CH_x + H_2$	CH_x formation	$1.04 \times 10^{-9}\, (300/T)^{0.00231} + T^{-1.5} \Sigma_{i=1}^{4}\, c_i\, \exp(-T_i/T)$; $c_i = [3.4 \times 10^{-8}, 6.97 \times 10^{-9}, 1.31 \times 10^{-7}, 1.51 \times 10^{-4}]$, $T_i = [7.62, 1.38, 26.6, 8110]$	2, 3
4	$H_3^+ + O \rightarrow OH_x + H_2$	OH_x formation	$1.99 \times 10^{9}\, T^{-0.190}$	3, 4
5	$CH_x + O \rightarrow CO + H$	CO formation	7.7×10^{-11}	1
6	$OH_x + C \rightarrow CO + H$	CO formation	$7.95 \times 10^{-10}\, T^{-0.339}\, \exp(0.108/T)$	1, 5
7	$H_3^+ + e^- \rightarrow H_2 + H$	Dissociative recombination	$4.54 \times 10^{-7} \times T^{-0.52}$	11, 12
8	$H_3^+ + e^- \rightarrow 3H$	Dissociative recombination	$8.46 \times 10^{-7} \times T^{-0.52}$	11, 12
9	$C^+ + e^- \rightarrow C + \gamma$	Radiative recombination	$\frac{2.995 \times 10^{-9}}{\alpha(1+\alpha)^{1-\gamma}\,(1+\beta)^{1+\gamma}}$	13,14
10	$C + \gamma \rightarrow C^+ + e^-$	Photoionisation	$3.5 \times 10^{-10}\, (G_0/1.7)\, \exp(-3.76\, A_V)\, f_{s,C}(N_C, N_{H_2})$	6, 7
11	$H_2 + \gamma \rightarrow 2H$	Photodissociation	$4.2 \times 10^{-11}\, G_0\, \exp(-4.18\, A_V)\, f_{s,H_2}(N_{H_2})$	6, 7
12	$CO + \gamma \rightarrow C + O$	Photodissociation	$2.4 \times 10^{-10}\, (G_0/1.7)\, \exp(-3.88\, A_V)\, f_{s,CO}(N_{CO}, N_{H_2})$	6, 8
13	$CH_x + \gamma \rightarrow C + H$	Photodissociation	$9.1 \times 10^{-10}\, (G_0/1.7)\, \exp(-2.12\, A_V)$	6
14	$OH_x + \gamma \rightarrow O + H$	Photodissociation	$3.8 \times 10^{-10}\, (G_0/1.7)\, \exp(-2.66\, A_V)$	6
15	$H_2 + CR \rightarrow {H_2}^+ + e^-$	Cosmic-ray ionisation	$2\zeta_H$	20
16	$C + CR \rightarrow C^+ + e^-$	Cosmic-ray ionisation	$3.85\zeta_H$	20
17	$CO + CR \rightarrow CO^+ + O$	Cosmic-ray ionisation	$6.52\zeta_H$	20,22
18	$C^+ + e^- + grain \rightarrow C$	Grain-assisted recombination	$4.558 \times 10^{-13}\, [1 + 6.089 \times 10^{-3}\, \psi^{1.128} \times (1 + 433.1\, T^{0.04845}\, \psi^{-0.8120 - 1.333 \times 10^{-4} \ln T})]^{-1}$;	15
19	$H + H + grain \rightarrow H_2 + grain$	Grain-assisted formation of H_2	3×10^{-17}	17, 18
20	$He + CR \rightarrow He^+ + e^-$	Cosmic-ray ionisation	$1.1\zeta_H$	20,22
21	$He^+ + H_2 \rightarrow H^+ + He + H$	Dissociative charge exchange	$1.26 \times 10^{-13}\, \exp(-22.5/T)$	3,27
22	$He^+ + H_2 \rightarrow He + H_2^+$	Charge exchange	7.2×10^{-15}	25, 26
23	$He^+ + CO \rightarrow He + C^+ + O$	CO destruction	1.6×10^{-9}	23
24	$He^+ + e^- \rightarrow He + \gamma$	Radiative recombination	$10^{-11}\, T^{-0.5} \times [11.19 - 1.676 \log_{10}\mathrm{T} - 0.2852\, (\log_{10}\mathrm{T})^2 + 0.04433\, (\log_{10}\mathrm{T})^3]$	21,22
25	$He^+ + e^- + grain \rightarrow He + \gamma$	Grain-assisted recombination	$5.572 \times 10^{-14}\, [1 + 3.185 \times 10^{-7} \psi^{1.512} \times (1 + 5115\, T^{3.903 \times 10^{-7}}\, \psi^{-0.4956 - 5.494 \times 10^{-7} \ln T})]^{-1}$;	
26	$H_3^+ + CO \rightarrow HCO^+ + H_2$	CO destruction, HCO^+ formation	1.7×10^{-9}	1
27	$C^+ + OH_x \rightarrow HCO^+$	C^+ destruction, HCO^+ formation	$9.15 \times 10^{-10}\, (0.62 + 45.41\, T^{-1/2}$	1
28	$HCO^+ + e^- \rightarrow CO + H$	Dissociative recombination	$1.06 \times 10^{-5}\, T^{-0.64}$	24
29	$HCO^+ + \gamma \rightarrow CO + H^+$	Photodissociation	$5.4 \times 10^{-12}\, (G_0/1.7)\, \exp(-3.3\, A_V)$	26

Notes. The rate coefficients are in $cm^3\ s^{-1}$ for reactions 1-9, 21-24, and 26-28; in s^{-1} for reactions 10-17, 20, and 29; in $cm^3\ s^{-1}\ Z_d^{-1}$ for 18, 19, and 25, where Z_d is the dust abundance relative to the solar neighbourhood value of 0.01. ζ_H is the value of the cosmic ray ionisation rate in units of $s^{-1}\ H^{-1}$; G_0 is the flux of the radiation field in the Habing units; A_V is the visual extinction defined in Eq. (10). In reaction 9, $\alpha = \sqrt{T/(6.67 \times 10^{-3})}$, $\beta = \sqrt{T/(1.9436 \times 10^{6})}$, and $\gamma = 0.7849 + 0.1597 \exp(49550/T)$. In reactions 18 and 25, $\psi = G_0 \exp(-1.87\, A_V)\ \sqrt{T}/(n_{e^-}/cm^{-3})$. We note that although Heays et al. (2017) provided an updated value of the unshielded photodissociation rate for reaction 11 of $5.7 \times 10^{-11}\ (G_0/1.7)$, here we use the old rate from Draine & Bertoldi (1996), that is $\approx 25\%$ lower, for a fair comparison with T15. Reactions 20-29 are only part of extended HYACINTH.

References. (1) Wakelam et al. (2012); (2) Vissapragada et al. (2016); (3) Gong et al. (2017); (4) de Ruette et al. (2016); (5) Zanchet et al. (2009); (6) Heays et al. (2017); (7) Draine & Bertoldi (1996); (8) Visser et al. (2009); (9) Glassgold & Langer (1974); (10) Liszt (2003); (11) McCall et al. (2004); (12) Woodall et al. (2007); (13) Badnell et al. (2003); (14) Badnell (2006); (15) Weingartner & Draine (2001); (16) Draine (2003); (17) Wolfire et al. (2008); (18) Hollenbach et al. (2012); (19) Anicich & Huntress (1986); (20) Le Teuff et al. (2000); (21) Hummer & Storey (1998); (22) Glover et al. (2010); (23) Kim et al. (1975); (24) Geppert et al. (2005); (25) Barlow (1984); (26) McElroy et al. (2013); (27) Schauer et al. (1989).

The fractional abundance f_X of a species X relative to hydrogen is defined as $f_X = n_X/n_H$, except for H_2 and H_3^+, where $f_{H_2} = 2n_{H_2}/n_H$ and $f_{H_3^+} = 3n_{H_3^+}/n_H$. For H_2, CO, and C^+, the differential equations for each sub-grid density n_H can be written as[9]

$$\frac{\mathrm{d}f_{H_2}}{\mathrm{d}t} = 2k_{H_2,gr}\, f_{H\,I}\, n_H - k_{\gamma,H_2}\, f_{H_2} - k_{c.r.,H_2}\, f_{H_2} - n_H\, k_{C^+,H_2}\, f_{C^+}\, f_{H_2}\, \alpha_{CH_x} + \frac{2}{3}\, n_H\, k_{e^-,H_3^+}\, f_{e^-}\, f_{H_3^+}\,; \quad \text{(A.1)}$$

$$\frac{\mathrm{d}f_{CO}}{\mathrm{d}t} = \frac{1}{3}\, n_H\, k_{O\,I,H_3^+}\, f_{O\,I}\, f_{H_3^+}\, \alpha_{OH_x} + \frac{1}{3}\, n_H\, k_{C\,I,H_3^+}\, f_{C\,I}\, f_{H_3^+}\, \alpha_{CH_x} + \frac{1}{2}\, n_H\, k_{C^+,H_2}\, f_{C^+}\, f_{H_2}\, \alpha_{CH_x} - k_{\gamma,CO}\, f_{CO}\,; \quad \text{(A.2)}$$

$$\frac{\mathrm{d}f_{C^+}}{\mathrm{d}t} = -n_H\, k_{C^+,e^-}\, f_{C^+}\, f_{e^-} - n_H\, k_{C^+,e^-,gr}\, f_{C^+}\, f_{e^-} - \frac{1}{2}\, n_H\, k_{C^+,H_2}\, f_{C^+}\, f_{H_2}\, \alpha_{CH_x} + k_{\gamma,C\,I}\, f_{C\,I}\,. \quad \text{(A.3)}$$

The reaction rate coefficient k for each reaction is listed in Table A.1. In the above equations, α_{CH_x} (α_{OH_x}) is the branching ratio (see Sect. B.1) for the formation of CH_x (OH_x) when H_3^+ reacts with atomic carbon (oxygen). The species O_2, H_2O, OH, OH^+, H_2O^+, and H_3O^+ are collectively referred to as OH_x in the chemical network (as originally done in NL99 based on the argument that these species have similar chemical reactions and reaction rates). Similarly, CH, CH_2, CH^+, and CH_2^+ are collectively referred to as CH_x. For H_3^+, we assume a local equilibrium (i.e. at each sub-grid density) between the formation and destruction pathways such that:

$$\frac{\mathrm{d}f_{H_3^+}}{\mathrm{d}t} = \frac{3}{2}\, k_{c.r.,H_2}\, f_{H_2} - n_H\, k_{C\,I,H_3^+}\, f_{C\,I}\, f_{H_3^+} - n_H\, k_{O\,I,H_3^+}\, f_{O\,I}\, f_{H_3^+} - n_H\, k_{e^-,H_3^+}\, f_{e^-}\, f_{H_3^+} = 0\,, \quad \text{(A.4)}$$

which implies

$$f_{H_3^+} = \frac{3}{2\, n_H}\, \frac{k_{c.r.,H_2}\, f_{H_2}}{k_{C\,I,H_3^+}\, f_{C\,I} + k_{O\,I,H_3^+}\, f_{O\,I} + k_{e^-,H_3^+}\, f_{e^-}}\,. \quad \text{(A.5)}$$

Similarly, an equilibrium abundance of OH_x and CH_x implies

$$\frac{\mathrm{d}f_{OH_x}}{\mathrm{d}t} = \frac{1}{3} k_{O\,I,H_3^+}\, f_{O\,I} f_{H_3^+} n_H - k_{C\,I,OH_x} f_{C\,I} f_{OH_x} n_H - k_{\gamma,OH_x}\, f_{OH_x} = 0\,; \quad \text{(A.6)}$$

$$\frac{\mathrm{d}f_{CH_x}}{\mathrm{d}t} = \frac{1}{3} k_{C\,I,H_3^+} f_{C\,I} f_{H_3^+} n_H + \frac{1}{2} k_{C^+,H_2} f_{C^+} f_{H_2} n_H - k_{O\,I,CH_x} f_{O\,I} f_{CH_x} n_H - k_{\gamma,CH_x}\, f_{CH_x} = 0\,. \quad \text{(A.7)}$$

These give

$$f_{OH_x} = \frac{\frac{1}{3} k_{O\,I,H_3^+} f_{O\,I} f_{H_3^+}}{k_{C\,I,OH_x} f_{C\,I} + k_{\gamma,OH_x}/n_H}\,; \quad \text{(A.8)}$$

$$f_{CH_x} = \frac{\frac{1}{3} k_{C\,I,H_3^+} f_{C\,I} f_{H_3^+} + \frac{1}{2} k_{C^+,H_2} f_{C^+} f_{H_2}}{k_{O\,I,CH_x} f_{O\,I} + k_{\gamma,CH_x}/n_H}\,. \quad \text{(A.9)}$$

[9] Atomic carbon is denoted as C I here.

Appendix B: Cell-averaged rate equations

Obtaining the differential equation for the evolution of each species within a given region (e.g. a cell in a cosmological simulation) requires calculating a PDF-weighted integral of each term on the right-hand side in Eqs. (A.1)–(A.3). This leads to the following set of differential equations for the cell-averaged abundances:

$$\begin{aligned}\frac{\mathrm{d}\langle f_{H_2}\rangle}{\mathrm{d}t} &= \langle 2k_{H_2,gr}\, f_{H\,I}\, n_H\rangle - \langle k_{\gamma,H_2}\, f_{H_2}\rangle - \langle k_{c.r.,H_2}\, f_{H_2}\rangle \\ &- \langle n_H\, k_{C^+,H_2}\, f_{C^+}\, f_{H_2}\, \alpha_{CH_x}\rangle + \left\langle \frac{2}{3}\, n_H\, k_{e^-,H_3^+}\, f_{e^-}\, f_{H_3^+}\right\rangle \\ &\underbrace{\int_0^{n_{crit,H_2}} 2k_{H_2,gr}\, n_H\, \mathcal{P}_M(n_H)\, \mathrm{d}n_H}_{A_1} - \underbrace{\int_{n_{crit,H_2}}^{\infty} k_{\gamma,H_2}\, \mathcal{P}_M(n_H)\, \mathrm{d}n_H}_{A_2} \\ &- \underbrace{\int_{n_{crit,H_2}}^{\infty} k_{c.r.,H_2}\, \mathcal{P}_M(n_H)\, \mathrm{d}n_H}_{A_3} - \underbrace{\int_{n_{crit,H_2}}^{n_{crit,C\,I}} n_H\, k_{C^+,H_2}\, f_{C,tot}\, \alpha_{CH_x}\, \mathcal{P}_M(n_H)\, \mathrm{d}n_H}_{A_4} \\ &+ \underbrace{\int_{n_{crit,H_2}}^{\infty} \frac{2}{3}\, n_H\, k_{e^-,H_3^+}\, f_{e^-}\, f_{H_3^+}\, \mathcal{P}_M(n_H)\, \mathrm{d}n_H}_{A_5}\,; \end{aligned} \quad \text{(B.1)}$$

$$\begin{aligned}\frac{\mathrm{d}\langle f_{CO}\rangle}{\mathrm{d}t} &= \left\langle \frac{1}{3}\, n_H\, k_{O\,I,H_3^+}\, f_{O\,I}\, f_{H_3^+}\, \alpha_{OH_x}\right\rangle + \left\langle \frac{1}{3}\, n_H\, k_{C\,I,H_3^+}\, f_{C\,I}\, f_{H_3^+}\, \alpha_{CH_x}\right\rangle \\ &+ \left\langle \frac{1}{2}\, n_H\, k_{C^+,H_2}\, f_{C^+}\, f_{H_2}\, \alpha_{CH_x}\right\rangle - \langle k_{\gamma,CO}\, f_{CO}\rangle \\ &= \underbrace{\int_{n_{crit,H_2}}^{\infty} \frac{1}{3}\, n_H\, k_{O\,I,H_3^+}\, f_{O,I}\, f_{H_3^+}\, \alpha_{CH_x}\, \mathcal{P}_M(n_H)\, \mathrm{d}n_H}_{B_1} \\ &+ \underbrace{\int_{\max(n_{crit,H_2}, n_{crit,C\,I})}^{n_{crit,CO}} \frac{1}{3}\, n_H\, k_{C\,I,H_3^+}\, f_{C,tot}\, f_{H_3^+}\, \alpha_{CH_x}\, \mathcal{P}_M(n_H)\, \mathrm{d}n_H}_{B_2} \\ &+ \underbrace{\int_{n_{crit,H_2}}^{n_{crit,C\,I}} \frac{1}{2}\, n_H\, k_{C^+,H_2}\, f_{C,tot}\, \alpha_{CH_x}\, \mathcal{P}_M(n_H)\, \mathrm{d}n_H}_{B_3} \\ &- \underbrace{\int_{n_{crit,CO}}^{\infty} k_{\gamma,CO}\, f_{C,tot}\, \mathcal{P}_M(n_H)\, \mathrm{d}n_H}_{B_4}\,; \end{aligned} \quad \text{(B.2)}$$

$$\begin{aligned}\frac{\mathrm{d}\langle f_{C^+}\rangle}{\mathrm{d}t} &= -\langle n_H\, k_{C^+,e^-}\, f_{C^+}\, f_{e^-}\rangle - \langle n_H\, k_{C^+,e^-,gr}\, f_{C^+}\, f_{e^-}\rangle \\ &- \langle \frac{1}{2}\, n_H\, k_{C^+,H_2}\, f_{C^+}\, f_{H_2}\, \alpha_{CH_x}\rangle + \langle k_{\gamma,C\,I}\, f_{C\,I}\rangle \\ &= -\underbrace{\int_0^{n_{crit,C\,I}} n_H\, (k_{C^+,e^-} + k_{C^+,e^-,gr})\, f_{C,tot}\, f_{e^-}\, \mathcal{P}_M(n_H)\, \mathrm{d}n_H}_{C_1} \\ &- \underbrace{\int_{n_{crit,H_2}}^{n_{crit,C\,I}} \frac{1}{2}\, n_H\, k_{C^+,H_2}\, f_{C,tot}\, \alpha_{CH_x}\, \mathcal{P}_M(n_H)\, \mathrm{d}n_H}_{C_2} \\ &+ \underbrace{\int_{n_{crit,C\,I}}^{n_{crit,CO}} k_{\gamma,C\,I}\, f_{C,tot}\, \mathcal{P}_M(n_H)\, \mathrm{d}n_H}_{C_3}\,. \end{aligned} \quad \text{(B.3)}$$

In the above equations, n_{crit,H_2}, $n_{crit,C\,I}$, and $n_{crit,CO}$ are the critical densities for the HI → H_2, C^+ → C, and C I → CO transitions, respectively

in a given cell (see Sect. 2.2.7 and Appendix C). The $f_{H_3^+}$ in these equations can be obtained from Eq. (A.5). We note that the limits of integrals A_5 and B_1 are from $n_{\rm crit,H_2}$ to ∞ as the formation of H_3^+ relies on the presence of H_2 (see Eq. A.4), but the exact expression for $f_{H_3^+}$ would be different for $n_{\rm H} < n_{\rm crit,C\,I}$ (where $f_{\rm C^+} = f_{\rm C,tot}$, $f_{\rm C\,I} = f_{\rm CO} = 0$), $n_{\rm crit,C\,I} \leq n_{\rm H} < n_{\rm crit,CO}$ (where $f_{\rm C\,I} = f_{\rm C,tot}$, $f_{\rm C^+} = f_{\rm CO} = 0$), and $n_{\rm H} \geq n_{\rm crit,CO}$ (where $f_{\rm CO} = f_{\rm C,tot}$, $f_{\rm C^+} = f_{\rm C\,I} = 0$). Furthermore, the integrals A_4, B_3 and C_2 will vanish for a given cell if $n_{\rm crit,H_2} > n_{\rm crit,C\,I}$.

In Eqs. (B.1)-(B.3), the resulting system of coupled differential equations is solved using the implicit integrator DASSL[10] (Petzold 1982). DASSL uses a variable order (between 1 and 5) backward differential formula to compute the solution of the coupled ODEs after one time step.

B.1. Branching ratios

Often there are multiple outcomes for the reaction between two species. The probability for a given outcome is represented as a branching ratio for that outcome and it denotes the fraction of times that particular outcome will occur. For example, the reaction between C^+ and H_2 leads to the formation of CH_2^+, which reacts with H_2 to give CH_3^+. CH_3^+ reacts with an electron to give CH or CH_2 in 70% of the cases (these two species are referred to as CH_x in reaction 1 in Table A.1). In the remaining 30% of the cases, $CH_3^+ + e^-$ gives C + 2H (reaction 2). The reaction rates in Table A.1 account for these branching ratios.

Appendix C: Calculation of $n_{\rm crit}$

The hydrogen in a region can exist in atomic (H I) and molecular (H_2) forms and their mean densities are related to the mean density $\langle n_{\rm H}\rangle$ of H nuclei in the cell as

$$\langle n_{\rm H\,I}\rangle + 2\langle n_{\rm H_2}\rangle = \langle n_{\rm H}\rangle\,. \tag{C.1}$$

The mean H_2 fraction in a region is defined as

$$f_{\rm H_2} = 2\frac{\langle n_{\rm H_2}\rangle}{\langle n_{\rm H}\rangle}\,. \tag{C.2}$$

Assuming that hydrogen shows a sharp transition from fully atomic to fully molecular at $n_{\rm H} = n_{\rm crit,H_2}$, Eq. (C.1) becomes

$$\int_0^{n_{\rm crit,H_2}} \mathcal{P}_{\rm M}\,{\rm d}n_{\rm H} + \underbrace{\int_{n_{\rm crit,H_2}}^{\infty} \mathcal{P}_{\rm M}\,{\rm d}n_{\rm H}}_{2\langle n_{\rm H_2}\rangle/\langle n_{\rm H}\rangle} = 1. \tag{C.3}$$

For a log-normal $\mathcal{P}_{\rm M}$, this becomes

$$\frac{1}{2}\left[1 + {\rm Erf}\left(\frac{\ln n_{\rm crit,H_2} - \mu}{\sqrt{2}\,\sigma}\right)\right] + 2\frac{\langle n_{\rm H_2}\rangle}{\langle n_{\rm H}\rangle} = 1. \tag{C.4}$$

If $\langle n_{\rm H}\rangle$ and $f_{\rm H_2}$ are known quantities, then the above equation can be written as

$$f_{\rm H_2} = \frac{1}{2}\left[1 - {\rm Erf}\left(\frac{\ln n_{\rm crit,H_2} - \mu}{\sqrt{2}\,\sigma}\right)\right], \tag{C.5}$$

where $\mu = \langle n_{\rm H}\rangle + 0.5\sigma^2$. The critical density $n_{\rm crit,H_2}$ can be obtained by finding the root of the above equation. This equation is equivalent to equation (10) in T15.

For a log-normal+power-law PDF, Eq.(C.3) becomes

$$\left[\int_0^{n_{\rm trans}} \mathcal{P}_{\rm M1}\,{\rm d}n_{\rm H} + \int_{n_{\rm trans}}^{n_{\rm crit,H_2}} \mathcal{P}_{\rm M2}\,{\rm d}n_{\rm H}\right] + 2\frac{\langle n_{\rm H_2}\rangle}{\langle n_{\rm H}\rangle} = 1\,. \tag{C.6}$$

We note that the second term in the brackets will be zero for $n_{\rm crit,H_2} < n_{\rm trans}$. While using an iterative root-finding method such as the Newton-Raphson method, one needs to evaluate the terms in the brackets differently depending on whether a given guess value for the root $n_{\rm crit,H_2}$ is smaller or larger than $n_{\rm trans}$. Thus, in this case, it is not possible to obtain an analytical expression relating $\langle n_{\rm H}\rangle$, $f_{\rm H_2}$, and $n_{\rm crit,H_2}$.

Appendix D: Effect of the CRIR on chemistry

In order to investigate the impact on the chemistry of our assumptions regarding the scaling of the CRIR, we extend the results presented in Sect. 4 by repeating the post-processing of the simulated galaxy and considering different options. In Fig. D.1, we plot the ratio of the CO abundance within a cell without and with the upper limit on the CRIR introduced in Sect. 2.2.5. At a given $\zeta_{\rm H}$, the effect of imposing an upper bound on $\zeta_{\rm H}$ is strongest at low densities and decreases with increasing density. Consequently, the mass-weighted ratio of the CO abundance in the two cases is very close to 1 for CRIR values up to $10^{-11}\,{\rm s^{-1}\,H^{-1}}$, which is larger than the highest observational estimate[11] of $3\times10^{-12}\,{\rm s^{-1}\,H^{-1}}$ reported in Yang et al. (2023, denoted by a blue arrow in Fig. D.1).

In Fig. D.2, we show the $f_{\rm H_2}$ obtained in each cell when using a variable $\zeta_{\rm H}$ in post-processing. We consider three different CRIRs: (i) $\zeta_{\rm H} \propto \chi^2$, with an upper limit of $3\times10^{-14}\,{\rm s^{-1}\,H^{-1}}$ on $\zeta_{\rm H}$ (left panel); (ii) $\zeta_{\rm H} \propto \chi^2$ without any upper limit (middle panel); (iii) the $\zeta_{\rm H} - \chi$ relation from Hu et al. (2021, right panel). The median $f_{\rm H_2}$ in each case is shown by a solid black line (also shown

[11] Till date and to the best of our knowledge.

