

3D chemical structure of the turbulent interstellar medium

Elena Bellomi

▶ To cite this version:

Elena Bellomi. 3D chemical structure of the turbulent interstellar medium. Astrophysics [astro-ph]. Université Paris sciences et lettres, 2021. English. NNT: 2021UPSLO001. tel-03273709

HAL Id: tel-03273709 https://theses.hal.science/tel-03273709

Submitted on 29 Jun 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE DE DOCTORAT DE L'UNIVERSITÉ PSL

Préparée à Observatoire de Paris

3D Chemical Structure of the Turbulent Interstellar Medium

Soutenue par Elena Bellomi Le 12 Février 2021

École doctorale nº127

Astronomie Astrophysique d'Ile de France

Spécialité

Astronomie Astrophysique

Composition du jury :

Rosine LALLEMENT DR, GEPI

Cécile GRY Astronome, LAM

Pierre HILY-BLANT Maitre de conférence, IPAG

Marc-Antoine MIVILLE-DESCHENES DR, CEA

Chiara FERRARI Astronome, OCA

Michel PERAULT DR, LPENS

Benjamin GODARD Astronome-Adjoint, LPENS

Patrick HENNEBELLE DR, CEA

Président

Rapporteur

Rapporteur

Examinateur

Examinateur

Directeur de thèse

Codirecteur de thèse, Invité

Codirecteur de thèse, Invité



Acknowledgements

First of all, I would like to thank the jury members, Dr. Rosine Lallement, Dr. Chiara Ferrari, Dr. Cécile Gry, Dr. Pierre Hily-Blant, Dr. Marc-Antoine Miville-Deschenes, for taking the time to read my manuscript and for the interesting discussion we had around it. I am deeply grateful to all the people that contributed to improving this manuscript with their corrections and words of advice. In particular, my Ph.D. co-supervisor Dr. Benjamin Godard for providing guidance and feedback throughout these last three and a half years. I would like to express my sincere gratitude to Dr. Michel Perault and Dr. Patrick Hennebelle for their extensive knowledge and invaluable advice.

I am grateful to the Astro-team at Ecole Normale Superieure de Paris, in particular to the post-docs for their insightful comments and suggestions. Andrea, Alba, and Andrew, you have always been there to assist me and support me when I needed it, which sometimes meant to force me to take a coffee break or go one hour to the gym. In the AstroENS, a special thank you goes to my lovely Céline, not only because of your great work as a secretary that made my administrative life very easy for a while but also to you as a person and friend.

Many others helped me along the way on this journey. I want to take a moment to thank some of them in their languages.

Merci à mes potes. Vous êtes beaucoup et formidables. Merci aux "lambdalcooliques", aux "dîners entre adultes", aux soirées choupinettes, aux pauses capus de 16h avec Cip et Ciop, aux fous rires. Merci pour les brunchs du dimanche, qui ont été rares mais mémorables. Merci pour les journées au parc pour profiter des rayons de soleil que Paris a à offrir.

Merci aux bières entre collègues et en particulier à Antoine qui a toujours su me soutenir et réconforter avec un bon verre de vin quand c'était nécessaire.

Merci à mes colocs, surtout à Marylou, à ces super-bons plats, à nos soirées jazz et à nos tisanes du soir.

Merci à mes potes astro à qui j'ai posé plein de questions, surtout au bar autour d'un verre pendant ces années.

Merci à mes potes skieurs/snowboarders, Lisa, Flo, Amaël, Manu, Yassine, Pierre, Simon, et Ted, avec qui dans les années on a passé des super aventures dans les montagnes enneigées.

Merci aux "adultes". À Flo avec qui on a découvert Paris en M1; à Max avec ces petits plats plus que délicieux; à Sohvi pour nos pauses-café à l'ENS qui nous aidaient à désamorcer notre rage contre les machines; et à Agathe, je ne sais même plus quand tu es arrivée dans le groupe, mais ce groupe serait beaucoup moins fun sans toi.

Merci à mon Yaya (amazing human/genius), sans qui j'aurais déjà pété un câble en M2 et j'aurais du vivre sous un pont pendant un moment, merci pour nos petits plats, pour les fous rires dès le réveil, mais surtout pour tout le faux drama.

Merci à MassMass, qui a toujours amélioré ma vie pendant ces années, il suffit d'un verre sur les quais avec toi pour que tout paraisse plus beau et facile, merci aussi de m'avoir hébergée et nourrie les dernières semaines avant la soutenance!

La liste de personnes à remercier est très très longue, surtout si on tient compte du fait que le doctorat n'est pas qu'un projet de 3 ans mais l'aboutissement d'un parcours qui a commencé des années auparavant. Je tiens donc à remercier toutes les personnes qui ont contribué à ma formation et à mon bien-être dans les dernières années.

E ora si cambia ancora lingua e pure registro! Beh, almeno parte di sta tesi sarà scritta in italiano!! Un immenso grazie a tutti quelli che mi hanno nutrito, e siete pochi (ma più di quanti mi aspettassi) ma mooolto buoni! Mi avete nutrito quando io giocavo a lasciarmi morire di fame alla Newton e non fosse stato per voi avrei mangiato solo yogurt e ogni tanto delle scatole di noodle precotte, quelle disgustose che la mia famiglia pensava esistessero solo nei film americani.

Grazie a tutte quelle persone che mi erano vicine anche essendomi lontane, diciamo anche a centinaia e migliaia di chilometri. Chi in Italia e chi dall'altra parte dell'oceano.

Grazie alla mia piattolina, ti ho conosciuta che non sapevo neanche ancora se avrei vinto una borsa di dottorato, e ora sono passati anni e noi siamo ancora vicine ma lontane, ma ancora per poco, perché in un batter d'occhio sarò lì a strapazzarti di coccole.

Grazie ad Alessia, conoscerti nel mezzo della mia tesi è stata una vera fortuna (ancora una volta grazie a Vero). Grazie per le nottate di chiacchiere, per le giostre più o meno adrenaliniche, mi hai aiutato a passare dei mesi molto belli.

Grazie a Chiara, perché la capitale parigina va vissuta comunque con un po' di italianità ogni tanto, e senza le nostre lunghe discussioni e le nostre mattinate in pigiama, questa esperienza sarebbe stata molto diversa e più triste.

Ai miei genitori, che sono stati fantastici come sempre, lontani o vicini sempre presenti, sempre positivi e sempre sostenitori. E il sostegno che mi hanno dato inizia ben prima dell'inizio di questa tesi di dottorato, inizia forse prima della mia nascita. Grazie per aver creduto in me anche quando io non ci credevo.

Grazie al mio amore, Dario, sei stato al top, specialmente durante il duro periodo della redazione, in cui lavoravo 80 ore a settimana e tu mi hai coccolato, mi hai sempre fatto trovare la tavola imbandita. Siamo sinceri, non ci fossi stato tu al mio fianco, sarei diventata un orso che comunicava solo più tramite versi, e invece la tua presenza mi ha mantenuto serena e sorridente. Sei speciale, e sono stata proprio fortunata ad incontrarti.

Grazie poi anche a tutte quelle cose che di solito non si ringraziano, ma sono invece importantissime. Grazie alla musica jazz, non esistesse io sarei già in un manicomio. Grazie al cioccolato, la birra e il tappetino di yoga, che sono riusciti separatamente a darmi un grande sostegno, soprattutto nel lockdown.. ehm lockdowns... Grazie alle passeggiate nei prati, nei boschi, tra le foglie, ma soprattutto nella neve.

Grazie al power thesaurus, sempre pronto a darmi una moltitudine, orda, schiera, montagna, marea, infinità, miriade, plenitudine, varietà (avete capito il concetto!?) di sinonimi. Senza di lui, la scrittura di questa tesi sarebbe stata ancora più dolorosa... e già così non c'era nulla da ridere.

Insomma, le persone e le cose da ringraziare sono tante, la cosa che di sicuro mi ha fatto meglio sono le risate fra amici. Non vi ho citati tutti, ma se siete fra le persone che mi hanno aiutato a rimanere sane negli ultimi anni lo sapete e sapete anche che siete nel mio cuore!

Contents

Ι	Inte	Interstellar Medium				
1 ISM Composition						
1.1 Dust			9			
	1.2	Phases	10			
	1.3	Reservoirs of Energy	11			
		1.3.1 Mechanical Energy and Turbulence	12			
		1.3.2 Magnetic Field	12			
		1.3.3 Radiation Field	12			
		1.3.4 Cosmic Rays	13			
		1.3.5 Equipartition	13			
2	The	ermodynamics	14			
	2.1	Heating Processes	14			
	2.2	Cooling Processes	17			
	2.3	Thermal Instability and Multi-phase Medium	21			
3	Rad	liation - photons	24			
3	Rad 3.1	liation - photons Interstellar Radiation Field	24 24			
3	Rad 3.1 3.2	liation - photons Interstellar Radiation Field	24 24 27			
3	Rad 3.1 3.2 3.3	liation - photons Interstellar Radiation Field	24 24 27 28			
3	Rad 3.1 3.2 3.3 Obs	liation - photons Interstellar Radiation Field Extinction OB Stars servations	 24 24 27 28 30 			
3	Rad 3.1 3.2 3.3 Obs 4.1	liation - photons Interstellar Radiation Field Extinction OB Stars Gervations Atomic and Molecular Hydrogen	 24 24 27 28 30 30 			
3	Rad 3.1 3.2 3.3 Obs 4.1	liation - photons Interstellar Radiation Field Extinction OB Stars Servations Atomic and Molecular Hydrogen	 24 24 27 28 30 30 31 			
3	Rad 3.1 3.2 3.3 Obs 4.1	liation - photons Interstellar Radiation Field Extinction OB Stars OB Stars Atomic and Molecular Hydrogen 4.1.1 Observational Results 4.1.2 Theoretical Interpretation and Open Question	 24 24 27 28 30 30 31 33 			
3	Rad 3.1 3.2 3.3 Obs 4.1	liation - photons Interstellar Radiation Field Extinction OB Stars OB Stars Servations Atomic and Molecular Hydrogen 4.1.1 Observational Results 4.1.2 Theoretical Interpretation and Open Question Neutral Carbon	 24 24 27 28 30 31 33 35 			
3	Rad 3.1 3.2 3.3 Obs 4.1	liation - photons Interstellar Radiation Field Extinction OB Stars OB Stars Atomic and Molecular Hydrogen 4.1.1 Observational Results 4.1.2 Theoretical Interpretation and Open Question Neutral Carbon 4.2.1 Observational Results	 24 24 27 28 30 30 31 33 35 36 			
3	Rad 3.1 3.2 3.3 Obs 4.1	liation - photons Interstellar Radiation Field Extinction OB Stars OB Stars Atomic and Molecular Hydrogen 4.1.1 Observational Results 4.1.2 Theoretical Interpretation and Open Question Neutral Carbon 4.2.1 Observational Results 4.2.2 Theoretical Interpretation and Open Questions	24 24 27 28 30 30 31 33 35 36 39			
3	Rad 3.1 3.2 3.3 Obs 4.1 4.2	liation - photons Interstellar Radiation Field	24 24 27 28 30 30 31 33 35 36 39 39			
3	Rad 3.1 3.2 3.3 Obs 4.1 4.2	liation - photons Interstellar Radiation Field Extinction OB Stars Atomic and Molecular Hydrogen 4.1.1 Observational Results 4.1.2 Theoretical Interpretation and Open Question Neutral Carbon 4.2.1 Observational Results 4.2.2 Theoretical Interpretation and Open Questions 4.3.1 Observational Results	 24 24 27 28 30 30 31 33 35 36 39 39 40 			
3	Rad 3.1 3.2 3.3 Obs 4.1 4.2	liation - photons Interstellar Radiation Field Extinction OB Stars OB Stars Atomic and Molecular Hydrogen 4.1.1 Observational Results 4.1.2 Theoretical Interpretation and Open Question Neutral Carbon 4.2.1 Observational Results 4.2.2 Theoretical Interpretation and Open Questions CH ⁺ 4.3.1 Observational Results 4.3.2 Theoretical Interpretation and Open Questions	24 24 27 28 30 30 31 33 35 36 39 39 40 42			

II Modeling

5	Sim	Simulations: RAMSES					
	5.1	MHD: Conservation Laws	46				
5.2		Thermodynamics	48				
		5.2.1 Radiative Transfer	49				
		5.2.2 Testing Heating and Cooling Processes	50				
	5.3	Turbulence	54				
		5.3.1 Overview on Theory	54				
		5.3.2 Implementation in RAMSES: Turbulent Forcing	55				
	5.4 Molecular Hydrogen						
		5.4.1 Formation of H ₂	56				
		5.4.2 Destruction of H ₂	57				
		5.4.3 H ₂ Shielding	58				
		5.4.4 Doppler parameter	59				
	5.5	Grid of Simulations	60				
		5.5.1 Steady-State	60				
		5.5.2 Parametric Study	60				
		5.5.3 Physical State of the Gas	62				
6 CI Excitation							
	6.1	Electronic Ground State: Fine-Structure Levels	69				
	6.2	CI Excitation	70				
		6.2.1 Radiative and Collisional (De-)Excitation	70				
		6.2.2 Chemical Pumping	74				
		6.2.3 Optical Pumping	74				
	6.3	CI radiative transfer	75				
	6.4	Resulting Line Profiles	75				
7	H_2 l	Excitation	78				
	7.1	H ₂ Levels	78				
	7.2	Rovibrational Excitation of H_2 in its Electronic Ground State	80				
		7.2.1 Radiative Transitions	81				
		7.2.2 Collisional Transition	82				
		7.2.3 Reactive Collisions	83				
		7.2.4 Chemical Pumping	85				
		7.2.5 Population as function of physical conditions	86				
	7.3	Validation of the H ₂ Excitation Code	86				
		7.3.1 Critical Density	87				
		7.3.2 Ortho-Para Ratio	88				
	7.4	Including Electronic Excited Levels	89				
		7.4.1 UV Pumping	89				

III Comparison with Observations

q	4
J	т

8	Rec	onstruction Algo	rithm		96
	8.1 Method				96
8.2 Application: HI-to-H ₂ Transition \ldots				97	
	8.3	Application: $N(0)$	CH^+) vs $N_{\rm H}$		99
	8.4	Tests and Caveat	ts		102
		8.4.1 The Rand	om Drawing Process		102
		8.4.2 The Choic	ce of Concatenated Patches		105
		8.4.3 Assumption	ons on the ISM Structure around the Sun		106
9	Kol	mogorov-Smirno	v Statistics		109
	9.1	Classic KS test .			109
	9.2	The Modified Ve	rsion of the KS Test		110
		9.2.1 Definition	and Interpretation		110
		9.2.2 Error on t	he Merit Function		111
		9.2.3 Influence	of the Distribution of Datapoints		112
	9.3	Application: HI-	to-H ₂		113
		9.3.1 Choices o	of the Axes and $n_{\rm thr}$		113
		9.3.2 Results .	····		114
	9.4	Application on C	CH^+		117
		9.4.1 Choice of	the Axes Orientation and $n_{\rm thr}$		117
		9.4.2 Results .			117
	9.5	Extension of KS	test to Non-Detections		118
		9.5.1 Applicatio	on on CH^+		119
		9.5.2 Different	Assumptions to count the contribution in a quadrant of	of a non-	
		detection			122
		9.5.3 Conclusio	ons		125
10	Ato	mic-to-Molecular	Transition		126
	10.1	H ₂ Self-shielding	g at High Temperature		126
		10.1.1 Impact of	the Inclusion of the Upperlimits in the KS Test		127
11	The	Elusive CH ⁺			163
	11.1	Chemical Consid	deration		163
11.2 Impact of Warm-H ₂					164
	11.3	Exploration of th	Parameters		167
	11.4	Conclusion			175
12	Exc	itation of Neutral	Carbon		176
	12.1	Influence of the	Parameters		178
	12.2	Comparison wit	h Observations		179
	12.3	Conclusions			182

IV	Conclusion and Perspectives	185
V	Appendix	189
13	Postprocessing: Chemistry 13.1 Reactions	190 191
14	Cosmic-Rays Ionization Rate	193
15	Comparison of Distances derived from Hipparcos and Gaia	195

Part I

Interstellar Medium



We look up at the night sky and, if we are lucky enough to be far from the cities' light pollution, we can observe a myriad of bright dots lighting up the sky's darkness. Those white, blue, red luminous points can be stars in our Galaxy or even entire galaxies, so far away from us that they appear like dots, or again, as the Greeks observed, if the position of these dots moved from one night to another they can be planets of our solar system (Planetes in ancient Greek meaning wanderer). Stars have played a fundamental part in human history. For instance, they helped the first explorers to navigate the seas and oceans, they ignited passionate debates between philosophers throughout the ages, and always made humankind dream. After all, hasn't the first scientific experiment been to observe the sky? Allowing us to grasp this mystery that contains us and inhabits us. As Carl Sagan said, "We are made of star-stuff. We are a way for the Cosmos to know itself". Indeed, in their life cycles, stars produce heavy nuclei in their cores thanks to multiple fusion processes and during supernovae explosions. These cataclysmic cosmic events are the "death" of certain stars. Last but not least, they produce light! Stars usually live in galaxies, where they are not quiet guests, they continuously change their surrounding environment, with their continuous feedback processes, radiation, and winds.

Contrary to human conception, the space separating the stars and filling the Galaxy is diluted but far from empty. This intermediate space called the interstellar medium (ISM) is composed of gas and dust grains (small solid particles mixed with the interstellar gas). The ISM has an average density of about one particle per cubic centimeter. The ISM is a very rich, turbulent, and magnetized medium. Gas and dust are in constant interactions with magnetic fields, cosmic rays, i.e. high energy particles, gravitational perturbations, turbulence, and electromagnetic radiations (photons): together, they govern the dynamics of the ISM. All these constituents have a great impact on the dynamic of the gas. For instance, magnetic pressure works against gravitational collapse in molecular clouds, while magnetic tension helps to confine filaments of the ISM increasing their lifetimes. Turbulence mixes the gas, transporting the energy to smaller and smaller scales, hence, connecting the different physical scales. Photons interact with matter thus heating the medium and changing its chemical content. The different constituents listed above are in almost equipartition of energy density, meaning that, generally speaking, one does not prevail on the others, and none is negligible compared to the others. Moreover, these constituents are all coupled with one another, influenced by each other, making the ISM a complex and highly dynamic medium.

The ISM is the birthplace of stars, their nursery, their home, and eventually their death bed and grave. Molecular clouds, formed from diffuse atomic gas, become increasingly denser subsequently condensing into sheets and filament structures and cores that in time collapse, under the effect of their own gravitational pull, into stars when gravity overcomes thermal and magnetic pressure. Stars will burn their fuel, usually hydrogen and helium at first through thermonuclear fusion. This nuclear process will synthesize heavier elements, and modify the surrounding environment through the injection of energy, momentum, and mass in the ISM. All these injections are called feedback processes, and examples are stellar winds, the ionizing radiation from massive stars, the HII regions surrounding such



Figure 1: Picture taken from https://chandra.harvard.edu/ "NGC 6357, a region where radiation from hot, young stars is energizing the cooler gas in the cloud that surrounds them. This composite image contains X-ray data from NASA's Chandra X-ray Observatory and the ROSAT telescope (purple), infrared data from NASA's Spitzer Space Telescope (orange), and optical data from the SuperCosmos Sky Survey (blue) made by the United Kingdom Infrared Telescope. Located in our galaxy about 5,500 light years from Earth, NGC 6357 is actually a "cluster of clusters," containing at least three clusters of young stars, including many hot, massive, luminous stars. The X-rays from Chandra and ROSAT reveal hundreds of point sources, which are the young stars in NGC 6357, as well as diffuse X-ray emission from hot gas. There are bubbles, or cavities, that have been created by radiation and material blowing away from the surfaces of massive stars, plus supernova explosions." Image credit: X-ray: NASA/CXC/PSU/L. Townsley et al; Optical: UKIRT; Infrared: NASA/JPL-Caltech



Figure 2: This image is taken with the NASA/ESA Hubble Space Telescope. "When a massive new star starts to shine while still within the cool molecular gas cloud from which it formed, its energetic radiation can ionize the cloud's hydrogen and create a large, hot bubble of ionized gas. Amazingly, located within this bubble of hot gas around a nearby massive star are dark compact globules of dust and gas, some of which are giving birth to low-mass stars."

stars, and UV radiation. The end of life of a star will differ depending on its mass. When massive enough they will burst in a supernova explosion thus injecting a large quantity of mechanical energy and heavy nuclei into the surrounding medium.

The interstellar medium interacts greatly with the stars, which are photon emitters. As mentioned before, photons and matter interact in a complicated dance. It is mainly thanks to photon-matter interactions that we get information on the medium in which they travel. Photons are the prime messengers of the Universe surrounding us. Everything we see is light or the absence of light. Our understanding of the ISM, the Galaxy, and even the Cosmos essentially come from photon measurements. However, photons are not the only messengers of our Universe. Neutrinos and more recently gravitational waves take now part in our new era of multi-messenger astrophysics. Indeed neutrinos which are fundamental particles that only interact through weak interaction are the second most abundant particles in the Universe and can, thanks to their properties, give us information on the very early Universe even before recombination times when the Universe stopped being opaque to photons. The same goes for gravitational waves which propagate at the speed of light in the space-time fabric and give us early warning of huge astrophysical phenom-

ena such as black-hole merging. It is important to notice that since almost all information comes from one messenger, and is two-dimensional, the interpretation of observations can be challenging. Many quantities observed are integrated quantities along the line of sight, e.g. column densities, and it is sometimes difficult to link the 3D physics with the 2D data coming from the observations.

A natural way to study the ISM on its whole, is to access to its properties though the observation of the local interstellar medium. Ongoing efforts to observe the local diffuse ISM unveiled a richness of features that characterize the space around us. Observations show a complex structure of matter characterized by dense, and cold, and almost fractal regions surrounded by hot and diffuse gas. These phases, defined by very different densities and temperatures, exchange matter and energy, and coexist because they are approximately in thermal pressure equilibrium. A detailed observation of the ISM, through spectroscopy also revealed a richness in molecules. Molecules not only exist in a medium that is very hostile to their survival because rich in UV photons, but they are also abundant enough to be detected. As we go around the cycle of matter, more and more complex chemical species are found, showing the importance of chemistry as an observable tracer of the evolution of the ISM. All these characteristics seem to be global properties of the local ISM because they appear to be observed in all directions.

These observations of the diffuse local ISM reveal a complex medium and require a strong theoretical counterpart to be explained and interpreted. Usually, the theoretical efforts are divided into two categories: on one side numerical simulations to explain the formation of the multi-scale structures; and on the other chemical models to explain the degree of ionization and the abundances of the different chemical species.

Ordinarily, the chemical solvers are one-dimensional models that solve a large and complex chemical network at steady-state, because it is virtually impossible to compute 3D time-dependent chemistry. An example is PDR codes (photons-dominated regions) which model regions in which the chemistry is driven by the UV photons. These codes usually explain well the abundances of many chemical species, however fail to explain the abundances of some species which are formed through highly endoenergetic reactions. Moreover, there is no coupling between the chemistry and the dynamics, therefore the structure of the medium and its subsequent permeability to photons are not considered. Other models, such as TDR (turbulent dominated regions) rose as attempts to couple chemistry and dynamics at small scales and explain the abundances of species with endothermic formation reactions. TDR codes give good results on the abundances of species whose formation is endothermic, nevertheless, the dynamics and the chemistry are not coupled at large scales, and the distribution of the positions and the velocities of the turbulent dissipation are introduced as ad hoc parameters.

To describe the evolution of physical quantities like the density, the temperature, the velocity, and the magnetic field, extensive numerical simulations are required. On one side, the magnetohydrodynamical equations describing the dynamics of the ISM are highly non-linear and therefore no analytical solution exists. On the other side, the physical con-



Figure 3: Picture taken from https://science.nasa.gov/. "The Great Nebula in Orion, an immense, nearby starbirth region, is probably the most famous of all astronomical nebulas. Here, glowing gas surrounds hot young stars at the edge of an immense interstellar molecular cloud only 1500 light-years away. In the featured deep image in assigned colors highlighted by emission in oxygen and hydrogen, wisps and sheets of dust and gas are particularly evident. The Great Nebula in Orion can be found with the unaided eye near the easily identifiable belt of three stars in the popular constellation Orion. In addition to housing a bright open cluster of stars known as the Trapezium, the Orion Nebula contains many stellar nurseries. These nurseries contain much hydrogen gas, hot young stars, and stellar jets spewing material at high speeds. Also known as M42, the Orion Nebula spans about 40 light years and is located in the same spiral arm of our Galaxy as the Sun." Image Credit & Copyright: Josep M. Drudis & Don Goldman

ditions of density and temperature are too extreme (e.g. $n_H \sim 0.01 \text{ cm}^{-3}$, $T \sim 10^5 \text{ K}$) to be able to conduct laboratory experiments. Even with powerful calculators, the challenge remains difficult, because on the one hand simulations need to cover a huge range of scales (from stars formation to the galaxy rotation), and on the other, they need to treat several competing and entangled physical processes.

These two facets – chemistry and dynamical simulations – were treated in parallel and independently from each other until recently. Recent theoretical efforts have started to couple the two approaches by implementing the time-dependent evolution of a simplified chemical network in the simulations and by recomputing the abundances of other chemical species at each point in post-treatment. This leads to a more dynamic and coherent treatment of the problem. In particular, it is important to include the processes of formation and destruction of molecular hydrogen in the MHD simulations of ISM mainly for three reasons. Firstly, H_2 is the most abundant molecule in the ISM, secondly, it has long timescales compared to other species, i.e. it is more likely to be out-of-equilibrium, and finally, it is the precursor to the rest of the chemistry.

In this Ph.D. thesis, we continue the effort of combining chemistry and dynamics in a more coherent picture, where at least part of the chemical and dynamical evolutions are coupled. Our goal is to understand the physical processes governing the evolution of the local diffuse ISM and the effect of the multiphase structure on its 3D chemical composition. Numerical simulations allow to conduct a series of experiments for different initial conditions and parameters. Even though these numerical experiments provide important information on their own, the real power resides in their connection with observations. Indeed, the combinations of observations and simulations is a powerful tool to constrain the physics of the simulations, to highlight the influence of different parameters, and to interpret the observational data. In this context, the rising amount of data collected by spectrometers from radio to ultraviolet (UV) wavelengths opens a new era where the statistical and chemical information contained in the observations can be used concomitantly. The originality of this work stands in developing new tools and defining a framework to effectively compare observations and simulations, taking advantage of the union of multiphase simulations and a chemical solver on one side, and the increasing number of observations on the other side. The tools developed in this thesis are a first step toward an N-dimensional investigation in which, multiple observables would be studied simultaneously maximizing the exploitation of the statistical power contained in the observations.

Structure of the Ph.D. Manuscript In the first chapters, i.e. Part I, we present the diffuse interstellar medium in general, its components (\$1), the thermodynamics of the multiphase diffuse gas (\$2), the interaction of matter and photons (\$3), and the three observables we concentrate on in this work (\$4): the hydrogen and its transition from its atomic to molecular form (\$4.1), the neutral carbon (\$4.2) and the elusive CH⁺ (\$4.3). In Part II, we present different aspects of the RAMSES simulations (\$5), the turbulent forcing, the out-of-equilibrium molecular hydrogen, and the computation of the excitation of neutral carbon (\$6) and molecular hydrogen (\$7). In Part III, we present the tools to compare ob-

servations and simulations: the reconstruction of lines of sight (§8), which allows to build the observed distribution of lines of sight from the simulations, and our modifications of the Kolmogorov-Smirnov test (§9), including some tests and application. In the last chapters of Part III, we present the results obtained on the HI-to-H₂ transition (§10), CH⁺ (§11), and CI excitation (§12).

Chapter 1

ISM Composition

The interstellar medium is a dilute mixture of gas, high energy particles, and dust grains. The ISM consists mainly of hydrogen which carries around 70% of the mass, followed by helium which accounts for 28% of the mass. The remaining 2% is carried by heavier elements that are commonly called metals (we will adopt this terminology).

Since the medium is rich in atomic hydrogen, which has an energy potential of 13.6 eV, UV photons with hv > 13.6 eV are effectively absorbed by the interstellar matter. This leads to the existence of a neutral phase in which species with an ionization potential smaller than 13.6 eV are in their singly ionized form while species with a potential larger than that of hydrogen are primarily neutral.

The state and evolution of the diffuse neutral ISM are at the root of the present thesis. In this chapter, we make a brief presentation of its composition, its structure related to the other phases of the ISM, and the energy sources that need to be considered.

1.1 Dust

Even though dust grain carry only 1% of the mass of the ISM, they play an important role in the physics and chemistry of the medium by locking elements in their mantles and cores, by affecting the thermodynamics through heating and cooling, by absorbing and scattering the UV photons coming from the stars and acting as catalysts for chemical reactions.

The existence of dust grains was first indicated by the dimming of starlight. In the 30s Trumpler tried to determine how much light was attenuated by the "absorbing medium" between us and the stars. Using a pair method (stars with same spectral type), the extinction due to the interstellar medium as a function of the wavelength could be traced. This curve is called the "extinction curve" and presents a bump at 217.5 nm that gives away the presence of dust grains: since this feature is too broad to be associated with atoms and small molecules, it traces the existence of large carboneus-rich complexes. Moreover, since different wavelengths are absorbed and scattered differently, and dust grains can only absorb photons with wavelengths smaller than their size, the observation of the extinction curve provides constraints on the size of grains.

Mathis et al. (1977) built a model of a grain size distribution that could reproduce well

the extinction curve, assuming graphite and silicate spheres. This study led to the well-known MRN model in which:

$$n(a) \propto a^{-3.5} \tag{1.1}$$

where n(a) is the density of grain of size *a* per unit size between a - da/2 and a + da/2. This distribution has the interesting property that the total surface of grains

$$\int_{a_{min}}^{a_{max}} a^2 n(a) da \propto a_{min}^{-1/2}$$

is mostly given by the small grains, while their total mass

$$\int_{a_{min}}^{a_{max}} a^3 n(a) da \propto a_{max}^{1/2}$$

is controlled by the large grains.

Nowadays the theory of dust leans on many other observational pieces of evidence, such as the abundance of elements in the gas phase compared to the solar abundances, the polarization of light, the infrared emission, and the abundances of several chemical species, like molecular hydrogen, which requires grains to be formed. Among those studies, polarization is particularly interesting. At the end of the 40s, a correlation between the polarization of the light and the reddening was pointed out: the stronger the reddening the more the light is polarized; in addition, stars close to one another show similar polarizations. These observations led to the conclusion that dust polarizes the light coming from stars. This is presently understood as a consequence of an alignment between dust and the magnetic field, showing that dust grains are not spherical and are coupled dynamically to the magnetic field lines.

1.2 Phases

Observations of the ISM performed at different wavelengths reveal a medium with different thermal and chemical states, characterized by very distinctive physical conditions. For instance, X-ray and UV reveal the presence of gas at temperatures larger than 10⁵ K, while observations at millimeter wavelengths highlight the presence of Giant Molecular Clouds at temperatures below 100 K. All these observations show that the ISM is highly inhomogeneous and can be classified in phases with different temperatures, densities, and ionization fractions (Lequeux 2005; Draine 2011).

HIM The Hot Ionized Medium, also called coronal gas is an ionized phase with a typical temperature $T \sim 10^5 \cdot 10^6$ K, and low density $n_{\rm H} \sim 10^{-2}$ cm⁻³. This very diffuse and hot medium is heated and ionized by supernovae explosions and stellar winds and can take up to ~ 50% of the volume of the disk in the galaxy. It is observed through UV absorptions by highly ionized species (e.g. CIV, SIV, NV, OVI), continuum emission generated by free electrons accelerated in the Coulomb field of positive ions or the radiative recombination of hydrogen and helium ions with free electrons, soft X-rays¹, and spectral emission lines (OVII, OVII) excited by collisions followed by radiative de-excitation.

¹Soft X-ray have typical energies below 1 keV

WIM The Warm Ionized Medium is an ionized phase, with characteristic temperature and density of $T \sim 8000$ K, and $n_{\rm H} \sim (0.2 -) 0.5$ cm⁻³. Its important ionization fraction is mainly due to photons (stellar radiation) and partially to shocks and collisions (ejection of matter from HII regions). This medium can take up to $\sim 10 - 25\%$ of the volume of the ISM and takes up $\sim 90\%$ of the ionized hydrogen (the rest is in HII regions). This medium is traced by the recombination of forbidden and optical line emissions of atoms and ions: H α (at 6563 Å), SII, NaII, OI, OIII, HeI (Sembach et al. 2000), and by ionic absorption in the visible and the ultraviolet.

WNM The Warm Neutral Medium is the atomic neutral phase characterized by the same temperature as the WIM $T \sim 8000$ K, and a density $n_{\rm H} \sim 0.5$ (- 0.6) cm⁻³. The WNM is believes to occupy ~ 30 – 40% of the volume of the Galactic disk and to carry a significant fraction of the total mass of the neutral hydrogen. Its main observational tracer is the emission line of HI at 21 cm.

CNM The Cold Neutral Medium is the atomic neutral phase characterized by $T \sim 70$ K (30-100K) and $n_{\rm H} \sim 50 - 100$ cm⁻³. It occupies a very small percentage of the volume of the galactic disk, only a few percent, but since it is quite dense, it takes up 30% - 60% of the mass of the ISM. The CNM is observable through absorption in UV and visible, and line emission of C⁺ which is an important coolant for this phase. Dense **molecular clouds** are colder ($T \sim 10-20$ K) and denser ($n_{\rm H} \sim 10^3-10^6$ cm⁻³) environment within the CNM bounded by self-gravitation. This fully molecular medium is the residence of complex molecules, the playground of dust chemistry, and the site where star formation occurs. Their main tracers are molecular absorption and emission lines observed in the millimeter domain.

The phases described above are not a universal classification and can be referred to with different names. Several authors prefer to use different criteria, such as the extinction or the chemical state of hydrogen. For instance, Snow & McCall (2006) classify different CNM and molecular clouds based on the value of the visual extinction A_V . Their classification is more refined and leads to the definition of the diffuse atomic, diffuse, translucent and dense molecular cloud.

1.3 Reservoirs of Energy

Non-isolated, the ISM is constantly pervaded by several energy sources. A major ingredient is the stars themselves. Stars not only irradiate the ISM with photons at different wavelengths but regularly inject mass and mechanical energy in the medium through stellar winds and supernovae explosions. A second important energy source is the cosmic rays. These relativistic particles accelerated in supernovae explosions propagate through the ISM with energy sufficient to heat and ionize the gas. At last, the galactic dynamics itself and the sustained large-scale magnetic field provide another important energy input. The magnetic field not only provides dynamical support to the gas through magnetic tension and magnetic pressure but also affects the propagation of cosmic rays.

1.3.1 Mechanical Energy and Turbulence

The ISM is a very dynamic medium. This dynamical state is maintained by the gravitational well of the Galaxy, its differential rotation, the stellar winds, the supernovae explosions of massive stars, and the protostellar outflows which constantly inject massive amounts of mechanical energy in the gas. As a result, the line profiles observed through spectroscopy exhibit supra-thermal broadenings, indicating that the medium is highly turbulent. Indeed, the Reynold number estimated from the velocity dispersion at a scale of 10 pc is of the order of 10^7 : the mechanical energy injected at this scale or above, necessarily induces a turbulent cascade that transport the kinetic energy to the viscous scales where it is dissipated. Because of this long dissipation process, mechanical energy is a major energy reservoir of the ISM. Quantitavely, the density of mechanical energy, or ram pressure, can be written as function of the mass density ρ and the speed v, as

$$P_{\rm dyn} = \rho v^2 \sim \left(\frac{n_{\rm H}}{10 \,{\rm cm}^{-3}}\right) \left(\frac{v}{3.5 \,{\rm kms}^{-1}}\right)^2 \times 1.7 \times 10^{-12} \quad {\rm erg \, cm}^{-3}.$$
 (1.2)

1.3.2 Magnetic Field

Magnetic fields are ubiquitous in the ISM, in the Galaxy, and in the Universe. Even though the origin and the evolution of the magnetic field are still open questions, its importance on the ISM is well established. Because the ISM is at least partly ionized, the magnetic field interacts with gas and dust through magnetic tension and magnetic pressure. This not only affects the entire turbulent cascade and the transport of mechanical energy between scales (Elmegreen & Scalo 2004), but also the structures and evolutions of the different phases of the ISM. Moreover, the magnetic field participates in the regulation of star formation: not only magnetic pressure opposes gravity but also the structure of the magnetic field is key for the evacuation of prestellar angular momentum during the phase of collapse.

Several indirect methods exist to measure the different components of the magnetic fields (parallel or perpendicular to the line of sight): the Zeeman splitting, the synchrotron emission, the Faraday rotation, and the dust polarization. With a typical magnetic field $B \sim 5 \mu m$ the magnetic pressure in the local ISM writes as

$$P_{\rm mag} = \frac{B^2}{8\pi} \sim \left(\frac{B}{5\,\mu\rm{m}}\right)^2 \times 10^{-12}\,\rm{erg}\,\rm{cm}^{-3} \tag{1.3}$$

and is therefore an important energy reservoir of the ISM.

1.3.3 Radiation Field

The photons constitute another important energy reservoir. The photons pervading the ISM originate from several sources including the Cosmic Microwave Background, the stars, and the dust emission, which all contribute to the radiation energy density at different

wavelengths. Separately, these sources amount to typical energy densities of $u_{\text{stars}} \sim 10^{-12}$ erg cm⁻³, $u_{\text{CMB}} \sim 4.2 \times 10^{-13}$ erg cm⁻³, and $u_{\text{IR,dust}} \sim 5 \times 10^{-13}$ erg cm⁻³ (see Table 12.1 of Draine 2011). The corresponding total energy density contained in photons therefore sums up to

$$P_{\rm rad} \sim 2.2 \times 10^{-12} \,{\rm erg} \,{\rm cm}^{-3}.$$
 (1.4)

which is comparable to the magnetic and mechanical energy densities.

Among all photons, the UV radiation field is of particular importance. Indeed, the interaction of the UV photons with interstellar matter plays a key role in the thermodynamics and chemistry of the ISM (see §2 and §3).

1.3.4 Cosmic Rays

The final main energy reservoir of the ISM is the cosmic rays. Cosmic rays (CR) are energetic (up to 10^{20} eV) particles that propagate in the ISM at relativistic speed. These particles are protons for the most part (H⁺), electrons (1%), and some heavy nuclei (10% alpha particles and 1% heavy nuclei). Since CR are very energetic particles, they are an important source of ionization, they can penetrate deeper than UV photons in clouds, and therefore dissociate molecules and heat the medium. The energy spectrum of CR is very sharp: the very energetic CR are rare and have little interaction with the medium, while the low-energy CR are numerous and have large interaction cross-sections with the gas. CR with galactic origins can be generated in supernovae and can reach energies up to 10^{15} eV through diffusive shock acceleration mechanism, i.e. the Fermi mechanism. OB-type stars may be the major source of the Galactic CR at low energy. These CR have small gyroradii and are easily deflected by the magnetic field, thus their propagation is approximated to a diffusive process (Strong et al. 2007). For these reasons, the low-energy CR distribution is often supposed to be isotropic (Draine 2011).

Overall, the contribution of this isotropic flux of CR to the energy density is about (Webber 1998)

$$P_{\rm CR} \sim 1.8 \, {\rm eV} \, {\rm cm}^{-3} \sim 2.9 \times 10^{-12} \, {\rm erg} \, {\rm cm}^{-3}.$$
 (1.5)

1.3.5 Equipartition

As shown in the previous sections, all the energy reservoirs presented above have comparable energy densities, which are also comparable to the thermal pressure of the gas

$$P_{\rm th} \sim nk_B T \sim \left(\frac{n_{\rm H}}{30\,{\rm cm}^{-3}}\right) \left(\frac{T}{100\,{\rm K}}\right) \times 10^{-12} {\rm erg\,cm}^{-3}.$$
 (1.6)

The physics of complex systems is often described by considering only the dominant processes and neglecting, at least in first approximation, all the other processes. The fact that all the energy densities derived above are comparable indicates that they could all play an important role in the structure and evolution of the interstellar matter, and should therefore not be neglected. Considerations on the energy densities alone are, however, not sufficient to draw a definitive conclusion. The importance of any energy reservoir depends on its capacity to interact with the matter and the corresponding interaction rates. These aspects will be described in more detail in the following chapters.

Chapter 2

Thermodynamics

The neutral diffuse ISM is not an isolated system but a system that exchanges energy with its surrounding environment, principally through the absorption and emission of photons. The interaction of this radiation and the gas is responsible for several physical processes taking place in the interstellar medium. These microscopic processes influence the dynamics of the medium and its macroscopic structure.

Indeed, observations show that the diffuse ISM is not homogeneous in density and temperature but presents different phases (see \$1.2) which are a consequence of thermodynamical processes. In the next sections, the different heating and cooling mechanisms are presented and modeled with formulae taken from Wolfire et al. (1995) and Wolfire et al. (2003), and shown in Fig. 2.3. The thermal processes indicated with filled triangles (\blacktriangle \checkmark) are those taken into account in the RAMSES code (see \$5) used to run all the simulations used in this work.

2.1 Heating Processes

The diffuse ISM is rich in photons that can photoionize atoms, molecules and grains, and produce electrons with an excess kinetic energy that heats the medium through collisions.

This heating process is mainly due to UV photons coming from OB stars. Since atomic hydrogen is abundant in the ISM and has an ionization potential of 13.6 eV, a neutral phase can develop. This neutral ISM is opaque to photons with energy greater than 13.6 eV. Photons less energetic can interact and ionized heavier species (ex: C, Mg, Si, Fe..) and dust that have lower ionization thresholds.

▲ Photoelectric Effect on Grains The process of ejection of an electron due to an incident radiation field is called the photoelectric effect. The UV photon, propagating through the diffuse ISM, frees an electron (often called photoelectron) from the dust grain surface. This electron has an excess energy that is the difference between the photon energy and the ionization potential (which is the energy barrier for the detachment of an electron). The surplus of energy is given to the surrounding environment via thermalization through collisions.

The photoelectric effect is a quantum process. Its probability to happen depends on the charge and the surface of the grains. The more positively charged are the grains the largest is the ionization potential that needs to be overcome by the incident photon. This has two negative effects on the heating: (1) a smaller fraction of photons have sufficient energy to ionize the grain, and (2) the ejected electrons have a smaller excess energy.

The heating rate of this process can be written as follow:

$$\Gamma_{\rm ph} = 1.3 \times 10^{-24} \, \epsilon \, G_{\rm eff} \, {\rm erg \ s}^{-1}$$
 (2.1)

where the heating efficiency is

$$\epsilon = \frac{4.9 \times 10^{-2}}{1.0 + [\kappa/1925]^{0.73}} + \frac{3.7 \times 10^{-2} (T/10^4 \text{ K})^{0.7}}{1.0 + [\kappa/5000]}$$
(2.2)

and

$$\kappa = \frac{G_{\rm eff} T^{1/2}}{n_{\rm e} \phi_{\rm PAH}}.$$
(2.3)

The efficiency of this heating process depends on the incident UV radiation field, G_{eff} (seen in more details in §3), the temperature *T* and the recombination parameter of electrons on PAHs, ϕ_{PAH} , discussed in Wolfire et al. (2003) and set to 0.5. The electronic density, n_{e} in the above formula, can be written at ionization equilibrium as (Wolfire et al. 2003):

$$n_{\rm e} = 2.4 \times 10^{-3} \left(\frac{\zeta_{\rm CR}}{10^{-16} {\rm s}^{-1}}\right)^{1/2} \left(\frac{T}{100 {\rm K}}\right)^{0.25} \frac{G_{\rm eff}^{1/2}}{\phi_{\rm PAH}} + n_{\rm H} x_{\rm C^+} {\rm ~cm}^{-3}$$
(2.4)

where ζ_{CR} is the total ionization rate including primary and secondary ionization of H by EUV photons, soft X-ray photons, and cosmic rays (set to $\zeta_{CR} = 1.3 \times 10^{-16} \text{ s}^{-1}$), x_{C^+} is the abundance of C⁺ relative to $n_{\rm H}$, $x_{C^+} = n(C^+)/n_{\rm H}$, which is set to 1.4×10^{-4} , i.e. its Solar Neighborhood value assuming a 40% depletion of the carbon onto grains and that the carbon remaining in the gas phase is singly ionized. This equation for the density of electrons differs from Wolfire et al. (2003) by the addition of C⁺ which is the most abundant ion in the diffuse and transluscent cold neutral medium (Snow & McCall 2006). Since the size distribution of the grains goes as $n(a) \propto a^{-3.5}$ (see §1.1), the majority of the surface that undergoes the photoelectric effect is provided by small grains and PAHs.

The efficiency of the photoelectric effect is often shown as a function of the parameter κ for different temperatures (see Fig. 2.1). But since κ and T are not two independent variables, and since we described the electronic density also as a function of the temperature, we also plot in Fig. 2.1 the photoelectric efficiency ϵ as a function of κ taking into account the dependence of this parameter on the temperature $\kappa(T)$, for two different density and a $G_{\text{eff}} = 1$ in Habing units.

▲ Cosmic Rays and X-rays ionization Cosmic rays and X-ray photons are important contributors to the heating. By ionizing the main constituents of the gas (i.e., H, H₂, and He which has a large cross-section with X-rays), CR and X-rays free electrons. These electrons have an excess energy which has a wide distribution with an average of 30-35 eV. The freed electron can either produce secondary ionizations – important in chemistry – or thermalize, heating the gas mainly through electron-electron collisions. X-rays decrease



Figure 2.1: Photoelectric efficiency as a function of the parameter κ (Bakes & Tielens 1994) for two temperature values in dashed lines. In dotted lines, the photoelectric efficiency as a function of $\kappa(T) = \frac{G_{\text{eff}}T^{1/2}}{n_e\phi_{\text{PAH}}}$ in which the electrons are computed as a function of G_{eff} and T, with eq. 2.4. The green dotted curve assumes a proton density $n_{\text{H}} = 1 \text{ cm}^{-3}$ while the orange curve $n_{\text{H}} = 100 \text{ cm}^{-3}$. The external radiation field in Habing units is set to $G_{\text{eff}} = 1$, consitent with Γ weakly dependent on T. In the blue and red curve κ is a parameter with no specific dependence on T. If T and κ are treated as two separated parameters (as in the blue and red curves), then there is a drop-off of the heating efficiency at $10^3 - 10^4$ K, which is not the case in reality.

as the column density increases, so the photoionization by X-rays becomes less important in shielded regions. In contrast, CR can penetrate and ionize the gas effectively in those regions. The resulting heating can be written as

$$\Gamma_{\rm CR} \sim 10^{-27} \left(\frac{\zeta_{\rm CR}}{10^{-16} {\rm s}^{-1}} \right) {\rm erg \ s}^{-1}.$$
 (2.5)

 \triangle Mechanical Energy Macroscopical movements of the gas inject mechanical energy at large scales that cascades down to small scales where it is dissipated (see §5.3). This dissipation contributes to the heating of the gas, but since the medium is magnetized, non-adiabatic, and highly compressive, mechanical heating is difficult to evaluate. In the past, several theoretical efforts have been made to try to model the turbulent heating rate. On average, the turbulent heating rate can be deduced from the Larson's law (Pan & Padoan 2009) and modeled as

$$\Gamma_{\rm turb} = 3 \times 10^{-27} \left(\frac{L}{\rm 1pc}\right)^{0.2} n \,\rm erg s^{-1} cm^{-3} \tag{2.6}$$

which gives a function of the scale considered. However, such a prescription for turbulent heating is a great misconception. Indeed, the turbulent heating, as turbulence itself, is highly intermittent. At a certain point in space and time, a lot of the energy can dissipate and a lot of heating can be produced. This dissipation by bursts can heat the gas through viscous dissipation and have strong consequences on the chemistry (Falgarone & Puget 1995; Joulain et al. 1998).

This type of heating is not directly included in RAMSES because the physical scale required to described dissipation regions is not resolved numerically.

 \triangle H₂ Formation and Destruction H₂ formation and destruction also release energy in the medium. The chemical reaction of formation of H₂ onto grains is exothermic and releases 4.5 eV that can heat up the surrounding gas, be partially stored in H₂ as internal energy and/or heat the grain on which H₂ forms. Since there is no consensus on the details and the fraction of energy going into these different processes, equipartition of energy is usually assumed. As a consequence, when H₂ is formed onto a grain, 1.5 eV is given to the thermal energy of the gas.

Destruction of H_2 usually happens after a radiative excitation in Lyman or Werner bands. This excitation leads 10% of the times to a radiative dissociation that releases on average 0.4 eV (see §5.4). In the other 90% of the cases H_2 de-excites to rovibrational excited states of the ground electronic state. This process is called fluorescence.

To correctly take into account these processes into the heating of the gas, it is not only necessary to compute the formation and destruction of molecular hydrogen on-the-fly in a simulation, but also to compute its excitation in order to obtain the dissociation rate. The implementation of these processes in RAMSES is an ongoing work.

2.2 Cooling Processes

Cooling processes correspond to the energy lost by the system, i.e. the amount of thermal energy converted into radiation that escapes the cloud. This conversion mainly happens

through collisions, which internally excite some species followed by the reemission of photons. The main cooling processes are, therefore, line emissions (radiative cooling) from the most abundant species.

▼ Lyman α The most abundant species in the diffuse ISM is atomic hydrogen which has permitted dipole transitions. At high temperatures, the first electronic level of neutral hydrogen can be collisionally excited by an electron. This process is followed by the emission of a Lyman α photon that escapes the medium. The associated cooling rate reads (Spitzer 1978):

$$\Lambda_{\rm HI} = 7.3 \times 10^{-19} x_{\rm e} \ {\rm e}^{-118400/T} \ {\rm erg} \ {\rm cm}^3 \ {\rm s}^{-1}, \tag{2.7}$$

where electrons are the main collisional parters of H and their abundance relative to the proton density, x_e , has been previously defined in eq. 2.4. The temperature for this transition is T = 118400 K which correspond to a photon of 10.2 eV and a wavelength of 1216 Å. This cooling is important at high temperatures and becomes dominant for T $\gtrsim 10^4$ K (cf Fig. 2.2).

When the temperatures become higher, the abundance of atomic hydrogen decreases, and Lyman α emission becomes less effective in cooling the gas. Even though temperature regimes over 10⁵ K are not relevant in our study, we show in Fig. 2.2 a sketch of the total cooling rates for $T > 10^5$ K taken from Gnat & Ferland (2012).

▼ Fine-Structure Lines Permitted electronic transitions are difficult to excite at temperatures below 10^4 K because the number of electrons with sufficient energy exponentially decreases. In this regime of temperatures, forbidden electrical-dipole transitions become the dominant coolants. Some species have a fine-structure splitting in their ground electronic state. This splitting is due to the coupling between the spin, i.e. the intrinsic magnetic momentum, and the orbital electronic momenta, i.e. the magnetic dipole due to the orbital motion of the electrons bound to the atom. The separation between the fine-structure energy levels is of the order of 10^{-2} eV, which corresponds to some tens or hundreds of K, depending on the species. This small separation between the fine-structure levels makes it possible to be excited at low temperatures. Species with such fine-structure splitting are, for instance, CI, OI, CII, and SiII. The associated transitions are forbidden dielectric dipole and as a consequence, the rates are orders of magnitude lower than the electric dipole permitted transitions, e.g. Lyman α transitions ¹. Both oxygen and carbon ion, C⁺, which are the most abundant species after hydrogen and helium, present fine-structure levels.

Since neutral **carbon** has an ionization potential of 11.26 eV, most of the carbon is in the form of C⁺ in the neutral diffuse ISM. The fine-structure of the ground state of C⁺ includes two energy levels separated by $\Delta E/k_B = 92$ K. It follows that the radiative de-excitation of the excited fine-structure level can be an effective coolant of the medium at temperatures

¹Fine-structure transitions are 10^{14} times weaker than Lyman α because (1) fine-structure transitions are magnetic dipole transitions, therefore their transition matrix elements are of the order of $\alpha_{FS}^2 \sim 5 \times 10^{-5}$, where $\alpha_{FS} = 1/137$ is the fine structure constant, thus 10^5 times smaller than a permitted transition; (2) the splitting between fine-structure levels is smaller and since the $A_{ij} \propto v_{ij}^3$, the Einstein coefficients of the fine-structure transitions are 10^9 smaller than Lyman α Einstein coefficients.

of a few tens of kelvins (cf Fig. 2.2). Including collisions by atomic hydrogen and electrons, the cooling rate reads:

$$\Lambda_{\rm CII} = \left[2.25 \times 10^{-23} + 10^{-20} \left(\frac{T}{100 \,\mathrm{K}} \right)^{-0.5} \frac{n_{\rm e}}{n_{\rm H}} \right] \,\mathrm{e}^{-92/T} \,x_{\rm C^+} \quad [\mathrm{erg}\,\mathrm{cm}^3\,\mathrm{s}^{-1}]. \tag{2.8}$$

where $x_{C^+} = 1.4 \times 10^{-4}$, i.e. the abundance of C⁺ is set to its Solar Neighborhood abundance and assuming a depletion of 40% in grains (Sofia et al. 1997; Sembach et al. 2000).

Because **oxygen** has an energy potential higher that the Lyman limit (13.6 eV), oxygen exists mainly in its neutral form in the neutral diffuse ISM. Neutral oxygen presents three fine-structure energy levels at 0 K, 228 K and 326 K. Taking the solar abundance of oxygen and a depletion of 37% we obtain $x_0 = 3.2 \times 10^{-4}$ (Meyer et al. 1998; Sembach et al. 2000). The cooling rate due to oxygen line emission can be written:

$$\Lambda_{\rm OI} = 7.81 \times 10^{-24} \left(\frac{T}{100 \,\mathrm{K}}\right)^{0.4} \,\mathrm{e}^{-228/T} x_{\rm O} \quad [\mathrm{erg}\,\mathrm{cm}^3\,\mathrm{s}^{-1}], \tag{2.9}$$

where collisions with the primary collider H are considered and the two terms arise from the two fine-structure excited levels of atomic oxygen. Wolfire et al. (2003) included only the transition from the first excited level at 228 K. However, oxygen has also a second excited level at 326 K. In §5.2.2, we discuss the effect of the inclusion of this level (see Fig.5.2).

Oxygen and carbon are the dominant coolants in the CNM phase. Even though oxygen is more abundant, its fine-structure levels have a separation in energy larger than C^+ , making this last species the most effective coolant at low temperatures (see Fig. 5.1).

✓ Metastable Fine-Structure Lines Both CII and OI also have transitions between metastable states. The cooling rates induced by these transitions can be expressed as (Hollenbach & McKee 1989):

$$\Lambda_{\text{CII,meta}} = 6.2 \times 10^4 \ k_B \ \times 9/5 (2.3 \times 10^{-8} \ x_e \times T_4^{-0.50} + 10^{-12}) e^{-6.2/T_4} \ x_{\text{C}^+}$$
(2.10)

$$\Lambda_{\rm OI,meta} = 2.3 \times 10^4 k_B \times 5/7 (5.1 \times 10^{-9} T_4^{0.57} x_e + 10^{-12}) e^{-2.3/T_4} x_{\rm O}$$
(2.11)

$$+4.9 \times 10^{4} k_{B} / 7 (2.5 \times 10^{-9} T_{4}^{0.57} x_{e} + 10^{-12}) e^{-4.9 / T_{4}} x_{O}$$
(2.12)

$$+2.6 \times 10^{4} k_{B} / 5(5.2 \times 10^{-9} T_{4}^{0.57} x_{e} + 10^{-12}) e^{-2.6/T_{4}} x_{O}$$
(2.13)

with $T_4 = T/10^4$ K. The main collisional partners are hydrogen atoms, for which we used a constant rate coefficient of 10^{-12} cm³ s⁻¹ given by Federman (1983), and electrons (Hollenbach & McKee 1989). Because the levels are separated by several 10^4 K, these metastable transitions play a role only at high temperature. As shown in §5.2.2, these cooling rates are important only if Lyman α is not included. We, therefore, neglect these terms in our simulations.

▼ Electron Recombination onto Charged Grains Grains and PAHs can be positively charged. When electrons recombine on a dust grain, the gas loses kinetic energy. As shown by Bakes & Tielens (1994), this cooling can be expressed as

$$\Lambda_{\rm rec} = 4.65 \times 10^{-30} \ T^{0.94} \ \kappa^{\beta} x_{\rm e} \phi_{\rm PAH} \ {\rm erg} \ {\rm cm}^3 \ {\rm s}^{-1}$$
(2.14)

with $\beta = 0.74/T^{0.068}$ and ϕ_{PAH} the recombination parameter.

As the photoelectric effect, the electron recombination depends on the parameter $\kappa = \frac{G_{\text{eff}}T^{1/2}}{n_e\phi_{\text{PAH}}}$ which takes into account the increase of the charge of the grain with the temperature, thus, the increase of Coulomb interaction. This cooling is important in all temperature regimes in the diffuse neutral ISM (see Fig. 2.2) and needs to be included in the computation of the thermal processes.

Molecular Lines Molecular species can also contribute to the cooling. Compared to atoms and ions, they have more degrees of freedom, thus they not only present electronic levels but also vibrational and rotational levels that can be excited. Similarly to the fine-structure splitting, the separation in energy of these levels is orders of magnitude smaller than for the electronic states.

▼ **Molecular hydrogen** does not have a permanent dipole (see §5.4 for more details). Consequently, the transitions between neighboring rotational levels are strongly forbidden (ΔJ must be even). The first allowed rotational transition is $J = 2 \rightarrow 0$, which is separated by $\Delta E/k_B = 510$ K. Because of this large energy separation, this transition is extremely difficult to excite at low temperatures (T < 100 K), i.e. in regions where molecular hydrogen is the most abundant. Glover & Clark (2014) showed that the cooling is dominated by CII and OI and that molecular hydrogen radiative cooling has an impact on the thermodynamics of the medium only for the two following scenarios: (1) if the metallicity of the local ISM is 90% lower than its solar value, i.e. metals like C and O are extremely rare compared to H₂, or (2) in molecular shocks, where the temperatures can be important enough to excite rovibrational levels of H₂ (Hollenbach & McKee 1979, 1989). H₂ cooling is very dependent on H₂ fraction, not only because the radiative cooling depends on the abundance of the species under consideration, but also because H₂-H₂ collisions are 10 times more effective than H-H₂ collisions in exciting molecular hydrogen, as long as T < 500 K.

∇ Hydrogen deuteride, **HD**, has a permanent dipole moment, therefore J = 1 → 0 transition is possible. Since the corresponding separation energy, $\Delta E/k_B$, is lower the J = 1 level can be collisionally excited at lower temperatures. One could think that this molecule could be effective at cooling the gas at T < 100 K, since its formation rate is exothermic, i.e. energetically favorable, and its destruction endothermic. However, its abundance is limited by deuterium abundance, which is smaller than 2.5×10^{-5} (Cooke et al. 2014). Consequently, this cooling process is less important than CII and OI even when considering a fully molecular medium, and will be ignored in our study. At high density, it could potentially be an important source of cooling, but those densities are beyond the scope of this Ph.D. thesis.

 \bigtriangledown Another molecule that can cool the medium is carbon monoxide, **CO**, which is the second most abundant molecule in the ISM. It is heavier than the two molecules presented before. Its energy levels are closer to each other, which implies that CO can be an effective coolant at low temperatures. In unshielded regions, the carbon is in its ionized form. For CO to be abundant, C⁺ must have transitioned to C and finally to CO. This requires



Figure 2.2: Different cooling (solid lines) and heating (dotted lines) rates as function of temperature. At temperatures of the CNM, C⁺ is the main coolant, followed by O that becomes more important at higher temperatures. In the WNM, and at $T \gtrsim 10^4$ K, Lyman α is the dominant coolant. The thick black line is the total cooling function $n_{\rm H}\Lambda$ taking into account all the cooling processes marked with a \checkmark (see main text). The gray curve is the prolongation of the black curve for higher temperatures described by Sutherland & Dopita (1993); Gnat & Ferland (2012) which includes cooling from ionized metals such as Mg, Fe, and Si. The cooling function is highly dependant on the temperature, while the heating curves are almost constant. The cooling rates are computed assuming $n_{\rm H} = 1$ cm⁻³. No extinction due to dust grains is taken into account for the photoelectric heating.

regions shielded from the UV photons of energy greater than 11.2 eV, i.e. relatively dense and opaque CNM clouds. Because the work presented below is focused on the most diffuse phases of the ISM, the cooling by CO is neglected. We note, however, that such an approximation is inappropriate for dense and translucent environments.

2.3 Thermal Instability and Multi-phase Medium

Taking into account the heating and cooling processes presented in the previous section, we can define the net cooling function:

$$\rho \mathscr{L} = n_{\rm H}^2 \Lambda - n_{\rm H} \Gamma \tag{2.15}$$

where $n_{\rm H}^2 \Lambda$ and $n_{\rm H}\Gamma$ are the cooling and heating rates of the medium (in erg cm⁻³ s⁻¹). The thermal equilibrium is reached when the heating and the cooling rates compensate one another, i.e. $n_{\rm H}^2 \Lambda = n_{\rm H}\Gamma$ and $\mathcal{L} = 0$.



Figure 2.3: A visual summary of the heating and cooling processes considered in this work taken directly from Wolfire et al. (2003). Heating and cooling curves as function of the total proton density in the local ISM. Dashed lines indicate heating processes: photoelectic effects onto dust grains and PAHS (PE), X-rays (XR), cormic rays (CR) and carbon photoionization (CI). The solid lines indicate cooling rates: CI, CII and OI fine-structure line emission, Lyman α and metastable transitions, and recombination onto small grains. Colored lines are those taken into account in our work.



Figure 2.4: Thermal equilibrium curve, $\mathscr{L} = 0$ or $n_{\rm H}^2 \Lambda = n_{\rm H} \Gamma$ in the $P/k_B - n_{\rm H}$ space. In a range of pressure, between $P_{\rm min}$ and $P_{\rm crit}$ three branches coexist. Two of these branches are stable, the CNM and the WNM, while the third one is unstable. The dotted gray lines are isotherms at 80000 K, 8000 K, 800 K, 800 K, 800 K, 80 K, starting from the left. Above the equilibrium curve, $\mathscr{L} > 0$, $n_{\rm H}^2 \Lambda > n_{\rm H} \Gamma$, i.e. the gas cools down. Below the equilibrium curve $\mathscr{L} < 0$, $n_{\rm H}^2 \Lambda < n_{\rm H} \Gamma$, i.e. the gas heats up. As shown in Fig. 2.4, the thermal equilibrium curve in the $P/k_B - n_H$ space is not monotonic and therefore reveals a thermal instability process. Branches where the pressure increases with the density correspond to stable equilibrium states, while branches where the pressure decreases when the density increases correspond to unstable equilibrium states.

Let's consider an isobaric line with pressure $P_{crit} > P > P_{min}$, indicated with a green dotted line in Fig. 2.4, and let's call A, B, C the intersections of this line and the thermal equilibrium curve. If we consider an isobaric perturbation of a parcel of gas at density $n_{\rm B}$, e.g. an increase of the density $n_{\rm H} > n_{\rm B}$, the parcel of gas enters the cooling domain. Its temperature decreases while its density increase until it reaches the stable equilibrium state C. If the density is perturbed again, $n_{\rm H} > n_{\rm C}$, the parcel enters the heating zone of the graph. Its temperature increases and its density decreases back to the state C. If the initial isobaric perturbation of the parcel of gas in B corresponds to a decrease in density, i.e. $n_{\rm H} < n_{\rm B}$, the parcel enters the heating zone. Its temperature increases and its density decreases and its density decreases until it reaches the stable equilibrium state A. This cartoon vision of the condensation and evaporation of a parcel of gas, starting from the B state, allows to understand the three regimes of the equilibrium curve which correspond to the different phases of the ISM. The two stable branches at high and low temperatures correspond to the WNM and CNM respectively, while the unstable branch and the surrounding states correspond to the Lukewarm Neutral Medium (LNM), i.e. gas in transition between the two stable states.

Chapter 3

Radiation - photons

The ISM is a complex medium, in which the different components and energy sources are coupled and entangled. A striking example is the coupling of radiation and matter that influence the dynamical evolution of the medium.

The physical state of the interstellar gas is determined in large part by the interaction of the gas and dust with the photons. As seen before, the thermal state also depends on this radiation field through different processes of heating and cooling which have dynamical consequences on the structure of the medium. Moreover, photons govern the chemical and ionization states of the gas, which greatly depends on the rates of photoionization and photodissociation.

In the following sections, we are presenting the radiation field at different wavelengths, focusing on the waveband of interest in our study, the extinction of this radiation field due mainly to dust grains, and finally, we cast an eye on the sources responsible for the UV part of the spectrum, OB stars.

3.1 Interstellar Radiation Field

There are different sources of radiation in the ISM, dominating the ISRF at different wavelengths in the Solar Neighborhood, which will be presented in the next paragraph by increasing frequency. All energy densities are taken from Draine (2011) book and the global spectrum is shown in Fig. 3.1. Particular attention will be given to the far-ultraviolet (FUV) photons which are major actors in the chemistry of the medium, being responsible for the photodissociation of most of the molecular species.

Galactic Synchrotron Emission Relativistic electrons interact with the magnetic field emitting synchrotron radiation, which is the dominating radiation in the radio waveband ($v \leq 1$ GHz). Relativistic electrons are not characterized by a Maxwellian velocity distribution, therefore the synchrotron emission is non-thermal. This emission dominates the radiation field at frequencies $v \leq 1$ GHz. Since the relativistic electrons are produced by supernovae and are accelerated through the Fermi process in shocks, the distribution of this emission can vary in space. A spatial mean of the synchrotron emission over the whole sky gives a total energy density integrated overall frequencies of 2.7×10^{-18} erg cm⁻³.



Figure 3.1: Interstellar continuum radiation field in an HI cloud in the Solar Neighborhood as presented in Draine's book. Spectral lines are not included. Solid line is the sum of all components for $hv \le 13.6$ eV.

Cosmic Microwave Background The CMB has nearly a constant temperature of 2.73 K (2.7255 ± 0.0006 K) and it is very close to emitting as a perfect blackbody. The radiation is basically isotropic. Even though it dominates the energy budget of the ISRF between $\lambda = 600 \ \mu m$ and $\lambda = 30 \ cm$, with an energy density of ~ 4 × 10⁻¹³ erg cm⁻³, it does not play an important role in the overall energy balance of the ISM, because the gas and the dust are typically warmer. The CMB interacts little with the ISM.

Infrared emission Dust grains absorb UV radiation coming from young stars and re-emit in IR. This radiation dominates at wavelengths between $\lambda \sim 600 \ \mu\text{m}$ and $\lambda \sim 5 \ \mu\text{m}$. This process is called extinction and depopulates the medium of UV photons, enriching it in infrared photons. Around 2/3 of the power of this radiation is emitted thermally by dust at $\lambda \sim 50 \ \mu\text{m}$, and 1/3 by PAH vibrational bands. The total energy density of dust radiation is $5 \times 10^{-13} \text{ erg cm}^{-3}$. Old stars as well contribute to the IR radiation.

Cold stars Cold stars are responsible for radiation in the visible and near-infrared part of the spectrum. Mathis et al. (1983) modeled this emission as a combination of three diluted blackbody spectra at 6184, 6123, and 2539 K. An update used by the Meudon PDR code (Le Petit et al. 2006) reads:

$$I_{\lambda} = 4\pi (2.07 \times 10^{-14} B_{\lambda} (6184K) + 1.09 \times 10^{-14} B_{\lambda} (6123K) + 1.52 \times 10^{-12} B_{\lambda} (2539K))$$
(3.1)

where

$$B_{\lambda}(T) = \frac{2hc^2/\lambda^5}{e^{hc/\lambda k_B T} - 1}$$
(3.2)



Figure 3.2: On the left an overview of the different prescriptions of the UV interstellar radiation field: Mathis' (green line) broken powerlaw (dotted line, eq. 3.3) and analytical fit (solid line, eq. 3.4), Habing (blue, eq. 3.6) and Draine (red, eq. 3.5). On the right, exploration of the Mathis radiation field at larger wavelengths.

is the blackbody spectrum as function of the wavelength λ and temperature *T*.

Hot Stars Far Ultraviolet photons come from young massive stars of O and B types. Photons with energy larger than the atomic hydrogen potential (hv > 13.6 eV) are absorbed by the hydrogen surrounding these young stars, i.e. the medium progressively becomes optically thick to photons with energy hv > 13.6 eV. This leads to the formation of HII bubbles, while the rest of the FUV photons can propagate further. FUV photons are of particular interest in H₂ chemistry since they are responsible for its dissociation.

Several prescriptions have been proposed for the FUV part of the ISRF and are shown in Fig. 3.2. Mathis et al. (1983) proposed a broken power-law fit:

$$vu_{\nu} = \begin{cases} 2.373 \times 10^{-14} \lambda^{-0.6678} & 1340 - 2450 \text{\AA} \\ 6.825 \times 10^{-13} \lambda & 1100 - 1340 \text{\AA} \\ 1.287 \times 10^{-9} \lambda^{4.4172} & 912 - 1100 \text{\AA} \end{cases}$$
(3.3)

with the wavelength λ in μ m. From a fit based on Mathis et al. (1983), Jacques Le Bourlot derived the expression of the FUV used in the Meudon PDR code:

$$(4\pi)I_{\lambda}\Big|_{\text{Mathis}} = 107.192\lambda^{-2.89} \left[\tanh\left(4.07 \times 10^{-3}\lambda - 4.5991\right) + 1 \right], \quad \lambda \le 8000 \text{ Å}.$$
(3.4)

In this expression, λ is the wavelength in Å and *I*, the specific intensity, in erg cm⁻² s⁻¹ sr⁻¹ Å⁻¹.

Draine & Salpeter (1978) estimated the UV background with a polynomial fit of observations up to 2000 Å:

$$I_{\lambda}\Big|_{\text{Draine}} = \frac{1}{4\pi} \left[\frac{6.3600 \times 10^7}{\lambda^4} - \frac{1.0237 \times 10^{11}}{\lambda^5} + \frac{4.0812 \times 10^{13}}{\lambda^6} \right], \quad \lambda \le 2000 \text{ Å}$$
(3.5)

	$\int u_{\lambda} d\lambda$ $[\mathrm{erg}\mathrm{cm}^{-3}]$	$\int u_{\lambda} c d\lambda$ $[\mathrm{erg}\mathrm{cm}^{-2}\mathrm{s}^{-1}]$	$\int u_{\lambda}/E_{\lambda} \ cd\lambda$ [photon cm ⁻² s ⁻¹]	scaling factor
Habing	8.011×10^{-15}	2.403×10^{-4}	1.224×10^7	1.00
Mathis	1.012×10^{-14}	3.035×10^{-4}	1.556×10^7	1.27
Draine	1.342×10^{-14}	4.027×10^{-4}	2.084×10^7	1.70

Table 3.1: Comparison of the UV radiation field integrated between 910 and 1110 Å, as described by Habing (1968), Mathis et al. (1983), and Draine & Salpeter (1978). $u_V = \frac{4\pi I_V}{c}$

with λ in Å.

Another prescription of the UV radiation field was proposed by Habing (1968). The spectrum of this radiation field has been fitted by Draine & Bertoldi (1996) and reads

$$\lambda u_{\lambda} \Big|_{\text{Habing}} = \left[-\frac{25}{6} \left(\frac{\lambda}{1000} \right)^3 + \frac{25}{2} \left(\frac{\lambda}{1000} \right)^2 - \frac{13}{3} \frac{\lambda}{1000} \right] \times 10^{-14} \quad \text{erg cm}^{-3}$$
(3.6)

with λ in Å.

In Table 3.1 and Fig. 3.2, we compare the properties of the UV ISRF prescribed by Mathis, Draine, and Habing in a range of wavelength going from 910 Å to 1110 Å. In the RAMSES simulations (see \$5) used in this work, the external radiation field illuminating the simulated gas at the integral scale is set to the Habing radiation field scaled with the parameter G_0 .

Plasma emission The hot ionized gas is characterized by free-free (bremsstrahlung), freebound (recombination), bound-bound (heavy ions not completely ionized) transitions that emit EUV and soft X-rays (13.6 eV < hv < 10³ eV). It is important to notice that the ISM is mostly opaque to EUV and transparent to high energy X-rays (> 500 eV). The energy density of X-rays between 1 – 2 keV is of 10⁻¹⁷ erg cm⁻³, i.e. very small compared to the energy densities of the radiation field in other bands, nevertheless, these photons can be an important source of ionization and electrons (see §2).

3.2 Extinction

The medium is not transparent to photons. The UV photons propagating through the ISM are absorbed by dust and re-emitted at longer wavelengths (IR), dimming the light coming from the stars. This extinction of the radiation depends on many factors as the wavelength of the incident radiation, the size distribution of the dust grains, and their spatial distribution, i.e. the structure of the matter that dictates the permeability to photons. The attenuation of light at different wavelengths is characterized by the extinction curve. The extinction is a function that increases from longer to shorter wavelength. Because the attenuation of the radiation due to the scattering and absorption of photons by dust grains is greater for shorter wavelengths, the propagating light appears redder than the emitted
light: this phenomenon is called reddening.

The shielding effect due to grains is crucial in the physics and chemistry of the ISM. It affects the thermodynamics, through the photoelectric effect, and the photodissociation of the chemical species in the medium. Since the permeability of photons depends on the structure and chemical composition of the ISM, radiation, chemistry, and dynamics are once again coupled.

The intensity of the radiation field at a wavelength λ , in the UV visible and near IR part of the spectrum, goes as:

$$I(\lambda) = I_0(\lambda)e^{-\tau(\lambda)}$$
(3.7)

where $I_0(\lambda)$ is the emitted intensity and $\tau(\lambda)$ is the optical depth that depend on the wavelength, λ . The extinction at a wavelength λ is defined as:

$$A(\lambda) = -2.5 \log \frac{I(\lambda)}{I_0(\lambda)} \approx 1.086\tau(\lambda)$$
(3.8)

where τ is related to the size of the absorbing particles via their cross sections.

The extinction in the visible, A_V , can be parametrized by the ratio of extinction in the *V* band¹ to the reddening:

$$R_V \equiv \frac{A_V}{A_B - A_V} = \frac{A_V}{E(B - V)} = \frac{1}{\tau_B / \tau_V - 1}.$$
(3.9)

 R_V depends on the dust properties and varies considerably from one line of sight to another. In the Solar Neighborhood the typical value is $R_V = 3.1$ (Schultz & Wiemer 1975; Fitzpatrick 1999). E(B - V) is the color excess is defined as:

$$E(B-V) = (B-V)_{\text{obseved}} - (B-V)_{\text{intrinsic}}$$
(3.10)

Since the color excess depends on the amount of dust on the line of sight, it correlates with the proton column density (Bohlin et al. 1978)

$$\frac{N_{\rm H}}{E(B-V)} \sim 5.8 \times 10^{21} \, \text{atoms cm}^{-2} \, \text{mag}^{-1}.$$
 (3.11)

3.3 OB Stars

As mentioned before, FUV photons correspond to the most relevant part of the spectrum for our study. UV radiation mainly comes from young stars. These stars belong to two spectral types called O and B. They are massive stars belonging to the main sequence characterized by high temperatures and short lifetimes. Their ionizing radiation changes the chemistry of the surrounding environment, injecting energy, and ionizing and heating the entire diffuse interstellar gas.

OB stars' spatial distribution is known to be in clusters, called OB associations, that disperse in time. Ambartsumian (1947) defined OB associations as regions in which the

 $^{^1}V$ it the visible passband at $\lambda\sim 550$ nm, and B the blue passband at $\lambda\sim 440$ nm

density of this type of stars is higher than the average field. Not gravitationally bound, these associations are expanding and dispersing in time.

As shown by the recent 3D studies of the distributions of stars based on the Hipparcos and GAIA Catalogs (e.g. Bouy & Alves 2015 and Zari et al. 2018), the typical distances separating two associations in the Solar Neighborhood range between 50 pc and a few hundreds of pc, i.e. several times the mean distance deduced from the integrated surface densities of OB stars ($\sim 1.6 \times 10^{-3}$ pc-2, Maíz-Apellániz 2001). Interestingly, such distances are not only comparable to the heights of the molecular gas (~ 75 pc) and the cold HI gas (~ 150 pc) above the galactic plane deduced from CO and HI all-sky surveys (e.g., Dame et al. 2001, Dickey & Lockman 1990, and Kalberla & Kerp 2009), but they also correspond to the typical size of HI superclouds (Elmegreen & Elmegreen 1987).

Chapter 4

Observations

Thanks to the development of a great variety of telescopes during the past century by the international community, the diffuse ISM has been observed at all wavelengths, not only in our Galaxy but also in extragalactic environments. This accumulation of data has led, over time, to a more complete view of the structure, composition, and evolution of the diffuse gas. In this context, spectroscopic observations are of particular importance. By giving access to atomic and molecular spectra, spectroscopic observations provide individual measurements, and sometimes even cartographies, of the chemical composition of the ISM. In turn, and because the chemical and excitation processes are highly dependent on local physical conditions, each observation provides formidable constrains on the thermodynamical properties of the gas.

As shown by Gudennavar et al. (2012), the local diffuse ISM has been observed in absorption or emission over thousands of lines of sight, at all Galactic longitudes and latitudes, with more than 20 different atomic and molecular tracers. The complete description of all observational data, chemical correlations, and theoretical deductions are far beyond the scope of the present thesis. In this chapter, we, therefore, focus on three different datasets which are at the center of our entire work: the observations of atomic and molecular hydrogen, the observations of the excited levels of the neutral carbon, and the observations of the methylidyne cation CH⁺.

4.1 Atomic and Molecular Hydrogen

Since hydrogen is the most abundant element in the ISM, the observations of HI and H_2 are of great value to characterize the structure and morphology of the diffuse matter. Moreover, because the conversion of the gas from atomic to molecular is a fundamental step to trigger a more complex chemistry, direct measurements of the molecular fraction is a primary goal for any coherent study of the diffuse gas.

In the past, extensive observations of molecular hydrogen have been done in absorption, starting from its first detection (Carruthers 1970), followed by short-term missions, such as the *HUT* (the Hopkins Ultraviolet Telescope, e.g., Blair et al. 1996), *ORFEUS* (the

Orbiting and Retrievable Far and Extreme Ultraviolet Spectrometer, e.g., Barnstedt et al. 2000), and *IMAPS* (e.g., Jenkins & Peimbert 1997). At the beginning of the 70s, a large survey on molecular hydrogen was performed with the *Copernicus* space mission (1972-1981), (e.g. Savage et al. 1977; Bohlin et al. 1983), followed by the *FUSE* mission (Far Ultraviolet Spectroscopic Explorer, e.g. Rachford et al. 2002b; Gillmon et al. 2006; Rachford et al. 2009), which was capable to observe translucent clouds thanks to its high spectral resolution.

In this section, we present the set of lines of sight where both the total hydrogen column density $N_{\rm H}$ and the molecular hydrogen column density $N({\rm H_2})$ have been observed. The goal is to study the combined distribution of these two quantities and derive observational conclusions regarding the molecular fraction. If HI and H₂ are observed along the same line of sight, $N_{\rm H}$ is simply computed as

 $N_{\rm H} = N({\rm H}) + 2N({\rm H_2});$

if not, $N_{\rm H}$ is derived from the reddening E(B - V) as

 $N_{\rm H} = 5.8 \times 10^{21} E(B - V) ~{\rm cm}^{-2}$,

assuming a standard Galactic extinction (Bohlin et al. 1978).

4.1.1 Observational Results

Column Densities All the column densities of H_2 presented below are obtained through direct observations of its FUV absorption lines. The sample results from a combination of the *Copernicus* survey of nearby stars (e.g., Savage et al. 1977; Bohlin et al. 1983) and the FUSE survey of the Galactic disk and the Galactic halo (e.g., Rachford et al. 2002a; Lehner et al. 2003; Cartledge et al. 2004; Pan et al. 2004; Gillmon et al. 2006; Jensen & Snow 2007a,b; Rachford et al. 2009) toward distant stars or even extragalactic sources.

In contrast, the HI column densities are derived from both direct methods, which include Ly α absorption studies, EUV observations of stellar spectra, and observations of the 21cm line, and indirect methods, which include the curve of growth of neutral and singly ionized atoms and optical interstellar absorption lines of NaI which are both found to correlate with *N*(H) (de Boer et al. 1986; Ferlet et al. 1985). The column densities of HI, therefore, result from a combination of FUV and optical studies for nearby stars (e.g., Bohlin et al. 1978, 1983; Diplas & Savage 1994; Fitzpatrick & Massa 1990; Jensen & Snow 2007a,b) and radio studies of the 21cm line for extragalactic sources at high Galactic latitude (Wakker et al. 2003; Gillmon et al. 2006).

Finally, the color excess E(B - V) results from a compilation which includes direct measurements of the star reddening compared to its intrinsic $(B - V)_0$ color (e.g., Savage et al. 1977; Fitzpatrick & Massa 1990; Diplas & Savage 1994; Rachford et al. 2002a) and dust emission maps from the IRAS telescope¹ (Schlegel et al. 1998).

¹Available from the NED (http://nedwww.ipac.caltech.edu) and on the more recent plateform https://irsa.ipac.caltech.edu.



Figure 4.1: Distances and positions, in an aitoff projection in Galactic latitude and longitude coordinates, of the 360 background sources of the observational sample of HI and H_2 deduced from absorption studies. The color code indicates the distance of the source. The distances come from the Catalog GAIA, or from Gudennavar et al. (2012) when Gaia data are not available. All unknown distances correspond to extragalactic sources: these are arbitrarily set to 1 Mpc and indicated with black points.

Sources The distribution of the sources in the sky is shown in Fig. 4.1 in Galactic coordinates. With comparable numbers of observations in all Galactic quadrants, the sources appear to be well distributed in Galactic longitudes. Oppositely, and while the sources cover almost all Galactic latitudes, about two-thirds of them are located toward the Galactic disk with latitudes smaller than 15°, and only one third is located above, crossing the Galactic halo.

Finally, and as shown in Fig. 4.1, the background sources are a combination of distant galaxies (including quasars and Seyfert galaxies) and nearby stars located in a radius of ~ 3 kpc around the sun. With all these characteristics, the column densities of H and H₂ and the measurements of the reddening E(B - V) correspond to a random sample across the diffuse matter located in the vicinity of the sun in our Galaxy.

HI-to-H₂ **Transition** The compiled data are shown in Fig. 4.2 which displays the observed column densities of H₂ as functions of the total proton column densities of the gas $N_{\rm H} = N({\rm H}) + 2N({\rm H}_2)$. This famous picture displays one of the most important characteristics of the diffuse gas, its molecular fraction, and shows several interesting features. (1) As noted by Goldsmith et al. (2009), almost no line of sight is either purely WNM or purely molecular. This implies that the observed gas is necessarily composed of a combination of phases and clouds of different extinctions with various contributions to the volume spanned by the different lines of sight. (2) Random lines of sight across the local diffuse gas appear to cross a great variety of total column density which varies over about three orders of

magnitude over the entire sample. (3) The column density of H_2 measured in the Solar Neighbourhood appears to be highly dependent on the total column density: while N_H varies over three orders of magnitude, $N(H_2)$ varies over eight orders of magnitude. Finally, (4) the molecular fraction, computed as

$$f_{\rm H_2} = \frac{2N({\rm H_2})}{N_{\rm H}} \tag{4.1}$$

reveals a sharp evolution as a function of $N_{\rm H}$: starting from a value of ~ 10^{-5} for $N_{\rm H} < 10^{20}$ cm⁻², the molecular fraction suddenly "jump" by three to four orders of magnitude at $N_{\rm H} \sim 3 \times 10^{20}$ cm⁻² (Gillmon et al. 2006). Because of this sharp and bimodal evolution of the molecular fraction, Fig. 4.2 is often referred to as "the HI-to-H₂" transitions and is interpreted as a signature of a transition from atomic to molecular hydrogen, and more generally from atomic to molecular gas.

In addition to this chemical information, Fig. 4.2 also contains several statistical information. To highlight these statistical features, we divide the observational sample into 5 subsamples A, B, C, D, and E, as follows

A) $5 \times 10^{18} \le N_{\rm H} \le 5 \times 10^{19}$, $5.6 \times 10^{-8} \le f_{\rm H_2} \le 10^{-3}$

B)
$$5 \times 10^{19} \le N_{\rm H} \le 7 \times 10^{20}$$
, $5.6 \times 10^{-8} \le f_{\rm H_2} \le 10^{-8}$

C) $1.6 \times 10^{20} \le N_{\rm H} \le 10^{21}$, $10^{-3} \left(\frac{N_{\rm H}}{10^{20}}\right)^{0.9} \le f_{\rm H_2} \le 1$

D)
$$10^{21} \le N_{\rm H} \le 10^{22}$$
, $10^{-3} \left(\frac{N_{\rm H}}{10^{20}}\right)^{0.9} \le f_{\rm H_2} \le 10^{-3}$

E)
$$10^{22} \le N_{\rm H} \le 10^{23}$$
, $10^{-3} \left(\frac{N_{\rm H}}{10^{20}}\right)^{0.9} \le f_{\rm H_2} \le 1$

which encompass almost all the observational points and whose statistical properties are summarized in Table 4.1. 48 lines of sight out of 360 are found to be not detected in H₂. Most of these upper limits are obtained for a total column density smaller than 10^{21} cm⁻² and about half of them provide strong constraints on the molecular fraction with $f_{\rm H_2} \leq 10^{-5}$. 3% of the 312 detections belong to subsample A, 13% to subsample B, 16% to subsample C, and 65% to subsample D. Interestingly, the subsample E is empty and no line of sight is observed with $N_{\rm H} > 10^{22}$ cm⁻². While the mean value of the logarithm of the molecular fraction strongly increases from subsamples A to D, the dispersion simultaneously decreases by about a factor of three, probably revealing an effect of average over long distances.

4.1.2 Theoretical Interpretation and Open Question

Over the last decades, great efforts have been devoted to propose analytical descriptions of the HI-to-H₂ transition in homogeneous clouds with plane-parallel or spherical geometries (e.g., Sternberg 1988; Krumholz et al. 2008; McKee & Krumholz 2010; Sternberg et al. 2014 and references therein), compute this transition in detailed 1D chemical models assuming chemical equilibrium (e.g., van Dishoeck & Black 1986; Abgrall et al. 1992; Le Bourlot et al. 2012; Bron et al. 2014) or not (e.g., Lee et al. 1996; Goldsmith et al. 2007; Lesaffre



Figure 4.2: H₂ column density as a function of the total column density of protons $N_{\rm H}$. Open circles correspond to detections of H₂ while arrows correspond to upper limits. The blue dashed line indicates the maximum value of N(H₂) derived from a purely molecular medium with an integrated molecular fraction $f_{\rm H_2} = 1$ (Eq. 4.1). The red dashed-dotted line indicates the theoretical molecular fraction derived in an unshielded WNM-type environment with a density of 0.5 cm⁻³ and a temperature of 8000 K, illuminated by a UV photon flux of 10^8 cm⁻² s⁻¹ (see Eqs. 5.14 & 5.18). The regions A, B, C, D, and E defined in the main text correspond to an arbitrary separation of the observational sample.

et al. 2007), treat the chemistry of H and H_2 in subgrid models applied to simulations of galaxy formations (e.g., Gnedin et al. 2009; Christensen et al. 2012; Thompson et al. 2014; Diemer et al. 2018), or solve it in 3D isothermal or multiphase simulations of the diffuse ISM using various treatments of the radiative transfer (e.g., Glover et al. 2010; Valdivia et al. 2016b; Hu et al. 2016; Bialy et al. 2017; Nickerson et al. 2018).

These studies led to the following conclusions. At the scale of a homogeneous cloud, the molecular content, the sharpness of the HI-to-H₂ transition, and the asymptotic column density of HI are controlled by the ratio of the intensity of the UV field to the gas density and the dust-to-gas ratio, or equivalently, the metallicity (Sternberg et al. 2014). At larger scales, the integrated column densities of HI and H₂ also depend on the distribution of clouds of various densities along the line of sight and on the porosity to the UV radiation field. Because of these effects the total column densities of H and H₂ necessarily depend on the strength, the scale, and the compressibility of the turbulent forcing which control the clumpiness of the CNM clouds (Gnedin et al. 2009; Christensen et al. 2012).

Despite these achievements, very few works have been dedicated so far to the analysis

region	number	%	$\mu \left[\log(f_{\text{H}_2}) \right]$	$\sigma [\log(f_{\rm H_2})]$
А	10	3	-5.02	0.99
В	41	13	-4.62	0.70
С	50	16	-1.10	0.49
D	204	65	-0.70	0.37
Е	0	0	_	_

Table 4.1: Statistical properties of H₂ observations in the subsamples A, B, C, D, and E defined in the main text and shown in Fig. 4.2. Only the lines of sight where H₂ has been detected (312 sources out of 360 are considered. The mean and dispersion values, μ and σ are computed on the logarithm of the molecular fraction f_{H_2} observed in the corresponding subsample.

of the HI-to-H₂ transition in a turbulent multiphase medium at a scale sufficient to resolve the formation of CNM structures. In addition and while the predictions of analytical models (e.g., Krumholz et al. 2008) and simulations (Gnedin et al. 2009; Valdivia et al. 2016b) were able to reproduce the trend of the HI-to-H2 transition observed by Copernicus and FUSE in the local ISM (e.g., Savage et al. 1977; Gillmon et al. 2006; Rachford et al. 2009), the LMC and the SMC (e.g., Browning et al. 2003; Gillmon et al. 2006; Leroy et al. 2007), no detailed comparison with the statistical properties of these observations have been proposed. As a result, the occurrence of lines of sight with large molecular fractions predicted by numerical simulations often exceeds what is deduced from the observations (Valdivia et al. 2016b). Finally, and while statistical studies of 1D probability distribution functions (PDF) have become a common tool to understand the formation and the dynamics of molecular clouds (Körtgen et al. 2019), few statistical studies have been performed to date on 2D probability distribution functions using combined observations of different molecular tracers. In this thesis, we revisit the interpretation of the HI-to-H₂ transition in terms of probabilistic ordering of CNM clouds along lines of sight across a turbulent multiphase environment.

4.2 Neutral Carbon

Carbon is one of the most abundant elements in the ISM. Neutral carbon presents a finestructure splitting of the electronic ground level into three levels. The fine-structure excited levels are at energies of $E/k_B = 23.6$ and 62.4 K and can be excited and de-excited by collisions with the other species at the typical densities and temperatures of the diffuse ISM. Since the balance between the collisional excitations and radiative decays determines the fine-structure populations, the population ratios of CI are a diagnostic of the local physical conditions, i.e. the density and the temperature of the medium containing neutral carbon. In a multiphase gas, where the cold and warm phases are believed to be in thermal pressure equilibrium, it is capital to measure the pressure. CI is an excellent probe for that.

CI can be observed in absorption in the UV part of the spectrum. The first large-survey



Figure 4.3: Normalized intensity observed by Jenkins & Tripp (2001) exhibiting multiple lines from different fine-structure levels of CI. The ground state is indicated with CI, and CI* and CI* are the first and second excited levels, respectively. The original data normalized to a continuum is shown in black for the multiplets in the spectrum of HD 210839. The best fit model is shown in red.

on CI absorption lines in the UV was conducted with the *Copernicus* satellite by Jenkins & Shaya (1979) using hot stars as background sources. This early observational study was extented by later surveys (e.g. Jenkins et al. 1983; Smith et al. 1991; Jenkins & Wallerstein 1995; Jenkins et al. 1998; Jenkins & Tripp 2001), the largest of which was done by Jenkins & Tripp (2011). These last authors used a broad wavelength coverage in the UV spectrum of STIS - the Space Telescope Imaging Spectrograph on Hubble Space Telescope which records many wavelengths elements at once, and is able of measuring 10 different CI multiplets simultaneously. The CI multiplets consist of complex blends of individual lines from different fine-structure levels. An example is given in Fig. 4.3, where both the original data from Jenkins & Tripp (2011) (black curves) and their reconstructions (red) are shown. Fig. 4.3 shows almost direct measurements of the apparent optical depth as a function of the radial velocity. Their spectra have a resolving power of 2.6 km s⁻¹ (or 1.5 km s⁻¹ for the sources of Jenkins & Tripp 2001) in radial velocity. To be able to differentiate the stellar features, from those originating from the diffuse gas, they chose background sources with large radial velocities and they rejected background sources with small radial velocities, ending up with 460 independent velocity samples.

4.2.1 Observational Results

Sources The positions and distances of the 89 background sources reported by Jenkins & Tripp (2011) are shown in Fig. 4.4. Oppositely to the HI and H₂ survey presented in the previous section, most of the background sources are located toward the Galactic disk: quantitatively, 90% of the sources have a latitude smaller than 20°. The target stars of the 21-stars survey carried out by Jenkins & Tripp (2001) are clustered around Galactic longitudes of 120° and 300° because they were chosen in the continuous viewing zone² of HST, while the larger survey of Jenkins & Tripp (2011) is more evenly spread at all longitudes. The distances of the background sources range between 120 pc and about 3 kpc.

²The continuous viewing zone, CVZ, exclude the zone where Earth occultations happen



Figure 4.4: Aitoff projection in Galactic latitude and longitude coordinates, of the 89 background sources of the observational sample of CI deduced from the absorption studies of Jenkins & Tripp (2011). The color code indicates the distance of the source. Distances are derived from GAIA's measurements, when available, or taken from Jenkins & Tripp (2011). Black points indicate distances larger than 5 kpc.

The resulting set of lines of sight, therefore, samples the local diffuse gas, mostly within the Galactic Disk.

CI excited populations From the multiplet spectrum, Jenkins & Tripp (2011) computed the column densities of the three fine-structure levels of CI in different velocity channels. Using the column densities of the ground-state of the fine structure, N(CI), and the first and the second excited levels of the fine-structure of the neutral carbon, $N(CI^*)$, and $N(CI^{**})$, Jenkins & Tripp (2001) defined,

$$f1 = \frac{N(\text{CI}^*)}{N(\text{CI}_{\text{tot}})}$$
 and $f2 = \frac{N(\text{CI}^{**})}{N(\text{CI}_{\text{tot}})}.$ (4.2)

with $N(CI_{tot}) = N(CI) + N(CI^*) + N(CI^{**})$.

The resulting measurements of f1 and f2 in each velocity channel and for each line of sight are shown in Fig. 4.5. The solid lines indicate the theoretical curves obtained by computing the excitation of CI at equilibrium, taking into account radiative decay (Froese Fischer & Saha 1985) and collisions with atomic hydrogen (Launay & Roueff 1977), molecular hydrogen (Schroder et al. 1991), helium (Staemmler & Flower 1991), ionized hydrogen (Roueff & Le Bourlot 1990) and electrons (Johnson et al. 1987). For a given temperature *T*, the couple f1 and f2 follow an upward bent segment as the density increases until reaching the Boltzmann population at temperature *T*. Comparison of observations and theoretical prediction should, therefore, allow to extract the density and the temperature in any velocity bin along any line of sight.



Figure 4.5: From Jenkins & Tripp (2011): Measurements of f1 and f2 for all velocity bins (each having a width of 0.5 km s⁻¹) for the 89 background stars in the survey. The area of each dot is proportional to the value of $N(\text{CI}_{\text{tot}})$, with a normalization in size indicated on the top portion of the plot. The white X located at f1 = 0.209, f2 = 0.068 represents the "center of mass" of all of the dots. The curves indicate the expected level populations for three different temperatures, for a CNM cloud with a molecular fraction $f(\text{H}_2) = 0.6$ and a fraction of helium x(He) = 0.094 (more detail on the computation of the theoretical curve is given in §6). The large open circles on the curves indicate integer values of $\log(P/k_B)$ with accompanying numbers to indicate their values.

4.2.2 Theoretical Interpretation and Open Questions

The striking result of Fig. 4.5 is that none or few of the observational points follow the theoretical predictions, regardless of the density, temperature, or pressure of the gas. Not only most of the mass of CI is found to have an f2 ratio far larger than the predicted value but large variations of this ratio are observed with no apparent correlation with the f1 ratio. To solve this issue, Jenkins & Tripp (2011) interpreted this result as the presence along any line of sight of two kinds of regions: a low-pressure component ($P \sim 10^3 - 10^4$ K cm⁻³) contributing to most of the mass and volume of CI and a large-pressure component ($P \sim 10^5 - 10^7$ K cm⁻³) contributing to a small fraction of the mass and volume of CI. By analyzing each datapoint as a mixture of these two components, Jenkins & Tripp (2011) inferred the distribution of pressure in the low-pressure component rich in neutral carbon. They found that the pressure of this component weighted by the mass density, follows a lognormal distribution, with an average of $\log_{10} \overline{P} / k_B = 3.58$ K cm⁻³. As shown by Saury et al. (2014), this distribution of pressure can be reproduced by numerical simulation of the multiphase turbulent ISM assuming a turbulent forcing with a corresponding large-scale velocity dispersion in agreement with the 1D velocity dispersion derived from HI profiles seen in emission.

These remarkable findings raise, however, a major issue: what is the origin of the highpressure component? So far, several scenarios have been proposed (e.g. Jenkins & Tripp 2011): the high-pressure component could originate from regions of dissipation of interstellar turbulence or from interstellar shocks at high velocities propagating in the neutral diffuse gas; it could be linked to the background star itself; it could be due to the recoil of H atoms following the photodissociation of H₂ at the border of molecular clouds; or, finally, it could be the trace of gravitational collapse of molecular clouds along the line of sight. In addition to this astrophysical riddle, a major question has been left out from theoretical interpretation: can the statistical distribution of the abundances of CI and of its excited states displayed in Fig. 4.5 be explained by the current astrochemical models? In this thesis, we explore these questions by computing the chemistry and excitation of CI in turbulent simulations of the multiphase ISM.

4.3 CH⁺

Diatomic molecules were detected for the first time in the diffuse medium at the end of the 1930s. Interestingly, the methylidyne cation, CH⁺, was one of the first molecules to be observed (Douglas & Herzberg 1941). Oppositely to HI, H₂, and CI, CH⁺ is neither a tracer of the mass, the distribution, nor the pressure of the entire volume of the multiphase ISM. At first sight, the study of such species could therefore appear to be secondary. Such considerations would be a mistake. Indeed, because of its chemical properties, the presence of CH⁺ in the ISM is a deep mystery rooted in the evolution of the interstellar matter. In the diffuse gas, the destruction of CH⁺ occurs through hydrogenation or dehydrogenation, two exothermic processes with very short timescales ~1 yr (50 cm⁻³/ $n_{\rm H}$). The only formation pathway efficient enough to balance these fast destruction mechanisms is the hydrogena-

tion of C⁺, C⁺ + H₂ \rightarrow CH⁺ + H, which is a highly endothermic reaction ($\Delta E/k_B \sim 4620$ K). This simple argument sets the mystery of CH⁺: its formation requires H₂ which is mostly formed in the cold neutral medium, i.e. in regions where the temperature is far too cold to activate the above endothermic reaction. The very presence of CH⁺ in the diffuse gas challenges the validity of the predictions of astrochemical models at equilibrium. Even if it does not provide information on the global properties of the diffuse ISM, CH⁺ can therefore be a formidable tracer of out-of-equilibrium processes induced by turbulent advection or the dissipation of suprathermal energy sources. In this section, we present the set of lines of sight where the total column density and the column density of CH⁺ have been observed.

4.3.1 Observational Results

Sources The column density of CH⁺ across the local diffuse gas have been measured over the past 80 years through its UV absorption lines detected towards nearby OB stars (e.g. Federman 1982; Allen 1994; Crane et al. 1995; Gredel 1997; Gredel et al. 2002; Sheffer et al. 2008; Weselak et al. 2008). The distance and the position of the 190 background sources observed so far are shown in Fig. 4.6. Similarly to the observational sample of CI, about 90% of the lines of sight are directed towards the Galactic disk at latitudes smaller than 15°. In contrast, these lines of sight are well distributed in Galactic longitudes covering homogeneously all the Galactic quadrants. The distance of the background sources ranges between 50 pc and 3.2 kpc. It follows that the resulting set of lines of sight sample the local diffuse gas located in a radius of about 3 kpc around the sun with very few data at high Galactic latitude.

CH⁺ Column Densities The resulting sample is shown in Fig. 4.7 which displays the column density of CH⁺ as a function of the total column density of the gas $N_{\rm H}$. As for HI, H₂, and CI, these observations reveal several chemical and statistical information. The column densities of CH^+ are found to be proportional to N_H and to display a very large dispersion. The relative abundance of CH^+ , $N(CH^+)/N_H$, ranges over more than two orders of magnitude from 2×10^{-10} to 8×10^{-8} with an average of 8.9×10^{-9} . Interestingly, the relative abundance of CH⁺ are not stacked around the mean but appear to be homogeneously distributed in logarithmic space. Analysis of the fraction of non-detections also provides invaluable information. Indeed four regimes can be identified as a function of $N_{\rm H}$: (1) CH⁺ is systematically not detected for $N_{\rm H} < 1.7 \times 10^{20}$ cm⁻²; (2) for $N_{\rm H}$ between 1.7×10^{20} cm⁻² and 9×10^{20} cm⁻², there is an equipartition between non-detections and detections of CH⁺; (3) for $N_{\rm H}$ between 9×10^{20} cm⁻² and 2.1×10^{21} cm⁻², the percentage of non-detections drops to around 3%; and finally, (4) CH⁺ is systematically detected for $N_{\rm H} \gtrsim 2.1 \times 10^{21} \ {\rm cm}^{-2}$. This very peculiar distribution has strong similarities with the HI-to-H₂ transition presented in \$10. Because the formation of CH⁺ requires molecular hydrogen, the range of total column density at which CH⁺ starts to be detected corresponds to the range of column density at which the HI-to-H₂ transition occurs.



Figure 4.6: Aitoff projection, in Galactic longitude and latitude coordinates, of the background sources of the observational sample of HI, H_2 and CH^+ deduced from absorption studies. The color code indicates the distance of the source. Black points indicate sources at distances larger than 3.2 kpc.



Figure 4.7: CH⁺ column densities as a function of the total column density of protons $N_{\rm H}$. Red open circles are detections of CH⁺ and the green triangles are the upperlimits. The vertical dashed lines correspont to $N_{\rm H} \sim 1.7 \times 10^{20}$ cm⁻², 9×10^{20} cm⁻², and 2.1×10^{21} cm⁻², respectively. The red dotted line indicates the median of the relative abundance of $N({\rm CH^+})/N_{\rm H} \sim 5.6 \times 10^{-9}$.

4.3.2 Theoretical Interpretation and Open Questions

Because CH⁺ cannot be formed in cold environments, the predictions of static astrochemical models taking into account heating and chemical processes induced by UV photons only have been shown to systematically underestimate the abundance of CH⁺ by two to three orders of magnitude (see Fig. 4.8). To reconcile theoretical predictions and observations several scenarios have been invoked. One possible scenario is that CH⁺ traces the dissipation of a suprathermal energy source such as magnetized turbulence (Falgarone et al. 1995). In this scenario, the large abundances of CH⁺ result from the out-ofequilibrium chemistry triggered in coherent and intermittent dissipative structures such as magnetized shocks (Draine & Katz 1986; Pineau des Forêts et al. 1986; Lesaffre et al. 2013; Godard et al. 2019), magnetized vortices (Joulain et al. 1998; Godard et al. 2009) or magnetic reconnection (Xie et al. 1995). As shown by Godard et al. (2014), such a scenario could explain the observed abundances of CH⁺ providing that all the energy contained in the turbulent cascade is dissipated by ambipolar diffusion in small-scale dissipative structures (~100 AU, see Fig. 4.8). Another viable scenario is that CH⁺ could trace the turbulent mixing or the turbulent transport of warm and cold gas at the interfaces between the WNM and the CNM (Lesaffre et al. 2007; Valdivia et al. 2017). In this case, the formation of CH⁺ would not be activated by the dissipation of a suprathermal energy source but by the outof-equilibrium H₂ transported by turbulence from the CNM to the warm LNM (Lesaffre et al. 2007; Valdivia et al. 2017). Although this scenario was shown to underestimate by a factor of five the observed abundances of CH⁺ (Valdivia et al. 2017), no parametric study was performed on the subject.

Despite these pioneer works, the origin of CH⁺ is still unknown. Indeed, while all the scenarios given above provide viable solutions, none really reach an unshakable consensus. On the one hand, the models describing the dissipation of suprathermal energy are limited by the fact that the dynamical structures responsible for the dissipation of interstellar turbulence are still unknown. On the other hand, the scenario advocating for turbulent transport and turbulent mixing is based on numerical simulations whose parameter domains are difficult to explore and which are ultimately limited by the spatial resolution.

In this thesis, we revisit the interpretation of the formation of CH^+ through turbulent transport of out-of-equilibrium H_2 from the CNM to the WNM by performing a parametric study of numerical simulations and by using simultaneously the chemical and statistical information contained in the observations of CH^+ .

Warm H₂ A possible related issue is the presence of warm molecular hydrogen in the diffuse ISM. Indeed, since the 70s, observations of molecular hydrogen have shown that the population of H₂ in its excited rotational levels does not follow a Boltzmann distribution at the kinetic temperature *T* of the gas (Spitzer et al. 1973, 1974; Spitzer & Cochran 1973), indicating the presence of what is now called warm H₂. This warm H₂ was later observed in cold molecular gas (e.g., Gry et al. 2002; Falgarone et al. 2005). Intermediate and high-*J* levels of H₂ can only be effectively populated through UV pumping or in gas with *T* \gtrsim 300 K. Since the formation of CH⁺ is closely related to the abundance of molecular hydrogen, and requires high temperatures, the problems of warm H₂ and of the abundance of CH⁺



Figure 4.8: From Godard et al. (2014): Observations (blue and red open circles) are compared to the predictions of PDR (top panel) and TDR (bottom panel) models. The CH⁺ and total hydrogen column densities inferred from visible absorption lines and extinction measurements are shown in red and are from Crane et al. (1995); Gredel (1997); Weselak et al. (2008). The CH⁺ and total hydrogen column densities inferred from the far-infrared rotational transitions of CH⁺ observed with Herschel/HIFI instrument and the 21 cm emission lines of HI are shown in blue and are from Falgarone et al. (2010), and Godard et al. (2012). The PDR model predictions (filled symbols, top panel) are computed for several densities: 10 (crosses), 30 (triangles), 50 (squares), and 100 cm^{-3} (circles). The TDR models (filled symbols, bottom panel) are computed for two different dissipation rates $\epsilon = 10^{-24}$ (green) and 10^{-23} (magenta) erg cm⁻³ s^{-1} , see Godard et al. 2014).

could be linked.

4.4 Summary

The observables presented in the previous section, trace different physical and chemical aspects of the diffuse ISM. The hydrogen traces the total amount of matter, its molecular fraction, and its division into phases. The populations of the neutral carbon trace the pressure of the cold phases. Finally, CH⁺ traces dynamical events taking place in small volumes and warmer phases.

In this work, we explore the effects of local physical conditions (turbulence, magnetic field, thermal instability, and gravity) on the observational tracers presented in this section. Our goal is to address the following questions. What physical processes control the HI-to-H₂ transition, the abundance of CH^+ , and the abundance and excitation of neutral carbon? What are the origins of the variations from one line of sight to another of the different observables? Does the distribution of lengths of the lines of sight influence the observed statistical properties? Can the observed heterogeneity of lines of sight be accounted for by the joint action of turbulence and the thermal instability?

To answer these questions we perform a parametric exploration of numerical simulations (see §5) of the multiphase ISM and compare the results to observations with methods detailed in §8.1 and §9.2.

Part II

Modeling



Chapter 5

Simulations: RAMSES

In this Ph.D. thesis, we explore the physical processes at play in the diffuse neutral ISM using state-of-the-art numerical simulations. The goal is to perform, for the first time, a large parametric study of the evolution and steady-state of a large patch of the multiphase diffuse gas as a function of its mean density, its size, the nature and the strength of the turbulent forcing, the strength of the magnetic field, and the strength of the UV radiation field.

All the simulations explored below and obtained with RAMSES (Teyssier 2002; Fromang et al. 2006), a well-known grid-based solver with Adaptative Mesh Refinement (AMR, Berger & Oliger 1984), and a fully-treated tree (Khokhlov 1998) data structure. The tree data structure is at the root of RAMSES efficiency and flexibility. In a nutshell, the grid is divided recursively into groups of 8 cells (octs) sharing the same parent cell. Each one of these octs has access to its parent cell address in memory, but also to the six neighbors of the parent cell and to its own children octs. The fully-treated tree data structure, not only makes the code suitable for parallelization with the Message Passing Interface (MPI) library, but it also allows access to distant cells by walking the tree.

In the following sections, we will address different aspects of the interstellar medium and how they are integrated into the RAMSES code. These include the heating and cooling processes (see §5.2), the turbulence (see §5.3), and the chemistry of molecular hydrogen (see §5.4). In the last section of this chapter, we will present the range of parameters studied in this work and the steady-state nature of the biphasic diffuse gas.

The diffuse matter in the Solar Neighborhood of our Galaxy is simulated over a box of size *L* with periodic boundary conditions. The matter, defined by a mean proton density $\overline{n_{\rm H}}$, is assumed to be illuminated on all sides by an isotropic spectrum of UV photons set to the standard interstellar radiation field (Habing 1968) and scaled with a factor *G*₀, and turbulence is induced through an artificial forcing.

5.1 MHD: Conservation Laws

Because the diffuse gas is a low density, the collision rates are low compared to many spontaneous rates. If we consider scales much larger than the mean free path of the species composing the medium, we can describe the medium as a fluid. The mean free path is

$$\lambda = \frac{1}{\sigma n} \sim 10^{15} \left(\frac{n_{\rm H}}{1\,{\rm cm}^{-3}}\right)^{-1} \,{\rm cm},\tag{5.1}$$

For a cross section of an atom or a small molecule of $\sigma = \pi (2 \cdot a_0)^2 \sim 10^{-15}$ cm² with $a_0 = 5.29 \times 10^{-9}$ cm, the Bohr radius of the hydrogen atom, we obtain $\lambda \sim 1$ AU in dense clouds, and 100 AU in the WNM¹. These lengths are much shorter than the typical scales of interest in our study and thus the continuum hypothesis holds. Since the condition of the fluid approximation is verified ($\lambda \ll L$), we can define macroscopic quantities like the mass density (ρ), velocity (\vec{v}), temperature (T) and pressure (P) and use fluid equations (Navier-Stokes equations) and magnetohydrodynamics equations to describe the ISM dynamics.

An important dimensionless number in hydrodynamics is the Reynolds number (Reynolds 1895). This parameter is the ratio between the advection and the viscosity term in the moment conservation equation

$$Re = VL/\nu \tag{5.2}$$

where *V* is the characteristic velocity of the flow, *L* the largest scale dynamically relevant, and *v* the kinematic viscosity. When the Reynolds number is small, Re \ll 1, the fluid is dominated by viscous stresses. As the Reynolds number increases, the flows remains laminar for a wide range² of value Re. When Re \gg 1, the advection term dominates over the viscous term and the flow is turbulent. It is important to notice that the advection term is highly non-linear and that the bigger it gets the more chaotic is the flow pattern. This case corresponds to a turbulent flow, characterized by strong velocity gradients and a partition of the kinetic energy on different spatial scales.

Since the interstellar medium is magnetized, the dynamical problem is described by magnetohydrodynamical (MHD) equations. In all its complexity, MHD should be treated with a multifluid approach, taking into account possible decoupling between charged and neutral particles and resistive processes. Unfortunately, such an approach would require a range of spatial and temporal scales far beyond the capabilities of modern computers.

To simplify the problem, RAMSES solves the equations³ of ideal MHD, an approximation which holds as long as the scales considered are far larger than the decoupling and resistive scales. In their conservative forms, those equations include the mass conservation,

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \, \vec{v}) = 0, \tag{5.3}$$

the momentum conservation,

$$\frac{\partial \rho \vec{v}}{\partial t} + \nabla \cdot (\rho \vec{v} \vec{v} - \vec{B} \vec{B}) + \nabla P = -\rho \nabla \Phi + \rho \vec{f}, \qquad (5.4)$$

the energy conservation,

$$\frac{\partial E}{\partial t} + \nabla \cdot \left[(E+P)\vec{v} - \vec{B}(\vec{B} \cdot \vec{v}) \right] = -\rho \vec{v} \cdot \nabla \Phi + \rho \vec{f} \cdot \vec{v} - \rho \mathscr{L}, \tag{5.5}$$

²The range depends on the geometry of the system

³These MHD equations are written in the code in a condensed form, using the definition of the tensor product as $\overline{\overline{A}} = \vec{B} \vec{v}$, where $\overline{\overline{A}}$ is a second rank tensor, also called dyatic tensor, $A_{ij} = B_i v_j$. Its flux is $\nabla \cdot \overline{\overline{A}} = \partial_i A_{ij}$.

¹As a comparison, the meam free path in air is $\lambda_{air} \sim 1 \ \mu m$

 $^{1 \}text{ AU} \sim 1,496 \times 10^{13} \text{ cm}$

and the induction equation,

$$\frac{\partial \vec{B}}{\partial t} + \nabla \cdot (\vec{v}\vec{B} - \vec{B}\vec{v}) = 0, \qquad (5.6)$$

$$\nabla \cdot \vec{B} = 0. \tag{5.7}$$

Where ρ , \vec{v} , \vec{B} , *P* and *E* are the mass density, the velocity field, the magnetic field, the total pressure, and the total energy density, respectively. \mathscr{L} is the net cooling function per unit mass (see §2), and *f* is the acceleration due to the turbulent driving (see §5.3).

The axis *x*, *y*, and *z* are chosen so that *z* corresponds to the direction perpendicular to the Galactic disk, and *x* corresponds to the direction of the mean magnetic field initially parameterized by a constant value of B_x .

Gravity To take all gravitational forces into account, including self-gravity and the action of stars and dark matter, the gravitational potential Φ is divided into two terms:

$$\Phi = \phi_{\text{gas}} + \phi_{\text{gal}}.$$
 (5.8)

The self-gravity potential, ϕ_{gas} , is deduced from the Poisson's equation:

$$\nabla^2 \phi = 4\pi G \rho. \tag{5.9}$$

Following Kuijken & Gilmore (1989) and Joung & Mac Low (2006), we assume that the Galactic potential along the direction *z* perpendicular to the Galactic disk can be written as

$$\phi_{\text{gal}}(z) = a_1 \left(\sqrt{z^2 + z_0^2} - z_0 \right) + 2a_2 z^2,$$
(5.10)

where the first term is the contribution of the stellar disk parametrized by $z_0 = 0.18$ kpc and $a_1 = 1.42 \times 10^{-3}$ kpc Myr⁻² and the second term is the contribution of the spherical dark halo parametrized by $a_2 = 5.49 \times 10^{-4}$ Myr⁻².

Diffusion In the above equations of conservation, there are no terms of diffusion included, either thermic, i.e. conduction, or of momentum, i.e. viscosity. This implies that the diffusion in these simulations is entirely due to the numerical scheme. The scales at which the diffusion takes place are much smaller than the ones resolved by the problem, and therefore the numerical diffusion takes place at scales bigger than the physical scale at which viscosity and conduction act.

5.2 Thermodynamics

As shown by Field (1965), the ISM has a multiphasic nature. This results from the thermal balance of the gas and thus from its net cooling function \mathcal{L} defined by

$$\rho \mathscr{L} = n_{\rm H}^2 \Lambda - n_{\rm H} \Gamma \tag{5.11}$$

where $n_{\rm H}^2 \Lambda$ and $n_{\rm H} \Gamma$ are the cooling and heating rates of the medium (in erg cm⁻³ s⁻¹) and $n_{\rm H}$ the proton density (in cm⁻³).

To correctly describe the thermal state of the diffuse ISM, we include the heating induced by the photoelectric effect and the decay of cosmic ray particles and the cooling induced by Lyman α photons, the recombination of electrons onto grains, and the finestructure lines of OI and CII. All these processes, described in §2, are modeled with the analytical formulae given by Wolfire et al. (2003). A few tests to validate the model are presented in the next section, and a comparison with (Wolfire et al. 2003) is shown in Fig. 5.3.

5.2.1 Radiative Transfer

The extinction of UV photons has a direct impact on the dynamics of the medium because it affects the heating. To describe the interaction of the radiation and the matter, a treatment of radiative transfer is, therefore, necessary. However, a full radiative transfer would be too expensive numerically. Several methods have been proposed to tackle the problem (e.g. Glover et al. 2010; Valdivia et al. 2016; Hu et al. 2016; Bialy et al. 2017; Nickerson et al. 2018). Here we present the tree-based method used in RAMSES and implemented by Valdivia & Hennebelle (2014).

To access the effective radiation field G_{eff} in each cell of our simulation we need to compute the absorption of UV photons by dust. With the tree-based method, we can compute the column densities around each cell of the simulation, using the information already stored in the tree structure of RAMSES code. This approach has the advantage of being fast numerically because the column densities are calculated while the tree is walked. This approach is also key for the calculation of the formation and destruction of molecular hydrogen in the simulation, in which the molecular hydrogen is subject to self-shielding and, thus, the column densities of H₂ around a point affect its destruction rate.

The column densities $N_{\rm H}$ (and $N({\rm H_2})$ for ${\rm H_2}$ self-shielding presented in §5.4.3) are estimated by adding all the contributions of the cells along each line of sight with decreasing resolution as a function of the distance with the general assumption that distant cells have a smaller contribution to the local extinction.

At each point the effective radiation field G_{eff} (in Habing units) is computed as

$$G_{\rm eff} = G_0 \langle e^{-2.5A_V} \rangle, \tag{5.12}$$

where A_V is the visual extinction along a given ray, deduced from the integrated proton column density⁴ computed from the border of the box to the current point

$$A_V = 5.34 \times 10^{-22} \left(\frac{N_{\rm H}}{\rm cm^{-2}}\right),$$
 (5.13)

and $\langle e^{-2.5A_V} \rangle$ is an average performed over 12 directions, treated as solid angles evenly spread in polar coordinates.

⁴In this work we use the relation between $N_{\rm H}$ and A_V deduced from the observations of the mean Galactic extinction curve (Fitzpatrick & Massa 1986).



Figure 5.1: Thermal equilibrium curves ($\mathcal{L} = 0$) on the $P/k_B - n_H$ space. Implicit functions $\Gamma = n_H \Lambda$ are computed taking or not taking into account different terms of cooling (dashed lines) and heating (dotted line). Isothermal curves are indicated in dotted gray lines (at 80000 K, 8000 K, 800 K, 80 K, from left to right).

5.2.2 Testing Heating and Cooling Processes

In \$2, we presented the main thermal processes occurring in the neutral diffuse ISM. In this section, we are analyzing how the inclusion or exclusion of some of these different processes in the thermodynamics changes the net cooling function. In Fig. 5.1, we show the thermal equilibrium curve, i.e. $\mathcal{L} = 0$, when different heating and cooling processes are excluded. This graphic allows us to access the processes responsible for the heating and cooling in different pressure and density regimes. The main heating process is the photoelectric effect onto grains. When this process is excluded (orange dotted curve) the bistability of the curve disappears. The medium would not, in such a case, present a warm and a cold phase in the typical conditions of pressure of the diffuse ISM. The main cooling process at high temperatures is the Lyman α line emission (turquoise dashed line), followed by the metastable lines of oxygen and carbon. At densities $n_{\rm H} \sim 0.5 - 3$ cm⁻³, the cooling is dominated by the fine-structure lines of oxygen. At higher densities, the main coolant is the carbon line emission (purple-blue dashed line). Without this cooling process the range of pressure over which the gas is biphasic shrinks, and the stable cold branch lies at higher temperatures. Other processes like cosmic ray heating and electric recombination have minor effects on the thermal equilibrium curve, when the other processes are included, affecting mainly the unstable part of the curve.



Figure 5.2: Net cooling function, $\mathcal{L} = 0$ or equivalently $n_{\rm H}^2 \Lambda = n_{\rm H} \Gamma$, in $P/k_B - n_{\rm H}$ space, traced for different values of the parameters at play in the different heating and cooling processes, e.g. $x_{\rm O}$, $x_{\rm C^+}$, G_0 , A_V , $\phi_{\rm PAH}$, $\zeta_{\rm CR}$. The black curve indicates the value used for all RAMSES simulations unless indicated otherwise in the main text.

It is also important to understand how certain assumptions made on the parameters at play in the thermodynamics, like ϕ_{PAH} , ζ_{CR} , x_{C^+} , x_0 , affect the thermal processes considered (see Fig. 5.2).

Since **oxygen** line emission is the main cooling at temperatures of a few hundred kelvins, its abundance influences the thermal equilibrium curve in the intermediate unstable branch, without affecting the stable branches. Oxygen presents two excited levels in the fine-structure, one at 228 K and the other at 326 K. In the second panel of Fig. 5.2, we show how the inclusion of the second level increases the total cooling rate. The thermal equilibrium curve lowers slightly towards lower pressures.

C⁺ line emission is an important coolant at intermediate temperatures and becomes the main coolant at lower temperatures because the excited level of its fine structure is at 92 K. Therefore, its abundance affects the cold stable branch. It is interesting to notice that, as for the oxygen, an increase in the abundance of this coolant lowers the equilibrium curve, but oppositely to the oxygen, this corresponds to an increase in the slope of the unstable branch, and consequently an increase of the pressure range over which the gas is bistable. For the oxygen the opposite happens, because the oxygen affects more the low-density end of the unstable branch lowers, while the high-density end is less affected.

The inclusion of **metastable**-states of O and C⁺ has a slight impact on the high temperature range. Both metastable states and Lyman α lines have the effect of lowering the low density (high temperature) stable branch of the equilibrium curve, but since Lyman α dominates the cooling in this range of parameters (see Fig. 5.1), the impact of the metastable state on the total net cooling function is negligible and thus ignored in our study.

Wolfire takes values of the **PAH recombination parameter** ϕ_{PAH} between 0.25, which is considered low PAH recombination efficiency, and 1.0, which is considered high PAH recombination efficiency. In Fig. 5.2 we show the effect of $\phi_{\text{PAH}} = 0.25, 0.5$ and 1 on the equilibrium curve. The impact of this parameter is to slightly reduce the pressure range (see Fig. 2.4) in which there is a coexistence of different stable phases, and raising this range towards larger pressures. As a recombination parameter, ϕ_{PAH} plays a role in the efficiency of the photoelectric effect. Higher ϕ_{PAH} implies less positively charged PAH, and in such a case the photoelectric efficiency increases, increasing the heating of the medium.

In Wolfire et al. (2003) (dashed blue line), the electronic density does not include the electrons coming from the ionization of the carbon atom. In contrast, the prescription we included in RAMSES code does (solid black line). This slightly affects the curve mainly at high densities, where a higher electronic density increases the efficiency of the photoelectric effect (see eq. 2.2).

As the reference value for the **cosmic ray** ionization rate, we take a fiducial value of $\zeta_{CR} = 10^{-16} \text{ s}^{-1}$ (see §1.3.4). An increase in the CR ionization rate results in an increase of the heating rates. ζ_{CR} not only affects CR heating rate, but also the efficiency of the photoelectric effect through its effect on electronic densities.

As shown by the last two panels of Fig. 5.2, the thermal equilibrium curve mainly depends on the local illumination of the gas by the UV **radiation field**. When the UV radiation field is weaker, i.e. G_0 is smaller, or the extinction is stronger, i.e. A_V is higher, the photoelectric effect onto grains decreases and the equilibrium curve moves towards lower

pressures.

In conclusion, even though it is important to keep in mind how each parameter affects the thermodynamics of the medium, most of the parameters presented here have little effect on the overall equilibrium curve compared to the strength of the UV radiation field. The intensity of the radiation field, G_0 , will be taken as a parameter that we can vary in our work, and the extinction, A_V , is a variable that changes from point to point in the simulation, as explained in §5.2.1.

In Fig. 5.3, we compare the thermal equilibrium curve computed with the processes of heating and cooling included in RAMSES code (solid lines) and the curve from Wolfire et al. (2003) (dashed lines). With no extinction included in the computation of the equilibrium curve (red solid line) and $\phi_{PAH} = 0.5$ (blue dashed lines), the two curves are almost in perfect agreement. A slight difference is present at high densities. As explained above, this difference is due to the different computation of the electron density (see Fig. 5.2). It is important to note that the analytical prescriptions used in RAMSES assume a constant abundance of the carbon ion C⁺, regardless of the environment. This is a shortcoming of this approach, because in dense regions carbon is more likely in its molecular form, CO, and the abundance of the carbon ion is reduced. Thus, on the one hand, we are overestimating the cooling due to C⁺, due to the overestimation of its abundance. On the other hand, CO is an important coolant of the medium in denser regions, and we do not include its cooling rate in RAMSES.



Figure 5.3: Thermal equilibrium curves ($\mathcal{L} = 0$) computed with RAMSES for $A_V = 0$ (red solid curve) and $A_V = 1$ (green solid curve) compared to those predicted by Wolfire et al. (2003) for $\phi_{PAH} = 0.5$ (black dashed curve) and 1.0 (blue dashed curve). A small difference is seen between RAMSES and Wolfire et al. (2003) in the high-density branch. This difference is due to the different computation of n_e in which we take into account the fact that in diffuse molecular clouds C⁺ is the main source of electrons. This difference in n_e affects both the recombination cooling and the photoelectric heating.

5.3 Turbulence

"According to an apocryphal story, Werner Heisenberg was asked what he would ask God, given the opportunity. His reply was: "When I meet God, I am going to ask him two questions: Why relativity? And why turbulence? I really believe he will have an answer for the first."

5.3.1 Overview on Theory

Molecular clouds show supra-thermal line widths. This non-thermal broadening is attributed to the turbulent motions of the gas. Turbulence is one of the greatest challenges in physics, that has been studied for centuries (e.g. Leonardo da Vinci). In 1922, Richardson proposed the concept of turbulent cascade: *"We realize that big whirls have little whirls that feed on their velocity, and little whirls have lesser whirls and so on to viscosity."* This theory was later developed by Kolmogorov (1941) who formalized it for an incompressible flow, under the hypotheses that the small-scale turbulent motions are statistically isotropic and that the statistics of those small scales are entirely determined by the kinematic viscosity and the rate of energy dissipation.

Turbulence has a heuristic definition which defines it as an instability of the laminar flow. This instability develops when the inertial forces, i.e. advection, become larger than the viscous ones. The ratio bewteen the advection and the viscous term is given by a dimensionless number, the Reynolds number (see eq. 5.2). The ISM has very high Reynolds numbers (Re $\sim 10^5 - 10^7$), meaning that the advection term in the momentum conservation is very large, i.e. the medium is highly turbulent, and that the viscosity is negligible at large scales.

Turbulence is a multiscale phenomenon that transfers energy from large scales (big eddies) to small scales (very small eddies) and it plays a central role in the evolution of the ISM and the structure formation. In the ISM, there are many sources of mechanical energy to which the medium is subject. As seen in §1.3, the mechanical energy is an important energy reservoir of the medium, with roughly the same energy density as the other energy reservoirs. Several sources can inject mechanical energy in the ISM: supernovae (Padoan et al. 2016a,b; Pan et al. 2016; Körtgen et al. 2016), stellar winds, superbubbles, expanding HII regions (Matzner 2002; Krumholz et al. 2006; Goldbaum et al. 2011), the differential rotation of the galaxy, gravitational instabilities (Federrath et al. 2011; Sur et al. 2012), jets and outflows (Li & Nakamura 2006; Wang et al. 2010; Federrath et al. 2014), and accretion (Klessen & Hennebelle 2010; Vázquez-Semadeni et al. 2010; Lee & Hennebelle 2016).

The mechanical energy is injected at large scales, and different energy sources can have different energy spectra, i.e. inject energy at different scales (Norman & Ferrara 1996). This injected energy is transferred to lower scales through the turbulent cascade until it reaches scales small enough so that the energy dissipates through ambipolar friction, viscosity, or resistivity. Many theoretical efforts have been done to characterize and describe turbulence and the turbulent cascade, and both are still debated.

In numerical simulations, there are different methods to inject turbulence. Some in-

clude the source of the mechanical energy at large scales, e.g. SN feedback. Others mimic the turbulence through a mathematical prescription of a forcing term in the momentum equation.

In this work, we adopt a pseudo-spectral approach to mimic the turbulence. Oppositely to converging flows studied by Audit & Hennebelle (2005) and Valdivia & Hennebelle (2014), this prescription allows to conserve the mass in the simulations to directly compare our results to other works, such as Seifried et al. (2011) and Saury et al. (2014), and study the influence of the injection scales, frequencies, and the relative power in solenoidal and compressive modes on the structure of the matter and in the chemical evolution.

5.3.2 Implementation in RAMSES: Turbulent Forcing

To mimic the injection of mechanical energy in the diffuse ISM, a large-scale turbulent forcing is applied. Following Schmidt et al. (2009) and Federrath et al. (2010), this forcing, modeled by an acceleration \vec{f} in the momentum conservation equation, is driven through an Ornstein-Uhlenbeck process using a pseudo-spectral method. At regular time intervals $\Delta \tau$, random fluctuations of the forcing term are generated and applied over an autocorrelation timescale τ . To excite only large-scale modes, the forcing is modeled as a paraboloid in Fourier space covering a small interval of wavenumbers $1 \le k \le 3$ and centered on k = 2. Using the notations of Seifried et al. (2011) and Saury et al. (2014), the total magnitude of these perturbations is set with either an acceleration parameter F or, equivalently, a velocity parameter V related by $F = V^2/L_{drive}$, where L_{drive} is the main driving scale, $L_{drive} = L/2$. A Helmholtz decomposition is finally applied, in order to control the powers injected in compressive and solenoidal modes. Using the classical notation, these powers are set with a parameter⁵ ζ ranging from a pure solenoidal field ($\zeta = 1$) to a pure compressive field ($\zeta = 0$).

5.4 Molecular Hydrogen

In addition to the thermodynamical properties described above, RAMSES also computes the chemistry of molecular hydrogen. The timescale required for the abundance of molecular hydrogen to reach its equilibrium value is known to range over several orders of magnitude, depending on the physical conditions of the ISM (e.g., Goldsmith et al. 2007; Tabone 2018). In the diffuse gas, this timescale varies typically between a few 10^3 yr and a few 10^7 yr (Valdivia et al. 2016b), hence over a range of values that often exceeds the dynamical timescales. To take this important aspect of the diffuse interstellar chemistry into account, the out-of-equilibrium abundance of H₂ is computed self-consistently in the simulation, using the formalism introduced in RAMSES by Valdivia et al. (2016b).

In this section, we will introduce the main formation and destruction processes, and the self-shielding of H_2 against photodissociation.

⁵We note that for $\zeta = 0.5$, the power of the compressive forcing corresponds to 1/3 of the total power.

5.4.1 Formation of H₂

In this section we will first address the different processes that can lead to the formation of molecular hydrogen, starting from the gas-phase reactions.

Radiative Association The formation of H₂ by radiative association,

$$H + H \rightarrow H_2 + h\nu$$
,

has an extremely small rate coefficient. This is due to the fact that the excess of energy of the two initial atomic hydrogens can only be radiated away through a quadripole transition, because molecular hydrogen is homonuclear, hence has no permanent dipole.

H⁻ **Route** One catalytic way to form H₂ is the H⁻ route that happens through a radiative attachement, $H + e^- \rightarrow H^- + hv$, followed by an associative detachment, $H^- + H \rightarrow H_2 + e^- + KE$, which is an exothermic reactions. Since ion-neutral reactions (e.g. H⁻ associative detachment) have bigger rates than neutral-neutral reactions (e.g. H radiative association), this route is more efficient in forming molecular hydrogen than the previous one. Nevertheless, H⁻ formed by the radiative attachment can be destroyed by a photodetachment, H⁻ + $hv \rightarrow H + e^-$, which has a rate proportional to the external radiation field, and therefore predominant in dense and dust-poor PDRs (Tabone et al. 2020). This route is then negligible in the diffuse medium under study in this work.

Three-body Association Three-body association (Palla et al. 1983) is another mechanism, $H+H+C \rightarrow H_2 + C$, where C is a third body that can be hydrogen, molecular hydrogen, helium, or another abundant species of the ISM. In this formation process, the excess of energy is carried out by a third atom. The density at which three-body collisions can be important is much higher than the densities of the diffuse ISM we are modeling here, and this formation route will be neglected in our studies.

Grain Surface Catalysis The dominant process of H_2 formation in our Galaxy is through dust-grain catalysis. The dust-grain surface acts like a catalyser for the formation of molecular hydrogen. There are two-pathways of this process: the physisorption (or the Langmuir-Hinshelwood mechanism) in which a hydrogen atom is weakly bonded and able to wander on the surface of the grain until it encounters another physisorbed hydrogen atom, and the chemisorption (the Eley-Rideal mechanism), in which the atomic hydrogen is chemically bonded to the surface of the grain and unable to move, but can react with a hydrogen atom in the gas phase that collides with the grain.