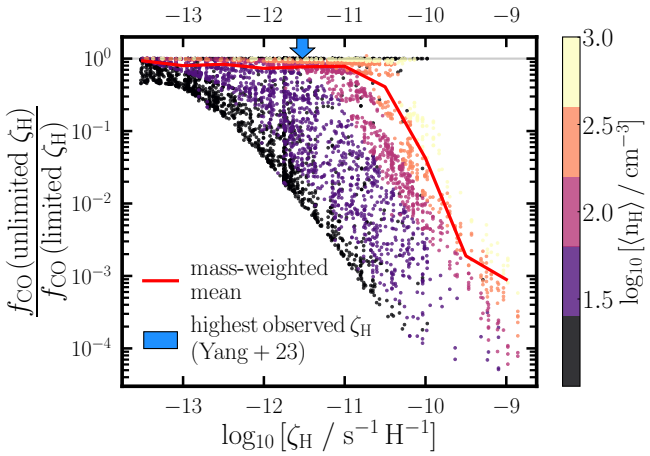


Fig. D.1. Ratio of the $f_{\rm CO}$ obtained when using a $\zeta_{\rm H} \propto \chi^2$without and with an upper limit on $\zeta_{\rm H}$, as a function of $\zeta_{\rm H}$ in the post-processed galaxy. Here we only show the grid cells with $\zeta_{\rm H}$ greater than the imposed upper limit of $3\times10^{-14}\,{\rm s^{-1}\,H^{-1}}$, since the two $f_{\rm CO}$ are identical for lower CRIR values. Each grid cell is colour-coded by the mean hydrogen density $\langle n_{\rm H}\rangle$ in the cell. The red line shows the mass-weighted mean of the cells in each $\zeta_{\rm H}$ bin. We also indicate by a blue arrow the highest observed $\zeta_{\rm H}$ of $3\times10^{-12}\,{\rm s^{-1}\,H^{-1}}$ (from Yang et al. 2023).

[10] https://www.osti.gov/servlets/purl/5882821

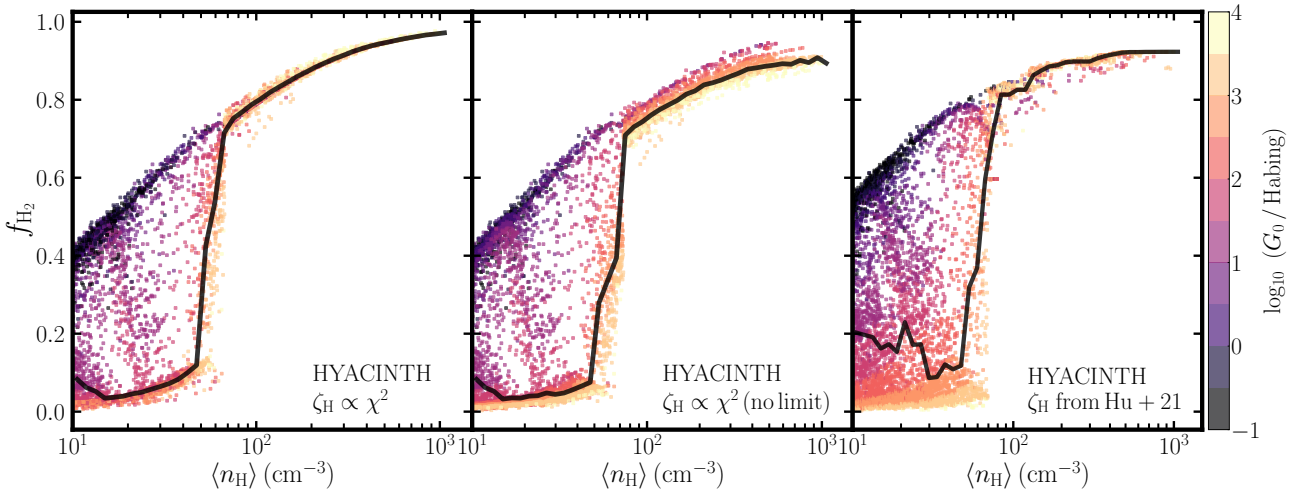


Fig. D.2. Same as the left panel of Fig. 6, but for a different CRIR used in post-processing (indicated in each panel). The solid black line shows the median f_{H_2} in a given $\langle n_H \rangle$ bin.

Table D.1. Total mass of the different chemical species in the post-processed galaxy for different choices of the CRIR. For reference, the (dynamically evolved) M_{H_2} from the simulation is $4.21 \times 10^{10}\,M_\odot$.

Method	ζ_H	M_{H_2} ($10^{10}\,M_\odot$)	M_{CO} ($10^7\,M_\odot$)	$M_{C\,I}$ ($10^7\,M_\odot$)	M_{C^+} ($10^7\,M_\odot$)
HYACINTH	$\zeta_{H,MW}$	4.70	9.54	0.63	0.62
HYACINTH	default	4.12	9.38	0.66	0.66
HYACINTH	$\propto \chi^2$ (no ceiling)	3.84	6.52	0.90	1.64
HYACINTH	Hu et al. (2021)	4.08	9.38	0.67	0.65

in the left panel of Fig. 6). We find slight variations among the three cases for the individual cells. Nevertheless, at low densities ($\langle n_H \rangle \lesssim 50\,\mathrm{cm}^{-3}$), f_{H_2} decreases with an increase in G_0 (the UV field in LW bands in Habing units, denoted by the colour of the points).

Finally, in Table D.1, we report the total mass of the different chemical species in the post-processed galaxy. Within the accuracy that can be expected from our simplified calculations, HYACINTH provides stable predictions for the masses that are not influenced much by the assumed scaling between the CRIR and the UV flux. The quadratic scaling with no upper limit gives the most discrepant results, with a 30% reduction in the CO mass compensated by an increase in the masses of neutral and ionised atomic carbon. In contrast, the H_2 mass is reduced only by ~7%. The results obtained with our default choice and the CRIR from Hu et al. (2021) are nearly identical.

APPENDIX C

The [C II] line emission as an interstellar medium probe in the MARIGOLD galaxies

The version of the paper *Khatri et al. arXiv:2411.09755* that has been submitted for publication to A&A is reproduced below in the original form.

Astronomy & Astrophysics manuscript no. main
November 18, 2024

The [C II] line emission as an ISM probe in the MARIGOLD galaxies

Prachi Khatri, Emilio Romano-Díaz, and Cristiano Porciani

Argelander Institute für Astronomie, Auf dem Hügel 71, D-53121 Bonn, Germany
e-mail: pkhatri@astro.uni-bonn.de

Received / Accepted

ABSTRACT

Context. The [C II] fine-structure line at 157.74 μm is one of the brightest far-infrared emission lines in galaxies and an important probe of galaxy properties such as the star formation rate (SFR) and the molecular gas mass ($M_{\rm mol}$).
Aims. Using high-resolution numerical simulations, we test the reliability of the [C II] line as a tracer of $M_{\rm mol}$ in high-redshift galaxies and investigate secondary dependences of the [C II] – $M_{\rm mol}$ relation on the SFR and metallicity. We also investigate the time evolution of the [C II] luminosity function (LF) and the relative spatial extent of [C II] emission and star formation.
Methods. We post-process galaxies from the MARIGOLD cosmological simulations at redshifts $3 \leq z \leq 7$ to obtain their [C II] emission. These simulations were performed with the sub-grid chemistry model, HYACINTH, to track the non-equilibrium abundances of H_2, CO, C and C^+ on the fly. Based on a statistical sample of galaxies at these redshifts, we investigate correlations between the [C II] line luminosity ($L_{\rm [C\,II]}$) and the SFR, the $M_{\rm mol}$, the total gas mass and the metal mass in gas phase ($M_{\rm metal}$).
Results. We find that accounting for secondary dependencies in the $L_{\rm [C\,II]} - M_{\rm mol}$ relation improves the $M_{\rm mol}$ prediction by a factor of 2.3 at all redshifts. Our simulations predict a mild evolution in the slope of the $L_{\rm [C\,II]}$ – SFR relation ($\lesssim 0.15$ dex) and an increase in the intercept by 0.5 dex in the above redshift range. Among the various galaxy properties we explore, the [C II] emission in our simulated galaxies shows the tightest correlation with $M_{\rm metal}$, indicating the potential of this line to constrain the metallicity of high-redshift galaxies. About 20% (10%) of our simulated galaxies at $z = 5$ ($z = 4$) have [C II] emission extending ≥ 2 times farther than the star formation activity. The [C II] LF evolves rapidly and is always well approximated by a double power law that does not show an exponential cutoff at the bright end. We record a 600-fold increase in the number density of $L_{\rm [C\,II]} \sim 10^9\,{\rm L}_\odot$ emitters in 1.4 Gyr.

Key words. methods: numerical – galaxies: high-redshift – galaxies: formation – galaxies: emission

1. Introduction

The fine-structure line of singly-ionised carbon (C^+) at a wavelength of 157.74 μm (hereafter [C II]), is an important coolant in the interstellar medium (ISM) of galaxies near and far. Being one of the brightest lines in the infrared, accounting for $\sim 0.1 - 1\%$ of the total infrared luminosity in star-forming galaxies (Díaz-Santos et al. 2013), it is particularly useful for observing high-redshift ($z \gtrsim 4$) galaxies, where it is conveniently redshifted to the transparent atmospheric window at millimetre wavelengths and is accessible from the ground with the Atacama Large Millimeter/submillimeter Array (ALMA) and Northern Extended Millimeter Array (NOEMA), among others.

The strength of this line has been shown to correlate well with both the galaxy-integrated star formation rate (SFR; Stacey et al. 2010; De Looze et al. 2011, 2014; Carniani et al. 2018; Matthee et al. 2019; Schaerer et al. 2020) and the spatially-resolved SFR (the $\Sigma_{\rm [C\,II]} - \Sigma_{\rm SFR}$ relation; Pineda et al. 2014; De Looze et al. 2014; Herrera-Camus et al. 2015). Additionally, in recent years, the line strength has been used as a tracer of other galaxy-integrated quantities such as the molecular gas mass (Hughes et al. 2017; Madden et al. 2020; Zanella et al. 2018), particularly of CO-dark molecular gas (Madden et al. 2020; Accurso et al. 2017), the H I mass (Heintz et al. 2021, 2022), the total gas mass (D'Eugenio et al. 2023), as well as the metal content (Heintz et al. 2023). However, it is known from observations of [C II] emission from galactic centres and luminous infrared galaxies, with infrared (IR) luminosities $L_{\rm IR} \gtrsim 10^{11}\,{\rm L}_\odot$, that the $L_{\rm [C\,II]}/L_{\rm IR}$ ratio decreases with increasing $L_{\rm IR}$ (e.g., Malhotra et al. 2001; Graciá-Carpio et al. 2011; Díaz-Santos et al. 2013), thereby hinting at a possible breakdown of the [C II]-SFR relation at high SFRs or high SFR surface densities (see e.g., Graciá-Carpio et al. 2011).

From the theoretical point of view, the observed correlations between [C II] line strength and the galaxy properties arise naturally as [C II] is a metal cooling line and is linked to both the metal content and the heating via star formation in regions where cooling is dominated by this line such as photon-dominated regions (PDRs) and the cold neutral medium. Moreover, the various galaxy properties are themselves correlated; for example, the Kennicutt-Schmidt relation connects the gas surface density and the SFR surface density (Kennicutt 1998; Leroy et al. 2008; Bigiel et al. 2008), while the mass-metallicity relation (Tremonti et al. 2004) connects the stellar mass and gas metallicity. This implies that any correlation of the [C II] line with another galaxy property will have secondary dependencies, often manifested as a scatter, that must be quantified to provide robust calibrations.

In this regard, the [C II]-SFR correlation has garnered a lot of attention on the theoretical front. Several studies have meticulously tested this correlation and its redshift evolution using chemical and radiative transfer modelling in individual galaxies (Vallini et al. 2015; Katz et al. 2019, 2022) or for entire simulation suite at targeted redshifts (Olsen et al. 2016, 2017; Pallottini et al. 2017; Lagache et al. 2018; Lupi et al. 2018; Popping et al. 2019; Leung et al. 2020).

However, unlike the [C II]-SFR relation, the [C II] – $M_{\rm mol}$ relation has received limited attention in simulations so far (see e.g., Vizgan et al. 2022), partly because current state-of-the-

art cosmological simulations do not self-consistently follow the evolution of the molecular gas component in galaxies, and they often rely on analytical relations to be used in post-processing (e.g., Lagos et al. 2015, 2016), which might not hold at high redshifts. For instance, Vizgan et al. (2022) found a shallower $L_{[\mathrm{C\,II}]} - M_{\mathrm{mol}}$ relation at $z \sim 6$ compared to the one obtained by Zanella et al. (2018) using a compilation of $M_* \gtrsim 10^{10}\,\mathrm{M}_\odot$ galaxies at $z = 0 - 5.5$, highlighting the need for robust testing of this calibration.

Therefore, while observations of the [C II] line have opened up an interesting avenue for probing the high-z ISM, several open questions still remain: does the [C II]-SFR relation evolve with redshift? Does the [C II]-M_{mol} relation show secondary dependencies on other galaxy properties? What is faint-end slope of the [C II] luminosity function and how does it evolve with redshift? To provide a theoretical insight on these, we have performed a suite of cosmological simulations, called the Marigold suite, wherein we follow the non-equilibrium abundance of H_2, CO, C, and C^+ on the fly using the sub-grid model Hyacinth (Khatri et al. 2024). The [C II] emission from the simulated galaxies is calculated by solving the radiative-transfer problem in post-processing. In this paper, we use this simulation suite to investigate the usefulness of this line as a probe of the interstellar medium (ISM) in high-redshift ($3 \leq z \leq 7$) galaxies. In particular, we provide a calibration for inferring the molecular gas mass of a galaxy from its [C II] luminosity, accounting for secondary dependencies in this relation across redshift.

In the past few years, observations of high-redshift galaxies have detected [C II] emission extending farther than the UV continuum emission from these galaxies (Fujimoto et al. 2019, 2020; Ginolfi et al. 2020; Fudamoto et al. 2022; Lambert et al. 2023; Posses et al. 2024), hinting at an extended gas reservoir rich in ionized carbon. This extended [C II] emission is often referred to as a [C II] halo, despite the misleading term. Satellites galaxies, galactic outflows, and gas stripped due to galaxy interactions are all plausible sources of an extended [C II] halo. Reproducing extended [C II] in numerical simulations has proven challenging so far (Fujimoto et al. 2019; Muñoz-Elgueta et al. 2024), making it difficult to pinpoint its exact origin(s). This is further complicated by the faint nature of the extended emission and the limited spatial resolution of high-z observations. In this study, we further explore the existence of extended [C II] emission using our simulations.

The rest of the paper is organised as follows: In Sect. 2, we describe the simulation suite and detail the modelling of the [C II] emission in Sect. 3. In Sect. 4, we investigate the redshift evolution of the [C II] luminosity function from the simulations. We investigate the [C II]-star formation rate correlation in our simulated galaxies on global and spatially resolved scales in Sect. 5. In Sect. 6, we examine the reliability of the [C II] line as a molecular gas tracer and quantify secondary dependencies of the $L_{[\mathrm{C\,II}]} - M_{\mathrm{mol}}$ relation on the star formation rate and the gas metallicity. Then, we explore the extended [C II] emission in Sect. 7. We compare our [C II]-SFR relation and [C II] luminosity function with those from previous numerical studies in Sect. 8 and conclude with a summary of our main results in Sect. 9.

2. Simulations

We use the Marigold suite of cosmological simulations (Khatri et al., in preparation), which comprises two hydrodynamical simulations – a $(25\,\mathrm{Mpc})^3$ comoving volume (`M25`) and a $(50\,\mathrm{Mpc})^3$ comoving volume (`M50`). The simulations and the statistical properties of the simulated galaxies are described in full detail in Khatri et al. (to be sumbitted). Here we briefly summarize some key details.

The simulations adopt the Planck cosmology (Planck Collaboration VI 2020) with $\Omega_\Lambda = 0.6847$, $\Omega_\mathrm{m} = 0.3153$, $\Omega_\mathrm{b} = 0.0493$, $\sigma_8 = 0.8111$, $n_\mathrm{s} = 0.9649$, and $h = 0.6736$. The simulations are started from uni-grid initial conditions (ICs) set at $z = 99$ generated with the code Music (Hahn & Abel 2011). The ICs have an initial refinement level $l_{\mathrm{initial}} = 10$ corresponding to 1024^3 grid cells and an equal number of dark matter particles. We allow the grid to refine naturally down $z = 3$, which results in a maximal spatial resolution (i.e., minimum grid cell size Δx^{min} in physical units) achieved during the course of the simulations of $\Delta x^{\mathrm{min}} = 32$ pc for `M25` and $\Delta x^{\mathrm{min}} = 64$ pc for `M50`. The simulation volumes have periodic boundary conditions and the dynamical evolution of dark matter, gas, and stars is tracked with the adaptive mesh refinement code Ramses (Teyssier 2002) down to $z = 3$. The simulation specification are given in Table 1.

These simulations were performed using the sub-grid model Hyacinth (Khatri et al. 2024) to evolve the non-equilibrium abundances of H_2, CO, C, and C^+. Hyacinth comprises a sub-grid chemical network based on our modified version of the widely used Nelson & Langer (1999) chemical network. It accounts for the unresolved density structure of the ISM by assuming a probability distribution function (PDF) of sub-grid densities. The PDF is designed to vary with the state of star formation. A metallicity-dependent temperature-density relation based on high-resolution molecular cloud simulations (Hu et al. 2021) assigns a (sub-grid) temperature to each sub-grid density. The chemical rate equations are solved at each sub-grid density and the cell-averaged chemical abundances are obtained by integrating over the density PDF. The model is described in full detail in Khatri et al. (2024).

We use the Amiga Halo Finder Knollmann & Knebe (AHF 2009) to identify halos and subhalos in each snapshot. AHF identifies halos by locating density peaks within the simulation and then iteratively determining the gravitationally bound particles that constitute each peak. Each resulting halo is a spherical region with virial radius R_{vir} and a mean matter density (i.e., including dark matter, gas, and stars) equal to 200 times the critical density ρ_{crit}. The virial mass of the halo can be written as $M_{\mathrm{vir}} = \frac{4}{3}\pi R_{\mathrm{vir}}^3\ 200\,\rho_{\mathrm{crit}}$, where the masses and sizes of haloes are calculated accounting for unbinding. These are referred to as 'main halos' in the following. Subhalos are defined as gravitationally bound objects within main halos and lying within common isodensity contours of the host halo. We impose that every halo is resolved with at least a 100 particles, unless otherwise specified (see e.g., Sect. 4). Galaxies are defined in terms of their parent halo. For main haloes, the stellar concentration at their centre is referred to as the main galaxy. For each main galaxy, we start with a spherical region of size $0.1\ R_{\mathrm{vir}}$ and calculate the stellar half-mass radius $r_{1/2,*}$ (that is, the radius containing half of the stellar mass within $0.1R_{\mathrm{vir}}$). The galaxy is defined in terms of $2r_{1/2,*}$ and all (galaxy-integrated) quantities are measured within this radius. Conversely, the stellar concentration residing at the centre of a subhalo is called a satellite galaxy whose size is defined by the radius corresponding to the maximum of the subhalo rotation curve, $R_{V_{\mathrm{max}}}$ (Klypin et al. 2011; Prada et al. 2012). In other words, $R_{\mathrm{V_{max}}}$ sets the boundary of a satellite galaxy. To ensure that the stellar component of the galaxies is well resolved, we impose a cut of 100 stellar particles (which results in different stellar mass cuts for the two simulations). Our `M25` simulation is capable of resolving low-mass galaxies down to a stellar mass of $10^6\,\mathrm{M}_\odot$, while our `M50` run is aimed at improving the statis-

Table 1: Specifications of the MARIGOLD simulation suite. From left to right, the columns list: the name of the simulation, the comoving box size, the number of dark-matter (DM) particles, the initial and final refinement levels, the minimum cell size achieved in the simulation in physical units, the DM and stellar particle masses, and the average gas mass per grid cell in the initial conditions.

Simulation	$L_{\rm box}$ (cMpc)	$N_{\rm DM}$	$l_{\rm initial}$	$l_{\rm final}$	$\Delta x^{\rm min}$ (pc)	$m_{\rm DM}$ ($\rm M_\odot$)	m_* ($\rm M_\odot$)	$m_{\rm gas}^{\rm ini}$ ($\rm M_\odot$)
`M25`	25	1024^3	10	17	32	5.0×10^5	7.2×10^3	9.3×10^4
`M50`	50	1024^3	10	17	64	4.0×10^6	5.8×10^4	7.4×10^5

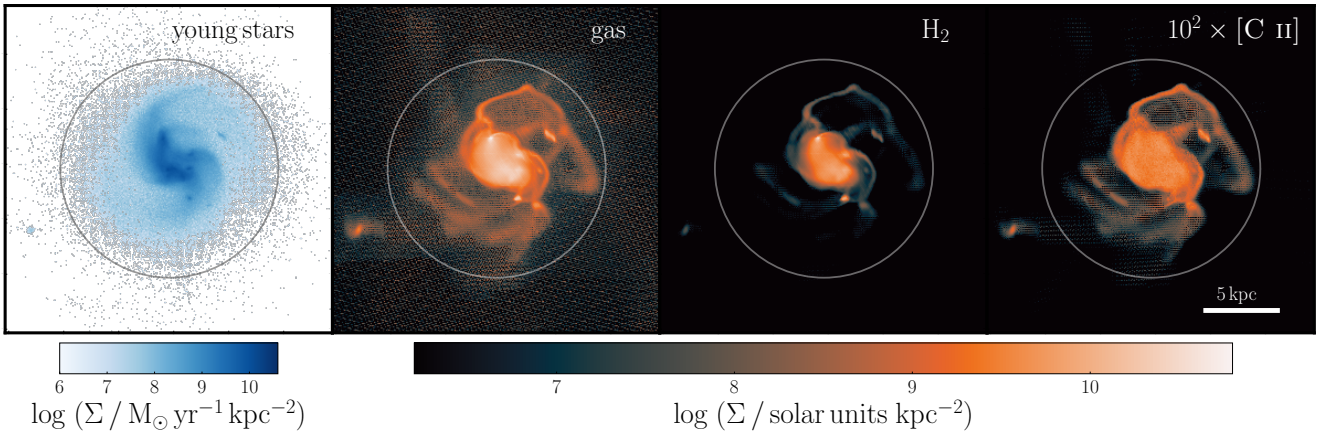


Fig. 1: Face-on view of a simulated galaxy at $z = 4$. From left to right, the columns show the surface density of young stars (with ages $\leq$ 200 Myr), total gas surface density, H_2 surface density, and [C II] surface brightness. In each panel the circle indicates 0.1 times the virial radius of the parent DM halo.

tical robustness of our results and its larger volume allows us to simulate more massive galaxies.

3. Modelling [C II] emission

The first ionization potential of atomic carbon (11.3 eV) is lower than that of hydrogen (13.6 eV). Consequently, singly-ionized carbon (C^+) exists in different ISM phases including molecular clouds, neutral atomic gas, and H II regions and emits the [C II] fine-structure line in a variety of conditions. Disentangling the contribution of the different phases to the total [C II] emission of a galaxy is therefore challenging and requires using other emission lines.

Observational studies of [C II] emission in $z \gtrsim 1$ galaxies have shown that the bulk of the [C II] emission originates from PDRs that represent a transition region between the H II region around a young massive star and a molecular cloud. For instance, Stacey et al. (2010) found that in a sample of 12 galaxies at $1 \lesssim z \lesssim 2$, more than 70% of the [C II] emission arises from PDRs. Also, Gullberg et al. (2015) found that in their sample of 20 dusty star-forming galaxies at $2.1 < z < 5.7$, the [C II] emission is consistent with arising from PDRs. This finding is further supported by theoretical studies that calculate the [C II] emission from simulated galaxies in post-processing, accounting for emission arising from different phases, e.g., Olsen et al. (2015), who employ a multi-phase ISM model for each gas particle in the simulation and use the photoionisation code CLOUDY (Ferland et al. 1992) to calculate the emission arising from each phase (also see Pallottini et al. 2017). The general consensus from these studies is that the bulk ($\gtrsim 70-90\%$) of the [C II] emission arises from neutral atomic and molecular gas. Moreover, based on a sample of low-metallicity dwarf galaxies from the *Herschel* Dwarf Galaxy Survey (Madden et al. 2013), Cormier et al. (2019) found that $\gtrsim$ 70% [C II] emission arises from PDRs. As such galaxies are expected to be similar to high-redshift galaxies in terms of their ISM structure, it is reasonable to assume that a similar fraction of the [C II] emission arises from PDRs in high-redshift galaxies as well.

We use HYACINTH to obtain the abundances of H_2, CO, C, and C^+ in our simulations. HYACINTH follows a simplified chemical network and predicts abundances that are consistent with PDR codes (Wolfire et al. 2010). This approach allows us to model the [C II] emission without relying on assumptions regarding the C^+ abundance that might not hold across the entire galaxy population at all redshifts.

The [C II] line arises from the ${}^2P_{3/2} \rightarrow {}^2P_{1/2}$ fine-structure transition of C^+, that can be excited by collisions with hydrogen molecules (H_2), hydrogen atoms (H), and electrons (e^-). The transition can also be excited by an external radiation field such as the cosmic microwave background (CMB), which becomes particularly important at higher redshifts (da Cunha et al. 2013). De-excitation can either happen spontaneously or be stimulated by the external radiation field.

When calculating the [C II] emission from a simulated galaxy, we account for optical depth effects within individual cells. For this we assume a plane-parallel configuration and divide each cell into N slices. The density in each slice is obtained from the underlying density PDF (same as in HYACINTH; see also Vallini et al. 2015). The level populations in each slice are computed accounting for the emission from all other slices. Conversely, the fraction of emission from a given slice that manages to reach the edge of the cell (where the optical depth $\tau = 0$) depends on the slice's position within the cell depends on its location within the cell. Solving this radiative transfer problem requires an iterative approach, which is detailed in Appendix A. We validate our model against CLOUDY in Appendix B.