The formation rate of these mechanisms depends on the rate of collisions between atomic hydrogen and dust grains, and on the probability of an atomic hydrogen to stick to the surface of the dust grain. The formation of H_2 onto grains in physisorption and chemisorption sites is the main formation processes of H_2 in the diffuse ISM and it is modeled in RAMSES with the simplified rate of Le Bourlot et al. (2012)

$$k_f = 3 \times 10^{-17} n_{\rm H} n({\rm H}) \sqrt{\frac{T}{100 \,{\rm K}}} S(T) \,{\rm cm}^{-3} {\rm s}^{-1},$$
 (5.14)

where $n_{\rm H}$ and $n({\rm H})$ are the local proton and atomic hydrogen densities, and

$$S(T) = \frac{1}{1 + (\frac{T}{T_2})^{\beta}}$$
(5.15)

is the sticking coefficient of H onto grain, parametrized by $T_2 = 464$ K and $\beta = 1.5$ (Bron et al. 2014).

5.4.2 Destruction of H₂

Photodissociation of H_2 In the diffuse medium and PDRs, the principal destruction process is photodissociation,

$$H_2 + h\nu \rightarrow H + H.$$

The photodissociation of H_2 can be direct, i.e. one molecular hydrogen is hit by a photon and dissociates in two atomic hydrogens, or it can be a two-step process, where the molecular hydrogen is excited by a photon and then it decays to the vibrational continuum.

The direct photodissociation from *X* ground state to *b* unbound state $(H_2 \rightarrow H(1s)+H(1s))$ is highly forbidden by selection rules ($\Delta S = 0$), while the direct dissociation from *X* electronic ground state to *B* or *C* vibrational continua requires a minimal energy of 14.7 eV. Since the diffuse ISM is rich in atomic hydrogen, whose ionization potential 13.6 eV, this photodissociation is negligible because the medium is not rich in photons with hv > 14.7 eV capable of directly photodissociating H₂.

The two-step photodissociation process, happens through a first step in which the molecular hydrogen absorbs a UV photon and goes from its ground electronic state to the Lyman band (if hv > 11.2 eV, i.e. photons with $\lambda < 1108$ Å, see Fig. 7.2) or the Werner band (if hv > 12.3 eV, i.e. photons with $\lambda < 1008$ Å, see Fig. 7.2), and a second step that can lead to a transition/decay to a vibrational excited level (~ 85-90% of the cases) of H₂ ground state or to the vibrational continuum of the ground state (~ 10-15% of the cases), leading to the dissociation of H₂. Both these 2nd steps imply an emission. Since the diffuse ISM is rich in UV photons, this photodissociation process is the main one in our domain of study.

Photoionization Another process that destroys hydrogen molecules is the photoionization,

$$H_2 + h\nu \rightarrow H_2^+ + e^-.$$

Since the photoionization threshold for molecular hydrogen is 15.4 eV (see Fig. 7.2), photons with hv > 15.4 eV can ionize H₂, destroying it. The diffuse ISM is opaque to such photons, and this destruction route of H₂ is negligeble in our study.

Collisional Dissociation of H₂ H₂ can also be destroyed by collisions with other species. Dalgarno & Roberge (1979) showed that the collisional dissociation can happen at both high temperatures and high densities (e.g. T > 4000 K and $n_{\rm H} > 10^4$ cm⁻³). Thus, this type of dissociation is negligible in the diffuse interstellar medium.

5.4.3 H₂ Shielding

Since the main process of destruction of H_2 is photodissociation by UV photons, the absorption and scattering of UV photons is a key ingredient in the computation of molecular hydrogen. The extinction of UV photons can lower the efficiency of the photodissociation of H_2 in certain regions, leading to an increase of molecular hydrogen. The UV radiation field is attenuated by dust, as explained in §3.2, but also by molecular hydrogen itself. H_2 molecules absorb UV photons in lines. These molecules are, thus, sources of opacity for photons at those wavelengths. In the inner regions of a cloud, the external radiation field is attenuated by all of the H_2 molecules in the outer regions of the cloud, leading to the so-called 'self-shielding' of H_2 .

Different methods exist to compute the shielding of H_2 due to dust and to molecular hydrogen itself. While an exact radiative transfer can be computed in PDR codes, this approach is very expensive numerically. As a consequence other approximative approaches are often used, such as the FGK transfer (Federman et al. 1979), which considers the absorption at the line center to compute the radiative transfer. In this work we use an analytical prescription to compute the self-shielding of H_2 (Draine & Bertoldi 1996):

$$f_{\text{shield}}(x) = \frac{0.965}{(1+x/b_D)^2} + \frac{0.035 \ \mathrm{e}^{-8.5 \times 10^{-4} \sqrt{1+x}}}{\sqrt{1+x}},$$
(5.16)

where

$$x = N(H_2)/5 \times 10^{14} \text{ cm}^{-2}$$
(5.17)

and b_D is the Doppler broadening parameter expressed in km s⁻¹. This prescription needs the amount of H₂ traversed by UV photons to compute an estimate of the self-shielding. In RAMSES, this quantity is computed with a tree-based method developed by Valdivia & Hennebelle (2014) to provide a fast estimate of H₂ column densities along 12 different directions. These amounts are then averaged to obtain $\langle f_{\text{shield}}(x) \rangle$ and compute the photodissociation rate of Eq. 5.18. Valdivia & Hennebelle (2014) method uses solid angles to compute the intensity of the incident external radiation field in each point in the simulation, taking into account the attenuation due to dust, $e^{-\sigma_d N_H}$, and due to the self-shielding coefficient, f_{shield} . For the shielding induced by dust, we adopt here an effective dust attenuation cross section at $\lambda = 1000$ Å, $\sigma_d = 2 \times 10^{-21}$ cm² (Sternberg et al. 2014).

Following all these considerations, the photodestruction rate of H₂ is modeled as

$$k_d = k_d^0 n(H_2) G_0 \langle e^{-\sigma_d N_H} \rangle \langle f_{\text{shield}}(x) \rangle \text{ cm}^{-3} \text{s}^{-1},$$
(5.18)

where $k_{d,0} = 3.3 \times 10^{-11} \text{ s}^{-1}$ is the inverse free-space dissociation timescale of H₂ in an isotropic Habing field (Glover & Mac Low 2007), $n(H_2)$ is the local density of the molecular hydrogen, $e^{-\sigma_d N_H}$ is the shielding induced by dust and f_{shield} the self-shielding function.

Note that the prescription of H₂ self-shielding used in this work (Eq. 5.16) is taken from Draine & Bertoldi (1996). As discussed by Wolcott-Green et al. (2011), such a prescription is reliable for diffuse gas at low temperature but becomes less and less reliable for high-temperature environments (T > 500 K) where efficient collisional excitation of H₂ in its rovibrational levels reduces the self-shielding. In Fig. 5.4, we compare the analytical



Figure 5.4: Comparison of different prescriptions of the self-shielding of H₂ as function of the molecular hydrogen column density. The results from the table of Lee et al. (1996) are shown with the blue curve, the analytical formula given by Draine & Bertoldi (1996) used in RAMSES (see eq. 5.16) is shown with the orange curve and the analytical formula of Wolcott-Green et al. (2011) is shown in purple. The curves are computed assuming $b_D = 8 \text{ km s}^{-1}$, a value discussed in the following section.

prescription of Wolcott-Green et al. (2011) (purple line) to Draine & Bertoldi (1996) formula used in RAMSES (orange line). In logarithmic scale, Wolcott-Green's self-shielding is smoother. Both prescriptions give the same results asymptotically. At $N(H_2)$ between a few 10^{15} cm⁻² and 10^{19} cm⁻², Draine's formula underestimates f_{shield} . The maximum difference between the two prescriptions is at $N(H_2) \sim 10^{17}$ cm⁻², where they almost differ by an order of magnitude. As shown in §10.1.1, using Wolcott-Green et al. (2011) prescription slightly affect the global results on the HI-to-H₂ transition analysis. However, it may affect more profoundly the results on the chemistry of CH⁺.

5.4.4 Doppler parameter

The chemical species composing the ISM have thermal motions and turbulent motion that result in a Doppler broadening of their line profiles. The Doppler broadening parameter b_D characterizes the widths of line profiles in km s⁻¹

$$b_D = \sqrt{u_{\rm th}^2 + u_{\rm turb}^2} \tag{5.19}$$

where u_{turb} is the turbulent broadening. The self-shielding of H₂ depends on the dispersion of the Lyman and Werner lines which is usually modeled with a turbulent Doppler broadening parameter b_D (Draine & Bertoldi 1996 and Eq. 5.16). In single cloud models, this parameter has the effect of shifting the HI-to-H₂ transition toward larger total column density without modifying any of the asymptotic states (e.g., Bialy et al. 2017). In this work, we identify this parameter with the turbulent velocity dispersion of the gas at large scales and therefore adopt $b_D = 8 \text{ km s}^{-1}$ for the fiducial simulation. This is done to prevent an overestimation of the H₂ self-shielding at large scales, at the cost of underestimating the self-shielding at the scale of a CNM cloud. The impact of this broadening parameter on the comparison of the HI-to- H_2 transition with the observations will be discussed in §10.

5.5 Grid of Simulations

5.5.1 Steady-State

All simulations start at an homogeneous density and temperature of T = 8000 K. The gas evolves under the joint actions of turbulence, gravity, magnetic field, radiation field, and thermal instability, following the MHD equations given in §5 over several tens of megayears. The outputs are saved periodically to be able to access the physical quantities in every cell of the cube at different times. The steady-state is characterized by a plateau reached by all mean physical quantities. Because of the turbulence and the thermal instability, the steady-state has a statistical nature: while the maps of the column densities can appear very different from one time to another, the PDFs remain constant.

During its evolution, the gas splits up in three different phases at thermal pressure equilibrium, the WNM, the CNM, and the LNM (see Fig. 5.5). The formation of dense environments well shielded from the destructive UV radiation field triggers the formation of H₂ and the medium jointly evolves from a purely atomic state to a partly molecular state (see the cyan line in Fig. 5.6). If the mass of the gas is conserved, as imposed by the periodic boundary conditions, the medium progressively tends towards a steady-state where the mean pressure (see the red curve in Fig. 5.6), $\langle P/k_B \rangle$, the mass fractions of the different phases, $f_{\rm WNM}$ and $f_{\rm CNM}$, their velocity dispersion, $\sigma_{\rm turb}$, and their mean molecular fractions are roughly constant. This steady-state is typically reached after a few turnover timescales, provided that the corresponding time is longer than the damping time, i.e. the typical timescale of decay of the turbulence.

The turbulent forcing and the subsequent turbulent cascade induce pressure variations and shear motions at all scales which trigger mass exchanges between the different phases. Any pressure or density structure is therefore a transient system which is usually described by its contribution to probability distribution functions. This steady yet everchanging environment is the reason for the sustained presence of a substantial amount of gas in the LNM at densities and temperatures out of thermal equilibrium (e.g. Marchal et al. 2019). All the results shown throughout this work are taken at steady-state, i.e. at times ranging from a few tens of Myr up to 100 Myr depending on the strength of the turbulent forcing.

5.5.2 Parametric Study

A parametric study, covering a wide range of physical conditions, is conducted to explore the impact of different physical parameters and unveil the joint effects of turbulence, magnetic field, radiation field, and gravity on the structure and the evolution of the ISM and the degeneracies between all these components, with the help of comparisons with different observables (e.g. HI-to-H₂ transitions, pressure distribution and correlations between different chemical species, see §4).



Figure 5.5: Evolution of the fraction of the different phases, CNM (bleu), LNM (yellow), WNM (red) as a function of time for different turbulent forcing strengths, *F*, and compression parameters, ζ . Top panel, fraction in volume, lower panel fraction in mass, i.e. weighted by the mass of each cell as defined in eq. 5.20. The final fraction of CNM is written in each panel, cf. values in Fig. 5.8



Figure 5.6: Evolution of the mean density of molecular hydroger (cyan) and the mean pressure (red) for the fiducial simulation as a function of time. Note that the early time of the simulations display sharp evolutions. This is because we only sample a reduced number of RAMSES outputs at irregular time intervals, saving more outputs at later time.

The framework, described in §5, lays on several independent parameters, which are all related to key physical ingredients of the ISM. The influence of each ingredient is studied here through several grids of simulations – including a total of 305 runs – covering a broad range of physical conditions and centered around a fiducial setup⁶. The reference value adopted for each parameter, and the range of values explored in this work, are summarized in Table 5.1. Among all parameters, L, $\overline{n_{\rm H}}$ and G_0 are of particular importance.

With our assumptions, *L* simultaneously corresponds to the scale of illumination of the gas by UV photons and twice the integral scale of turbulence, that is twice the scale of injection of mechanical energy L_{drive} . Taking into account the distances between of OB stars associations in the Solar Neighborhood, the height of molecular gas above the Galactic plane, and the typical size of HI superclouds discussed in §3.3, we adopt a fiducial simulation with L = 200 pc and explore values down to a few tens of pc.

The mean density of the gas, $\overline{n_H}$, represents the mass of the diffuse neutral ISM contained in a volume L^3 , and also controls the porosity of the matter to the impinging radiation field. In this work, we adopt a fiducial value $\overline{n_H} = 2 \text{ cm}^{-3}$, a value slightly larger than the standard Galactic midplane density of HI at a galactocentric distance of 8.5 kpc (Kalberla & Kerp 2009). With our fiducial value of *L*, the mass of gas contained in the box accounts for all the mass surface density of HI in the Solar Neighborhood ($\Sigma_{\text{HI}} \sim 10 \text{ M}_{\odot}$ pc⁻², Nakanishi & Sofue 2016; Miville-Deschênes et al. 2017).

 G_0 controls the intensity of the radiation field, hence the thermodynamical state of the gas. Since G_0 is normalized to the Habing field, we choose a fiducial value of $G_0 = 1$. The corresponding UV energy density of the fiducial setup is therefore slightly smaller than that contained in the standard UV radiation fields given by Draine & Salpeter (1978) and Mathis et al. (1983).

The fiducial values of the two parameters *F* and B_x are set to 9×10^{-4} kpc Myr⁻² (see vertical dotted line in Fig. 5.11) and 3.8 μ G, respectively. Those values are chosen so that the velocity dispersion of the gas and the strength of the magnetic field are close to the values observed in the diffuse ISM.

5.5.3 Physical State of the Gas

Mean Pressure As shown in Fig. 5.7, the mean pressure of the gas is primarily regulated by the thermal equilibrium curve (see §2) which depends on the local illumination of the gas by the UV radiation field G_{eff} . Since G_{eff} results from the absorption of the external UV field by the surrounding environments, $\langle P/k_B \rangle$ is not only sensitive to G_0 but also to the total mass of the simulation set by $\overline{n_{\text{H}}}$ and *L*. Larger values of G_0 or smaller values of *L* or $\overline{n_{\text{H}}}$ leads to larger $\langle P/k_B \rangle$.

The turbulent forcing induces pressure fluctuations and shearing motions at all scales.

 $^{^{6}}$ The grids have been run on the computing cluster Totoro funded by the ERC Advanced Grant MIST. The computational time of the fiducial simulation is ~ 6 days using 40 processors.

Parameter	Notation	Ref	Range	Units
box size	L	200	20 - 200	рс
mean density	$\overline{n_H}$	2	0.5 - 4	cm ⁻³
UV radiation field	G_0	1	0.5-4	Habing field
resolution	R	256 ³	$64^3 - 512^3$	_
turbulent forcing	F	9×10^{-4}	$10^{-5} - 10^{-2}$	kpc Myr ⁻²
compressive ratio	ζ	0.1	0.1 - 0.9	-
Doppler broadening	b_D	8	1 - 8	$\rm km~s^{-1}$
initial magnetic field	B_{X}	3.8	0 - 40	μG
self-gravity	_	on	on - off	_
galactic well	_	on	on - off	_

Table 5.1: The range of parameters explored in this work as well as the reference value adopted for each parameter, the ensemble of these reference values will be called fiducial model.



Figure 5.7: Colored tables of the mean pressure expressed in K cm⁻³. The first and second columns display these quantities as functions of $\overline{n_{\rm H}}$ and G_0 , for L = 50 pc and $F = 3.6 \times 10^{-3}$ kpc Myr⁻² (first column) and for L = 200 pc and $F = 9 \times 10^{-4}$ kpc Myr⁻² (second column). The third and fourth columns display these quantities as functions of the acceleration parameter *F* and the compressive ratio ζ , for L = 50 pc (third column) and L = 200 pc (fourth column). All other parameters are set to their fiducial values (see Table 5.1).

As shown by Seifried et al. (2011), this not only frequently perturbs the gas out of the thermal equilibrium states but also strongly reduces the time spent by any fluid element in the WNM, LNM, and CNM. Because of these two aspects, increasing the turbulent forcing reduces the mass of the CNM to the benefit of those of the LNM and WNM (right panels of Fig. 5.8 and Fig. 10 of Seifried et al. 2011). The mean pressure therefore increases, the 1D PDF of the density broadens and its bimodal nature progressively disappears (Piontek & Ostriker 2005; Walch et al. 2011). These effects can be magnified depending on the nature of the turbulent forcing and the power injected in the compressive and solenoidal modes. Because solenoidal motions are more efficient to prevent the gas to condensate back to the CNM phase, a pure solenoidal forcing naturally leads to larger pressure and smaller CNM fractions than those obtained with an equivalent kinetic energy injected in pure compressive modes.


Figure 5.8: Summary of the final fraction of the CNM and WNM for 4 grids of the simulation with different radiative radiation field and mean densities for two *L* on the left, and different forcing strengths and compression on the right.

Phases For the sake of simplicity, we assume that the WNM is composed of all cells with a temperature $T \ge 3000$ K, the CNM of all cells with $T \le 300$ K, and the LNM, has temperature comprised between those of the CNM and the WNM. Therefore, we define the mass-weighted fractions of CNM and WNM as

$$f_{\text{WNM}} = \frac{\sum \rho_{|T| \ge 3000\text{K}}}{\sum \rho}; \quad f_{\text{CNM}} = \frac{\sum \rho_{|T| \le 300\text{K}}}{\sum \rho}.$$
(5.20)

The fractions of mass contained in the WNM and the CNM are controlled by the mean pressure and the total mass of the gas. Larger pressure implies larger densities of the WNM (and the CNM). The fraction of mass of the CNM, therefore, decreases as $\langle P/k_B \rangle$ increases; eventually, if the density of the WNM becomes comparable to the mean density $\overline{n_H}$, the CNM almost entirely disappears (see left panels of Fig. 5.8). At last, and because the WNM occupies most of the volume, f_{CNM} necessarily increases as a function of the total mass of the gas, or equivalently $\overline{n_H}$, even at constant pressure.

Large pure solenoidal forcing efficiently prevents the unstable gas to condensate back to the cold and dense environment, compared to the pure compressive forcing, which can lead to an underestimation of the amount of molecular gas (see Figs. 5.5, 5.9). Without turbulence or with a weak turbulent forcing, CNM clouds are found to evolve toward large-scale entities which evaporate slowly. If the turbulent forcing increases, the CNM becomes progressively organized into a distribution of structures of different sizes down to the numerical resolution.

The initial conditions of the simulations are important and turbulence on its own cannot divide the medium into phases if the initial density is too low. The turbulent forcing will push the low-density gas into the LNM phase, but not the CNM.

Velocity dispersion To guide our choice of parameters like $\Delta \tau$ (the time interval at which the forcing is applied) and *F* (the acceleration parameter), and to evaluate the impact of



Figure 5.9: Total column densities for simulations with different turbulent forcing strengths, *F*, and compression factors, ζ . The statistical properties of the column density are found to depend on the strength, the scale, and the compressibility of the turbulent forcing in simulation of the multi-phase neutral gas.

the turbulent forcing, we introduce the velocity dispersion. Scaling laws link the velocity dispersion to the size of molecular clouds (Larson 1981; Hennebelle & Falgarone 2012).

The velocity dispersion of the WNM is mostly given by the strength of the turbulent forcing and the driving scale ($L_{drive} \sim L/2$), with a slight dependence on $\overline{n_{\rm H}}$ and ζ . As proposed by Saury et al. (2014), a realistic value for the turbulent velocity dispersion of the WNM can be estimated by looking at the HI 21 cm emission spectra with the fewest components observed at high Galactic latitude (Kalberla et al. 2005). Towards these directions, Haud & Kalberla (2007) derive a total velocity dispersion $\sigma_{tot} = (\sigma_{turb}^2 + \sigma_{th}^2)^{1/2} \sim 10 \text{ km s}^{-1}$, where σ_{th} is the 1D thermal velocity dispersion (~ 8.2 km s⁻¹ for the WNM). In the present work, the turbulent forcing applied to the fiducial simulation (see Table. 5.1 and Fig. 5.11) is chosen so that $\sigma_{turb} \sim 4 - 6 \text{ km s}^{-1}$ (see dotted lines in Fig. 5.11), in fair agreement with the observations at high Galactic latitude. While this value is chosen as a reference, the velocity dispersions obtained in all the simulations explored in this work range between 1 and 15 km s⁻¹.

There are many ways to define the turbulent velocity dispersion σ_{turb} . It could be defined as the volume-weighted or mass-weighted velocity dispersion computed over the entire volume (Audit & Hennebelle 2005; Federrath et al. 2010), the average of the mass-weighted velocity dispersion computed along individual lines of sight (Miville-Deschenes & Martin 2007; Saury et al. 2014), or the dispersion of the mass-weighted velocity centroids computed along independent directions (Henshaw et al. 2019). All these definitions give velocity dispersions that differ from one another by factors of a few and are not equally relevant for the comparison with observed quantities. To relate the velocity dispersion to a kinetic energy and provide values that could be compared to the observations of broad HI emission profiles at high Galactic latitude, we choose to compute σ_{turb} in the WNM only as

$$\sigma_{\text{turb}}^2 = \frac{1}{3} \frac{\sum \rho ||\vec{v} - \overline{\vec{v}}||^2}{\sum \rho}$$
(5.21)

with

$$\overline{\vec{v}} = \frac{\sum \rho \, \vec{v}}{\sum \rho}.$$
(5.22)

Correlation Time and Damping Time In Fig. 5.10, we show the impact of different correlation times on the 1D turbulent velocity dispersion for different box sizes for $F = 1.5 \times 10^{-4}$ and 1.5×10^{-3} kpc Myr⁻². The curves are approximately flat as a function of the correlation time, and they depend more strongly on the driving strength and the size of the box.

Theoretical studies and numerical simulations (McKee & Ostriker 1977; Korpi et al. 1999; Mac Low & Klessen 2004) show that the main source of turbulence is the mechanical energy injected by supernovae explosions. To follow this trail, we set the time scale to $\Delta \tau \sim 0.4$ Myr, which is the typical interval separating two SN explosions in a box of (200 pc)³. The characteristic damping time of the turbulence τ is approximately set to the turnover timescale of the diffuse ISM $\tau \sim 33(L/200\text{pc})^{0.6}$ Myr (Larson 1981; Hennebelle & Falgarone 2012).

Driving Strength In Fig. 5.11, we show how different box sizes *L* and different driving strengths, *F*, affect the turbulent velocity dispersion. These 15 simulations are run with-



Figure 5.10: Turbulent velocity dispersion along one axis as a function of the correlation time, for different box physical size and driving strength. The two horizontal lines indicate the range of the observed turbulent velocity dispersion. The driving strength in this graph is in machine units, that can be converted into physical units by the factor 1.5×10^{-7} kpc Myr⁻².

out magnetic field and with a mean density $\overline{n_{\rm H}} = 1 \text{ cm}^{-3}$ and an external radiation field of $G_0 = 1$. The turbulent velocity dispersion is computed following eq. 5.21.

Another parameter left free in this work is ζ , that imposes the fraction in solenoidal and compressible modes. In Fig. 5.12, we show the dependence of the turbulent velocity dispersion in 1D to the driving strength *F* for three different compression parameters, $\zeta =$ 0.1,0.5,0.9. In this case, instead of using the acceleration parameter, *F*, we trace the velocity dispersion as function of the velocity parameter $V = \sqrt{FL_{drive}}$, to show how the velocity dispersion scales roughly linearly with this parameter. These 15 simulations are run on a box of *L* = 200 pc, i.e. the main driving scale is $L_{drive} = 100$ pc, and all the rest of the parameters (apart from *F* and ζ) are set to their reference value in Table 5.1. Compare to the driving strength and the size of the box, the effect of ζ is small on the velocity dispersion.



Figure 5.11: Turbulent velocity dispersion along one axis as function of the turbulent driving strength, *F*, in physical units of an acceleration, kpc Myr⁻², for diffent box physical size, *L*, in which the magnetic field is set to zero to avoid having a privileged direction. The two horizontal lines indicate the range of the observed turbulent velocity dispersion. The external radiation field is set to $G_0 = 1$ and the mean density $\overline{n_{\rm H}} = 1 \text{ cm}^{-3}$. The dotted gray lines indicate $\sigma_{\rm turb} \sim 4 - 6 \text{ km s}^{-1}$ and the vertical line is the acceleration parameter used in the fiducial simulation of a box of L = 200 pc.

Figure 5.12: Turbulent velocity dispersion as defined in 5.21 as a function of the velocity parameter, *V*. The 15 simulations are run with the fiducial parameters (see Table 5.1). The velocity dispersion is computed in one direction, as in equation 5.21, for different driving strengths and compressions. We can see that the velocity dispersion increases with the square root of the driving strength, i.e. linearly with the velocity parameter $V = \sqrt{FL_{\text{drive}}}$.



Chapter 6

CI Excitation

Carbon is an abundant element in the ISM, and one of the most abundant metals¹ with oxygen. In the diffuse neutral ISM, carbon is mainly carried by three different species: C, C^+ , CO. Since the populations of the fine-structure levels of CI are often used as a measure of pressure (see §4), we describe in this section its energy structure and transitions along with the code we built to compute the excitation of neutral carbon in post-processing.

The neutral carbon atom has an electronic ground state configuration of $1s^22s^22p^2$: the two outer electrons are in p orbitals, meaning that they have an orbital momentum l = 1. These two orbital momenta build up to a total orbital momentum, L, that can have values 0, 1, or 2. Both these electrons have a spin of 1/2 which can build up a total spin momentum S of 0 or 1. From Pauli's exclusion principle the total wavefunction needs to be asymmetric with respect to the exchange of electrons, which is equivalent to say that either both S and L are odd or even. Looking at all possible combinations gives the electronic levels ${}^{1}S_{0}$ (S = 0, L = 0), ${}^{1}D_{2}$ (S = 0, L = 2) and ${}^{3}P_{0,1,2}$ (S = 1, L = 1). The superscript indicates the degeneracy of the level due to the total spin momentum (subscript = 2S + 1). Both ${}^{1}S$ and ${}^{1}D$ are singlets states, while ${}^{3}P$ is a triplet that shows a fine-structure splitting (see Fig. 6.1).

6.1 Electronic Ground State: Fine-Structure Levels

A good quantum number to differentiate the fine-structure levels is J = |L+S|, which in this case has 3 possible values: J = 0, 1, 2, as shown in Fig. 6.1. In the following, we call the first excited level CI^{*} (³P₁) and the highest level CI^{**} (³P₃). Because they lie at energies 23 K and 62 K, these levels can be populated collisionally at low temperatures.

In §2, we have already discussed the fine-structure transitions (the one responsible for cooling at $T < 10^4$ K) and pointed out that these transitions are highly forbidden. Selection

¹We remind that in astronomy all elements heavier than helium are called metals

Figure 6.1: CI fine-structure energy levels in ground state. This splitting is due to LS coupling.



rules² in atomic physics dictate that allowed electronic dipole transitions are given by:

$$\Delta S = 0 \tag{6.1}$$

$$\Delta L = \pm 1 \tag{6.2}$$

$$\Delta J = 0, \pm 1 \tag{6.3}$$

All CI fine-structure levels have the same orbital momentum L = 1: the transitions are therefore forbidden. It follows that absorption of radiation and induced emission are probably negligible processes in the excitation and de-excitation of the fine-structure levels.

6.2 CI Excitation

6.2.1 Radiative and Collisional (De-)Excitation

Collisions with other species present in the medium can excite and de-excite the atom and therefore change the population in the fine-structure levels of the ground electronic state of CI. The main collisional partners of CI are hydrogen (whatever its forms: H⁺, H, H₂), helium and electrons. The collision rate for each partner³ can be written as $n_C k_{ij}^C$ (in s⁻¹), where n_C is the density of the collisional partner and k_{ij}^C is the collision rate coefficient (in cm³ s⁻¹) and depends on the cross-section, the relative velocity of the two species and their reduced mass.

Spontaneous emission can also depopulate the higher levels. An atom can emit a photon and radiatively de-excite with a rate which is given by the Einstein coefficient of spontaneous emission $A_{u\ell}$ (in s⁻¹).

²Selection rules, or transition rules constrain the possible transitions from one quantum state to another of a system

³ indicated with the letter *C* for collisional partner

At steady-state $\left(\frac{dn_i}{dt} = 0\right)$, we can write

$$n_i \left[\sum_{j \neq i} n_C k_{ij}^C \right] = \sum_{j \neq i} n_j n_C k_{ji}^C + \sum_{j > i} n_j A_{ji}.$$

$$(6.4)$$

Dividing by the abundance of the element and putting all terms depending on the abundances on one side we obtain

$$AX = B$$

with:

$$A = M_{\text{rad}} + M_{\text{coll}}; \quad X = \begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix}; \quad B = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix};$$
$$M_{\text{coll}} = \begin{pmatrix} -C_{01} - C_{02} & C_{10} & C_{20} \\ C_{01} & -C_{10} - C_{12} & C_{21} \\ C_{02} & C_{12} & -C_{20} - C_{21} \end{pmatrix}; \quad M_{\text{rad}} = \begin{pmatrix} 0 & A_{10} & A_{20} \\ 0 & -A_{10} & A_{21} \\ 0 & 0 & -A_{20} - A_{21} \end{pmatrix}.$$

The collisional matrix is composed of the sum of the collisions for each collision partner

$$C_{ik} = \sum_{c} n_c k_{ij}^c$$

where the colliders are H, H^+ , H_2 , He, and e^- . The collision excitation rates are linked to the collision de-excitation rates by the detailed balance principle:

$$C_{\ell u} = C_{u\ell} \frac{g_u}{g_\ell} e^{\frac{-\Delta E_{u\ell}}{k_B T}}$$
(6.5)

where $\Delta E_{u\ell} = E_u - E_\ell$.

This system of equations is not closed, i.e. the lines forming the A matrix are not linearly independent. Therefore the system needs to be closed by a conservation relation which is

$$\sum_{i} x_i = 1$$

or equivalently $n(CI) = \sum_i n_i$. This equation can replace any equation of the system. For numerical reasons, we choose to introduce this closure equation instead of the steady-state equation of the level *i* with the largest abundance x_i .

We built the excitation matrix taking into account radiative decay (Froese Fischer & Saha 1985) and collisions with atomic hydrogen (Launay & Roueff 1977), molecular hydrogen (Schroder et al. 1991), helium (Staemmler & Flower 1991), ionized hydrogen (Roueff & Le Bourlot 1990) and electrons (Johnson et al. 1987).



Figure 6.2: *Left:* Population of the first excited fine-structure level ${}^{3}P_{1}$ of CI relative to the ground state ${}^{3}P_{0}$ in logarithmic scale as a function of temperature and density. *Right:* Population of the two first excited fine-structure level CI^{*} (filled points) and CI^{**} (crosses) of CI relative to the total density of CI as a function of the temperature and the density of the medium.

Dendence of the Level Population on the Physical Conditions of the Gas The population of the levels of neutral carbon depends on the collision partners, and therefore on the physical conditions of the medium. For this reason, the population of the levels in the fine structure of neutral carbon is a tracer of the pressure and density of the neutral ISM.

To validate our code, we compute the populations of the fine-structure levels of the ground state considering only collisions with hydrogen and electrons with an abundance $n_e/n_{\rm H} = 10^{-3}$ – which are the prime collisional partners – and compare the results to Fig 17.6 in Draine (2011).

As shown in Fig. 6.2, at low density, the populations of the first and second excited levels naturally increase with the density and the temperature. Because the critical densities of theses two excited states are of the order of 10^3 cm^{-3} , the system naturally tends toward a Boltzmann distribution at $n_{\rm H} \gtrsim 10^3 \text{ cm}^{-3}$. In this regime, the LTE applies and the ratios between the populations of the excited levels of CI only depend on the degeneracies of the levels and the temperature (if $T \leq 200 \text{ K}$).

Following Jenkins & Tripp (2011) we compute the theoretical prediction of f2 as a function of f1 (see eqs. 4.2) at contant temperatures for a large range of densities. As done by Jenkins & Tripp (2011) we assume a molecular fraction of 0.6 ($n(H_2) = 0.3n_H$), an ionization fraction of $2 \times 10^{-4}n_H$ and 10% of helium. The results are shown in Fig. 6.3. With a fixed molecular fraction, the pressure is simply proportional to n_HT . Fig. 6.3 shows how both the temperature and the density have a great impact on f1 and f2. Following an isothermal curve at high temperature, we note that f1 increases with density at a larger rate than the increase of f2, until a certain pressure. When this pressure is reached, these behaviors are inverted and f2 varies more strongly than f1. As seen in Fig. 6.2, at high temperature, as the density increases f1 increases and then decreases, while f2 only in-

Figure 6.3: Theoretical curves of f2 vs f1. The fraction of the population in the second excited fine-structure level as a function of the fraction of population in the first excited fine-structure level plotted for T = 15, 30, 80, 300, 1000 K and proton densities from 10 to 10^6 cm⁻³. The abundance of molecular hydrogen, helium and electrons are fixed to $n(H_2) =$ $0.3n_{\rm H}$, $n({\rm He}) = 0.1n_{\rm H}$, and $n_e = 2 \times 10^{-4}$ to follow Jenkins & Tripp (2011) prescription. The two white and two black crosses indicate pairs of points with the same pressure. Two regimes can be identified: at low pressures, the value of f^2 does not vary greatly while f1 does; at high pressure these behaviors are inverted.



creases until it becomes larger than f_1 : this behavior makes the curve look like an upward and backward bent arch. Interestingly, in the low-pressure regime, one given pressure can result in markedly different values of f_1 and very close values of f_2 . This behavior is inversed at high pressure. All these results are in perfect agreement with theoretical prediction of Draine (2011) and Jenkins & Tripp (2011). This validates the post-treatment excitation code of neutral carbon. In the following, we estimate the impact of the chemical pumping and optical pumping on the populations of the fine-structure levels of CI.

6.2.2 Chemical Pumping

Until now we only considered excitation from collisions and de-excitation due to collisions and spontaneous emission. However, excitation and de-excitations can also occur during chemical reactions, i.e. at the formation and destruction of neutral carbon. In the diffuse ISM, neutral carbon is mainly formed through the recombination of C⁺ with electrons and PAHs, and mainly destroyed by photoionization.

These processes add a formation and destruction term for every level. At steady-state, the equation system is modified to

$$n_{i}\left[\sum_{C}\sum_{j\neq i}n_{C}k_{ij}^{C}+k_{D}\right]=\sum_{C}\sum_{j\neq i}n_{j}n_{C}k_{ji}^{C}+\sum_{j>i}n_{j}A_{ji}+K_{F}^{i}.$$
(6.6)

Given the large exothermicity of recombination reactions, we assume that the formation rate of a level, K_F^i , is proportional to the degeneracy of the level *i*. Assuming that all levels are destroyed at a single rate k_D , the system of equations to solve reads AX = B, where $A = M_{\text{rad}} + M_{\text{coll}} + M_D$,

$$M_D = \begin{pmatrix} -k_D & 0 & 0 \\ 0 & -k_D & 0 \\ 0 & 0 & -k_D \end{pmatrix} \text{ and } B = \begin{pmatrix} -p_0 k_D \\ -p_1 k_D \\ -p_2 k_D \end{pmatrix}.$$

6.2.3 Optical Pumping

Through the absorption of optical photons neutral carbon, can be excited to higher electronic states and then radiate to end up in one of the three fine-structure levels of the ground state. This process, referred to here as optical pumping, can be easily modeled by adding an optical pumping matrix to *A* which reads (Jenkins & Tripp 2011).

$$M_{\text{OptPump}} = \begin{pmatrix} -P_{01} - P_{02} & P_{10} & P_{20} \\ P_{01} & -P_{10} - P_{12} & P_{21} \\ P_{02} & P_{12} & -P_{20} - P_{21} \end{pmatrix} \times \frac{G_{\text{eff}}}{1.27}.$$

Since we express the radiation field in Habing units, and the tabulated optical pumping values are given for a Mathis et al. (1983) radiation field, we use the conversion factor (1.27) calculated in Table 3.1 between the energy densities of these two radiation fields. Interestingly, the inclusion of the chemical and optical pumping has almost no impact on the excitation and de-excitation of the fine-structure levels of CI. This result is in line with the conclusion of Jenkins & Tripp (2011). Because of the relatively large density of the CNM

and the weak energy density of optical photons in the ISRF, both processes are negligible compared to collisional excitations.

6.3 CI radiative transfer

Computing the excitation of CI is not enough to allow a comparison with the observations described in §4.2. Indeed, comparing simulated data with observations requires to access the relative population of CI in velocity channels. To do so, it is necessary to implement the radiative transfer and create line profiles of the excited states of CI.

The evolution of the specific intensity I_v (in erg cm⁻² s⁻¹ sr⁻¹ Hz⁻¹) can be written as

$$\frac{dI_{\nu}}{dz} = -k_{\nu}I_{\nu} + \eta_{\nu} \tag{6.7}$$

where k_v is the atomic absorption coefficient, (in cm⁻¹), and η_v is the emissivity (in erg cm⁻³ s⁻¹ sr⁻¹ Hz⁻¹). Assuming that there is no scattering and that the optical emission of CI is neglible, the transfer equation simplifies into

$$\frac{dI_{\nu}}{dz} = -k_{\nu}I_{\nu} \tag{6.8}$$

where the absorption coefficient reads

$$k_{\nu} = \frac{\lambda_{ij}^2}{\sqrt{8\pi}} g_i A_{ij} \frac{n_j}{g_j} \phi_{ij}^{\nu}.$$
(6.9)

with a Gaussian line profile,

$$\phi_{ij}^{\nu} = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(\nu - \nu_z)^2}{2\sigma^2}\right)$$
(6.10)

where σ is the total velocity dispersion, which is a quadratic sum of the 1D thermal velocity dispersion and the micro-turbulent velocity dispersion, *b*:

$$\sigma^2 = \frac{k_B T}{m} + b^2. \tag{6.11}$$

6.4 Resulting Line Profiles

Our procedure to compute the line profiles of CI across simulations of the diffuse multiphase ISM is based on the chemical post-process code developed by Valdivia et al. (2017) (see §13). This code provides the atomic and molecular abundances of 151 chemical species including neutral carbon and all its collisional partners. Implementing the excitation of CI in this code along with the simple radiative transfer described above, not only gives access to the densities n_i of the three fine-structure levels of CI but also their column densities in any velocity channel. In practice, the dynamical information from RAMSES output, e.g. the velocity in each direction and the temperature, are used to determine the line profile per velocity, ϕ_{ii}^v , defined above. The quantity $n_i \phi_{ii}^v$ is then integrated along the line of sight in velocity intervals of 0.5 km s⁻¹ (see Jenkins & Tripp 2011), to find the column density of the level *i* per velocity channel

$$N_i^{\nu} = \int_{\nu-0.25}^{\nu+0.25} \int_0^{l_{\rm los}} n_i \phi_{ij}^{\nu} d\ell d\nu.$$
(6.12)

Each cell of the simulation, therefore, contributes as a gaussian to N_i^v with a width that depends on the temperature *T* of the cell, a height proportional to the density n_i , and a position that depends on the velocity along the line of sight v_z of the cell. In Fig. 6.4, we show an example of those variables along one line of sight of our fiducial simulation. We note that, on that particular line of sight, both CI* and CI** exhibit two density peaks, which correspond to two different velocities at around 2 km s⁻¹ and -13 km s⁻¹, and to two different phases, the CNM at low temperature and WNM with $T \sim 8000$ K.

The corresponding column density per velocity channel $N_{CI^*}^{\nu}$ is shown in the left panel of Fig. 6.5 assuming a microturbulent velocity dispersion of 1 km s⁻¹. The range of velocity, from -30 to 30 km s⁻¹, is determined by the minimal and maximal velocity in the simulation and the maximal velocity dispersion, $\min(v_z) - \max(\sigma) < v < \max(v_z) + \max(\sigma)$. $N_{CI^*}^{\nu}$ exhibits two peaks centered at the velocities given above: the peak at high temperature appears broader, because of the thermal broadening. The thinner colored lines are the contributions of each cell integrated in velocity bins of 0.5 km s⁻¹, $\int n_i \phi_{ij}^{\nu} d\nu$, and the thicker line is the integration over the entire line of sight. In the right panels of Fig. 6.5, we show other examples of profiles integrated over other lines of sight. The top panel displays a line profile with two separated peaks in velocity because the corresponding peaks of $n(CI^*)$ along the line of sight are at velocities that differ by 10 km s⁻¹ and belong to the CNM. In contrast, the bottom panel shows a line profile that appears as a mixture of a large component at high temperature and a narrow component at a lower temperature, with a separation in velocity similar to the previous case. All these results show that the combination of the chemical solver, the CI excitation code, and the radiative transfer code provide a powerful tool to study the column densities of neutral carbon integrated in separate velocity channels. Such a tool allow a direct comparison of the simulations with observations of neutral carbon across the diffuse neutral ISM collected by Jenkins & Tripp (2011) (see §4), but also allows to study the kinematic signature of CI. The comparison of the simulated data and the observations will be shown in §12.



Figure 6.4: One line of sight of the fiducial simulation showing the velocity along the *z*-axis, temperature, density of CI^* , CI^* and CI_{tot} .



Figure 6.5: Examples of $N_{\text{CI}^*}^{\nu}$ for three different lines of sight of the fiducial simulation. The panel on the left shows $N_{\text{CI}^*}^{\nu}$ for the line of sight showed in 6.4.

Chapter 7

H₂ Excitation

Molecular hydrogen, H_2 , is a homonuclear molecule formed by two atoms of hydrogen linked by a covalent bond, with a binding energy of 4.476 eV which corresponds to a temperature of 51942 K. Since this molecule is symmetric, it lacks of a permanent dipole moment. The excitation of H_2 can be an important diagnostic of the turbulent multiphase diffuse ISM. Although the goal of computing absorption and emission profiles of H_2 was not completely achieved in this thesis, we present in the next sections the levels and transitions of H_2 , and the code we built to compute H_2 excitation in post-processing.

7.1 H_2 Levels

Rotational Levels In first approximation, H_2 can be treated as a linear rigid rotor with a moment of inertia I_m that takes into account the reduced mass μ of the molecule and the nuclear distance through $I_m = \mu r_0^2$, where r_0 is the inter-nuclei distance at rest. The energy levels are quantified and the rotational energy of a level *i* can be written as function of the moment of inertia of the molecule

$$E(J) = \frac{J(J+1)\hbar^2}{2I_m} = BJ(J+1)$$
(7.1)

where *J* is the rotational quantum number, \hbar is Planck constant devided by 2π , and $B = \frac{\hbar^2}{2I_m}$ the rotational constant, which has a value of 85.25 K for H₂. The distance between two neighboring rotational levels can be written as

$$\Delta E = 2BJ \tag{7.2}$$

with J the rotational quantum number of the upper level.

Ortho and Para Molecular hydrogen has two isomeric forms. Each hydrogen forming the molecule has a nuclear spin of 1/2. The two proton nuclear spins can be parallel forming ortho-hydrogen with total nuclear spin I = 1, or the two proton spins can be antiparallel forming para-hydrogen. Para-H₂ has an anti-symmetric nuclear wavefunction and since the total wavefunction must be anti-symmetric with respect to exchange of the two protons (from Pauli's principle), the rotational wavefuction needs to be symmetric with respect to permutation of the two protons, hence *J* must be even. Ortho-H₂, on the other

hand, has a symmetric nuclear wavefunction therefore J needs to be odd. The transitions ortho-H₂ \leftrightarrow para-H₂ are highly forbidden, because they need a change in the nuclear spins of the molecule. For these reasons, ortho-H₂ and para-H₂ are often considered as two independent species.

Consequently, only rotational transitions with even ΔJ are possible and therefore the first rotational accessible excited level from the ground state is J = 2 at 510 K. The transitions between ortho-H₂ and para-H₂ can only occur through reactive collisions with H, H⁺, H₂ and H₃⁺ that are able to change the orientation of nuclear spins.

Vibrational Levels Vibrational levels are the result of the quantification of the different vibrational modes and their quantum number is v which corresponds to the number of nodes in the vibrational wavefunction. Diatomic and linear molecules can vibrate along their internuclear axis and the quantified energy of the vibrational levels can be written as

$$E(v) = hv_0(v + \frac{1}{2})$$
(7.3)

where v_0 is the fundamental vibrational frequency.

Rovibrational Transition One ro-vibrational level (v, J) can be written as the sum of its rotational and vibrational levels:

$$E(v,J) = hv(v + \frac{1}{2}) + BJ(J+1)$$
(7.4)

The transition between different levels are usually indicated with letters to indicate the jump in rotational levels as indicated in the table 7.1.



Figure 7.1: Vibration-rotation energy levels of the ground electronic state of molecular hydrogen with $J \le 29$ taken from Draine (2011) with indication of the 1 - 0S(1) transition, taken as an example. 1 - 0S(1)transition refers to $v = 1 \rightarrow 0$ and $\Delta J = 2$ ending in J = 1so the rotational transition J = $3 \rightarrow 1$. This transition occurs at 2.1218 μ m

D	$J_u - J_\ell = -\Delta J$	Type of transition
$\mathrm{O}(J_\ell)$	-2	Electric quadripole transition
$\mathrm{P}(J_\ell)$	-1	Electric dipole transition
$Q(J_\ell)$	0	Electric dipole or quadripole transition
$\mathrm{R}(J_\ell)$	+1	Electric dipole transition
$\mathrm{S}(J_\ell)$	+2	Electric quadripole transition

Table 7.1: Notation for the rotational transitions



Figure 7.2: Electronic potential well as function of the internuclei distance (in Å) with energies written in eV on the extreme right, in cm⁻¹ on the left, where also the transition wavelength in Å is indicated. Beside every well there is the term symbol indicating the electronic quantum state. Ground state is indicated by the notation $X^1\Sigma_g^+$, the first electronic excited state, i.e. the Lyman band by $B^1\Sigma^+_{\mu}$, and the second excited electronic state, i.e. Werner band by $C^1\Pi_u$. Between the ground state and the first electronic excited state there is an unbound state $b^3 \Sigma_{\mu}^+$.

Electronic Levels On the top of ro-vibrational states, there are the electronic states, that are more separated in energy.

Each electronic state of a molecule, is actually a superposition of a multitude of rovibrational levels. The different electronic transitions of H_2 are observed as broad bands which are the results of the superposition of the ro-vibrational lines.

The selection rules for molecular electronic transitions are $\Delta S = 0$, $\Delta \Lambda = \pm 1$, $\Sigma^+ \leftrightarrow \Sigma^+$, $\Sigma^- \leftrightarrow \Sigma^-$ and $g \rightleftharpoons u$ for homonuclear molecules. Therefore, the transition from the bound electronic state *X* to the unbound electronic state *b* is highly forbidden, while electronic transitions to Lyman and Werner bands are allowed. Photons with energy hv > 11.2eV, or equivalently wavelength of $\lambda < 1108$ Å, can excite molecular hydrogen from the ground state to the Lyman Band, while photons with $\Delta E > 12.3$ eV, or equivantely $\lambda < 1008$ Å can excite the Werner band.

7.2 Rovibrational Excitation of H₂ in its Electronic Ground State

Where is the molecular hydrogen? Looking at simulations, with the help of the 2D probability histograms in the space of P/k_B vs n_H weighted by the mass of H₂, we can characterize the physical conditions in which we find molecular hydrogen (see right panl Fig.



Figure 7.3: 2D probability histogram of the proton mass (on the left) and the mass of H_2 (on the right) in a pressure vs density diagram. Results are taken from the fiducial simulation (see table 5.1).



Figure 7.4: Einstein coefficients as function of J for the pure rotational radiative de-excitation transitions (i.e. $\Delta v = 0$, $\Delta J =$ 2 pure rotational S transitions). A_{ij} values are small, because the only allowed transitions are quadripolar. As a consequence, the critical densities is expected to be relatively low and increase rapidely with *J*, cf fig. 7.8.

7.3).

While in the local diffuse ISM most of H_2 is formed in a low-temperature and highdensity gas, large amount of warm molecular hydrogen are observed. We can notice that H_2 is not only present in the CNM, but that a non-negligible fraction is also in the LNM. This warm H_2 , and its excited levels, can have important consequences for the chemistry of the medium. Some of the reactions in which H_2 is actor, are endothermic, and if the molecular hydrogen is not in its ground state, the energy barrier to overcome is lower, i.e. the excitation of H_2 reduces the energy gap.

For the above motivations, we coded the excitation of H_2 which can be applied as a post-processing in each cell of our simulation. In the following subsection, we will introduce the main excitation mechanisms, the collisions and the rates with the main collisional partners and some tests to validate the computation of the excitation of molecular hydrogen.

7.2.1 Radiative Transitions

The transition probability from one level *i* to another at lower energy *j*, or equivalently the strength of the radiative transition, is given by the Einstein spontaneous emission coeffi-

cient A_{ij} in s⁻¹, which is proportional to the absorption oscillator strength f_{ji}

$$A_{ij} = \frac{8\pi^2 v_{ij}^2 e^2}{m_e c^3} \frac{g_j}{g_i} f_{ji}.$$
(7.5)

As shown in 7.4, the Einstein coefficient of spontaneous emission for a pure rotational de-excitation with $\Delta v = 0$ and $\Delta J = -2$ increases with J_i (e.g. for v = 0, $A_{ij} \in [2.9 \times 10^{-11}, 5.5 \times 10^{-6}] \text{ s}^{-1}$). The transition can also link rotational states with different vibrational values. The rovibrational transitions ($\Delta v \neq 0$) considered are quadripole transitions (O, Q, S). We do not consider induced emission and absorption (B_{ij} and B_{ji}), because the occupation number of the photons in the wavelength of interest is low. As seen in the previous section, the rovibrational transitions are in the infrared waveband, and the ISM medium is not rich enough in such photons.

7.2.2 Collisional Transition

Even in the diffuse ISM, collisions can be an important source of excitation and de-excitation, especially in warm gas.

Since the collisional rates are proportional to the density of the partner and the crosssection is inversely proportional to the square root of the reduced mass, the most important collisional partners are the ones that are both abundant and light. We will, thus, consider collisions between H₂ and atomic hydrogen (Wrathmall & Flower 2007), helium(Flower et al. 1998), and molecular hydrogen (Flower 1998).

The rate of de-excitation per unit volume of a level due to a collision with the collisional parter *C* is written as

$$n_i n_C k_{ij}^C \,[\mathrm{cm}^{-3} \mathrm{s}^{-1}]$$
 (7.6)

where n_i is the density of the level *i* considered, n_c is the density of the collisional partner, k_{ij}^C is the rate coefficient of the collision that depends on the collisional partner and the starting and ending level of H₂. $k_{ij}^C = \langle \sigma v \rangle_{i \to j}$ is the integrated cross-section, σ the cross-section, and *v* the relative velocity distribution.

The rate coefficients used in this work for collisional de-excitation result from a fit using three parameters *a*, *b*, *c*:

$$q(T) = a + \frac{b}{t} + \frac{c}{t^2}$$
(7.7)

$$k_{i\,i}^{\rm H}(T) = 10^{q(T)} \tag{7.8}$$

where $t = T(K)/10^3 + \delta t$, and $\delta t = 1.0$ to prevent the divergence of the rate coefficient when t goes to zero. ij indicate the transition form the state i = (v, J) to a state at lower energy j = (v', J'). The excitation rate coefficients k_{ji} is given by the detailed balance principle:

$$k_{ji} = k_{ij} \frac{g_i}{g_j} e^{(-\Delta E_{ij}/k_B T)} \,\mathrm{cm}^3 \mathrm{s}^{-1}$$
(7.9)



Figure 7.5: Comparison of rate coefficients k_{ij}^C for two rotovibrational transitions for non-reactive collisions with H₂ + H₂, H₂ + He and H₂ + H. For this last collision we compare two different fits for the rates, Martin & Mandy (1995) in orange and Wrathmall & Flower (2007) in blue.

7.2.3 Reactive Collisions

Reactive collisions are chemical reactions during which the molecular hydrogen exchanges one of its hydrogen nuclei with the collisional partner (e.g. H, H⁺, H₂, H₃⁺). The energy barrier for this process is important, because the covalent bond between the two hydrogen atoms needs to be broken, nevertheless it can be overcome at temperature over 1000 K which are relevant in our study: H₂ + H has a barrier of 5000 K (Siegbahn & Liu 1978). When the collision takes place and the covalent bond is broken and recreated with a hydrogen nucleus of the collisional partner, the final molecular hydrogen can be formed with a relative orientation which is different from the original one. The rate of transformation ortho-H₂ \leftrightarrow para-H₂ is known to be very low, mainly for two reasons: (1) reactive collisions have a high energy barrier, that need to be overcome to break the chemical bond, and (2) proton exchange does not necessarily result in an ortho \leftrightarrow para conversion, i.e. para \leftrightarrow para and ortho \leftrightarrow ortho transitions are also possible.

 $H_2 + H^+$ proton exchange plays a key role in the ortho-para conversion. Gerlich (1990) gives the rate coefficients for the ortho- $H_2 \leftrightarrow$ para- H_2 conversion due to a proton exchange with H^+ for the first nine rotational levels:

$$k_{ij}^{\rm H^+} = k_{0,ij}^{\rm H^+} \exp\left(-\frac{E_{A,ij}}{k_B T}\right) \quad \rm{cm}^3 \rm{s}^{-1} \tag{7.10}$$

where each transition has its rate $k_{0,i}$ in cm³ s⁻¹ and its energy activation barrier $E_{A,ij}$. In Fig. 7.6, we plot the rate coefficients for H₂-H⁺ in units of 10^{-10} cm³s⁻¹ as given by Gerlich (1990).



Figure 7.6: Rate coefficients for H_2 -H⁺ reactive collision, in units of 10^{-10} cm³s⁻¹ from Gerlich (1990) tables. In the top left panel, we show the rate coefficient for the excitation from the para level at J = 0 and on the right from the first ortho level, J = 1. The crosses indicates and o-o or p-p transition, while the points indicate an o-p transition. The bottom panels display the de-excitation rates from a para level (left) and an ortho level (right) that are basically constant.

 $H_2 + H$ Rate coefficients for collisional de-excitations in reactive scattering due to collisions with H are taken from Le Bourlot et al. (1999):

$$k_{\rm r}^{\rm H}(\nu J \to \nu' J') = k_{\rm nr}^{\rm H}(\nu J \to \nu' J') \times \mathscr{E} \qquad (|J - J'| \text{ even })$$

$$k_{\rm r}^{\rm H}(\nu J \to \nu' J') = \frac{1}{2} [k_{\rm nr}^{\rm H}(\nu J \to \nu' J' - 1) + k_{\rm nr}^{\rm H}(\nu J \to \nu' J' + 1)] \times \mathscr{E} \qquad (|J - J'| \text{ odd and } J \text{ even })$$

$$(7.12)$$

$$k_{\rm r}^{\rm H}(vJ \to v'J') = \frac{1}{6} \left[k_{\rm nr}^{\rm H}(vJ \to v'J' - 1) + k_{\rm nr}^{\rm H}(vJ \to v'J' + 1) \right] \times \mathscr{E} \quad \left(\left| J - J' \right| \text{ odd and } J \text{ odd} \right)$$

$$(7.13)$$

where

$$\mathscr{E} = \exp\left[-\max\left\{0, \left[\frac{3900 - \left(E_{\nu J} - E_{\nu' J'}\right)}{T}\right]\right\}\right]$$
(7.14)

The factor of 3900 K accounts for the activation energy barrier, k_r^{H} indicates the reactive collision rate with H, and k_{nr}^{H} non-reactive collision rate with H (see previous section).

7.2.4 Chemical Pumping

Until now we have considered excitation from collisions, both reactive or not, and deexcitation due to collisions and spontaneous emission. However, de-excitation and excitation can also occur during chemical reactions, i.e. at chemical formation and destruction of molecular hydrogen.

Assumptions The formation of molecular hydrogen has an exothermicity of 4.4781 eV, i.e. the binding energy of H_2 . A common assumption is that one third of this energy is tranformed into internal energy of H_2 , one third into the grain on which H_2 forms and one third heats the medium (see §2). Under these assumption, the energy that internally excites H_2 is usually assumed to be spread among the levels following a Boltzmann distribution with a temperature T_F . The probability of forming H_2 in a certain level *i* therefore reads:

$$p_i = \mathcal{N} g_i \exp(-E_i / k_B T_F), \qquad (7.15)$$

where \mathcal{N} , is the normalization factor

$$\mathcal{N} = 1 / \sum_{i} g_{i} e^{(-E_{i}/k_{B}T_{F})}$$
(7.16)

which basically corresponds to the inverse of the partition function.

If the total energy distributed in the internal rovibrational levels of H_2 is equal to 1/3 of the formation energy,

$$\frac{\sum_{i=1}^{N_{\text{lvl}}} g_i E_i \exp(-E_i/k_B T_F)}{\sum_{i=1}^{N_{\text{lvl}}} g_i e^{(-E_i/k_B T_F)}} = 4.4781/3$$
(7.17)

where E_i is in eV and the Boltzmann constant $k_B = 8.617 \times 10^{-5}$ eV K⁻¹. We can easily see that T_F depends on the number of level included N_{lvl} . To avoid this problem, we choose a constant $T_F = 17323$ K which correspond to one third of the formation energy in kelvins.

The downside of this approach is that including only a small number of levels in the computation, implies that the energy recovered is far smaller than 4.4781 eV/3 (e.g. if $N_{lvl} = 10$, the recovered energy is 0.75 eV). Respectively, if the number of levels included is large, the energy distributed in the internal levels of H₂ is larger than 1/3 of 4.4781 eV.

7.2.5 Population as function of physical conditions

Determining the population of the different levels, under the steady-state assumption, means solving the equation:

$$n_{i}\left[\sum_{C}\sum_{i\neq j}n_{C}k_{ij}^{C}(T) + \sum_{i< j}A_{ij} + k_{D}\right] = \sum_{C}\sum_{i\neq j}n_{C}n_{j}k_{ij}^{C}(T) + \sum_{i>j}A_{ji}n_{j} + K_{F}^{i}$$
(7.18)

for each level *i* of H₂. n_i and n_j are the densities of the levels *i* and *j*, n_c the density of a collisional parner and k_{ij}^c the rate coefficient to go from the level *i* to level *j* due to a collision with a partner *C*, A_{ij} the spontaneous Einstein coefficient, k_D the destruction rate of the level *i* (in s⁻¹), K_F^i is the formation rate of the level *i* (in cm⁻³s⁻¹). On the left side of the equation we have all the processes de-populating the level *i* and on the right side the processes populating *i*. To solve this system of equations, we can write it in matrix form

$$AX = B$$

and look for the solution $X = A^{-1}B$. A is the matrix including all excitation and deexcitation rates

$$A = M_{\rm rad} + M_{\rm col} + M_{\rm D}$$

where M_D is the destruction matrix, which contains the k_D on the diagonal and zeros elsewhere. M_{rad} is the radiation matrix containing the Einstein coefficients, and M_{col} the collisional matrix containing $\sum_C n_C k_{ij}^C$ at the position ij. X is the population vector composed of the abundances of the levels x_i , or the density vector composed of n_i , and B is a vector composed of $-p_i k_D$. For the matrix to be inversible and the solution unique, this system needs to be closed by the conservation equation $\sum x_i = 1$, or $\sum n_i = n(H_2)$. The solution X is then found with a Newton-Raphson algorithm.

7.3 Validation of the H₂ Excitation Code

In this section, we perform standard tests on the population of H₂ computed with the code.

1. At Local Thermal Equilibrium (LTE), the population of the levels should follow a Boltzmann distribution

$$\frac{n_i}{n(H_2)} = \frac{g_i e^{-E_i/k_B T}}{\sum_i g_i e^{-E_i/k_B T}}$$
(7.19)

where g_i and E_i are the degeneracy and the energy of the level *i* and $\sum_i n_i = n(H_2)$.

2. In the ISM, the densities are generally lower than the critical densities, and the levels can be far from LTE, depending on the species considered and the physical conditions. We expect the levels with low critical densities, $n_{i,crit}$ to be thermalized.



Figure 7.7: Effect of the chemical pumping: H_2 abundances in the two first vibrational states divided by the degeneracy of the level as function of the energy of the level in kelvin. The assumed fraction of molecular hydrogen is $f(H_2) = 0.6$, $x_e = 2 \times 10^{-4}$, x_{He} is 10%, and the abundance of $H^+ = 10^{-4}$. We show here the effect of chemical pumping on the population of H_2 . Equilibrium between the formation and the photodissociation is assumed, therefore the external radiation field acts in the chemical pumping. Here we show the population for $n_{\rm H} = 200 \text{ cm}^{-3}$ and T = 1000 K on the left, and $n_{\rm H} = 2000 \text{ cm}^{-3}$ and T = 3000 K on the right, for a radiation field of 1 and 10 in Habing unit. The crosses indicates the levels that have a critical density lower than $n_{\rm H}$ and so that are supposed to be thermalized. The dotted line indicates the Boltzman population at LTE condition. Including chemical pumping at low temperature change drastically the population of the levels of H₂.

3. Reactive scattering with H and H⁺ can change the ortho-to-para ratio.

7.3.1 Critical Density

Each level has a critical density which depends on the ratio of the radiative de-excitation to the total collisional de-excitation:

$$n_{\text{crit,i}} = \frac{\sum_{j < i} A_{ij}}{\sum_{j < i} \sum_{C} k_{ij}^{C}(T) x_{C}}$$
(7.20)

where $k_{ij}^C(T)$ is the rate coefficient of the transition *i* to *j* due to the collisional partner *C*, $k_{ij}^C(T)$ is a function of temperature of the medium, and increases with the temperature, leading to lower critical densities. When $n_{\rm H} \gg n_{\rm crit,i}$ the level *i* is thermalized, thus, the critical densities are a measure of how easily a level can reach LTE. We can also define a critical density for a transition, which is basically eq. 7.20 without the sum of the levels *j* lower than *i*:

$$n_{\rm crit,ij} = \frac{A_{ij}}{\sum_C k_{ij}^C(T) x_C}.$$
(7.21)

In Fig. 7.7, we compare the computed level populations of H_2 (dots) with the Boltzmann distribution (red dotted line). When no pumping is included (purple points), the levels with critical densities lower than the density of the medium appear to be thermalized (red crosses), while the levels with $n < n_{crit}$ are subthermally distributed. If the density increases, the number of thermalized levels naturally increases (right panel). When the



Figure 7.8: Critical densities computed with eq. 7.20, for different robivrational levels of H_2 (as function of the energy of their level E_i/k_B in K), the different colors indicate different vibrational levels and the different points in the same color indicate the critical densities of the different rotational levels in the same vibrational state. Low *J* rotational levels have lower critical densities and thus thermalized first. Since collisional rate coefficients are functions of the temperature, these critical densities depend on the temperature of the medium, the graph on the right shows the critical densities for a temperature 2 order of magnitude bigger than that on the left. Notice that the scale on the *y*-axis is not the same, and that the critical densities decrease when the temperature increases.

chemical pumping is included, higher levels of H_2 become super-thermally populated. These results contribute to the validation of the code.

7.3.2 Ortho-Para Ratio

We define the $o-H_2/p-H_2$ ratio as ratio of all populations in ortho levels over all populations in para:

$$o/p = \frac{\sum_{J \text{odd}} x_i}{\sum_{J \text{even}} x_i}.$$
(7.22)

For high densities we can assume LTE and we can compute the ratio analytically. The population of each level is given by:

$$x(v, J) = g_I g_J \exp(-E(v, J)/k_B T)$$
(7.23)

where g_Ig_J is the total degeneracy of the level (v, J) that depends on the nuclear and rotational quantum numbers, $g_Ig_J = 3(2J + 1)$ for ortho levels that have parallel nuclear spin and $g_Ig_J = (2J_i + 1)$ for para levels. We can then write the o/p ratio at LTE, i.e. assuming a Boltzmann population for the levels as:

$$o/p = \frac{\sum_{J \text{odd}} 3g_J \exp(-E(v, J)/k_B T)}{\sum_{J \text{even}} g_J \exp(-E(v, J)/k_B T)} = \frac{Z_{\text{ortho}}}{Z_{\text{para}}}.$$
(7.24)

At LTE the o/p ratio is the ratio between the two partition functions *Z*. At temperatures close to zero, only the ground level is populated, so the o/p ratio at LTE is close to zero. At low temperatures, only the two first excited levels are populated and so the o/p ratio at LTE gives $9\exp(-170K/T)$ where 9 is the ratio between the degeneracies of the two levels and -170 K their energy difference. At high temperatures we expect the ratio to be equal



Figure 7.9: Critical densities computed in our code for the excitation of H2, for different rotational lines including the collisional partners: H₂, H, H⁺, He and reactive collisions. In red is the rovibrational transition 1-0 *S*(1) shown previously in Fig. 7.1, in the energy diagram of the rovibrational levels of the electronic ground state. This graph assumes relative abundances x(He) = 0.1, $x(\text{H}_2) = 0.3$, and $x(\text{H}^+) = 10^{-4}$.

to the ratio of the degeneracies, which depends on the number of levels included in the calculation, but should be close to 3.

In Fig. 7.10 we compare the results of the H_2 excitation code with analytical calculations. As long as the chemical pumping is not included, the predictions of the code are in perfect agreement with analytical calculations. If the chemical pumping is included: H_2 is formed into a level *i* with a probability p_i that is proportional to the degeneracy of the level. Since the degeneracy of o- H_2 levels is 3 times larger than the para level, the o/p ratio can be close to 3 also at low temperature.

7.4 Including Electronic Excited Levels

When UV photons are absorbed, H_2 can be electonically excited, i.e. the Lyman and Werner bands get populated (see Fig. 7.2). The excited H_2 will than radiate back to the rovibrational excited levels of the electronic ground state, through a process called fluorescence, which occurs with an average probability of 85-90%. The other 10% leads to the dissociation of H_2 , that instead of radiating to a rovibrational ground state, radiates to the vibrational continuum of the ground electronic state. In this section, we study the effect of the fluorescence on the excitation of the rotational levels of H_2 .