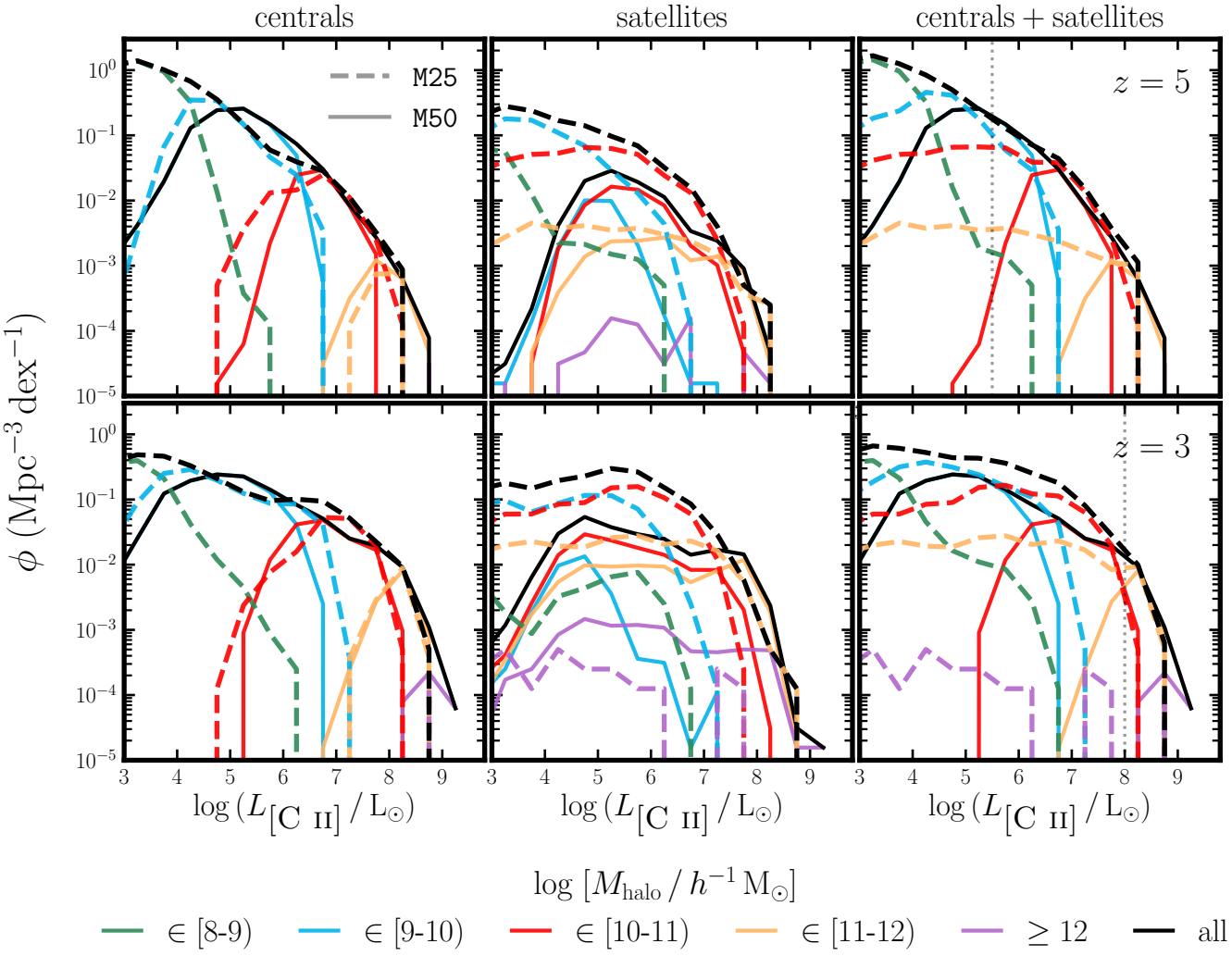


Fig. 2: Conditional [C II] LF from the M25 (dashed lines) and M50 (solid lines) simulations at redshifts $z = 5$ and 3 for central galaxies (left panels), satellites galaxies (middle panels), and all galaxies (right panels). The coloured lines show the CLF of emitters residing in DM halos in different M_{halo} bins and the black lines show the total LFs. The dotted grey line in the right panels denotes the luminosity threshold, L_{thr}, below which the total LFs from two simulations differ significantly due to resolution effects.

However, when calculating the total [C II] luminosity of a galaxy, we assume that the cells are radiatively decoupled from each other. This means that the [C II] emission that escapes the cell of origin, travels unattenuated to the edge of the galaxy. This assumption is valid whenever the velocity difference between neighbouring cells is larger than the intrinsic line width due to the gas motions within a cell: the emitted spectra is shifted out of the frequency range where it could be absorbed by another cell. This is a common approximation in the literature (see e.g. Olsen et al. 2015; Vallini et al. 2015). The total [C II] luminosity of the galaxy is then calculated as the sum of the luminosities of each cell within the galaxy.

Fig. 1 shows the different baryonic components and the [C II] surface brightness of a simulated galaxy at $z = 4$.

4. [C II] luminosity function

In this section, we examine the [C II] luminosity function (LF) from the MARIGOLD simulations and investigate how it evolves with redshift.

4.1. Conditional luminosity function and resolution effects

The difficulty we have to face is to combine simulations with different spatial and mass resolutions in a consistent way while accounting for sample variance given the relatively small computational volumes. For these reasons, we first perform a consistency check between the outputs of `M25` and `M50` by measuring the conditional luminosity function (CLF, Yang et al. 2003), i.e. the luminosity function of emitters hosted by DM halos within a narrow mass bin. We only consider (main) halos containing more than 300 DM particles and examine three different cases: (i) central galaxies only, (ii) satellite galaxies only, and (iii) centrals+satellites. Our results for $z = 5$ and 3 are shown in Fig. 2. These cases are representative of what happens in the redshift ranges $5 \leq z \leq 7$ and $3 \leq z < 5$, respectively.

For all halo masses, the CLFs of the central galaxies in the two simulations are in excellent agreement. The CLF has a characteristic bell shape and its width grows with decreasing halo mass. Conversely, satellites have a much broader CLF and `M25` presents many more faint satellites than `M50` at fixed halo mass. This dis-

crepancy reflects the different mass and spatial resolutions in the simulations that regulate the abundance of DM satellites and the [C II] emission from their gas, respectively. Inspection of the total LF reveals that the two simulations show small differences (which are compatible with sample variance) above a threshold luminosity $L_{\rm thr}$ and substantial systematic differences for $L_{\rm [C\,II]} < L_{\rm thr}$. We find that $L_{\rm thr} \simeq 10^{5.5}$ L$_\odot$ for $5 \leq z \leq 7$ and $L_{\rm thr} \simeq 10^{8}$ L$_\odot$ for $3 \leq z < 5$ (dotted grey lines in the right panels of Fig. 2). This confirms that, as we stated before, our high-resolution `M25` simulation is excellent for probing the faint end of the LF while the `M50` simulation is ideally suited for probing the bright end because of its larger volume.

Finally, we note that at $L_{\rm [C\,II]} \geq 10^5$ L$_\odot$, the total luminosity function is fully represented by halos with masses $M_{\rm halo} \geq 10^9\,h^{-1}$ M$_\odot$. Therefore, in the following, we restrict our analysis to these ranges of luminosities and masses.

4.2. Bayesian curve fitting

At each redshift, we fit the simulated LF, $\phi(L_{\rm [C\,II]}) = {\rm d}n/{\rm d}\log L_{\rm [C\,II]}$[1], with different functional forms. Based on the discussion above, we consider all galaxies with $L_{\rm [C\,II]} \geq 10^5$ L$_\odot$ hosted by halos with $M_{\rm halo} \geq 10^9\,h^{-1}$ M$_\odot$ from `M25` but only those with $L_{\rm [C\,II]} > L_{\rm thr}$ from `M50`. For the functional forms, we consider a Schechter function,

$$\phi(L_{\rm [C\,II]}) = \ln(10)\ \phi_* \left(\frac{L_{\rm [C\,II]}}{L_*}\right)^{\alpha+1} \exp\left(-\frac{L_{\rm [C\,II]}}{L_*}\right), \quad (1)$$

and a double power-law (DPL),

$$\phi(L_{\rm [C\,II]}) = \ln(10)\ \phi_* \left[\left(\frac{L_{\rm [C\,II]}}{L_*}\right)^{-(\alpha+1)} + \left(\frac{L_{\rm [C\,II]}}{L_*}\right)^{-(\beta+1)}\right]^{-1}. \quad (2)$$

where all parameters have the usual meaning and $\beta < \alpha$.

We follow a Bayesian approach and sample the parameter space using a Markov-chain Monte-Carlo (MCMC) method implemented with the python package `emcee` (Foreman-Mackey et al. 2013). We assume that the counts N_i in each logarithmic bin of luminosity follow a Poisson distribution and write the log-likelihood function for each simulation as

$$\ln \mathcal{L} = \sum_{i=1}^{N_{\rm bins}} N_i\ \ln(N_{{\rm model},i}) - N_{{\rm model},i} + \text{constant}. \quad (3)$$

Eventually, we sum the results obtained for `M25` and `M50`.

To account for the sample variance of the simulated volumes, we follow an approach that builds upon the method originally proposed by Trenti & Stiavelli (2008) to obtain the LF from the combination of observational data with different depths. Briefly, we fit the LF data from the two simulations with exactly the same shape but (slightly) different normalisation factors that can be written as $\log\phi_{*,j} = \log\phi_* + \Delta_j$, where ϕ_* represents the 'cosmic' normalization and Δ_j is the correction due to sample variance in the $j^{\rm th}$ simulation. We impose independent Gaussian priors on each Δ_j, i.e., $\Delta_j \sim \mathcal{N}(0, (\sigma_{{\rm v},j}/\ln 10)^2)$, where $\sigma^2_{{\rm v},j}$ is the sample variance of the overdensity within the respective simulation volume. The latter is estimated from the calculations presented in Somerville et al. (2004), considering the halo mass bin that gives the dominant contribution to the counts of emitters around L_*. We adopt (broad) flat priors on all other parameters.

As an example, we show in Appendix C, the resulting posterior distribution for the DPL fit at $z = 3$. In what follows, we present results obtained after marginalising the posterior distributions over the parameters, Δ_j.

Since computing the model evidence from the Markov chains is problematic, for simplicity, we use the deviance information criterion (DIC, Spiegelhalter et al. 2002) to identify whether the Schechter function or the DPL provide a better fit to the simulated data. Deviance is a measure of goodness of fit defined as $D = -2\ln\mathcal{L}(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ indicates the parameters of a model. The DIC is obtained correcting the deviance with a penalty, $p_{\rm D}$, for the complexity of the model which is quantified in terms of the effective number of fit parameters. There are two common approaches to estimate $p_{\rm D}$ from a Markov chain: $p_{\rm D}^{(a)} = \overline{D(\boldsymbol{\theta})} - D(\overline{\boldsymbol{\theta}})$ and $p_{\rm D}^{(b)} = {\rm Var}(D)/2$, where an overbar denotes the average over the posterior distribution (i.e. over the sampled chain) and the symbol Var denotes the corresponding variance. The DIC is then defined as DIC$= D(\overline{\boldsymbol{\theta}}) + 2p_{\rm D}$. Models with smaller DIC are better supported by the data. Typically, differences (ΔDIC) above 5 are considered substantial and the model with the higher DIC is ruled out if the difference grows above 10. Based on this test, we find that the LF from our simulations is better represented by a DPL. The corresponding parameters are listed in Table 2 along with the ΔDIC values computed with both estimators for $p_{\rm D}$. We provide in Appendix C a comparison of the two functional forms with the simulation data.

Fig. 3 shows the resulting [C II] luminosity function obtained using a DPL at different redshifts. The shaded band indicates the central 68% credibility region obtained with the MCMC method. The solid line represents the best fit (evaluated at the posterior mean $\overline{\boldsymbol{\theta}}$). We see a clear redshift evolution in the LF. The turnover luminosity (L_*) increases (almost) monotonically with time and the faint-end slope (α) tends to flatten at late times. In contrast, the bright-end slope (β) does not show a clear evolutionary trend. The large jump in the LF at $L_{\rm [C\,II]} \lesssim 10^8$ L$_\odot$ is partially due to the spatial refinement that happens shortly after $z = 5$ in both simulations, that allows them, especially `M25`, to resolve more emitters. We also see a significant evolution in the number density of bright emitters. The number density of galaxies at $L_{\rm [C\,II]} \sim 10^9$ L$_\odot$ increases from $\sim 10^{-6}$ dex^{-1} Mpc^{-3} at $z = 7$ to 6×10^{-4} dex^{-1} Mpc^{-3} at $z = 3$, indicating a 600-fold increase in the number of $L_{\rm [C\,II]} \sim 10^9$ L$_\odot$ galaxies in a less than 1.5 Gyr timespan.

Observational constraints on the [C II] luminosity function at these redshifts come from Yan et al. (2020) for targeted detections at $4.5 \lesssim z \lesssim 5.9$ in the ALMA Large Program to Investigate C$^+$ at Early Times[2] (ALPINE Le Fèvre et al. 2020) and a lower limit at $z \sim 4.4$ reported by Swinbank et al. (2012) based on [C II] detections in two observed galaxies. We find that our LFs at $z = 5$ and 4 are in excellent agreement with both estimates. We compare against previous numerical work in Sect. 8.

4.3. [C II] luminosity density

Measuring the cosmic star formation rate density (SFRD) at different epochs has been the subject of several studies (see Madau & Dickinson 2014, for a complete review). Similarly, estimating the cosmic molecular gas density from blind and targeted surveys of molecular gas tracers such as CO and dust continuum

[1] Note that here and throughout the text, we use 'log' to denote $\log_{10}$; for the natural logarithm $\log_{\rm e}$, we use 'ln' instead.

[2] http://alpine.ipac.caltech.edu/

Table 2: Best-fit parameters to the LF for the DPL function given in Eq. (2) . The last two columns show the values of $\Delta\mathrm{DIC}^a$ and $\Delta\mathrm{DIC}^b$ between the Schechter function and the DPL (see text).

z	$\log(\phi_* / \mathrm{Mpc^{-3}\,dex^{-1}})$	$\log(L_* / \mathrm{L_\odot})$	α	β	$\Delta\mathrm{DIC}^{(a)}$	$\Delta\mathrm{DIC}^{(b)}$
7	$-1.84^{+0.15}_{-0.15}$	$6.44^{+0.10}_{-0.11}$	$-1.54^{+0.06}_{-0.05}$	$-2.75^{+0.08}_{-0.09}$	105	107
6	$-1.26^{+0.12}_{-0.12}$	$6.20^{+0.10}_{-0.10}$	$-1.24^{+0.06}_{-0.06}$	$-2.30^{+0.04}_{-0.05}$	222	223
5	$-1.52^{+0.13}_{-0.13}$	$6.72^{+0.11}_{-0.11}$	$-1.42^{+0.03}_{-0.03}$	$-2.31^{+0.05}_{-0.05}$	159	161
4	$-0.91^{+0.08}_{-0.07}$	$7.03^{+0.04}_{-0.04}$	$-1.28^{+0.01}_{-0.01}$	$-2.57^{+0.05}_{-0.05}$	449	451
3	$-1.00^{+0.08}_{-0.08}$	$7.37^{+0.04}_{-0.04}$	$-1.22^{+0.01}_{-0.01}$	$-2.65^{+0.05}_{-0.05}$	174	176

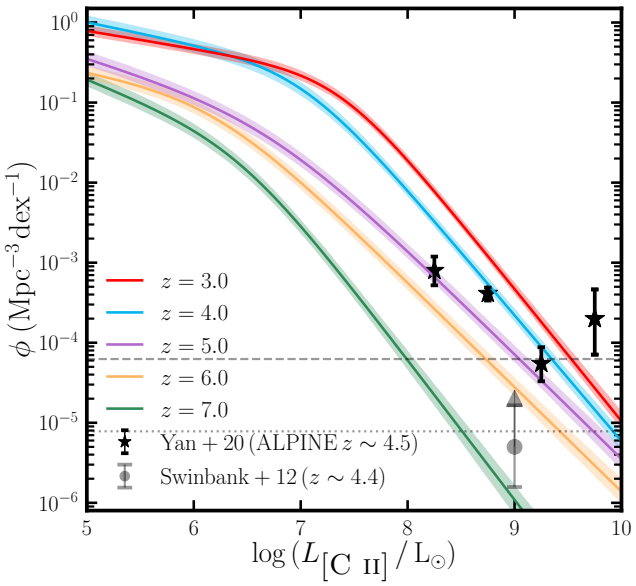


Fig. 3: The simulated [C II] LF compared with observational estimates. The coloured lines represent the best-fit DPL – Eq. (2) – to the simulated LF and the shaded area represents the central 68% credibility range obtained using the MCMC chains. Black stars represent the observational estimates at $z \sim 4.5$ from the ALPINE survey (Yan et al. 2020) and the grey arrow shows the lower limit from Swinbank et al. (2012) based on observations of two galaxies at $z \sim 4.4$. The dashed and dotted horizontal lines represent a number count of 1 per dex in the entire simulation volume of `M25` and `M50`, respectively.

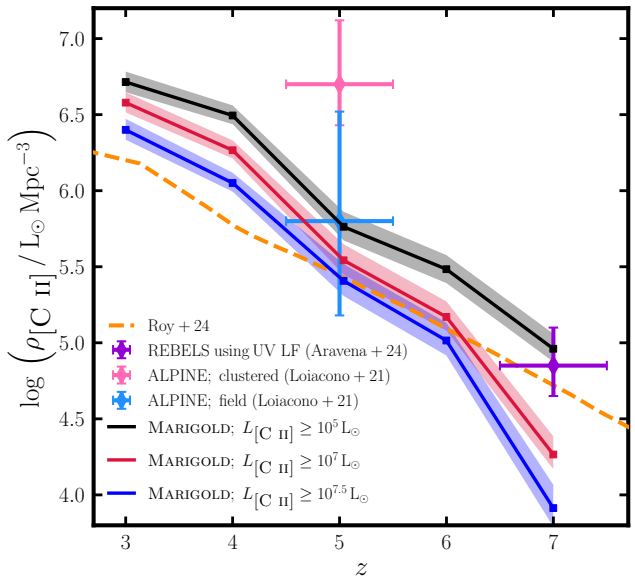


Fig. 4: Comparison of the cosmic [C II] luminosity density ($\rho_{[C\,II]}$) for different luminosity cuts in the Marigold simulations with observational estimates from ALPINE (Loiacono et al. 2021) – clustered and field estimates in pink and blue, respectively; from REBELS (Aravena et al. 2024) in purple and from a semi-empirical model by Roy & Lapi (2024) shown as an orange dashed line. The $\rho_{[C\,II]}$ from the simulations is obtained by integrating the LFs shown in Fig. 3 down to $\log(L_{\rm min} / \mathrm{L_\odot}) = 5$ (black), 7 (red), and 7.5 (blue). The shaded regions represent the 16-84 percentile range of $\rho_{[C\,II]}$ obtained from the LFs.

(see e.g. Walter et al. 2020; Riechers et al. 2019; Scoville et al. 2017; Magnelli et al. 2020, among others) has also gained significant interest over the last decade. Owing to the correlation between the [C II] luminosity with both the SFR and molecular gas mass across redshift, estimates of the cosmic [C II] luminosity density ($\rho_{[C\,II]}$) can be used to infer the cosmic SFRD and the cosmic molecular gas density (see e.g., Yan et al. 2020; Loiacono et al. 2021 for the ALPINE survey and Aravena et al. 2024 for the Reionization Era Bright Emission Line Survey[3], REBELS, Bouwens et al. 2022).

In Fig. 4, we show the redshift evolution of $\rho_{[C\,II]}$ predicted by our simulations and compare with observational estimates at these redshifts. For this we integrate the LF in Fig. 3 down to $\log(L_{\rm min} / \mathrm{L_\odot}) = 5$. This is shown by a black line in Fig. 4 and referred to as our fiducial estimate in the following. For a fair comparison with observational estimates of $\rho_{[C\,II]}$ from the ALPINE (Loiacono et al. 2021) and REBELS (Aravena et al. 2024) surveys, we show using red and blue lines, respectively, the $\rho_{[C\,II]}(z)$ for $\log(L_{\rm min} / \mathrm{L_\odot}) = 7$ and $\log(L_{\rm min} / \mathrm{L_\odot}) = 7.5$, which correspond to the luminosity cuts adopted in the two surveys for integrating the luminosity function. Note that Loiacono et al. (2021) provide two estimates for $\rho_{[C\,II]}$ based on serendipitously detected galaxies in ALPINE at $z \sim 5$; one of these is obtained by integrating the [C II] LF based on a sample of galaxies that seem to be part of a local overdensity and are therefore not representative of the galaxy population at the targeted redshift (this is referred to as the 'clustered estimate'). The $\rho_{[C\,II]}$ for the field population (the 'field estimate') can be estimated by only considering emitters detected outside the aforementioned overdensity. Since there is only such emitter detected in their sample, Loiacono et al. (2021) obtain the field estimate by multiplying the clustered estimate by the ratio between the number of emitters in the field and clustered subsamples (=1/11), on account of the similar volumes estimated for the two subsamples. We find that the $\rho_{[C\,II]}$ predicted by our

[3] https://rebelsalma.wordpress.com/

simulations at $z = 5$ shows an excellent agreement with that from ALPINE for their field sample, irrespective of the luminosity cut. Conversely, as expected, the $\rho^{\rm clustered}_{\rm [C\,II]}$ from ALPINE in almost an order of magnitude higher than our predicted $\rho_{\rm [C\,II]}$. Our $\rho_{\rm [C\,II]}$ at $z = 7$ is lower than the REBELS one when considering their luminosity cut of $10^{7.5}$ $L_\odot$ (blue line), but our $\rho_{\rm [C\,II]}$ with lower luminosity cuts (red and black lines) bracket the REBELS estimate. Overall, our estimates show a good agreement with observations. At $z \lesssim 4$, the impact of a luminosity cut is not as severe and alters the predicted $\rho_{\rm [C\,II]}$ by at most a factor of 2.

We also include in Fig. 4, the $\rho_{\rm [C\,II]}(z)$ estimate from Roy & Lapi (2024), based on a semi-empirical model that performs an abundance matching of the theoretical halo mass function and the observed stellar mass function, and assigns a [C II] luminosity to every halo based the empirical SFR-stellar-mass relation and the [C II]-SFR relation from Vallini et al. (2015). The resulting $\rho_{\rm [C\,II]}(z)$ from their approach is similar in shape to our fiducial estimate but consistently lower by a factor of ~ 2 in the redshift range shown here.

5. The $L_{\rm [C\,II]}$ – SFR relation

We now turn our attention to investigating how the [C II] luminosity correlates with the SFR (this section) and the molecular gas mass (Sect. 6) using a statistical sample of simulated galaxies. For this, we only consider central galaxies as the [C II] LFs of the centrals from the two simulations show an excellent agreement across redshift (see the left panels of Fig. 2).

The luminosity of the [C II] line correlates strongly with the SFR in normal star-forming galaxies (Stacey et al. 2010; De Looze et al. 2014). It is one of the brightest emission lines in high-z galaxies and is unaffected by dust obscuration. Thus, it can provide an estimate of the total (unobscured+obscured) SFR in distant galaxies. Moreover, if the [C II]-SFR calibration does not evolve significantly with redshift, then it can be robustly used across cosmic time. In this section, we aim to investigate the correlation between [C II] luminosity ($L_{\rm [C\,II]}$) and the SFR in the Marigold galaxies on both galaxy-wide (Sects. 5.1- 5.2) and spatially-resolved (Sect. 5.3) scales.

5.1. Galaxy-integrated [C II]-SFR relation

Fig. 5 shows our simulated galaxies in the $L_{\rm [C\,II]}$ – SFR plane for different redshifts. We calculate the SFR averaged over the last 200 Myr to be consistent with observations that commonly use a combination of SF tracers to estimate the total obscured + unobscured SF (see e.g. Kennicutt & Evans 2012; De Looze et al. 2014; Herrera-Camus et al. 2015; Lomaeva et al. 2022). For comparison, we show the best-fit relation from (De Looze et al. 2014) for their high-redshift sample ($0.5 \leq z \leq 6.6$, orange solid line). We further compare with [C II]-detected galaxies at $4.4 \lesssim z \lesssim 5.9$ from the ALPINE survey (Béthermin et al. 2020). The best-fit $L_{\rm [C\,II]}$–SFR for these galaxies (Schaerer et al. 2020) is shown using a blue solid line. We also include a literature sample of [C II]-detected galaxies compiled by Olsen et al. (2015) (pink symbols) and the galaxy REBELS-25 (black diamond Rowland et al. 2024) observed as part of REBELS. Many of our galaxies occupy the same region in the [C II]-SFR plane as the observed galaxies (shown as blue, pink, and black scatter points).

At each redshift, we fit a relation of the form $\log(L_{\rm [C\,II]}/L_\odot) = a \log({\rm SFR}_{200}/M_\odot\,{\rm yr}^{-1}) + b$ (where $\{a, b\} \in \mathbb{R}^2$) to our simulated galaxies using an ordinary least squares linear regression (shown as a solid red line in Fig. 5). We report in Table 3 the resulting parameters along with the 1σ dispersion, i.e., the standard deviation of the residuals around the best fit in each case.

Firstly, from the distribution of simulated galaxies at different redshifts in the [C II]-SFR plane, we immediately see an increase in $L_{\rm [C\,II]}$ at a given SFR from $z = 7$ to $z = 3$, which is likely fueled by the buildup of the metal content of galaxies over time. We quantify this redshift evolution in Sect. 5.2. At $z = 3$, our galaxies are well-distributed around the De Looze et al. (2014) relation and our best fit is in excellent agreement with theirs, while at $z = 4$, we obtain a similar relation as Schaerer et al. (2020) (considering conservative upper limits on [C II] non-detections to the ALPINE galaxy sample). The value of the Spearman's rank correlation coefficient, indicated in each panel in Fig. 5, increases with time and remains high ($\gtrsim 0.86$) at all redshifts. We further find that the scatter $L_{\rm [C\,II]}$ – SFR relation increases with increasing redshift, similar to previous findings by Carniani et al. (2018) for a sample of ~ 50 galaxies at $5 \lesssim z \lesssim 7$. In contrast, based on a semi-analytical galaxy formation model coupled to the spectral synthesis code Cloudy (Ferland et al. 1992), Lagache et al. (2018) reported a slight decrease in the scatter of the $L_{\rm [C\,II]}$ – SFR relation with redshift.