7.4.1 UV Pumping

We consider the excitation from a rovibrational initial level of the electronic ground state $X(v_i, J_i) = X_i$ to the Lyman (or Werner) band B(v', J') (or C(v', J')) and the subsequent cascade to a final rovibrational state of the electronic ground state $X(v_f, J_f) = X_f$. The goal of this section is to create a matrix containing the probabilities to go from any initial states X_i to any final state X_f .



Figure 7.10: o/p ratio of H₂ computed for 100 levels, including radiative deexcitation, elastic collision with H⁺, H, H₂, H, reactive collisions with H⁺ and H, and chemical pumping. The blue line is the theoretical curve at LTE (see eq. 7.24). At low temperature, the o/p ratio is given by the lowest energy levels: J = 0 and J = 1. The purple curve is the theoretical ratio of the these two levels: $9\exp(-170K/T)$ The ratio at low densities (green crosses) is close to 3. This is due to the chemical pumping: H₂ is formed into a level with a probability p_i that is proportional to the degeneracy of the level, and the degeneracy of o-H₂ levels is 3 times higher than the para level. The green dots indicate the o/p ratio for a density of 10^{-2} cm⁻³ if the chemical pumping is not included.



Figure 7.11: Increasing the temperature the population of J = 0 decreases while population J = 1 increases. Starting from a certain temperarature population J=1 starts decreasing again, but we can see that thats because population J = 2 starts increasing. E(J = 1) = 170 K, E(J = 2) = 510 K. Crosses indicate the ground level J = 0, dots indicate the first excited rotational level J = 1 and the squares indicate J = 2, all in the ground vibrational level, i.e. v = 0. The solid line indicate the Boltzmann population of these levels as function of the temperature, and we can see that it is in good agreement with the result at high densities (densities > 100 cm⁻³). At low densities though, there are too few collisions to be LTE and the level J = 1 is more populated that the level J = 0 also at low temperatures, because we assumed that the formation of H₂ happens with a Boltzmann distribution on the levels.

Figure 7.12: Cartoon of the absorption of a UV photon from the molecule for hydrogen that transition from its ground electronic state $X(v_i, J_i)$ to Lyman B(v', J') or Werner C(v'', J'') band (in blue) and then radiates away through fluorescence and end up in the final state $X(v_f, J_f)$. This is the UV pumping, in which the population of the levels are modified by the UV radiation field illuminating the ISM.



Let's consider a Lyman or a Werner transition

$$X(v_i, J_i) \to B(v', J') \to X(v_f, J_f)$$
 or $X(v_i, J_i) \to C(v'', J'') \to X(v_f, J_f).$

We note $p_B^e(v', J')$ (or $p_C^e(v'', J'')$) the probability to excite an H₂ molecule from X_i to B(v', J')(or C(v'', J'')) and $p_B^d(v', J')$ (or $p_C^d(v'', J'')$) the probability to deexcite from B(v', J') (or C(v'', J'')) to X_f . The probability to go from X_i to X_f through B(v', J') (or C(v'', J'')) is simply $p_B^e(v', J') p_B^d(v', J')$ (or $p_C^e(v'', J'') p_C^d(v'', J'')$), if we then sum on all possible rovibration levels (v', J') (or (v'', J'')for Werner) we obtain the probability to go from X_i to X_f through any intermediate level of the Lyman (or Werner) band. The corresponding element in the UV excitation matrix reads

$$M_{UV}(i,f) = \sum_{\nu'} \sum_{J'} p_B^d(\nu',J') p_B^e(\nu',J') + \sum_{\nu''} \sum_{J''} p_C^d(\nu'',J'') p_C^e(\nu'',J'')$$
(7.25)

where p^e clearly depends also on the initial level X_i and p^d on the final level X_f , even though not indicated explicitally in the formula to simplify the notation.

In the following, we only describe the treatment applied to the Lyman band, for simplicity, but the same treatment is also applied to Werner. The probability of de-excitation from a level $u \equiv B(v', J')$ to $i \equiv X(v_f, J_f)$ for a dipole transition is defined as the ratio of the rate of that transition over the sum of all the possible transitions from u:

$$p_B^d(v',J') = \frac{A_{uf}}{\sum_f A_{uf}}$$
(7.26)

$$A_{uf} = 10^{16} \times \frac{8\pi^2 e^2}{m_e \lambda_{uf}^2 c} \frac{g_f}{g_u} f_{uf}.$$
 (7.27)

The rate of $u \to f$ is given by the Einstein coefficient for spontaneous¹, f_{uf} is the oscillator strength of the transition, λ_{uf} the wavelength of the emitted photon in Angstrom, m_e and e the mass and the charge of the electron, and the ratio of the statistical weights $\frac{g_f}{g_u} = \frac{2J_f+1}{2J_u+1}$

¹Einstein coefficients of absorption, induced emission and spontaneous emission in gauss units for which $1/4\pi\epsilon_0 = 1/4\pi\epsilon_0$



Figure 7.13: *Left:* H₂ level populations considering excitation from collisions, both reactive or not, and de-excitation due to collisions and spontaneous emission (blue), and chemical pumping (cyan), and UV pumping assuming a Habing field (red). *Right:* H₂ excitation including different radiation field prescriptions (see §3).

because the ortho levels in Lyman band correspond to even J, as introduced in $\$7.1^2$.

The rate of excitation from $X \to B$, $p_B^e(v', J')$ comes from the Einstein absorption coefficient from the initial level, *i*, to the upper level *u*, B_{iu} , and can be written as:

$$p_{B}^{e}(v',J') = B_{iu}u_{\nu} = \overline{n_{\gamma}}\frac{g_{u}}{g_{i}}A_{ui} = \frac{\lambda^{3}u_{\nu}}{8\pi h}\frac{g_{u}}{g_{i}}A_{ui} = \frac{\pi e^{2}\lambda u_{\nu}}{m_{e}hc}f_{iu} = 10^{-16}\frac{\pi e^{2}\lambda^{3}u_{\lambda}}{m_{e}hc^{2}}f_{iu}$$
(7.31)

where the mean number of photons averaged over all directions reads

$$\overline{n_{\gamma}} = \frac{c^2}{2h\nu^3} \overline{I_{\nu}} = \frac{c^3}{8\pi h\nu^3} u_{\nu}$$
(7.32)

with u_{λ} the energy density per wavelength units, and λ the wavelength in Å.

In Fig. 7.13, we show the effect of the chemical and UV pumping on the populations of H_2 . We can notice that the pumpings affect the populations of the levels by orders of magnitudes.

In this chapter we presented some tests of the excitation code against some analytical calculations. The results not only prove that the H_2 excitation code is reliable but also that it is important to take into account the ultraviolet and chemical pumping in the computation of the populations. The future goal of this excitation code is to become an integral

$$B_{\ell u} = \frac{\pi e^2}{m_e h \nu} f_{\ell u} \tag{7.28}$$

$$B_{u\ell} = \frac{\pi e^2}{m_e h v} \frac{g_\ell}{g_u} f_{\ell u} \tag{7.29}$$

$$A_{u\ell} = \frac{8\pi^2 v^2 e^2}{m_e c^3} \frac{g_\ell}{g_u} f_{\ell u}$$
(7.30)

²Therefore, the dipolar transition $\Delta J = \pm 1$ between *X* and *B* correspond to o-o or p-p transitions.

part of the chemical solver and be applied in post-processing to the ensemble of the simulations. This chapter presented the first step toward this more ambitious goal.

Part III

Comparison with Observations



The main goal of this work is to compare the outcomes of the simulations with the observed samples introduced in Part I (§4). In order to achieve this goal, we need to create a simulated sample of lines of sight statistically comparable to the observed sample. This is achieved by creating a set of synthetic lines of sight, patching together extractions from our MHD simulations (§8). The comparison is optimally achieved with a modified KS test that optimizes the comparison of our observed and simulated samples, on a multivariate statistical basis (§9). These comparisons are then analyzed in the subsequent chapters of this last part of the manuscript.

Chapter 8

Reconstruction Algorithm

This chapter presents and discusses how we have constructed sets of simulated lines of sight through neutral diffuse ISM, using the outputs of the parametric study presented in §5, with the goal to perform reasonably accurate comparisons with the observed samples presented in §4. While we had adopted a fiducial size of 200 pc for the MHD simulated lumps of diffuse neutral gas, we now have to take two important facts into account: the distribution of the lengths of the lines of sight through the ISM in the observed sample, and the heterogeneity of the medium along these lines of sight, as only a fraction of the lines of sight is expected to be neutral medium.

We first describe the method used to create our sets of simulated lines of sight (reconstruction). We then apply this method to two of our observed column density samples: the atomic vs molecular hydrogen sample (the HI-to-H₂ transition, see §8.2), and the CH⁺ molecular ion vs total hydrogen sample (see §8.3). In the last part of this chapter (see §8.4), we discuss some of the limitations of the method adopted. The scientific analysis and the conclusions will be drawn later in §10, §11, and §12.

8.1 Method

Let's consider an observational sample. The first step is to extract the distribution of lengths of the line of sight, l_{los} . Since the observational sample may contain extragalactic sources or galactic sources at high latitude, and since the amount of molecular gas in the Milky Way decreases exponentially as a function of the distance from the midplane, the length l_{los} occupied by the observed diffuse gas cannot always be identified to the distance of the background source. For the sake of simplicity, adopting a scale height of 100 pc above the galactic midplane for the cold ISM component, we estimate the length of one line of sight through the ISM as

$$l_{\rm los} = \min\left(\frac{1"}{p}, \frac{100}{\sin(|b|)}\right) \,\mathrm{pc},\tag{8.1}$$

where *b* is the Galactic latitude of the background source and *p* is its parallax.

The distribution of these estimated lengths for a given sample is built as a histogram regularly sampled on a logarithmic length scale. A simulated sample is then built with



Figure 8.1: Schematic view of the reconstruction of individual lines of sight over a distance l_{los} . The medium between the observer and the source is assumed to be composed of hot and warm ionized material (light blue cubes) with a volume filling factor φ and of uncorrelated pieces of diffuse neutral gas of individual size *L* (simulated boxes) with a volume filling factor (1- φ).

a distribution of lengths following this histogram. In practice for the construction of a sample of *N* synthetic lines of sight, we assign a length equal to the central value l_i of the histogram box *i* to a number

$$N_i = w_i N, \tag{8.2}$$

where the coefficients w_i are the normalised histogram values.

For a line of sight of length l_i , the length through the neutral medium is $(1 - \varphi)l_i$ on average, where φ is the still debated value of the fraction of the line of sight occupied by the ionized medium. We adopt a fixed value of $\varphi = 0.5$ (Hill et al. 2018) for this study and discuss the consequences of alternative choices in §8.4. We also use this value for all simulated lines of sight of length l_i , neglecting the intrinsic variations of the neutral to ionized medium along different lines of sight. Not doing so would only have a second-order effect on the statistical comparisons, as further discussed in §8.4.

The process of creating a synthetic line of sight of length l_{los} is sketched in Fig. 8.1 where patches of neutral and ionized media are stacked up in this proportion, up to the targeted length, from the observer towards the source. For the sake of economy, we chose to pick all neutral patches from a single simulation output, a cube of size L = 200 pc but concatenating different randomly selected cuts through the cube. The number of cuts to concatenate is equal to the fractional number $(1 - \varphi)l_{los}/L$. Alternatives to this simplifying choice are also discussed in §8.4.

8.2 Application: HI-to-H₂ Transition

As described in §4.1, the positions of the sources in the sky and their distances deduced from GAIA and Hipparcos measurements of parallaxes (Perryman et al. 1997; Gaia Collaboration et al. 2018) are shown in Fig. 4.1, where unknown distances of extragalactic sources (see paper in §10) are arbitrarily set to 1 Mpc.

The distribution of the lengths of the lines of sight through the diffuse medium, resulting from the computation of the length of the observed diffuse material, is shown in Fig. 8.2. The shortest lines of sight are found to extend over ~ 100 pc and the largest over ~ 3.5



Figure 8.2: Distribution of lengths l_{los} of the intercepted diffuse material computed with Eq. 8.1 along all lines of sight of the observational sample of $N_{\rm H}$ and $N({\rm H_2})$. The tomato sample corresponds to lines of sight where H₂ is detected and the sea-green sample to those for which an upper limit on $N({\rm H_2})$ has been derived.

kpc. Remarkably, and because of the combined distributions of distances and Galactic latitudes of the sources, we find that $\log(l_{los})$ follows a flat distribution up to $l_{los} \sim 2$ kpc with about 50 sources per bin and drops by about a factor of two for $l_{los} \sim 3$ kpc. Oppositely, and as expected, the distribution of lengths of non-detections of H₂ is not flat but decreases rapidly up to 1 kpc. Long lines of sight are finally not limited to the first and fourth Galactic quadrants but are found to spread over all Galactic longitudes and mostly depend on the Galactic latitude of the background source.

The reconstruction of the predicted probability histograms of atomic and molecular hydrogen were computed with m = 6 lengths of lines of sight $l_{los} = 100, 200, 400, 800, 1600,$ and 3200 pc, and the weights deduced from the distribution of distances in the observed sample (see Fig. 8.2), $w_1 = 0.14$, $w_2 = 0.21$, $w_3 = 0.16$, $w_4 = 0.20$, $w_5 = 0.18$, $w_6 = 0.11$.

The outcome of the reconstruction algorithm on the transition HI-to-H₂ is shown in Fig. 8.3 which displays the 2D probability histogram of the total proton column density $N_{\rm H}$ and the column density of molecular hydrogen $N({\rm H}_2)$ obtained with the fiducial simulation (see Table 5.1). Because of the flat distribution of distances in logarithmic space (see Fig. 8.2), the peaks of the reconstructed probability histogram are found to be shifted and spread toward both the large and the low values of $N_{\rm H}$ compared to those of the initial distribution (top panel of Fig. 8.3). This naturally enhances the initial bimodality and many lines of sight are found to have either low (~ 10⁻⁵) or large (~ 10⁻¹) integrated molecular fractions. In addition, it induces an inclination toward large column densities; more than half of the lines of sight are found to have $N_{\rm H} > 10^{21}$ cm⁻². By virtue of the central limit theorem, the molecular fraction obtained in those lines of sight tends toward the mean H₂ molecular fraction of the initial simulation with a dispersion that decreases as a function of $N_{\rm H}$. Detailed comparisons with observational data will be performed later in the manuscript (see §10).



Figure 8.3: 2D probability histogram of the total proton column density $N_{\rm H}$ and the column density of molecular hydrogen $N({\rm H}_2)$ obtained with the fiducial simulation (see Table 5.1). The upper panel shows the original data where all lines of sight have a size L = 200 pc. The bottom panel shows the outcome of the reconstruction algorithm described in §8 that produces a sample of lines of sight ranging from 100 pc to 3200 pc. The color code indicates the fraction of lines of sight (in logarithmic scale) contained in each bin. Dotted lines are isocontours of the molecular fraction for $f_{\rm H_2} = 10^{-8}$, 10^{-6} , 10^{-4} , 10^{-2} , and 1.

8.3 Application: $N(CH^+)$ vs N_H

In §4.3, we presented the observations of CH⁺ and we explained how the total column densities have been computed. For 190 of the 213 sources observed, we have information on the position and the distance of the sources, and the total column densities. The observational sample of $N(CH^+)$ vs $N_{\rm H}$ thus counts 190 sources, 32 of which are found to be not detected in CH⁺, and gives us upper limits.

We want to study the correlation of the column density, $N(CH^+)$ with the total proton column density $N_{\rm H}$, so, proceeding as before we reconstruct the 2D probability histogram of $N(CH^+)$ vs $N_{\rm H}$ generating a simulated sample with the same distribution of lengths $l_{\rm los}$ as for the observed sample.


Figure 8.4: The distribution of lengths l_{los} of the intercepted diffuse material computed with Eq. 8.1 along all lines of sight of the observational of CH⁺. The tomato sample corresponds to lines of sight where CH⁺ is detected and the green sample to those for which an upper limit on $N(CH^+)$ has been derived.

Extracting the positions and the distances of the sources, we notice that most of the sources are close to the Galactic Disk, with about 90% of the latitudes smaller than 15°. High latitude observations are towards galactic sources at a distance smaller than 100 pc (see Fig. 8.4). The distribution of the lengths of the lines of sight, resulting from the computation of the length of the observed diffuse material with the help of eq. 8.1, is shown in Fig. 8.4. For bins centered at $l_{los} = 100$, 200, 400, 800, 1600, and 3200 pc, the distribution has a peak with about 40 sources in the bin around 800 pc, and drops off at $l_{los} \sim 3$ kpc. In the first two bins, around 50% of the observations are non-detections, then, as for the observations of H₂, the distribution of the non-detections decreases quickly up to $l_{los} \sim 1$ kpc. Long lines of sight are not limited to the first and fourth Galactic quadrants but are found to spread over all Galactic longitude and mostly close to the Galactic equator.

We use eq. 8.2 to generate simulated lines of sight with the same distribution of lengths as the observed sample, i.e. adopting the weights $w_1 = 0.15$, $w_2 = 0.15$, $w_3 = 0.16$, $w_4 = 0.25$, $w_5 = 0.20$, $w_6 = 0.09$. The outcome of our reconstruction algorithm for CH⁺ using the fiducial simulation is shown in Fig. 8.5, assuming a volume filling factor of the ionized gas of $\varphi = 0.5$. The 2D probability histogram of the total proton column density $N_{\rm H}$ and the column density of CH⁺, $N({\rm CH^+})$ in the raw simulation (top panel Fig. 8.5) of L = 200 pc shows a bimodality, i.e. two peaks of probability, one at low column densities and the other at high column densities of CH⁺, which was not obviously expected. The peaks of the reconstructed probability histogram (bottom panel Fig. 8.5) are found to be spread towards both the large and the low values of the total proton column densities $N_{\rm H}$, enhancing the initial bimodality. As for H₂, detailed comparisons with observations are performed later in the manuscript (see §11).



Figure 8.5: 2D probability histogram of the total proton column density $N_{\rm H}$ and the column density $N({\rm CH^+})$ obtained with the fiducial simulation (see Table 5.1). The upper panel shows the original data where all lines of sight have a size L = 200 pc. The bottom panel shows the outcome of the reconstruction algorithm described in this section that produces a sample of lines of sight ranging from 100 pc to 3200 pc. The color code indicates the fraction of lines of sight (in logarithmic scale) contained in each bin.

8.4 Tests and Caveats

The construction of synthetic lines of sight, explained in §8.1, is based on the concatenation of randomly selected cuts through a single simulation output cube, assuming a uniform filling factor φ for the ionized gas in the local interstellar medium. The consequences of these simplifying assumptions are successively examined in the present section through the effects of changing the random concatenation process, patching cuts from different simulations or from independent outputs of a single simulation, and varying the value of the assumed filling factor.

8.4.1 The Random Drawing Process

Reconstruction by Translation of the Box We introduce here an alternative method to generate lines of sight with the same distribution of lengths as the observational sample. When the column densities are generated as explained previously, the random process involved decorrelates the spatial structure (see the left panel in Fig.8.7).

Instead of drawing independent segments for each of the reconstructed lines of sight, a long patch can be created once with the concatenation of random translations of the same cube. All lines of sight can then be drawn from this long patch. Indeed, since the boundary conditions of the simulation are periodic, the simulated box repeats itself every period *L* in the *x*, *y*, *z* directions and each point of the box (X,Y) can be seen as the origin of a new translated box. Fig. 8.6 shows a schematic view on how the box-translation is built. A point (X, Y) is drawn randomly in the simulation box and divides the cube into 4 parallelepipeds (A,B,C,D). These parallelepipeds are rearranged to create a new simulation box (C,D,A,B) which is a simple translation of the original box. In the box-translation reconstruction, the column densities of the generated lines of sight are computed as an integration over boxes with different origin points. Following the explanation in the precedent section, if $(1-\varphi) l_{los} > L$, we create $L/[(1-\varphi) l_{los}]$ translated boxes by drawing random origin points (see Fig. 8.6) and then we add the respective individual column densities to create the column densities of lengths l_{los} , therefore, the building blocks of the generated lines of sight are boxes translated by a random distance the one compared to the other.

In the right panel of Fig. 8.7, we can see how the reconstruction with the box-translation algorithm keeps information on the structure of the matter. This alternative method should be considered for studies involving the spatial structure.

Comparison Translation and Random Reconstruction In Fig. 8.7, we show the reconstructed total column density of length $l_{los} = 800$ pc for the fiducial simulation. On the left with the random method explained in the previous section, and on the right with the box-translation method. We notice that they both extend on the same range of column densities, but the random one does not contain any information on the original structure. In Fig. 8.8, we show the isocontours of the 2D probability distributions found with both methods. The two methods give the same statistical results on the column densities, allowing us to compute the probability histograms, necessary to our study, with the random method, which has the advantage of being simpler. Despite the fact that this reconstructions.



Figure 8.6: Schematic view of the translation of a simulation. The original simulation cube (magma colors) is seen from its face (x,y) and the z-axis goes into the plane. Because of the periodic boundary conditions of the simulation, the original simulated box repeats itself in each directions with a period *L* (blue colors). A point (X, Y) of the original simulation is drawn randomly and divides the cube into 4 parallelepipeds (A,B,C,D). A new translated cube (the bright one) is created with (X,Y) as origin.



Figure 8.7: Maps of the total column densities of a $l_{los} = 800$ pc, reconstructed from the fiducial simulation, assuming a filling factor of the ionized gas of $\varphi = 0.5$. On the left side with a random reconstruction and on the right side using a translated box. The colorbar indicate the total integrated proton densities in logarithmic scale. In the random case, neighbor pixels are completely uncorrelated and we lose all information on the structure of the neutral medium, while in the translation case we keep information on the structure coming from the original box of L = 200 pc. We note that in the both maps we have the range of simulated column densities and the same statistical distribution of the reconstructed total column densities.



Figure 8.8: 2D probability histogram of the reconstructed column densities of $N_{\rm H}$ and $N({\rm H}_2)$ with the random method explained in the main text. The solid and dashed lines are the isocontours for probabilities 10^{-4} (black), 10^{-3} (dark gray), 10^{-2} (gray) for the random and box-translation methods, respectively.

tion method has no use in cases where the information on the structure is key, it is a very useful method in our work, because compared to the box-translation method it has another perk. The different lines of sight to the observed sources may start from the same local ISM patch, but further from the observer, the patches traversed become less and less correlated. Hence drawing the successive cuts independently from each other is better than drawing all lines of sight from a single concatenation/realization. For these reasons, the random reconstruction seems a better method to model the observed lines of sight in which no correlation between the different neutral clouds is assumed/observed.

8.4.2 The Choice of Concatenated Patches

Different Simulations Yet a better way of generating lines of sight of length l_{los} would be to take different simulation runs with the same set of parameters and stock them one behind the other to reconstruct our lines of sight. This procedure would allow to have a coherent structure in each piece of neutral medium and conserve structure information in the column densities map, without assuming any spatial correlation between contiguous pieces of the neutral medium along each line of sight. This method has a high computational cost because the number of simulations required (that need to be run and post-processed) to perform the parametric study would be at least 8 times larger than the current one. To address this issue, but still be able to analyze a possible result, we used different outputs of the same simulation. We chose two outputs separated by one correlation time because after a correlation time the initial conditions are erased. Since the correlation time of the turbulent forcing was set to the turbulent turnover time in our simulations, the time difference between the two outputs needs to be larger than $\tau_l = l/\sigma_l$ where *l* can be any scale. Since we want the turnover time of the driving scale, $l = L_{drive}$, the velocity dispersion at $L_{drive} = 100$ pc is of $\sigma_{100pc} \sim 6$ km s⁻¹, giving a correlation time of around 16 Myr.

Effect of Different Outputs on the Random Reconstruction To see the effect of this approach, we analyze the random reconstruction done on two different outputs. In the reconstruction explained in the previous paragraph, each output would model a piece of neutral interstellar medium along lines of sight, therefore there would be a contribution of different outputs along one line of sight. Here we chose to apply our original method to the two different, assumingly independent, outputs, in order to assess the sensitivity of the simulated histograms to the choice of the simulation output.

In the left panel of Fig. 8.9, we can see that the total column density maps integrated over L = 200 pc essentially show no correlation in the structure of the matter. The right panel shows both histograms for the reconstructed lines of sight using the later output as a building block for the colored probability histogram and the solid lines, and the early output for the dashed lines. We notice that the difference between the dashed and solid curves is insignificant, validating the random reconstruction approach. We remind the reader that the more complex method would consist in taking not one building block (as done here), but $L/[(1 - \varphi)l_{los}]$ different building blocks, each one corresponding to a different output, thus any difference visible on the right panel of Fig. 8.9 would be diminished.



Figure 8.9: On the left the total column densities integrated over L = 200 pc for the fiducial simulation at two different times, 20 Myr (up) and 36 Myr (down), on the right the 2D probability histogram of the reconstructed column densities of $N_{\rm H}$ and $N({\rm H}_2)$. The solid and dashed lines are the isocontours for probabilities 10^{-4} (black), 10^{-3} (dark gray), 10^{-2} (gray) of the outputs at 20 Myr and 36 Myr, respectively.

In conclusion, using different simulations with the same set of parameters, assures both the spatial uncorrelation between pieces of neutral medium along the lines of sight and the possibility of computing column density maps, because the structure stays coherent. Even with the simplifications, we adopted (using different outputs of the same simulation which are separated by at least one correlation time), the computational cost of this analysis stays large compared to the random method, it would necessitate the extraction of the physical quantities needed for the analysis of 8 different outputs at 16 Myr of difference at least, which is time consuming and asks for longer simulations, which are computationally expensive. Moreover, the difference between the results using the two methods is small and the increase of analytical cost does not justify the use of this method.

8.4.3 Assumptions on the ISM Structure around the Sun

Hot and Warm Ionized Gas Volume Filling Factor The fraction of volume occupied by the hot and warm ionized gas plays a significant role in our recontruction algorithm and its subsequent analysis, however, its value is still debated. The consensus is that the volume of the HIM far exceeds that of the WIM and results from an interplay between supernovae explosions, which regularly produce hot gas in the Galactic disk, and buoyancy, which lifts this gas into the halo, releasing the pressure in the midplane. Early analytical studies neglecting buoyancy (McKee & Ostriker 1977) predicted a large fraction of HIM in the midplane ($\varphi \sim 95\%$). In contrast, early numerical simulations, including the cycle of matter and energy between the disk and the halo, led to considerably smaller predictions with $\varphi \sim 25\%$ (de Avillez & Breitschwerdt 2004). This value is now considered as a lower limit by the most recent numerical simulations which reveal the importance of the driving mode of supernovae explosions (Walch et al. 2015) and of the photoelectric heating (Li et al. 2015;



Figure 8.10: 2D probability histogram of the total proton column density N_H and the column density of molecular hydrogen $N(H_2)$ obtained with the fiducial simulation (see Table 5.1). The upper panel shows the original data where all lines of sight have a size L = 200pc. The bottom panels shows the outcome of the reconstruction algorithm that produces a sample of lines of sight ranging from 100 pc to 3200 pc assuming different volume filling factors for the ionized medium. The color code indicates the fraction of lines of sight (in logarithmic scale) contained in each bin. Dotted lines are isocontours of the molecular fraction for $f_{\rm H_2} = 10^{-8}$, 10^{-6} , 10^{-4} , 10^{-2} , and 1.

Hill et al. 2018) on the volume of the HIM. These latest studies estimate that $20\% \le \varphi \le 90\%$ in the Galactic midplane.

Highly uncertain, the volume filling factor of the HIM can also vary from one line of sight to another (Fig. 1 of Hill et al. 2018). In Fig. 8.10, we tested the results of the reconstruction algorithm assuming different values of φ . Increasing φ has the effect of stretching the predicted 2D probability histogram toward low values of $N_{\rm H}$ along the *x*-axis, and gives a wider HI-to-H₂ transition, that can extend on almost 3 orders of magnitude from $N_{\rm H} \sim 10^{18} \text{ cm}^{-2}$ to $N_{\rm H} \sim 10^{21} \text{ cm}^{-2}$ in the extreme case in which most of the line of sight is composed of the hot and warm ionized medium. The balance of probabilities of occurrence of lines of sight at high and low molecular fraction is also modified, and the occurrence of low column densities increases as a function of φ .

Because of the lack of additional information and since no major difference is seen in the probability histogram distributions for value of φ given in the litterature, i.e. $0.25 < \varphi < 0.75$, we assume here a conservative value of $\varphi = 0.5$. Taking into account a realistic distribution of φ would require to simulate the Galactic disk and halo over a scale of several kiloparsecs and to properly model and follow the impact of supernovae explosions. This is far beyond the scope of the present work.

One Realization We note that, while spatially uncorrelated, the pieces of diffuse neutral gas used in the reconstruction algorithm correspond to random realizations obtained with a single simulation. The approach of assuming that the local ISM can be built out of a single simulation was chosen in order to highlight the natural variations induced by turbulence and thermal instability alone in a diffuse neutral gas with a known averaged density and UV illumination factor. However, because the medium probed by the observations extends in all directions around the sun over a maximum distance of 3 kpc (see Fig. 8.2), it stands to reason that potential variations of all parameters should be taken into account, not only from one line of sight to another but also along a single and outstretched line of sight. We ignore this aspect here because it would drastically increase the parameter space to explore. However, such consideration needs to be kept in mind for the comparison with the observations.

Chapter 9

Kolmogorov-Smirnov Statistics

As shown in the previous chapter, the originality of the work presented here is to compare simultaneously the chemical and statistical information of observed and simulated samples by the mean of probability histograms (PHs), or equivalently, probability distribution functions (PDF). In many ways, the direct comparison of probability histograms, such as those shown in Figs. 8.8 and 8.10, provides valuable information regarding the statistical nature of the correlation between different species. Unfortunately, such an approach is impractical to use for the comparison of an observational sample to a large set of simulations (obtained with a parametric study) or if the chemical correlations are studied in dimensions larger than two.

For all these reasons, we developed a procedure to measure with a single value, the degree of similarity between two probability distribution functions. This procedure, based on the well-known Kolmogorov-Smirnov (KS) test, can be used in any dimension and can take into account all observational information including actual detections and lower or upper limits. In the following sections we will introduce the classic Kolmogorov-Smirnov test (9.1), its extension to N-dimensional samples, the modification, and definition of the KS distance, R_{KS} , and its application on two observational samples.

9.1 Classic KS test

The KS test is a cumulative test originally built to compare two PDFs in one dimension. In a nutshell, the KS test consists in computing the KS difference which measures the maximum value of the absolute distance between two distribution functions (two-sample KS test),

$$D = \max_{-\infty < x < \infty} \left| S_{N_1}(x) - S_{N_2}(x) \right|$$
(9.1)

or between one sample and a model (one-sample KS test),

$$D = \max_{-\infty < x < \infty} |S_N(x) - P(x)|.$$
(9.2)

 $S_N(x)$, $S_{N_1}(x)$ and $S_{N_2}(x)$ are cumulative distribution functions, i.e. functions giving the fraction of datapoints smaller than *x*, and *P*(*x*) is a known cumulative probability distribution.

Figure 9.1: Two-dimensional distributions of 20 observational points. Each observational datapoint, identified by a pair of variables (X^{obs} , Y^{obs}), divides the space into four quadrants A, B, C, D.



Widely used by all communities, the original KS test has several advantages and disadvantages. On the one hand, the definition of the test is independent of the underlying cumulative distribution function and it can be applied to samples of any size. Moreover, the KS test not only gives a measure of the degree of similarity between the distribution (the KS distance) but also the position *x* where the similarity is minimal: it therefore partly captures the region (in the *x*-space) where the two distributions are different and can be used for identifying shifts between distributions. On the other hand, because it is applied on cumulative distributions, the KS test is more sensitive to differences near the median of the distribution than differences in the tails. In addition, such a test cannot be generalized in dimension larger than one, because there is no unique way to order the point so that distances between two cumulative distributions can be computed.

To overcome this issue, we modify the definition of the KS test following the works of Fasano & Franceschini (1987) and Peacock (1983). The idea is to replace the cumulative probability distribution, which is not well defined in a dimension larger than one, with the integrated probability in each quadrant surrounding a datapoint (see Fig. 9.1). Such a consideration, detailed below, allows to define a KS distance which measures how two distribution functions differ from one another in N dimensions.

9.2 The Modified Version of the KS Test

9.2.1 Definition and Interpretation

The methods and illustration described here are shown in 2D but they can all be applied in larger dimensions. Let's consider two datasets, one observed (labeled 'obs') and one simulated (labeled 'sim').

As illustrated in Fig. 9.1, each observational datapoint is identified by a pair of variables $(X^{\text{obs}}, Y^{\text{obs}})$ which divide the space into four quadrants:

A : $X \leq X^{\text{obs}} \& Y \leq Y^{\text{obs}}$,



Figure 9.2: Example of the KS procedure in 2D. Twodimensional distributions of 20 red squares (the observational dataset) and 100 lightblue dots (the simulated dataset). Our modification of the KS ratio finds the datapoint and the quadrant that maximize the merit function. The darker red squares, indicate the subsample of the observational dataset (coral squares) \mathcal{O} used to compute the KS distance, this subsample was chosen such that all quadrants scanned contain at least 1 observational datapoint.

- B : $X \leq X^{\text{obs}} \& Y > Y^{\text{obs}}$,
- C : $X > X^{obs} \& Y > Y^{obs}$,
- D : $X > X^{\text{obs}} \& Y \le Y^{\text{obs}}$.

As shown in Fig. 9.2, each quadrant thus contains two fractions f_{obs} and f_{sim} of the entire observed and simulated datasets, which correspond to the cumulative probability in 1D. To compare these values, we define a merit function

$$M = \left| \log_{10} \left(\frac{f_{\rm sim}}{f_{\rm obs}} \right) \right|,\tag{9.3}$$

and the KS distance between the two distributions as the maximum value of M computed over all quadrants of an observational dataset \mathcal{O} ,

$$R_{\rm KS} = \max_{\mathcal{O}}(M). \tag{9.4}$$

This procedure, not only provides a measurement of the difference between the two distributions, but also the datapoint and the quadrant that maximize the merit function. For instance in Fig. 9.2, we show the datapoint and the quadrant (in gray) that maximizes the merit function. The interpretation is also straightforward. For example, a KS distance $R_{\text{KS}} = 1$ implies that one of all the quadrants scanned contains 10 times fewer or 10 times more detections than simulated lines of sight, and that all the other quadrants have smaller distances.

9.2.2 Error on the Merit Function

The stability of the procedure depends on the errors made on the merit function and therefore on the absolute numbers of observed and simulated lines of sight contained in each quadrant. The observational dataset \mathcal{O} used to compute the KS distance is chosen as the subsample such that all quadrants scanned contain at least n_{thr} observed lines of sight. With this assumption, the error on the merit function is calculated by taking into account only the statistical errors on the number of simulated lines of sight. For each quadrant, we assume that the "true" merit function ranges between

$$M_{\rm minus} = \left| \log_{10} \left(\frac{f_{\rm sim} - 3\sqrt{(f_{\rm sim}/S)}}{f_{\rm obs}} \right) \right| \text{ and}$$
(9.5)

$$M_{\rm plus} = \left| \log_{10} \left(\frac{f_{\rm sim} + 3\sqrt{(f_{\rm sim}/S)}}{f_{\rm obs}} \right) \right|,\tag{9.6}$$

where *S* is the total number of simulated lines of sight. Because the errors are asymmetric, M_{plus} or M_{minus} can tend toward infinity. If so, the KS distance is a lower limit, even if the infinite error bar is obtained for another quadrant than the one that maximizes *M*. In short, for each observational datapoint, we compute M_{minus} , *M*, M_{plus} , and R_{KS} . If one of the M_{minus} goes to infinity, R_{KS} is considered as a lower limit. The lower end of the errorbar is given by the minimum value between M_{minus} and M_{plus} evaluated in the quadrant that maximizes the KS difference, R_{KS} .

$$M_{\min} = \min_{max(M)} (M_{\min us}, M_{\text{plus}}), \tag{9.7}$$

$$M_{\max} = \begin{cases} \max_{\mathscr{O}} (M_{\min us}, M_{plus}) & \text{if either} \to +\infty \\ \max_{max(M)} (M_{\min us}, M_{plus}) & \text{if both finite} \end{cases}$$
(9.8)

The definition adopted above, not only allows to compute the merit function value but also to compute the errors on the merit function. We should bear in mind that the number of scanned points during the KS procedure, and the errors on the merit function depend on the minimal number of observations per quadrant, n_{thr} . We will show the impact of n_{thr} in the following section.

9.2.3 Influence of the Distribution of Datapoints

Another important aspect for the stability of the KS test is linked to the shape of the observational dataset. Since the cumulative probability is defined as the fraction of points per quadrant and since the division of quadrants is performed on a cartesian grid, the result of the KS test, explained in the previous section, depends on the choice of the axes.

To explain this concept, let's imagine we observe two quantities *X* and *Y*, which are strongly correlated, as in the left panel of Fig. 9.3. We can see a real correlation between *X* and *Y* and if we impose $n_{\text{thr}} = 1$ datapoint, we scan only 15% of the observed datapoints. If we perform a rotation on the observational sample (see right panel of Fig. 9.3), and thus we plot Y - aX, where *a* is the slope of the correlation between *X* and *Y*, we see that the distribution is now flat and the percentage of the scanned point increases greatly (to 90% in the example) and therefore the subsample over which the KS distance is computed is more populated.

For any observed quantities, *X* and *Y*, the best mathematical option would be to identify the principal components of the observational sample using proper orthogonal decomposition or singular value decomposition algorithms. Such a method could even be



Figure 9.3: Red points correspond to the scanned points, if X and Y are strongly correlated as in the left panel, the number of scanned points is low, while when turning the axis (right panel) the dispersion of the points increases and the number of scanned points increases, imposing a certain threshold, here of 1 point per quadrant.

applied to subsamples in order to adaptively rotate the system of axes and follow more precisely the distribution of observations. In any case, the resulting system would be a linear combination of *X* and *Y* which could be difficult to relate to the underlying physical properties.

In the following sections, we will show the choices we made for the application to real observations, where we decide not to apply the orthogonal decomposition to stay close to the physical interpretation.

9.3 Application: HI-to-H₂

In this section, we apply the KS test on the HI-to- H_2 transition, after choosing a judicious orientation of the axes and choice of n_{thr} .

9.3.1 Choices of the Axes and $n_{\rm thr}$

Because the division in quadrants is performed on a cartesian grid, the first step of the procedure is to choose the orientation of the axes in such a way that the distribution shows no visible correlation between the *x* and *y* axes. As explained above, the best choice of the axes orientation is a linear combination such that the points are spread in both directions. We choose here $N_{\rm H}$ and $f({\rm H}_2)$ as primary variables to ensure some homogeneity without compromising the physical interpretation of the results.

In Fig. 9.4, we show how this choice corresponds to a rotation of the *y*-axis and ensures some homogeneity of the spread of the observational data in all quadrants. When we consider the fraction of H₂, $f(H_2) = 2N(H_2)/N_H$, instead of the column density of H₂, $N(H_2)$, we increase the number of points in the subsample over which the KS test is performed, \mathcal{O} , that goes from 106 to 150.

In Fig. 9.5, we see how the choice of n_{thr} affects the number of points in the subsample used for the KS test. The subsample \mathcal{O} counts 287 points with $n_{\text{thr}} = 1$; 150 points for $n_{\text{thr}} = 10$; and 81 points for $n_{\text{thr}} = 20$. The optimal choice of n_{thr} comes from a compromise of



Figure 9.4: Observational dataset for the transition of HI-to-H₂. On the left, $N(H_2)$ as a function of $N_{\rm H}$, and on the right, $f(H_2)$ as a function of $N_{\rm H}$. The red points indicate the subsample used to compute the KS difference imposing a threshold of 10 observations in each scanned quadrant, and the ratio indicate the fraction of points in the subsample. The dotted lines are iso-fraction of molecular hydrogen.



Figure 9.5: The complete observational dataset in black and the subsample of the observational datased used to compute the KS distance in red. The subsample is chosen so that that all quadrants scanned contain at least a fraction of all observed lines of sight equal to the threshold written in gray. The red number into bracket indicate how many observational datapoints are scanned assuming the threshold written above.

having points canvassing the whole space on one side, i.e. scan also point belonging to the low column densities bump, while having small errors on the merit function on the other side. Taking into account those consideration, for the HI-to-H₂ transition, we take $n_{\text{thr}} = 10$ (central panel in Fig. 9.5).

9.3.2 Results

In Fig. 9.6, we show the result of the modified KS test applied to the HI-to-H2 transition for the fiducial model (see Table 5.1). We can notice a remarkable agreement between the simulation set and the observable sample, with a KS difference of the order of 1. This implies that in the gray rectangle, there are 10 times more observations than simulated lines of sight. All the other quadrants defined by the red points as origin have a smaller $R_{\rm KS}$. The KS procedure not only gives the $R_{\rm KS}$ value but also the quadrant that maximizes the merit function. This last piece of information has a physical relevance and indicates that the simulated dispersion of the fraction of molecular hydrogen is slightly lower than the observed.



Figure 9.6: Results of the modified KS test applied to the HI-to-H2 transition for fiducial simulation. The black squares are the observational data, the red square indicate the dataset used for the estimations of the merit function M, and the 2D histogram the simulated data. The gray star and rectangle indicate the observational point and the quadrant that maximize M (see main text). The fiducial simulation has a KS distance of 0.97. The corresponding quadrant contains 0.49% and 4.49% of the entire simulated and observed datasets.



Figure 9.7: KS distance between the simulations and the observational sample as a function of the mean density $\overline{n_{\rm H}}$, the UV scaling factor $G_0 = 0.5$ (red), 1 (green), 2 (orange), and 4 (blue), and for 4 different outputs. All other parameters are set to their fiducial values (see Table 5.1). Points correspond to reliable measurements of the KS distances. Triangles indicate lower limits corresponding to simulations where the upper error bar on $R_{\rm KS}$ tends toward infinity. The gray area indicate the errorbars going from $M_{\rm min}$ to $M_{\rm max}$ that include the maximal value of the merit function, $R_{\rm KS}$. The variation between the different outputs is small and lies in the range defined by the errorbars.

Since the KS test reduces the comparison of a simulation and an observational sample to a single number, we can use the KS difference to evaluate a whole grid of simulations. The physical interpretation of the results deduced from the KS distances applied to the HI-to-H₂ on the whole grid of simulations presented in Table 5.1 and counting 305 simulations, is given in the paper (see \$10).

In Fig. 9.7, the results of the KS test applied to a grid of 16 simulations obtained for different values of $\overline{n_{\rm H}}$ and G_0 and for different outputs of the same simulation. We can notice that the KS test offers a measure that is sensitive to the distribution and changes by orders of magnitudes from different values of the parameters of the grid. To verify the stability of the KS test, and the steady-state of the simulation, we computed the KS distance for the last 4 outputs of each of the simulation of the grid. The values of the KS test for the different outputs vary inside the respective estimated errorbars. This shows that the state of the simulation is actually stationary, as predicted in §5.5.1, and shows the stability of the KS test because the variation obtained from one output to the other is of the order of magnitude of the estimated error.

The KS test turns out to be a valuable tool for estimating the distance between two distributions without performing a detailed comparison of the samples.

9.4 Application on CH⁺

In this section, we compare the probability histogram distributions of the reconstructed column densities $N_{\rm H}$ and $N({\rm CH^+})$ with the observational sample. The simulated probability histogram for $N_{\rm H}$ and $N({\rm CH^+})$, as for $N_{\rm H}$ and $N({\rm H}_2)$, shows two peaks, one at high and one at low column densities, with a transition of about three orders of magnitude. Nevertheless, oppositely to what was seen for the transition HI-to-H₂, these two bumps do not appear in the observational sample (see Fig.9.8). The instruments are not sensitive enough to detect column densities of CH⁺ in the low densities peak, i.e. $N({\rm CH^+}) \sim {\rm afew} 10^7 - {\rm afew} 10^9 {\rm cm}^{-2}$, and therefore, in the low column density peak there are only non-detections.



Figure 9.8: Comparison of the observational dataset (black points) to the 2D probability histogram reconstruction algorithm (see §8.1) applied to the column densities of CH^+ and the total proton column densities (colored histogram). Observations include detections of CH^+ (circles) and upper limits on $N(CH^+)$ (triangles). The color code indicates the fraction of lines of sight (in logarithmic scale) contained in each bin.

9.4.1 Choice of the Axes Orientation and *n*_{thr}

To perform the KS test, we need to choose the right orientation of the axes and a threshold that grant us to avoid the divergence in the merit function and at the same time allows us to scan enough points. Since CH⁺ exhibits a great dispersion in the observational sample, there is no advantage in taking the fraction of CH⁺ vs $N_{\rm H}$ instead of the column density of CH⁺. Therefore, we perform the KS test of $N(\rm CH^+)$ vs $N_{\rm H}$. We chose a $n_{\rm thr}$ that allow to scan around 50% of the points, as for HI-to-H₂, i.e. $n_{\rm thr} = 6$.

9.4.2 Results

In Fig.9.9, we show the result of the KS test for the fiducial simulation. The KS distance has a value of 0.86, i.e. in the quadrant that maximizes the merit functions, the fraction

Figure 9.9: Results of the modified KS test applied on the probability histogram of $N(CH^+)$ vs N_H for the fiducial simulation. The black points are the observational data, the red dots indicate subsample, i.e. the dataset used for the estimations of the merit function M, the white triangles are the upperlimit that are ignored by the KS test, and the 2D histogram the simulated data. The gray star and rectangle indicate the observational point and the quarant that maximize M (see eq.9.3)). The fiducial simulation has a KS distance of 0.86. The corresponding quadrant contains 46.45% and 6.37% of the entire simulated and observed datasets.



of simulated lines of sight is 7 times bigger than the fraction of detections, and all the other quadrants give better ratios. Since there are two bumps in the simulated probability histogram and this distribution is not followed by the observational sample, the value that maximizes the merit function is always found in the A quadrant (see Fig.9.9) which is characterized by a richness in simulated lines of sight and very little detections. We can also notice that quadrant A contains most of the upper limits and that if those upper-limits were taken into account the percentage of observational points in quadrant A would go from 6.37% to around 20%.

Fig. 9.9 shows a big caveat of the KS test, in which only information from detections are taken into account. This could lead to wrong conclusions when there is a bias in the observational sample, for instance, if the telescope does not have the sensitivity to detect the column densities of CH^+ , because they are actually of a few 10^8 cm^{-2} . For these reasons, the results from the KS test applied to certain chemical species can be weak. This problem will be addressed in §9.5.

9.5 Extension of KS test to Non-Detections

The major limitation of the method described in this chapter is that we only compare the simulation set to the detection sample and we do not use the information in the non-detections. The goal of this section is to try to address this problem, by using the information contained in the non-detections in the modified version of the KS test.

In the study of the correlation of CH⁺ column densities with the total column densities, Fig. 9.9, we notice that the results from the KS test on detections are not strong and are biased toward the range of column densities in which there are only upper-limits. To overcome this issue, we could either limit our study to the column densities spanned by the observations – which limits the goal and strength of the comparison of simulations to observations – or try to extract some information from the non-detections. To do so, and integrate the non-detections in the KS distance, we assume that a non-detection is equivalent to an upper limit, Y_{upl} , i.e. it signals a detection at an unknown column density $Y < Y_{upl}$ and $X = X_{upl}$.

The subsample \mathcal{O} of the scanned datapoints assuring a minimal fraction of observations per quadrant does not change from the previous version, i.e. only detections give receivable couples of values (X^{obs}, Y^{obs}) . The difference between the previous version of the KS test resides in the fraction of observed datapoints per quadrant, which now can include also upper-limits. The contribution to the fraction of observations, f_{obs} , of each non-detection depends on its value Y_{upl} . For all non-detections with $Y_{upl} < Y^{obs}$, we can assume that the "real" detection belongs to the corresponding quadrant and, therefore, contribute to the fraction of observations in a quadrant for $1/n(\mathcal{O})$. For non-detections in the upper quadrants (B and D), i.e. $Y_{upl} > Y^{obs}$, nothing sure can be deduced to such a value and it is therefore excluded, i.e. it does not count in any quadrant. In Fig. 9.10, we give an example of the method. The red points are the detections that are scanned imposing a threshold on the number of observations - both detections and non-detections - in each quadrant. When a detection point is scanned (X^{obs}, Y^{obs}) , it divides the space into four quadrants (dark red). Non-detections are indicated with triangles, the upper limits with $Y_{\rm upl} < Y^{\rm obs}$ are counted in quadrants A or D as detections (light green triangles), the upper limits with $Y_{upl} > Y^{obs}$ are excluded (gray triangles). This is one way to tackle the problem, in §9.5.2 we discuss different assumptions and subsequent methods to count for the contribution of an upper-limit in a quadrant.

9.5.1 Application on CH⁺

We applied the modified KS test, including the non-detections as explained in the previous section and we show the result of this modified KS test in Fig. 9.11. The orientation of the axis is $N(CH^+)$ vs N_H and $n_{thr} = 6$, as before. Taking into account also information from the upper-limits the KS distance has improved (cf Fig. 9.9). While from the previous KS test on detections, we could only conclude that the main difference between the observational dataset and the simulated one was due to the non-detections (see Fig. 9.9), in the new KS test case we can extract more valuable information. The upper-limits are located in the same range of N_H as the low- $N(CH^+)$ bump, so the A quadrant is not the quadrant that maximizes the merit function. The quadrant that maximizes the merit function. The quadrant that maximizes the merit function is now quadrant D (indicated with a gray rectangle in 9.11), where the fraction of sight. This indicates that either the dispersion of the simulated lines of sight in the high- $N(CH^+)$ bump is larger than the observed one, or the simulated low- $N(CH^+)$ bump extends toward unrealistically large N_H , which could suggest that the transition between the two simulated bumps happens at sightly too larger total column densities.

In general, the KS distance is smaller when the upper limits are taken into account. Not all distances computed including the upper limits show as much variation as the fiducial case. In §11 we show the KS differences computed with and without the non-detections.

When including the upper-limits in the KS test applied to CH⁺ we obtain a value of 0.47



Figure 9.10: A 2D distribution of observational datapoints composed of detections (red points) and upperlimits (triangles). Each observational datapoint, (X^{obs} , Y^{obs}), indicated here with a dark red dot, divides the space into four quandrants A, B, C, D. Each detection contributes to the fraction of simulation, f_{obs} , for $1/n(\mathcal{O})$ in the quadrants where it belongs and 0 in the other three. The same is supposed to be true for the non-detections. We assume that only upper-limits belonging to the lower quadrants (A and D) count (green points) and non-detections happening at larger column densities, i.e. $Y_{upl} > Y^{obs}$, are excluded (light gray points).



Figure 9.11: Result of the modified KS test including the non-detections applied to compare the observational sample of the CH⁺ observations to the simulated lines of sight created with the fiducial simulation. Detections are indicated with squares and upperlimits with triangles. The red squares indicate the subsample \mathcal{O} scanned for the estimation of the merit function *M*, assuring a number n_{thr} of observations in each quadrant. The gray star and rectangle indicate the observational point and quadrant that maximize the merit function *M*. The fiducial simulation has a KS distance of 0.47. The corresponding quadrant contains 19.14% of the simulations and 6.42% of the observations.

for the KS distance. This means that in the worst quadrant the fractions of simulations and observations differ by a factor of less than three.

9.5.2 Different Assumptions to count the contribution in a quadrant of a non-detection

The KS test including the non-detections is a great improvement on the method. Here we proposed two additional more convoluted methods.

Half Method An attempt to take into account all upper-limits, also the one with $Y_{upl} > Y^{obs}$, i.e. belonging to B or C quadrants (see 9.12), would be to assume that the corresponding detection has the same probability of being larger or smaller column densities than Y^{obs} . This translates into giving a weight of one half in both A and B (or C and D) quadrants to all detections that belong to B (or C). Taking the same example as before, we show the concept in Fig. 9.12. Detections (red points) in each quadrants count as $1/n(\mathcal{O})$ in f_{obs} , upper-limits with $Y_{upl} < Y^{obs}$ counts as detections, i.e. $1/n(\mathcal{O})$ (light green triangles), and upperlimits with $Y_{upl} < Y^{obs}$ counts as half in both the upper and lower quadrant (A-B or C-D depending if $X_{upl} < X^{obs}$ or $X_{upl} > X^{obs}$), i.e. a contribution of $0.5/n(\mathcal{O})$ in each (dark green points).

Distance Method We could imagine that a non-detected species has the same probability, in logarithmic space, to have whatever value between Y_{min} up to Y_{upl} , where Y_{min} is an arbitrary column density and can be set to whatever value between zero and the minimal column density of the observed species.

With these assumptions, we can define a contribution to the fraction of observations due to a non-detection in each quadrant, defined by the datapoint (X^{obs} , Y^{obs}). Each Y_{upl} counts as a fraction of $1/n(\mathcal{O})$ in the quadrant A and B (or C and D) if $X_{upl} \leq X^{obs}$ (or $X_{upl} > X^{obs}$) that is normalized to $1/n(\mathcal{O})$. The contribution to the fraction of observations, f_{obs} , of each non-detection, Y_{upl} , depends on the distance of the non-detection in question to the origin of the quadrant, $Y_{upl} - Y^{obs}$. Fig. 9.13 shows that if $Y_{upl} > Y^{obs}$ the non-detection contributes to the fraction of observations of both the upper quadrant (B or C) and the lower quadrant (A or D), and this contribution is proportional to $Y_{upl} - Y^{obs}$ and $Y^{obs} - Y_{min}$, respectively. Normalizing, we obtain the contribution of each upper-limit in the case $Y_{upl} > Y^{obs}$

$$\frac{1}{n(\mathcal{O})} \left[\frac{Y_{\text{upl}} - Y^{\text{obs}}}{Y_{\text{upl}} - Y_{\text{min}}} \right] \qquad \text{for B and C quadrants,} \qquad (9.9)$$
$$\frac{1}{n(\mathcal{O})} \left[1 - \frac{Y_{\text{upl}} - Y^{\text{obs}}}{Y_{\text{upl}} - Y_{\text{min}}} \right] \qquad \text{for A and D quadrants,} \qquad (9.10)$$

while if $Y_{upl} < Y^{obs}$ the contribution of that upper-limit to the fraction of observations in a quadrant counts as $\frac{1}{n(0)}$ in the quadrants A or D, depending if $X_{upl} < X^{obs}$ or not.

Comparisons on CH⁺ In Fig. 9.14, we apply the "half" (left panel) and the "distance" (right panel) method to the 2D probability histogram of $N(CH^+)$ vs N_H for the fiducial sim-



Figure 9.12: A 2D distribution of observational datapoints composed of detections (red points) and upperlimits (green triangles). Each observational datapoint, (X^{obs} , Y^{obs}), indicated here with a dark red dot, divides the space into four quandrants A, B, C, D. Each detection contributes to the fraction of simulation, f_{obs} , for $1/n(\mathcal{O})$ in the quadrants where it belongs and 0 in the other three. Each upper limit counts as a detection if it belongs surely to a quadrant (light green points), i.e. the upper-limits in A and D. If the non-detection happens at larger column densities, i.e. $Y_{upl} > Y^{obs}$ (dark green points), the "real" detection corresponding to that upperlimit, Y_{upl} could be at $Y > Y^{obs}$ and belong to B (or C) or $Y < Y^{obs}$ and belong to A (or D). Therefore, those non-detections (dark green points), count as $0.5/n(\mathcal{O})$ in both A and B (or D and C).



Figure 9.13: A 2D distribution of observational datapoints composed of detections (red points) and upperlimits (green triangles). Each observational datapoint, (X^{obs} , Y^{obs}), indicated here with a dark red dot, divides the space into four quandrants A, B, C, D. Each detection contributes to the fraction of simulation, f_{obs} , for $1/n(\mathcal{O})$ in the quadrants where it belongs and 0 in the other three. Each non-detection counts as a fraction of $1/n(\mathcal{O})$ in the quadrants it crosses and zero in the others. Let's take the non-detection (X_{upl} , Y_{upl}) indicated with a dark green triangle in the plot, the corresponding detection can be anywhere on the green segment between Y_{upl} and Y_{min} , and the probability to be in C is proportional to the amount of the green segment in the C quandrant, i.e. it is proportional to ($Y_{upl} - Y^{obs}$), and the probability to be in D is proportional to ($Y^{obs} - Y_{min}$).



Figure 9.14: Result of the modified KS test including the non-detections with two different methods ("half" method on the left and the "distance" method on the right) applied to $N(CH^+)$ vs N_H for the fiducial simulation. Detections are indicated with squares and upperlimits with triangles. The red squares indicate the subsample \mathcal{O} scanned for the estimation of the merit function M, assuring a number $n_{thr} = 6$ of observations in each quadrant. The gray star and rectangle indicate the observational point and quadrant that maximize the merit function M.

ulation. We can notice that; in the case of the fiducial simulation, all three methods give similar results, in both the KS distance and the quadrant responsible for such a distance.

We presented in this section alternative methods to take into account the non-detections into the KS test for completeness. Nevertheless, these methods appear to have little to no impact on the results of the test performed. For this reason, we conclude that the best method is by far the first one because, oppositely to the last two methods, it does not rely on additional information provided by the user.

9.5.3 Conclusions

In this section, we presented a preliminary study to try and integrate the statistics of the non-detections. The inclusion of upper-limits in our modified version of the KS test can be capital, especially for certain chemical species that present two bumps in the probability histograms but not in the observational sample, probably due to observational limitations. Since taking into account the non-detections is still ongoing work, the results presented here should be explored and developed further.

Chapter 10

Atomic-to-Molecular Transition

The new and original analysis tools presented in the previous chapters have been applied to several sets of observations (see §4) in combination with a parametric study of the diffuse, neutral, and multiphasic ISM. The first application dedicated to the analysis of the HI-to-H₂ transition has been published in Bellomi et al. (2020). In a nutshell, the grid of simulations run with RAMSES (see §5) were used to compute the distributions of the total column densities, $N_{\rm H}$ and molecular hydrogen column densities, $N({\rm H}_2)$ taking into account the distribution of the observational sample (see §4.1) through the reconstruction method (see §8.1). The Kolmogorov-Smirnov difference (see §9.2) was then computed to facilitate the comparison of the simulations with the observations. The results are presented and discussed in the paper which is included below. In the next section, we simply present two additional results. The first one, discussed but not shown in the paper, focuses on the comparison of the predictions of simulations assuming two different prescriptions for the self-shielding of H₂. The second section shows the impact of including the upper-limits in the statistical analysis of the HI-to-H₂ transitions.

10.1 H₂ Self-shielding at High Temperature

The prescription of H_2 self-shielding used in the paper (Eq. 5.16) is taken from Draine & Bertoldi (1996). As discussed by Wolcott-Green et al. (2011), such a prescription is reliable for diffuse gas at low temperature but becomes less and less reliable for high-temperature environments (T > 500 K) where efficient collisional excitation of H_2 in its rovibrational levels reduces the self-shielding. To estimate the impact of this process, we ran the fiducial simulation with the alternative self-shielding function proposed by Wolcott-Green et al. (2011):

$$f_{\text{shield}}(x) = \frac{0.965}{(1+x/b_D)^{\alpha}} + \frac{0.035 \ \mathrm{e}^{-8.5 \times 10^{-4} \sqrt{1+x}}}{\sqrt{1+x}},\tag{10.1}$$

where $\alpha = 1.1$ and

$$x = N(H_2)/5 \times 10^{14} \text{ cm}^{-2}$$
(10.2)

and b_D is the Doppler broadening parameter expressed in km s⁻¹. In Fig. 10.1, we compare the 2D probability histograms of the HI-to-H₂ transition with the self-shielding computed



Figure 10.1: Comparison of the HI-to-H₂ transition obtained with the self-shieding function of Draine & Bertoldi (1996) (on the left) and of Wolcott-Green et al. (2011) (on the right).

with Draine & Bertoldi (1996) prescription (left panel) and Wolcott-Green et al. (2011) formula (on the right). Both prescriptions lead to similar probability histograms. Since the shielding function f_{shield} of Wolcott-Green et al. (2011) varies more smoothly as a function of the column density of molecular hydrogen, $N(H_2)$, compared to that of Draine & Bertoldi (1996) (see Fig. 5.4), the two regimes at high and low $N(H_2)$ are less populated, to the benefit of the transition region. Interestingly, we note that adopting Wolcott-Green et al. (2011) alternative prescription, not only induces more lines of sight at intermediate molecular fraction (region B, see Fig. 4.2), but also increases the width of the HI-to-H₂ transition. This also affects the dispersion at medium-high column densities (region C and D, see Fig. 4.2). Both these effects lead to a better agreement with the observations, as shown in Fig. 10.2. The KS distance appears to be lower than that computed to the self-shielding function of Draine & Bertoldi (1996).

10.1.1 Impact of the Inclusion of the Upperlimits in the KS Test

As already noticed in Fig. 9.6, the dispersion in the fraction of H_2 is underestimated in the simulations. This fact was detected by the KS test in the high- $N(H_2)$ bump and it is even more evident when upper limits are included in the KS test applied to the HI-to- H_2 transition. In Fig. 10.3, we show the modified KS test applied to the fiducial model for the transition of HI-to- H_2 in which we include the non-detections in the observational sample. The fraction of observations in the quadrant that maximizes the merit function (gray quadrant) is more than 60 times larger than the fraction of the simulated lines of sight in that same quadrant, indicating that the dispersion at low column densities seen in the observational sample is larger than the one simulated.

In the left panel of Fig. 10.4, we present the KS distances computed on the fiducial grid including only detections (dashed lines) or both detections and upper limits (solid lines). One of the simulations that shows the most difference of results between the two tests is the fiducial model, due to the underestimation of the number of simulated lines of sight at low column densities. With this new set up for the KS test, the best agreement between the observational dataset and the simulation is given by $G_0 = 1$ and $\overline{n_H} = 1$ cm⁻³ (right panel of Fig. 10.4), for which the dispersion in the low- $N(H_2)$ bump is high in the simulation.



Figure 10.2: Results of the modified KS test applied to the HI-to-H₂ transition for the fiducial simulation, run with Wolcott-Green et al. (2011) prescription for the computation of H₂ self-shielding (see eq. 10.1). The black squares are the observational data, the red square indicate the dataset used for the estimations of the merit function M, and the 2D histogram the simulated data. The gray star and rectangle indicate the observational point and the quadrant that maximize M (see §9.2). The fiducial simulation has a KS distance of 0.61. The corresponding quadrant contains 12.93% and 3.21% of the entire simulated and observed datasets.



Figure 10.3: Results of the modified KS test applied to the fiducial simulation taking into consideration the upperlimits. The black circles and the triangles indicate the detections and upperlimits, the red dots indicate the dataset used for the estimations of the merit function *M*, and the 2D histogram the simulated data. The gray star and rectangle indicate the observational point and the quadrant that maximize *M*.

We can see that detection and the quadrant that maximizes the merit function (gray) is at high column densities, indicating that this simulation underestimates the fraction of H_2 compared to observations.



Figure 10.4: *Left panel.* KS distance between the simulations and the observational sample as a function of the mean density $\overline{n_{\rm H}}$, the UV scaling factor $G_0 = 0.5$ (red), 1 (green), 2 (orange), and 4 (blue), and for both versions: the one that only consider detections (dashed line) and the one that includes upper-limits (solid). Points correspond to reliable measurements of the KS distances. Triangles indicate lower limits corresponding to simulations where the upper error bar on $R_{\rm KS}$ tends toward infinity. *Right panel.* Results of the modified KS test that consider also upper-limits applied to $G_0 = 1$ and $\overline{n_{\rm H}} = 1$ cm⁻³, the simulation giving the smallest KS difference, i.e. the best agreement with the observational dataset. The black points are the observational data, the red dots indicate the dataset used for the estimations of the merit function M, and the 2D histogram the simulated data. The gray star and rectangle indicate the observational point and the quadrant that maximize M.

3D chemical structure of diffuse turbulent ISM

I. Statistics of the HI-to-H₂ transition

E. Bellomi^{1,2}, B. Godard^{1,2}, P. Hennebelle³, V. Valdivia³, G. Pineau des Forêts^{4,1}, P. Lesaffre², and M. Pérault²

¹ Observatoire de Paris, PSL University, Sorbonne Université, LERMA, 75014 Paris, France e-mail: elena.bellomi@obspm.fr

- ² Laboratoire de Physique de l'Ecole normale supérieure, ENS, Université PSL, CNRS, Sorbonne Université, Université de Paris, 75005 Paris, France
- ³ Laboratoire AIM, CEA/IRFU, CNRS/INSU, Université Paris Diderot, CEA-Saclay, 91191 Gif-Sur-Yvette, France

⁴ Université Paris-Saclay, CNRS, Institut d'Astrophysique Spatiale, 91405 Orsay, France

Received 5 June 2020 / Accepted 8 August 2020

ABSTRACT

Context. The amount of data collected by spectrometers from radio to ultraviolet (UV) wavelengths opens a new era where the statistical and chemical information contained in the observations can be used concomitantly to investigate the thermodynamical state and the evolution of the interstellar medium (ISM).

Aims. In this paper, we study the statistical properties of the HI-to- H_2 transition observed in absorption in the local diffuse and multiphase ISM. Our goal is to identify the physical processes that control the probability of occurrence of any line of sight and the origins of the variations of the integrated molecular fraction from one line of sight to another.

Methods. The turbulent diffuse ISM is modeled using the RAMSES code, which includes detailed treatments of the magnetohydrodynamics, the thermal evolution of the gas, and the chemistry of H_2 . The impacts of the UV radiation field, the mean density, the turbulent forcing, the integral scale, the magnetic field, and the gravity on the molecular content of the gas are explored through a parametric study that covers a wide range of physical conditions. The statistics of the HI-to-H₂ transition are interpreted through analytical prescriptions and compared with the observations using a modified and robust version of the Kolmogorov-Smirnov test.