The [C II]-SFR connection on galaxy-wide scales has been investigated in previous numerical studies. We compare with these in Sect. 8.

5.2. Redshift evolution of the [C II] – SFR relation

An important question concerning the [C II]-SFR relation is whether it shows any evolution with redshift. For instance based on the ALPINE survey, Schaerer et al. (2020) found that star-forming galaxies at $4 \lesssim z \lesssim 6$ follow the same or a slightly steeper [C II]-SFR relation compared to local galaxies (De Looze et al. 2014, shown as an orange line in Fig. 5), depending on the treatment of non-detections in their galaxy sample. For the Marigold galaxies, the slope of the $L_{\rm [C\,II]}$ – SFR relation shows little redshift evolution in the range $3 \leq z \leq 7$ ($\lesssim 0.15$ dex variation). As noted before, the best fit for our lowest redshift galaxy sample ($z = 3$) is in excellent agreement with the De Looze et al. (2014), but deviations towards lower $L_{\rm [C\,II]}$ are evident at $z \gtrsim 5$. Consequently, at a given SFR, $L_{\rm [C\,II]}$ decreases with increasing redshift. This trend can be attributed to the lower metallicity of galaxies at higher redshifts. For reference, the median metallicity of our galaxy sample, expressed as $12 + \log({\rm O/H})$ increases from 7.4 ± 0.2 at $z = 7$ to 7.9 ± 0.3 at $z = 3$ (the errors on the median metallicities denote the interquartile range). Similarly, for galaxies with $\log({\rm SFR}\,/\,M_\odot\,{\rm yr}^{-1}) = 1.0 \pm 0.25$, the median metallicity increases by ≈ 0.4 dex in this redshift range. The same is also reflected in the monotonically increasing values of the intercept b from $z = 7$ to $z = 3$ (see Table 3), which increases by a factor of ~ 3 (0.5 dex) in this redshift range. To conclude, we find a redshift evolution in the intercept of the [C II]-SFR relation, as evident from the offset between our best-fit relation and the De Looze et al. (2014) relation that is the same in all panels.

5.3. Spatially resolved [C II] – SFR relation

Resolved observations of [C II] and SF in high-redshift galaxies have found that these galaxies exhibit a "[C II] deficit" with respect to the local $\Sigma_{\rm [C\,II]} - \Sigma_{\rm SFR}$ relation. For example, Carniani et al. (2018) found that the galaxy-integrated [C II]-SFR relation at $5 \lesssim z \lesssim 7$ is similar to the local one but in the $\Sigma_{\rm [C\,II]} - \Sigma_{\rm SFR}$ plane, the galaxies have a substantially lower $\Sigma_{\rm [C\,II]}$ compared to the local galaxies at any given $\Sigma_{\rm SFR}$. They attributed the offset

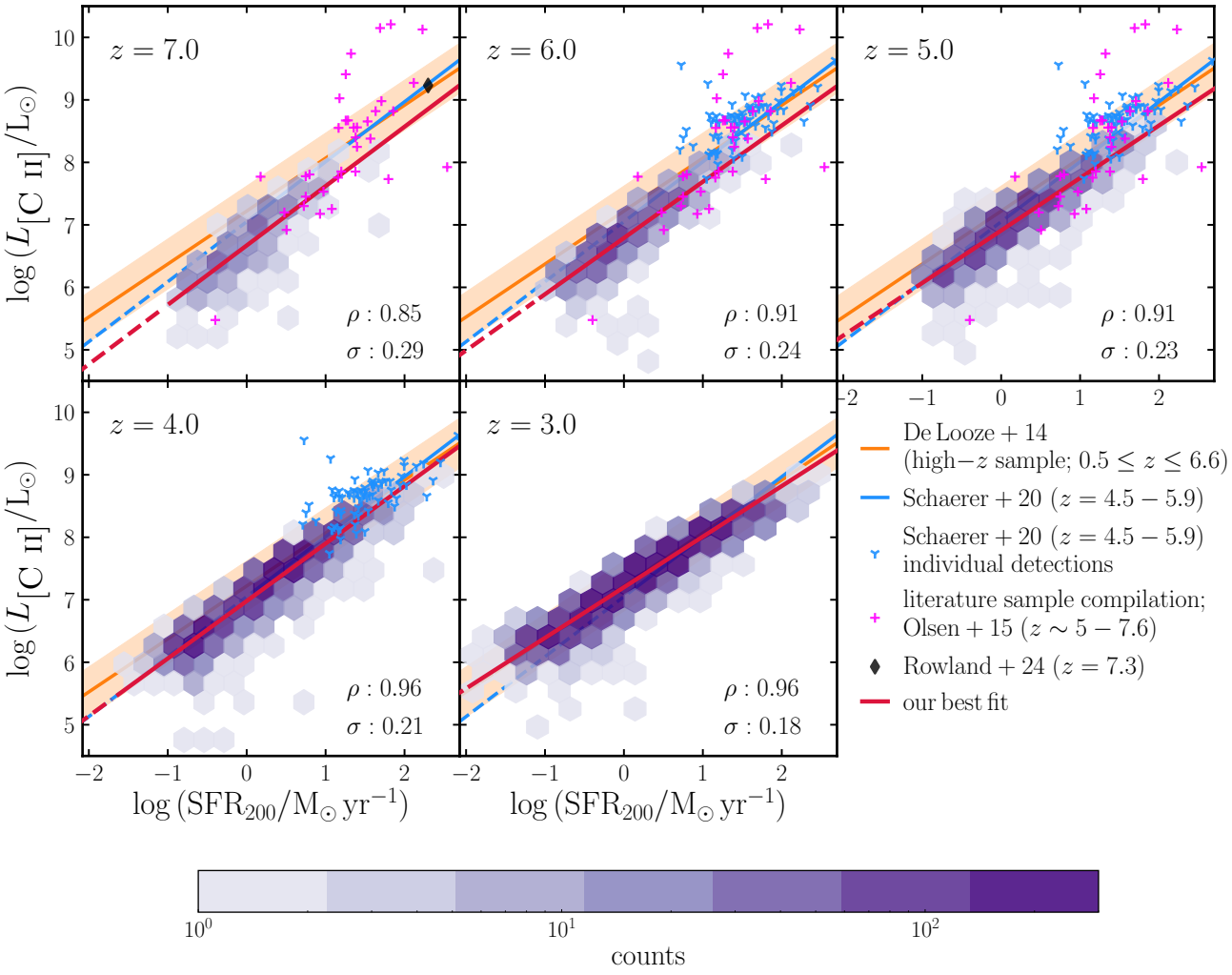


Fig. 5: The [C II] − SFR relation from the MARIGOLD simulations at $3 \leq z \leq 7$ compared with observations. The simulated galaxy population is represented as purple hexbins, with the colour indicating the galaxy counts per bin. The red line showing the best-fit to these galaxies (see Table 3 for the fit parameters). In each panel, we report the Spearman's rank correlation coefficient (ρ) and the 1σ scatter around the best-fit relation. The best-fit relation from De Looze et al. (2014) for their high−z ($0.5 < z < 6.6$) sample is shown in orange, with the shaded area representing the 1σ scatter. The blue line indicates the best-fit relation for $4.5 \lesssim z \lesssim 5.9$ galaxies from the ALPINE survey (Schaerer et al. 2020). The best-fits mentioned above are extrapolated beyond the range constrained by the respective studies using a dashed line of the same colour. The individual ALPINE galaxies (at $4.5 \lesssim z \lesssim 5.9$), the literature sample (at $5 \lesssim z \lesssim 7.6$) taken from Olsen et al. (2015), and REBELS-25 from the REBELS survey Rowland et al. (2024) are shown with blue, pink, and black symbols, respectively.

mainly to the lower metallicity of these galaxies and stressed that the more extended [C II] emission in high-z galaxies could also play a role.

Here we examining the (spatially-resolved) $\Sigma_{[\mathrm{C\,II}]} - \Sigma_{\mathrm{SFR}}$ relation in our simulated galaxies. For simplification, we present results based on $z = 4$ galaxies, although, qualitatively similar results are obtained for other redshifts as well. In each galaxy, the [C II] surface brightness and SFR surface density are obtained from a face-on projection of a cube centred on the galaxy and of side length equal to twice the size of the galaxy. The spatial resolution (minimum grid cell size; see Table 1) of our simulations is better than that achieved in current high-z [C II] observations that can resolve kpc-scale regions within $z \gtrsim 4$ galaxies (e.g., Posses et al. 2024). Therefore, we apply a 2D Gaussian smoothing to our simulated surface brightness and surface density maps to mimic observations. For this analysis, we adopt the beam sizes (in terms of the full-width at half-maximum, FWHM) for the SFR surface density and the [C II] surface brightness measurements from Posses et al. (2024), which are 0.2″ and 0.17″, respectively. At $z = 4$, these correspond to ∼ 1.4 kpc and ∼ 1.2 kpc.

In the left panel of Fig. 6, we show the $\Sigma_{[\mathrm{C\,II}]} - \Sigma_{\mathrm{SFR}}$ relation for our simulated galaxies. We split our simulated galaxies into three groups according to their stellar mass (M_*), with the number of galaxies in each group indicated in the legend. For each group, solid lines show the median $\Sigma_{[\mathrm{C\,II}]}$ as a function of Σ_{SFR} and the shaded area represents the interquartile range. The median $\Sigma_{[\mathrm{C\,II}]}$ shows slight variations among the different M_* bins. Most notably, towards higher Σ_{SFR}, lower M_* galaxies exhibit a lower $\Sigma_{[\mathrm{C\,II}]}$, likely because of their (relatively) lower metallicity. The median $\Sigma_{[\mathrm{C\,II}]}$ in all M_* bins is lower than the local relation and

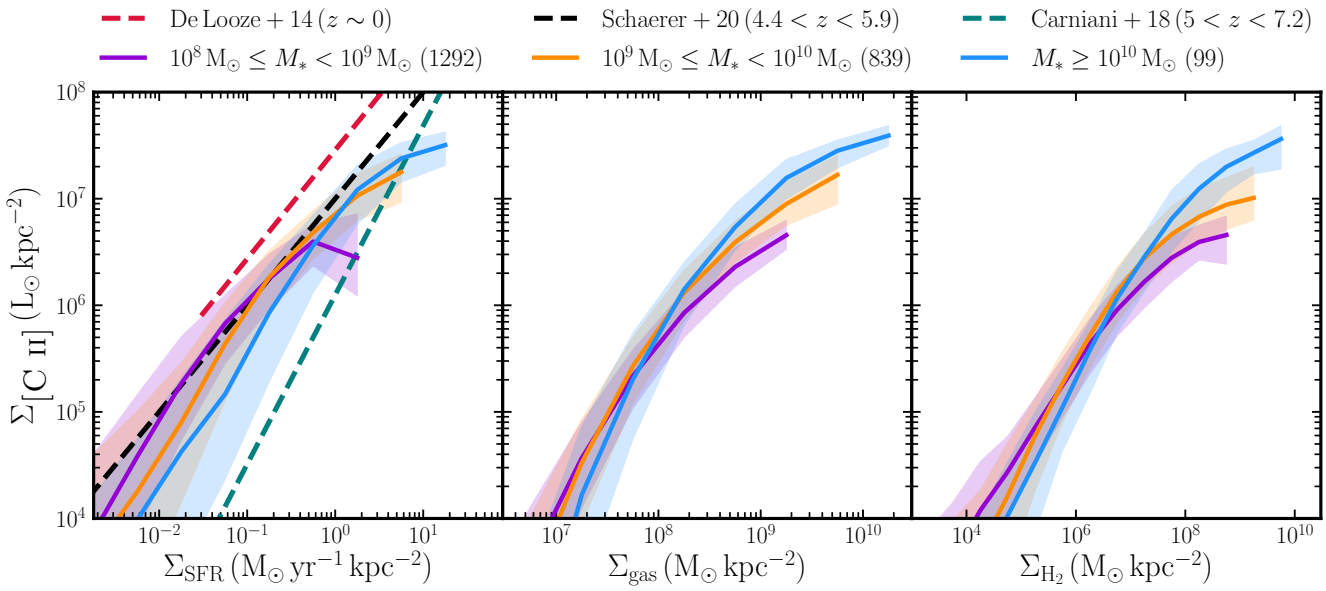


Fig. 6: The spatially-resolved [C II] − SFR relation for the MARIGOLD galaxies at $z = 4$. The galaxies are divided into different stellar-mass bins (the number of galaxies in each bin is indicated in the legend). The solid lines show the median $\Sigma_{[C\,II]}$ as a function of the SFR surface density (*left*), of the the gas surface density (*middle*), and of the H_2 surface density (*right*). The shaded areas represent the 16-84 percentile range. In the left panel, dashed lines indicate the empirical relations from De Looze et al. (2014, red) based on local dwarf galaxies from the *Herschel* Dwarf Galaxy Survey, for ALPINE galaxies in black (based on global values only Schaerer et al. 2020, black), and for a sample of galaxies at $5 < z < 7.2$ from Carniani et al. (2018, teal). For the simulated galaxies, the [C II] surface brightness, and the SFR, gas, and H_2 surface densities are obtained from a face-on projection of a cube centred on the galaxy and of side length equal to twice the radius of the galaxy.

similar to the relation from Schaerer et al. (2020, same as in Figure 12 of Posses et al. 2024). At $\Sigma_{SFR} \lesssim 1\,M_\odot\,yr^{-1}\,kpc^{-2}$, our medians for all M_* bins are consistent with the Schaerer et al. (2020) relation, but at higher Σ_{SFR}, they start to deviate towards lower $\Sigma_{[C\,II]}$ values and the shaded area overlaps with the Carniani et al. (2018). This trend continues towards higher Σ_{SFR}, and for $\Sigma_{SFR} \gtrsim 10\,M_\odot\,yr^{-1}\,kpc^{-2}$, the median $\Sigma_{[C\,II]}$ from our simulations is lower than the Schaerer et al. (2020) and Carniani et al. (2018) relations by a factor of ≈ 2.5.

For the same galaxies, we also show the $\Sigma_{[C\,II]} - \Sigma_{gas}$ and $\Sigma_{[C\,II]} - \Sigma_{H_2}$ in Fig. 6. The gas surface densities are smoothed with a Gaussian beam of FWHM of $\sim$ 1.0 kpc (same as for $\Sigma_{[C\,II]}$). Once again, there are no strong variations among the different M_* bins except at the highest gas surface densities. In the $\Sigma_{[C\,II]} - \Sigma_{H_2}$ plane, a similar trend is observed although with larger differences, similar to those in the $\Sigma_{[C\,II]} - \Sigma_{SFR}$ plane. We also observe that in all panels, the slope decreases towards higher surface densities.

5.3.1. Possible [C II]-deficit?

At $\Sigma_{SFR} \gtrsim 1\,M_\odot\,yr^{-1}\,kpc^{-2}$, our $\Sigma_{[C\,II]} - \Sigma_{SFR}$ relation deviates from the Schaerer et al. (2020) relation towards lower $\Sigma_{[C\,II]}$ and exhibit values similar to the Carniani et al. (2018) relation, albeit with a shallower slope. As regions of high SF are expected to be rich in dust and thereby bright in FIR emission, this trend is similar to the observed '[C II]-deficit' i.e., the decrease in the ratio of the [C II] to the FIR luminosity in luminous infrared galaxies.

We investigate the underlying cause of this deficit by examining the surface density of C^+, and CO as a function of Σ_{SFR}, Σ_{gas}, and the surface density of metals (Σ_{metals}) in our galaxies in Appendix D. We see that Σ_{C^+} plateaus towards high Σ_{SFR} and Σ_{gas}, while Σ_{CO} continues to grow. Firstly, a Σ_{gas} leads to a higher H_2 abundance. Second, a higher Σ_{SFR} is associated with a higher dust abundance; together these provide better shielding of CO against the dissociating UV radiation in dense environments. Therefore, in regions with a high SFR surface density, the bulk of the carbon is locked up in CO, leading to a flattening of Σ_{C^+} towards high Σ_{SFR}, and consequently a slower increase of the [C II] surface brightness towards high SFR and gas surface densities. Further note that our simulations do not account for the UV radiation from active galactic nuclei (AGN), that can contribute to the first ionization of carbon (and subsequently to [C II] emission). While some studies suggest that AGN are not a significant source of [C II] emission (see e.g., De Breuck et al. 2022), their contribution has not yet been estimated using a large sample of galaxies at high redshifts. However, the harder ionizing radiation from AGN can also ionize C^+, resulting in a decrease in the [C II] emission (Langer & Pineda 2015). In this case, the [C II] luminosity could be even lower than predicted by our model.

A similar decline in [C II] was also reported by De Looze et al. (2014) (see their Figure 2) who found that the slope of the $\Sigma_{SFR} - \Sigma_{[C\,II]}$ relation steepens for $\Sigma_{[C\,II]} \gtrsim$ few times $10^6\,L_\odot\,kpc^{-2}$ indicating that the [C II] line is not the dominant coolant in intensely star-forming regions. Such a deficit would manifest as a shallower slope of the $\Sigma_{[C\,II]} - \Sigma_{SFR}$ relation towards higher $\Sigma_{[C\,II]}$ (and higher Σ_{SFR}).

Overall we find that for our simulated galaxies at $z = 4$, the median $\Sigma_{[C\,II]}$ in a given Σ_{SFR} bin shows an excellent agreement with the best-fit to ALPINE galaxies at $4.4 \leq z \leq 5.9$ for $\Sigma_{SFR} \lesssim 1\,M_\odot\,yr^{-1}\,kpc^{-2}$. Towards higher Σ_{SFR}, however, the

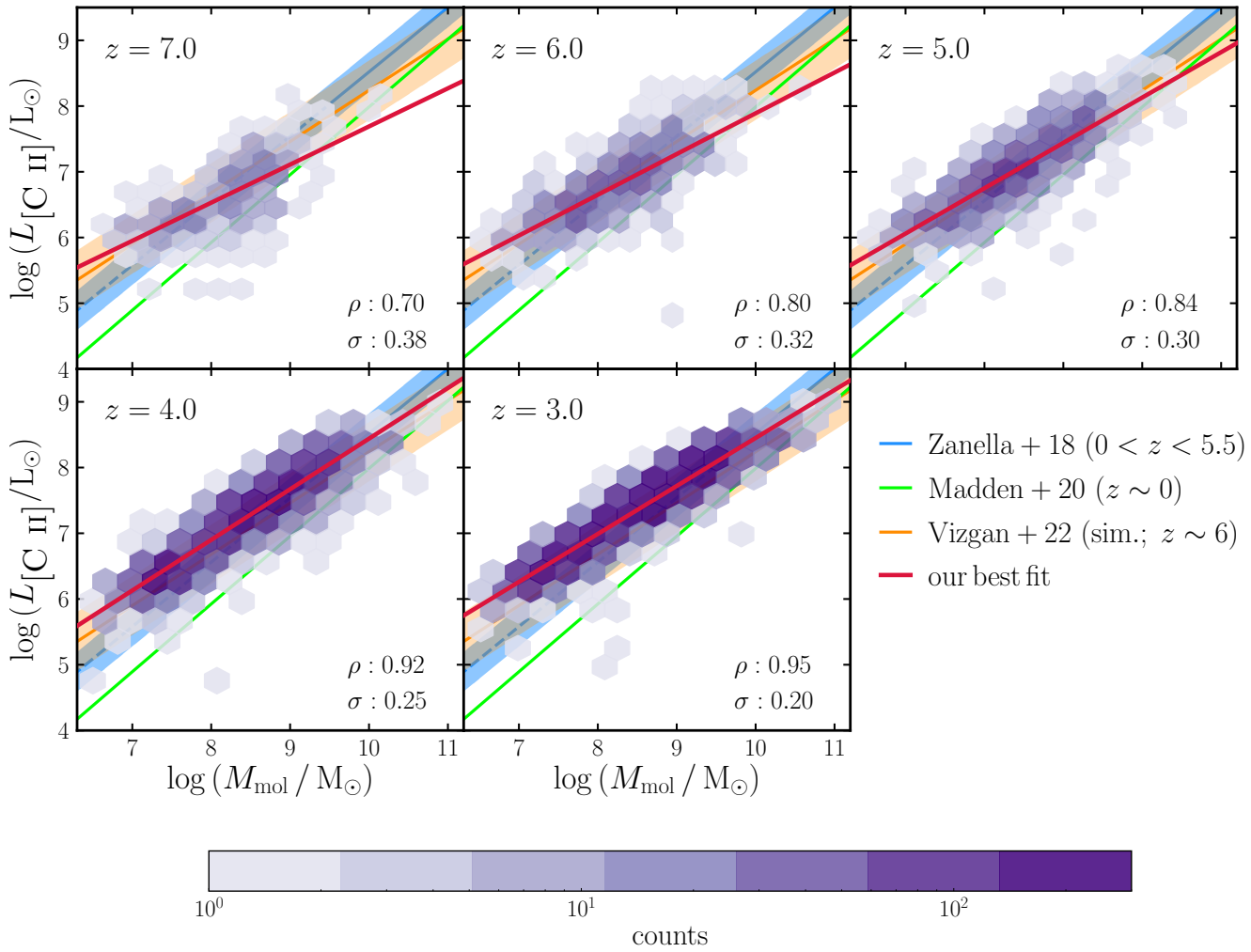


Fig. 7: The [C II] − M_{mol} relation from our simulations compared with observations. The simulated galaxies are represented by purple hexbins, where the colour indicates the number of galaxies in each bin. The solid red line gives the ordinary least squares linear fit to these galaxies, with the fit parameters listed in Table 3. In each panel, we report the Spearman's rank correlation coefficient (ρ) and the 1σ scatter around the best-fit relation. The best fit to the observed galaxy sample at $z = 0 - 5.5$ by Zanella et al. (2018) is shown in blue and the fit to the $z \sim 0$ dwarf galaxies (Madden et al. 2020) is shown in lime. The relation from Simba simulations at $z = 6$ (Vizgan et al. 2022) is shown in orange. As in Fig. 5, the extrapolated Zanella et al. (2018) relation is shown as a dashed line of the same colour.

median shows a flattening. This is driven by the increased abundance of CO at the expense of C^+ along high surface density lines of sight.

6. [C II] emission as a molecular gas tracer

Zanella et al. (2018) identified a strong correlation between the luminosity of the [C II] line and the molecular gas mass based on a compilation of galaxies in the redshift range $0 \leq z \leq 5.5$. Their sample includes local dwarf galaxies, main-sequence galaxies at $0 \leq z \leq 5.5$, starburst galaxies at $0 \lesssim z \lesssim 2$ and local luminous/ultra-luminous infrared galaxies. The best fit to their galaxy sample is given by:

$$\log\left(\frac{L_{[\mathrm{C\,II}]}}{\mathrm{L}_\odot}\right) = -1.28(\pm0.21) + 0.98(\pm0.02)\log\left(\frac{M_{\mathrm{mol}}}{\mathrm{M}_\odot}\right), \quad (4)$$

with a scatter of ≈ 0.3 dex around the best fit. Consequently, [C II] emission is now routinely used as a molecular gas tracer in high-redshift galaxies and it seems to be as good as conventional tracers like CO rotational lines and dust continuum emission. For instance, Dessauges-Zavadsky et al. (2020) found that the [C II]-based molecular gas mass estimates for ALPINE galaxies were consistent with those derived from dynamical mass estimates and (rest-frame) infrared luminosities. Aravena et al. (2024) found similar results for REBELS galaxies at $6.5 \lesssim z \lesssim 7.5$.

However, this relation and its redshift evolution have not been extensively explored in numerical simulations. While Vizgan et al. (2022) investigated the [C II]-M_{mol} relation in Simba simulations at a specific redshift of $z \sim 6$, a detailed study at different epochs is still lacking. To address this, here we examine the [C II]-molecular gas connection in the Marigold galaxies at different redshifts and develop a prescription for inferring the molecular gas mass of a galaxy from its [C II] emission. In Fig. 7, we compare the $L_{[\mathrm{C\,II}]} - M_{\mathrm{mol}}$ relation for our simulated galaxies at different redshifts against the calibrations from Zanella et al. (2018), Madden et al. (2020), and Vizgan et al. (2022).

Table 3: The best-fit scaling relations between the [C II] luminosity and the SFR (averaged over the last 200 Myr, SFR_{200}) and the molecular gas mass (M_{mol}) in our simulated galaxies at different redshifts.

z	No. of galaxies	$\log(L_{[CII]}/L_\odot) = a\log(SFR_{200}/M_\odot\, yr^{-1}) + b$			$\log(L_{[CII]}/L_\odot) = a\log(M_{mol}/M_\odot) + b$		
		a	b	1σ	a	b	1σ
7	330	0.947 ± 0.034	6.670 ± 0.016	0.29	0.580 ± 0.034	1.893 ± 0.277	0.38
6	692	0.907 ± 0.018	6.786 ± 0.009	0.24	0.624 ± 0.018	1.690 ± 0.148	0.32
5	1210	0.852 ± 0.011	6.905 ± 0.007	0.22	0.695 ± 0.013	1.223 ± 0.108	0.30
4	2787	0.921 ± 0.006	6.985 ± 0.004	0.21	0.772 ± 0.006	0.742 ± 0.052	0.25
3	4458	0.805 ± 0.004	7.205 ± 0.003	0.18	0.733 ± 0.004	1.160 ± 0.032	0.20

The molecular gas mass, M_{mol}, is obtained by scaling up the H_2 mass directly obtained from the simulations by 1.36 to account for the contribution of helium and heavier elements confined with H_2. At each redshift, we perform an ordinary least squares regression to fit a linear relation of the form: $\log(L_{[CII]}) = a\log(M_{mol}) + b$ to our galaxies. The corresponding best-fit parameters and the 1σ dispersion around the best fit are listed in Table 3. We observe that the slope increases with time from $z = 7$ to $z = 4$ and thereafter decreases slightly (by ≈ 0.04 dex). The overall change of ≈ 0.2 dex in the slope between $z = 7$ and $z = 3$ is similar to the change of ≈ 0.15 dex in the slope of the [C II] − SFR relation. At all redshifts, our best-fit slope is shallower than that of the Zanella et al. (2018) relation. As a result only our high-mass ($M_{mol} \gtrsim 10^9\,M_\odot$) galaxies follow their relation, while our low-mass galaxies exhibit higher $L_{[CII]}$ than expected from extrapolation of their relation.

We also report in In Fig. 7, the Spearman's rank correlation coefficient (ρ) between $L_{[CII]}$ and M_{mol} at each redshift. From the monotonically increasing values of ρ across redshift, we see that the [C II]-M_{mol} correlation is relatively weak at $z \gtrsim 5$ and becomes progressively stronger over time. This is in contrast with the [C II]-SFR correlation that exhibits a high value ($\rho \gtrsim 0.86$) out to $z = 7$. This trend is also evident from the decreasing values of the scatter (σ) around the best-fit linear relation between $L_{[CII]}$ and M_{mol} with decreasing redshift.

6.1. The conversion factor $\alpha_{[CII]}$

Now we compute the conversion factor between the [C II] luminosity and the molecular gas mass: $\alpha_{[CII]} \equiv M_{mol} / L_{[CII]}$[4]. Fig. 8 shows the correlation between $\alpha_{[CII]}$ and various galaxy properties for our simulated galaxies, namely the gas-phase metallicity, 12 + log(O/H), the SFR averaged over the last 5 Myr (SFR_5) and the ratio of the SFRs averaged over the last 5 and 200 Myr (hereafter R_{5-200}). The use of the latter is motivated by Lomaeva et al. (2022), who proposed an SFR change diagnostic derived from the ratio of the SFRs averaged over the past 5 and 200 Myr to quantify the current rate of change of the SFR. As different SFR indicators are sensitive to the SF happening on different timescales, their ratio can be used to quantify the SFR change over time. Lomaeva et al. (2022) proposes the ratio of Hα to FUV emission as a proxy for SFR_5/SFR_{200}. This quantity has also been used recently to quantify the bursty SF in galaxies observed with the James Webb Space Telescope (see e.g., Atek et al. 2024; Clarke et al. 2024).

[4] Note that the symbol α is used in Sect. 4 to denote the faint-end slope of the [C II] LF. However, throughout the text, the two are never discussed in the same context.

Table 4: Results of PCA at $z = 4$. The PCA is performed in the 5D parameter space of scaled variables as expressed in Eq.(5), where Z stands for 12 + log (O/H).

	M_{mol}	$L_{[CII]}$	SFR_5	SFR_{200}	Z	% of variance
PC1	0.46	0.46	0.43	0.46	0.42	88.10
PC2	-0.30	0.12	-0.60	0.10	0.72	7.33
PC3	0.05	-0.60	0.45	-0.38	0.55	3.15
PC4	-0.20	-0.57	0.03	0.80	-0.07	1.06
PC5	-0.81	0.32	0.49	0.01	0.03	0.35

In each panel of Fig. 8, we show the median $\alpha_{[CII]}$ in different bins of the quantity on the x-axis along with the 16-84 percentile range (denoted by error bars). Firstly, we observe that the $\alpha_{[CII]}$ values span about two orders of magnitude at all redshifts, particularly at $z \geq 6$. The Spearman's rank correlation coefficient (denoted by ρ) for the galaxy sample at different redshifts is reported in each panel. We find a weak correlation between $\alpha_{[CII]}$ and gas metallicity at all redshifts and observe a large scatter at all values of 12 + log(O/H), in agreement with Zanella et al. (2018) who found little systematic variation of $\alpha_{[CII]}$ with metallicity.

In the middle panel of Fig. 8, we observe that $\alpha_{[CII]}$ increases with SFR_5 at all redshifts, with the highest variation occurring at $z = 7$. At SFR $\gtrsim 10\,M_\odot\,yr^{-1}$, our $\alpha_{[CII]}$ values at all redshifts are in good agreement with the range from Zanella et al. (2018), but deviate towards lower values at lower SFRs. Therefore, using a constant $\alpha_{[CII]}$ (calibrated on the high-SFR galaxies alone) would lead to an overestimate of the molecular gas mass in low-SFR galaxies. A similar trend is observed in the $\alpha_{[CII]}$-R_{5-200} plane, with $\rho \in [0.68, 0.83]$. Interestingly, in this case the correlation is the strongest at high redshifts and slowly decays at lower redshifts, highlighting the increased sensitivity of $\alpha_{[CII]}$ on the star-formation history of galaxies at $z \gtrsim 6$, as quantified by SFR change parameter R_{5-200}.

6.2. Secondary dependence of the [C II]-M_{mol} relation

After examining how $\alpha_{[CII]}$ varies with other galaxy properties, we now develop a prescription for inferring the molecular gas mass of a galaxy from its [C II] emission, taking into account these secondary dependences. For this, we perform a principal component analysis (PCA) in the 5D space of parameters – M_{mol}, $L_{[CII]}$, SFR_5, SFR_{200} and 12 + log(O/H). PCA identifies dominant patterns and correlations between the parameters, reducing the dimensionality while approximately preserving variance. This technique has been previously used to identify secondary

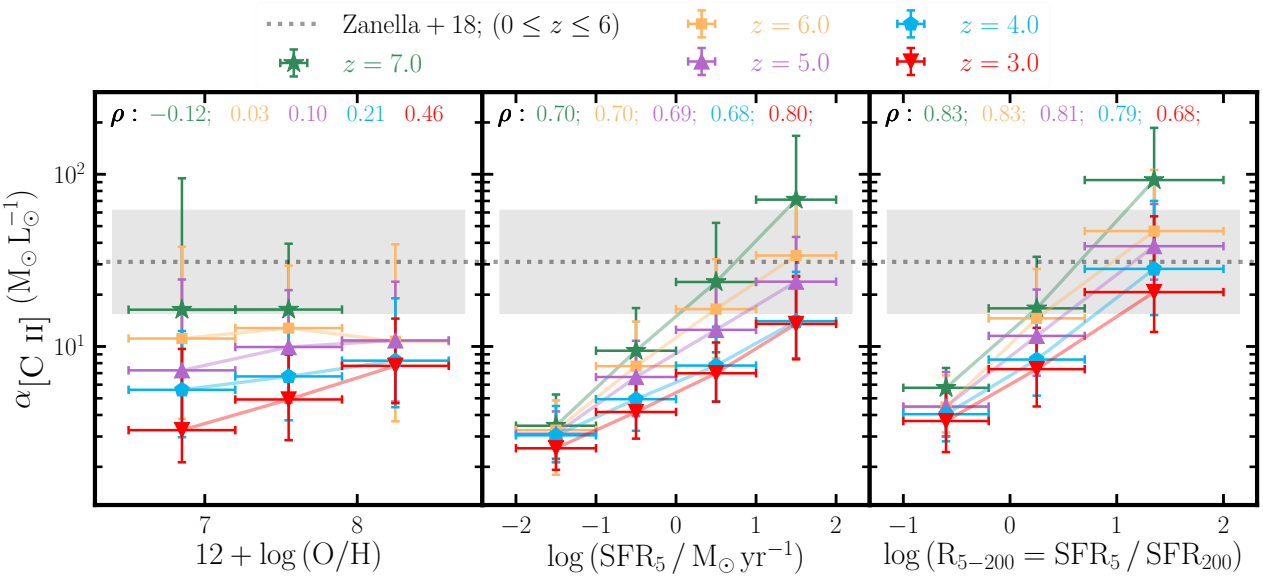


Fig. 8: The conversion factor $\alpha_{[\mathrm{C\,II}]}$ in our simulated galaxies as a function of gas metallicity (left), the SFR averaged over the last 5 Myr (middle) and the SFR change diagnostic R_{5-200} (right, see text) at different redshifts. The coloured symbols represent the median $\alpha_{[\mathrm{C\,II}]}$ in each bin, while the error bars represent the 16-84 percentiles. The Spearman's rank correlation coefficient between the variables on the y and x axes are shown in each panel as well. The grey dotted line denotes the mean $\alpha_{[\mathrm{C\,II}]}$ from Zanella et al. (2018) and the shaded band represents the corresponding scatter.

Table 5: Coefficients for the PCA-based prescriptions for estimating the molecular gas mass at different redshifts using the equation: $\log(M_{\rm mol}/\mathrm{M}_\odot) = a\log(L_{[\mathrm{C\,II}]}/\mathrm{L}_\odot) + b\log(\mathrm{SFR}_5/\mathrm{M}_\odot\,\mathrm{yr}^{-1}) + c\log(\mathrm{SFR}_{200}/\mathrm{M}_\odot\,\mathrm{yr}^{-1}) + d\,[12+\log(\mathrm{O/H})] + e$. Errors on the coefficients are obtained from bootstrapping with replacement with a 1000 iterations. The seventh and eighth columns enlist the standard deviation of the offset between the true and predicted $M_{\rm mol}$ when using the PCA-based and the best fit $M_{\rm mol}$-$L_{[\mathrm{C\,II}]}$ relations, respectively (see Fig. 9), while the last two (denoted by a *) enlist the same when accounting for typical observational uncertainties (see text).

z	a	b	c	d	e	1 σ dispersion (dex)			
						PCA	$M_{\rm mol}$-$L_{[\mathrm{C\,II}]}$	PCA*	$M_{\rm mol}$-$L_{[\mathrm{C\,II}]}$*
7	0.50 ± 0.15	0.79 ± 0.08	−0.16 ± 0.31	+0.04 ± 0.26	4.39 ± 2.80	0.20	0.45	0.28	0.47
6	0.42 ± 0.08	0.73 ± 0.02	−0.02 ± 0.10	−0.03 ± 0.07	5.46 ± 0.87	0.17	0.41	0.24	0.42
5	0.45 ± 0.04	0.65 ± 0.01	−0.00 ± 0.04	−0.03 ± 0.03	5.27 ± 0.35	0.13	0.36	0.2	0.37
4	0.47 ± 0.03	0.59 ± 0.01	+0.01 ± 0.03	+0.09 ± 0.02	4.11 ± 0.26	0.13	0.30	0.2	0.32
3	0.57 ± 0.02	0.59 ± 0.01	−0.06 ± 0.02	+0.13 ± 0.02	3.12 ± 0.17	0.11	0.25	0.19	0.28

dependencies in the stellar-mass-metallicity relation on SFR and other parameters (Mannucci et al. 2010; Lara-López et al. 2010; Hunt et al. 2012; Bothwell et al. 2016).

Since PCA is highly sensitive to extremes in the data, we first scale all parameters by their respective mean and standard deviation before performing the analysis such that for a parameter X,

$$X^{\rm scaled} = \frac{\log X - \langle \log X \rangle}{\sqrt{\mathrm{variance}(\log X)}}. \quad (5)$$

The results of this analysis are shown in Table 4. About 88% of the variance is explained by the first principal component and ∼ 95 % is explained by the first two principal components. The first component contains nearly equal contributions from all variables, while PC2 is dominated by metallicity. The last principal component, PC5 is dominated by $M_{\rm mol}$ while containing only 0.35% of the sample variance. Therefore, we set PC5 to zero to obtain an expression for $M_{\rm mol}$ in terms of the other quantities:

$$\log(M_{\rm mol}/\mathrm{M}_\odot) = 4.11 + 0.47\log(L_{[\mathrm{C\,II}]}/\mathrm{L}_\odot) + 0.59\log(\mathrm{SFR}_5/\mathrm{M}_\odot\,\mathrm{yr}^{-1}) + 0.01\log(\mathrm{SFR}_{200}/\mathrm{M}_\odot\,\mathrm{yr}^{-1}) + 0.09\,[12+\log(\mathrm{O/H})]\,. \quad (6)$$

The $M_{\rm mol}$ obtained from Eq. (6) versus the true $M_{\rm mol}$ is shown in Fig. 9. We also contrast this with the $M_{\rm mol}$ obtained from the best-fit relation between $M_{\rm mol}$ and $L_{[\mathrm{C\,II}]}$[5]. The latter shows ≈ 2.3 times higher scatter. The 1σ standard deviation between the true $M_{\rm mol}$ and the predicted $M_{\rm mol}$ using the PCA-based relation is 0.13 implying that for most (95%) of the galaxies, the PCA relation predicts the true molecular gas within a factor of ∼ 1.8 while using the two variable linear best-fit, the molecular gas mass is predicted within a factor of 4. It is worth noting that

[5] Note that this best fit is different from the one listed in Table 3 as the dependent and independent variables are reversed.

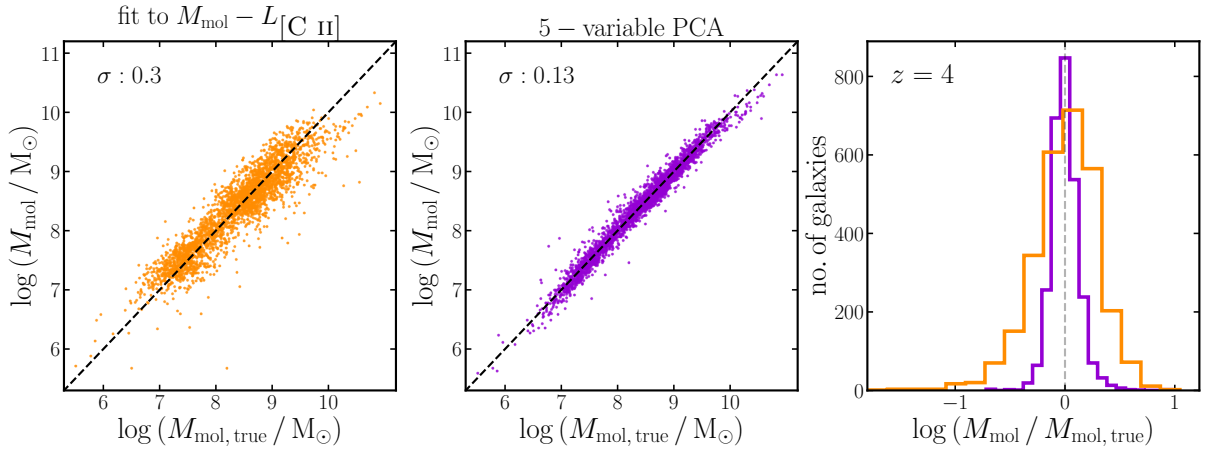


Fig. 9: Comparison of the predicted molecular gas mass estimates with the true molecular gas mass using two different approaches for simulated galaxies at $z = 4$. The left panel shows the performance of the best-fit relation between $M_{\rm mol}$ and $L_{\rm [C\,II]}$, while the middle panel shows the same using the PCA-based relation (Table 5). In the top left corner, σ denotes the standard deviation of the offset between the true and predicted $M_{\rm mol}$ (both in log scale). The right panel compares the distribution of the offsets from the two approaches.

while the linear relation systematically underpredicts $M_{\rm mol}$ for $M_{\rm mol} \gtrsim 10^{10}\,{\rm M}_\odot$ (as the linear fit is heavily influenced by the more numerous low-mass galaxies), the PCA-based prediction does not suffer from this discrepancy.

We obtain similar results at other redshifts; these are listed in Table 5. To estimate the uncertainties in the coefficients of the PCA-based relation, we perform a bootstrapping analysis with replacement, using 1000 iterations at each redshift. We find that the coefficient a, that quantifies the dependency of $M_{\rm mol}$ on $L_{\rm [C\,II]}$, increases from $z = 6$ to $z = 3$, while the coefficient b, that represents the dependency of $M_{\rm mol}$ on ${\rm SFR}_5$ decreases such that at $z = 3$, $a \approx b$. Interestingly, the dependence of $M_{\rm mol}$ on ${\rm SFR}_{200}$ and metallicity is relatively low at all times and does not evolve significantly with redshift.

6.2.1. Accounting for observational uncertainties

To estimate the impact of observational uncertainties in measurements of $L_{\rm [C\,II]}$, ${\rm SFR}_5$, ${\rm SFR}_5$, and 12+log (O/H) in estimating $M_{\rm mol}$, we add a random perturbation, δ, to each of the four quantities, where δ is drawn from a normal distribution with a mean of zero and a standard deviation corresponding to the typical errors reported in observations of high-z galaxies. For instance we adopt an error of 0.1 dex for $L_{\rm [C\,II]}$ and 0.24 dex for ${\rm SFR}_5$ and ${\rm SFR}_{200}$, both based on ALPINE galaxies (Béthermin et al. 2020). For 12 + log(O/H), we adopt an error of 0.05 dex (Sanders et al. 2015). Then we employ the PCA-based relations shown in Table 5 to obtain a prediction for $M_{\rm mol}$ and compute the offset from between our prediction and the true $M_{\rm mol}$. We then compute the standard deviation of this offset (as before, the offset is computed from the log of the quantities). We repeat the procedure for the best-fit $M_{\rm mol} - L_{\rm [C\,II]}$ relation. The resulting σ values are reported in the last two columns of Table 5. Overall, even accounting for observational uncertainties, the PCA-based relation provides a significant gain in the precision/accuracy of molecular gas mass estimates and is capable of predicting the true molecular gas mass within a factor of 2 at $3 \leq z \leq 6$.

6.3. Correlation with other quantities

In addition to the SFR and the molecular gas mass, the [C II] emission at high redshifts has been shown to correlate with the metal mass in the gas-phase of high-z galaxies (Heintz et al. 2023). In Khatri et al. (2024), we found that the C^+ distribution in our post-processed galaxy shows a grater similarity with the total gas distribution while CO and atomic carbon correlate better with the molecular gas. Inspired by these findings, we further investigate the correlation between $L_{\rm [C\,II]}$ and the total gas mass ($M_{\rm gas}$) and the mass in metals ($M_{\rm metal}$). At each redshift, we perform an ordinary least squares regression and the resulting parameters and the 1σ scatter around the best fit are listed in Table 6.

We find that among the galaxy properties explored so far, namely the SFR, the molecular gas mass, the total gas mass, and the metal mass, the [C II] emission in our simulated galaxies shows the strongest/tightest correlation with $M_{\rm metal}$ across redshifts. This is evident from the lowest scatter in this relation. In other words, we can say that the gas-phase metal mass is the most robust predictor of the [C II] emission in any given galaxy. Interestingly, the scatter in the $L_{\rm [C\,II]} - M_{\rm metal}$ relation is lower than the scatter in the $L_{\rm [C\,II]} - M_{\rm gas}$ relation at all times. A stronger correlation between [C II] emission and total gas mass simply highlights the multi-phase origin of the emission.

After exploring the correlation between $L_{\rm [C\,II]}$ and other galaxy properties in our simulations, we now turn our attention to examining the spatial extent of the [C II] emission in these galaxies and how it varies across the galaxy population.

7. Extended [C II] emission

In recent years, several observations of high-z ($z \gtrsim 4$) galaxies have revealed that the [C II] emission extends 2 − 3 times farther than the UV continuum emission. These findings come from both stacked galaxy samples (e.g., Fujimoto et al. 2019; Ginolfi et al. 2020; Fudamoto et al. 2022) and individual galaxies (e.g., Fujimoto et al. 2020; Lambert et al. 2023; Posses et al. 2024), with some studies, like Posses et al. (2024), resolving emission on $\sim$ 1 kpc scales within galaxies. Potential sources of this extended emission include unresolved satellites, outflows, and extended

Table 6: The best-fit scaling relations between the [C II] luminosity and the total gas mass ($M_{\rm gas}$) and the metal mass in the gas phase ($M_{\rm metal}$) in our simulated galaxies at different redshifts.

z	$\log(L_{\rm [C\,II]}/{\rm L}_\odot) = a\log(M_{\rm gas}/{\rm M}_\odot) + b$			$\log(L_{\rm [C\,II]}/{\rm L}_\odot) = a\log(M_{\rm metal}/{\rm M}_\odot) + b$		
	a	b	1σ	a	b	1σ
7	1.080 ± 0.032	-3.437 ± 0.298	0.25	0.888 ± 0.016	-0.379 ± 0.128	0.17
6	1.064 ± 0.019	-3.209 ± 0.179	0.22	0.851 ± 0.009	-0.044 ± 0.072	0.14
5	1.116 ± 0.014	-3.641 ± 0.132	0.22	0.843 ± 0.005	+0.048 ± 0.045	0.12
4	1.193 ± 0.006	-4.169 ± 0.057	0.16	0.851 ± 0.003	+0.059 ± 0.024	0.11
3	1.158 ± 0.004	-3.753 ± 0.040	0.14	0.798 ± 0.002	+0.481 ± 0.018	0.10

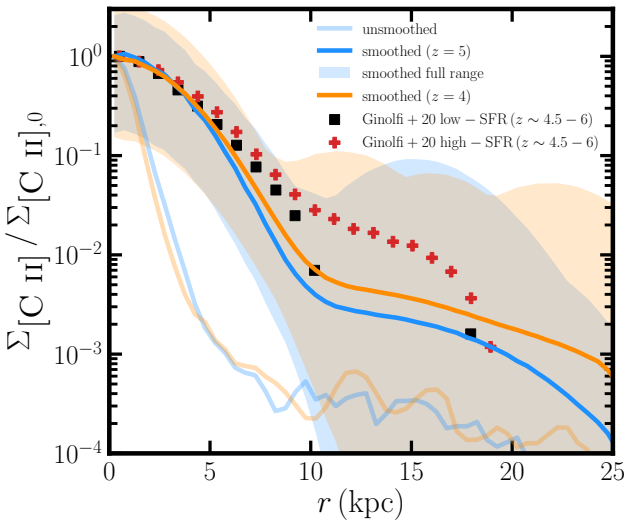


Fig. 10: Comparison of the simulated and observed stacked (radial) surface brightness profiles of the [C II] emission. The observed profiles are from Ginolfi et al. (2020) for two samples of 25 galaxies each from the ALPINE survey at $z = 4.5 - 5.9$ – a low-SFR (SFR $< 25\,{\rm M}_\odot\,{\rm yr}^{-1}$, black squares) sample and high-SFR (SFR $\geq 25\,{\rm M}_\odot\,{\rm yr}^{-1}$, red pluses) sample. The solid lines represent the stacked profiles of the 50 highest star-forming (central) galaxies from the simulation at $z = 5$ (in blue) and $z = 4$ (in orange). The shaded areas represent the full range spanned by the individual profiles, which are constructed from the 2D projection of a 50 kpc cube centred on the galaxy along three orthogonal lines of sight. All profiles are smoothed with a 2D Gaussian beam of FWHM 0.9 ″ (as in Ginolfi et al. (2020)) and normalized by the peak value of the stack. For reference, the unsmoothed stacked profiles are shown in lighter shades.

PDRs (see, e.g., Figure 12 in Fujimoto et al. 2019). In Khatri et al. (2024), we examined the extent of C^+ in a simulated galaxy by calculating the H_2, CO, C, and C^+ abundances in post-processing using the sub-grid model HYACINTH, and found that the C^+ surface density profile is more extended than other components such as H_2 and CO, and closely resembles the total gas distribution and that of young stars (with ages $\leq$ 20 Myr). Here we extend our analysis to a statistical sample of galaxies from the MARIGOLD suite and examine the extent of their [C II] emission. For this, we first look at the stacked emission from a sample of galaxies at two different redshifts (Sect. 7.1). Then we inspect the relative extent of the [C II] emission with respect to SF in individual galaxies and investigate possible causes of an extended [C II] emission (Sect. 7.2).

7.1. Stacked [C II] emission

In this section, we compare the stacked [C II] emission from our simulated galaxies with that from Ginolfi et al. (2020) based on 50 [C II]-detected galaxies from ALPINE at $z = 4.5 - 5.9$. The stacked [C II] emission from these galaxies extends out to $\sim$ 15 kpc. The authors further split their sample into low star-forming (SFR $< 25\,{\rm M}_\odot\,{\rm yr}^{-1}$) and high star-forming (SFR $\geq 25\,{\rm M}_\odot\,{\rm yr}^{-1}$ galaxies) and reported that the extended [C II] emission is more prominent in the latter.

To compare with these observations, we selected the top 50 highest star-forming (central) galaxies each at redshifts $z = 4$ and $z = 5$. For each galaxy, we obtained the [C II] surface brightness map from the 2D projection of a cylinder centred on the galaxy with a line-of-sight velocity cut of $v_{\rm los} \in [-200, 200]\,{\rm km\,s}^{-1}$ (same as in Ginolfi et al. 2020) along three orthogonal lines of sights (these are taken to be the coordinate axes of the simulation box). Note that in this analysis, we adopted the stellar centre of the galaxy to be the spatial centre of the [C II] emission[6]. We smoothed the resulting surface brightness maps with a 2D Gaussian beam of FWHM 0.9″ to mimic the synthesized ALMA beam in Ginolfi et al. (2020). We then stack the individual smoothed surface brightness maps. Note that, since we do not add noise to the galaxy images, our stacking procedure is unweighted by construction and is different from the variance-weighted stacking employed in observations.

From the stacked surface brightness map at each redshift, we extract radial profiles in bins of size 0.5 kpc. The resulting profiles are shown as solid lines in Fig. 10. The shaded areas represents the full range covered by the individual profiles. We also include for reference, the profiles obtained without smoothing using a faint line of the same colour. We see that while the true [C II] distribution in the simulated galaxies is rather compact, smoothing extends the profiles out to larger radii. We observe that our stacked profiles show a remarkable agreement with the Ginolfi et al. (2020) profile for their low-SFR sample. However our stacked profile is not as extended as their high-SFR sample. Nevertheless, some individual galaxies exhibit a similar extent as the high-SFR Ginolfi et al. (2020) profile (as indicated by the spread of our stacked profiles).