Results. The analysis of the observed background sources shows that the lengths of the lines of sight follow a flat distribution in logarithmic scale from ~100 pc to ~3 kpc. Without taking into account any variation of the parameters along a line of sight or from one line of sight to another, the results of one simulation, convolved with the distribution of distances of the observational sample, are able to simultaneously explain the position, the width, the dispersion, and most of the statistical properties of the HI-to-H₂ transition observed in the local ISM. The tightest agreement is obtained for a neutral diffuse gas modeled over ~200 pc, with a mean density $\overline{n_{\rm H}} = 1-2 \text{ cm}^{-3}$, illuminated by the standard interstellar UV radiation field, and stirred up by a large-scale compressive turbulent forcing. Within this configuration, the 2D probability histogram of the column densities of H and H₂, poetically called the kingfisher diagram, is remarkably stable and is almost unaltered by gravity, the strength of the turbulent forcing, the resolution of the simulation, or the strength of the magnetic field B_x , as long as $B_x < 4 \,\mu$ G. The weak effect of the resolution and our analytical prescription suggest that the column densities of HI are likely built up in large-scale warm neutral medium and cold neutral medium (CNM) structures correlated in density over ~20 pc and ~10 pc, respectively, while those of H₂ are built up in CNM structures between ~3 and ~10 pc. *Conclusions*. Combining the chemical and statistical information contained in the observations of HI and H₂ sheds new light on the study of the diffuse matter. Applying this new tool to several atomic and molecular species is a promising perspective to understanding the effects of turbulence, magnetic field, thermal instability, and gravity on the formation and evolution of molecular clouds.

Key words. ISM: structure – ISM: molecules – ISM: kinematics and dynamics – ISM: clouds – methods: numerical – methods: statistical

1. Introduction

The multiphase nature of the interstellar medium (ISM) is at the root of the regulation of star formation in galaxies (e.g., Hill et al. 2018). As shown by the emission profiles of the HI 21 cm line (Heiles & Troland 2003a,b; Murray et al. 2015, 2018), the diffuse neutral ISM is composed of two stable thermal states at thermal pressure equilibrium (Jenkins & Tripp 2011), the warm neutral medium (WNM, $T \sim 7000$ K) and the cold neutral medium (CNM, $T \sim 70$ K), coexisting with a third unstable state, the lukewarm neutral medium (LNM), whose temperature is comprised between those of the CNM and the WNM (e.g., Marchal et al. 2019). Through condensation and evaporation processes, turbulent transport, and turbulent mixing, the diffuse

matter flows from one stable state to the other eventually leading to the formation of dense and cold clouds massive enough to trigger gravitational collapse (e.g., Ostriker et al. 2010). While this picture is widely accepted, the intricated effects of turbulence, gravity, radiation field, and magnetic field on the exchange of mass and energy between the different phases and on the formation of structures at all scales has yet to be unveiled.

Following the illustrious analytical descriptions of the thermal instability process (Field 1965; Wolfire et al. 1995, 2003; Bialy & Sternberg 2019), several analytical and numerical studies have been dedicated to understand the dynamical evolution of the gas, focusing on the formation of CNM structures, molecular clouds, and collapsing cores (e.g., Hennebelle & Pérault 1999; Koyama & Inutsuka 2002a,b; Audit & Hennebelle 2005; Vázquez-Semadeni et al. 2007; Hennebelle et al. 2008), as well as on the stability of clouds of various geometries under evaporation and condensation conditions (e.g., Inoue et al. 2006; Stone & Zweibel 2009; Kim & Kim 2013; Nagashima et al. 2005; Iwasaki & Inutsuka 2014). These show that large-scale turbulence combined with thermal instability is sufficient to explain several features of the neutral ISM, including the fractions of mass observed in the different thermal states (Seifried et al. 2011; Hennebelle & Iffrig 2014; Hill et al. 2018), the distribution of thermal pressure (Saury et al. 2014), and the mass spectrum, the mass-size relation, and the velocity dispersion-size relation of molecular clouds (e.g., Audit & Hennebelle 2010; Padoan et al. 2016; Iffrig & Hennebelle 2017).

To extend the predictions of simulations to a larger set of observational diagnostics, recent numerical studies have undertaken the challenging task of solving the chemical evolution of turbulent and/or multiphase environments. Originally dedicated to the formation of CO in molecular clouds and to the analysis of the CO-to-H₂ conversion factor in galaxies and the CO dark-gas (e.g., Glover et al. 2010; Smith et al. 2014; Richings & Schaye 2016a; Seifried et al. 2017; Gong et al. 2018), numerical simulations are now used to study a variety of atomic and molecular tracers, including CII, CI, CH⁺, OH⁺, H₂O⁺, and ArH⁺ (e.g., Richings & Schaye 2016b; Valdivia et al. 2017; Clark et al. 2019; Bialy et al. 2019). All these works demonstrate the predictive power of astrochemistry. The column density distribution of each atom and molecule has a unique signature that provides detailed information on the thermodynamical state of the diffuse matter (Clark et al. 2019). In turn, the confrontation with the predictions of numerical simulations can be used to estimate the scale and strength of the injection of mechanical energy by stellar feedback (Bialy et al. 2019), the large-scale turbulent transport and the interfaces between CNM and WNM (Valdivia et al. 2017), and the nature of the turbulent dissipation processes (Lesaffre et al. 2020).

In this context, understanding the formation and survival of molecular hydrogen has long been recognized as a major topic of investigation. As the most abundant molecule in space, H₂ is at the root of interstellar chemistry and the growth of molecular complexity. In addition, and because its formation preferentially occurs in dense environments, H₂ naturally correlates with the star formation rate of galaxies (e.g., Lupi et al. 2017) and therefore offers a valuable proxy to understand the limit in the Kennicutt-Schmidt relation above which star formation occurs (e.g., Bigiel et al. 2008, 2011; Schruba et al. 2011; Leroy et al. 2013).

Over the last decades, great efforts have thus been devoted to propose analytical descriptions of the HI-to-H₂ transition in homogeneous clouds with plane-parallel or spherical geometries (e.g., Sternberg 1988; Krumholz et al. 2008; McKee & Krumholz 2010; Sternberg et al. 2014 and references therein), compute this transition in detailed 1D chemical models assuming chemical equilibrium (e.g., van Dishoeck & Black 1986; Abgrall et al. 1992; Le Bourlot et al. 2012; Bron et al. 2014) or not (e.g., Lee et al. 1996; Goldsmith et al. 2007; Lesaffre et al. 2007), treat the chemistry of H and H₂ in subgrid models applied to simulations of galaxy formations (e.g., Gnedin et al. 2009; Christensen et al. 2012; Thompson et al. 2014; Diemer et al. 2018), or solve it in 3D isothermal or multiphase simulations of the diffuse ISM using various treatments of the radiative transfer (e.g., Glover et al. 2010; Valdivia et al. 2016; Hu et al. 2016; Bialy et al. 2017; Nickerson et al. 2018).

Thanks to all these works, a global picture of the formation of H_2 in galaxies is now emerging. At the scale of a homogeneous

cloud, the molecular content, the sharpness of the HI-to-H₂ transition, and the asymptotic column density of HI are controlled by the ratio of the intensity of the ultraviolet (UV) field to the gas density and the dust-to-gas ratio, or equivalently, the metallicity (Sternberg et al. 2014). At larger scales, the integrated column densities of HI and H₂ also depend on the distribution of clouds of various densities along the line of sight and on the porosity to the UV radiation field. Because of these effects, the statistical properties of the total column density are found to depend on the strength, the scale, and the compressibility of the turbulent forcing in simulations of CNM gas (Micic et al. 2012; Bialy et al. 2017). The amount of molecular gas depends on the "clumpiness factor" used for the subgrid models in simulations of galaxy formation (Gnedin et al. 2009; Christensen et al. 2012).

Despite these achievements, very few works have been dedicated so far to the analysis of the HI-to-H₂ transition in a turbulent multiphase medium at a scale sufficient to resolve the formation of CNM structures. In addition and while the predictions of analytical models (e.g., Krumholz et al. 2008) and simulations (Gnedin et al. 2009; Valdivia et al. 2016) were able to reproduce the trend of the HI-to-H₂ transition observed by Copernicus and FUSE in the local ISM (e.g., Savage et al. 1977; Gillmon et al. 2006; Rachford et al. 2009), the LMC and the SMC (e.g., Browning et al. 2003; Gillmon et al. 2006; Leroy et al. 2007), no detailed comparison with the statistical properties of these observations have been proposed. As a result, the occurrence of lines of sight with large molecular fractions predicted by numerical simulations often exceed what is deduced from the observations (Valdivia et al. 2016). Finally, and while statistical studies of 1D probability distribution functions (PDF) have become a common tool to understand the formation and the dynamics of molecular clouds (Körtgen et al. 2019), few statistical studies have been performed to date on 2D probability distribution functions using combined observations of different molecular tracers. In that perspective, the recent work of Bialy et al. (2019) opens new horizons for the analysis of chemistry in the diffuse matter.

In the first paper of this series, we extend these pioneer statistical studies to the measurements of the atomic-to-molecular transition observed in the diffuse and translucent ISM located in a radius of \sim 3 kpc around the sun. We perform a parametric exploration of numerical simulations of the multiphase ISM and compare the results with the observed 2D probability histogram (PH) of total and molecular hydrogen column densities in order to identify the physical processes that control the molecular content of CNM clouds and the probability of occurrence of lines of sight. The observational dataset and the distribution of sizes of the sampled medium are presented in Sect. 2. The different setups of the simulations and the method used to reconstruct the 2D PH are described in Sect. 3. The comparisons with the observations are shown in Sect. 4 which also highlights the influences of the different parameters. The paper finally ends with Sects. 5 and 6 where we discuss the validity of our approach and summarize our main conclusions.

2. Observations of the HI-to-H₂ transition

2.1. Observational sample and distances

The observational sample studied in this work is built from the database of Gudennavar et al. (2012) who compiled existing data of atomic and molecular lines observed in absorption toward several thousand sources, including stars and AGNs. Limiting this catalog to observations or tentative detections of HI, H_2 ,

E. Bellomi et al.: 3D chemical structure of diffuse turbulent ISM. I.



Fig. 1. Aitoff projection, in Galactic longitude and latitude coordinates, of the background sources of the observational sample of HI and H_2 deduced from absorption studies and used in this work (see Appendix A and Table A.1). The color code indicates the distance of the source. All unknown distances correspond to extragalactic sources (see Table A.1): these are arbitrarily set to 1 Mpc and indicated with black points.

and of the reddening E(B-V), and removing the data associated to the Magellanic Cloud or high redshift extragalactic environments (e.g., Tumlinson et al. 2002; Cartledge et al. 2005; Welty & Crowther 2010; Noterdaeme et al. 2007), we obtain a sample of 360 sources which form, to date, the most complete set of observations of the HI-to-H₂ transition in the local diffuse ISM. A more detailed description of this set, the list of the background sources, and the values of the column densities of HI and H₂ toward each source, N(H) and $N(H_2)$ are given in Appendix A and Table A.1.

The positions of the sources in the sky and their distance deduced from Gaia and HIPPARCOS measurements of parallaxes (Perryman et al. 1997; Gaia Collaboration 2018) are shown in Fig. 1, where unknown distances of extragalactic sources (see Table A.1) are arbitrarily set to 1 Mpc. With comparable numbers of observations in all Galactic quadrants, the sources appear to be well distributed in Galactic longitudes. Oppositely, and while the sources cover almost all Galactic latitudes, about twothirds of them are located toward the Galactic disk with latitudes smaller than 15°, and only one third is located above, crossing the Galactic halo. Since the sample contains extragalactic sources, and since the amount of molecular gas in the Milky Way decreases exponentially as a function of the distance from the midplane, the length of the line of sight $l_{\rm los}$ occupied by the observed diffuse gas cannot always be identified to the distance of the background source. For the sake of simplicity, we assume here a molecular height above the midplane of 100 pc and compute the length of the observed diffuse Galactic material as

$$l_{\rm los} = \min\left(\frac{1''}{p}, \frac{100}{\sin(|b|)}\right) \,\mathrm{pc},\tag{1}$$

where b is the Galactic latitude of the background source and p is its parallax.

The resulting distribution of the lengths of the lines of sight is shown in Fig. 2. The shortest lines of sight are found to extend over ~100 pc and the largest over ~3.5 kpc. Remarkably, and because of the combined distributions of distances and Galactic latitudes of the sources, we find that $\log(l_{\rm los})$ follows a flat distribution up to $l_{\rm los} \sim 2$ kpc with about 50 sources per bin and



Fig. 2. Distribution of lengths l_{los} of the intercepted diffuse material computed with Eq. (1) along all lines of sight of the observational sample. The orange sample corresponds to lines of sight where H₂ is detected and the green sample to those for which an upper limit on $N(H_2)$ has been derived (see Table A.1).

drops by about a factor of two for $l_{\rm los} \sim 3$ kpc. Oppositely, and as expected, the distribution of lengths of non-detections of H₂ is not flat but decreases rapidly up to 1 kpc. Long lines of sight are finally not limited to the first and fourth Galactic quadrants but are found to spread over all Galactic longitudes and mostly depend on the Galactic latitude of the background source.

2.2. Physical and statistical properties

The compiled data are shown in Fig. 3 which displays the observed column densities of H₂ as functions of the total proton column densities of the gas $N_{\rm H} = N({\rm H}) + 2N({\rm H}_2)$. As shown in Fig. 3 and as already noted by Goldsmith et al. (2009), almost no line of sight is either purely WNM or purely molecular. This implies that the observed gas is necessarily composed of a combination of phases and clouds of different extinctions with various contributions to the volume spanned by the different lines of sight. As a result, the integrated molecular fraction computed as

$$f_{\rm H_2} = \frac{2N({\rm H_2})}{N_{\rm H}}$$
(2)

shows a large dispersion in the observational sample covering about seven orders of magnitude. While the molecular fraction averaged over all the lines of sight is found to be 0.20, the mass averaged molecular fraction is 0.27, a value similar to the results obtained by Miville-Deschênes et al. (2017) at 8.5 kpc based on the analysis of molecular clouds observed in CO in the entire Galactic disk. The position of the HI-to-H₂ transition, on the other hand, is found to extend over about one order of magnitude of total column density from $N_{\rm H} \sim 10^{20}$ to 10^{21} cm⁻² and occurs, on average, at $N_{\rm H} \sim 3 \times 10^{20}$ cm⁻² (Gillmon et al. 2006). In order to highlight the statistical features of this transi-

In order to highlight the statistical features of this transition, we divide the observational sample into 5 subsamples A, B, C, D, and E, shown in Fig. 3, which encompass almost all the observational points and whose statistical properties are summarized in Table 1. 48 lines of sight out of 360 are found to be not detected in H₂. Most of these upper limits are obtained for a total column density smaller than 10^{21} cm⁻² and about half of them provide strong constraints on the molecular fraction with $f_{H_2} \leq 10^{-5}$. 3% of the 312 detections belong to the subsample A, 13% to subsample B, 16% to subsample C, and



Fig. 3. H₂ column density as a function of the total column density of protons $N_{\rm H}$. Open circles correspond to detections of H₂ while arrows correspond to upper limits (see Table A.1). The blue dashed line indicates the maximum value of N(H₂) derived from a purely molecular medium with an integrated molecular fraction $f_{\rm H_2} = 1$ (Eq. (2)). The red dashed-dotted line indicates the theoretical molecular fraction derived in an unshielded WNM-type environment with a density of 0.5 cm⁻³ and a temperature of 8000 K, illuminated by a UV photon flux of $10^8 \text{ cm}^{-2} \text{ s}^{-1}$ (see Eqs. (13) and (15)). The regions A, B, C, D, and E defined in Table 1 correspond to an arbitrary separation of the observational sample used for quantitative comparisons with the results of simulations (see Sect. 4).

Table 1. Statistical properties of H_2 observations in the subsamples A, B, C, D, and E defined in footnote and shown in Fig. 3.

Region	Number	%	$\mu \left[\log(f_{\rm H_2}) \right]$	$\sigma [\log(f_{\rm H_2})]$
А	10	3	-5.02	0.99
В	41	13	-4.62	0.70
С	50	16	-1.10	0.49
D	204	65	-0.70	0.37
E	0	0	_	_

Notes. Only the lines of sight where H₂ has been detected (312 sources out of 360, see Table A.1) are considered. The mean and dispersion values, μ and σ are computed on the logarithm of the molecular fraction f_{H_2} observed in the corresponding subsample. Definitions of subsamples: region A: $5 \times 10^{18} \leq N_{\text{H}} \leq 5 \times 10^{19}$, $5.6 \times 10^{-8} \leq f_{\text{H}_2} \leq 10^{-3}$; region B: $5 \times 10^{19} \leq N_{\text{H}} \leq 7 \times 10^{20}$, $5.6 \times 10^{-8} \leq f_{\text{H}_2} \leq 10^{-3}$; region C: $1.6 \times 10^{20} \leq N_{\text{H}} \leq 10^{21}$, $10^{-3} \left(\frac{N_{\text{H}}}{10^{20}}\right)^{0.9} \leq f_{\text{H}_2} \leq 1$; region D: $10^{21} \leq N_{\text{H}} \leq 10^{22}$, $10^{-3} \left(\frac{N_{\text{H}}}{10^{20}}\right)^{0.9} \leq f_{\text{H}_2} \leq 1$; region D: $10^{-3} \left(\frac{N_{\text{H}}}{10^{20}}\right)^{0.9} \leq f_{\text{H}_2} \leq 1$.

65% to subsample D. Interestingly, the subsample E is empty and no line of sight is observed with $N_{\rm H} > 10^{22}$ cm⁻². While the mean value of the logarithm of the molecular fraction strongly increases from subsamples A to D, the dispersion simultaneously decreases by about a factor of three, probably revealing an effect of average over long distances.

All these statistical properties, and more precisely the probability of occurrence of a given line of sight, are the subject of this paper. What physical processes control the HI-to-H₂ transition? How does the distribution of lengths of the lines of sight influence its observed statistical properties? What are the origins of the variations of the molecular fraction from one line of sight to another?

3. Physics and numerical method

To study the physical processes at play in the HI-to- H_2 transition, we performed numerical simulations of the multiphase diffuse ISM, using the RAMSES code (Teyssier 2002; Fromang et al. 2006), a grid-based solver with adaptative mesh refinement (Berger & Oliger 1984). The methodology applied in this paper follows the works of Seifried et al. (2011), Saury et al. (2014), and Valdivia et al. (2015, 2016).

The diffuse matter in the Solar Neighborhood of our galaxy is simulated over a box of size L with periodic boundary conditions. The matter, defined by a mean proton density $\overline{n_{\rm H}}$, is assumed to be illuminated on all sides by an isotropic spectrum of UV photons set to the standard interstellar radiation field (Habing 1968) and scaled with a factor G_0 .

3.1. Fluid equations

Within this framework, RAMSES computes the evolution of the gas solving the classic equations of ideal magnetohydrodynamics (MHD):

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \boldsymbol{v}) = 0, \tag{3}$$

$$\frac{\partial \rho \boldsymbol{v}}{\partial t} + \nabla \cdot (\rho \boldsymbol{v} \boldsymbol{v} - \boldsymbol{B} \boldsymbol{B}) + \nabla P = -\rho \nabla \Phi + \rho \boldsymbol{f}, \tag{4}$$

$$\frac{\partial E}{\partial t} + \nabla \cdot \left[(E + P)\boldsymbol{v} - \boldsymbol{B}(\boldsymbol{B}\boldsymbol{v}) \right] = -\rho \boldsymbol{v} \cdot \nabla \Phi + \rho \boldsymbol{f} \cdot \boldsymbol{v} - \rho \boldsymbol{\mathcal{L}}, \text{ and}$$
(5)

$$\frac{\partial \boldsymbol{B}}{\partial t} + \nabla \cdot (\boldsymbol{v}\boldsymbol{B} - \boldsymbol{B}\boldsymbol{v}) = 0, \tag{6}$$

where ρ , v, B, P and E are the mass density, the velocity field, the magnetic field, the total pressure, and the total energy density, respectively. The net cooling function per unit mass, \mathcal{L} , and the acceleration due to the turbulent driving, f, are described in Sects. 3.2 and 3.3. The axis x, y, and z are chosen so that z corresponds to the direction perpendicular to the Galactic disk, and x corresponds to the direction of the mean magnetic field initially parametrized by a constant value B_x .

To take into account all gravitational forces, including selfgravity and the action of stars and dark matter, the gravitational potential Φ is divided into two terms:

$$\Phi = \phi_{\rm gas} + \phi_{\rm gal}.\tag{7}$$

The self-gravity potential, ϕ_{gas} , is deduced from the Poisson's equation:

$$\nabla^2 \phi = 4\pi G\rho. \tag{8}$$

Following Kuijken & Gilmore (1989) and Joung & Mac Low (2006), we assume that the Galactic potential along the direction z perpendicular to the Galactic disk can be written as

$$\phi_{\text{gal}}(z) = a_1 \left(\sqrt{z^2 + z_0^2} - z_0 \right) + 2a_2 z^2, \tag{9}$$

where the first term is the contribution of the stellar disk parametrized by $z_0 = 0.18$ kpc and $a_1 = 1.42 \times 10^{-3}$ kpc Myr⁻² and the second term is the contribution of the spherical dark halo parametrized by $a_2 = 5.49 \times 10^{-4}$ Myr⁻².

3.2. Thermal processes and radiative transfer

As shown by Field (1965), the multiphase nature of the ISM results from the thermal balance of the gas and thus from its net cooling function \mathcal{L} defined by

$$\rho \mathcal{L} = n_{\rm H}^2 \Lambda - n_{\rm H} \Gamma, \tag{10}$$

where $n_{\rm H}^2 \Lambda$ and $n_{\rm H} \Gamma$ are the cooling and heating rates of the medium (in erg cm⁻³ s⁻¹) and $n_{\rm H}$ is the proton density. To correctly describe the thermal state of the diffuse ISM, we include in this work the heating induced by the photoelectric effect and the decay of cosmic ray particles and the cooling induced by Lyman α photons, the recombination of electrons onto grains, and the fine structure lines of OI and CII. All these processes, described in Appendix B, are modeled with the analytical formulae given by Wolfire et al. (2003).

The absorption of UV photons by dust, and its subsequent impact on the photoelectric effect, is treated with the treebased method proposed by Valdivia & Hennebelle (2014). At each point the effective radiation field G_{eff} (in Habing units) is computed as

$$G_{\rm eff} = G_0 \langle e^{-2.5A_V} \rangle, \tag{11}$$

where A_V is the visual extinction along a given ray, deduced from the integrated proton column density¹ computed from the border of the box to the current point

$$A_V = 5.34 \times 10^{-22} \left(\frac{N_{\rm H}}{\rm cm^{-2}}\right),$$
 (12)

and $\langle e^{-2.5A_V} \rangle$ is an average performed over 12 directions, treated as solid angles evenly spread in polar coordinates.

3.3. Turbulence forcing

To mimic the injection of mechanical energy in the diffuse ISM, a large scale turbulent forcing is applied. Following Schmidt et al. (2009) and Federrath et al. (2010), this forcing, modeled by an acceleration f in the momentum conservation equation, is driven through an Ornstein-Uhlenbeck process using a pseudo-spectral method. At regular time intervals $\Delta \tau$, random fluctuations of the forcing term are generated and applied over an autocorrelation timescale τ . To excite only large scale modes, the forcing is modeled as a paraboloid in Fourier space covering a small interval of wavenumbers $1 \le k \le 3$ and centered on k = 2. Using the notations of Seifried et al. (2011) and Saury et al. (2014), the total magnitude of these perturbations is set with either an acceleration parameter F or, equivalently, a velocity parameter V related by $\bar{F} = V^2 / L_{\text{drive}}$, where L_{drive} is the main driving scale, $L_{\text{drive}} = L/2$. A Helmholtz decomposition is finally applied, in order to control the powers injected in compressive and solenoidal modes. Using the classical notation, these powers are set with a parameter² ζ ranging from a pure solenoidal field $(\zeta = 1)$ to a pure compressive field $(\zeta = 0)$.

Throughout this work, we adopt $\Delta \tau \sim 0.4$ Myr which roughly corresponds to the time interval separating two supernova events occurring in a volume of $(200 \text{ pc})^3$. The characteristic damping time of the turbulence τ is approximately set to the turnover

timescale of the diffuse ISM $\tau \sim 33(L/200 \text{pc})^{0.6}$ Myr (Larson 1981; Hennebelle & Falgarone 2012). *F* (or *V*) and ζ are left as free parameters.

3.4. H₂ chemistry

The timescale required for the abundance of molecular hydrogen to reach its equilibrium value is known to range over several orders of magnitude, depending on the physical conditions of the ISM (e.g., Goldsmith et al. 2007; Tabone 2018). In the diffuse gas, this timescale varies typically between a few 10^3 yr and a few 10^7 yr (Valdivia et al. 2016), hence over a range of values that often exceeds the dynamical timescales. To take into account this important aspect of the diffuse interstellar chemistry, the outof-equilibrium abundance of H₂ is computed self-consistently in the simulation, using the formalism introduced in RAMSES by Valdivia et al. (2015, 2016).

The formation of H_2 onto grains in physisorption and chemisorption sites is modeled with the simplified rate of Le Bourlot et al. (2012)

$$k_f = 3 \times 10^{-17} n_{\rm H} n({\rm H}) \sqrt{\frac{T}{100 \,{\rm K}}} S(T) \,{\rm cm}^{-3} {\rm s}^{-1},$$
 (13)

where $n_{\rm H}$ and $n({\rm H})$ are the local proton and atomic hydrogen densities, and

$$S(T) = \frac{1}{1 + \left(\frac{T}{T_2}\right)^{\beta}} \tag{14}$$

is the sticking coefficient of H onto grain, parametrized by $T_2 = 464$ K and $\beta = 1.5$.

The destruction of H_2 by UV photons is computed using the formalism, described by Draine & Bertoldi (1996) and Sternberg et al. (2014), which is classically introduced in many astrochemical models (e.g., Lesaffre et al. 2013; Gong et al. 2017; Bialy et al. 2017). In each cell, the photodestruction rate of H_2 is modeled as

$$k_d = k_d^0 n(\mathrm{H}_2) G_0 \langle e^{-\sigma_d N_\mathrm{H}} \rangle \langle f_{\mathrm{shield}}(x) \rangle \mathrm{cm}^{-3} \mathrm{s}^{-1}, \qquad (15)$$

where $x = N(H_2)/5 \times 10^{14} \text{ cm}^{-2}$, $k_{d,0} = 3.3 \times 10^{-11} \text{ s}^{-1}$ is the inverse freespace dissociation timescale of H₂ in an isotropic Habing field, $n(H_2)$ is the local density of the molecular hydrogen, $e^{-\sigma_d N_{\text{H}}}$ is the shielding induced by dust and f_{shield} the selfshielding function. We adopt here an effective dust attenuation cross section at $\lambda = 1000 \text{ Å}$, $\sigma_d = 2 \times 10^{-21} \text{ cm}^2$ (Sternberg et al. 2014). Following Draine & Bertoldi (1996), the self-shielding function is computed as

$$f_{\text{shield}}(x) = \frac{0.965}{(1+x/b_D)^2} + \frac{0.035 \,\mathrm{e}^{-8.5 \times 10^{-4} \,\sqrt{1+x}}}{\sqrt{1+x}},\tag{16}$$

where b_D is the Doppler broadening parameter expressed in km s⁻¹. As done for the photoelectric heating rate (see Sect. 3.2), both the shielding by dust and the self-shielding are calculated along 12 different directions and then averaged to obtain the photodissociation rate of Eq. (15).

3.5. Fiducial model and grids of parameters

The framework described above lays on several independent parameters, which are all related to key physical ingredients of the ISM. The influence of each ingredient on the HI-to- H_2

¹ In this work we use the relation between $N_{\rm H}$ and A_V deduced from the observations of the mean Galactic extinction curve (Fitzpatrick & Massa 1986).

² We note that for $\zeta = 0.5$, the power of the compressive forcing corresponds to 1/3 of the total power.

Parameter	Notation	Ref	Range	Units
Box size	L	200	20-200	рс
Mean density	$\overline{n_H}$	2	0.5–4	cm ⁻³
UV radiation field	G_0	1	0.5–4	Habing field
Resolution	R	256^{3}	$64^3 - 512^3$	_
Turbulent forcing	F	9×10^{-4}	$10^{-5} - 10^{-2}$	kpc Myr ⁻²
Compressive ratio	ζ	0.1	0.1-0.9	-
Doppler broadening	$\dot{b_D}$	8	1-8	$\rm km~s^{-1}$
Initial magnetic field	B_x	3.8	0–40	μG
Self-gravity	_	on	on–off	-
Galactic well	-	on	on-off	_

Table 2. Fiducial model and range of parameters explored in this work.

transition is studied here through several grids of simulations – including a total of 305 runs – covering a broad range of physical conditions and centered around a fiducial setup³. The reference value adopted for each parameter, and the range of values explored in this work, are summarized in Table 2. Among all parameters, L, $\overline{n_{\rm H}}$ and G_0 are of particular importance.

With our assumptions, L simultaneously corresponds to the scale of illumination of the gas by UV photons and twice the integral scale of turbulence, that is twice the scale of injection of mechanical energy L_{drive} . OB stars, which are the dominant sources of the interstellar UV field, are not uniformly distributed in the sky but are known to be clustered in associations (Ambartsumian 1947). As shown by the recent 3D studies of the distributions of stars based on the HIPPARCOS and Gaia Catalogs (e.g., Bouy & Alves 2015, Zari et al. 2018), the typical distances separating two associations in the Solar Neighborhood range between 50 pc and a few hundreds of pc, that is several times the mean distance deduced from the integrated surface densities of OB stars (~ 1.6×10^{-3} pc⁻², Maíz-Apellániz 2001). Interestingly, such distances are not only comparable to the heights of the molecular gas (\sim 75 pc) and the cold HI gas (~150 pc) above the Galactic plane deduced from CO and HI all-sky surveys (e.g., Dame et al. 2001, Dickey & Lockman 1990, Kalberla & Kerp 2009), but they also correspond to the typical size of HI superclouds (Elmegreen & Elmegreen 1987). For all these reasons, we, therefore, adopt a fiducial simulation with L =200 pc and explore values down to a few tens of pc.

The mean density of the gas, $\overline{n_H}$, represents the mass of the diffuse neutral ISM contained in a volume L^3 , and also controls the porosity of the matter to the impinging radiation field. In this work, we adopt a fiducial value $\overline{n_H} = 2 \text{ cm}^{-3}$, a value slightly larger than the standard Galactic midplane density of HI at a galactocentric distance of 8.5 kpc (Kalberla & Kerp 2009). With our fiducial value of *L*, the mass of gas contained in the box accounts for all the mass surface density of HI in the Solar Neighborhood ($\Sigma_{\text{HI}} \sim 10 M_{\odot} \text{ pc}^{-2}$, Nakanishi & Sofue 2016, Miville-Deschênes et al. 2017).

 G_0 controls the intensity of the radiation field, hence the thermodynamical state of the gas. Since G_0 is normalized to the Habing field, we choose a fiducial value of $G_0 = 1$. The corresponding UV energy density of the standard setup is therefore slightly smaller than that contained in the standard UV radiation fields given by Draine (1978) and Mathis et al. (1983). The standard values of the two parameters F and B_x are set to 9×10^{-4} kpc Myr⁻² and $3.8 \,\mu$ G, respectively. As we will show later, those values are chosen so that the velocity dispersion of the gas and the strength of the magnetic field are close to the values observed in the diffuse ISM.

3.6. Steady-state

The evolution of the multiphase environments simulated here is identical to the description already given in many papers (e.g., Seifried et al. 2011; Saury et al. 2014; Valdivia et al. 2015, 2016). Starting from an homogeneous density $n_{\rm H} = \overline{n_{\rm H}}$, the gas evolves under the joint actions of turbulence, gravity, and thermal instability, and splits up in three different phases at thermal pressure equilibrium, the WNM, the CNM, and the LNM. The formation of dense environments well shielded from the destructive UV radiation field triggers the formation of H₂ and the medium jointly evolves from a purely atomic state to a partly molecular state. If the mass of the gas is conserved, as imposed by the periodic boundary conditions, the medium progressively tends toward a steady-state where the mean pressure, the volume filling factors of the different phases, their velocity dispersion, and their mean molecular fractions are roughly constant. This steady-state is typically reached after a few turnover timescales, providing that the corresponding time is longer than the damping time (see Sect. 3.3).

Because of turbulence and thermal instability, the steadystate has a statistical nature. The turbulent forcing and the subsequent turbulent cascade induce pressure variations and shear motions at all scales which trigger mass exchanges between the different phases. Any pressure or density structure is therefore a transient system which is usually described by its contribution to probability distribution functions. At steady-state, only PDFs remain constant. This steady yet ever changing environment is the reason for the sustained presence of a substantial amount of gas in the LNM at densities and temperatures out of thermal equilibrium (e.g., Marchal et al. 2019). All the results shown throughout this paper are taken at steady-state, at times ranging from a few tens of Myr up to 100 Myr depending on the strength of the turbulent forcing.

3.7. Properties of the multiphase medium

The steady-state values of the mean pressure $\langle P/k \rangle$, the turbulent velocity dispersion σ_{tur} , and the fractions of mass of the WNM and the CNM, f_{WNM} and f_{CNM} , are shown in Fig. 4 for a set of 60 different simulations. For the sake of simplicity, we assume

 $^{^3}$ The grids have been run on the computing cluster Totoro funded by the ERC Advanced Grant MIST. The computational time of the standard simulation is ~6 days using 40 processors.


E. Bellomi et al.: 3D chemical structure of diffuse turbulent ISM. I.

Fig. 4. Colored tables of the mean pressure expressed in K cm⁻³ (*first line*), the turbulent velocity dispersion σ_{tur} (*second line*), and the fractions of mass f_{WNM} and f_{CNM} contained in the WNM phase (*third line*) and the CNM phase (*fourth line*). *First and second columns*: these quantities as functions of $\overline{n_H}$ and G_0 , for L = 50 pc and $F = 3.6 \times 10^{-3}$ kpc Myr⁻² (*first column*) and for L = 200 pc and $F = 9 \times 10^{-4}$ kpc Myr⁻² (*second column*). *Third and fourth columns*: these quantities as functions of the acceleration parameter F and the compressive ratio ζ , for L = 50 pc (*third column*). All other parameters are set to their standard values (see Table 2).

that the WNM is composed of all cells with a temperature $T \ge 3000$ K, the CNM of all cells with $T \le 300$ K, and the LNM of all cells with 300 K < T < 3000 K, hence

$$f_{\rm WNM} = \frac{\sum \rho_{|T \ge 3000\,\rm K}}{\sum \rho} \tag{17}$$

and

$$f_{\rm CNM} = \frac{\sum \rho_{|T \le 300 \,\rm K}}{\sum \rho}.$$
(18)

The mean pressure $\langle P/k \rangle$ (with k the Boltzmann constant) is classically computed as an average over the entire volume. While the above conventions are well established, there are many ways to define the turbulent velocity dispersion σ_{tur} . It could be defined as the volume-weighted or mass-weighted velocity dispersion computed over the entire volume (Audit & Hennebelle 2005; Federrath et al. 2010), the average of the mass-weighted velocity dispersion computed along individual lines of sight (Miville-Deschênes & Martin 2007; Saury et al. 2014), or the dispersion of the mass-weighted velocity centroids computed along independent directions (Henshaw et al. 2019). All these definitions give velocity dispersions that differ from one another and are not equally relevant for the comparison with observed quantities. To relate the velocity dispersion to a kinetic energy and provide values that could be compared to the observations of broad HI emission profiles at high Galactic latitude (see below), we chose to compute σ_{tur} in the WNM only as

$$\sigma_{\rm tur}^2 = \frac{1}{3} \frac{\sum \rho ||\boldsymbol{v} - \bar{\boldsymbol{v}}||^2}{\sum \rho}$$
(19)

with

$$\overline{\boldsymbol{v}} = \frac{\sum \rho \boldsymbol{v}}{\sum \rho}$$
(20)

and where the sums are performed over all cells with $T \ge 3000$ K.

The results displayed in Fig. 4 are very similar to those obtained in the parametric studies of Seifried et al. (2011) and Saury et al. (2014) and are in line with the expectations of models of turbulent multiphase environments (Wolfire et al. 2003; Ostriker et al. 2010). The mean pressure of the gas is primarily regulated by the thermal equilibrium curve (see Appendix B) which depends on the local illumination of the gas by the UV radiation field G_{eff} . Since G_{eff} results from the absorption of the external UV field by the surrounding environments, $\langle P/k \rangle$ is not only sensitive to G_0 but also to the total mass of the simulation set by $\overline{n_{\text{H}}}$ and *L*. Larger values of G_0 or smaller values of *L* or $\overline{n_{\text{H}}}$ leads to larger $\langle P/k \rangle$. In turn, the fractions of mass contained in the WNM and the CNM are controlled by the mean pressure and the total mass of the gas. Larger pressure implies larger densities of the WNM (and the CNM). The fraction of mass of the

CNM therefore decreases as $\langle P/k \rangle$ increases; eventually, if the density of the WNM becomes comparable to the mean density $\overline{n_{\rm H}}$, the CNM almost entirely disappears (see bottom left panels of Fig. 4). At last, and because the WNM occupies most of the volume, $f_{\rm CNM}$ necessarily increases as a function of the total mass of the gas, or equivalently $\overline{n_{\rm H}}$, even at constant pressure.

The turbulent forcing induces pressure fluctuations and shearing motions at all scales. As shown by Seifried et al. (2011), this not only frequently perturbs the gas out of the thermal equilibrium states but also strongly reduces the times spent by any fluid elements in the WNM, LNM, and CNM. Because of these two aspects, increasing the turbulent forcing reduces the mass of the CNM to the benefit of those of the LNM and WNM (right panels of Fig. 4 and Fig. 10 of Seifried et al. 2011). The mean pressure therefore increases, the 1D PDF of the density broadens and its bimodal nature progressively disappears (Piontek & Ostriker 2005; Walch et al. 2011). These effects can be magnified depending on the nature of the turbulent forcing and the power injected in the compressive and solenoidal modes. Because solenoidal motions are more efficient to prevent the gas to condensate back to the CNM phase, a pure solenoidal forcing naturally leads to larger pressure and smaller CNM fractions than those obtained with an equivalent kinetic energy injected in pure compressive modes.

As expected, the velocity dispersion of the WNM is mostly given by the strength of the turbulent forcing and the driving scale ($L_{\text{drive}} \sim L/2$, see Sect. 3.3), with a slight dependence on $\overline{n_{\text{H}}}$ and ζ . As proposed by Saury et al. (2014), a realistic value for the turbulent velocity dispersion of the WNM can be estimated by looking at the HI 21 cm emission spectra with the fewest components observed at high Galactic latitude (Kalberla et al. 2005). Toward these directions, Haud & Kalberla (2007) derive a total velocity dispersion $\sigma_{\text{tot}} = (\sigma_{\text{tur}}^2 + \sigma_{\text{thr}}^2)^{1/2} \sim 10 \text{ km s}^{-1}$, where the σ_{thr} is the 1D thermal velocity dispersion (~8.2 km s^{-1} for the WNM). In the present paper, the turbulent forcing applied to the standard simulation (see Table 2 and Fig. 4) is chosen so that $\sigma_{\text{tur}} \sim 4-5 \text{ km s}^{-1}$, in fair agreement with the observations at high Galactic latitude. While this value is chosen as a reference, the velocity dispersions obtained in all the simulations explored in this work range between 1 and 15 km s^{-1} (see Fig. 4).

3.8. Reconstruction of lines of sight

As shown in Sect. 2, the medium observed in absorption at UV and visible wavelengths extends over a very broad range of distances, from ~ 100 pc to several kpc (see Fig. 2). The targeted lines of sight may therefore contain several isolated diffuse neutral phases but also hot and warm ionized material (McKee & Cowie 1977; de Avillez & Breitschwerdt 2004). Indeed, such a superimposition of independent components is particularly well seen in submillimeter and infrared observations of the Galactic disk where the gas seen in absorption is found to cluster in several velocity components associated to known Galactic structures (e.g., Gerin et al. 2016). Since our setup only follows a piece of diffuse neutral material of size L, an additional treatment regarding the lengths of the lines of sight is therefore required in order to compare the results of the simulations to the distribution of observations. We apply here a methodology similar to that proposed by Bialy et al. (2019) and schematized in Fig. 5. We assume that a given simulation corresponds to a building block of neutral diffuse ISM. Depending on its length, any random line of sight necessarily intercepts parts or several of these building elements and an unknown mass of diffuse ionized gas



Fig. 5. Schematic view of the reconstruction of individual lines of sight over a distance l_{los} . The medium between the observer and the source is assumed to be composed of hot and warm ionized material (light blue cubes) with a volume filling factor φ and of uncorrelated pieces of diffuse neutral gas of individual size *L* (simulated boxes) with a volume filling factor (1- φ).

parametrized by its volume filling factor φ . A total sample of N simulated lines of sight is then generated as follows.

Based on the results of Sect. 2 (Fig. 2), we consider six lengths of lines of sight homogeneously distributed in log space: $l_{\rm los}$ = 100, 200, 400, 800, 1600, and 3200 pc. For each length, we generate a sample of $N_l = \frac{1}{6}w_l N$ lines of sight, where w_l are normalized weights deduced from the distribution of distances in the observed sample: $w_1 = 0.14$, $w_2 = 0.21$, $w_3 = 0.16$, $w_4 = 0.14$ 0.20, $w_5 = 0.18$, $w_6 = 0.11$ (see Fig. 2). The column densities of H and H₂ along each lines of sight are finally reconstructed by comparing the length occupied by the neutral medium $(1 - \varphi)l_{los}$ (see Fig. 5) and the size of the box L. If $(1 - \varphi)l_{los} = L$, we draw a random line of sight in the simulation and extract the corresponding column densities. If $(1 - \varphi)l_{los} < L$, we draw a random line of sight and integrate the column density over a reduced distance of $(1 - \varphi)l_{\text{los}}$. If $(1 - \varphi)l_{\text{los}} > L$, we draw $L/[(1 - \varphi)l_{\text{los}}]$ random lines of sight and add the respective individual column densities.

For the sake of simplicity, we assume here that any line of sight intercepts a constant fraction of diffuse ionized gas with $\varphi = 0.5$ (Hill et al. 2018). Similarly, we note that, while spatially uncorrelated, the pieces of diffuse neutral gas used in the reconstruction algorithm correspond to random realizations obtained with a single simulation. Potential variations of the mean density, of the external radiation field, or of the turbulent forcing that naturally follow the Galactic structure depending on the position of the source (see Appendix A) are not taken into account. All these limitations are discussed in Sect. 5.

The outcome of the reconstruction algorithm is shown in Fig. 6 which displays the 2D PH of the total proton column density $N_{\rm H}$ and the column density of molecular hydrogen $N({\rm H}_2)$ obtained with the standard simulation. Because of the flat distribution of distances in log space (see Fig. 2), the peaks of the reconstructed PH are found to be shifted toward both the large and the low values of $N_{\rm H}$ compared to those of the initial distribution (top panel of Fig. 6). This naturally enhances the initial bimodality and many lines of sight are found to be either at low (~10⁻⁵) or large (~10⁻¹) integrated molecular fractions. In addition, it induces an inclination toward large column densities; more than half of the lines of sight are found to have $N_{\rm H} > 10^{21} {\rm cm}^{-2}$. By virtue of the central limit theorem, the



Fig. 6. 2D probability histogram of the total proton column density $N_{\rm H}$ and the column density of molecular hydrogen $N({\rm H}_2)$ obtained with the standard simulation (see Table 2). *Upper panel*: original data where all lines of sight have a size L = 200 pc. *Bottom panel*: outcome of the reconstruction algorithm described in Sect. 3.8 that produces a sample of lines of sight ranging from 100 to 3200 pc. The color code indicates the fraction of lines of sight (in logarithmic scale) contained in each bin. Dotted lines are isocontours of the molecular fraction for $f_{\rm H_2} = 10^{-8}$, 10^{-6} , 10^{-4} , 10^{-2} , and 1.

molecular fraction obtained in those lines of sight tends toward the mean H_2 molecular fraction of the initial simulation with a dispersion that decreases as a function of $N_{\rm H}$.

Building simulated lines of sight from the observed distribution of sizes allows us to perform a statistical comparison of both samples and to limit the impact of observational biases to lines of sight at large column density (region E, see Sect. 5). The combined 2D distributions of $N_{\rm H}$ and $N({\rm H}_2)$ deduced from the simulations are the main focus of this paper and will be shown several times in the following sections. For obvious aesthetical reasons, and to simplify the descriptions of the figures, we will often refer to this representation as a kingfisher diagram.

3.9. Interpretative modeling

The computation of column densities over various lines of sight and the subsequent kingfisher diagrams are the outcome of three main factors: the local conditions of the gas ($n_{\rm H}$, $G_{\rm eff}$, T, and the self-shielding) which control the local abundance of H₂; the probabilistic ordering of these local conditions along any random line of sight of size $l_{\rm los}$; and finally, the distribution of sizes $l_{\rm los}$ used for the reconstruction algorithm (see Sect. 3.8). In order to separate these effects, in particular the impact of local conditions from the probabilistic aspects, and propose a physical interpretation of the behaviors shown in this paper, we developed a semi-analytical approach. The resulting model and the confrontations of its predictions to the results of numerical simulations are described in details in Appendix C. To keep the paper concise, we only summarize here its basic ingredients and our main deductions.

Following the works of Vázquez-Semadeni & García (2001), Bialy et al. (2017), we assume that any line of sight can be understood as a succession of density fluctuations. These fluctuations are supposed to be fully correlated over a distance called the "decorrelation scale", and completely uncorrelated over larger distances. Because of the biphasic nature of the neutral gas, we adopt two different decorrelation scales: y_{dec}^{diff} if $n_{\rm H} < n_{\rm H}^{\rm lim}$ and $y_{dec}^{\rm dens}$ if $n_{\rm H} \ge n_{\rm H}^{\rm lim}$. The limit $n_{\rm H}^{\rm lim}$ separating the diffuse and dense components is chosen as the inflection point between the two log-normal distributions classically found in the PH of the gas density (see Fig. C.1). Analyzing the production of H₂ and the integration of column densities in this framework leads to the following conclusions.

1. The comparisons with numerical simulations performed at four different scales and for fifteen different simulations show that the analytical model reproduces to an outstanding level the 1D PHs of the total column density $N_{\rm H}$ (see Fig. C.2), assuming that the diffuse gas is correlated over a scale $y_{\rm dec}^{\rm diff} = 0.2 L_{\rm drive}$ (i.e., 20 pc for the fiducial simulation), and that the dense gas is correlated over a scale $y_{\rm dec}^{\rm dens} = 10 \,{\rm pc} \,(\overline{n_{\rm H}}/2 \,{\rm cm}^{-3})^{1/3}$. Interestingly, the value of $y_{\rm dec}^{\rm diff}$ is similar to that obtained by Bialy & Burkhart (2020) in a set of isothermal MHD simulations with different driving scales and Mach numbers. This confirms that the WNM, which fills most of the volume, behaves like an isothermal gas perturbed by a sustained turbulent forcing.

2. Similar comparisons performed on the 2D PHs of $N_{\rm H}$ and $N({\rm H}_2)$ (Figs. C.3 and C.4 and Appendix C.4) indicate that the dense gas is necessarily composed of a distribution of structures of different sizes. Indeed, while most of the mass and volume of the cold HI can be accurately modeled with a single scale $y_{\rm dec}^{\rm dens} = 10 \,{\rm pc} \,(\overline{n_{\rm H}}/2 \,{\rm cm}^{-3})^{1/3}$, H₂ is required to be built up in smaller components. $y_{\rm dec}^{\rm dens}$ should therefore be considered as the maximum decorrelation scale of the dense gas.

3. The local production of H₂ mostly depends on the density: low density components are atomic while high density components are molecular. The threshold $n_{\rm H}^{\rm tr}$ triggering the transition between the two regimes is set by the local self-shielding (induced by the component itself) and the large-scale selfshielding (induced by the surrounding environment). The local self-shielding alone in a component of size $y_{\rm dec} \leq y_{\rm dec}^{\rm dens}$ implies

$$n_{\rm H}^{\rm tr} \sim 8 \,{\rm cm}^{-3} \,G_0^{1/2} (y_{\rm dec}/10\,{\rm pc})^{-1/2}.$$
 (21)

The large-scale self-shielding can be seen as a stochastic process that lowers this limit: for L = 200 pc, $n_{\rm H}^{\rm tr}$ is found to be reduced by a factor of two on average.

4. As schematized in Fig. 7, the distribution of normalized column densities $N_{\rm H}/l_{\rm los}$ and $N({\rm H}_2)/l_{\rm los}$ obtained for a given $l_{\rm los}$ can be divided in three categories. Lines of sight with low integrated molecular fraction $(f_{\rm H_2} < 10^{-4})$ exclusively contain components with $n_{\rm H} < n_{\rm H}^{\rm tr}$ (case a). In contrast, lines of sight with high integrated molecular fraction $(f_{\rm H_2} > 10^{-2})$ necessarily intercept at least one large or several small components at high density $n_{\rm H} > n_{\rm H}^{\rm tr}$ (case c). In spite of what intuition dictates, lines of sight with intermediate integrated molecular fraction $(10^{-4} \le f_{\rm H_2} \le 10^{-2})$ do not result from components at moderate densities $(n_{\rm H} \sim n_{\rm H}^{\rm tr})$ but from the combination of low density material and a small number of small components at high density $n_{\rm H} > n_{\rm H}^{\rm tr}$ (case b).

5. The proportions of lines of sight of types (a), (b), and (c) (Fig. 7) are given by the volume filling factors of the diffuse and dense gas and the length of the lines of sight $l_{\rm los}$. If $l_{\rm los}$ is comparable to $y_{\rm dec}^{\rm diff}$ and $y_{\rm dec}^{\rm dens}$, the 2D PH of $N_{\rm H}/l_{\rm los}$ and $N({\rm H}_2)/l_{\rm los}$ is spread and contains a large number of lines of sight of type (a). Larger $l_{\rm los}$ naturally favor lines of sight of types (b) and (c). If $l_{\rm los}$



Fig. 7. Schematic view of lines of sight of fixed length l_{los} inferred from the analytical model described in Appendix C, and corresponding contributions to the histogram of the normalized column densities $N_{\rm H}/l_{los}$ and $N({\rm H}_2)/l_{\rm los}$. *Bottom panels*: the white stars correspond to all lines of sight while the black stars correspond to the specific cases illustrated above. Any line of sight is intercepting several components of constant density $n_{\rm H}$. Diffuse components ($n_{\rm H} < n_{\rm H}^{\rm im}$) have a size $y_{\rm dec}^{\rm def}$; dense components ($n_{\rm H} \ge n_{\rm H}^{\rm im}$) have a distribution of sizes $y_{\rm dec} \le y_{\rm dec}^{\rm dens}$. Only components with densities larger than $n_{\rm H}^{\rm tr}$ are molecular (see main text). The red star in the *bottom panels* indicates the mean value of $N_{\rm H}/l_{\rm los}$ and $N({\rm H}_2)/l_{\rm los}$ computed over a large sample of lines of sight (white stars), hence the expected mean molecular fraction. The dotted line indicates a fully molecular medium.

becomes large compared to to y_{dec}^{diff} and y_{dec}^{dens} , the central limit theorem applies. The spread PH described above is squeezed along the x and y axis as both $N_{\rm H}/l_{\rm los}$ and $N({\rm H_2})/l_{\rm los}$ progressively tend toward Gaussian distributions centered on the means (red star in Fig. 7).

4. Comparison with observations

4.1. Fiducial simulation

In Fig. 8, we compare the observational dataset to the 2D PH of $N_{\rm H}$ and $N({\rm H}_2)$ (i.e., the kingfisher diagram) obtained with the reconstruction algorithm applied to the fiducial simulation for two resolutions, $R = 64^3$ and $R = 512^3$. In each panel, the color code indicates, in logarithmic scale, the fraction of lines of sight predicted for any couple $(N_{\rm H}, N({\rm H_2}))$. Quantitative comparisons of the observed and predicted fractions of lines of sight, mean molecular fractions, and dispersions in the regions A, B, C, D, and E defined in Sect. 2 (see Table 1) are given in Fig. 9. Unexpectedly, the sample of lines of sight built from the fiducial simulation reproduces, to an outstanding level, the global trend of the HI-to-H₂ transition and its statistical properties. Without taking into account any possible variation of the parameters along the lines of sight or from one line of sight to the next, the structures induced by the joint actions of turbulence and thermal instability alone are found to produce a wealth of lines of sight whose probabilities of occurrence match those derived from the observations.

In particular, the integrated molecular fraction is predicted to have a bimodal distribution with a transition occurring at $N_{\rm H} \sim 3 \times 10^{20}$ cm⁻² and extending over one order of magnitude of total column density. More quantitatively, the fractions of lines of sight simulated and observed in regions A, B, C, and D are found to differ by 50% at the most (see Fig. 9). Similarly,



Fig. 8. Comparison of the observational dataset (black points) to the 2D probability histogram reconstruction algorithm (see Sect. 3.8) applied to the fiducial simulation (colored histogram). Results are shown for two resolutions, $R = 64^3$ (*top panel*) and $R = 512^3$ (*bottom panel*). Observations include detections of H₂ (circles) and upper limits on N(H₂) (arrows). The color code indicates the fraction of lines of sight (in logarithmic scale) contained in each bin. As a reminder, contours of the regions A, B, C, D, and E defined in Sect. 2 (see Table 1) are also displayed.

the observed and simulated mean molecular fractions and their corresponding dispersions are found to be comparable and to differ by less than a factor of three in region B, and less than a factor of two in regions C, and D. Because of the distribution of background sources, the reconstructed ensemble predicts a strong increase in probability from region C to D, which contains a large fraction of the entire sample, and a decrease in the dispersion of $f_{\rm H_2}$ as a function of $N_{\rm H}$. At last, the probability of occurrence of lines of sight with $N_{\rm H} < 3 \times 10^{19}$ and $f_{\rm H_2} > 10^{-3}$ or $N_{\rm H} > 3 \times 10^{21}$ and $f_{\rm H_2} < 10^{-3}$ are rare to nonexistent. All these features are also found in the observational dataset.

Notwithstanding, Figs. 8 and 9 also reveal a few discrepancies. Firstly, about 16 observed lines of sight (out of 360) lay at the border of the simulated distribution, in regions where the predicted probability is $\leq 10^{-4}$ (or even smaller than 10^{-5} for the white regions of Fig. 8), a value far smaller than the inverse number of observations $\sim 3 \times 10^{-3}$. This implies that the simulation used here somehow fails to explain on its own part of the diversity observed in the Solar Neighborhood. Secondly, the mean molecular fractions predicted in regions A and B are found to be noticeably smaller than that derived from the observations, by about a factor of ten and three, respectively. It is important to note, however, that these molecular fractions deduced from the observations are probably overestimated as a third of the observed lines of sight contained in these regions correspond



E. Bellomi et al.: 3D chemical structure of diffuse turbulent ISM. I.

Fig. 9. *Top frame*: comparisons of the observational dataset (black points) with the 2D probability histograms of $N_{\rm H}$ and $N({\rm H}_2)$ computed from the reconstruction algorithm (see Sect. 3.8) applied to the simulations. *Bottom frame*: fraction of lines of sight (%), and mean value μ and dispersion σ of the logarithm of the molecular fraction computed from the simulated histograms in the regions A, B, C, D, and E defined in Table 1. Numbers correspond to the values of %, μ , and σ . The color code indicates a measure of distance (in arbitrary units) between the observed and simulated values in order to guide the eye. These comparisons are shown in each frame for 15 different simulations with G_0 varying from 0.5 (*left panels*) to 4 (*right panels*). All other parameters are set to their fiducial values (see Table 2).

to upper limits on $N(H_2)$. Finally, the simulated sample clearly shows that a substantial fraction of the lines of sight are in the region E, with an integrated probability of 4%. If the observational sample of 360 targets is unbiased, between 10 and 18 lines of sight should have been observed in this region, which is not the case. These discrepancy will be discussed in more details in Sect. 5.

4.2. Impact of the resolution

The comparison of the two panels of Fig. 8 shows that the kingfisher diagram is independent of the resolution over about one order of magnitude of scales, from 64^3 to 512^3 . Even if not systematically shown, this unusual result is not limited to the fiducial setup but is a general feature of all the simulations explored in this work (see Fig. D.2 for instance). Our interpretation is based on the analytical model presented in Appendix C and summarized in the previous section.

Evidently, high resolution simulations are important to accurately describe small scale structures. In particular, large resolution are required for modeling the formation of dense and gravitationally bound environments and follow their collapse. The fact that the kingfisher diagrams are independent of resolution for $R \ge 64^3$ therefore suggests that the cold and dense structures between 0.4 and 3 pc have no influence on the distributions of $N_{\rm H}$ and $N({\rm H}_2)$ for the fiducial simulations and are insignificant in the mass and the volume budgets of H and H₂. This is in line with the conclusions deduced from the analytical model. Indeed, as explained in the previous section, the turbulent forcing at large scale induces density fluctuations in the diffuse gas that extend over ~20 pc and a distribution of dense structures with sizes smaller than ~ 10 pc. While the total quantity of matter is inferred to be contained in the diffuse and the largest dense components, H₂ is exclusively built up in dense components smaller than ~ 10 pc. The results of Fig. 8 combined with conclusions deduced from the analytical model therefore imply that the structures contributing the most to the mass and volume of H_2 are above 3 pc and below 10 pc for the fiducial simulation.

Beside the physical insights on the typical scales participating to the build-up of column densities, this result also provides a strong justification for using simulations with moderate numerical resolution for the study of the HI-to-H₂ transition. In all this work we therefore adopt a standard resolution of $R = 256^3$ unless indicated otherwise.

4.3. Impact of G_0 and $\overline{n_H}$

Figures 9 and 10 show comparisons between the observational dataset and the 2D PHs extracted from the simulations for $0.5 \text{ cm}^{-3} \le \overline{n_{\text{H}}} \le 4 \text{ cm}^{-3}$, $0.5 \le G_0 \le 4$, and two sets of values of the box size and the turbulent forcing, L = 200 pc and $F = 9 \times 10^{-4} \text{ kpc Myr}^{-2}$ (Fig. 9) and L = 50 pc and $F = 3.6 \times 10^{-3} \text{ kpc Myr}^{-2}$ (Fig. 10). While the trend and the statistics of the HI-to-H₂ transition weakly depend on the resolution, they strongly depend on the total mass of the gas parametrized by $\overline{n_{\text{H}}}$ and on the UV illumination factor. As G_0 increases or $\overline{n_{\text{H}}}$ decreases, (1) the fraction of lines of sight with large $f(\text{H}_2)$ drops to the benefit of lines of sight with low $f(\text{H}_2)$, (2) the transition is shifted toward larger total column density and its width increases, and (3) the molecular fraction globally decreases over all lines of sight while its dispersion increases. Interestingly, and because $\overline{n_{\text{H}}}$ and G_0 have opposite effects, the simulations that reproduce the most accurately the observed statistics of the HI-to-H₂ transition follow a trend with $G_0/\overline{n_{\text{H}}} \sim 0.5-1$. While

similar, the effect of these two parameters are, however, not symmetrical. In particular, $\overline{n_{\rm H}}$ has an obvious and strong impact on the fraction of lines of sight in region E, regardless of G_0 . Likewise, the fraction of lines of sight at low column densities (regions A and B) and the mean molecular fraction at large column densities (regions C and D) are not constant for a given $G_0/\overline{n_{\rm H}}$ ratio but respectively decrease and increase with $\overline{n_{\rm H}}$. All these properties effectively break the degeneracies between the two parameters. All things considered, the tightest concordance between observed and simulated data is obtained for $\overline{n_{\rm H}} \sim 1-2$ cm⁻³, in agreement with the Galactic midplane density deduced from HI surveys in the Solar Neighborhood (Kalberla & Kerp 2009).

At first sight, the results described above seem obvious as they somehow mimic the dependencies of the HI-to-H2 transition found with detailed models of photodissociation regions (e.g., Krumholz et al. 2009; Sternberg et al. 2014). Such an interpretation is, however, a dangerous misconception. Indeed, while G_0 is tightly linked to the effective radiation field G_{eff} that locally illuminates the gas, the mean density $\overline{n_{\rm H}}$ should not be mistaken for the local density. Moreover, the results displayed in Figs. 9 and 10 cannot be compared to PDR models because they are statistical in nature. For instance, simulations at high G_0 do not preclude the existence of clouds with high molecular fractions. Indeed, increasing G_0 may lead to denser local environments with larger molecular fractions whose probability of occurrence along a line of sight is simply reduced. It implies that the results shown here are very different from PDR model predictions. They rather reflect the complex link between the global properties of the simulation (mass, illumination, driving scale) on the one side, and the local conditions and their probability distribution functions on the other side.

Because the local abundance of H₂ is inversely proportional to G_0 , increasing G_0 naturally reduces the local self-shielding. Similarly, increasing G_0 or decreasing $\overline{n_H}$ reduce the large-scale self-shielding. As a result, the density threshold n_H^{tr} required to produce highly molecular environments (see item 3. of Sect. 3.9) rises by a factor of three when either G_0 is multiplied or $\overline{n_H}$ is divided by a factor of eight. While significant, such an effect on the local conditions of the gas is, however, too shallow to fully explain the variations observed in Figs. 9 and 10.

Indeed, regardless of local conditions, G_0 and $\overline{n_{\rm H}}$ have a major impact on the probabilistic reconstruction of lines of sight. As shown in Sect. 3.7, increasing G_0 or decreasing $\overline{n_{\rm H}}$ strongly reduce the fractions of mass and volume occupied by the dense and cold gas. This not only reduces the molecular fraction averaged over the entire simulation (red star in Fig. 7) but also favors the occurrence of lines of sight with low or intermediate $f_{\rm H_2}$ (cases (a) and (b) in Fig. 7). When combined with the distribution of sizes l_{los} , the HI-to-H₂ transition is naturally shifted toward larger $N_{\rm H}$, and the asymptotic molecular fraction at high $N_{\rm H}$ drops. Moreover, because the central limit theorem requires larger lines of sight to apply, the HI-to-H₂ transition is naturally wider and the dispersion of lines of sight at large molecular fraction increases. This final property is particularly well seen in the kingfisher diagram obtained for the simulation at $G_0 = 0.5$ and $\overline{n_{\rm H}}$ = 4 cm⁻³ where most of the lines of sight follow an homothetic transformation of the mean normalized column densities $N_{\rm H}/l_{\rm los}$ and $N({\rm H_2})/l_{\rm los}$ (red star in Fig. 7).

4.4. Impact of the box size L

The impacts of the box size revealed by comparing Figs. 9 and 10 partly follow the results of the previous section. Reducing L by a



Fig. 10. Same as in Fig. 9 for simulations with a box size of L = 50 pc instead of 200 pc. The turbulent forcing is adjusted as in Fig. 4 ($F = 3.6 \times 10^{-3}$ kpc Myr⁻²) to obtain similar velocity dispersions for L = 50 pc and L = 200 pc.

factor of four drastically reduces the total amount of matter in the simulation, hence the absorption of the impinging UV radiation field and the large-scale self-shielding. As shown in Sect. 3.7, the mean pressure of the gas rises while the mass and volume occupied by the CNM decreases. All the local and statistical effects described in the previous section therefore apply and modify the kingfisher diagrams accordingly. Changing L has, however, two additional and specific consequences.

Because L_{drive} is four times smaller in the simulations displayed in Fig. 10 than in those displayed in Fig. 9, the decorrelation scales y_{dec}^{diff} and y_{dec}^{dens} are correspondingly smaller. According to the interpretation given in Sect. 3.9, all the reconstructed lines of sight are therefore considerably larger than individual density fluctuations. This not only favors the occurrence of lines of sight at intermediate and high molecular fractions (cases (b) and (c) of Fig. 7) but also magnify the impact of the central limit



Fig. 11. KS distance between the simulations and the observational sample as a function of the mean density $\overline{n_{\rm H}}$ and the UV scaling factor $G_0 = 0.5$ (red), 1 (green), 2 (orange), and 4 (blue). Results obtained with and without gravity are shown with solid lines and dashed lines, respectively. All other parameters are set to their standard values (see Table 2). Points correspond to reliable measurements of the KS distances. Triangles indicate lower limits corresponding to simulations where the upper error bar on $R_{\rm KS}$ tends toward infinity (see Appendix D).

theorem (see item 5. of Sect. 3.9), as already illustrated by Bialy et al. (2019). The kingfisher diagrams shown in Fig. 10 are thus squeezed along the *x* and *y* axis compared to those of Fig. 9. Consequently, the simulations at L = 50 pc predict almost no line of sight at low total column density ($N_{\rm H} \le 3 \times 10^{19}$ cm⁻²) and systematically underestimate the proportion of line of sight in region A compared to the observations.

Because it controls the total amount of matter, changing L has finally a major effect on gravitational forces. For L = 50 pc, the size of the largest dense clouds are almost always smaller than the Jean length computed in the CNM. Consequently, selfgravity plays almost no role in the physical state or the evolution of the gas for all simulations at L = 50 pc. The specific impact of gravity on the probability histograms of column densities is described in more details in the following section.

4.5. Impact of gravity

To facilitate the comparison between simulations and observations and avoid a pedestrian repetition of the kingfisher diagrams, we developed a modified version of the Kolmogorov-Smirnov test. This test, fully explained in Appendix D and validated over 30 different simulations, defines a value $R_{\rm KS}$, called the KS distance, that measures how two 2D PDFs (or 2D PHs) differ from one another. In a nutshell, any observational datapoint in a 2D diagram can be used to divide the diagram into four different regions which each contain different fractions of observed and simulated data. The modified KS test simply searches for the observational datapoint and the region that maximize the ratio between the fractions of simulated and observed lines of sight. The distance $R_{\rm KS}$ is the absolute value of the logarithm of this ratio: a value $R_{\rm KS} = 1$ therefore implies that some region in the 2D diagram contains tens times more or ten times less simulated data than required to explain the observations, and that all the other regions have smaller ratios.

The results of the KS test applied to 15 simulations run with and without gravity are shown in Fig. 11. As expected from Sect. 4.3, the KS distance strongly depends on both $\overline{n_{\rm H}}$ and G_0 . In comparison, gravity appears to have a limited impact on the kingfisher diagrams. While including gravity seems to slightly reduce the KS distance, the trends as functions of $\overline{n_{\rm H}}$ and G_0 and the set of simulations found to minimize $R_{\rm KS}$ remain



Fig. 12. KS distances between the simulations and the observational sample computed for five values of the acceleration parameter $F = 4.5 \times 10^{-5}$, 1.5×10^{-4} , 4.5×10^{-4} , 1.5×10^{-3} , and 4.5×10^{-3} kpc Myr⁻², and three values of the compressive ratio $\zeta = 0.1$ (blue points), 0.5 (orange points), and 0.9 (green points), which set the balance between compressive and solenoidal forcing (see Sect. 3.3). All other parameters are set to their standard values (see Table 2). Points correspond to reliable measurements of the KS distances. The triangle indicates a lower limit corresponding to a simulation where the upper error bar on R_{KS} tends toward infinity (see Appendix D).

unaltered. These results can be understood by simple statistical considerations.