[6] Note that the centres of the different baryonic components do not necessarily overlap. However, since we apply Gaussian smoothing to the surface brightness maps with a 0.9 ″ beam, which is equivalent to 6.4 kpc (5.5 kpc) at $z = 5$ ($z = 4$), we do not expect this offset to significantly impact our results.

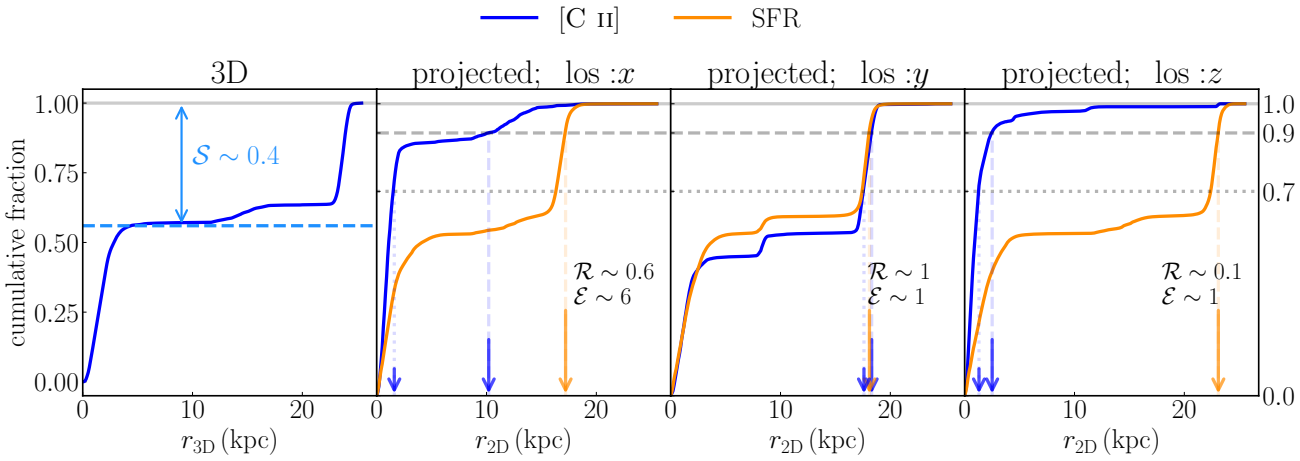


Fig. 11: An example illustrating the calculation of the $\mathcal{S}$, $\mathcal{R}$, and $\mathcal{E}$ parameters in a simulated galaxy. The left panel shows the the cumulative profile for the 3D distribution of [C II], with the dashed line marks the contribution of the central galaxy to the total [C II] emission (as evident from the flattening of the cumulative profile). The remaining fraction (denoted by $\mathcal{S}$) represents the contribution of satellites. The other panels show cumulative profile constructed from the [C II] surface brightness (blue) and SFR surface density (orange). For the profiles obtained from projections, the value of the $\mathcal{R}$ and $\mathcal{E}$ parameters are indicated in each panel. In all but the leftmost panel, the dotted, dashed and solid horizontal lines denote cumulative fractions of 70%, 90%, and 100%, respectively. The small and large blue arrows mark $r_{70,\,[\mathrm{C\,II}]}$ and $r_{90,\,[\mathrm{C\,II}]}$, respectively and the the orange arrow denotes $r_{90,\,\mathrm{SFR}}$. The parameters are $\mathcal{R} \equiv r_{90,\,[\mathrm{C\,II}]}/r_{90,\mathrm{SFR}}$ is calculated from the ratio of the r values denoted by the large blue and orange arrows, while the parameter $\mathcal{E} \equiv r_{90,\,[\mathrm{C\,II}]}/r_{70,[\mathrm{C\,II}]}$ from the ratio of the large and small blue arrows.

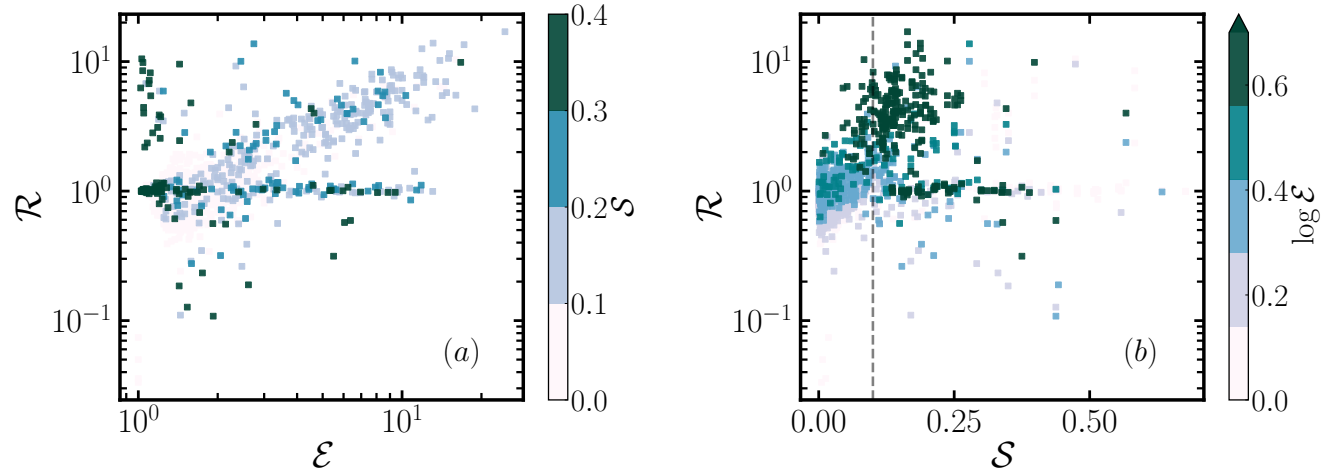


Fig. 12: The distribution of our galaxies at redshift $z = 4$ in the $\mathcal{R}$-$\mathcal{E}$ (*left*), $\mathcal{R}$-$\mathcal{S}$ (*right*) planes. The scatter points are colour-coded by the parameter $\mathcal{S}$ in the left panel and by the log of parameter $\mathcal{E}$ in the right panel. The dashed grey lines in the right panel indicates the threshold value of $\mathcal{S}$ used to separate galaxies into low and high $\mathcal{S}$.

7.2. Extent of [C II] emission in individual objects

Now we aim to quantify the extent of the [C II] emission relative to the SF activity in our simulated galaxies. For this analysis, we include all galaxies that meet the following criteria: (i) $M_* \geq 10^{8.5}\,\mathrm{M}_\odot$ and (ii) SFR $\geq 3\,\mathrm{M}_\odot\,\mathrm{yr}^{-1}$. These are chosen to match the range of M_* and SFR values in ALPINE galaxies (Le Fèvre et al. 2020). For each galaxy in our sample, we take a sphere of radius 25 kpc centred on the galaxy and obtain projections of the same along three orthogonal lines of sights (we take these to be the coordinate axes of the simulation box) to obtain the [C II] surface brightness and SFR surface density maps. Then we compute the (cumulative) radial surface brightness profiles for each projection, from which we derive the radius enclosing 70% and 90% of the total [C II] emission – these are denoted as $r_{70,[\mathrm{C\,II}]}$ and $r_{90,[\mathrm{C\,II}]}$, respectively. Similarly, from the (cumulative) SFR surface density profile, we derive $r_{90,\,\mathrm{SFR}}$ (as in Sect. 5, we use the SFR averaged over the last 200 Myr). For each galaxy, we also compute the cumulative [C II] luminosity profile from the full 3D distribution of [C II] within the region.

Based on these, we calculate the following parameters for each galaxy for the three orthogonal projections. As an example, we show in Fig. 11, cumulative profiles for a simulated galaxy and how these are used for calculating the three parameters and to ease their interpretation.

1. The multicomponent extent parameter $\mathcal{E} \equiv r_{90,\,[\mathrm{C\,II}]}\,/\,r_{70,\,[\mathrm{C\,II}]}$ that measures the spread in the [C II] emission. A higher $\mathcal{E}$ indicates a relatively larger extent of

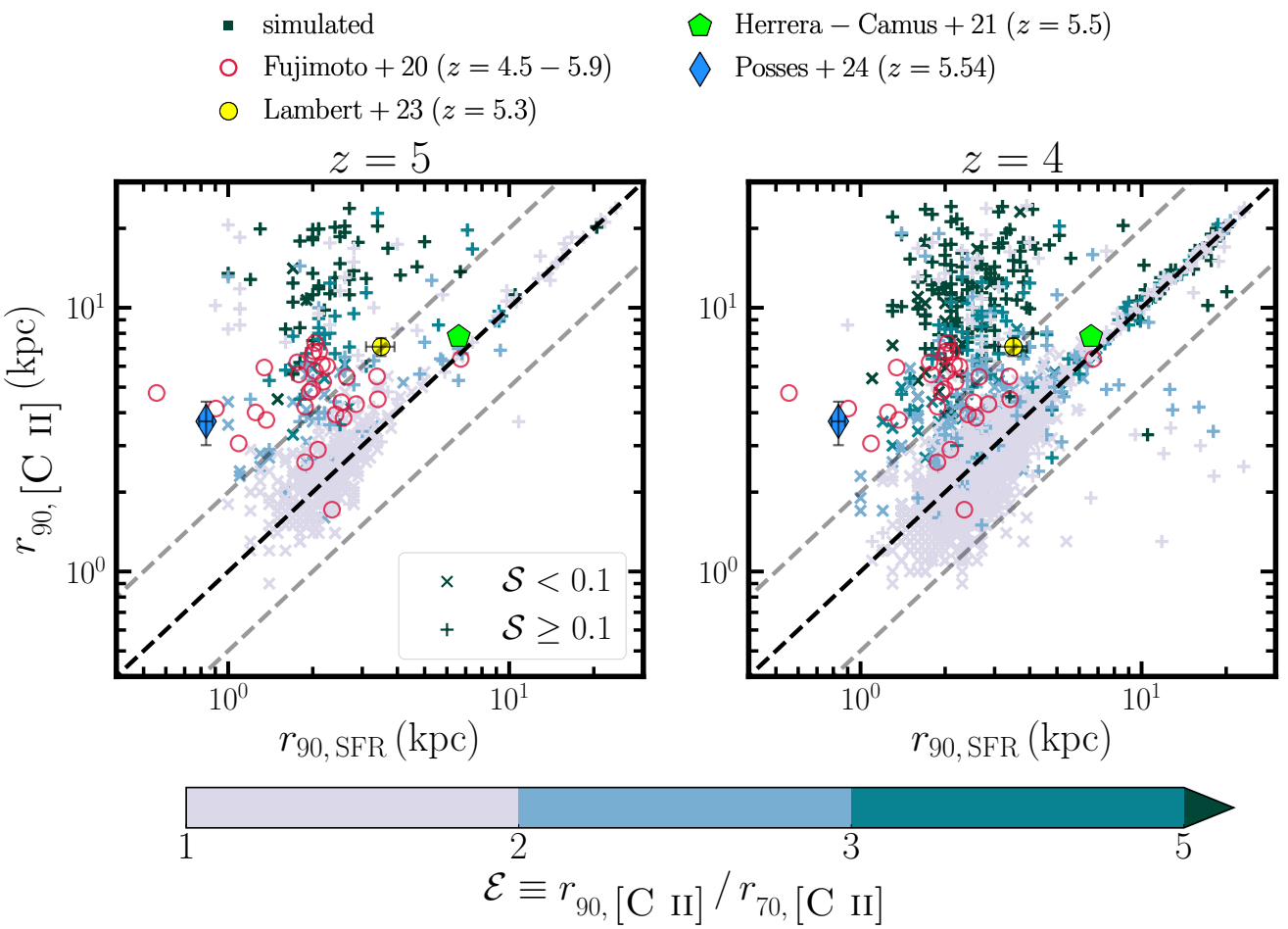


Fig. 13: Comparison of $r_{90,\,[\mathrm{C\,II}]}$ and $r_{90,\,\mathrm{SFR}}$ for simulated galaxies at redshifts $z = 5$ (left) and 4 (right). The galaxies are colour-coded by their multicomponent extent parameter $\mathcal{E}$ defined as the ratio of the r_{90} and r_{70} values of the [C II] surface brightness profile. The shape of the symbol reflects the $\mathcal{S}$ parameter that quantifies the satellite contribution to the total [C II] emission (see text for details). We use 'low $\mathcal{S}$' and 'high $\mathcal{S}$', respectively to denote galaxies with < 10% and ≥ 10% satellite contribution. These $r_{90,\,[\mathrm{C\,II}]}$ versus $r_{90,\,\mathrm{SFR}}$ values of observed galaxies are shown as red open circles (Fujimoto et al. 2020), a yellow plus (Lambert et al. 2023), a green pentagon (Herrera-Camus et al. 2021), and a blue diamond (Posses et al. 2024). The black dashed line indicates a 1:1 relation, while the top and bottom grey dashed lines indicate 2:1 and 1:2 relations, respectively. Note that the error bars are not shown for Fujimoto et al. (2020) galaxies for the sake of clarity.

the diffuse [C II] component. This would occur in galaxies where 70% of the emission is relatively confined, while the remaining 10 − 30% is more spread out, likely due to the presence of satellite galaxies. Thus, a higher $\mathcal{E}$ denotes more extended emission relative to the bulk of the emission.

2. The parameter $\mathcal{R} \equiv r_{90,\,[\mathrm{C\,II}]} \,/\, r_{90,\,\mathrm{SFR}}$ that quantifies the relative extent of the [C II] emission compared to SF. In this analysis, we are particularly interested in galaxies where the [C II] emission is at least twice as extended as the SF activity (i.e., $\mathcal{R} \geq 2$).
3. The parameter $\mathcal{S}$ that quantifies the fraction of the total [C II] emission that arises from outside the central galaxy and represents the contribution from satellites Unlike, $\mathcal{R}$ and $\mathcal{E}$, this parameter is computed from the true 3D distribution of [C II] emission and is therefore agnostic to the projection axis by construction.

Note that all three parameters are agnostic to whether the [C II] emission at a given location represents a C^+ ion formed in situ or transported there, e.g., via outflows/inflows.

From Fig. 11, it is evident that the parameters $\mathcal{R}$ and $\mathcal{E}$ are sensitive to the orientation of the galaxy. For instance, the y-projection of the galaxy has $\mathcal{R} \sim 1$, indicating equal extent of the [C II] emission and SFR. In contrast, in the x- and z- projections, the SFR profile is more extended than the [C II] profile, resulting in $\mathcal{R} < 1$. Note that in this case the $\mathcal{E}$ value is relatively high , i.e., the $r_{70,\,[\mathrm{C\,II}]}$ and $r_{90,\,[\mathrm{C\,II}]}$ are well separated, but owing to a more extended SFR profile, the $\mathcal{R}$ value is low.

To better understand how the three parameters are correlated, we show in Fig. 12, how $\mathcal{R}$ varies with $\mathcal{E}$ and $\mathcal{S}$ for our galaxy sample at $z = 4$. Each scatter point represents one of the orthogonal projections of a galaxy and is colour-coded by value of the third parameter. Firstly, in panel (a), we see that the parameter $\mathcal{R}$ increases in general with $\mathcal{E}$. However, some galaxies[7] with $\mathcal{R} \gtrsim 2$ exhibit a low $\mathcal{E}$ ($\lesssim 2$). In contrast, in panel (b), we see that all galaxies with a high $\mathcal{R}$ generally have a high $\mathcal{S}$ as well, although the reverse is not always true – some galaxies have a high $\mathcal{S}$ but $\mathcal{R} \sim 1$. In these galaxies, satellites contribute similarly to the [C II] emission and the SF, driving the r_{90} of both quantities

[7] Note that while $\mathcal{R}$ and $\mathcal{E}$ are properties of the projection of a given galaxy, for simplicity, we associate these parameters directly with the galaxy itself.

to high values, and thereby reducing the $\mathcal{R}$ (the y-projection in Fig. 11 is an example of this).

In Fig. 13, we compare the extent of the r_{90} values of the [C II] emission and the SFR. The galaxies are colour-coded according to their $\mathcal{E}$ parameter while the shape of the symbol represents the value of the $\mathcal{S}$ parameter: we split the galaxies into 'low $\mathcal{S}$' ($\mathcal{S} < 0.1$) and 'high $\mathcal{S}$' ($\mathcal{S} \geq 0.1$) subsamples. The threshold of 0.1 or 10% is inspired by Springel et al. (2008) who found that $\approx$ 11% of the mass fraction in virialized haloes is present in substructures. For reference, we also show the r_{90} values for [C II] and SFR from observations of individual galaxies: to obtain these, we use the half-light radii r_e for the [C II] emission and UV continuum emission, reported in the respective observations. In all observations shown in Fig. 13, $r_{e,\,[C\,II]}$ and $r_{e,\,UV}$ are obtained from fitting exponential disk-profiles to the [C II] emission and UV continuum emission, respectively. We scale these r_e values by 2.318 to obtain the respective r_{90} values[8]. Since the bulk of the UV emission from galaxies arises from stars younger than $\approx$ 100 Myr (Kennicutt & Evans 2012), we further assume that the $r_{90,\,SFR}$ in the observed galaxies can be approximated by the $r_{90,\,UV}$ values. A possible caveat is that the UV sizes might be larger than the corresponding SFR sizes (e.g., see Szomoru et al. 2013, for a comparison of the half-light and half-mass radii of $0.5 < z < 2$ galaxies.).

From Fig. 13, we see that the [C II] emission in many of our simulated galaxies have a similar extent as the observed galaxies at $4.5 \lesssim z \lesssim 5.9$ from Fujimoto et al. (2020), Herrera-Camus et al. (2021), and Lambert et al. (2023). However, our simulated galaxies occupy a larger area of the parameter space compared to observations. Our galaxies where the [C II] emission is at least twice as extended as the SF activity (i.e., above the top grey dashed line, $\mathcal{R} \geq 2$) exhibit preferentially higher $\mathcal{E}$ values (darker colours of the symbol) and a higher contribution from satellites, compared to galaxies lying along the diagonal (i.e. $\mathcal{R} \sim 1$). This is also evident from the higher median of the $\mathcal{E}$ and $\mathcal{S}$ parameters for the galaxies with $\mathcal{R} \geq 2$ (see Table 7).

To summarize, we find that the inferred detection of an extended [C II] emission in a given galaxy is sensitive to the orientation of the galaxy. Nevertheless, some statistical differences emerge between galaxies with extended [C II] emission compared to their SF activity (i.e., with $\mathcal{R} \equiv r_{90,\,[C\,II]} \,/\, r_{90,\,SFR} \geq 2$) and those without (i.e., $\mathcal{R} < 2$). Galaxies with $\mathcal{R} \geq 2$ tend to have a higher contribution from satellites. They also exhibit higher $\mathcal{E}$ values compared to the latter, indicating that while the bulk ($\sim$ 70%) of the [C II] emission is relatively concentrated, the remaining fraction can extend out to 4-5 times larger radii. Overall, about 10% of our simulated galaxies at $z = 4$ have $\mathcal{R} \geq 2$ i.e., their [C II] emission extends ≥ 2 times farther than the SF activity; this fraction increases to 20% at $z = 5$. This is in agreement with recent observations pointing out the increased prevalence of extended [C II] emission towards higher redshifts (Carniani et al. 2018; Fujimoto et al. 2019; Ginolfi et al. 2020; Fudamoto et al. 2022).

8. Comparison with previous work

In Fig. 14, we compare our best-fit [C II]-SFR relation (Table 3) with previous numerical work in the literature, briefly described in the following. Lagache et al. (2018) used a semi-analytical galaxy-formation model coupled to the photoionisation code CLOUDY (Ferland et al. 1992) to obtain the [C II] luminosity for a statistical sample of galaxies at redshifts $4 \leq z \leq 8$. From their full sample of galaxies, they obtain a [C II]-SFR relation of the form: $\log(L_{[C\,II]}/L_\odot) = (1.4 - 0.07z)\log(SFR/M_\odot\,yr^{-1}) + 7.1 - 0.07z$, which is shown as solid lime line in the figure. in their galaxies. Vallini et al. (2015) calculated the [C II] emission from a single $z \sim 7$ galaxy from a high-resolution ($\approx$ 60 pc) SPH simulation (Pallottini et al. 2014) assuming optically-thin emission. To do so, they adopt a log-normal sub-grid density distribution with a variable Mach number. In this regard, their approach is similar to ours except that we adopt a log-normal+power-law PDF in regions where self-gravity cannot be neglected (see Khatri et al. 2024, for details). They investigate how the total [C II] luminosity changes with metallicity, assuming a uniform metal distribution. To derive the [C II]-SFR relation, they scale the [C II] luminosity with the molecular gas mass ($L_{[C\,II]} \propto M_{H_2}$), which is assumed to scale with the SFR in accordance with the Kennicutt-Schmidt relation: $\Sigma_{SFR} \propto \Sigma_{H_2}^N$. In Fig. 14, we show their results for $N = 1$ (in blue) and $N = 2$ (in orange). For both cases, the dotted line is for the median gas-metallicity (Z_{gas}) of our galaxies at a given redshift, while the shaded area represents the range adopting the 16 and 84 percentiles of the metallicity distribution. Previously, Lagache et al. (2018) found that their [C II]-SFR agrees well with the one from Vallini et al. (2015) for $N = 2$. We also include the results from Leung et al. (2020) and Vizgan et al. (2022), both of which were obtained by post-processing $z \sim 6$ galaxies from the SIMBA simulations using different versions of the emission line tool SÍGAME (Olsen et al. 2015, 2017). SÍGAME includes a multi-phase ISM model and accounts for the contribution of different phases (molecular gas phase, cold neutral medium, and H II regions) to the [C II] emission of a galaxy and employs CLOUDY to obtain the chemical abundances in the different phases. For reference, we also show the [C II]-SFR relation based on observations of $0.5 \leq z \leq 6.6$ galaxies (De Looze et al. 2014) as a solid blue line with a scatter of 0.4 dex (the same is shown as a solid orange line in Fig. 5).

Table 7: Median values of the parameters $\mathcal{E}$ and $\mathcal{S}$ for galaxies with and without extended [C II] emission at $z = 4$ and $z = 5$.

	$z = 4$		$z = 5$	
	median $\mathcal{E}$	median $\mathcal{S}$	median $\mathcal{E}$	median $\mathcal{S}$
$\mathcal{R} \geq 2$	5.5	0.14	3.99	0.17
$\mathcal{R} < 2$	1.5	0.02	1.5	0.03

Overall, we see that the variation among the different models increases with redshift, and can span up to an order of magnitude at SFR $\sim 1 - 10\,M_\odot\,yr^{-1}$. This is enhanced towards lower and higher SFRs. We find that at all redshifts our [C II]-SFR relation differs significantly from Vallini et al. (2015). In this regard, while their approach for calculating the [C II] emission accounting for the sub-grid densities is similar to ours (albeit with different sub-grid density PDFs), their method to derive the [C II]-SFR relation relies on scaling relations between $L_{[CII]}$ and M_{H_2} and $\Sigma_{SFR} - \Sigma_{H_2}$, while we follow the dynamical evolution of these quantities in our simulations. Moreover, as shown by our PCA analysis (Sect. 6.2), the $L_{[C\,II]} - M_{mol}$ relation has secondary dependencies on the SFR, that evolve with redshift. This provides a natural explanation for the different results from the two studies. At $5 \leq z \leq 7$, our results, as well as those from Lagache et al. (2018), show slightly steeper slopes compared to those from Leung et al. (2020) and Vizgan et al. (2022). However, these differences remain within the scatter of $\sim 0.3 - 0.6$. Notably, at

[8] We remind the reader that for an exponential disk-profile $\Sigma = C\exp(-r/h)$, where h is the scale length of the disk, the half-light radius $r_e \approx 1.678\,h$ and $r_{90} \approx 3.890\,h$. Therefore $r_{90} \approx 2.318\,r_e$.

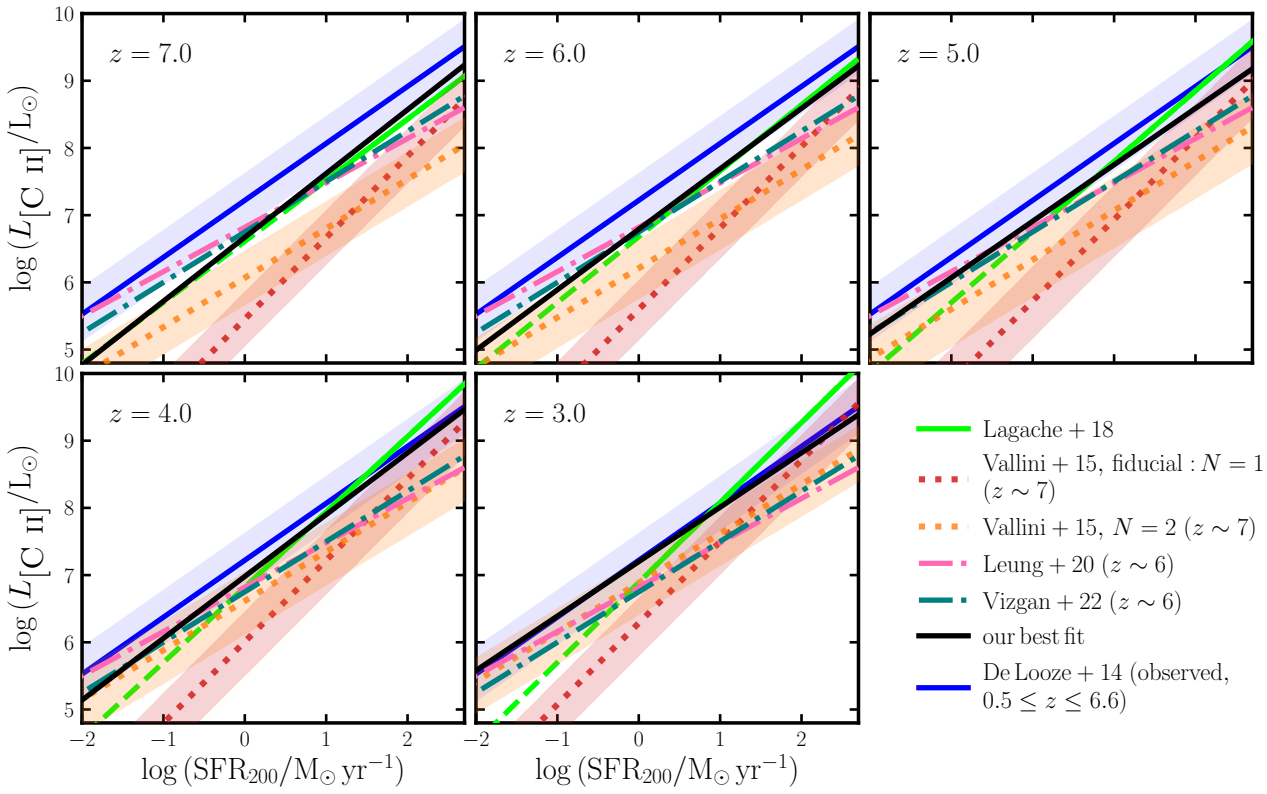


Fig. 14: Comparison of our [C II]-SFR relation (black line) at different redshifts with other relations from the literature. The relations from Lagache et al. (2018) are shown in lime. The red and orange dotted lines along with the scatter represent the Vallini et al. (2015) relations with $N = 1$ and $N = 3$, respectively for the Kennicutt-Schmidt relation. In both cases we show with the dotted line the relation assuming the median Z_{gas} of our simulated galaxies at a given redshift, while the shaded area encloses the relations with 16-84 percentile of our Z_{gas} values. The dash-dotted lines in pink and teal represent the relations from Leung et al. (2020) and Vizgan et al. (2022), respectively, both obtained by post-processing the $z \sim 6$ snapshot of the SIMBA simulations (Davé et al. 2020) with different versions of SÍGAME (Olsen et al. 2015, 2017). For reference, we include the De Looze et al. (2014) relation based on observations of $0.5 \leq z \leq 6.6$ galaxies as a blue line with a scatter of 0.4 dex shown by the shaded region.

these redshifts, the numerical predictions fall slightly below the empirical De Looze et al. (2014) relation. At $z = 3$, our best-fit shows an excellent agreement with De Looze et al. (2014) and exhibits a slightly higher offset compared to others except Lagache et al. (2018).

In Fig. 15, we compare our [C II] LF against the results from Popping et al. (2016) and Lagache et al. (2018), both based a semi-analytical galactic formation models. Popping et al. (2016) fit a Schechter functions to their predicted LFs at different redshifts. At $z = 6$, our predicted LF is very similar to Popping et al. (2016), despite the differences in our methods. However, deviations start to appear at late times. For instance, despite similar faint-end slopes at redshifts $z = 4$, our simulations predict a high number density of emitters at the all luminosities. Moreover, our LFs extend out to higher luminosities, similar to Lagache et al. (2018), who fit their LFs with a single power law with a slope of -1.0 at all redshifts. In this regard, the underabundance of bright galaxies in semi-analytical models (SAMs) that track the H_2 abundance is a well-known problem and is related to the underabundance of cold gas in the galaxies simulated with SAMs (Popping et al. 2015). At all redshifts, we find a steeper bright-end slope in comparison to Lagache et al. (2018).We further compare with the LF from Garcia et al. (2024) obtained by post-processing the SIMBA simulations at $z = 5$ and 6. We find that although consistent within 1 σ uncertainty, their predictions are consistently below ours and can be up to an order of magnitude lower at the bright end.

Discrepancies in the predicted LFs from different studies can have important consequences for predicting the power spectrum of the line intensity mapping (LIM) signal. LIM is an emerging technique that measures the integrated line emission from galaxies and the intergalactic medium without resolving individual sources (see Bernal & Kovetz 2022, for a recent review). The first moment of the LF governs the overall amplitude of the LIM power spectrum, while the second moment influences the shot noise component. As bright emitters are expected to have a dominant contribution to the [C II] LIM power spectrum at $z \gtrsim 4$ (e.g. Marcuzzo et al. 2025), current and upcoming LIM surveys can prove extremely useful in constraining the different numerical models.

9. Summary

Based on a sample of simulated galaxies with molecular cloud chemistry evolved on the fly and [C II] 158 μm line emission calculated in post-processing, we have investigated the reliability of this line as a tracer of the SF activity and molecular gas content

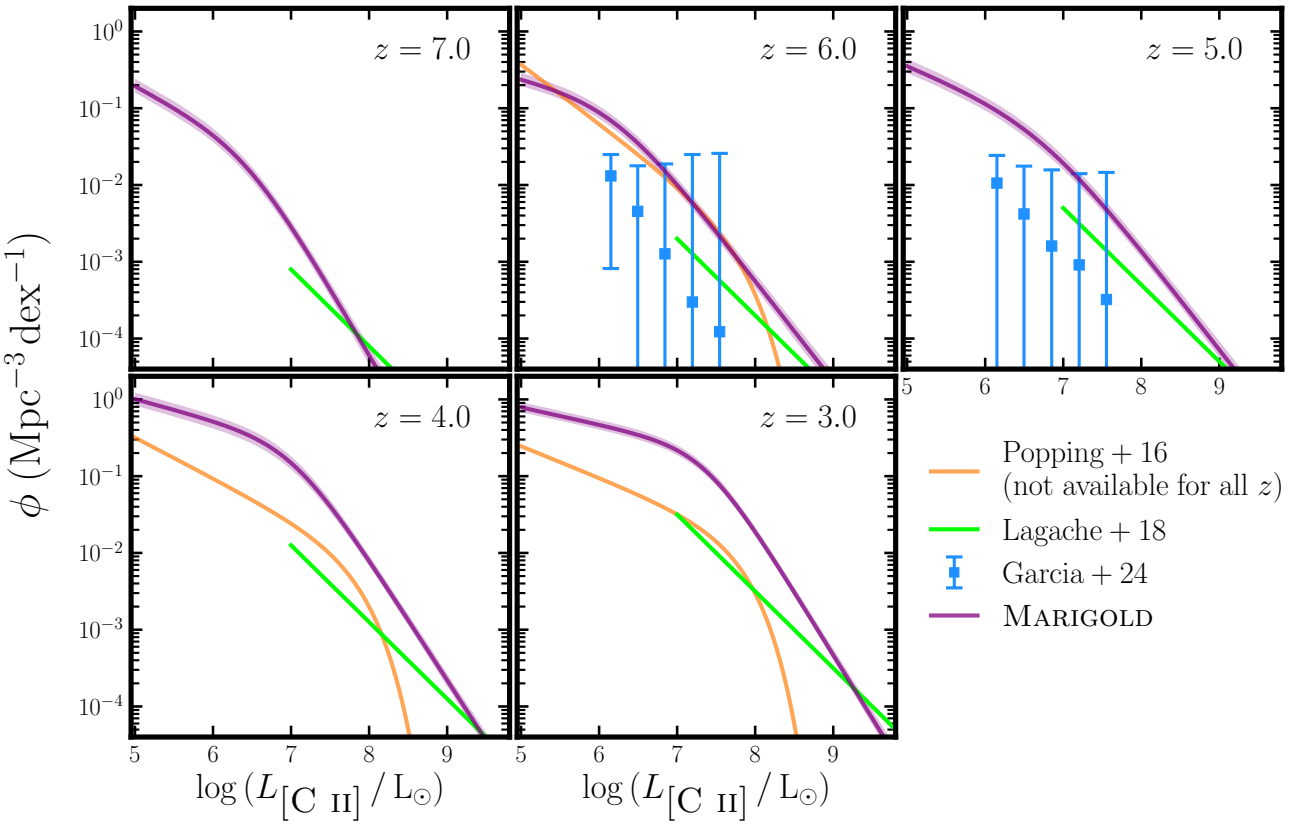


Fig. 15: Comparison of the [C II] LF from the MARIGOLD simulations at different redshifts with those from Popping et al. (2016) and Lagache et al. (2018) both based on a semi-analytical galaxy formation model, and from Garcia et al. (2024) obtained by post-processing the SIMBA simulations at $z = 6$ and 5.

in galaxies at redshifts $3 \leq z \leq 7$. Here we briefly summarize our findings:

1. **Redshift evolution of the [C II] luminosity function:** Our simulations predict a strong time evolution in the number density of [C II] emitters, especially at the bright end. The number density of $L_{[C\,II]} \sim 10^9\,L_\odot$ galaxies increases by 600 times in the above redshift range. At all redshifts, a double power-law provides a better fit to our simulated LFs (Table 2).

2. **Redshift evolution of the [C II]-SFR relation:** The slope $L_{[C\,II]}$ − SFR relation shows little evolution in the redshift range $3 \leq z \leq 7$ while the intercept increases by ≈ 0.5 dex in this interval, indicating that the [C II] luminosity at given SFR increase roughly by a factor of three from $z = 7$ to $z = 3$. Notably, the scatter in the relation increases towards higher redshifts (Table 3).

3. **The conversion factor $\alpha_{[C\,II]}$:** The conversion factor $\alpha_{[C\,II]}$ between the [C II] luminosity and the molecular gas mass in our simulated galaxies ranges from $\sim 1 - 200\,M_\odot\,L_\odot^{-1}$ and does not show a systematic dependence on metallicity in agreement with the findings from Zanella et al. (2018) based on a compilation of galaxies from $z = 0 - 5.5$. Across redshifts, $\alpha_{[C\,II]}$ shows a strong correlation with the SFR averaged over 5 Myr and with the SFR change diagnostic $R_{5-200} = \mathrm{SFR}_5/\mathrm{SFR}_{200}$ (Fig. 8).

4. **Secondary dependences in the $L_{[C\,II]} - M_{mol}$ relation:** We performed a principal component analysis to quantify secondary dependences in the [C II]-M_{mol} relation and found a strong dependence on the SFR_5 (the star formation rate measured on a 5 Myr timescale) that evolves with redshift and a weak dependence on metallicity across redshifts (Table 5). The resulting 5-variable PCA relation predicts the true molecular gas within a factor of 1.7 (2.5) at $z = 3$ ($z = 7$). When accounting for typical observational uncertainties on $L_{[C\,II]}$, SFR, and gas metallicity, the PCA-based relation predicts the true molecular gas mass within a factor of ~ 2.5 at $3 \leq z \leq 5$.

5. **What does the [C II] emission really trace?:** We investigated the correlation of $L_{[C\,II]}$ with several galaxy-integrated properties, namely the SFR, the molecular gas mass, the total gas mass, and the metal mass in gas phase (M_{metal}). Among these, the [C II] emission in our simulated galaxies shows the tightest correlation with M_{metal} across redshifts.

6. **Extended [C II] emission:** We observe that our stacked [C II] surface brightness profiles show a similar extent as the low-SFR galaxy sample from Ginolfi et al. (2020), although some individual galaxies also exhibit a similar extent as their high-SFR sample. We further find that galaxies where the [C II] emission extends twice or more farther than the SF activity preferentially exhibit a spatial distribution wherein the bulk ($\gtrsim 70\%$) of the [C II] emission is confined to the central galaxy, while the remaining $\lesssim 30\%$ extends out to larger distances because of the presence of satellites. The typical fractional contribution of satellites to the total [C II]

emission in these galaxies is $\approx 5-7$ times higher than that in galaxies without extended emission (see Table 7).

Acknowledgements. The authors thank R. Teyssier for making the RAMSES code publicly available. They gratefully acknowledge the Collaborative Research Center 1601 (SFB 1601 sub-project C5) funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 500700252. Early stages of this work were carried out within the Collaborative Research Centre 956 (SFB 956 sub-project C4), funded by the DFG – 184018867. The authors acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for funding this project by providing computing time on the GCS Supercomputer SuperMUC-NG at the Leibniz Supercomputing Centre (http://www.lrz.de). They are thankful to the Max Planck Society (MPG) for providing computing time on the supercomputer Raven at the Max Planck Computing and Data Facility (https://www.mpcdf.mpg.de), where part of the MARIGOLD simulations were performed. PK is a part of the International Max Planck Research School in Astronomy and Astrophysics, the Bonn Cologne Graduate School, and a guest at the Max Planck Institute for Radio Astronomy in Bonn.

References

Accurso, G., Saintonge, A., Catinella, B., et al. 2017, MNRAS, 470, 4750
Aravena, M., Heintz, K., Dessauges-Zavadsky, M., et al. 2024, A&A, 682, A24
Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, ARA&A, 47, 481
Atek, H., Labbé, I., Furtak, L. J., et al. 2024, Nature, 626, 975
Bernal, J. L. & Kovetz, E. D. 2022, A&A Rev., 30, 5
Béthermin, M., Fudamoto, Y., Ginolfi, M., et al. 2020, A&A, 643, A2
Bigiel, F., Leroy, A., Walter, F., et al. 2008, AJ, 136, 2846
Bothwell, M. S., Maiolino, R., Cicone, C., Peng, Y., & Wagg, J. 2016, A&A, 595, A48
Bouwens, R. J., Smit, R., Schouws, S., et al. 2022, ApJ, 931, 160
Carniani, S., Maiolino, R., Smit, R., & Amorín, R. 2018, ApJ, 854, L7
Clarke, L., Shapley, A. E., Sanders, R. L., et al. 2024, arXiv e-prints, arXiv:2406.05178
Cormier, D., Abel, N. P., Hony, S., et al. 2019, A&A, 626, A23
da Cunha, E., Groves, B., Walter, F., et al. 2013, ApJ, 766, 13
Davé, R., Crain, R. A., Stevens, A. R. H., et al. 2020, MNRAS, 497, 146
De Breuck, C., Lundgren, A., Emonts, B., et al. 2022, A&A, 658, L2
De Looze, I., Baes, M., Bendo, G. J., Cortese, L., & Fritz, J. 2011, MNRAS, 416, 2712
De Looze, I., Cormier, D., Lebouteiller, V., et al. 2014, A&A, 568, A62
Dessauges-Zavadsky, M., Ginolfi, M., Pozzi, F., et al. 2020, A&A, 643, A5
D'Eugenio, C., Daddi, E., Liu, D., & Gobat, R. 2023, A&A, 678, L9
Díaz-Santos, T., Armus, L., Charmandaris, V., et al. 2013, ApJ, 774, 68
Ferland, G. J., Peterson, B. M., Horne, K., Welsh, W. F., & Nahar, S. N. 1992, ApJ, 387, 95
Foreman-Mackey, D., Conley, A., Meierjurgen Farr, W., et al. 2013, emcee: The MCMC Hammer, Astrophysics Source Code Library, record ascl:1303.002
Fudamoto, Y., Smit, R., Bowler, R. A. A., et al. 2022, ApJ, 934, 144
Fujimoto, S., Ouchi, M., Ferrara, A., et al. 2019, ApJ, 887, 107
Fujimoto, S., Silverman, J. D., Bethermin, M., et al. 2020, ApJ, 900, 1
Garcia, K., Narayanan, D., Popping, G., et al. 2024, ApJ, 974, 197
Gehrels, N. 1986, ApJ, 303, 336
Ginolfi, M., Jones, G. C., Béthermin, M., et al. 2020, A&A, 633, A90
Goldsmith, P. F., Langer, W. D., Pineda, J. L., & Velusamy, T. 2012, ApJS, 203, 13
Gong, Y., Cooray, A., Silva, M., et al. 2012, ApJ, 745, 49
Graciá-Carpio, J., Sturm, E., Hailey-Dunsheath, S., et al. 2011, ApJ, 728, L7
Gullberg, B., De Breuck, C., Vieira, J. D., et al. 2015, MNRAS, 449, 2883
Hahn, O. & Abel, T. 2011, MNRAS, 415, 2101
Heintz, K. E., Oesch, P. A., Aravena, M., et al. 2022, ApJ, 934, L27
Heintz, K. E., Shapley, A. E., Sanders, R. L., et al. 2023, A&A, 678, A30
Heintz, K. E., Watson, D., Oesch, P. A., Narayanan, D., & Madden, S. C. 2021, ApJ, 922, 147
Herrera-Camus, R., Bolatto, A. D., Wolfire, M. G., et al. 2015, ApJ, 800, 1
Herrera-Camus, R., Förster Schreiber, N., Genzel, R., et al. 2021, A&A, 649, A31
Hu, C.-Y., Sternberg, A., & van Dishoeck, E. F. 2021, ApJ, 920, 44
Hughes, T. M., Ibar, E., Villanueva, V., et al. 2017, A&A, 602, A49
Hunt, L., Magrini, L., Galli, D., et al. 2012, MNRAS, 427, 906
Katz, H., Galligan, T. P., Kimm, T., et al. 2019, MNRAS, 487, 5902
Katz, H., Rosdahl, J., Kimm, T., et al. 2022, MNRAS, 510, 5603
Kennicutt, Robert C., J. 1998, ApJ, 498, 541
Kennicutt, R. C. & Evans, N. J. 2012, ARA&A, 50, 531
Khatri, P., Porciani, C., Romano-Díaz, E., Seifried, D., & Schäbe, A. 2024, A&A, 688, A194
Klypin, A. A., Trujillo-Gomez, S., & Primack, J. 2011, ApJ, 740, 102
Knollmann, S. R. & Knebe, A. 2009, ApJS, 182, 608
Lagache, G., Cousin, M., & Chatzikos, M. 2018, A&A, 609, A130
Lagos, C. d. P., Crain, R. A., Schaye, J., et al. 2015, MNRAS, 452, 3815
Lagos, C. d. P., Theuns, T., Schaye, J., et al. 2016, MNRAS, 459, 2632
Lambert, T. S., Posses, A., Aravena, M., et al. 2023, MNRAS, 518, 3183
Langer, W. D. & Pineda, J. L. 2015, A&A, 580, A5
Lara-López, M. A., Cepa, J., Bongiovanni, A., et al. 2010, A&A, 521, L53
Le Fèvre, O., Béthermin, M., Faisst, A., et al. 2020, A&A, 643, A1
Leroy, A. K., Walter, F., Brinks, E., et al. 2008, AJ, 136, 2782
Leung, T. K. D., Olsen, K. P., Somerville, R. S., et al. 2020, ApJ, 905, 102
Loiacono, F., Decarli, R., Gruppioni, C., et al. 2021, A&A, 646, A76
Lomaeva, M., De Looze, I., Saintonge, A., & Decleir, M. 2022, MNRAS, 517, 3763
Lupi, A., Bovino, S., Capelo, P. R., Volonteri, M., & Silk, J. 2018, MNRAS, 474, 2884
Madau, P. & Dickinson, M. 2014, ARA&A, 52, 415
Madden, S. C., Cormier, D., Hony, S., et al. 2020, A&A, 643, A141
Madden, S. C., Rémy-Ruyer, A., Galametz, M., et al. 2013, PASP, 125, 600
Magnelli, B., Boogaard, L., Decarli, R., et al. 2020, ApJ, 892, 66
Malhotra, S., Kaufman, M. J., Hollenbach, D., et al. 2001, ApJ, 561, 766
Mannucci, F., Cresci, G., Maiolino, R., Marconi, A., & Gnerucci, A. 2010, MNRAS, 408, 2115
Matthee, J., Sobral, D., Boogaard, L. A., et al. 2019, ApJ, 881, 124
Muñoz-Elgueta, N., Arrigoni Battaia, F., Kauffmann, G., et al. 2024, arXiv e-prints, arXiv:2407.17359
Nelson, R. P. & Langer, W. D. 1999, ApJ, 524, 923, (NL99)
Olsen, K., Greve, T. R., Narayanan, D., et al. 2017, ApJ, 846, 105
Olsen, K. P., Greve, T. R., Brinch, C., et al. 2016, MNRAS, 457, 3306
Olsen, K. P., Greve, T. R., Narayanan, D., et al. 2015, ApJ, 814, 76
Pallottini, A., Ferrara, A., Gallerani, S., Salvadori, S., & D'Odorico, V. 2014, MNRAS, 440, 2498
Pallottini, A., Ferrara, A., Gallerani, S., et al. 2017, MNRAS, 465, 2540
Pineda, J. L., Langer, W. D., & Goldsmith, P. F. 2014, A&A, 570, A121
Planck Collaboration VI. 2020, A&A, 641, A6
Popping, G., Behroozi, P. S., & Peeples, M. S. 2015, MNRAS, 449, 477
Popping, G., Narayanan, D., Somerville, R. S., Faisst, A. L., & Krumholz, M. R. 2019, MNRAS, 482, 4906
Popping, G., van Kampen, E., Decarli, R., et al. 2016, MNRAS, 461, 93
Posses, A., Aravena, M., González-López, J., et al. 2024, arXiv e-prints, arXiv:2403.03379
Prada, F., Klypin, A. A., Cuesta, A. J., Betancort-Rijo, J. E., & Primack, J. 2012, MNRAS, 423, 3018
Riechers, D. A., Pavesi, R., Sharon, C. E., et al. 2019, ApJ, 872, 7
Rowland, L. E., Hodge, J., Bouwens, R., et al. 2024, MNRAS[arXiv:2405.06025]
Roy, A. & Lapi, A. 2024, arXiv e-prints, arXiv:2407.19007
Sanders, R. L., Shapley, A. E., Kriek, M., et al. 2015, ApJ, 799, 138
Schaerer, D., Ginolfi, M., Béthermin, M., et al. 2020, A&A, 643, A3
Scoville, N., Lee, N., Vanden Bout, P., et al. 2017, ApJ, 837, 150
Somerville, R. S., Lee, K., Ferguson, H. C., et al. 2004, ApJ, 600, L171
Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. 2002, Journal of the Royal Statistical Society Series B: Statistical Methodology, 64, 583
Springel, V., Wang, J., Vogelsberger, M., et al. 2008, MNRAS, 391, 1685
Stacey, G. J., Hailey-Dunsheath, S., Ferkinhoff, C., et al. 2010, ApJ, 724, 957
Swinbank, A. M., Karim, A., Smail, I., et al. 2012, MNRAS, 427, 1066
Szomoru, D., Franx, M., van Dokkum, P. G., et al. 2013, ApJ, 763, 73
Tajiri, Y. & Umemura, M. 1998, ApJ, 502, 59
Teyssier, R. 2002, A&A, 385, 337
Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, ApJ, 613, 898
Trenti, M. & Stiavelli, M. 2008, ApJ, 676, 767
Vallini, L., Gallerani, S., Ferrara, A., Pallottini, A., & Yue, B. 2015, ApJ, 813, 36
Vizgan, D., Greve, T. R., Olsen, K. P., et al. 2022, ApJ, 929, 92
Walter, F., Carilli, C., Neeleman, M., et al. 2020, ApJ, 902, 111
Wolfire, M. G., Hollenbach, D., & McKee, C. F. 2010, ApJ, 716, 1191
Yan, L., Sajina, A., Loiacono, F., et al. 2020, ApJ, 905, 147
Yang, X., Mo, H. J., & van den Bosch, F. C. 2003, MNRAS, 339, 1057
Zanella, A., Daddi, E., Magdis, G., et al. 2018, MNRAS, 481, 1976

Appendix A: Modelling [C II] emission

For a two-level system such as the C^+ fine-structure transition, the excitation temperature ($T_{\rm ex}$) captures the relative population in the upper (u) and lower (l) energy levels of the transition:

$$\frac{n_u}{n_l} = \frac{g_u}{g_l}\, e^{-T_*/T_{\rm ex}}\,. \tag{A.1}$$

Here g_u and g_l are the statistical weights of the upper and lower levels with an energy difference of $k_{\rm B}T_*$ ($k_{\rm B}$ being the Boltzmann constant), and $n_u + n_l = n_{\rm C^+}$. In statistical equilibrium, level populations are determined by the balance between excitation and deexcitation processes:

$$n_l(B_{lu}U + C_{lu}) = n_u(A_{ul} + B_{ul}U + C_{ul}), \tag{A.2}$$

where U is the energy density at the transition frequency ν; A_{ul}, B_{ul} and B_{lu} are the Einstein coefficients for spontaneous decay, stimulated decay, and stimulated excitation, respectively, C_{ul} and C_{lu} are the collision deexcitation and excitation rates and, in the case of multiple collision partners, can be obtained from the respective collision rate coefficients and number densities as:

$$C_{ul} = \sum_{i=1}^{N} R_{ul,\,i}\, n_i\,. \tag{A.3}$$

The upward and downward collision rate coefficients are related as:

$$R_{lu} = R_{ul}\frac{g_u}{g_l} e^{-T_*/T_{\rm kin}}, \tag{A.4}$$

where $T_{\rm kin}$ is the kinetic temperature related to the thermal motions of the collision partner. The collision rate coefficients are taken from Goldsmith et al. (2012):

$$R_{ul}(e^-) = 8.7\times10^{-8}(T_e/2000)^{-0.37}\,{\rm cm^3\,s^{-1}}; \tag{A.5}$$

$$R_{ul}({\rm H}) = 7.6\times10^{-10}(T_{\rm kin}/100)^{0.14}\,{\rm cm^3\,s^{-1}}; \tag{A.6}$$

$$R_{ul}({\rm H_2}) = 3.8\times10^{-10}(T_{\rm kin}/100)^{0.14}\,{\rm cm^3\,s^{-1}}\,, \tag{A.7}$$

where T_e is the electron temperature. In the following, we assume $T_e = T_{\rm kin}$. We obtain the kinetic temperature at each sub-grid density using the temperature density relation (from Hu et al. 2021) and is identical to the one adopted in HYACINTH (see Khatri et al. 2024, for details). The C^+ and H_2 number densities are obtained directly from the simulations. To obtain the number density of atomic hydrogen (H), we assume that gas at densities $n_{\rm H,tot} \gtrsim 0.013\,{\rm cm^{-3}}$ is well-shielded and $n_{\rm H^+} = 0$ (Tajiri & Umemura 1998); below these densities, we assume all hydrogen to be ionised, i.e., $n_{\rm H^+} = n_{\rm H,tot}$. The electron number density follows from charge conservation i.e., $n_{\rm e^-} = n_{\rm C^+} + n_{\rm H^+}$. Note that, similar to Gong et al. (2012) and Vallini et al. (2015), we do not consider the pumping effect from the soft UV background from stars at 1330 Å.