The impact of gravity in multiphase simulations is to produce self-gravitating environments which appear as a high density tail in the PDF of the gas density. Because these self-gravitating clumps are dense and fully molecular, they often dominate the integrated column densities of both HI and H2 along any line of sight that intercept them, and therefore favor lines of sight at high molecular fraction (case (c) in Fig. 7). This not only induces a tail at high $N_{\rm H}$ and $N({\rm H}_2)$ in the kingfisher diagram but may also contribute to shift the HI-to-H₂ transition to lower $N_{\rm H}$ as faint but highly molecular lines of sight starts to appear. The importance of these two effects depends on the area filling factor of self-gravitating clumps and on whether they carry or not a substantial fraction of the mass of the dense gas. For $\overline{n_{\rm H}} = 0.5 \text{ cm}^{-3}$, self-gravitating clumps occupy less than 0.001% of the entire volume and carry less than a percent of the mass of the gas. These fractions increase, however, as functions of $\overline{n_{\rm H}}$: for $\overline{n_{\rm H}} = 4 \text{ cm}^{-3}$, self-gravitating environments occupy 0.01% of the volume and contain as much as 30% of the total mass. Therefore, while the impacts of gravity on the kingfisher diagram is negligible for most simulations, they become important at high $\overline{n_{\rm H}}$: this property is effectively captured by the KS distances displayed in Fig. 11.

4.6. Impact of turbulent forcing

As done in the previous section, the impact of the turbulent forcing is discussed through the results of the Kolmogorov-Smirnov test. The KS distance obtained for various configurations of the turbulent forcing (Fig. 12) shows that the strength of turbulence affects differently the HI-to-H₂ transition depending on its nature: highly compressive turbulent forcing produces virtually identical column density distributions over almost two decades of the turbulent acceleration parameter *F*; oppositely the strength of the forcing significantly modifies these distributions if a substantial fraction of the kinetic energy is injected in pure solenoidal modes. The tightest agreement with the observational sample is obtained for $F \sim 1.5 \times 10^{-4}$ kpc Myr⁻² if $\zeta \ge 0.5$ and for all $F \ge 1.5 \times 10^{-4}$ kpc Myr⁻² if $\zeta = 0.1$. According to Sect. 3.7 (Fig. 4), these values corresponds to a WNM velocity dispersion $\sigma_{tur} \sim 2 \text{ km s}^{-1}$ if $\zeta \ge 0.5$, and $\sigma_{tur} \ge 2 \text{ km s}^{-1}$ for $\zeta = 0.1$. In any case, a small amount of turbulence is always required.

All these characteristics are consequences of the fractions of volume and mass contained in the CNM structures and their size distribution. Without turbulence or with a weak turbulent forcing, CNM clouds are found to evolve toward large-scale entities which evaporate slowly. Such a configuration favor lines of sight at high molecular fraction (type (c) in Fig. 7) and leads to an overestimation of the global amount of H₂ in the local diffuse ISM. If the turbulent forcing increases, the CNM becomes progressively organized into a distribution of structures of different sizes down to the numerical resolution. Simultaneously, the fraction of mass located in the CNM phase diminishes to the benefit of the LNM (see Sect. 3.7). Both effects reduce the mean molecular fraction of the gas and favor the occurrence of lines of sight of type (b) (see Fig. 7), in better agreement with the observational sample. Larger turbulent forcing ultimately lead to an underestimation of the global amount of H_2 in the local ISM. However, and as shown in Sect. 3.7, this last effect is much more pronounced for a pure solenoidal forcing which efficiently prevents the unstable gas to condensate back to cold and dense environments compared to a pure compressive forcing.

Here again, the probabilistic information contained in the kingfisher diagram proves to provide a valuable tool to analyze the nature of turbulence in the diffuse local ISM. Indeed, the velocity dispersion deduced from the comparison of observations and simulations at high solenoidal forcing is substantially smaller than the velocity dispersion observed at high Galactic latitude (see Sect. 3.7) and the velocity dispersion deduced from the observations of CO at a scale of 200 pc (Hennebelle & Falgarone 2012). The fact that the observed statistics of the HIto-H₂ transition is reproduced over a broader range of velocity dispersion for a compressive forcing suggests that the large-scale turbulence of the diffuse local ISM is dominated by compressive modes. This picture is coherent with the results of Saury et al. (2014) who found that a large-scale compressive forcing induces a distribution of thermal pressure in excellent agreement with that derived from the observations of the fine structure lines of CI (Jenkins & Tripp 2011). It is also coherent with the fact that 200 pc corresponds to the width of the Galactic disk seen in CO and therefore to the limit above which the equipartition between compressive and solenoidal modes switches from the values expected in a 3D fluid to those expected in a 2D medium, hence values of ζ smaller than 0.5. At last, it concurs with the conclusions of Iffrig & Hennebelle (2017) who found that compressible modes dominates at low altitudes, close to the equatorial plane, in simulations where the ISM turbulence is self-consistently driven by supernovae explosions.

4.7. Impact of the magnetic field

The impact of the initial magnetic field on the kingfisher diagram simply reveals the competition between thermal instability which induces the production of dense environments, and the magnetic pressure which acts against this evolution. As shown in Fig. 13, the KS distance obtained for different magnetic field intensities is found to be constant until a critical value of $B_x \sim 4 \,\mu\text{G}$ above which the predicted and observed distribution of $N_{\rm H}$ and $N({\rm H}_2)$ significantly differ from one another. The initial homogeneous magnetic field adopted in the standard simulation, $B_x = 3.8 \,\mu\text{G}$, is just under this critical value and corresponds to the case where the energy density of the large scale coherent magnetic field and



Fig. 13. KS distances between the simulations and the observational sample computed for six values of the initial magnetic field B_x . All other parameters are set to their standard values (see Table 2). Points correspond to reliable measurements of the KS distances. The triangle indicates a lower limit corresponding to a simulation where the upper error bar on R_{KS} tends toward infinity (see Appendix D).

the thermal energy density of the WNM are equivalent. It follows that reproducing the observed 2D PH of $N_{\rm H}$ and $N({\rm H}_2)$ requires a magnetic field intensity below or at equipartition.

Interestingly, the value of B_x adopted for the fiducial simulation leads, at steady-state, to a constant magnetic field intensity $B \sim 5 \ \mu\text{G}$ for $n_{\rm H} < 10 \ {\rm cm}^{-3}$, and a field intensity that scales as $B \propto (n_{\rm H})^{0.3}$ for $n_{\rm H} > 10 \ {\rm cm}^{-3}$. Those values are comparable with the magnetic field intensities obtained in diffuse and molecular environments from Zeeman measurements (Crutcher et al. 2010) and those obtained in the most diffuse phases from Faraday rotation measurements toward extragalactic radio sources (e.g., Frick et al. 2001). It appears that these information on the interstellar magnetic field are encoded, at least partly, in the statistical properties of the HI-to-H₂ transition.

5. Discussion

5.1. Observational biases

All the results presented in Sect. 4 are discussed under the assumption that the observational dataset is unbiased, meaning that the underlying lines of sight correspond to a random sample with no selection effect. This is not true. As Krumholz et al. (2008) already noticed, "the FUSE and the Copernicus lines of sight have been specifically chosen to probe a certain range of column densities with a selection biased against high extinction which makes determining column densities very costly or altogether impossible." Indeed, very few stars emit a UV radiation field strong enough to be detected through large visual extinction material. The problem is not due to the increase in H_2 absorption, which can be overcome by focusing on fainter bands, but to dust absorption itself and the complication of the structure of NaI often used as a proxy to derive the column density of HI (Rachford et al. 2002). Moreover, because such bias depends on the sensitivity of the instrument, it necessarily applies at different column densities for FUSE and Copernicus.

This selection effect obviously complicates the comparison between simulations and observations. As shown in Sect. 4.3 and Figs. 9 and 10, several simulations, including the fiducial setup, predict a significant fraction of lines of sight at high column densities ($N_{\rm H} \ge 10^{22}$ cm⁻², region E), in apparent contradiction with the observations. However, the fact that the fraction of stars dismissed by selection is unknown makes it difficult to assess whether this result reveals an actual and important statistical discrepancy or a simple observational limit. It implies that the likelihood of a simulation to be representative of the local diffuse gas cannot be estimated from this criterion alone. It must involve other observational signatures such as the average and the dispersion of the molecular fraction in regions A, B, C, and D, or even the chemical and statistical signatures of other atomic and molecular lines. This latter aspect is currently under development and will be the subject of the next paper of this series.

The fact that the limitations differ between FUSE and *Copernicus* surveys finally raises the question of the validity of studying the two samples simultaneously. We find, however, that performing comparisons with simulations on the two samples separately gives very similar results and does not impact any of our conclusions. It is so because the observational bias discussed above occurs at an extinction which increases as the natural logarithm of the square root of the instrument sensitivity. The largest total column density probed by FUSE is thus only three times larger than that observed by *Copernicus* (Gillmon et al. 2006; Rachford et al. 2009). Moreover, the number of lines of sight observed by FUSE that are above the maximum extinction seen by Copernicus corresponds to a small fraction of the entire sample (Gillmon et al. 2006).

5.2. Variations of physical conditions in the local ISM

The reconstruction of the simulated sample of lines of sight and the subsequent comparisons with the observations are done assuming that the local ISM can be built out of a single simulation (see Sect. 3.8). This approach was chosen in order to highlight the natural variations induced by turbulence and thermal instability alone in a diffuse neutral gas with a known averaged density and UV illumination factor. However, because the medium probed by the observations extends in all directions around the sun over a maximum distance of 3 kpc (see Fig. 2), it stands to reason that potential variations of all parameters should be taken into account, not only from one line of sight to another but also along a single and outstretched line of sight. Indeed, such considerations would offer a natural explanation for observations whose existence is in contradiction with the corresponding simulated probability of occurrence (see Sect. 4.1).

The total proton mass surface density deduced from HI and CO all sky surveys appears to be rather constant in the Galactic layer located between 5 and 12 kpc from the Galactic center (Fig. 9 of Miville-Deschênes et al. 2017). Taking into account variations of the ISM scale height above the Galaxy (Kalberla & Kerp 2009), the midplane proton density is expected to vary by less than a factor of two over the corresponding volume.

Using the Galactic star distribution of Wainscoat et al. (1992) and the grains composition and size distribution of Weingartner & Draine (2001), Porter & Strong (2005) and Moskalenko et al. (2006) estimated the energy density of the radiation field across the Galaxy. Similarly to the midplane density, the mean UV radiation field is estimated to vary by a factor two over the volume considered in this paper (Fig. 2 of Porter & Strong 2005). Interestingly, this estimation is far smaller than the variations derived by Jenkins & Tripp (2011) (Fig. 8 in their paper) from the observation of the fine structure line of CI in the local gas. Such discrepancies could be explained by the fact that Jenkins & Tripp (2011) perform local measurements: the observed gas could thus be located close to or far from an irradiating star. Alternatively, we note that the results obtained by Jenkins & Tripp (2011) are derived from models at equilibrium which do not take into account the uncorrelated perturbations of pressure and density in a turbulent multiphase medium: this naturally favors large fluctuations of the UV radiation field.

The cosmic ray ionization rate inferred from submillimeter observations of several molecular ions, including OH^+ , H_2O^+ , ArH^+ , and H_3^+ , shows a wide distribution across the Galactic disk (Indriolo et al. 2015). A recent estimation performed by Neufeld & Wolfire (2017) suggests that this rate could vary by a factor of five in the gas located between 5 and 12 kpc from the Galactic center.

Finally, potential variations of the composition and the size distribution of grains in the local diffuse gas should also be considered. While the extinction curve is found to be surprisingly uniform in the Milky Way (Schlafly et al. 2016), the local distribution of grains could change, not only along the line of sight but between the atomic and ionized phase and the molecular clouds. This would modify the efficiency of the photoelectric effect and the equilibrium between the two stable states of the neutral ISM, and would also have an impact the H₂ formation rate.

Interestingly, if uncorrelated, the expected variations of $\overline{n_{\rm H}}$ and G_0 alone would help to explain the slight discrepancies described in Sect. 4.3 but it would also lead to a dispersion of lines of sight far larger than those observed (see Figs. C.2 and C.3). The fact that the predicted statistics of the HI-to-H₂ transition is close to the observed sample therefore suggests that the variations of all the parameters described above cannot be considered independently but must follow strong correlations which apply locally (as discussed, for instance, by Bialy et al. 2019) but also across the Galactic disk.

5.3. Fraction of ionized gas

The fraction of volume φ occupied by the ionized phases of the ISM, the warm ionized medium (WIM) and the hot ionized medium (HIM), plays an important role in our reconstruction algorithm. As illustrated in Fig. 5, this parameter controls the filling factor of the neutral medium along a line of sight of length $l_{\rm los}$. Unfortunately, its value in the Milky Way is highly uncertain and still debated.

The consensus is that the volume of the HIM far exceeds that of the WIM and results from an interplay between supernovae explosions, which regularly produce hot gas in the Galactic disk, and buoyancy, which lifts this gas into the halo, releasing the pressure in the midplane. Early analytical studies neglecting buoyancy (McKee & Ostriker 1977) predicted a large fraction of HIM in the midplane ($\varphi \sim 95\%$). In contrast, early numerical simulations, including the cycle of matter and energy between the disk and the halo, led to considerably smaller predictions with $\varphi \sim 25\%$ (de Avillez & Breitschwerdt 2004). This value is now considered as a lower limit by the most recent numerical simulations which reveal the importance of the driving mode of supernovae explosions (Walch et al. 2015) and of the photoelectric heating (Li et al. 2015; Hill et al. 2018) on the volume of the HIM. These latest studies estimate that $20\% \le \varphi \le 90\%$ in the Galactic midplane.

Highly uncertain, the volume filling factor of the HIM can also vary from one line of sight to another (Fig. 1 of Hill et al. 2018). In this paper, we adopt a constant and conservative value $\varphi = 0.5$ for every line of sight. Changing φ would have the effect of either squeeze or stretch the predicted 2D PHs displayed in Figs. 9 and 10 along the x axis, and to modify the balance of probabilities of occurrence of lines of sight at high and low molecular fraction. Taking into account a realistic distribution of φ would require to simulate the Galactic disk and halo over a scale of several kiloparsecs and to properly model and follow the impact of supernovae explosions. This is far beyond the scope of the present paper.

5.4. Galactic vertical structure

The simulated 2D PH of the HI-to- H_2 transition are found to slightly depend on the Galactic gravitational potential. This odd result is nothing but an artifact of the physical scales considered here. Since the sizes of all simulations are below 200 pc, the gas expands, at the most, over 100 pc above the Galactic plane, a distance far smaller than the characteristic scale of variation of the thermal pressure expected for a turbulent gas in hydrostatic equilibrium. It follows that the column densities show no significant variation as a function of the position of the line of sight or its angle with the Galactic plane.

This setup, initially chosen to favor the physical resolution, is a strong shortcoming which prevents us from using and studying the information carried by the Galactic latitude of each observation. Indeed, the comparison between the FUSE halo survey and the data collected by FUSE and *Copernicus* in the Galactic disk indicates that the HI-to-H₂ transition at high latitude occurs at a total hydrogen column density ~2 times smaller than that in the disk (Gillmon et al. 2006). Similarly to the previous section, studying this effect would requires to compute the local vertical structure of our Galaxy over several hundreds of parsecs, taking into account the hot and ionized component of the ISM.

5.5. Doppler broadening parameter

The self-shielding of H₂ depends on the dispersion of the Lyman and Werner lines which is usually modeled with a turbulent Doppler broadening parameter b_D (Draine & Bertoldi 1996 and Eq. (16)). In single cloud models, this parameter has the effect of shifting the HI-to-H₂ transition toward larger total column density without modifying any of the asymptotic states (e.g., Bialy et al. 2017). In this paper, we identify this parameter with the turbulent velocity dispersion of the gas at large scales and therefore adopt $b_D = 8 \text{ km s}^{-1}$ for the fiducial simulation. This is done to prevent an overestimation of the H₂ self-shielding at large scales, at the cost of underestimating the self-shielding at the scale of a CNM clouds.

To estimate the effect of this parameter, all the grids presented in this paper were also run assuming $b_D = 2 \text{ km s}^{-1}$, which roughly corresponds to the velocity dispersion expected at a scale of 10 pc for the fiducial simulation. While locally important, b_D is found to have a relatively small impact on the 2D PHs of $N_{\rm H}$ and $N({\rm H}_2)$: increasing b_D by a factor of four slightly increases the width of the HI-to-H₂ transition and the fraction of lines of sight in region B. We interpret this limited effect as a consequence of the fact that the asymptotic molecular states of any cloud are independent of b_D .

Even so, it should be stressed that b_D has a strong impact on the local molecular fraction in transition regions. Therefore, and while inconsequential for the global statistics of H and H₂, the value of the Doppler broadening parameter might be paramount for any chemical species preferentially formed at the border of molecular clouds. As proposed by Bialy et al. (2019), the H₂ selfshielding could be computed self-consistently using the velocity and density fields of the simulation and a cost effective radiative transfer method. This would prevent the dilemma of favoring large-scale or small-scale self-shielding.

5.6. H₂ self-shielding at high temperature

The prescription of H_2 self-shielding used in this paper (Eq. (16)) is taken from Draine & Bertoldi (1996). As discussed by Wolcott-Green et al. (2011), such a prescription is reliable for diffuse gas at low temperature but becomes less and less reliable for high temperature environments (T > 500 K) where efficient collisional excitation of H2 in its rovibrational levels reduces the self-shielding. To estimate the impact of this process, we ran the fiducial simulation with the alternative self-shielding function proposed by Wolcott-Green et al. (2011) (Eq. (12) in their paper). This prescription leads to a similar probability histogram and therefore does not influence the global analysis of the kingfisher diagram presented in this paper. However, we note that adopting this alternative prescription slightly increases the width of the HI-to-H2 transition, and induces more lines of sight at intermediate molecular fraction (region B, see Fig. 3), in better agreement with the observations.

6. Summary and conclusions

This paper presents an exhaustive parametric study of the transition from atomic to molecular hydrogen in the local diffuse ISM. Using state-of-the-art MHD simulations, and an ensemble of 305 runs, we quantify separately the impacts of the mean density, the UV radiation field, the integral scale, the resolution, the turbulent forcing, the magnetic field, and the gravity on the molecular content of multiphase environments. The original feature of this work is to not only focus on the production of individual column densities but also on their statistics, meaning the probabilities of occurrence of these column densities along random lines of sight. For the first time, both the chemical and statistical information are used concomitantly, through the socalled kingfisher diagrams, to interpret the distribution of H and H₂ observed toward 360 lines of sight across the local interstellar matter.

The results of the simulations are interpreted with a semianalytical model which attempts to separate the effects of local conditions from those induced by the probabilistic reconstruction of individual lines of sight. To compare the kingfisher diagrams to the observational sample, we propose a new version of the Kolmogorov-Smirnov test which can be generalized and used for the comparison of two probability histograms or distribution functions in any dimension larger than one.

Taking into account the distance of each background source and simulating random lines of sight over the same distribution of distances is paramount to explain the range of observed column densities and their corresponding statistics. Once this aspect is included, the joint actions of thermal instability and large-scale turbulence in the standard simulation are found to produce a wealth of lines of sight which reproduce the observed position and width of the HI-to-H₂ transition, and whose probabilities of occurrence match those derived from the observations. The agreement is so remarkable that it is almost unnecessary to invoke variations of physical conditions along lines of sight or from one line of sight to another.

The minimal KS distance obtained over the entire grid is ~ 0.5 . Such a value implies that there exists a small group of lines of sight in the observational sample whose probability of occurrence is under- or over-predicted by about a factor of three. However, it also implies that the probability of occurrence of any other group of observed lines of sight, small or large, is reproduced to a better level.

The distribution of column densities computed from the simulations strongly depends on the Galactic midplane density parametrized by the mean density $\overline{n_{\rm H}}$, the density of OB stars parametrized by the UV scaling factor G_0 , and the scale of neutral diffuse clouds parametrized by the box size L. It is so because these three parameters not only regulate the mean pressure of the gas, hence the fractions of mass and volume occupied by the CNM and WNM, but also control the typical scale of density fluctuations in the WNM and the distribution of sizes of the CNM structures where H₂ forms. The tightest concordance between the observed and simulated samples is obtained for a mean density $\overline{n_{\rm H}} = 1-2$ cm⁻³ and a UV radiation field scaling factor $G_0 = 1$ (in Habing units), in good agreement with the values deduced from HI and CO all sky surveys and from direct observations of the UV radiation field in the Solar Neighborhood. The range of observed column densities of H and H_2 requires a box size L = 200 pc which corresponds to the estimated scale of HI superclouds.

Within this setup, the column densities of HI are inferred to be built up in large-scale WNM and CNM structures correlated in densities over ~ 20 and ~ 10 pc, respectively. In contrast H₂ is inferred to be built up at smaller scales. However, the fact that the kingfisher diagram is independent from the resolution of the simulation suggests that most of the mass and volume of H₂ is contained in CNM structures between ~3 and ~10 pc. All these values are given for the standard simulation (L = 200 pc and $L_{\rm drive} \sim 100 \text{ pc}$) but naturally depend on the size of the box and the mean density of the gas.

In spite of the strong influences of $\overline{n_{\rm H}}$, G_0 , and L, the statistical properties of the HI-to-H₂ transition are otherwise remarkably stable. Admittedly, the kingfisher diagram depends on the strength of the turbulence if most of the forcing is injected in solenoidal modes; however such a configuration prevents to reproduce the observational sample unless the large-scale velocity dispersion of the gas is unrealistically small. In contrast, if most of the kinetic energy is injected in compressive modes, the kingfisher diagram is found to weakly depend on the strength of the forcing. Similarly, the HI-to-H₂ transition is almost not affected by gravity and is found to weakly depend on the Doppler broadening parameter and the strength of the magnetic field, as long as $B_x \leq 4 \mu$ G. The 2D PH of the column densities of H and H₂ is therefore a valuable tool to constrain the nature of the turbulent forcing at large scale; however, it provides few or no information regarding the velocity dispersion of the gas, the amount of gravitationally bound environments and the strength of the magnetic field. Other observational tracers are required.

All these results open new perspectives for the study of the chemical state of the ISM in which any observation must be understood through the combination of local physical conditions and the probabilistic ordering of these conditions along the line of sight. In particular, similar studies should be applied to all atomic and molecular species with observational samples large enough to conduct statistical analysis. It also invites to expend the study of PHs to higher dimensions, taking into account simultaneously the joint information contained in the column densities of several species. In this context, the generalization of the Kolmogorov-Smirnov test proposed in this paper will be very valuable. All these aspects are currently under development and will be the subjects of the following papers of this series.

from the Programme National "Physique et Chimie du Milieu Interstellaire" (PCMI) of CNRS/INSU with INC/INP co-funded by CEA and CNES. We would finally like to thank E. Falgarone and M. Gerin for the stimulating discussions we had and their precious comments regarding this work.

References

- Abgrall, H., Le Bourlot, J., Pineau des Forêts, G., et al. 1992, A&A, 253, 525
- Ambartsumian, V. A. 1947, The evolution of stars and astrophysics
- Andersson, B. G., Wannier, P. G., & Crawford, I. A. 2002, MNRAS, 334, 327
- André, M. K., Oliveira, C. M., Howk, J. C., et al. 2003, ApJ, 591, 1000
- Audit, E., & Hennebelle, P. 2005, A&A, 433, 1
- Audit, E., & Hennebelle, P. 2010, A&A, 511, A76
- Berger, M. J., & Oliger, J. 1984, J. Comput. Phys., 53, 484
- Bialy, S., & Burkhart, B. 2020, ApJ, 894, L2 Bialy, S., & Sternberg, A. 2019, ApJ, 881, 160
- Bialy, S., Burkhart, B., & Sternberg, A. 2017, ApJ, 843, 92
- Bialy, S., Neufeld, D., Wolfire, M., Sternberg, A., & Burkhart, B. 2019, ApJ, 885, 109
- Bigiel, F., Leroy, A., Walter, F., et al. 2008, AJ, 136, 2846
- Bigiel, F., Leroy, A., & Walter, F. 2011, IAU Symp., 270, 327
- Bohlin, R. C., Savage, B. D., & Drake, J. F. 1978, ApJ, 224, 132
- Bohlin, R. C., Hill, J. K., Jenkins, E. B., et al. 1983, ApJS, 51, 277
- Bouy, H., & Alves, J. 2015, A&A, 584, A26
- Bowen, D. V., Jenkins, E. B., Tripp, T. M., et al. 2008, ApJS, 176, 59
- Bron, E., Le Bourlot, J., & Le Petit, F. 2014, A&A, 569, A100
- Browning, M. K., Tumlinson, J., & Shull, J. M. 2003, ApJ, 582, 810
- Burgh, E. B., France, K., & McCandliss, S. R. 2007, ApJ, 658, 446
- Burgh, E. B., France, K., & Jenkins, E. B. 2010, ApJ, 708, 334
- Cartledge, S. I. B., Meyer, D. M., & Lauroesch, J. T. 2003, ApJ, 597, 408
- Cartledge, S. I. B., Lauroesch, J. T., Meyer, D. M., & Sofia, U. J. 2004, ApJ, 613, 1037
- Cartledge, S. I. B., Clayton, G. C., Gordon, K. D., et al. 2005, ApJ, 630, 355
- Cartledge, S. I. B., Lauroesch, J. T., Meyer, D. M., Sofia, U. J., & Clayton, G. C. 2008, ApJ, 687, 1043
- Christensen, C., Quinn, T., Governato, F., et al. 2012, MNRAS, 425, 3058
- Clark, P. C., Glover, S. C. O., Ragan, S. E., & Duarte-Cabral, A. 2019, MNRAS,
- 486, 462
- Crutcher, R. M., Wandelt, B., Heiles, C., Falgarone, E., & Troland, T. H. 2010, ApJ, 725, 466
- Dame, T. M., Hartmann, D., & Thaddeus, P. 2001, ApJ, 547, 792
- de Avillez, M. A., & Breitschwerdt, D. 2004, A&A, 425, 899
- de Boer, K. S., Lenhart, H., van der Hucht, K. A., et al. 1986, A&A, 157, 119
- Dickey, J. M., & Lockman, F. J. 1990, ARA&A, 28, 215
- Diemer, B., Stevens, A. R. H., Forbes, J. C., et al. 2018, ApJS, 238, 33
- Diplas, A., & Savage, B. D. 1994, ApJS, 93, 211
- Draine, B. T. 1978, ApJS, 36, 595
- Draine, B. T., & Bertoldi, F. 1996, ApJ, 468, 269
- Elmegreen, B. G., & Elmegreen, D. M. 1987, ApJ, 320, 182
- Fasano, G., & Franceschini, A. 1987, MNRAS, 225, 155
- Federman, S. R. 1982, ApJ, 257, 125
- Federman, S. R., Strom, C. J., Lambert, D. L., et al. 1994, ApJ, 424, 772
- Federrath, C. 2013, MNRAS, 436, 1245 Federrath, C., Klessen, R. S., & Schmidt, W. 2008, ApJ, 688, L79
- Federrath, C., Roman-Duval, J., Klessen, R. S., Schmidt, W., & Mac Low, M.-M. 2010, A&A, 512, A81
- Ferlet, R., Vidal-Madjar, A., & Gry, C. 1985, ApJ, 298, 838
- Field, G. B. 1965, ApJ, 142, 531
- Fitzpatrick, E. L. 1999, PASP, 111, 63
- Fitzpatrick, E. L., & Massa, D. 1986, ApJ, 307, 286
- Fitzpatrick, E. L., & Massa, D. 1990, ApJS, 72, 163
- Frick, P., Stepanov, R., Shukurov, A., & Sokoloff, D. 2001, MNRAS, 325, 649
- Fromang, S., Hennebelle, P., & Teyssier, R. 2006, A&A, 457, 371
- Fruscione, A., Hawkins, I., Jelinsky, P., & Wiercigroch, A. 1994, ApJS, 94, 127 Gaia Collaboration (Brown, A. G. A., et al.) 2018, A&A, 616, A1
- Garmany, C. D., & Stencel, R. E. 1992, A&AS, 94, 211
- Gerin, M., Neufeld, D. A., & Goicoechea, J. R. 2016, ARA&A, 54, 181
- Gillmon, K., Shull, J. M., Tumlinson, J., & Danforth, C. 2006, ApJ, 636, 891
- Girichidis, P., Konstandin, L., Whitworth, A. P., & Klessen, R. S. 2014, ApJ, 781, 91
- Glover, S. C. O., Federrath, C., Mac Low, M.-M., & Klessen, R. S. 2010, MNRAS, 404, 2
- Gnacinski, P., & Krogulec, M. 2006, Acta Astron., 56, 373
- Gnedin, N. Y., Tassis, K., & Kravtsov, A. V. 2009, ApJ, 697, 55
- Goldsmith, P. F., Li, D., & Krčo, M. 2007, ApJ, 654, 273
- Goldsmith, P. F., Velusamy, T., Li, D., & Langer, W. 2009, ASP Conf. Ser., 417, 177

Acknowledgements. We thank the referee for his careful reading and his valuable comments and suggestions. The research leading to these results has received fundings from the European Research Council, under the European Community's Seventh framework Programme, through the Advanced Grant MIST (FP7/2017-2022, No 742719). We would also like to acknowledge the support

E. Bellomi et al.: 3D chemical structure of diffuse turbulent ISM. I.

- Gong, M., Ostriker, E. C., & Wolfire, M. G. 2017, ApJ, 843, 38
- Gong, M., Ostriker, E. C., & Kim, C.-G. 2018, ApJ, 858, 16
- Gudennavar, S. B., Bubbly, S. G., Preethi, K., & Murthy, J. 2012, ApJS, 199, 8
- Habing, H. J. 1968, Bull. Astron. Inst. Netherlands, 19, 421 Haud, U., & Kalberla, P. M. W. 2007, A&A, 466, 555
- Heiles, C., & Troland, T. H. 2003a, ApJ, 586, 1067
- Heiles, C., & Troland, T. H. 2003b, VizieR Online Data Catalog: J/ApJS/145/329
- Hennebelle, P., & Falgarone, E. 2012, A&ARv, 20, 55
- Hennebelle, P., & Iffrig, O. 2014, A&A, 570, A81
- Hennebelle, P., & Pérault, M. 1999, A&A, 351, 309
- Hennebelle, P., Banerjee, R., Vázquez-Semadeni, E., Klessen, R. S., & Audit, E. 2008, A&A, 486, L43
- Henshaw, J. D., Ginsburg, A., Haworth, T. J., et al. 2019, MNRAS, 485, 2457
- Hill, A. S., Mac Low, M.-M., Gatto, A., & Ibáñez-Mejía, J. C. 2018, ApJ, 862, 55
- Hobbs, L. M. 1978, ApJS, 38, 129
- Hu, C.-Y., Naab, T., Walch, S., Glover, S. C. O., & Clark, P. C. 2016, MNRAS, 458, 3528
- Iffrig, O., & Hennebelle, P. 2017, A&A, 604, A70
- Indriolo, N., Neufeld, D. A., Gerin, M., et al. 2015, ApJ, 800, 40
- Inoue, T., Inutsuka, S.-i., & Koyama, H. 2006, ApJ, 652, 1331
- Iwasaki, K., & Inutsuka, S.-i. 2014, ApJ, 784, 115
- Jenkins, E. B., & Tripp, T. M. 2011, ApJ, 734, 65
- Jenkins, E. B., Savage, B. D., & Spitzer, L., J. 1986, ApJ, 301, 355
- Jensen, A. G., & Snow, T. P. 2007a, ApJ, 669, 378
- Jensen, A. G., & Snow, T. P. 2007b, ApJ, 669, 401
- Joung, M. K. R., & Mac Low, M.-M. 2006, ApJ, 653, 1266
- Kalberla, P. M. W., & Kerp, J. 2009, ARA&A, 47, 27
- Kalberla, P. M. W., Burton, W. B., Hartmann, D., et al. 2005, A&A, 440, 775
- Kim, J.-G., & Kim, W.-T. 2013, ApJ, 779, 48
- Koyama, H., & Inutsuka, S.-i. 2002a, ApJ, 564, L97
- Koyama, H., & Inutsuka, S. 2002b, 8th Asian-Pacific Regional Meeting, Volume II, eds. S. Ikeuchi, J. Hearnshaw, & T. Hanawa, 159
- Krumholz, M. R., McKee, C. F., & Tumlinson, J. 2008, ApJ, 689, 865
- Krumholz, M. R., McKee, C. F., & Tumlinson, J. 2009, ApJ, 693, 216
- Kuijken, K., & Gilmore, G. 1989, MNRAS, 239, 571
- Körtgen, B., Federrath, C., & Banerjee, R. 2019, MNRAS, 482, 5233
- Larson, R. B. 1981, MNRAS, 194, 809
- Le Bourlot, J., Le Petit, F., Pinto, C., Roueff, E., & Roy, F. 2012, A&A, 541, A76
- Le Coupanec, P., Rouan, D., Moutou, C., & Léger, A. 1999, A&A, 347, 669
- Lee, H.-H., Herbst, E., Pineau des Forêts, G., Roueff, E., & Le Bourlot, J. 1996, A&A, 311, 690
- Lehner, N., Jenkins, E. B., Gry, C., et al. 2003, ApJ, 595, 858
- Lenz, D., Hensley, B. S., & Doré, O. 2017, ApJ, 846, 38
- Leroy, A., Bolatto, A., Stanimirovic, S., et al. 2007, ApJ, 658, 1027
- Leroy, A. K., Walter, F., Sandstrom, K., et al. 2013, AJ, 146, 19
- Lesaffre, P., Gerin, M., & Hennebelle, P. 2007, A&A, 469, 949
- Lesaffre, P., Pineau des Forêts, G., Godard, B., et al. 2013, A&A, 550, A106
- Lesaffre, P., Todorov, P., Levrier, F., et al. 2020, MNRAS, 495, 816
- Li, M., Ostriker, J. P., Cen, R., Bryan, G. L., & Naab, T. 2015, ApJ, 814, 4
- Linsky, J. L., Redfield, S., Wood, B. E., & Piskunov, N. 2000, ApJ, 528, 756 Liszt, H. 2014, ApJ, 783, 17
- Lupi, A., Volonteri, M., & Silk, J. 2017, MNRAS, 470, 1673
- Maíz-Apellániz, J. 2001, AJ, 121, 2737
- Marchal, A., Miville-Deschênes, M.-A., Orieux, F., et al. 2019, A&A, 626, A101
- Mathis, J. S., Mezger, P. G., & Panagia, N. 1983, A&A, 500, 259
- McKee, C. F., & Cowie, L. L. 1977, ApJ, 215, 213
- McKee, C. F., & Krumholz, M. R. 2010, ApJ, 709, 308
- McKee, C. F., & Ostriker, J. P. 1977, ApJ, 218, 148
- Micic, M., Glover, S. C. O., Federrath, C., & Klessen, R. S. 2012, MNRAS, 421, 2531
- Miville-Deschênes, M. A., & Martin, P. G. 2007, A&A, 469, 189
- Miville-Deschênes, M.-A., Murray, N., & Lee, E. J. 2017, ApJ, 834, 57
- Moskalenko, I. V., Porter, T. A., & Strong, A. W. 2006, ApJ, 640, L155
- Murray, C. E., Stanimirović, S., Goss, W. M., et al. 2015, ApJ, 804, 89
- Murray, C. E., Stanimirović, S., Goss, W. M., et al. 2018, ApJS, 238, 14
- Nagashima, M., Koyama, H., & Inutsuka, S.-i. 2005, MNRAS, 361, L25 Nakanishi, H., & Sofue, Y. 2016, PASJ, 68, 5
- Neckel, T., & Klare, G. 1980, A&AS, 42, 251
- Neufeld, D. A., & Wolfire, M. G. 2017, ApJ, 845, 163
- Nickerson, S., Teyssier, R., & Rosdahl, J. 2018, MNRAS, 479, 3206
- Noterdaeme, P., Petitjean, P., Srianand, R., Ledoux, C., & Le Petit, F. 2007,
- A&A, 469, 425 Oegerle, W. R., Jenkins, E. B., Shelton, R. L., Bowen, D. V., & Chayer, P. 2005, ApJ, 622, 377
- Ostriker, E. C., McKee, C. F., & Leroy, A. K. 2010, ApJ, 721, 975
- Padoan, P., Nordlund, A., & Jones, B. J. T. 1997, MNRAS, 288, 145
- Padoan, P., Pan, L., Haugbølle, T., & Nordlund, Å. 2016, ApJ, 822, 11

- Palazzi, E., Mandolesi, N., & Crane, P. 1992, ApJ, 398, 53
- Pan, K., Federman, S. R., Cunha, K., Smith, V. V., & Welty, D. E. 2004, ApJS, 151.313
- Pan, K., Federman, S. R., Sheffer, Y., & Andersson, B. G. 2005, ApJ, 633, 986 Passot, T., & Vázquez-Semadeni, E. 1998, Phys. Rev. E, 58, 4501
- Peacock, J. A. 1983, MNRAS, 202, 615
- Perryman, M. A. C., Lindegren, L., Kovalevsky, J., et al. 1997, A&A, 500, 501
- Piontek, R. A., & Ostriker, E. C. 2005, ApJ, 629, 849
- Porter, T. A., & Strong, A. W. 2005, ICRC, 4, 77
- Rachford, B. L., Snow, T. P., Tumlinson, J., et al. 2002, ApJ, 577, 221
- Rachford, B. L., Snow, T. P., Destree, J. D., et al. 2009, ApJS, 180, 125
- Richings, A. J., & Schaye, J. 2016a, MNRAS, 460, 2297
- Richings, A. J., & Schaye, J. 2016b, MNRAS, 458, 270
- Ritchey, A. M., Martinez, M., Pan, K., Federman, S. R., & Lambert, D. L. 2006, ApJ, 649, 788
- Roth, K. C., & Blades, J. C. 1995, ApJ, 445, L95
- Ryu, K. S., Dixon, W. V. D., Hurwitz, M., et al. 2000, ApJ, 529, 251
- Saury, E., Miville-Deschênes, M.-A., Hennebelle, P., Audit, E., & Schmidt, W. 2014, A&A, 567, A16
- Savage, B. D., Bohlin, R. C., Drake, J. F., & Budich, W. 1977, ApJ, 216, 291
- Savage, B. D., Massa, D., Meade, M., & Wesselius, P. R. 1985, ApJS, 59, 397
- Savage, B. D., Meade, M. R., & Sembach, K. R. 2001, ApJS, 136, 631
- Schlafly, E. F., Meisner, A. M., Stutz, A. M., et al. 2016, ApJ, 821, 78
- Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525
- Schmidt, W., Federrath, C., Hupp, M., Kern, S., & Niemeyer, J. C. 2009, A&A, 494, 127
- Schruba, A., Leroy, A. K., Walter, F., et al. 2011, AJ, 142, 37
- Seifried, D., Schmidt, W., & Niemeyer, J. C. 2011, A&A, 526, A14
- Seifried, D., Walch, S., Girichidis, P., et al. 2017, MNRAS, 472, 4797
- Sheffer, Y., Rogers, M., Federman, S. R., Lambert, D. L., & Gredel, R. 2007, ApJ, 667, 1002
- Sheffer, Y., Rogers, M., Federman, S. R., et al. 2008, ApJ, 687, 1075
- Shull, J. M., & van Steenberg, M. E. 1985, ApJ, 294, 59

Snow, T. P., Destree, J. D., & Jensen, A. G. 2007, ApJ, 655, 285

Snow, T. P., Ross, T. L., Destree, J. D., et al. 2008, ApJ, 688, 1124

Smith, R. J., Glover, S. C. O., Clark, P. C., Klessen, R. S., & Springel, V. 2014, MNRAS, 441, 1628

Spitzer, L. 1978, Physical Processes in the Interstellar Medium (Weinheim:

Tabone, B. 2018, PhD thesis, Université de recherche Paris Sciences et Lettres,

Valdivia, V., Hennebelle, P., Gerin, M., & Lesaffre, P. 2015, EAS Publ. Ser., 75,

Vázquez-Semadeni, E., Gómez, G. C., Jappsen, A. K., et al. 2007, ApJ, 657, 870

Wainscoat, R. J., Cohen, M., Volk, K., Walker, H. J., & Schwartz, D. E. 1992,

Walch, S., Wünsch, R., Burkert, A., Glover, S., & Whitworth, A. 2011, ApJ,

Welsh, B. Y., Vedder, P. W., & Vallerga, J. V. 1990, ApJ, 358, 473 Welsh, B. Y., Vedder, P. W., Vallerga, J. V., & Craig, N. 1991, ApJ, 381, 462

Wolfire, M. G., Hollenbach, D., McKee, C. F., Tielens, A. G. G. M., & Bakes,

Wolfire, M. G., McKee, C. F., Hollenbach, D., & Tielens, A. G. G. M. 2003,

Zari, E., Hashemi, H., Brown, A. G. A., Jardine, K., & de Zeeuw, P. T. 2018,

A36, page 19 of 33

Wolcott-Green, J., Haiman, Z., & Bryan, G. L. 2011, MNRAS, 418, 838

Valdivia, V., Hennebelle, P., Gérin, M., & Lesaffre, P. 2016, A&A, 587, A76

Sternberg, A., Le Petit, F., Roueff, E., & Le Bourlot, J. 2014, ApJ, 790, 10

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

Tumlinson, J., Shull, J. M., Rachford, B. L., et al. 2002, ApJ, 566, 857

Valdivia, V., Godard, B., Hennebelle, P., et al. 2017, A&A, 600, A114

Wakker, B. P., Savage, B. D., Sembach, K. R., et al. 2003, ApJS, 146, 1

Walch, S., Girichidis, P., Naab, T., et al. 2015, MNRAS, 454, 238

Weingartner, J. C., & Draine, B. T. 2001, ApJ, 548, 296

Welty, D. E., & Crowther, P. A. 2010, MNRAS, 404, 1321

E. L. O. 1995, ApJ, 443, 152

York, D. G. 1976, ApJ, 204, 750

Wolff, B., Koester, D., & Lallement, R. 1999, A&A, 346, 969

Snow, T. P., & McCall, B. J. 2006, ARA&A, 44, 367

Stone, J. M., & Zweibel, E. G. 2009, ApJ, 696, 233

Valdivia, V., & Hennebelle, P. 2014, A&A, 571, A46

van Dishoeck, E. F., & Black, J. H. 1986, ApJS, 62, 109

van Dishoeck, E. F., & Black, J. H. 1989, ApJ, 340, 273

van Steenberg, M. E., & Shull, J. M. 1988, ApJS, 67, 225

Vázquez-Semadeni, E., & García, N. 2001, ApJ, 557, 727

Wiley-VCH)

France

393

ApJS, 83, 111

ApJ, 587, 278

A&A, 620, A172

733, 47

Sternberg, A. 1988, ApJ, 332, 400

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

Appendix A: Observations of HI and H₂ in the local diffuse ISM

The complete observational dataset used in this work is given in Table A.1, including the sources identifiers, the coordinates, and the column densities of HI and H₂. The column densities of H₂ are obtained through direct observations of its FUV absorption lines. The sample presented in Table A.1 therefore results from a combination of the Copernicus survey of nearby stars (e.g., Savage et al. 1977; Bohlin et al. 1983) and the FUSE surveys of the Galactic disk and the Galactic halo (e.g., Rachford et al. 2002, 2009; Lehner et al. 2003; Cartledge et al. 2004; Pan et al. 2004: Gillmon et al. 2006: Jensen & Snow 2007a.b). In contrast, and as reviewed by Fruscione et al. (1994), the HI column densities are derived from both direct methods, which include $Ly\alpha$ absorption studies, EUV observations of stellar spectra, and observations of the 21cm line, and indirect methods, which include curve of growth of neutral and singly ionized atoms and optical interstellar absorption lines of NaI which are both found to correlate with N(H) (de Boer et al. 1986; Ferlet et al. 1985). The column densities of HI given in Table A.1 therefore result from a combination of FUV and optical studies for nearby stars (e.g., Bohlin et al. 1978, 1983; Diplas & Savage 1994; Fitzpatrick & Massa 1990; Jensen & Snow 2007b,a) and radio studies of the 21cm line for extragalactic sources at high Galactic latitude (Wakker et al. 2003; Gillmon et al. 2006).

As reported by all these authors, the indirect methods relying on the observations of metals can be subjects to uncertainties mostly due to the assumptions made regarding the elemental abundances. In addition, the column densities of HI derived from the emission profiles of the 21cm line can also be highly uncertain because the measurements are done over a beam far larger than the pencil-beam sampled by H₂ data but also because it requires to identify in the HI profiles the components associated to the molecular gas. It should thus be kept in mind that while the errors on the column densities of H₂ are somehow limited, those on HI can be sometimes larger than a factor of five, in particular for the lines of sight at high Galactic latitude (Gillmon et al. 2006). The color excess E(B-V) given in Table A.1 finally results from a compilation which includes direct measurements of the star reddening compared to its intrinsic $(B-V)_0$ color (e.g., Savage et al. 1977; Fitzpatrick & Massa 1990; Diplas & Savage 1994; Rachford et al. 2002) and dust emission maps from the IRAS telescope⁴ (Schlegel et al. 1998).

With all these data at hand, we adopt the following methodology to derive the total proton column densities $N_{\rm H}$: if the column densities of HI and H₂ are available, then $N_{\rm H}$ is computed as $N({\rm H}) + 2N({\rm H}_2)$; if not, $N_{\rm H}$ is derived from the reddening E(B-V) as $N_{\rm H} = 5.8 \times 10^{21}E(B-V)$ cm⁻² assuming a standard Galactic extinction curve and the average interstellar ratio $R_V = A_V/E(B-V) = 3.1$ (Fitzpatrick & Massa 1986; Fitzpatrick 1999). It should be noted that the $N_{\rm H}/E(B-V)$ ratio observed at low column density is larger than the standard value used in this work (Liszt 2014; Lenz et al. 2017). Because of this and the uncertainties on the HI column densities discussed above, the values of $N_{\rm H}$ derived here should be considered as estimates. Examples of the underlying uncertainties on $N_{\rm H}$ can be seen in Table A.1 where $N({\rm H})$ sometimes exceeds slightly the total column density derived from E(B-V).

 Table A.1. Observational dataset used in this work.

Source ID	Longitude [°]	Latitude [°]	Distance [kpc]	E(B-V)	$\log_{10}(N(H))$	$\log_{10}(N(H_2))$	$\log_{10}(N_{\rm H})$
BD +35 4258	77.190	-4.740	2.000	0.290 (9)	21.28 (41)	19.56 (41)	21.30
BD +48 3437	93.560	-2.060	2.639	0.350 (5)	21.36 (18)	20.42 (43)	21.45
BD +53 2820	101.240	-1.690	3.521	0.330 (45)	21.35 (32)	20.01 (32)	21.39
CPD -59 2603	287.590	-0.690	4.098	0.460 (8)	21.46 (18)	20.15 (39)	21.50
CPD -69 1743	303.710	-7.350	3.817	0.300 (18)	21.12 (18)	19.99 (43)	21.18
ESO 141-G55	338.180	-26.710	_	0.111 (21)	20.70 (36)	19.32 (36)	20.73
HD 000886	109.430	-46.680	0.255	0.010 (3)	20.04 (3)	<14.20 ⁽²⁾	19.76
HD 001383	119.020	-0.890	3.344	0.510 (18)	21.36 (18)	20.45 (32)	21.46
HD 002905	120.840	0.140	0.521	0.350 (3)	21.20 (3)	20.27 (2)	21.29
HD 005394	123.580	-2.150	0.190	0.210 (45)	19.99 (13)	17.51 (32)	19.99
HD 010516	131.320	-11.330	0.151	0.200 (3)	20.54 (3)	19.08 (2)	20.57

Notes. The distance of each source is computed from the parallax measured by *Gaia* if the data is given in the DR2 catalog (Gaia Collaboration 2018); otherwise, the distance of the source is taken from Gudennavar et al. (2012). Column densities are expressed in cm⁻². The total proton column densities $N_{\rm H}$ are computed as $N(\rm H) + N(\rm H_2)$ if the column densities of HI and H₂ are available, or derived from the reddening E(B - V) as $N_{\rm H} = 5.8 \times 10^{21} E(B - V)$ cm⁻² assuming a standard Galactic extinction curve and the average interstellar ratio $R_V = A_V/E(B - V) = 3.1$ (Fitzpatrick & Massa 1986; Fitzpatrick 1999).

References. ⁽¹⁾ York (1976); ⁽²⁾Savage et al. (1977); ⁽³⁾Bohlin et al. (1978); ⁽⁴⁾Hobbs (1978); ⁽⁵⁾Neckel & Klare (1980); ⁽⁶⁾Federman (1982); ⁽⁷⁾Bohlin et al. (1983); ⁽⁸⁾Shull & van Steenberg (1985); ⁽⁹⁾Savage et al. (1985); ⁽¹⁰⁾Jenkins et al. (1986); ⁽¹¹⁾van Steenberg & Shull (1988); ⁽¹²⁾van Dishoeck & Black (1989); ⁽¹³⁾Welsh et al. (1990); ⁽¹⁴⁾Welsh et al. (1991); ⁽¹⁵⁾Garmany & Stencel (1992); ⁽¹⁶⁾Palazzi et al. (1992); ⁽¹⁷⁾Federman et al. (1994); ⁽¹⁸⁾Diplas & Savage (1994); ⁽¹⁹⁾Fruscione et al. (1994); ⁽²⁰⁾Roth & Blades (1995); ⁽²¹⁾Schlegel et al. (1998); ⁽²²⁾Wolff et al. (1999); ⁽²³⁾Le Coupanec et al. (1999); ⁽²⁴⁾Linsky et al. (2000); ⁽²⁵⁾Ryu et al. (2000); ⁽²⁶⁾Savage et al. (2001); ⁽²⁷⁾Rachford et al. (2002); ⁽²⁸⁾Andersson et al. (2002); ⁽²⁹⁾André et al. (2003); ⁽³⁰⁾Lehner et al. (2003); ⁽³¹⁾Cartledge et al. (2003); ⁽³²⁾Cartledge et al. (2004); ⁽³³⁾Oegerle et al. (2005); ⁽³⁴⁾Pan et al. (2005); ⁽³⁵⁾Gnacinski & Krogulec (2006); ⁽³⁶⁾Gillmon et al. (2006); ⁽³⁷⁾Ritchey et al. (2006); ⁽³⁸⁾Snow et al. (2007); ⁽³⁹⁾Burgh et al. (2007); ⁽⁴⁰⁾Sheffer et al. (2007); ⁽⁴¹⁾Jensen & Snow (2007a); ⁽⁴²⁾Cartledge et al. (2008); ⁽⁴³⁾Sheffer et al. (2008); ⁽⁴⁴⁾Snow et al. (2008); ⁽⁴⁵⁾Bowen et al. (2008); ⁽⁴⁶⁾Rachford et al. (2009); ⁽⁴⁷⁾Burgh et al. (2010); ⁽⁴⁸⁾Gudennavar et al. (2012).

⁴ Available from the NED (http://nedwww.ipac.caltech.edu) and on the more recent plateform https://irsa.ipac.caltech.edu.

Table A.1. continued.

Source ID	Longitude [°]	Latitude [°]	Distance [kpc]	E(B-V)	$\log_{10}(N(\mathrm{H}))$	$\log_{10}(N(H_2))$	$\log_{10}(N_{\rm H})$
HD 012323	132.910	-5.870	2.809	0.410 (32)	21.18 (32)	20.32 (43)	21.29
HD 013268	133.960	-4.990	1.692	0.411 (21)	21.34 (18)	20.51 (43)	21.45
HD 013745	134.580	-4.960	2.268	0.460 (18)	21.26 (18)	20.67 (43)	21.44
HD 014434	135.080	-3.820	2.558	0.480 (18)	21.45 (18)	20.43 (43)	21.53
HD 014633	140.780	-18.200	5.051	0.070 (45)	20.56 (3)	<19.11 (2)	20.61
HD 015137	137.460	-7.580	3.704	0.310 (18)	21.11 (18)	20.32 (43)	21.23
HD 015558	134.720	0.920	2.151	0.830 (44)	21.52 (18)	20.89 (44)	21.69
HD 017040	198.380	-62.380	0.211	0.480 (16)	-	20.81 (16)	21.44
HD 021278	147.520	-6.190	0.178	0.100 (8)	21.28 (8)	19.48 (2)	21.29
HD 021483	158.870	-21.300	0.533	0.560 (17)	-	20.81 (2)	21.51
HD 021856	156.320	-16.750	0.466	0.190 (3)	21.04 (3)	20.04 (2)	21.12
HD 022928	150.280	-5.770	0.113	0.040 (8)	<21.11 (8)	19.30 (2)	20.37
HD 022951	158.920	-16.700	0.330	0.240 (3)	21.04 (3)	20.46 (2)	21.22
HD 023180	160.360	-17.740	0.245	0.300 (3)	20.90 (3)	20.61 (2)	21.21
HD 023408	166.170	-23.510	0.106	0.070 (37)	_	19.75 ⁽²⁾	20.61
HD 023478	160.760	-17.420	0.288	0.250 (23)	21.01 (42)	20.48 (42)	21.21
HD 023480	166.570	-23.750	0.106	0.100 (37)	_	20.12 (2)	20.76
HD 023630	166.670	-23.460	0.125	0.050 (45)	20.08 (14)	19.54 ⁽²⁾	20.28
HD 024190	160.390	-15.180	0.413	$0.300^{(5)}$	$21.18^{(42)}$	$20.38^{(42)}$	21.30
HD 024398	162.290	-16.690	0.294	$0.273^{(35)}$	$20.81^{(10)}$	$20.67^{(2)}$	21.20
HD 024534	163.080	-17.140	0.810	0.560 (11)	$20.73^{(18)}$	$20.92^{(27)}$	21.34
HD 024760	157 350	-10.090	0.082	$0.100^{(8)}$	$20.45^{(18)}$	$19.52^{(10)}$	20.54
HD 024912	160 370	-13 110	0.725	$0.291^{(35)}$	$21.11^{(10)}$	$20.53^{(2)}$	21.29
HD 026571	172 420	-20550	0.723	0.290 (17)	19 65 ⁽¹⁶⁾	20.85 (16)	21.22
HD 027778	172.720	-17390	0.274	0.400 ⁽¹⁷⁾	21 10 ⁽²⁷⁾	20.79 (27)	21.19
HD 028497	208 780	-37400	0.224	0.400	$20.30^{(8)}$	$14.82^{(2)}$	20.30
HD 020427	199 310	-31 380	0.400	$0.020^{(7)}$	20.50 20.45 ⁽¹⁰⁾	$< 17.41^{(7)}$	20.06
HD 029240	174 050	-13350	0.155	$1.040^{(16)}$	20.45 20.16 ⁽¹⁶⁾	$2154^{(16)}$	21.85
HD 030122	176 620	-14.030	0.155	$0.603^{(21)}$	20.10	$20.70^{(43)}$	21.03
HD 030614	144 070	14.030	0.237	0.003	$20.90^{(3)}$	20.70 20.34 ⁽²⁾	21.04
HD 031237	196 270	-24 560	0.750	0.020	20.90 20.41 ⁽⁷⁾	$< 17.45^{(7)}$	20.54
HD 034078	172 080	-2,260	0.205	$0.000^{(8)}$	20.41 21.30 ⁽⁸⁾	20 88 (43)	20.54
HD 034816	214 830	-26240	0.270	$0.030^{(7)}$	$20.30^{(8)}$	$< 15.04^{(7)}$	20.24
HD 034989	194 620	-15 610	0.534	0.030	20.50 21.11 ⁽³⁾	$<18.45^{(2)}$	20.24
HD 035149	199 160	-17860	0.368	0.130	$20.74^{(3)}$	$18 30^{(32)}$	20.00
HD 035439	201.960	-18 290	0.257	0.050 ⁽⁷⁾	20.74 20.46 ⁽¹⁰⁾	14 78 ⁽¹⁰⁾	20.74
HD 035715	200.090	-17220	0.259	$0.050^{(10)}$	20.40	14.70 14 78 ⁽¹⁰⁾	20.40
HD 036166	200.070	_17.220	0.257	0.000	20.37 20.32 ⁽⁷⁾	$<15.00^{(7)}$	20.37
HD 036486	201.070	-17 700	0.420	0.030	20.32 20.18 ⁽⁸⁾	$14.68^{(2)}$	20.24
HD 036822	195 400	-12290	0.420	0.070	20.10 20.81 ⁽³⁾	19.32 ⁽²⁾	20.10
HD 036861	195.100	-12.000	0.271	0.100 (45)	20.01 $20.87^{(8)}$	19.12 (10)	20.01
HD 037022	209.010	-19 380	0.369	$0.320^{(18)}$	20.67	$< 17.55^{(2)}$	20.87
HD 037043	209.520	-19 580	0.502	0.020	$20.00^{(8)}$	$14.69^{(2)}$	20.30
HD 037128	205.320	-17.300	0.463	$0.000^{(3)}$	20.50	$16.57^{(2)}$	20.30
HD 037202	185 600	-5.640	0.405	$0.050^{(3)}$	20.43	$-17.67^{(2)}$	20.45
HD 037202	170 040	-5.040	0.080	$0.050^{(32)}$	20.04	$20.61^{(43)}$	20.40
HD 037307	206 820	-1.030	0.989	$0.400^{(7)}$	21.26	$\sim 18 \ 30^{(7)}$	21.43
HD 037408	200.820	-17.340	0.358	$0.000^{(3)}$	20.32	(10.30^{+1})	20.34
HD 037003	200.430	-10.390	0.332	$0.080^{(18)}$	20.41 (32)	20.85 (31)	20.41
IID 037903	200.830	-10.340	0.401	$0.330^{(21)}$	21.10 (41)	20.85	21.40
пр 03008/ Пр 020444	207.070	-10.200	0.339	$0.717^{()}$	10.75 (3)	20.04 (2)	21.23 10.75
	257.290	-27.100	0.400	$0.040^{(3)}$	19.75	15.51 (2)	19.73
HD 038//I	214.510	-18.500	0.520	$0.070^{(3)}$	$20.52^{(3)}$	10.52 (38)	20.52
HD 039680	194.070	-5.880	5.5/8	$0.300^{(-3)}$	21.30 (10)	19.53 (30)	21.31
HD 040111	183.970	0.840	0.663	$0.150^{(3)}$	$21.08^{(0)}$	$19.73^{(2)}$	21.12
HD 040893	180.090	4.340	4.000	$0.460^{(+0)}$	$21.50^{(41)}$	20.58 (41)	21.59
HD 041117	189.650	-0.860	1.000	$0.450^{(0)}$	$21.40^{(0)}$	20.69 (41)	21.54
HD 042087	187.750	1.770	1.400	0.290 (45)	21.40 (18)	20.52 (41)	21.50

Table A.1. continued.

Source ID	Longitude [°]	Latitude [°]	Distance [kpc]	E(B-V)	$\log_{10}(N(\mathrm{H}))$	$\log_{10}(N(\mathrm{H}_2))$	$\log_{10}(N_{\rm H})$
HD 043384	187.990	3.530	2.494	0.580 (46)	21.27 (46)	20.87 (46)	21.52
HD 044506	241.630	-20.780	0.616	0.020 (45)	20.30 (7)	<14.85 ⁽⁷⁾	20.06
HD 044743	226.060	-14.270	0.153	0.030 (35)	<18.70 ⁽³⁾	<17.30 ⁽²⁾	20.24
HD 045314	196.960	1.520	0.827	0.370 (45)	21.04 (18)	20.60 (44)	21.28
HD 046056	206.340	-2.250	1.524	$0.490^{(8)}$	21.38 (18)	20.68 (41)	21.53
HD 046202	206.310	-2.000	1.350	0.380 (45)	21.58 (18)	20.68 (41)	21.68
HD 047129	205.880	-0.310	1.520	$0.360^{(3)}$	$21.08^{(3)}$	$20.54^{(2)}$	21.28
HD 047839	202.940	2.200	0.950	$0.070^{(3)}$	$20.31^{(3)}$	$15.54^{(2)}$	20.31
HD 048099	206.210	0.800	1.916	$0.270^{(3)}$	$21.15^{(3)}$	$20.29^{(2)}$	21.26
HD 050896	234,760	-10.080	2.427	$0.140^{(3)}$	$20.54^{(3)}$	19.30 ⁽²⁾	20.59
HD 052089	239.830	-11.330	0.188	$0.010^{(3)}$	17.95 (24)	<17.66 ⁽²⁾	19.76
HD 052918	218.010	0.610	0.384	0.060 ⁽⁷⁾	20.35 (19)	14.78 (7)	20.35
HD 053367	223.710	-1.900	0.129	0.740 ⁽¹⁷⁾	21.32 (41)	$21.04^{(41)}$	21.63
HD 053975	225.680	-2.320	1.247	$0.220^{(3)}$	$21.15^{(3)}$	19.23 ⁽²⁾	21.16
HD 054662	224.170	-0.780	1.170	$0.350^{(3)}$	$21.38^{(3)}$	$20.00^{(2)}$	21.41
HD 055879	224 730	0 350	1 011	$0.120^{(8)}$	$20.85^{(8)}$	<18 90 (2)	20.84
HD 057060	237.820	-5 370	1 477	$0.120^{(3)}$	20.81 ⁽⁸⁾	15 78 ⁽²⁾	20.81
HD 057061	238 180	-5540	1 514	$0.130^{(35)}$	20.01 $20.74^{(8)}$	15.48 ⁽²⁾	20.01
HD 057682	224 420	2 630	1.311	$0.120^{(3)}$	20.96 ⁽⁸⁾	<18 95 (2)	20.71
HD 058510	235 520	-2470	3 333	0.120	21 23 ⁽¹⁹⁾	$20.23^{(43)}$	21.31
HD 062542	255,920	-9.240	0 390	0.370 ⁽²⁷⁾	$20.90^{(27)}$	20.23 20.81 ⁽²⁷⁾	21.31
HD 063005	242 470	-0.930	13 699	0.300 (45)	20.90 21 24 ⁽³²⁾	20.31 (32)	21.32
HD 064740	263 380	-11 190	0 214	$0.000^{(45)}$	21.24 20.26 ⁽⁷⁾	14 95 (7)	20.26
HD 064760	262,060	-10.420	0.363	0.050 (45)	20.26 ⁽⁷⁾	$\sim 14.00^{(7)}$	20.20
HD 065575	266.680	-12320	0.139	$0.020^{(7)}$	~ 20.20	<14.00	20.40
HD 065818	263.480	-10.280	1 117	0.020	20.74	15 08 (7)	20.00
HD 066788	205.400	2 050	5 747	$0.000^{(41)}$	20.34 21 23 ⁽⁴¹⁾	19.00 19.72 ⁽⁴¹⁾	20.34
HD 066811	255 980	-4 710	0.668	0.200	10 05 (18)	17.72 14.45 ⁽²⁾	10.05
HD 068273	262,800	-7.690	0.008	0.040 (10)	19.95 10 78 (10)	14.43 14.23 ⁽²⁾	19.95
UD 060106	254 520	1.330	1 202	0.040	$21.06^{(42)}$	10.73 (41)	21.10
HD 072754	254.520	-1.330 -5.820	1.292	$0.180^{(8)}$	21.00 21 18 (32)	20 35 (32)	21.10
HD 072194	200.850	-5.820	0.131	$0.300^{(48)}$	21.10	$20.33^{(48)}$	21.29
HD 073882	245.090	0.640	0.131	$0.710^{(17)}$	21 11 (27)	20.94	21.01
HD 07/375	200.180	-10.860	0.347	0.120	21.11 20.82 ⁽³⁾	$\sim 18 34^{(2)}$	21.59
HD 074575	275.820	-10.800	0.330	$0.100^{(7)}$	20.62	$<15.04^{(7)}$	20.70
HD 074373	254.990	2.610	1 3 2 6	$0.070^{(45)}$	20.00	$20.30^{(38)}$	20.01
HD 074020	205.740	-2.010	2.874	0.230 (45)	- 21 15 (18)	20.30^{-10}	21.10
HD 075300	205.290	-1.930	2.074	$0.280^{(32)}$	21.13 (32)	$20.20^{(32)}$	21.23
HD 070186	205.800	-1.900	2.041	$0.270^{(32)}$	$21.06^{(32)}$	$20.20^{(32)}$	21.18
HD 079351	207.500	-7 370	0.151	$0.300^{(7)}$	20.78 (7)	$\sim 17.00^{(7)}$	20.37
HD 080077	271.630	-7.370	0.151	1 520 (16)	20.78	$21 40^{(16)}$	20.37
UD 080077	271.030	-0.070	2.551	0.050 (10)	$20.48^{(7)}$	$\sim 17.70^{(7)}$	21.95
HD 087001	275.880	-3.340	0.102	0.030	20.46 ···	$<11.70^{(2)}$	20.40
HD 08/901	220.430	40.930	0.020	$0.100^{(45)}$	< 10.00 (18)	$(14.96)^{(29)}$	20.70
HD 000087	265.520	-3.330	4.000	$0.120^{(26)}$	$21.00^{(18)}$	19.30	21.02
HD 090087	265.100	-2.130	5.205	0.280 (3)	21.13 (3)	19.92	21.20
HD 091510	234.890	32.770	0.303	$0.080^{(45)}$	$20.20^{(8)}$	10.70 (41)	20.20
HD 091397	280.800	-2.370	8.090	$0.300^{(18)}$	$21.34^{(1)}$	$19.70^{(41)}$	21.50
UD 001024	200.330	-1.720	1.934	$0.300^{(32)}$	21.13 (32)	10.07 (32)	21.10 21.14
HD 091824	203.700	0.070	2.331 1 255	$0.270^{(32)}$	21.12 (32)	20.14 (32)	21.10 21.24
пD 091983	203.000	0.050	4.233	$0.200^{(32)}$	$21.17^{(02)}$	$20.14^{(02)}$	21.24
пD 092334	287.000	-2.020	4.38/	$0.340^{(10)}$	21.28 (10) 21.20 (3)	10.93	21.28
HD 092/40	287.170	-0.850	2.332	0.330 (21)	21.20	19.97	21.23
пD 092809	280.780	-0.030	2.801	$0.220^{(21)}$	-	20.23 (30)	21.11
HD 093030	209.000	-4.900	0.207	$0.000^{(0)}$	20.28 (18)	<17.03 (38)	20.54
HD 093102	207.510	-0./10	2.101	0.020 (8)	21.33 (10) 21.40 (8)	19.63 (44)	21.37
пD 093204	287.370	-0./10	2.221	0.420	21.40	19.77	21.42

Table A.1. continued.

Source ID	Longitude [°]	Latitude [°]	Distance [kpc]	E(B-V)	$\log_{10}(N(H))$	$\log_{10}(N(H_2))$	$\log_{10}(N_{\rm H})$
HD 093205	287.570	-0.710	2.688	0.370 (18)	21.34 (8)	19.83 (43)	21.37
HD 093206	287.670	-0.940	1.101	0.330 (45)	21.34 (18)	19.52 (44)	21.35
HD 093222	287.740	-1.020	2.941	0.370 (26)	21.40 (29)	19.84 (43)	21.42
HD 093237	297.180	-18.390	0.318	0.090 (28)	_	19.80 (43)	20.72
HD 093521	183.140	62.150	1.949	0.050 (45)	20.15 (8)	<18.54 (2)	20.46
HD 093840	282.140	11.100	3.521	0.160 (5)	21.04 (18)	19.28 (43)	21.05
HD 093843	288.240	-0.900	2.625	0.340 (8)	21.33 (18)	19.61 (41)	21.35
HD 094454	295.690	-14.730	0.267	0.180 (28)	_	20.70 (42)	21.02
HD 094473	272.830	29.170	0.387	0.140 ⁽²⁵⁾	20.90 (25)	19.06 (25)	20.91
HD 094493	289.010	-1.180	1.852	0.200 (18)	$21.11^{(18)}$	20.12 (38)	21.19
HD 096675	296.620	-14.570	0.163	$0.310^{(27)}$	20.66 (27)	$20.82^{(27)}$	21.25
HD 099171	286.330	17.380	0.555	$0.050^{(3)}$	$20.65^{(3)}$	$15.25^{(2)}$	20.65
HD 099857	294,780	-4.940	2.326	0.330 ⁽¹⁸⁾	$21.31^{(18)}$	$20.25^{(41)}$	21.38
HD 099872	296 690	-10.620	0.230	0.360 (43)		$20.55^{(42)}$	21.30
HD 099890	291 750	4 430	1957	$0.150^{(45)}$	20.85 (19)	19 47 ⁽⁴¹⁾	20.88
HD 101131	294 780	-1.620	2 632	0.130 $0.280^{(45)}$		20 27 (44)	21.21
HD 101191	294.780	-1.490	3 367	0.200 (45)	21 04 (8)	20.27 20.42 (44)	21.21
HD 101/13	205.030	-1 710	1 887	$0.320^{(45)}$	21.04 21.23 (18)	20.42 20.38 (44)	21.21
UD 101415	295.030	-1.710	3.067	0.320	21.23 21.23 (18)	20.38 (44)	21.34
HD 101450	295.040	-1.710	0.104	0.310	21.23 (27)	$20.58^{(27)}$	21.34
HD 102003	206.850	-18.000	0.194	$0.170^{(18)}$	20.34	$20.30^{(41)}$	20.99
HD 103779	290.830	-1.020	2.301	$0.210^{(18)}$	$21.10^{(18)}$	19.62	21.20
HD 104/03	297.430	-0.540	2.313	$0.220^{(45)}$	21.11 (7)	19.98	21.17
HD 100490	298.230	5.790	0.080	0.020 (5)	20.04	$< 14.08^{(43)}$	20.00
HD 100943	298.960	1.140	0.355	$0.145^{(0)}$	_	$19.81^{(43)}$	20.92
HD 108002	300.100	-2.480	2.770	$0.310^{(0)}$	-	20.34 (12)	21.20
HD 108248	300.130	-0.360	0.114	$0.030^{(1)}$	19.00	$< 14.18^{(1)}$	20.24
HD 108610	300.280	0.880	0.503	$0.155^{(0)}$	-	19.86 (13)	20.95
HD 108639	300.220	1.950	1.825	0.250 (45)	$21.35^{(+2)}$	19.95 (+2)	21.38
HD 108927	301.920	-15.360	0.341	$0.220^{(40)}$	$20.86^{(27)}$	$20.49^{(27)}$	21.13
HD 109399	301.710	-9.880	2.755	$0.260^{(13)}$	$21.04^{(17)}$	$20.04^{(41)}$	21.12
HD 110432	301.960	-0.200	0.420	$0.520^{(12)}$	20.85 (27)	$20.64^{(27)}$	21.20
HD 110434	302.070	-3.600	0.423	$0.050^{(20)}$	_ (%)	19.90 ⁽⁴³⁾	20.46
HD 112244	303.550	6.030	1.167	$0.340^{(3)}$	21.11 (8)	$20.14^{(2)}$	21.19
HD 112999	304.170	2.180	0.747	$0.161^{(3)}$	-	$19.99^{(42)}$	20.97
HD 113904	304.670	-2.490	2.786	$0.290^{(3)}$	21.08 (3)	19.83 ⁽²⁾	21.13
HD 114886	305.520	-0.830	1.045	$0.400^{(9)}$	$21.34^{(42)}$	$20.34^{(43)}$	21.42
HD 115071	305.760	0.150	2.101	0.490 (18)	21.36 (42)	$20.63^{(42)}$	21.50
HD 115455	306.060	0.220	2.268	$0.400^{(45)}$	21.41 (18)	20.58 (43)	21.52
HD 116538	308.230	10.680	1.675	0.130 (45)	21.04 (18)	19.63 (38)	21.07
HD 116658	316.000	51.000	0.084	$0.030^{(3)}$	18.83 (13)	12.95 (2)	18.83
HD 116781	307.050	-0.070	2.045	0.340 (41)	21.18 (41)	20.08 (41)	21.24
HD 116852	304.880	-16.130	22.727	$0.210^{(31)}$	$20.96^{(32)}$	19.83 (43)	21.02
HD 118716	310.190	8.720	0.168	0.040 (10)	19.90 (10)	<14.08 (7)	20.37
HD 120315	100.700	65.320	0.030	$0.080^{(45)}$	<20.90 ⁽⁸⁾	13.38 (2)	20.67
HD 120324	314.240	19.120	0.119	0.100 (7)	20.40 (7)	<14.78 (7)	20.76
HD 121263	314.070	14.190	0.120	$0.020^{(45)}$	19.28 ⁽¹⁹⁾	12.80 ⁽²⁾	19.28
HD 121968	333.970	55.840	3.425	0.090 (26)	20.71 (18)	18.70 ⁽³⁹⁾	20.72
HD 122451	311.770	1.250	0.160	0.060 (45)	19.52 ⁽³⁾	12.80 (2)	19.52
HD 122879	312.260	1.790	2.387	0.298 (35)	21.26 (32)	20.31 (42)	21.35
HD 124314	312.670	-0.420	1.808	0.530 (18)	21.39 (42)	20.52 (43)	21.49
HD 127972	322.770	16.670	0.095	0.050 (7)	20.11 (7)	<14.18 ⁽⁷⁾	20.46
HD 135591	320.130	-2.640	0.835	0.220 (3)	21.08 (3)	19.77 (2)	21.12
HD 135742	352.020	39.230	0.093	_	19.38 (19)	14.34 (6)	19.38
HD 136298	331.320	13.820	0.115	0.020 (10)	20.18 (7)	<14.26 ⁽⁷⁾	20.06
HD 137595	336.720	18.860	0.822	0.250 (5)	21.00 (42)	20.56 (42)	21.24
HD 138690	333.190	11.890	0.137	0.030 (7)	20.23 (7)	<14.26 ⁽⁷⁾	20.24

Table A.1. continued.

Source ID	Longitude [°]	Latitude [°]	Distance [kpc]	E(B-V)	$\log_{10}(N(H))$	$\log_{10}(N(\mathrm{H}_2))$	$\log_{10}(N_{\rm H})$
HD 140037	340.150	18.040	0.402	0.090 (28)	_	19.34 (43)	20.72
HD 141637	346.100	21.710	0.145	0.200 (2)	21.18 (18)	19.23 (2)	21.19
HD 143018	347.210	20.230	0.580	0.070 (8)	20.74 (8)	19.32 (2)	20.77
HD 143275	350.100	22.490	0.155	0.190 (8)	$21.15^{(10)}$	19.41 ⁽²⁾	21.17
HD 144217	353 190	23 600	0.161	$0.210^{(35)}$	21.09 (10)	19 83 ⁽²⁾	21.13
HD 144470	352 750	22 770	0.142	$0.220^{(10)}$	21.09	$20.04^{(2)}$	21.13
HD 144965	339.040	8 420	0.142	0.220 0.350 ⁽²⁸⁾	21.10 21.07 ⁽⁴²⁾	20.04 20.77 ⁽⁴²⁾	21.24
HD 145502	354 610	22 700	0.135	0.330	21.07 21.15 ⁽³⁾	10 80 (2)	21.37
HD 147165	351 310	17,000	0.100	0.270	21.13 21.28 (18)	10.70 ⁽²⁾	21.20
HD 147103	351.510	17.000	0.100	$0.400^{(16)}$	21.30	20.78 (16)	21.40
ПD 147545	244.960	17.030	0.181	0.040	21.45	20.78 (42)	21.39
HD 147085	344.800	10.090	0.293	0.390	21.41 (16)	20.08 (16)	21.55
HD 147700	352.250	16.850	0.140	$0.740^{(10)}$	$21.50^{(10)}$	$20.90^{(10)}$	21.68
HD 14/888	353.650	17.710	0.092	$0.520^{(32)}$	$21.71^{(32)}$	$20.57^{(32)}$	21.77
HD 147889	352.860	17.040	0.139	$1.090^{(10)}$	21.46 (10)	$21.37^{(10)}$	21.88
HD 147933	353.690	17.690	0.174	$0.470^{(3)}$	21.81 (10)	$20.57^{(2)}$	21.86
HD 148184	357.930	20.680	0.122	0.530 (5)	21.15 (5)	20.63 (2)	21.36
HD 148379	337.250	1.580	3.012	0.720 (16)	-	20.41 (16)	21.62
HD 148422	329.920	-5.600	6.944	0.230 (45)	21.15 (26)	20.13 (44)	21.23
HD 148594	350.930	13.940	0.193	$0.210^{(32)}$	21.80 (32)	19.88 (32)	21.81
HD 148605	353.100	15.800	0.117	$0.100^{(10)}$	20.95 (10)	18.74 ⁽²⁾	20.96
HD 148937	336.370	-0.220	1.135	0.660 (18)	21.60 (18)	20.71 (40)	21.70
HD 149038	339.380	2.510	0.842	0.370 (3)	21.12 (18)	20.44 (2)	21.27
HD 149404	340.540	3.010	1.316	0.680 (8)	21.40 (8)	20.79 (46)	21.57
HD 149438	351.530	12.810	0.195	0.060 (3)	20.43 (8)	15.50 (2)	20.43
HD 149757	6.280	23.590	0.172	0.320 (3)	20.78 (8)	20.65 (2)	21.17
HD 149881	31.370	36.230	2.439	0.050 (45)	20.65 (3)	<19.00 ⁽²⁾	20.46
HD 150898	329,980	-8.470	0.882	0.110 (45)	$20.95^{(3)}$	19.81 ⁽²⁾	21.01
HD 151804	343 620	1 940	1 629	$0400^{(3)}$	$21.08^{(3)}$	$20.26^{(2)}$	21.19
HD 151805	343 200	1 590	1.672	0 190 (45)	21.00 $21.32^{(42)}$	20.36 (42)	21.19
HD 151800	346 120	3 910	0.268	0.050 ⁽⁷⁾	$20.40^{(7)}$	$< 14.26^{(7)}$	20.46
HD 152233	343 480	1 220	2 300	$0.000^{(45)}$	20.40 21.35 ⁽⁸⁾	20 29 (44)	20.40
HD 152235	343 030	0.870	2.500	0.400	21.55 21.77 (18)	20.27 20.73 (46)	21.42
HD 152230	242.050	1 120	1.403	$0.080^{(45)}$	21.77	$20.73^{(44)}$	21.04
HD 152248	244.090	1.180	1.098	0.420 (3)	-	20.29 (2)	21.39
HD 152408	244.080	1.490	2.242	$0.480^{(3)}$	$21.20^{(3)}$	$20.38^{(32)}$	21.50
HD 152590	344.840	1.830	1.03/	$0.380^{(3-)}$	$21.37^{(02)}$	$20.47^{(32)}$	21.47
HD 152623	344.620	1.610	1.500	$0.330^{(18)}$	$21.28^{(18)}$	$20.21^{(30)}$	21.35
HD 152723	344.810	1.610	16.667	$0.460^{(10)}$	21.43 (10)	$20.33^{(3)}$	21.49
HD 154368	349.970	3.220	1.217	$0.820^{(10)}$	$21.00^{(27)}$	$21.16^{(27)}$	21.59
HD 155806	352.590	2.870	0.994	0.230 (43)	21.08 (10)	19.92 (2)	21.14
HD 157246	334.640	-11.480	0.267	$0.050^{(8)}$	20.74 (18)	19.24 (2)	20.77
HD 157857	12.970	13.310	3.968	0.370 (43)	21.30 (18)	20.69 (43)	21.47
HD 158408	351.270	-1.840	0.134	$0.020^{(3)}$	<19.26 ⁽³⁾	<14.11 (2)	20.06
HD 158926	351.740	-2.210	0.220	0.080 (45)	19.23 (20)	12.70 (2)	19.23
HD 160578	301.040	-4.720	0.202	0.083 (35)	20.19 (18)	<14.23 (7)	20.68
HD 161807	351.780	-5.850	1.319	$0.140^{(45)}$	_	19.86 ⁽⁴⁴⁾	20.91
HD 163758	355.360	-6.100	3.876	0.350 (8)	21.23 (18)	19.85 ⁽³⁹⁾	21.26
HD 164284	30.990	13.370	0.143	0.190 (7)	20.82 (7)	19.85 ⁽⁷⁾	20.90
HD 164353	29.730	12.630	0.566	0.110 ⁽³⁾	21.00 (3)	20.26 (2)	21.13
HD 164402	7.160	-0.030	1.672	0.280 (3)	21.11 (3)	19.49 ⁽²⁾	21.13
HD 164740	5.970	-1.170	1.109	$0.870^{(46)}$	21.95 (46)	20.19 (41)	21.96
HD 164816	6.060	-1.200	1.185	0.310 (26)	21.18 (18)	20.00 (38)	21.23
HD 164906	6.050	-1.330	1.235	0.380 (45)	21.20 (26)	20.22 (38)	21.28
HD 165024	343 330	-13 820	0 279	0.060 (45)	$20.85^{(3)}$	18 95 (2)	20.86
HD 165052	6 120	-1 480	1 276	0 360 (45)	21 36 (18)	20 20 (38)	21.00
HD 165246	6 400	_1 5 60	1 996	-	21.50 21 41 ⁽⁴²⁾	20.15 (42)	21.42
HD 165055	357 /10	_7/30	1.205	0 120 (45)	21.11 (32)	16 53 ⁽³²⁾	21.40
105955	557.410	7.450	1.205	0.120	21,11	10.55	41,11

Table A.1. continued.

Source ID	Longitude [°]	Latitude [°]	Distance [kpc]	E(B-V)	$\log_{10}(N(\mathrm{H}))$	$\log_{10}(N(\mathrm{H}_2))$	$\log_{10}(N_{\rm H})$
HD 167263	10.760	-1.580	2.079	0.310 (3)	21.08 (3)	20.18 (2)	21.18
HD 167264	10.460	-1.740	0.861	0.330 (18)	21.15 (10)	20.28 (2)	21.25
HD 167971	18.250	1.680	2.033	1.040 (27)	21.60 (27)	20.85 (27)	21.73
HD 168076	16.940	0.840	2.100	0.790 (27)	21.65 (18)	20.68 (27)	21.73
HD 168941	5.820	-6.310	2.488	0.240 (45)	21.11 (18)	20.10 (41)	21.19
HD 169454	17.540	-0.670	2.128	1.120 (16)	>19.95 (16)	21.16 (16)	21.81
HD 170740	21.060	-0.530	0.231	0.480 (46)	21.15 (18)	20.86 (27)	21.46
HD 175191	9.560	-12.440	0.070	0.050 (45)	<19.48 (3)	<14.00 ⁽²⁾	20.46
HD 177989	17 810	-11880	2.538	$0.250^{(18)}$	20.95 (18)	20 23 (39)	21.09
HD 179406	28 230	-8 310	0.283	0 499 (21)	21 23 ⁽⁴⁶⁾	20.23 (41)	21.05
HD 184915	31 770	-13290	0.466	$0.770^{(18)}$	20.85 (18)	20.73 (2)	21.05
HD 185418	53 600	-2 170	0.755	0.380 (45)	20.00 21.11 ⁽²⁷⁾	20.51	21.05
HD 186004	78 620	10.060	1 965	$0.130^{(45)}$	$20.90^{(3)}$	10 50 ⁽⁴¹⁾	20.94
HD 188200	80.000	10.000	1.705	0.130	$20.90^{(3)}$	$20.01^{(2)}$	20.94
HD 188/130	81 770	10.000	1.47	0.210	20.90 20.85 ⁽⁸⁾	10 05 (2)	20.05
UD 100018	72 650	2 070	1.147	0.140	20.85	10.05 (43)	20.95
HD 101765	72.030	2.070	1.955	0.400	21.40	20.27 (38)	21.45
IID 191703	73.430 61.570	6.450	1.045	0.430	$21.30^{(18)}$	20.27 (38)	21.00
HD 1910//	01.370	-0.430	1.401	$0.140^{(18)}$	$20.90^{(19)}$	$20.02^{(43)}$	21.00
HD 192055	85.550	1.490	2.232	$0.530^{(45)}$	$21.20^{(13)}$	$20.08^{(12)}$	21.41
HD 192639	74.900	1.480	2.597	$0.560^{(10)}$	$21.32^{(10)}$	$20.69^{(2)}$	21.49
HD 193322	/8.100	2.780	0.989	$0.400^{(3)}$	$21.15^{(0)}$	$20.08^{(2)}$	21.22
HD 193924	340.900	-35.190	0.056	$0.020^{(3)}$	$<19.30^{(3)}$	$<14.30^{(2)}$	20.06
HD 195965	85.710	5.000	0.861	$0.190^{(43)}$	$20.90^{(10)}$	$20.36^{(30)}$	21.10
HD 197512	87.890	4.630	1.664	$0.330^{(27)}$	$21.26^{(27)}$	$20.66^{(27)}$	21.44
HD 198478	85.750	1.490	1.176	0.439 (33)	$21.32^{(32)}$	20.87 (32)	21.55
HD 198781	99.940	12.610	0.935	0.350 (32)	20.91 (32)	20.48 (32)	21.15
HD 199579	85.700	-0.300	0.941	0.310 (45)	21.04 (8)	20.53 (27)	21.25
HD 200120	88.030	0.970	0.399	$0.180^{(3)}$	$20.26^{(3)}$	19.30 (10)	20.34
HD 200775	104.060	14.190	0.361	$0.570^{(23)}$	-	21.15 (43)	21.52
HD 201345	78.440	-9.540	3.195	0.191 (21)	20.87 (18)	19.43 (39)	20.90
HD 202347	88.220	-2.080	0.931	0.170 (9)	20.99 (41)	20.00 (38)	21.07
HD 202904	80.980	-10.050	0.187	0.130 (35)	20.68 (7)	19.15 ⁽⁷⁾	20.70
HD 203064	87.610	-3.840	0.587	0.320 (8)	21.00 ⁽³⁾	20.29 (2)	21.14
HD 203374	100.510	8.620	2.611	$0.600^{(18)}$	21.20 (42)	20.60 (42)	21.38
HD 203532	309.460	-31.740	0.292	0.280 (47)	21.27 (32)	20.64 (32)	21.44
HD 203938	90.560	-2.330	0.223	0.720 (27)	21.48 (27)	21.00 (27)	21.70
HD 204172	83.390	-9.960	1.927	0.170 (3)	21.00 (3)	19.60 (2)	21.03
HD 206165	102.270	7.250	0.746	0.470 (17)	-	20.78 (34)	21.44
HD 206267	99.290	3.740	1.117	0.510 (16)	21.30 (27)	20.86 (27)	21.54
HD 206773	99.800	3.620	0.958	0.440 (32)	21.09 (32)	20.44 (32)	21.25
HD 207198	103.140	6.990	1.025	0.590 (8)	21.34 (8)	20.83 (27)	21.55
HD 207308	103.110	6.820	1.026	0.520 (15)	21.20 (41)	20.76 (41)	21.44
HD 207538	101.600	4.670	0.838	0.640 (46)	21.34 (18)	20.91 (27)	21.58
HD 208266	102.710	4.980	0.911	0.520 (40)	_	$20.87^{(34)}$	21.48
HD 208440	104.030	6.440	0.829	$0.290^{(31)}$	21.23 (18)	$20.34^{(34)}$	21.33
HD 208905	103 530	5 170	1.030	$0.290^{(5)}$		20.31 (43)	21.33
HD 209339	104 580	5 870	0.845	0 380 (15)	21 16 (41)	20.13 20.21 ⁽³⁸⁾	21.33
HD 209381	101.010	2 180	1 101	0.370 ⁽⁵⁾	21.10 21.11 ⁽¹⁸⁾	20.21 20.54 (43)	21.20
HD 200833	84 400	-21 270	0.008	0.089 (21)		<13.68 (48)	20.10
HD 200053	350 000	_52 470	0.020	0.060 (45)	10 04 (19)	<13.68 (2)	20.10
HD 209952	104 970	-52.470	0.050	0.000 (18)	21 11 (10)	20.15 (43)	20.34
11D 2099/3	56 000	J.390 11 160	0.030	0.300 (27)	21.11 (27)	20.15 (27)	21.20 21.10
HD 210121	JU.88U	-44.400	0.342	$0.380^{(-1)}$	$20.03^{(-1)}$	20.73 (2)	21.19
пD 210191	57.130	-31.700	0.002	$0.070^{(3)}$	$20.70^{(3)}$	<10.00	20.01
HD 210809	99.850	-3.130	4.529	$0.550^{(52)}$	21.23 (32)	$20.00^{(32)}$	21.30
HD 210839	103.830	2.610	0.617	$0.570^{(3)}$	$21.11^{(10)}$	$20.84^{(27)}$	21.43
HD 212791	101.640	-4.300	0.998	$0.050^{(32)}$	21.21 (32)	19.42 (32)	21.22

Table A.1. continued.

	X 1 1 103	T			1 ()////		1 () ()
Source ID	Longitude [°]	Latitude [°]	Distance [kpc]	E(B-V)	$\log_{10}(N(H))$	$\log_{10}(N(H_2))$	$\log_{10}(N_{\rm H})$
HD 214080	44.810	-56.920	1.203	$0.050^{(8)}$	$20.60^{(8)}$	<19.00 ⁽²⁾	20.46
HD 214680	96 650	-16980	0 359	$0.110^{(3)}$	$20.70^{(3)}$	$19\ 22\ ^{(2)}$	20.73
HD 214993	97.650	-16180	0.321	0.110 ⁽⁷⁾	20 79 (7)	19 63 (7)	20.85
HD 216532	109 650	2 680	0.751	0.860 (17)	20.75	21 10 (34)	20.00
HD 216808	109.030	2.000	0.840	0.850 (17)	_	21.10	21.70
HD 210090	110 250	2.390	0.840	0.850	21.46(18)	21.05 20.05 (34)	21.09
HD 217055	110.230	2.800	0.829	$0.700^{(17)}$	21.40 (18)	$20.93^{(34)}$	21.07
HD 21/312	110.560	2.950	1.631	$0.660^{(1)}$	21.48 (10)	$20.80^{(31)}$	21.63
HD 21/615	332.370	-56.680	0.279	$0.247^{(21)}$	-	19.67 (10)	20.94
HD 21/6/5	102.210	-16.100	0.109	0.050 (10)	-	$19.67^{(2)}$	20.46
HD 218376	109.950	-0.780	0.374	$0.220^{(3)}$	20.95 (3)	20.15 (2)	21.07
HD 218915	108.060	-6.890	8.065	0.210 (45)	21.20 (8)	20.15 (39)	21.27
HD 219188	83.030	-50.170	2.268	$0.080^{(10)}$	20.85 (3)	19.34 ⁽²⁾	20.88
HD 220057	112.130	0.210	0.392	$0.270^{(32)}$	21.17 (32)	20.28 (32)	21.27
HD 224151	115.440	-4.640	1.898	0.420 (44)	21.32 (18)	20.57 (41)	21.45
HD 224572	115.550	-6.360	0.292	0.200 (4)	20.88 (3)	20.23 (2)	21.04
HD 232522	130.700	-6.710	11.905	0.180 (32)	21.08 (32)	20.22 (32)	21.19
HD 303308	287.590	-0.610	2.457	0.430 (44)	21.40 (8)	20.24 (44)	21.46
HD 308813	294,790	-1.610	5,291	$0.260^{(45)}$	21.15 (18)	20.29 (44)	21.26
HD 315021	6120	-1340	1 292	$0.310^{(18)}$	21 28 (18)	19 99 (38)	21.32
HE 0226-4110	253 940	-65 780	-	$0.016^{(21)}$	>19 50 (36)	$-14.29^{(36)}$	19.97
HE 11/3 1810	281 850	41 710	_	0.030 (36)	20 47 (36)	16 54 (36)	20.47
HE 1143-1010	201.030	41.710	-	0.009 (21)	20.47	10.54	20.47
HS 0024+0907	145.710	25.550	-	0.098	20.80 (36)	19.82 (36)	20.88
MRC 2251-178	46.200	-61.330	-	0.039 (30)	$20.39^{(36)}$	14.54 (36)	20.39
Mrk 0009	158.360	28.750	-	0.059 (21)	20.64 (36)	19.36 (36)	20.68
Mrk 0106	161.140	42.880	-	0.028 (21)	20.35 (30)	16.23 (30)	20.35
Mrk 0116	160.530	44.840	-	0.032 (36)	20.41 (36)	19.08 (36)	20.45
Mrk 0205	125.450	41.670	-	0.042 (36)	$20.40^{(36)}$	16.53 (36)	20.40
Mrk 0209	134.150	68.080	-	0.015 (36)	>19.73 (36)	<14.48 (36)	19.94
Mrk 0290	91.490	47.950	-	0.015 (36)	20.11 (36)	16.18 ⁽³⁶⁾	20.11
Mrk 0335	108.760	-41.420	-	0.035 (21)	20.43 (36)	18.83 ⁽³⁶⁾	20.45
Mrk 0421	179.830	65.030	_	0.013 (21)	19.94 ⁽³⁶⁾	14.63 (36)	19.94
Mrk 0478	59.240	65.030	_	0.013 (21)	>19.21 (36)	<14.56 (36)	19.88
Mrk 0501	63.600	38.860	_	0.019 (21)	20.24 (36)	14.78 (36)	20.24
Mrk 0509	35.970	-29.860	_	0.057 (21)	20.58 (36)	17.87 (36)	20.58
Mrk 0817	100 300	53 480	_	0.007 (36)	>19.83 (36)	<14 03 (36)	19.61
Mrk 0876	98 270	40 380	_	0.007 (21)	20.36 (36)	16 64 (36)	20.36
Mrk 1005	201.600	-21 130	_	0.027 0.128 ⁽²¹⁾	20.50 20.95 ⁽³⁶⁾	18 76 ⁽³⁶⁾	20.50
Mrlz 1292	201.090	-21.130	_	0.120	20.95	14 25 (36)	20.90
Mrl- 1512	549.220	20.070	-	0.032	20.40	14.55	20.40
MIK 1515	05.070	-29.070	-	0.044	20.32 (36)	10.42 (36)	20.32
MS 0/00.7+6338	152.470	25.630	-	$0.051^{(21)}$	20.43 (36)	18.75 (36)	20.45
NGC 0985	180.840	-59.490	-	$0.033^{(21)}$	$20.52^{(30)}$	16.05 (36)	20.52
NGC 1068	172.100	-51.900	-	0.034 (21)	19.61 (30)	18.13 (30)	19.64
NGC 1705	261.080	-38.740	-	$0.008^{(21)}$	>19.66 (30)	<14.17 (36)	19.67
NGC 4151	155.080	75.060	-	0.028 (21)	20.20 (36)	16.70 (36)	20.20
NGC 4670	212.690	88.630	-	0.015 (21)	19.95 ⁽³⁶⁾	14.72 (36)	19.95
NGC 7469	83.100	-45.470	-	0.069 (21)	20.59 ⁽³⁶⁾	19.67 ⁽³⁶⁾	20.68
PG 0804+761	138.280	31.030	-	0.035 (21)	20.54 (36)	18.66 (36)	20.55
PG 0844+349	188.560	37.970	_	0.037 (21)	20.34 (36)	18.22 (36)	20.35
PG 0953+414	179.790	51.710	-	0.013 (21)	20.00 (36)	15.03 (36)	20.00
PG 1116+215	223.360	68.210	-	0.023 (21)	>19.70 (36)	<14.16 (36)	20.13
PG 1211+143	267.550	74.320	_	0.035 (21)	20.25 (36)	18.38 (36)	20.26
PG 1250±503	120 560	58 050	_	0.008 (21)	19 67 ⁽³⁶⁾	14 75 (36)	19.67
PG 1207 107	308 500	52 160	_	0.043 (21)	20 12 (36)	15 62 (36)	20.42
DKS 0405 12	204.020	_/1 760	—	0.059 (21)	20.42 (36)	15 70 (36)	20.42
T KS 0403-12 DVC 0550 504	204.930 257.060	-41.700	-	0.030	20.41	15 14 (36)	20.41
FKS UJJ8-304	251.900	-28.370	-	0.044 (21)	20.35 (36)	15.44 (36)	20.33
PKS 2003-489	550.570	-32.000	—	0.050 (21)	20.00 (30)	15.07 (30)	20.60

E. Bellomi et al.: 3D chemical structure of diffuse turbulent ISM. I.

Table A.1. con	tinued.
----------------	---------

Source ID	Longitude [°]	Latitude [°]	Distance [kpc]	E(B-V)	$\log_{10}(N(\mathrm{H}))$	$\log_{10}(N(H_2))$	$\log_{10}(N_{\rm H})$
PKS 2155-304	17.730	-52.250	_	0.022 (21)	20.06 (36)	14.17 (36)	20.06
QSO B1226+0219	289.950	64.360	_	0.021 (21)	20.22 (36)	15.74 ⁽³⁶⁾	20.22
QSO J1104+7658	130.390	38.550	_	0.034 (21)	20.25 (36)	18.98 ⁽³⁶⁾	20.29
QSO J1821+6420	94.000	27.420	_	0.043 (21)	20.43 (36)	17.91 (36)	20.43
Ton S180	138.990	-85.070	_	0.014 (21)	>20.08 (36)	<14.37 (36)	19.91
Ton S210	224.970	-83.160	-	0.017 (21)	20.19 (36)	16.57 ⁽³⁶⁾	20.19
TY CrA	359.990	-17.780	0.136	0.480 (16)	-	21.10 (16)	21.44
VII Zw 118	151.360	25.990	-	0.038 (21)	20.56 (36)	18.84 ⁽³⁶⁾	20.58
WD 0004+330	112.480	-28.690	0.097	0.049 (21)	19.68 ⁽¹⁹⁾	14.46 (30)	19.68
WD 1636+351	56.980	41.400	0.145	0.026 (21)	19.57 ⁽¹⁹⁾	15.05 (30)	19.57
WD 1800+685	98.730	29.780	0.159	0.054 (21)	18.86 ⁽³³⁾	14.75 ⁽³⁰⁾	18.86
WD 2247+583	107.640	-0.640	0.107	1.310 (21)	19.89 (22)	15.11 (30)	19.89

Appendix B: Heating and cooling equations

The analytical equations of the heating and cooling rates used in this work are taken from Wolfire et al. (1995, 2003).

The electronic density is calculated from the ionization equilibrium

$$n_{\rm e} = 2.4 \times 10^{-3} \zeta_{\rm CR}^{1/2} \left(\frac{T}{100 \text{ K}}\right)^{0.25} \frac{G_{\rm eff}^{1/2}}{\phi} + n_{\rm H} x_{\rm C^+} \text{ cm}^{-3}, \qquad (B.1)$$

where ζ_{CR} is the total ionization rate (including primary and secondary ionizations) of H by soft X-rays and cosmic ray particles (expressed in unit of 10^{-16} s⁻¹) and G_{eff} is the local effective radiation field in Habing units (see Sect. 3.2). ϕ_{PAH} is a recombination parameter of electrons on PAHs, discussed in Wolfire et al. (2003), and set to 0.5. x_{C^+} is the abundance of C⁺ relative to $n_{\rm H}$, $x_{C^+} = n(C^+)/n_{\rm H}$. Throughout this work, we adopt $x_{C^+} = 1.4 \times 10^{-4}$, which corresponds to the value derived in the Solar Neighborhood assuming 40% depletion of carbon onto grains and that the remaining carbon is singly ionized. This equation for the density of electrons differs from that of Wolfire et al. (2003) by the addition of C⁺ which is the most abundant ion in the diffuse and transluscent CNM (Snow & McCall 2006).

Following Wolfire et al. (2003), we include the heating induced by the photoelectric effect on grains and PAHs and by cosmic ray ionizations. The former is modeled with a rate

$$\Gamma_{\rm ph} = 1.3 \times 10^{-24} \,\epsilon \, G_{\rm eff} \,\, {\rm erg \ s^{-1}},$$
(B.2)

where the heating efficiency

$$\epsilon = \frac{4.9 \times 10^{-2}}{1.0 + [\kappa/1925]^{0.73}} + \frac{3.7 \times 10^{-2} (T/10^4 \text{ K})^{0.7}}{1.0 + [\kappa/5000]}$$
(B.3)

and

$$\kappa = \frac{G_{\rm eff} T^{1/2}}{n_{\rm e} \phi_{\rm PAH}}.$$
(B.4)

The latter is modeled with a rate

$$\Gamma_{\rm CR} \sim 10^{-27} \left(\frac{\zeta_{\rm CR}}{10^{-16} \, {\rm s}^{-1}} \right) \, {\rm erg} \, {\rm s}^{-1}.$$
 (B.5)

Regarding the cooling, we include the fine-structure emission of CII and OI, the emission of Lyman α photons by HI, and



Fig. B.1. Thermal equilibrium curves ($\mathcal{L} = 0$) computed with RAMSES for $A_V = 0$ (red solid curve) and $A_V = 1$ (green solid curve) compared to those predicted by Wolfire et al. (2003) for $\phi_{PAH} = 0.5$ (black dashed curve) and 1.0 (blue dashed curve).

the recombination of electrons onto charged grains and PAHs. The cooling rate due to collisional excitation of the fine-structure levels of C^+ by atomic hydrogen and electrons is given by

$$\Lambda_{\rm CII} = \left[2.25 \times 10^{-23} + 10^{-20} \left(\frac{T}{100 \text{ K}} \right)^{-0.5} \frac{n_{\rm e}}{n_{\rm H}} \right] \\ \times \text{ e}^{-92/T} x_{\rm C^+} \text{ erg cm}^3 \text{ s}^{-1}.$$
(B.6)

The cooling rate by collisional excitation of the fine-structure level of OI by atomic hydrogen is computed as

$$\Lambda_{\rm OI} = 7.81 \times 10^{-24} \left(\frac{T}{100 \text{ K}}\right)^{0.4} \text{ e}^{-228/T} x_{\rm O} \text{ erg cm}^3 \text{ s}^{-1}, \qquad (B.7)$$

where $x_{\rm O} = n({\rm O})/n_{\rm H}$ is the relative abundance of atomic oxygen. Throughout this paper, we adopt $x_{\rm O} = 3.2 \times 10^{-4}$, which corresponds to the value derived in the Solar Neighborhood assuming 37% depletion of oxygen onto grains and that the remaining oxygen is in its atomic form. Those two lines are the dominant cooling terms of the CNM phase. The cooling induced by the excitation of the Lyman α line, which is the dominant cooling at $T \gtrsim 8000$ K (WNM), is taken from Spitzer (1978)

$$\Lambda_{\rm HI} = 7.3 \times 10^{-19} x_{\rm e} \ {\rm e}^{-118400/T} \ {\rm erg} \ {\rm cm}^3 \ {\rm s}^{-1}. \tag{B.8}$$

Finally, the cooling rate due to electron recombination onto charged grains and PAHs is set to

$$\Lambda_{\rm rec} = 4.65 \times 10^{-30} T^{0.94} \kappa^{\beta} x_e \phi_{\rm PAH} \,\,{\rm erg} \,\,{\rm cm}^3 \,\,{\rm s}^{-1}, \tag{B.9}$$

with $\beta = 0.74/T^{0.068}$. To validate the calculations of the heating and cooling terms, we compare in Fig. B.1 the thermal equilibrium curve obtained with RAMSES to the predictions of Wolfire et al. (2003).

Appendix C: Analytical description of 1D and 2D probability histograms

In order to interpret the results found in Sect. 4, we propose a semi-analytical prescription to predict the 1D and 2D PHs of the total column density and the column density of H_2 obtained with numerical simulations. This prescription is based on the work of Vázquez-Semadeni & García (2001) and Bialy et al. (2017, 2019) who showed that lines of sight across isothermal simulations of turbulence can be accurately modeled as a series of random density fluctuations.

C.1. Decorrelation scales

In Fig. C.1, we display an example of the volume-weighted distribution of the proton density computed for the fiducial simulation (see Table 2). The total column density integrated along the *x* direction over a random line of sight of size $l \le L$ is

$$N_{\rm H}(l) = \int_{O}^{l} n_{\rm H} dx = \sum_{i=i_0}^{i_l=i_0+\frac{l}{L}R^{1/3}} n_{\rm H}(i) dx, \qquad (C.1)$$

where *R* is the resolution of the box of size *L* (see Table 2), i_0 is a random starting index for the integration of the column densities, and i_l is the final index deduced from *l*. Because of spatial correlations of the density $n_{\rm H}$, this computation is not equivalent to the sum of random realizations of $n_{\rm H}$ drawn out of the 1D probability distribution (Fig. C.1). It depends, instead, on how and over which distance the values of $n_{\rm H}$ are correlated. As proposed by Bialy et al. (2017, 2019), and since the correlations of density in a turbulent medium decrease over long distances, we assume that these correlations can be modeled with a parameter $y_{\rm dec}$, called the decorrelation scale. The density is supposed to be constant over distances smaller than $y_{\rm dec}$, and uncorrelated over larger distances.

In this framework, if y_{dec} is the same for all densities, the total column density integrated over a distance *l* becomes equivalent to the sum of $1 + l/y_{dec}$ random realizations of $n_{\rm H}$. In isothermal simulations, the 1D probability distribution of the gas density is found to follow a lognormal distribution with a dispersion proportional to the Mach number, dependent on the nature of the turbulent driving, and independent on the mean density of the gas (e.g. Padoan et al. 1997; Federrath et al. 2008). Because of this property, Bialy et al. (2017) were able to establish a relation between the dispersion $\sigma_{N_{\rm H}/l}$ of the distribution of the proton density:

$$\frac{\sigma_{N_{\rm H}/l}}{\sigma_{n_{\rm H}}} = \left(1 + \frac{L_{\rm drive}}{y_{\rm dec}} \frac{l}{L_{\rm drive}}\right)^{-1/2},\tag{C.2}$$

where L_{drive} is the turbulence driving scale. Fitting the $\sigma_{N_{\text{H}}/l}/\sigma_{n_{\text{H}}}$ ratio as a function of l/L_{drive} , Bialy et al. (2017) estimated $y_{\text{dec}} = 0.2 \times L_{\text{drive}}$.



Fig. C.1. Probability histogram of the proton density $n_{\rm H}$ extracted from the fiducial simulation (see Table 2). The black histogram correspond to the extracted data. The red dashed curve shows an example of the sum of two log-normal components and a power-law tail at high density, for comparison. The blue line indicates the inflection point of the PH between the diffuse and the dense components.

Unfortunately, this method cannot be applied here. Indeed, oppositely to isothermal simulations and as illustrated in Fig. C.1, the PH of $n_{\rm H}$ derived from simulations of the multiphase ISM is usually described by the sum two log-normal distributions plus a power-law tail at high density that could be a signature of the CNM which is known to behave like a polytropic gas with an exponent $\gamma < 1$ (Passot & Vázquez-Semadeni 1998) or a signature of gravity (Federrath 2013; Girichidis et al. 2014). To overcome this issue, and for the sake of simplicity, we assume that a two phase medium is described by two decorrelation scales. All densities below a limit $n_{\rm H}^{\rm lim}$ are supposed to belong to the diffuse log-normal component and to be correlated over a scale y_{dec}^{diff} . Similarly, all densities above n_{H}^{lim} are supposed to belong to the dense log-normal component and to be cor-related over a smaller scale y_{dec}^{dens} . We identified here the limit r_{h}^{lim} between the two log normal distributions with the inflac $n_{\rm H}^{\rm lim}$ between the two log-normal distributions with the inflection point of the 1D probability histogram of the gas density (blue line in Fig. C.1). Since the diffuse component behave like an isothermal gas at the temperature of the WNM, we assume that y_{dec}^{diff}/L_{drive} is constant for all simulations and adopt the value given by Bialy et al. (2017), $y_{dec}^{diff} = 0.2 \times L_{drive}$, where $L_{drive} = L/2$ is the main driving scale used for the turbulent forcing (see Sect. 3.3). In contrast, and because the diffuse component occupies most of the volume, we state that y_{dec}^{dens} depends on the total mass of the gas or equivalently its mean density $\overline{n_{\rm H}}$. To simplify, we propose that

$$y_{\rm dec}^{\rm dens} \propto \overline{n_{\rm H}}^{1/3},$$
 (C.3)

which means that the impact of changing the total mass of the gas is to change the typical volume of the dense structures by the same factor. Within this framework, the semi-analytical model proposed here therefore depends on a single parameter: the decorrelation scale of the dense component for the fiducial simulation. In the following, we assume $y_{dec}^{dens} = 10$ pc for $\overline{n_{\rm H}} = 2$ cm⁻³ which implies a decorrelation length of 6.3, 7.9, and 12.6 pc for $\overline{n_{\rm H}} = 0.5$, 1, and 4 cm⁻³, respectively.

It is quite optimistic to believe that the dense and cold component of the ISM can be modeled by a single decorrelation scale y_{dec}^{dens} . This component is indeed likely to follow a distribution



E. Bellomi et al.: 3D chemical structure of diffuse turbulent ISM. I.

Fig. C.2. Comparisons of the 1D probability histograms of the total column density normalized to the integration scale, $N_{\rm H}/l$, extracted from the simulations (black histograms) and constructed with the semi-analytical model described in the main text (green histograms) for l = 200 (*top left*), 100 (*top right*), 50 (*bottom left*), and 25 (*bottom right*) pc. Each of these four main panels displays the comparisons performed for 15 simulations with different $\overline{n_{\rm H}}$ and G_0 around the fiducial setup defined by $\overline{n_{\rm H}} = 2 \text{ cm}^{-3}$ and $G_0 = 1$. All probability histograms inferred from the semi-analytical model are obtained assuming a fixed correlation length of the diffuse component $y_{\rm dec}^{\rm diff} = 0.2 \times L_{\rm drive} = 20$ pc and a correlation length of the dense component $y_{\rm dec}^{\rm diff} = 10 \times (\overline{n_{\rm H}}/2\text{ cm}^{-3})^{1/3}$ pc (see main text).

of sizes which decrease with the local pressure, hence the local density $n_{\rm H}$. However, and as we show below, such an assumption allows us to highlight important features of the simulations. The essential impact of the distribution of sizes is proven and discussed in Appendix C.4.

C.2. Comparison with simulations

The goal of the model is to offer an explanation on how the PHs of local densities translate into PHs of column densities in a simulation of the multiphase ISM. To test its validity, we generate a series of N fictitious lines of sight of size l and compare the PHs to those obtained with an equivalent sample of lines of sight extracted from numerical simulations. For each line of sight, we draw a sequence of random realizations of $n_{\rm H}$ out of its known 1D PH using the rejection method. For each draw, the density is supposed to be constant over a distance $y_{\rm dec}^{\rm diff}$ if $n_{\rm H} < n_{\rm H}^{\rm lim}$, and over a distance $y_{\rm dec}^{\rm dens}$ otherwise⁵. The contribution of this piece of

gas to the total column density is therefore computed as $n_{\rm H}y_{\rm dec}^{\rm diff}$ or $n_{\rm H}y_{\rm dec}^{\rm dens}$. In contrast, the contribution of this piece to the column density of H₂ is inferred from the expected density profile of H₂ over a 1D slab of size $y_{\rm dec}^{\rm diff}$ or $y_{\rm dec}^{\rm dens}$. Throughout the slab, the density of H₂ is calculated at equilibrium taking into account local and large-scale extinction and self-shielding as

$$\left\langle e^{-\sigma_{\rm d}(N_{\rm H}^{\rm loc}+N_{\rm H}^{\rm ext})} \right\rangle \left\langle f_{\rm shield} \left(\frac{N^{\rm loc}({\rm H}_2)+N^{\rm ext}({\rm H}_2)}{5\times 10^{14} {\rm cm}^{-2}} \right) \right\rangle,\tag{C.4}$$

where $N_{\rm H}^{\rm loc}$ and $N^{\rm loc}({\rm H_2})$ are the local column densities of protons and H₂ computed from the border of the slab. The mean values in this expression are calculated from six random realizations of the large-scale column densities $N_{\rm H}^{\rm ext}$ and $N^{\rm ext}({\rm H_2})$ drawn out of the sample a lines of sight under construction. The construction of each line of sight ends when its size reaches the integration scale *l*. The entire sample of lines of sight is finally reconstructed until convergence of the large-scale shielding processes described above.

The comparisons between the 1D and 2D PHs reconstructed from the semi-analytical model and extracted from the simulations are shown in Figs. C.2 and C.3 for samples of $N = 64\,000$

⁵ To ensure that the resulting sample matches the original PH, the probability associated to each density is weighted by the inverse of the component size $1/y_{dec}^{diff}$ or $1/y_{dec}^{dens}$.



Fig. C.3. Comparisons of the 2D probability histograms of the total column density normalized to the integration scale, $N_{\rm H}/l$, and the column density of H₂ normalized to integration scale, $N({\rm H}_2)/l$ extracted from the simulations and constructed with the semi-analytical model described in the main text for l = 200 (top left), 100 (top right), 50 (bottom left), and 25 (bottom right) pc. The results of the simulations are indicated with contour plots with isoprobabilities of 10^{-4} , 10^{-3} , and 10^{-2} . The colored histograms correspond to the results obtained with the semi-analytical model. As in Fig. C.2, each of the four main panels displays the comparisons performed for 15 simulations with different $\overline{n_{\rm H}}$ and G_0 around the fiducial setup defined by $\overline{n_{\rm H}} = 2 \,\mathrm{cm}^{-3}$ and $G_0 = 1$. All probability histograms inferred from the semi-analytical model are obtained assuming a fixed correlation length of the diffuse component $y_{\rm dec}^{\rm dens} = 0.2 \times L_{\rm drive} = 20$ pc and a correlation length of the dense component $y_{\rm dec}^{\rm dens} = 10 \times (\overline{n_{\rm H}}/2 \,\mathrm{cm}^{-3})^{1/3}$ pc (see main text).

lines of sight, 15 different simulations, and different values of the integration scale *l*, from *L* down to y_{dec}^{diff} . Unexpectedly, setting the decorrelation length of the dense component to ~10 pc for the fiducial simulation ($\overline{n_{\rm H}} = 2 \text{ cm}^{-3}$) leads to a remarkable agreement between all the fictitious and actuals PHs. Once this parameter is set, the model not only reproduces surprisingly well the shapes of the 1D PHs and of the HI-to-H₂ transition, but also their global trends depending on $\overline{n_{\rm H}}$ and G_0 and their deformations depending on the chosen integration scale *l*. This result is not straightforward and highly depends on the decorrelation lengths y_{dec}^{diff} and y_{dec}^{dens} . The agreement observed in Figs. C.2 and C.3 therefore suggests that the model somehow captures an essential property of the simulations, namely some characteristic lengths of the diffuse and dense components of a multiphase gas with a turbulence driven at a scale L_{drive} .

C.3. Interpretation of the model

The statistics of the HI-to- H_2 transition derived from the model result from a combination of effects. Locally, the fraction of H_2

of a given slab depends on the density and the size of the slab y_{dec}^{diff} and y_{dec}^{dens} , which set the local self-shielding, and on the surrounding environment, which sets the large-scale self-shielding. How these local properties contribute to the integrated quantities $N_{\rm H}$ and $N({\rm H}_2)$ depends, in turn, on the sizes of the slabs and on the PH of the density which both control the reconstruction of the line of sight.

1. Since \tilde{y}_{dec}^{dens} is fixed, the HI-to-H₂ transition induced by the local self-shielding alone occurs in any slab with a density larger than

$$n_{\rm H}^{\rm tr} \propto G_0^{1/2} (y_{\rm dec}^{\rm dens})^{-1/2} \propto G_0^{1/2} (\overline{n_{\rm H}})^{-1/6}.$$
 (C.5)

This equation can be obtained from the expression of the column densities of HI envelopes (Sternberg et al. 2014, Eq. (40)) in the weak field limit. The large-scale self-shielding not only increases the fraction of H₂ in atomic slabs (i.e., for $n_{\rm H} < n_{\rm H}^{\rm tr}$) but also shifts $n_{\rm H}^{\rm tr}$ toward lower values, two effects which depend on the size of the box *L*. For *L* = 200 pc, including the large-scale self-shielding is found to reduce $n_{\rm H}^{\rm tr}$ by a factor of two.



Fig. C.4. Same as Fig. C.3, assuming that $y_{dec}^{dens} = 10 \times (\overline{n_H}/2 \text{ cm}^{-3})^{1/3} (n_H/n_H^{lim})^{-0.35}$ pc. The exponent used here corresponds to a simple test to show the effect of a dense decorrelation scale y_{dec}^{dens} varying with density.

2. Lines of sight with very low molecular fraction necessarily result from the combination of slabs with $n_{\rm H} < n_{\rm H}^{\rm tr}$. The occurrence of such events depends on the volume filling factor of the diffuse gas, hence on the 1D PH of low density material, and on the integration length *l*: as *l* increases, their likelihood decreases. Such a scenario occurs for a maximum normalized column density of

$$N_{\rm H}/l = n_{\rm H}^{\rm tr}.$$
 (C.6)

For $l \sim y_{dec}^{diff}$, such a high normalized column density of atomic gas is a likely event. As *l* increases, it becomes, however, unlikely to throw a line of sight composed of components with identical densities $n_{\rm H}^{\rm tr}$. Therefore, while the above limit is still valid, the maximum normalized column density of weakly molecular gas appears to decrease.

3. Lines of sight with large molecular fraction necessarily contain at least one slab with $n_{\rm H} > n_{\rm H}^{\rm tr}$. Oppositely to the previous case, the occurrence of such events depends on the volume filling factor of the dense gas, hence on the 1D PH of large density material, and on the integration length *l*: as *l* increases, their likelihood increases. Such a scenario occurs for a minimum normalized column density

$$N_{\rm H}/l = n_{\rm H}^{\rm tr} y_{\rm dec}^{\rm dens}/l. \tag{C.7}$$

4. These two limits for lines of sight with low and large f_{H_2} (items 2. and 3.), set the width of the HI-to-H₂ transition seen in

column densities. As shown in Fig. C.3, this width is somehow smaller than that obtained from the numerical simulations. We will discuss this point in the next section.

5. At last, lines of sight with intermediate H₂ fraction $(10^{-4} \le f_{H_2} \le 10^{-2})$ mostly result from a combination of slabs of low and moderate densities $(n_H \le n_H^t)$. Because such events are unlikely, the model predicts a small fraction of lines of sight at intermediate f_{H_2} , in contradiction with results extracted from the simulations. This point will also be discussed further in the next section.

All these properties fully explain the behaviors of the analytical model observed in Fig. C.3. The transition density for the fiducial simulation is $n_{\rm H}^{\rm tr} = 4 \text{ cm}^{-3}$. As expected, the corresponding lower and upper limits of $N_{\rm H}/l$ required to activate the HI-to-H₂ transition are in agreement with the limits found for the lowest integration scale l = 25 pc (bottom right panels of Fig. C.3). Increasing the integration length has three effects: (a) to squeeze the 2D PH along the *x* and *y* axis as both $N_{\rm H}$ and $N({\rm H}_2)$ progressively tend toward Gaussian distributions centered on the means, (b) to shift the HI-to-H₂ transition to lower $N_{\rm H}/l$ according to the limits derived above, and (c) to increase the occurrence of lines of sight with large $f_{\rm H_2}$ to the detriment of lines of sight with low $f_{\rm H_2}$ as the probability of intercepting a slab with $n_{\rm H} > n_{\rm H}^{\rm tr}$ rises. The dependence of the distributions on G_0 and $\overline{n_{\rm H}}$ are also straightforward. As G_0 increases or $\overline{n_{\rm H}}$ decreases, the HI-to-H₂ transition is shifted according to the dependence of

 $n_{\rm H}^{\rm tr}$ on these parameters. The occurrences of high or low molecular fraction lines of sight simply depend on the volume filling factor of the dense gas, and are therefore a direct consequence of the 1D PH of the gas density. As G_0 increases or $\overline{n_{\rm H}}$ decreases, the mean thermal pressure rises and the mass fraction of the CNM diminishes. This favors the occurrence of lines of sight with low molecular fraction.

C.4. Discrepancies and conclusions

Despite the surprising agreement between the simulated and the modeled PHs, in particular regarding the 1D PH of the total column density $N_{\rm H}$, Figs. C.2 and C.3 also reveal important discrepancies. Most notably, and as mentioned above, the modeled 2D PHs systematically underestimate the widths of the HI-to-H₂ transition and underestimate the proportion of lines of sight at intermediate integrated molecular fraction. These strong discrepancies are entirely due to the hypothesis of a constant decorrelation scale of the dense component.

Obviously, the dense and cold ISM are not characterized by a unique scale but a distribution of sizes which likely decrease with the gas pressure and local density. Such a distribution would increase the probability of occurrence of small components of high density along any line of sight (see footnote 5) and therefore solve the discrepancy between the analytical model and the simulations. Indeed, as schematized in Fig. 7, small and dense components surrounded by diffuse material favor the occurrence of lines of sight at intermediate molecular fractions. This configuration not only reduces the mean molecular fraction predicted by the model but also necessarily widens the HI-to-H₂ transition, in closer agreement with the simulations. To illustrate this point, we display in Fig. C.4 the 2D PHs reconstructed from the semianalytical model assuming that y_{dec}^{dens} is a power law function of the gas density.

In other words, the excellent agreement observed in Fig. C.2 indicates that setting constant decorrelation scales y_{dec}^{diff} and y_{dec}^{dens} is sufficient to reproduce the distribution of the total quantity of matter $N_{\rm H}$. It is so because the chosen $y_{dec}^{\rm dens}$ probably describes the largest sizes of the dense components which capture most of its mass and volume. However, the model also proves that choosing a single value of $y_{dec}^{\rm dens}$ is inappropriate to accurately describe the distribution of H₂. It is so because the mass and volume of H₂ in the simulation is likely built in smaller components. $y_{dec}^{\rm dens}$ should therefore be interpreted as a maximum length scale of the dense gas.

All these considerations show that the simplistic model developed here is very useful to interpret the results of the simulations. It successfully separates local properties and probabilistic effects in the integration of column densities. Moreover it provides estimations of the decorrelation scale of the diffuse gas and the maximum decorrelation scale of the dense gas. Finally, its flaws clearly highlight the importance of a distribution of sizes of the dense and cold ISM and the necessary existence of small dense clouds which produce lines of sight with intermediate integrated molecular fraction. The findings of this section are synthesized in Sect. 3.9 and Fig. 7 and used in the rest of the paper as a major tool for interpreting the results of the simulations.

Appendix D: Kolmogorov-Smirnov test

The results of this paper rely on the comparison of 2D probability histograms of observed and simulated data. To facilitate this comparison and the underlying parametric study, we apply here



Fig. D.1. Results of the modified KS test applied to the fiducial simulation. The black points are the observational data, the red dots indicate the dataset used for the estimations of the merit function M, and the 2D histogram the simulated data. The violet star and rectangle indicate the observational point and the quadrant that maximize M (see main text). The fiducial simulation has a KS distance of 0.98. The corresponding quadrant contains 0.47 and 4.49% of the entire simulated and observed datasets.



Fig. D.2. KS distance between the simulations and the observational sample as a function of the mean density $\overline{n_{\rm H}}$, the UV scaling factor $G_0 = 0.5$ (red), 1 (green), 2 (orange), and 4 (blue), and for a resolution of 256³ (solid lines) and 128³ (dashed lines). All other parameters are set to their standard values (see Table 2). Points correspond to reliable measurements of the KS distances. Triangles indicate lower limits corresponding to simulations where the upper error bar on $R_{\rm KS}$ tends toward infinity (see main text).

a modified version of the Kolmogorov-Smirnov (KS) test. This test, originally developed for the study of 1D samples, searches for the maximum cumulative difference between two distributions. Fasano & Franceschini (1987) generalized the KS test to 2D samples following Peacock (1983) idea of replacing the cumulative probability distribution, which is not well defined in a dimension larger than one, with the integrated probability in each of the four quadrants surrounding a datapoint. Such a consideration allows us to define a KS distance which measures how two 2D distribution functions differ from one another.

As illustrated in Fig. D.1, each observational datapoint is identified by a pair of variables $(N_{\rm H}^{\rm obs}, f^{\rm obs}({\rm H}_2))$ which divide the space into four quadrants: (1) $N_{\rm H} \le N_{\rm H}^{\rm obs} \& f_{\rm H_2} \le f^{\rm obs}({\rm H}_2)$, (2) $N_{\rm H} \le N_{\rm H}^{\rm obs} \& f_{\rm H_2} > f^{\rm obs}({\rm H}_2)$, (3) $N_{\rm H} > N_{\rm H}^{\rm obs} \& f_{\rm H_2} > f^{\rm obs}({\rm H}_2)$, and (4) $N_{\rm H} > N_{\rm H}^{\rm obs} \& f_{\rm H_2} \le f^{\rm obs}({\rm H}_2)$. Each quadrant thus contains two fractions $f_{\rm obs}$ and $f_{\rm sim}$ of the entire observed and

simulated datasets. To compare these values, we define a merit function

$$M = \left| \log_{10} \left(\frac{f_{\rm sim}}{f_{\rm obs}} \right) \right|,\tag{D.1}$$

and the KS distance between the two distributions as the maximum value of *M* computed over all quadrants of an observational dataset *O*,

$$R_{\rm KS} = \max_{\mathcal{O}}(M). \tag{D.2}$$

This procedure, not only provides a measurement of the difference between the two distributions, but also the datapoint and the quadrant that maximize the merit function (see Fig. D.1). The interpretation is also straightforward. For instance, a KS distance of 1 implies that one of all the quadrants scanned contains 10 times fewer or 10 times more observations than simulated lines of sight, and that all the other quadrants have smaller distances.

The stability of the procedure depends on the errors made on the merit function and therefore on the absolute numbers of observed and simulated lines of sight contained in each quadrant. The observational dataset O used to compute the KS distance (red points in Fig. D.1) is chosen as the subsample such that all quadrants scanned contain at least 10 observed lines of sight. With this assumption, the error on the merit function is calculated by taking into account only the statistical errors on the number of simulated lines of sight. For each quadrant, we assume that the "true" merit function ranges between

$$M_{\min} = \left| \log_{10} \left(\frac{f_{\sin} - 3\sqrt{(f_{\sin}/S)}}{f_{obs}} \right) \right| \text{ and } (D.3)$$
$$M_{\max} = \left| \log_{10} \left(\frac{f_{\sin} + 3\sqrt{(f_{\sin}/S)}}{f_{obs}} \right) \right|, \quad (D.4)$$

where S is the total number of simulated lines of sight. Because the errors are asymmetric, M_{\min} can tend toward infinity. If so, the KS distance is a lower limit, even if the infinite error bar is obtained for another quadrant than the one that maximizes M. In short, for each observational datapoint, we compute M_{\min} , M, M_{\max} , and R_{KS} . If one of the M_{\min} goes to infinity, R_{KS} is considered as a lower limit.

Because the division in quadrants is performed on a Cartesian grid, the procedure also depends on the choice of the axes. Mathematically, the best option would be to identify the principal components of the observational sample using proper orthogonal decomposition or singular value decomposition algorithms. Such a method could even be applied to subsamples in order to adaptively rotate the system of axes and follow more precisely the distribution of observations. In any case, the resulting system would be a linear combination of $N_{\rm H}$ and $N({\rm H}_2)$ which could be difficult to relate to the underlying physical properties. Because the molecular fraction is bimodal as a function of $N_{\rm H}$, we choose here $N_{\rm H}$ and $f_{\rm H_2}$ as primary variables. This choice facilitates the physical interpretation of the KS test while ensuring some homogeneity of the spread of the observational data in all quadrants (see Fig. D.1).

As a proof of concept, we display in Fig. D.2 the results of the KS test applied to a grid of simulations obtained for different values of $\overline{n_{\rm H}}$ and G_0 and two different resolutions. Despite its simplicity, this test appears to capture the main behaviors described in Sects. 4.2 and 4.3. In particular, the KS distance between the simulations and the observations is found to strongly depend on both $\overline{n_{\rm H}}$ and G_0 , and more loosely on the resolution. Moreover, the simulations that minimize $R_{\rm KS}$ are found to be identical to those identified in Sect. 4.3 to be in closest agreement with the observations. Interestingly, the value of $R_{\rm KS}$ obtained for $\overline{n_{\rm H}}$ = 4 cm⁻³ and G_0 = 4 is relatively small, in apparent contradiction with the conclusions of Sect. 4.3. This is due to the limit imposed on the minimum number of observations contained in each quadrant. Because of this limit, several observations at large $N_{\rm H}$ are not included in the analysis which reduces the merit function at large column density (region E, see Table 1). Keeping in mind these border effects, the KS test turns out to be a valuable tool for estimating the distance between two distributions without performing a detailed comparison of the samples. In this paper, we apply this method to display our results in a synthetic manner (see Sects. 4.5-4.7) and only give additional details when necessary.

Chapter 11

The Elusive CH⁺

In this chapter, we are addressing the problem of the abundance of CH^+ in the diffuse ISM. As explained in §4, the very presence of CH^+ is a long-standing conundrum of the physics of the ISM. What are the physical processes that control its abundance? Can the dispersion observed be explained by the joint action of the thermal instability and turbulence? Without taking into account the details of turbulent dissipation, can part of the observed column densities of CH^+ be reproduced? To try and answer those questions, we perform a parametric study (§5), that we post-process chemically (§13). The lines of sight are reconstructed with the algorithm presented in §8.3, so that the simulated lines of sight have the same distribution in lenghts as the observed sample. The distributions are finally compared through a modified version of the KS test (§9).

11.1 Chemical Consideration

CH⁺ **formation** The formation of CH⁺ is a slow process because the routes of its formation have either very low rates, e.g. the radiative association Barinovs & Hemert (2006)

$$C^{+} + H \rightarrow CH^{+} + h\nu \quad k_{C^{+}} \sim 2 \times 10^{-17} \left(\frac{T}{300 \,\mathrm{K}}\right)^{-0.42} \,\mathrm{cm}^{3} \mathrm{s}^{-1}$$
 (11.1)

or they are very endothermic, e.g. the ion-neutral reaction (Hierl et al. 1997)

$$C^+ + H_2 \rightarrow CH^+ + H \qquad k_{H_2} \sim 7.4 \times 10^{-10} \exp(-4537/T) \,\mathrm{cm}^3 \mathrm{s}^{-1}.$$
 (11.2)

where 7.4×10^{-10} cm³s⁻¹ and $E_A/k_B = 4537$ K are the best fit parameters found by Hierl et al. (1997), and E_A is the activation energy of this reaction. There are other formation routes for CH⁺, involving species that are orders of magnitude less abundant than the ones considered above. For this reason, they are usually secondary in the standard conditions of diffuse ISM, and therefore we do not discuss them here. Nevertheless, they are taken into account in the chemical solver.

CH⁺ destructions In the diffuse ISM, the principal destruction routes for CH⁺ are the photodissociation by UV photons (van Dishoeck & Black 1988)

$$CH^+ + h\nu \to C + H^+ \qquad k_{\gamma} \sim 2.5 \times 10^{-10} G_{\text{eff}} \,\text{s}^{-1},$$
 (11.3)

the dissociative recombination with electrons (Florescu-Mitchell & Mitchell 2006)

$$CH^{+} + e^{-} \rightarrow C + H \qquad k_{d,e^{-}} \sim 1.5 \times \times 10^{-7} \left(\frac{T}{300 \,\mathrm{K}}\right)^{-0.42} \,\mathrm{cm}^{3} \mathrm{s}^{-1},$$
 (11.4)

and the ion-neutral reaction collision of CH⁺ with atomic hydrogen,

$$CH^{+} + H \rightarrow C^{+} + H_{2} \qquad k_{d,H} \sim 7.5 \times 10^{-10} \,\mathrm{cm}^{3} \mathrm{s}^{-1},$$
 (11.5)

and molecular hydrogen:

$$CH^{+} + H_{2} \rightarrow CH_{2}^{+} + H \qquad k_{d,H_{2}} \sim 1.2 \times 10^{-9} \,\mathrm{cm}^{3} \mathrm{s}^{-1},$$
 (11.6)

Because of its rapid destruction, CH⁺ is a short-lived species, i.e. a species that exists exclusively where it forms because it cannot be transported.

11.2 Impact of Warm-H₂

In each simulation, the abundance of $\rm CH^+$ can be computed with the chemical solver presented in Appendix \$13

As explained in §13, the chemical composition of the gas can be computed with a known density of molecular hydrogen, calculated out-of-equilibrium with RAMSES, or not. In this case, the density of H₂ is then calculated at equilibrium with the rest of the chemistry. In Fig. 11.1, we show the reconstructed column density of CH⁺ vs the total column density $N_{\rm H}$ (see §8.3), for those two cases applied to the fiducial simulation and we compare them to the observational dataset (black circles indicate detections and pink triangles indicate upper limits). When H₂ is computed out of equilibrium, the 2D probability histogram shows a bimodality, i.e. two peaks of probability, one at low column densities, and one at high column densities of CH⁺. The high- $N(CH^+)$ regime extends over 1.5 orders of magnitude in $N_{\rm H}$ and 2 orders of magnitude in fraction of CH⁺ (dotted lines), and matches the bulk of detections (black circles). The bulk of non-detections, on the other hand, is located at the same total column densities as the peak at low- $N(CH^+)$, between $N_{\rm H} \sim 2 \times 10^{19}$ and $N_{\rm H} \sim 9 \times 10^{20}$ cm⁻² (see 4.7). While this low- $N(CH^+)$ regime is at an almost constant fraction of CH⁺, $f(CH^+) = N(CH^+)/N_{\rm H} \sim 10^{-11}$ and is similar in the right and left panels, the difference between the location of the high- $N(CH^+)$ regime is striking.