Now, suppose that a given grid cell of sidelength L can be split into N plane-parallel slices as shown in Fig. A.1. In the following, we take $N = 3$ for simplicity, but in practice, use 20 slices in each slice, which were sufficient to reach convergence. The width and density of each slice are determined by the sub-grid density PDF (same as in HYACINTH) within the cell. If $T_{{\rm ex},i}$ is the excitation temperature in slice i, then the energy density generated in the slice at the transition frequency ν can then be written as

$$U_\nu(T_{{\rm ex},i}) = \frac{4\pi}{c}\,B_\nu(T_{{\rm ex},i}) = \frac{8\pi h\nu^3}{c^3}\,\frac{1}{\exp(h\nu/k_{\rm B}T_{{\rm ex},i})-1}\,. \tag{A.8}$$

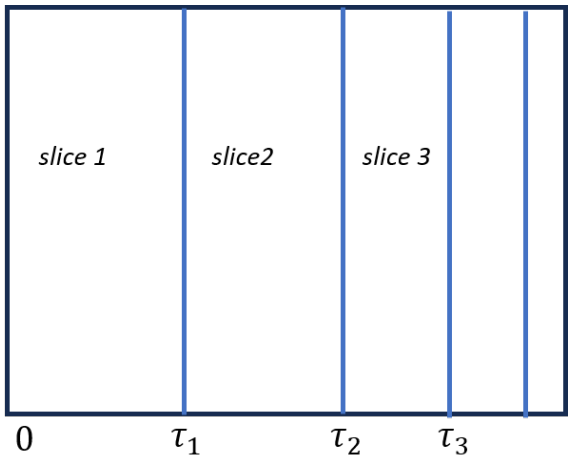


Fig. A.1: A schematic of the plane-parallel slices in a grid cell.

Assuming the entire region is permeated by a background like the CMB at temperature $T_{\rm bg}$ with energy density $U_\nu(T_{\rm bg})$, the total energy density in a slice will have three contributions:

1. the energy density generated in the slice weighted by the fraction of the energy that does not escape the slice (self-absorption, denoted by κ_{ii});
2. the energy density from all other slices where the energy density of the emitting slice i is weighted by the fraction that is absorbed by the absorbing slice j (denoted by κ_{ij});
3. the energy density from the background (CMB in this case).

Following Goldsmith et al. (2012), the total energy density in the slice at the [C II] frequency can be written as:

$$\begin{aligned} U_1 = {} & (1-\kappa_{11})\,U_\nu(T_{\rm bg}) + \kappa_{11}\,U_\nu(T_{{\rm ex},1}) \\ & + \kappa_{21}\,U_\nu(T_{{\rm ex},2}) + \kappa_{31}\,U_\nu(T_{{\rm ex},3})\,. \end{aligned} \tag{A.9}$$

Similarly,

$$\begin{aligned} U_2 = {} & (1-\kappa_{22})\,U_\nu(T_{\rm bg}) + \kappa_{12}\,U_\nu(T_{{\rm ex},1}) \\ & + \kappa_{22}\,U_\nu(T_{{\rm ex},2}) + \kappa_{32}\,U_\nu(T_{{\rm ex},3})\,; \end{aligned} \tag{A.10}$$

and

$$\begin{aligned} U_3 = {} & (1-\kappa_{33})\,U_\nu(T_{\rm bg}) + \kappa_{13}\,U_\nu(T_{{\rm ex},1}) \\ & + \kappa_{23}\,U_\nu(T_{{\rm ex},2}) + \kappa_{33}\,U_\nu(T_{{\rm ex},3})\,. \end{aligned} \tag{A.11}$$

If U_1, U_2, U_3 are known, they can be used to evaluate the level population in each slice by balancing the upward and downward transitions as (e.g., for slice 1):

$$n_{u,1}\,(A_{ul} + B_{ul}U_1 + C_{ul}) = n_{l,1}\,(B_{lu}U_1 + C_{lu})\,, \tag{A.12}$$

and the excitation temperature $T_{{\rm ex},1}$ can be obtained using Eq. (A.1). Following Goldsmith et al. (2012), the optical depth for slice 1 can be calculated from the excitation temperature $T_{{\rm ex},1}$ as:

$$\begin{aligned} \tau_1 &= \frac{hB_{lu}N({\rm C^+})}{\delta v}\,\frac{1-e^{-h\nu/k_{\rm B}T_{{\rm ex},1}}}{1+(g_u/g_l)\,e^{-h\nu/k_{\rm B}T_{{\rm ex},1}}} \\ &= \frac{g_u}{g_l}\,\frac{A_{ul}\lambda_{ul}^3 N({\rm C^+})}{8\pi\sqrt{8\,ln(2)}\,\sigma_v}\,\frac{1-e^{-h\nu/k_{\rm B}T_{{\rm ex},1}}}{1+(g_u/g_l)\,e^{-h\nu/k_{\rm B}T_{{\rm ex},1}}}\,, \end{aligned} \tag{A.13}$$

where λ_{ul} is the wavelength of the [C II] line and $N(C^+)$ denotes the column density of C^+ from the edge of the cell to the slice i. The above expression approximates the line profile function ϕ_ν at line center by δv^{-1}. where δv is the line width ($\delta_v = \sqrt{8\ \ln 2}\,\sigma_v$ for a Gaussian velocity distribution with 1-D velocity dispersion σ_v). We obtain the 1-D velocity dispersion σ_v for the cells in our simulations following the approach of Olsen et al. (2015): the velocity dispersion in a giant molecular cloud (GMC) of mass $M_{\rm GMC}$ and radius $R_{\rm GMC}$ is given as

$$\sigma_v = 1.2 {\rm km\,s^{-1}} \left(\frac{R_{\rm GMC}}{\rm pc}\right)^{-1/2} \left(\frac{M_{\rm GMC}}{290\,{\rm M_\odot}}\right)^{1/2}, \tag{A.14}$$

where we approximate $R_{\rm GMC} \approx \frac{1}{2}\Delta L$ and $M_{\rm GMC} = M_{\rm gas,\,cell}$.

The system of equations (A.1), (A.9)-(A.13) can be solved iteratively. We start with the optically thin case where all emitted radiation freely escapes (i.e., $\kappa_{ij} = 0\ \forall\ \{i, j\}$), and obtain an initial estimate of $n_{u,i}$ using Eq. (A.12). These estimates are then used to obtain a first estimate of $T_{\rm ex,1}, T_{\rm ex,2}, T_{\rm ex,3}$ using Eq. (A.1). These determine the optical depths τ_1, τ_2, τ_3 using Eq. (A.13). Using these, the κ values can be updated and the entire procedure is repeated until convergence[9].

Once we have a set of self-consistent $T_{\rm ex}$ and κ values, we can calculate the emissivity of each slice. Following Goldsmith et al. (2012), the emissivity in slice 1 can be written as

$$\epsilon_1 = n_{u,1}\, A_{ul}\, \gamma_1\, h\,\nu \left[1 - \frac{e^{(h\nu/k_{\rm B}T_{\rm ex,1})} - 1}{e^{(h\nu/k_{\rm B}T_{\rm bg})} - 1}\right], \tag{A.15}$$

where

$$\gamma_i = 1.0 - \sum_{j=1}^{3} \kappa_{ij} \tag{A.16}$$

denotes the final escape fraction for slice i, i.e., the fraction of photons emitted in slice i that manage to escape the cell and accounts for absorption from all intervening slices.

The total luminosity from the cell can be written as:

$$L_{\rm [C\,II]} = \sum_{i=1}^{N} \gamma_i\, \epsilon_i\, \Delta V_i\,, \tag{A.17}$$

where $\Delta V_i = \mathcal{P}_V(n_i)\,\Delta n_i\, V_{\rm cell}$ is the volume of each slice, and N is the number of slices in the cell. The galaxy luminosity is obtained by summing over the [C II] emission from all cells within the galaxy. We note that our method assumes that the [C II] emission from the different grid cells are radiatively decoupled and the total [C II] emission from a galaxy is the sum of the emission from each cell.

Appendix B: Model Validation

We validate our [C II] emission model by comparing its output with the photoionization code Cloudy (version 17.02; Ferland et al. 1992). To do this, we compute the luminosity emerging from a plane-parallel slab with a side length of $\Delta x = 38$pc, corresponding to the minimum spatial resolution achieved in our `M25` run at $z = 4$. The CMB is included for $z = 4$.

To span the range of gas densities and metallicities exhibited by PDRs and molecular clouds, we compute the luminosity for a 2D grid with hydrogen number density $n_{\rm H} \in [10^0, 10^4]\,{\rm cm^{-3}}$ and gas metallicity $Z_{\rm gas} \in [10^{-1.8}, 10^0]\,{\rm Z_\odot}$. The slabs have a uniform density, metallicity and temperature. We assume that the kinetic temperature is uniform throughout the slab. For each set of ($n_{\rm H}$, $Z_{\rm gas}$) values, we obtain the temperature using the metallicity-dependent temperature density relation from Hu et al. (2021, same as in HYACINTH). We assume that the elemental abundance of carbon $f_{\rm C,tot}$ scales as $f_{\rm C,tot} = 2.9\times10^{-4} Z_{\rm gas}/{\rm Z_\odot}$ (Asplund et al. 2009). For each of these models, Cloudy computes the abundances of different chemical species as a function of the depth into the slab. The C^+, H_2, and H I abundances are used as inputs to our model compute the emergent luminosity from the model cells. We further assume $n_{e^-} = n_{C^+}$.

The results of the comparison are shown in Fig. B.1. In the following, we denote the [C II] luminosity predicted by our model as L and the one from Cloudy as $L_{\rm Cloudy}$. At all values of $T_{\rm kin}$, the distribution of L is very similar to $L_{\rm Cloudy}$, with L from both approaches increasing with density at a fixed metallicity. Conversely, at fixed densities $n_{\rm H} \lesssim 10^3\,{\rm cm^{-3}}$, L increases with metallicity and the variation with metallicity is minimal at higher densities. At $\log(Z_{\rm gas}\,/\,{\rm Z_\odot}) \gtrsim 10^{-0.6}$ (i.e., $Z_{\rm gas} \gtrsim 0.25\,{\rm Z_\odot}$), our model overpredicts $L_{\rm [C\,II]}$ at low densities ($n_{\rm H} \lesssim 10^2\,{\rm cm^{-3}}$) and underpredicts at intermediate densities ($10^2\,{\rm cm^{-3}} \lesssim {\rm n_H} \lesssim 10^3\,{\rm cm^{-3}}$). In most of the parameter space there is a $\leq 30\%$ deviation between the $L_{\rm [C\,II]}$ from the two approaches, particularly at high densities ($n_{\rm H} \gtrsim 10^3\,{\rm cm^{-3}}$).

Appendix C: Fit to the simulated luminosity function

Fig. C.1 shows a comparison between the Schechter function and double power-law fits to the predicted [C II] LFs from our simulations at $3 \leq z \leq 7$.

In Fig. C.2, we show the covariance distributions obtained from the Markov Chain Monte Carlo (MCMC) fitting of our predicted [C II] luminosity function at $z = 4$ with a DPL given in Eq. 2. The fitting was performed using the python package `emcee` and includes two parameters – $\Delta_{\rm M25}$ and $\Delta_{\rm M50}$ that represent variation of the log ϕ of the two simulation volumes from the cosmic log $\phi*$, because of sample variance. We see that the posterior distributions of the all parameters are well-behaved and smooth, and that the three parameters are highly correlated. We obtain similar results at other redshifts.

Appendix D: Surface densities

In Fig. D.1, we show a scatter plot of the surface densities of CO and C^+ as a function of the SFR (left), gas (middle), and metal (right) surface density. In all panels, we find that while $\Sigma_{\rm CO}$ continues to increase at high surface densities, $\Sigma_{\rm C^+}$ shows a plateau. This results in a decrease in the slope of the log $\Sigma_{\rm [C\,II]}$ versus log $\Sigma_{\rm SFR}$ curve at high $\Sigma_{\rm SFR}$.

[9] Here convergence is defined as the point when the difference between the norm of κ matrices of successive iterations is $\leq 10^{-4}$.

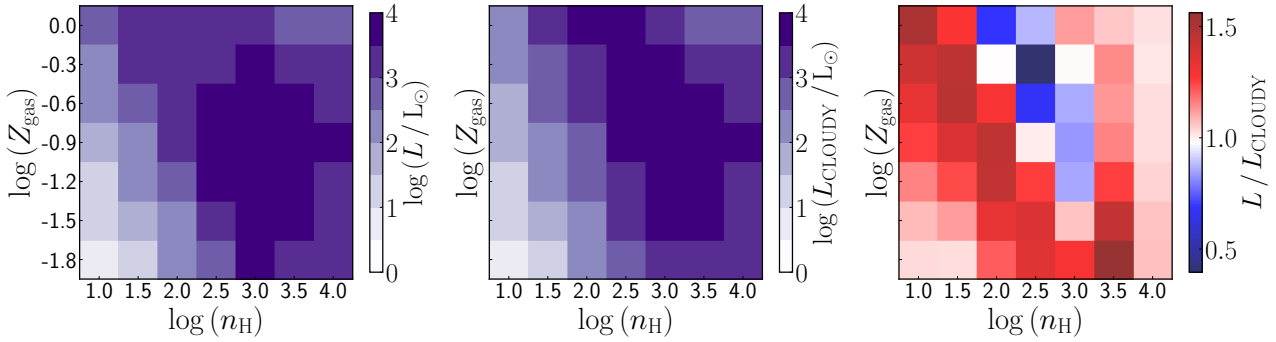


Fig. B.1: Tests of [C II] emission model with CLOUDY. The left and middle panels show, respectively, the luminosity from our model and CLOUDY, while the right panel shows the ratio of the two. For each set of ($n_{\rm H}$, $Z_{\rm gas}$) values, the temperature is obtained using the temperature-density relation from Hu et al. (2021).

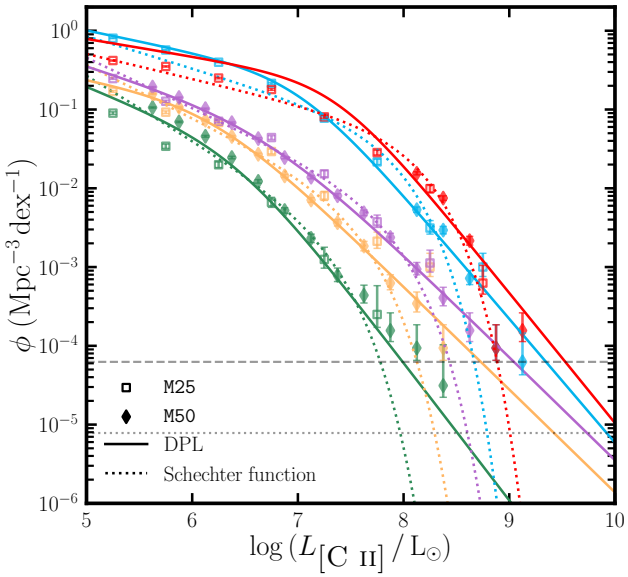


Fig. C.1: Comparison of the Schechter function (dotted lines) and DPL (solid lines) fits to our simulated LFs from the two simulations at different redshifts (from top to bottom, z increases from 3 to 7). The open squares and diamonds represent the separate LFs from the simulations that are used to obtain the fit parameters using an MCMC analysis. The error bars represent the 16 CL upper and lower Poisson uncertainties on number counts (Gehrels 1986). The dashed and dotted horizontal lines represent a number count of 1 per dex in the entire simulation volume of `M25` and `M50`, respectively.

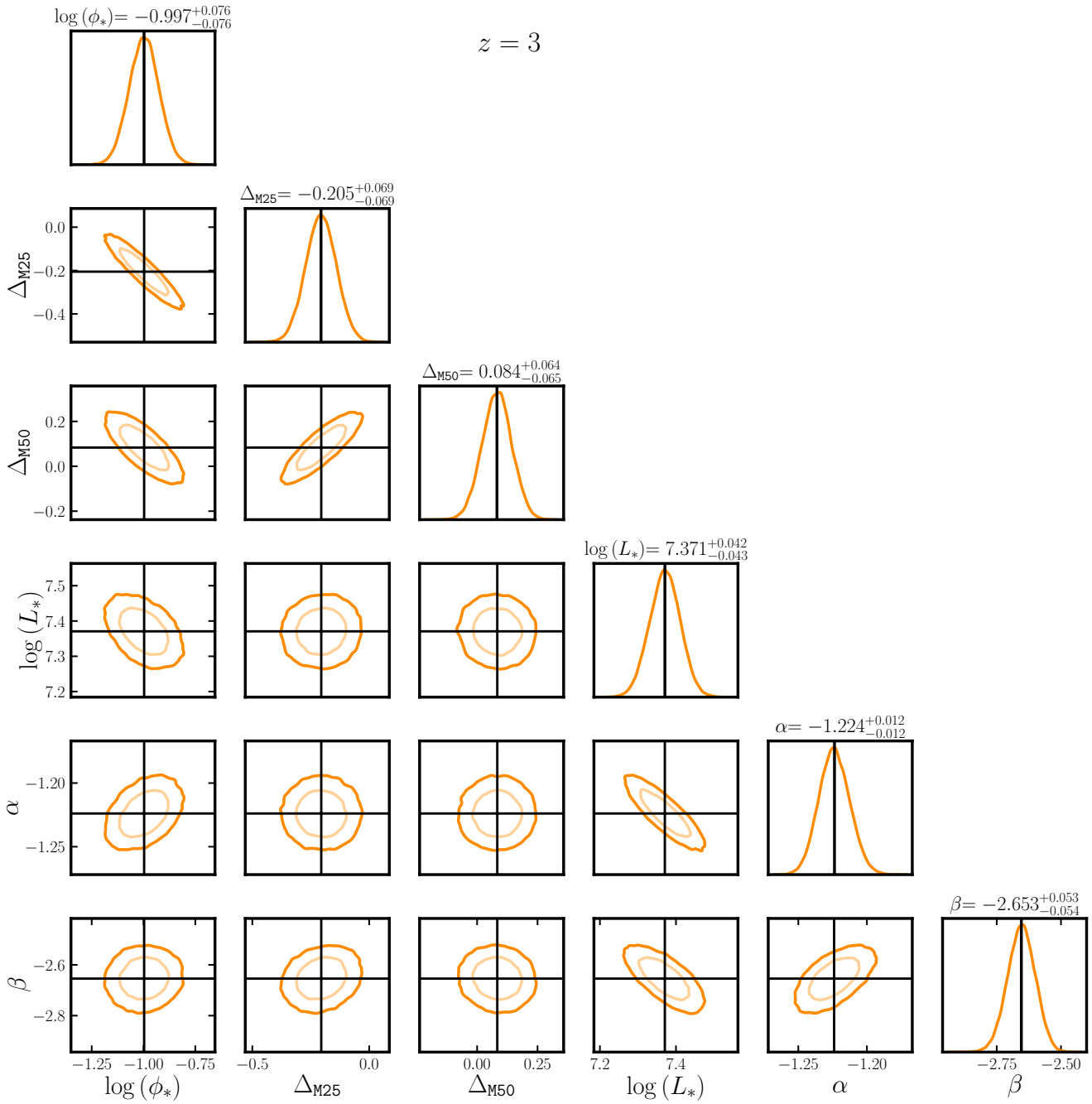


Fig. C.2: Covariance distributions and PDFs of the double power-law function (Eq. 2) parameters: ϕ_*, L_*, α, β, and two additional parameters: $\Delta_{\mathtt{M25}}$ and $\Delta_{\mathtt{M50}}$, from MCMC runs using the python package `emcee` for simulated LF at $z = 3$.

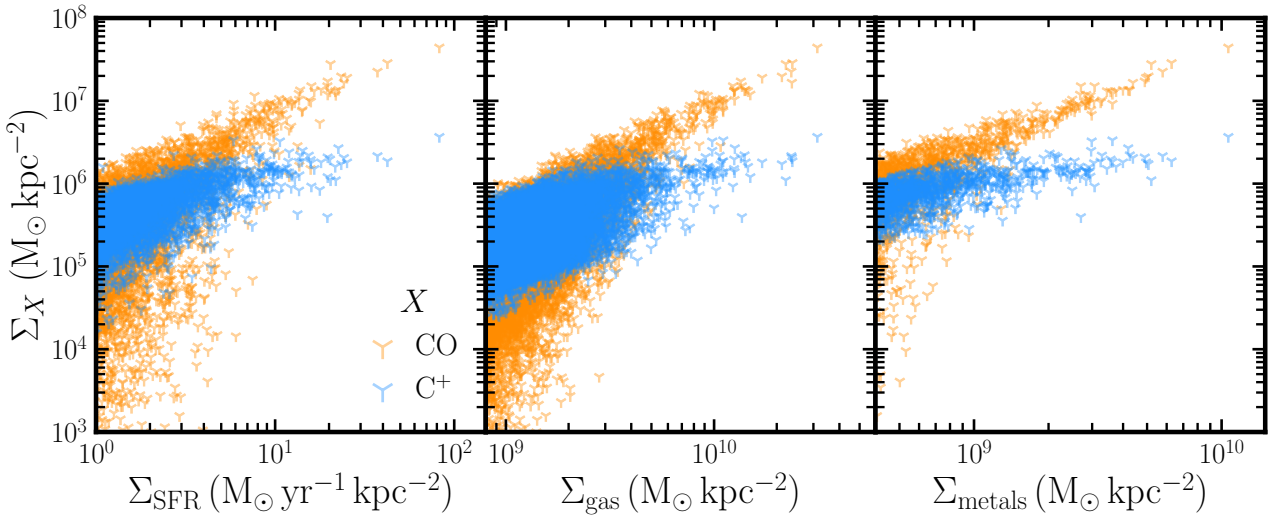
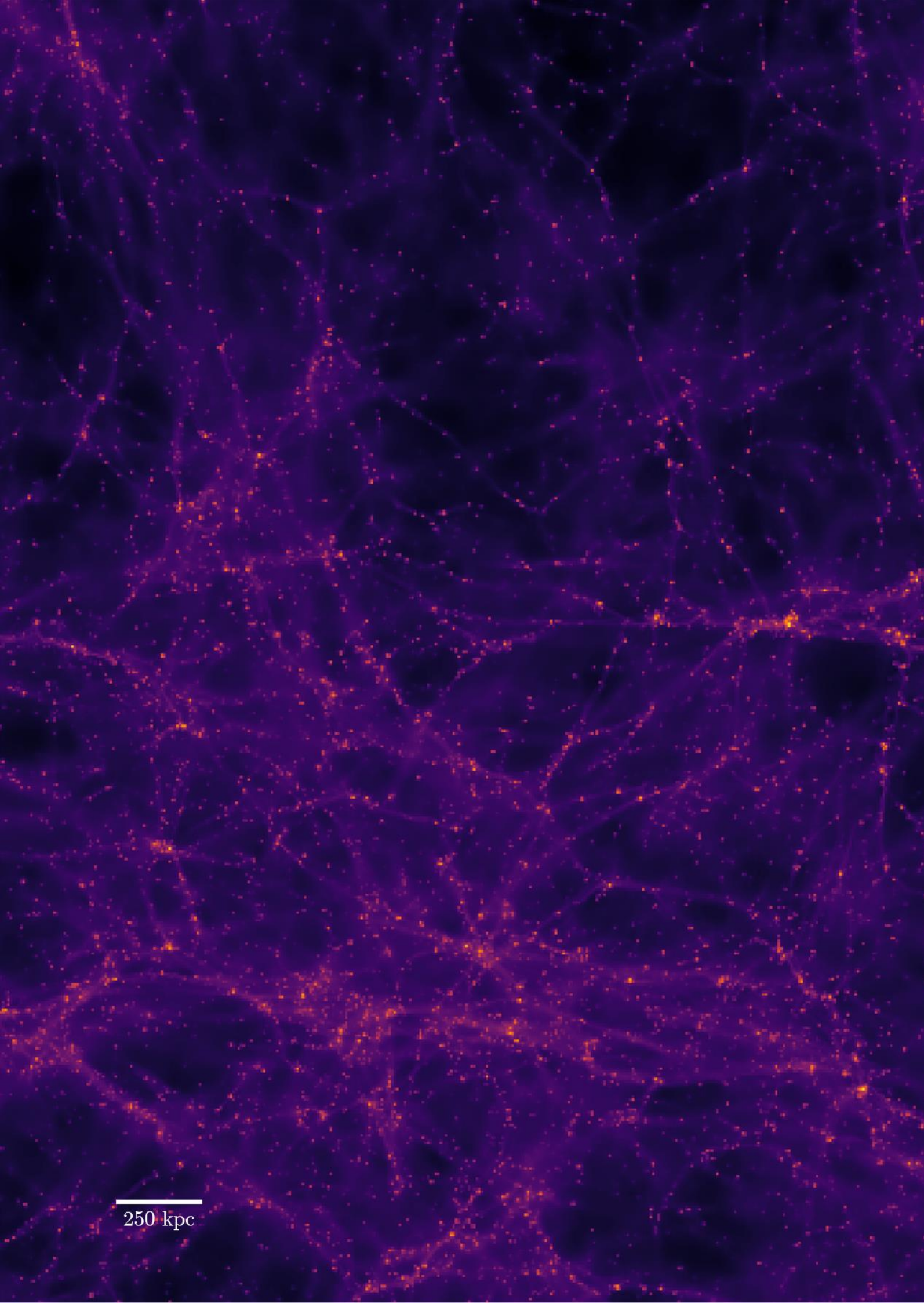


Fig. D.1: The surface density of CO and C^+ as a function of the SFR surface density (left), the total gas surface density (middle), and the total metal surface density (right), for galaxies used in Fig. 6.

List of Figures

List of Tables


250 kpc