If the abundance of molecular hydrogen is post-processed at equilibrium (Fig. 11.1, right panel), the simulated lines of sight fail to reproduce the column densities of CH⁺ by at least two orders of magnitude, and the bimodality in the distribution of $N(CH^+)$ disappears. Since the output of the RAMSES simulation is the same in both panels, the structure of the matter is the same and, therefore, the PDF of the total column is the same. The difference on the distribution of $N(CH^+)$ shown in Fig. 11.1 indicates that the observed column densities of CH⁺ highly depend on the presence of out-of-equilibrium molecular hydrogen. Where does the CH⁺ come from? In which physical condition is it produced and what chemical species are involved?



Figure 11.1: Comparison of the observational dataset (black circles and pink triangles) to the 2D probability histograms reconstruction algorithm (see §8.1) applied to the column densities of CH⁺ and the total proton column densities (colored histogram). Observations include detections of CH⁺ (black circles) and upper limits on $N(CH^+)$ (pink triangles). The color code indicates the fraction of lines of sight (in logarithmic scale) contained in each bin. The panel on the left shows results when $n(H_2)$ is computed out-of-equilibrium with RAMSES, while the panel on the right show the probability histograms when the abundance of CH⁺ is computed with H₂ postprocessed at equilibrium. The dotted lines indicates $f(CH^+) = 10^{-6}$, 10^{-7} , 10^{-8} , 10^{-9} , 10^{-11} , and 10^{-12} .



Figure 11.2: 2D PDF in thermal pressure and density weighted by the density of CH⁺. Solid curves are the thermal equilibrium curve for $G_{\text{eff}} = G_0 \exp(-2.5A_V)$ for $G_0 = 1$ and two values of A_V (0 and 1, the higher and lower curve respectively). The dotted lines are two isothermal curves at 8000 K and 500 K.

Physical Conditions Favorable to CH⁺ In Fig. 11.2, we plot the 2D histogram of the density of CH⁺ in a thermal pressure-density diagram. As in the previous graph, these results are shown using the out-of-equilibrium density of H₂ computed with RAMSES, or not. CH⁺ is formed both in the stable CNM branch and in the warm unstable LNM. When H₂ is computed out-of-equilibrium (left panel), CH⁺ is particularly abundant in the unstable LNM phase between 500 K and 8000 K (dotted lines). A high percentage of the total mass of CH⁺ belongs to pressure conditions above the thermal equilibrium curve (blue curves).

The main processes of formation of CH^+ in the warm and in the cold phase are different. In the LNM and WNM, CH^+ is produced through the endothermic reaction $C^+ + H_2$ and is mainly destroyed by photodissociation and collisions with electrons and atomic hydrogen, which are abundant in those conditions (see Fig. 11.3). As a result, the density $n(CH^+)$ is proportional to the fraction of H_2 . In colder regions, CH^+ is either produced through the radiative association $C^+ + H$, or through a chain of reactions initiated by the radiative association $C^+ + H_2$, which have low reaction rates. In these regions, CH^+ is effectively destroyed by atomic and molecular hydrogen. Because of the lack of an efficient formation process, the abundance of CH^+ is small there.

When H_2 is post-processed at equilibrium, H_2 is mainly present in the CNM, and only a small fraction of molecular hydrogen is at "high" temperatures. As a result, the total mass of CH⁺ is small and its column densities are underestimated by two or three orders of magnitude compared with observations (see Fig. 11.1). Conversely, the out-of-equilibrium H_2 advected from the CNM to the LNM boosts the production of CH⁺ by several orders of magnitudes. The resulting column densities of CH⁺ not only match those derived from the observations but also, and most remarkably, explain their probability distribution. In particular, the three regimes discussed in §4 appear to be a natural outcome of random lines of sight across the multi-phase environment.



Figure 11.3: Pressure vs density histograms weighted by the density of the species involved in CH⁺ formation and destruction reactions (atomic and molecular hydrogen, electrons and carbon ions), showing in which physical conditions the different species are abundant. Note that for an easy comparison of the abundances, the colorcode is the same one for all species. Both *x* and *y* axes are in logarithmic scale, and with units of cm⁻³ and cm⁻³K, respectively.

11.3 Exploration of the Parameters

Impact of the Numerical Resolution In Fig. 11.4, we show the effect of the resolution on the correlation of $N(CH^+)$ with the total column density. Differences are found at low resolution (< 256³). In particular, the simulation with the lowest resolution (64³) predicts column densities of CH⁺ more than three times larger than those predicted at high resolution. This result comes from an overestimation of the volume occupied by the out-of-equilibrium H₂, and, therefore, the total mass of warm molecular hydrogen. Interestingly, and in agreement with the results of Valdivia et al. (2016a, 2017), the column densities of CH⁺ appear to be converged at large resolution ($\geq 256^3$). This result justifies the use of simulations with a moderate resolution to study the CH⁺ formed by advection of the gas at the interface of a multi-phase environment.

Effects of G_0 **and** $\overline{n_{\rm H}}$ Fig. 11.5 shows comparisons between the observational dataset and the 2D probability histograms extracted from the simulations for $0.5 \text{ cm}^{-3} \le \overline{n_{\rm H}} \le 4 \text{ cm}^{-3}$, and $0.5 \le G_0 \le 4$. The trend and the statistics of the probability histograms of $N(\text{CH}^+)$ vs $N_{\rm H}$ strongly depend on the total mass of the gas parametrized by $\overline{n_{\rm H}}$ and on the UV illumination factor, G_0 . This was expected because those two parameters have a major role in the column densities of H₂ (Bellomi et al. 2020), which is the precursor of CH⁺ formation.

As G_0 increases or $\overline{n_{\rm H}}$ decreases,

- 1. the fraction of lines of sight with large $N(CH^+)$ decreases to the benefit of lines of sight with low $N(CH^+)$, which directly follows the behavior of the $f(H_2)$ presented by Bellomi et al. (2020);
- 2. the fraction of CH^+ in the high- $N(CH^+)$ regime decreases, progressively removing



Figure 11.4: Comparison of the 2D probability histogram reconstruction algorithm (see §8) applied to the fiducial simulation for three resolutions, $R = 64^3$ (top panel), $R = 256^3$ (central panel), and $R = 512^3$ (bottom panel). The column densities of CH⁺ and the total column densities are plotted in logarithmic scale.



Figure 11.5: Comparison of the observational dataset (black and pink points) with the 2D probability histograms of $N_{\rm H}$ and $N({\rm CH^+})$ computed from the reconstruction algorithm §8 applied to the fiducial grid. Both axis are in logarithmic scale.



Figure 11.6: Probability histogram of the thermal pressure vs the local density weighted by the density of molecular hydrogen for the fiducial grid. The colorcode is the same for all the panel.

the bimodality;

- 3. the low- $N(CH^+)$ regime extends toward higher $N_{\rm H}$;
- 4. all these effects together shift the transition toward larger total column density.

While the simulations that reproduced the HI-to-H₂ transition were found to follow a trend $G_0/\overline{n_{\rm H}} \sim 0.5 - 1$ the probability histograms of CH⁺ exhibit large differences when $G_0/\overline{n_{\rm H}}$ is constant. For instance, the fraction of lines of sight in the low- $N(CH^+)$ regime and the mean fraction of CH⁺ in the high- $N(CH^+)$ regime are not constant for a given $G_0/\overline{n_{\rm H}}$ ratio but respectively decreases and increases with $\overline{n_{\rm H}}$. The effect is so large that for $G_0 = 0.5$ and $\overline{n_{\rm H}} = 0.5$ cm⁻³ there is almost no line of sight with large $f(CH^+)$. The fact that for a given $G_0/\overline{n_{\rm H}}$ ratio the probability histograms appear so different reflects the importance of the large-scale structure of matter, which controls the fraction of mass advected from CNM to WNM, and the porosity to the UV radiation field.

In Fig. 11.6, we show how the mass of H₂ changes with the external radiation field and the mean density. We can see that along a constant $G_0/\overline{n_H}$ ratio, the amount of warm H₂ strongly increases with $\overline{n_H}$. While the mean molecular fraction is rather constant for a given $\overline{n_H}/G_0$ ratio, the fraction of warm H₂ is not. This explains why we see more differences on $N(CH^+)$ vs N_H than on the HI-to-H₂ transition along a constant $G_0/\overline{n_H}$.

In addition, in Fig. 11.7 we show the KS difference between the observational data and the probability histograms shown in Fig.11.5. The solid lines indicate the KS test applied



Figure 11.7: KS distance between the simulations and the observational sample as a function of the mean density $\overline{n_{\rm H}}$ and the UV scaling factor $G_0 = 0.5$ (red), 1 (green), 2 (orange), and 4 (blue). Results obtained taking into account all observations or only upperlimits are shown with solid lines and dashed lines, respectively (see §9). All other parameters are set to their standard values (see Table 5.1). Points correspond to reliable measurements of the KS distances. Triangles indicate lower limits corresponding to simulations where the upper error bar on $R_{\rm KS}$ tends toward infinity.

taking into account the whole observational sample – detections and non-detections – while the dashed lines are the KS distances taking into account only the detections. All things considered, the tightest concordance between observed and simulated data is obtained for $\overline{n_{\rm H}} \sim 2 \,{\rm cm}^{-3}$, in agreement with our results on HI-to-H₂ from the previous chapter.

Size of the Box The amount of matter and the subsequent extinction of UV photons is a key ingredient in the study of the column density of CH⁺. The amount of matter in the box is controlled also by another parameter of the simulation, the size of the box *L*. In Fig. 11.8, we compare the probability histograms of $N_{\rm H}$ and $N({\rm CH^+})$ with a box size of 200 pc (left panel, fiducial simulation) and L = 50 pc (right panel). The forcing applied to the simulation at L = 50 pc is such to obtain the same velocity parameter $V = \sqrt{FL_{\rm drive}}$ (see Fig. 5.12) for both simulations under analysis. Reducing *L* by a factor of four drastically reduces the total amount of matter in the simulation, hence the absorption of the impinging UV radiation field and the large-scale self-shielding of H₂. The resulting impact on the averaged molecular fraction is transmitted to the fraction of warm H₂, hence the production of CH⁺. Moreover, because $L_{\rm drive} = L/2$ is consequently four times smaller in the simulations displayed in the right panel of Fig. 11.8 than in that displayed on the left panel, the decorrelation scales of the diffuse and denser medium are correspondingly smaller (see the semi-analytical model in the Appendix C of the paper in §10). Therefore, all the reconstructed lines of sight are therefore considerably larger than individual density fluctuations


Figure 11.8: Comparison of the observational dataset (black and pink points) with the 2D probability histograms of $N_{\rm H}$ and $N({\rm CH^+})$ with a box size of 200 pc (left panel) and L = 50 pc (right panel). The turbulent forcing is adjusted ($F = 3.6 \times 10^{-3}$ kpc Myr⁻²) to obtain similar velocity dispersions for L = 50 pc and L = 200 pc. All other parameters are set to their fiducial values.

induced by the turbulent forcing. This magnifies the impact of the central limit theorem, and the histogram with L = 50 pc results squeezed along the x and y axes compared to the probability histogram at L = 200 pc. Consequently, the simulation at L = 50 pc predicts almost no line of sight at low total column density ($N_{\rm H} \le 3 \times 10^{19}$ cm⁻²). Even though it is not systematically shown, this result is not limited to the fiducial setup but is a general feature of all the simulations explored in this work.

Effects of Turbulence Forcing While the effect of turbulence was limited on the kingfisher diagram of the HI-to-H₂ transition, especially for high-compressive forcing, it appears to affect greatly the probability histogram of $N(CH^+)$ vs N_H . In Fig. 11.10, we show the probability histogram of the reconstructed lines of sight for three different forcing strengths, *F*, and three different compression parameters, ζ .

As *F* increases or, to a lesser extent, ζ decreases:

- 1. the fraction of lines of sight with large $N(CH^+)$ increases to the deficit of lines of sight with low $N(CH^+)$, increasing the fraction of CH⁺ globally;
- 2. the transition is shifted toward smaller $N_{\rm H}$ and its width is reduced in the $N_{\rm H}$ -axis;
- 3. all these effects increases the bimodality.

Oppositely to the effect of G_0 and $\overline{n_H}$, the turbulence affects greatly the width of the PDF of $N(CH^+)$. The PDF of N_H is larger for a compressive forcing, because compressive modes induced larger pressure variations, hence, large variations in density. Both the volume of WNM and fraction of mass of CNM increase (see Fig. 5.5), therefore N_H appear broader.

In Fig. 11.10, the effect of the acceleration parameters and ζ are shown on the pressure vs density diagram weighted by the density of CH⁺. Turbulence induces pressure fluctuations, that appear in these graphs as a larger dispersion in the *y*-axis. These pressure variations increase with the acceleration parameter, the fraction of mass located in the CNM



Figure 11.9: Comparison of the observational dataset (black and pink points) with the 2D probability histograms of $N_{\rm H}$ and $N({\rm CH^+})$ of different simulations with three values of the acceleration parameter $F = 4.6 \times 10^{-4}$, 1.5×10^{-3} , and 4.6×10^{-3} kpc Myr⁻², and three values of the compressive ratio $\zeta = 0.9, 0.5, 0.1$ which set the balance between solenoidal and compressive forcing (from the more solenoidal to the more compressive). All other parameters are set to their standard values (see Table 5.1).

phase diminishes to the benefit of the out-of-thermal-equilibrium gas (see Fig. 5.5). This increases the abundance of CH^+ along the lines of sight, increasing the fraction of CH^+ globally on all lines of sight as *F* increases. As shown by Bellomi et al. (2020), a substantial fraction of the kinetic energy injected in pure solenoidal modes effectively prevents the unstable gas to condensate back to cold and dense environments. This leads to a decrease of the mass fraction in the CNM and of the global amount of H₂. This naturally affects the fraction of CH⁺ that appears to be lower for solenoidal forcing than compressive forcing.

Overall, simulations with a substantial fraction of the kinetic energy injected in pure compressive modes, lead to a better agreement with the observational sample. The closest agreement is obtained for a simulation with a turbulent forcing of $F \sim 1.5 \times 10^{-3}$ kpc Myr⁻², which correspond to a turbulent velocity dispersion of $\sigma \sim 3$ km s⁻¹ (Bellomi et al. 2020). For large turbulence, especially if compressive modes, the probability histograms show a clear overestimation of the column density of CH⁺, which is reflected by the KS test with a big KS difference.



Figure 11.10: 2D probability histograms of the thermal pressure vs the local density weighed by the density of CH⁺. Results are shown for different turbulent forcing strengths, *F*, and spectal parameters, ζ . Both axes are in logarithmic scale.

Magnetic Field In Fig. 11.11, we show the results of the KS test applied to simulation with different initial magnetic field B_x .

The impact of the initial magnetic field reveals the competition between thermal instability which induces the production of dense environments and the magnetic pressure which acts against this evolution and expands it. As shown in Fig. 11.11, the KS distance obtained for different magnetic field intensities is found to be constant until a critical value of $B_x \sim 4 \ \mu$ G, for both the computation of the KS test taking into account only detections (dotted) or both detections and upper limits (solid). This result was already found for the HI-to-H₂ transition. Oppositely to what was found for the HI-to-H₂ transition, at a stronger magnetic field, the KS difference increases smoothly and gently, suggesting that



Figure 11.11: KS distances between the simulations and the observational sample computed for six values of the initial magnetic field B_x . All other parameters are set to their standard values (see Table 5.1). Dashed lines indicates the KS distance computed on detections, while solid lines indicate the KS distance computed taken into account the upperlimits as well (see §9).

the B_x does not necessarily need to be lower than the equipartition to reproduce the observed 2D probability histograms of $N(CH^+)$ vs N_H .

11.4 Conclusion

The transport induced by large-scale turbulence in the multi-phase ISM is a viable scenario that explains simultaneously the large column densities of CH⁺ detected in the diffuse medium, its trend as a function of $N_{\rm H}$, and its observed probability distribution function. The variations of the integrated fraction of CH⁺ from one line of sight to another can be obtained with one set of parameters, i.e. without taking into account variations of large-scale physical conditions along one line of sight or from one line of sight to another. The tightest agreement is obtained for a neutral diffuse gas modeled over ~ 200 pc, with a mean density $\overline{n_{\rm H}} = 2 \text{ cm}^{-3}$, illuminated by the standard interstellar UV radiation field $G_0 = 1$, and stirred up by a large-scale compressive turbulent forcing. This was also found to be the best set of parameters required to explain the properties of the HI-to-H₂ transition observed in the local ISM.

Combining the chemical and statistical information contained in the observations of CH^+ shows that out-of-equilibrium H_2 is capital in our study. CH^+ proves to be a tracer of the warm and out-of-equilibrium phase of the gas, warm- H_2 , and turbulence. A further improvement would be to take into account the excitation of the warm H_2 . Indeed, the excitation of H_2 would reduce the energy barrier of the formation of CH^+ , making it more effectively produced in warm gas. This is especially true in regions where the radiation field is high, i.e. where UV pumping effectively populates the high energy levels of molecular hydrogen.

Chapter 12

Excitation of Neutral Carbon

As explained in §4.2, Jenkins & Tripp (2011) derived the pressure through the observation of the levels populations of CI, by dividing the spectrum in velocity bins and computing the abundance of each level relative to the total amount of CI in each velocity channel. In this chapter, we explore the production and excitation of neutral carbon in the diffuse ISM in our simulations. Our goal is to analyze if the variation of pressure, density, and temperature due to the joint action of thermal equilibrium and turbulence are sufficient to reproduce the dispersions of *f*1 and *f*2 defined in eqs. 4.2. To achieve such a goal, we compute the abundance of CI with the chemical solver (see §13) and the population of the fine-structure levels, including all the excitation and de-excitation processes presented in §6. We reconstruct the lines of sight (see §8) with the same distribution of l_{los} observed (see Fig. 12.1) and we then evaluate the integrated line profile (see §6.4) to have the population fraction of the fine structure per velocity channel.

Distribution of lengths of LOS From the positions (see Fig. 4.4) and the distances (see Fig. 15.2) of the background sources observed by Jenkins & Tripp (2011), we compute the length of lines of sight with eq. 8.1. The normalized histogram (see Fig. 12.1) give us the weights for each l_{los} to be used in the recontruction algorithm §8.

Reconstruction We create a dataset of f1 and f2 which contain the $N(CI^*)/N(CI_{tot})$ ratio and the $N(CI^{**})/N(CI_{tot})$ ratio for each velocity channel of width 0.5 km s⁻¹. In Fig. 12.2, we show a comparison of the results with the raw simulation of $l_{los} = L = 200$ pc (on the left panel) and taking into account the distribution of lengths of the lines of sight of the observational sample (right panel). The entire probability histogram appear to belong to the low-pressure regime of the f2 vs f1 graph (see Fig. 6.3). Moreover, the bulk of the probability histogram is at low values of f1 and f2. The peak of the reconstructed probability histogram (right panel) is found to be shifted and spread toward larger f1 and f2, compared to those of the initial distribution (left panel). This is particularly true for models with a higher Doppler parameter b (see Fig. 12.3). All these results were expected. Indeed, long lines of sight or large Doppler parameter favor the mixing in any velocity channel of gas at different physical conditions. This naturally favors larger f1 and f2 values compared to the original distribution.



Figure 12.1: Normalized histogram of the lenghts of the lines of sight (see eq. 8.1) in bins in logarithmic space centered at $l_{los} = 100,200,400,800,1600$, and 3200 pc. The resulting weights used for the reconstruction algorithm are 0.06, 0.12, 0.19, 0.24, 0.21, and 0.18, respectively.



Figure 12.2: Probability histogram of f_2 vs f_1 in velocity bins of 0.5 km s⁻¹ for the fiducial simulation. The raw result of a simulation is shown on the left panel. On the right panel, the lines of sight have been reconstructed with the algorithm explained in §8.1 taking into account the weights shown in Fig. 12.1. The Doppler parameter is set to $b = 8 \text{ km s}^{-1}$ (see eqs. 6.10 and 6.11).



Figure 12.3: Probability histogram of f_2 vs f_1 in velocity bins of 0.5 km s⁻¹ for the fiducial simulation, for different Doppler parameters b = 1, 2, and 8 km s⁻¹ used in the calculation of the line profiles (see eqs. 6.10 and 6.11).

12.1 Influence of the Parameters

Doppler parameter In Fig. 12.3, we show a comparison of the results with different values for the Doppler parameter b, used in the formula of the line profile for the computation of f1 and f2 (see eqs. 6.10 and 6.11). When b is low, the broadening of the absorption line is dominated by the thermal broadening, while a large b corresponds to a broadening dominated by turbulence. Evidently, a large value of b has a larger impact on CNM cells, which have small thermal broadenings, due to their low temperatures. Since the Doppler parameter controls the width of the gaussian profile (see §6.4), when b is large, any cell with high f1 or f2 affects more velocity bins, smearing out its high population into more velocity channels. As a result, the Doppler parameter affects the bulk of the probability distribution of f2 vs f1, spreading it toward larger values of f1 and f2.

Resolution In contrast to the HI-to-H₂ transition and as shown in Fig. 12.4, the resolution is of great importance for the level populations of CI. Between resolutions of 64^3 and 512^3 , the larger the resolution the more spread is the probability histogram in both f1 and f2. While results derived for $R = 256^3$ and $R = 512^3$ are comparable, the distribution of f1 and f2 has clearly not converged, even at the highest resolution. This indicates that both f1 and f2 are highly sensitive to the physical conditions and structures at scales as small and smaller than ~ 0.4 pc.

Fig. 12.5 shows the conditions of pressure and density where the excited levels of CI (in this case CI^{*}) are abundant. Fig. 12.5 contains all the information required to explain the dependence of the probability distribution of f_1 and f_2 on the numerical resolution. While neutral carbon may be built up in the LNM where the activation of the production of CH⁺ favors the production of CI, CI is mostly built up in CNM gas, i.e. in regions where the UV radiation field is sufficiently absorbed to allow the C⁺-to-CI transition. This result is in agreement with the predictions of static models of photodissociation (e.g. Le Petit et al. 2006) which show that neutral carbon is mostly formed, not at the border but in a shell inside CNM clouds, where the visual extinction $A_V \sim 1$. In contrast to the HI-to-H₂

transition, the C⁺-to-CI transition requires either larger sizes of CNM clouds (for a given density) or larger densities (for a given size). The size distribution of CNM clouds set by the combined action of turbulence and thermal instability implies that H_2 is built up in CNM clouds of typical sizes comprised between 3 and 10 pc (Bellomi et al. 2020). Because the size distribution is fixed, the mass of CI is necessarily built up in denser clouds which are considerably smaller than these scales. Increasing the resolution has two combined effects. It allows the gas to form denser entities which are favorable for the production of CI. Because these denser entities occupy a small fraction of the volume but also have a high surface filling factor, increasing the resolution increases the probability of intercepting them along random lines of sight. This is why the probability distributions shown in Fig. 12.5 shift towards larger densities as the resolution increases and why the 2D distribution of the *f*1 and *f*2 ratio sample larger values of *f*1 and *f*2 at larger resolution (Fig. 12.2).

Turbulence In Fig. 12.6, we show the effect of turbulence on the populations of CI, for turbulent strengths, $F = 4.6 \times 10^{-4}$, 1.5×10^{-3} , and 4.6×10^{-3} kpc Myr⁻², and compressibility factors, $\zeta = 0.1, 0.5$, and 0.9. The corresponding distributions of $n(CI^*)$ in a pressure-density diagram are displayed in Fig. 12.7. As shown by Seifried et al. (2011), Saury et al. (2014), and Bellomi et al. (2020), increasing the turbulent forcing quickens the cycle between the CNM, LNM, and the WNM and favors the exchange of mass between these different phases. The mean pressure therefore increases, the 1D PDF of the density broadens and its bimodal nature progressively disappears (Piontek & Ostriker 2005; Walch et al. 2011). These broader variations of the pressure and the density naturally imply larger variations of the f_1 and f_2 ratios displayed in Fig. 12.6. Because high turbulent forcing drastically increases the mass and volume occupied by the LNM phase, the ratio f2/f1 slightly increases compared to low and moderate turbulent forcing. The effect on the f2/f1 ratio is only slight because, while the mass of CI produced in the LNM increases with the turbulent forcing, it is still low compared to the mass of CI built up in the CNM phase (see Fig. 12.7). Interestingly, the compressibility factor ζ has only a slight effect on the combined distribution of f_1 and f_2 . This is surprising because ζ was found to have a large impact on the HI-to-H₂ transition (Bellomi et al. 2020). The main reason is that f_1 and f_2 respectively correspond to column density ratios $N(CI^*)/N(CI_{tot})$ and $N(CI^{**})/N(CI_{tot})$ per velocity channel and are therefore less sensitive to the total mass of CI.

12.2 Comparison with Observations

In Fig. 12.8, we compare the f_1 and f_2 ratios observed by Jenkins & Tripp (2011) to the results of the fiducial simulation at a resolution of 512^3 and all the other parameters set to their fiducial values. All the values of f_1 and f_2 computed in the simulation are found between the theoretical curves at low and high temperature T = 15 K and 10^4 K (dotted and dashed lines) and to be confined in the low-density regime, hence the low-pressure regime (see Fig. 6.3). Despite the variations of physical conditions induced by turbulence and the thermal instability, the predictions of simulations clearly fail to reproduce the distribution



Figure 12.4: Probability histogram of f_2 vs f_1 in velocity bins of 0.5 km s⁻¹ for simulations with different resolutions and all physical parameters set to their standard values.



Figure 12.5: Combined histograms of the thermal pressure as function of the local densities for different resolutions. The histograms are weighted by the density of the first excited level, $n(CI^*)$, thus indicate in which pysical conditions CI^* is abundant.



Figure 12.6: Probability histogram of f2 vs f1 in velocity bins of 0.5 km s⁻¹ for simulation run with different turbulent strengths, $F = 4.6 \times 10^{-4}$, 1.5×10^{-3} , and 4.6×10^{-3} kpc Myr⁻², and compressibility, $\zeta = 0.1, 0.5$, and 0.9.



Figure 12.7: Histograms of the thermal pressure as function of the local density for different turbulent strengths and compressions. The histograms are weighted by the density of the first excited level, $n(CI^*)$, and therefore indicate in which pysical conditions of temperature and density CI^* is abundant.

of observed f1 and f2 ratios. In particular, the values of f2 predicted by numerical simulations are found to be always smaller than 0.05, in evident contradiction with the observations where f2 reaches values as large as 0.5. As explained by Jenkins & Tripp (2011) and summarized in §4.2, the observed distribution of f1 and f2 could be explained if any line of sight systematically intercepts a low ($P \sim 10^3 - 10^4$ K cm⁻³) and a high-pressure component ($P \sim 10^5 - 10^6$ K cm⁻³). The results displayed in Fig. 12.8 show that turbulence and thermal instability alone fail in explaining the presence of the high-pressure component. With the turbulent forcing explored in this work, the pressure of the gas hardly goes above 10^5 . Moreover, the corresponding cells are either at low density and high temperature or large density and low temperature (see Fig. 12.7) which are both unfavorable conditions to obtain large f1 and f2 ratios. Finally, and as shown by Fig. 12.8, the high-pressure component cannot be ascribed to the presence of self-gravitating or collapsing clumps – which effectively form in the simulations – because those are simply too cold to efficiently excite the second level of CI.

12.3 Conclusions

It should be kept in mind that the discrepancies found here only concern the high-pressure component. As shown by Saury et al. (2014) and as found in this work, the predicted distribution of thermal pressure is in excellent agreement with the distribution of pressure derived by Jenkins & Tripp (2011) for the low-pressure component (see Fig. 15 of Saury



Figure 12.8: Comparison between the observations of Jenkins & Tripp (2011) (blu dots) of f2 vs f1 and the probability histogram for the fiducial simulation at resolution 512^3 . The dotted and dashed lines are theoretical curves at temperatures of 15 K and 10^4 K, respectively.

et al. 2014). This being said, the results of this chapter are unambiguous: the origin of the high-pressure component seen in the observations of CI is a deep and fundamental mystery that cannot be explained by thermal instability, turbulent transport between phases, or gravitational collapse. Several alternative scenarios need to be explored.

- The high-pressure component could originate from high-velocity J-type shocks propagating in the WNM. High-velocity J-type shocks ($V_S > 40 \text{ km s-1}$) propagating in a low-magnetized environment could bring the gas to pressure and density regimes high enough to significantly populate the second level of CI. In principle, the probability of occurrence of these events and their effect on the local chemistry and excitation processes should be captured by numerical simulations of the diffuse multiphase ISM. In practice, because the scales considered (~ 0.4 200 pc) are much larger than the viscous scale (~ 0.3 mpc for the WNM) and because the numerical resolution is finite, such structures are smoothed out by numerical simulations. There is almost no way around that because simulations performed at a much smaller scale will fail to capture the thermal instability process. The only way to study these events would be to run 1D state-of-the-art models of interstellar shocks and ascribe a probability distribution function to such events based on the analysis of turbulence at larger scales.
- As proposed by Jenkins & Tripp (2011), the high pressure component could also be due to the intermittent dissipation of interstellar turbulence in CNM clouds (Falgarone et al. 1995; Joulain et al. 1998; Godard et al. 2009; Lesaffre et al. 2013; Godard et al. 2014; Lesaffre et al. 2020). Dissipative coherent structures in the CNM could locally increase the temperature (hence the pressure) of the gas to a level sufficient to excite CI. In particular, dissipation through ambipolar diffusion, i.e. the differential velocities between neutral and charged species, would have the effect of enhancing the collisional excitation rate of CI in collisions with charged species such as H⁺ or e⁻. In principle, such a scenario could explain the observations providing that a CNM cloud of density of ~ 100 cm⁻³ is heated up to temperatures of about 1000 K or if the ambipolar diffusion drift is larger than $\sim 4 \text{ km s}^{-1}$. Again, the ideal MHD simulations presented in this work do not resolve dissipation processes and dissipative scales. Moreover, the simulations only consider a single fluid and therefore do not follow the decoupling between ions and neutrals. One way to study this effect would be to run state-of-the-art numerical simulations of the CNM designed to capture dissipative processes (Lesaffre et al. 2020, Richard et al., in prep.) including the decoupling between ions and neutrals.
- At last, and as proposed by Jenkins & Tripp (2011), the high-pressure component could be uncorrelated with the diffuse interstellar gas and rather originate from the gas associated with the background sources. If so, this component can not be captured by the present simulations which neglect the presence of stars and their surrounding environments.

Part IV

Conclusion and Perspectives



The study of the diffuse interstellar medium is rich in terms of spectroscopic observations. In this Ph.D. work, we attempt not only to use the chemical information contained in the observations but also to take advantage of their statistical properties.

The local ISM is an open and complex system. One of the main challenges to be faced is to understand the joint action of the thermal instability and the turbulent motions, and the roles of different components of the ISM on observed quantities. To quantify separately the impacts of the mean density, the UV radiation field, the integral scale, the resolution, the turbulent forcing, the magnetic field, and the gravity on the observables, we used stateof-the-art MHD simulations, and an ensemble of 305 runs (see §5). Taking into account the distance of each background source and simulating random lines of sight over the same distribution of distances is paramount to explain the range of observed column densities and their corresponding statistics (see §8). To compare the results of the simulations to the observational sample, we propose a new version of the Kolmogorov-Smirnov test (see §9) which is a valuable tool for estimating the distance between two distributions and the errors on the estimated distance. Moreover, it provides information on where the maximal difference between the distributions occurs, bringing key insights into the origin of the discrepancies between observations and simulations. The original feature of this work is to not only focus on the production of individual column densities but also on their statistics, meaning the probabilities of occurrence of these column densities along random lines of sight.

This exhaustive parametric study, combined with a chemical solver and the tools mentioned above, allows the investigation of the thermodynamical state and the evolution of the local diffuse interstellar medium. In this work, we applied the study to three chemical observables: H_2 , CH^+ , and CI.

The HI-to-H₂ **transition** The joint actions of thermal instability and large-scale turbulence in the fiducial simulation are found to produce a wealth of lines of sight which reproduce the observed position and width of the HI-to-H₂ transition, and whose probabilities of occurrence match those derived from the observations. The agreement is so remarkable that it is almost unnecessary to invoke variations of physical conditions along lines of sight or from one line of sight to another.

The distribution of column densities computed from the simulations strongly depends on the Galactic midplane density parametrized by the mean density $\overline{n_{\rm H}}$, the density of OB stars parametrized by the UV scaling factor G_0 , and the scale of neutral diffuse clouds parametrized by the box size *L*. It is so because these three parameters not only regulate the mean pressure of the gas, hence the fractions of mass and volume occupied by the CNM and WNM, but also control the typical scale of density fluctuations in the WNM and the distribution of sizes of the CNM structures where H₂ forms. The tightest concordance between the observed and simulated samples is obtained for a mean density $\overline{n_{\rm H}} = 1 - 2$ cm⁻³ and a UV radiation field scaling factor $G_0 = 1$ (in Habing units), in good agreement with the values deduced from HI and CO all-sky surveys and from direct observations of the UV radiation field in the Solar Neighborhood. The range of observed column densities of H and H₂ requires a box size *L* = 200 pc which corresponds to the estimated scale of HI superclouds. Within this setup, the column densities of HI are inferred to be built up in large-scale WNM and CNM structures correlated in densities over ~ 20 pc and ~ 10 pc, respectively. In contrast, H₂ is inferred to be built up at smaller scales. However, the fact that the kingfisher diagram is independent of the resolution of the simulation suggests that most of the mass and volume of H₂ is contained in CNM structures between ~ 3 and ~ 10 pc. All these values are given for the standard simulation (L = 200 pc and $L_{drive} \sim 100$ pc) but naturally depends on the size of the box and the mean density of the gas.

In spite of the strong influences of $\overline{n_{\rm H}}$, G_0 , and L, the statistical properties of the HI-to-H₂ transition are otherwise remarkably stable. Admittedly, the kingfisher diagram depends on the strength of the turbulence if most of the forcing is injected in solenoidal modes; however, such a configuration prevents to reproduce the observational sample unless the large-scale velocity dispersion of the gas is unrealistically small. In contrast, if most of the kinetic energy is injected in compressive modes, the kingfisher diagram is found to weakly depend on the strength of the forcing. Similarly, the HI-to-H₂ transition is almost not affected by gravity and is found to weakly depend on the Doppler broadening parameter and the strength of the magnetic field, as long as $B_x \leq 4 \mu G$. The 2D probability histogram of the column densities of H and H₂ is, therefore, a valuable tool to constrain the nature of the turbulent forcing at large scales; however, it provides few or no information regarding the velocity dispersion of the gas, the amount of gravitationally bound environments and the strength of the magnetic field. Other observational tracers are required.

The elusive CH⁺ The distribution of column densities $N(CH^+)$ vs N_H computed from the simulations not only strongly depends on the mean density $\overline{n_H}$, the UV scaling factor G_0 , and L, but also on the turbulent forcing, which greatly affects the width of the PDF of $N(CH^+)$. As for the HI-to-H₂ transition, the observed variations of the integrated fraction of CH⁺ from one line of sight to another can be obtained with a single simulation computed with one set of parameters, i.e. without taking into account variations of large-scale physical conditions along one line of sight or from one line of sight to another. The best agreement is obtained for a neutral diffuse gas modeled over ~ 200 pc, with a mean density $\overline{n_H} = 2 \text{ cm}^{-3}$, illuminated by the standard interstellar UV radiation field $G_0 = 1$, and stirred up by a large-scale compressive turbulent forcing. The study we conducted indicates that out-of-equilibrium H₂ is capital for reproducing the observed column densities of CH⁺.

The Neutral Carbon With the CI excitation code (see §6) we computed the populations of the fine-structure levels of neutral carbon and the fractions f1 and f2 per velocity channel, which is a measure of the pressure of the diffuse ISM. The probability histograms of f2 vs f1 are greatly affected by the resolution, the turbulent forcing strength, and the doppler parameter. Contrary to the case of CH⁺ and molecular hydrogen, the distribution of pressure derived by Jenkins & Tripp (2011) cannot be fully reproduced by the combination of the thermal instability and the turbulent transport between the phases, or gravitational collapse. While the distribution of pressure in the low-pressure regime is in agreement with that deduced from the observed populations of the first two excited fine-structure levels of CI, the high-pressure component is not reproduced. These discrepancies require the exploration of other scenarios such as high-velocity J-type shocks, or intermittent dis-

sipation of interstellar turbulence.

Perspectives All these results open new perspectives for the study of the chemical state of the ISM in which any observation must be understood through the combination of local physical conditions and the probabilistic ordering of these conditions along the line of sight. In particular, similar studies should be applied to all atomic and molecular species with observational samples large enough to conduct statistical analysis. Since the modifications of the Kolmogorov-Smirnov test make it applicable to N dimensions, the study can be expended to probability histograms in higher dimensions, taking into account simultaneously the joint information contained in the column densities of several species.

Part V

Appendix

Chapter 13

Postprocessing: Chemistry

To explore the effects of the local physical conditions (density, temperature, and local radiation field) and the global properties of the ISM (turbulence, magnetic field, mean radiation field, thermal instability, and gravity) on observational tracers such as CH⁺ and the level population of CI, we need to compute the chemical composition of the multiphase medium. To achieve such a goal, we post-process the simulation output with a state-ofthe-art chemical solver that computes the density of 151 species in each cell of the simulation, by solving a chemical network, at equilibrium, that includes around 2700 reactions.

The chemical network includes gas-phase reactions (e.g., neutral-neutral, ion-neutral), reactions onto grains (e.g., the formation of H₂), photoreactions due to UV photons (photodissociation and photoionization), and reaction with cosmic rays. The code takes as input the local conditions of a fluid cell, i.e. the temperature *T*, the proton density $n_{\rm H}$, the radiation field G_0 , the extinction by dust grains A_V , the self-shielding of H₂, $f_{\rm shield}$, and optionally the abundance of H₂. The gas-phase chemical abundances are taken to be solar abundance with a depletion factor that depends on the element (see Table 13.1).

The chemical solver, fully described by Valdivia et al. (2017) (see Appendix A) computes the abundances of all chemical species at equilibrium assuming two different prescriptions depending on the choice of the user. If the out-of-equilibrium of H₂ computed in RAMSES is given in input, the code computes the chemistry at equilibrium except for H and H₂. If not, the code computes the entire chemistry at equilibrium, including the abundances of H and H₂. Such a method allows to compute the chemistry partly out-of-equilibrium and is justified by consideration on chemical timescales. Indeed as found by Valdivia et al. (2017) the equilibrium timescales of H₂ are larger than that of any other species over a wide range of physical parameters, as long as $n_{\rm H} > 3$ cm⁻³ or $f_{\rm shield} < 10^{-2}$. Although these conditions do not cover the entire physical conditions of the simulations, it covers most of the physical conditions over which the chemistry is activated by the formation of H₂ not only provides the entire chemical composition of the diffuse neutral matter but also allows to estimate the impact of out-of-equilibrium molecular hydrogen on the chemistry of the ISM.

Element	Abundance	Eion
Н	1	13.59 eV
He	0.1	24.58 eV
Ν	$7.5 imes 10^{-5}$	14.53 eV
0	$3.19 imes 10^{-4}$	13.61 eV
C^+	1.32×10^{-4}	11.26 eV
S^+	1.86×10^{-5}	10.36 eV
Ar	3.29×10^{-6}	29.2 eV
Fe ⁺	$1.5 imes 10^{-8}$	7.90 eV
PAHs	4.189×10^{-7}	
grains	4.8×10^{-10}	

Table 13.1: The elements included in the chemical code, and their initial state (neutral or ionic). The elemental abundances take into account the solar abundances and a depletion factor. On the third column the first ionization potential of the neutral species in eV. Species with ionization potential lower than the Lyman limit (13.6 eV) are in initially their ionic state.

13.1 Reactions

The chemical solver takes as inputs not only the physical condition of each cell of the simulation but also an input file describing a chemical network, i.e. the γ_i , α_i , β_i , E_i for each of the reactions *i* included in the code. This network is used to create the system of equations which is then solved using the Newton-Raphson algorithm to compute the densities of all 151 species.

 γ_i , α_i , β_i , E_i are parameters given by quantum mechanical calculations or experiments that allow to compute the reaction rates k_i . Chemical reaction rates, k_i , for **two-body col-lisions** (e.g. ion-neutral, dissociative recombination, radiative association) are modeled with an Arrhenius-Kooij's law:

$$k_i = \gamma_i \left(\frac{T}{300}\right)^{\alpha_i} \exp\left(-E_i/k_B T\right)$$
(13.1)

an expressed in cm³s⁻¹, where *T* is the effective temperature of the collision between the two reactants. In the above formuma, E_i is the activation energy for endothermic reactions, while γ_i and α_i are the coefficients setting the amplitude and temperature dependence of the reaction rate. The total reaction rate between two species *j* and *k* with densities n_j and n_k in cm⁻³, reads $k_i n_j n_k$, and is expressed in cm⁻³s⁻¹.

Cosmic-ray processes, both direct or that include secondary photons are computed as

$$k_{\rm CR} = \gamma_i \zeta_{\rm H_2} \tag{13.2}$$

where k_{CR} is the reaction rate in s⁻¹ and ζ_{H_2} is the cosmic ray dissociation rate of molecular hydrogen, and γ_i is the scaling parameter for each atom and molecule compared to H₂.

Photoreaction rates are finally computed as

$$k_{\gamma} = \gamma_i G_0 \exp\left(-\beta_i A_V\right) \tag{13.3}$$

where k_{γ} is the rate in s⁻¹, G_0 the scaling factor of the standard ISRF and β_i describes the dependence of the rate on the visual extinction A_V .

Chapter 14

Cosmic-Rays Ionization Rate

Cosmic rays are high-energy particles presented in §1.3.4. Since low-energy cosmic rays are not observed directly, their ionization rate needs to be deduced indirectly, constrained by other chemical observers, e.g. H_3^+ , which is the most direct tracer of CR ionization, due to its rather simple chemistry. The formation of H_3^+ happens in two steps: CR photodissociate H_2 creating a H_2^+ ion that through a neutral-ion reaction with another H_2 creates H_3^+ . The H_2^+ formed by the first reaction can also be destroyed by dissociative recombination with electrons or charge transfer to atomic hydrogen, but both these processes are slow in comparison with the reaction that forms H_3^+ . The first step of the formation of H_3^+ is called the rate-limiting step because it is very slow compared to the second, and so its reaction rate can be taken as the formation rate of H_3^+ . For the destruction of H_3^+ the predominant process in the diffuse medium is the dissociative recombination with an electron, with a rate of k_e . If we assume steady state (0), the formation of H_3^+ it is equal to its destruction:

$$\zeta_{\rm H_2} \stackrel{(0)}{=} k({\rm H}_3^+|{\rm e}^-) \frac{n_e n({\rm H}_3^+)}{n({\rm H}_2)} \stackrel{(1)}{=} k({\rm H}_3^+|{\rm e}^-) \frac{n_e N({\rm H}_3^+)}{N({\rm H}_2)} \stackrel{(2)}{=} k({\rm H}_3^+|{\rm e}^-) \frac{n(C^+)N({\rm H}_3^+)}{N({\rm H}_2)}$$
(14.1)

where $k(H_3^+|e^-) \sim 2 \times 10^{-7} \text{ cm}^3 \text{ s}^{-1}$ is the dissociative recombination rate which measured experimentally by McCall et al. (2004), $N(H_3^+)$ and $N(H_2)$ are the column densities of H_2 and H_3^+ which can be observed. Equation (1) assumes that the ratio between the density of H_3^+ and H_2 does not change much on a line of sight and equation (2) takes into account that C⁺ is the main source of electrons because it has an ionization potential lower than Lyman limit and it is the most abundant element after hydrogen (Sofia et al. 2004).

Indriolo & McCall (2012) computed this calculation also for a more complete and complex chemical reaction networks taking into account the destruction of H₃⁺ by CO and O, showing that in diffuse clouds the previous prescription gives good results. McCall et al. (2003) found a CR primary ionization ratio for atomic hydrogen $\zeta_{\rm H} \sim 5 \times 10^{-16}$ along sightline to ζ Persei. Indriolo & McCall (2012) found that the mean value of the ionization rate is $\overline{\zeta_{\rm H_2}} \sim 3.5 \times 10^{-16} \, {\rm s}^{-1}$ with big variations around this mean value, values can vary greatly from one line to another from 10^{-17} to $10^{-15} \, {\rm s}^{-1}$, apparently due to small-scale and local effects, e.g. the distance between the acceleration the ionization site. $\zeta_{\rm H_2}$ was expected to decrease with increasing column density $N_{\rm H}$ (Padovani et al. 2009) because CR would lose their energy entering the cloud and ionizing more and more matter. However, no correlation seems to appear between $\zeta_{\rm H_2}$ and $N_{\rm H}$, but a drop happens once the cloud goes from diffuse to dense and the low-energy CR are not energetic enough to penetrate into the cloud and "stop" in the outer region of the cloud.

Chapter 15

Comparison of Distances derived from Hipparcos and Gaia

As shown in §8, the lengths of the neutral material sampled in an observational dataset is at the heart of the statistical and chemical comparison between the observational dataset and the predictions of the simulations. These lengths depend on the Galactic latitudes of the background sources but also, and evidently, on their distances. As explained in §4, the distances of the background sources are deduced from measurements of their parallaxes with the Gaia satellite, if available, or from the data collected by Gudennavar et al. (2012) and Jenkins & Tripp (2011) from the HIPPARCOS catalog. To check the validity of the data used in our successive samples, we compare here the distances deduced from Gaia to those given by Gudennavar et al. (2012) and Jenkins & Tripp (2011).



Figure 15.1: Correlation between the distances of the sources computed from the parallax from GAIA or taken from Gudennavar et al. (2012) catalog. The histograms of the distributions of those distances are also plotted. There are 270 out of the 360 sources that have both distances. Pearson's correlation coefficient (see eq. 15.1) is R = 0.90. The gray line is a 1:1 curve indicating perfect correlation. Figs. 15.1, 15.2, and 15.3 show these two sets of distances for the three observational samples studied in this work: the HI-to-H₂ transition, the observations of the excitation of neutral carbon, and the column densities of CH⁺ measured in absorption in the local diffuse gas. Each figure also shows the corresponding distribution of distances deduced from Gaia's data and from the previous studies cited above. Errors on the parallaxes were not extracted for the sample of the HI-to-H₂ transition but were extracted for the other samples, hence the error bars on the *x*-axis in Figs. 15.2 and 15.3.

Figure 15.2: Histograms of the distributions of distances computed from the parallax from GAIA (light blue) or taken from Jenkins & Tripp (2011) paper (gray). There are 88 out of the 89 sources that have both distances. Pearson's correlation coefficient (see eq. 15.1) is R = 0.90. The gray line is a 1:1 curve indicating a perfect correlation. T



For all the three samples, the distances derived from Gaia appear to follow those obtained with previous studies over about two orders of magnitude. Although the correlations are satisfactory, a large dispersion is observed around the perfect correlation (shown in black in all the figures). As a result, the distribution of distances deduced from Gaia and from previous studies are similar but not entirely equivalent. As shown in Figs. 15.2 and 15.3, the large deviation from the perfect correlation are linked to background sources with large errors on the parallax measured by Gaia (and probably by Hipparcos as well). Since the distances in all samples cover two orders of magnitude, we estimate the degrees of correlation and the dispersions on the logarithm of the distances rather than on the distances themselves. In other word, the standard Pearson's coefficient

$$R = \frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{\sigma_x} \right) \left(\frac{y_i - \bar{y}}{\sigma_y} \right),\tag{15.1}$$

is estimated from $x_i = \log(d_i)$ where d_i is the distance of the source *i*. σ_x , σ_y , \overline{x} , and \overline{y} are the standard deviations and mean values of x_i and y_i , respectively.

The correlation coefficients found for the three samples range from 0.83 to 0.90. This confirms the use of either the Gaia data or the distances given by previous studies to extract the distribution of lengths sampled by the observed lines of sight used for the reconstruction algorithm. The dispersion from the perfect correlation range between 0.38 and 0.47,



Figure 15.3: Correlation between the distances of the sources computed from the parallax from GAIA or taken from Gudennavar et al. (2012) paper. There are 151 out of the 190 sources that have both distances. Pearson's correlation coefficient (see eq. 15.1) is R = 0.83. The gray line is a 1:1 curve indicating perfect correlation.

which corresponds to an uncertainty on the distances of about a factor of 2.5. This level of uncertainty is the reason why the width of the bins of distances used for the reconstruction algorithm (see §8) is set to a factor of two, or equivalently to ~ 0.3 in logspace.

Bibliography

- Abgrall, H., Le Bourlot, J., Pineau des Forêts, G., et al. 1992, A&A, 253, 525 33
- Allen, M. M. 1994, ApJ, 424, 754 40
- Ambartsumian, V. A. 1947, The evolution of stars and astrophysics 28
- Audit, E. & Hennebelle, P. 2005, A&A, 433, 1 55, 66
- Bakes, E. L. O. & Tielens, A. G. G. M. 1994, ApJ, 427, 822 16, 19
- Barinovs, G. & Hemert, M. C. v. 2006, ApJ, 636, 923 163
- Barnstedt, J., Gringel, W., Kappelmann, N., & Grewing, M. 2000, A&AS, 143, 193 31
- Bellomi, E., Godard, B., Hennebelle, P., et al. 2020, A&A, 643, A36 126, 167, 173, 179
- Berger, M. J. & Oliger, J. 1984, Journal of Computational Physics, 53, 484 46
- Bialy, S., Burkhart, B., & Sternberg, A. 2017, ApJ, 843, 92 34, 49, 59
- Blair, W. P., Long, K. S., & Raymond, J. C. 1996, ApJ, 468, 871 30
- Bohlin, R. C., Hill, J. K., Jenkins, E. B., et al. 1983, ApJS, 51, 277 31
- Bohlin, R. C., Savage, B. D., & Drake, J. F. 1978, ApJ, 224, 132 28, 31
- Bouy, H. & Alves, J. 2015, A&A, 584, A26 29
- Bron, E., Le Bourlot, J., & Le Petit, F. 2014, A&A, 569, A100 33, 57
- Browning, M. K., Tumlinson, J., & Shull, J. M. 2003, ApJ, 582, 810 35
- Carruthers, G. 1970, Space Sci Rev, 10 30
- Cartledge, S. I. B., Lauroesch, J. T., Meyer, D. M., & Sofia, U. J. 2004, ApJ, 613, 1037 31
- Christensen, C., Quinn, T., Governato, F., et al. 2012, MNRAS, 425, 3058 34
- Cooke, R. J., Pettini, M., Jorgenson, R. A., Murphy, M. T., & Steidel, C. C. 2014, ApJ, 781, 31 20
- Crane, P., Lambert, D. L., & Sheffer, Y. 1995, ApJS, 99, 107 40, 43
- Dalgarno, A. & Roberge, W. G. 1979, ApJ, 233, L25 57

- Dame, T. M., Hartmann, D., & Thaddeus, P. 2001, ApJ, 547, 792 29
- de Avillez, M. A. & Breitschwerdt, D. 2004, A&A, 425, 899 106
- de Boer, K. S., Lenhart, H., van der Hucht, K. A., et al. 1986, A&A, 157, 119 31
- Dickey, J. M. & Lockman, F. J. 1990, ARA&A, 28, 215 29
- Diemer, B., Stevens, A. R. H., Forbes, J. C., et al. 2018, ApJS, 238, 33 34
- Diplas, A. & Savage, B. D. 1994, ApJS, 93, 211 31
- Douglas, A. E. & Herzberg, G. 1941, ApJ, 94, 381 39
- Draine, B. T. 2011, Physics of the Interstellar and Intergalactic Medium 10, 13, 24, 72, 74, 79
- Draine, B. T. & Bertoldi, F. 1996, ApJ, 468, 269 27, 58, 59, 126, 127
- Draine, B. T. & Katz, N. 1986, ApJ, 310, 392 42
- Draine, B. T. & Salpeter, E. E. 1978, Nature, 271, 730 26, 27, 62
- Elmegreen, B. G. & Elmegreen, D. M. 1987, ApJ, 320, 182 29
- Elmegreen, B. G. & Scalo, J. 2004, ARA&A, 42, 211 12
- Falgarone, E., Godard, B., Cernicharo, J., et al. 2010, A&A, 521, L15 43
- Falgarone, E., Pineau des Forêts, G., & Roueff, E. 1995, A&A, 300, 870 42, 184
- Falgarone, E. & Puget, J. L. 1995, A&A, 293, 840 17
- Falgarone, E., Verstraete, L., Pineau Des Forêts, G., & Hily-Blant, P. 2005, A&A, 433, 997 42
- Fasano, G. & Franceschini, A. 1987, MNRAS, 225, 155 110
- Federman, S. R. 1982, ApJ, 257, 125 40
- Federman, S. R. 1983, ApJ, 269, 5 19
- Federman, S. R., Glassgold, A. E., & Kwan, J. 1979, ApJ, 227, 466 58
- Federrath, C., Chabrier, G., Schober, J., et al. 2011, Phys. Rev. Lett., 107, 114504 54
- Federrath, C., Roman-Duval, J., Klessen, R. S., Schmidt, W., & Mac Low, M.-M. 2010, A&A, 512, A81 55, 66
- Federrath, C., Schrön, M., Banerjee, R., & Klessen, R. S. 2014, ApJ, 790, 128 54
- Ferlet, R., Vidal-Madjar, A., & Gry, C. 1985, ApJ, 298, 838 31
- Field, G. B. 1965, ApJ, 142, 531 48
- Fitzpatrick, E. L. 1999, PASP, 111, 63 28

- Fitzpatrick, E. L. & Massa, D. 1986, ApJ, 307, 286 49
- Fitzpatrick, E. L. & Massa, D. 1990, ApJS, 72, 163 31
- Florescu-Mitchell, A. I. & Mitchell, J. B. A. 2006, Physics Reports, 430, 277 164
- Flower, D. R. 1998, MNRAS, 297, 334 82
- Flower, D. R., Roueff, E., & Zeippen, C. J. 1998, J. Phys. B: At. Mol. Phys., 31, 1105 82
- Froese Fischer, C. & Saha, H. P. 1985, Physica Scripta, 32, 181 37, 71
- Fromang, S., Hennebelle, P., & Teyssier, R. 2006, A&A, 457, 371 46
- Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1 97
- Gerlich, D. 1990, Journal of Chemical Physics, 92, 2377 83, 84
- Gillmon, K., Shull, J. M., Tumlinson, J., & Danforth, C. 2006, ApJ, 636, 891 31, 33, 35
- Glover, S. C. O. & Clark, P. C. 2014, MNRAS, 437, 9 20
- Glover, S. C. O., Federrath, C., Mac Low, M.-M., & Klessen, R. S. 2010, MNRAS, 404, 2 34, 49
- Glover, S. C. O. & Mac Low, M.-M. 2007, ApJS, 169, 239 58
- Gnat, O. & Ferland, G. J. 2012, ApJS, 199, 20 18, 21
- Gnedin, N. Y., Tassis, K., & Kravtsov, A. V. 2009, ApJ, 697, 55 34, 35
- Godard, B., Falgarone, E., & Forêts, G. P. D. 2014, A&A, 570, A27 42, 43, 184
- Godard, B., Falgarone, E., Gerin, M., et al. 2012, A&A, 540, A87 43
- Godard, B., Falgarone, E., & Pineau des Forêts, G. 2009, A&A, 495, 847 42, 184
- Godard, B., Pineau des Forêts, G., Lesaffre, P., et al. 2019, A&A, 622, A100 42
- Goldbaum, N. J., Krumholz, M. R., Matzner, C. D., & McKee, C. F. 2011, ApJ, 738, 101 54
- Goldsmith, P. F., Li, D., & Krčo, M. 2007, ApJ, 654, 273 33, 55
- Goldsmith, P. F., Velusamy, T., Li, D., & Langer, W. 2009, ASP Conf. Ser., 417, 177 32
- Gredel, R. 1997, A&A, 320, 929 40, 43
- Gredel, R., Pineau des Forêts, G., & Federman, S. R. 2002, A&A, 389, 993 40
- Gry, C., Boulanger, F., Nehme, C., et al. 2002, A&A, 391, 675 42
- Gudennavar, S. B., Bubbly, S. G., Preethi, K., & Murthy, J. 2012, ApJS, 199, 8 30, 32, 195, 197
- Habing, H. J. 1968, Bull. Astron. Inst. Netherlands, 19, 421 27, 46
- Haud, U. & Kalberla, P. M. W. 2007, A&A, 466, 555 66

- Hennebelle, P. & Falgarone, E. 2012, A&A Rev., 20, 55 66
- Henshaw, J. D., Ginsburg, A., Haworth, T. J., et al. 2019, MNRAS, 485, 2457 66
- Hierl, P. M., Morris, R. A., & Viggiano, A. A. 1997, J. Chem. Phys., 106, 10145 163
- Hill, A. S., Mac Low, M.-M., Gatto, A., & Ibáñez-Mejía, J. C. 2018, ApJ, 862, 55 97, 108
- Hollenbach, D. & McKee, C. F. 1979, ApJS, 41, 555 20
- Hollenbach, D. & McKee, C. F. 1989, ApJ, 342, 306 19, 20
- Hu, C.-Y., Naab, T., Walch, S., Glover, S. C. O., & Clark, P. C. 2016, MNRAS, 458, 3528 34, 49
- Indriolo, N. & McCall, B. J. 2012, ApJ, 745, 91 193
- Jenkins, E. B., Jura, M., & Loewenstein, M. 1983, ApJ, 270, 88 36
- Jenkins, E. B. & Peimbert, A. 1997, ApJ, 477, 265 31
- Jenkins, E. B. & Shaya, E. J. 1979, ApJ, 231, 55 36
- Jenkins, E. B. & Tripp, T. M. 2001, ApJS, 137, 297 36, 37
- Jenkins, E. B. & Tripp, T. M. 2011, ApJ, 734, 65 36, 37, 38, 39, 72, 73, 74, 76, 176, 179, 182, 183, 184, 187, 195, 196
- Jenkins, E. B., Tripp, T. M., Fitzpatrick, E. L., et al. 1998, ApJ, 492, L147 36
- Jenkins, E. B. & Wallerstein, G. 1995, ApJ, 440, 227 36
- Jensen, A. G. & Snow, T. P. 2007a, ApJ, 669, 378 31
- Jensen, A. G. & Snow, T. P. 2007b, ApJ, 669, 401 31
- Johnson, C. T., Burke, P. G., & Kingston, A. E. 1987, J. Phys. B: At. Mol. Phys., 20, 2553 37, 71
- Joulain, K., Falgarone, E., Pineau des Forêts, G., & Flower, D. 1998, A&A, 340, 241 17, 42, 184
- Joung, M. K. R. & Mac Low, M.-M. 2006, ApJ, 653, 1266 48
- Kalberla, P. M. W., Burton, W. B., Hartmann, D., et al. 2005, A&A, 440, 775 66
- Kalberla, P. M. W. & Kerp, J. 2009, ARA&A, 47, 27 29, 62
- Khokhlov, A. M. 1998, Journal of Computational Physics, 143, 519 46
- Klessen, R. S. & Hennebelle, P. 2010, A&A, 520, A17 54
- Kolmogorov, A. 1941, Akademiia Nauk SSSR Doklady, 30, 301 54
- Korpi, M. J., Brandenburg, A., Shukurov, A., Tuominen, I., & Nordlund, A. 1999, ApJ, 514, L99 66

- Krumholz, M. R., Matzner, C. D., & McKee, C. F. 2006, ApJ, 653, 361 54
- Krumholz, M. R., McKee, C. F., & Tumlinson, J. 2008, ApJ, 689, 865 33, 35
- Kuijken, K. & Gilmore, G. 1989, MNRAS, 239, 571 48
- Körtgen, B., Federrath, C., & Banerjee, R. 2019, MNRAS, 482, 5233 35
- Körtgen, B., Seifried, D., Banerjee, R., Vázquez-Semadeni, E., & Zamora-Avilés, M. 2016, Mon Not R Astron Soc, 459, 3460 54
- Larson, R. B. 1981, MNRAS, 194, 809 66
- Launay, J. M. & Roueff, E. 1977, A&A, 56, 289 37, 71
- Le Bourlot, J., Le Petit, F., Pinto, C., Roueff, E., & Roy, F. 2012, A&A, 541, A76 33, 56
- Le Bourlot, J., Pineau des Forêts, G., & Flower, D. R. 1999, MNRAS, 305, 802 85
- Le Petit, F., Nehmé, C., Le Bourlot, J., & Roueff, E. 2006, ApJS, 164, 506 25, 178
- Lee, H.-H., Herbst, E., Pineau des Forêts, G., Roueff, E., & Le Bourlot, J. 1996, A&A, 311, 690 33, 59
- Lee, Y.-N. & Hennebelle, P. 2016, A&A, 591, A30 54
- Lehner, N., Jenkins, E. B., Gry, C., et al. 2003, ApJ, 595, 858 31
- Lequeux, J. 2005, The Interstellar Medium 10
- Leroy, A., Bolatto, A., Stanimirovic, S., et al. 2007, ApJ, 658, 1027 35
- Lesaffre, P., Gerin, M., & Hennebelle, P. 2007, A&A, 469, 949 33, 42
- Lesaffre, P., Pineau des Forêts, G., Godard, B., et al. 2013, A&A, 550, A106 42, 184
- Lesaffre, P., Todorov, P., Levrier, F., et al. 2020, MNRAS, 495, 816 184
- Li, M., Ostriker, J. P., Cen, R., Bryan, G. L., & Naab, T. 2015, ApJ, 814, 4 106
- Li, Z.-Y. & Nakamura, F. 2006, ApJ, 640, L187 54
- Mac Low, M.-M. & Klessen, R. S. 2004, Reviews of Modern Physics, 76, 125 66
- Marchal, A., Miville-Deschênes, M.-A., Orieux, F., et al. 2019, A&A, 626, A101 60
- Martin, P. G. & Mandy, M. E. 1995, ApJ, 455, L89 83
- Mathis, J. S., Mezger, P. G., & Panagia, N. 1983, A&A, 128, 212 25, 26, 27, 62, 74
- Mathis, J. S., Rumpl, W., & Nordsieck, K. H. 1977, ApJ, 217, 425 9
- Matzner, C. D. 2002, ApJ, 566, 302 54
- Maíz-Apellániz, J. 2001, AJ, 121, 2737 29

- McCall, B. J., Huneycutt, A. J., Saykally, R. J., et al. 2004, Phys. Rev. A, 70, 052716 193
- McCall, B. J., Huneycutt, A. J., Saykally, R. J., et al. 2003, Nature, 422, 500 193
- McKee, C. F. & Krumholz, M. R. 2010, ApJ, 709, 308 33
- McKee, C. F. & Ostriker, E. C. 1977 66, 106
- Meyer, D. M., Jura, M., & Cardelli, J. A. 1998, ApJ, 493, 222 19
- Miville-Deschenes, M.-A. & Martin, P. G. 2007, A&A, 469, 189 66
- Miville-Deschênes, M.-A., Murray, N., & Lee, E. J. 2017, ApJ, 834, 57 62
- Nakanishi, H. & Sofue, Y. 2016, PASJ, 68, 5 62
- Nickerson, S., Teyssier, R., & Rosdahl, J. 2018, MNRAS, 479, 3206 34, 49
- Norman, C. A. & Ferrara, A. 1996, ApJ, 467, 280 54
- Padoan, P., Juvela, M., Pan, L., Haugbølle, T., & Nordlund, A. 2016a, ApJ, 826, 140 54
- Padoan, P., Pan, L., Haugboelle, T., & Nordlund, A. 2016b, ApJ, 822, 11 54
- Padovani, M., Galli, D., & Glassgold, A. E. 2009, A&A, 501, 619 193
- Palla, F., Salpeter, E. E., & Stahler, S. W. 1983, ApJ, 271, 632 56
- Pan, K., Federman, S. R., Cunha, K., Smith, V. V., & Welty, D. E. 2004, ApJS, 151, 313 31
- Pan, L. & Padoan, P. 2009, ApJ, 692, 594 17
- Pan, L., Padoan, P., Haugbølle, T., & Nordlund, A. 2016, ApJ, 825, 30 54
- Peacock, J. A. 1983, MNRAS, 202, 615 110
- Perryman, M. A. C., Lindegren, L., Kovalevsky, J., et al. 1997, A&A, 323, L49 97
- Pineau des Forêts, G., Flower, D. R., Hartquist, T. W., & Dalgarno, A. 1986, MNRAS, 220, 801 42
- Piontek, R. A. & Ostriker, E. C. 2005, ApJ, 629, 849 63, 179
- Rachford, B. L., Snow, T. P., Destree, J. D., et al. 2009, ApJS, 180, 125 31, 35
- Rachford, B. L., Snow, T. P., Tumlinson, J., et al. 2002a, ApJ, 577, 221 31
- Rachford, B. L., Snow, T. P., Tumlinson, J., et al. 2002b, ApJ, 577, 221 31
- Reynolds, O. 1895, Philosophical Transactions of the Royal Society of London. (A.), 186, 123 47
- Roueff, E. & Le Bourlot, J. 1990, A&A, 236, 515 37, 71

- Saury, E., Miville-Deschênes, M.-A., Hennebelle, P., Audit, E., & Schmidt, W. 2014, A&A, 567, A16 39, 55, 66, 179, 182
- Savage, B. D., Bohlin, R. C., Drake, J. F., & Budich, W. 1977, ApJ, 216, 291 31, 35
- Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525 31
- Schmidt, W., Federrath, C., Hupp, M., Kern, S., & Niemeyer, J. C. 2009, A&A, 494, 127 55
- Schroder, K., Staemmler, V., Smith, M. D., Flower, D. R., & Jaquet, R. 1991, J. Phys. B: At. Mol. Opt. Phys., 24, 2487 37, 71
- Schultz, G. V. & Wiemer, W. 1975, A&A, 43, 133 28
- Seifried, D., Schmidt, W., & Niemeyer, J. C. 2011, A&A, 526, A14 55, 63, 179
- Sembach, K. R., Howk, J. C., Ryans, R. S. I., & Keenan, F. P. 2000, ApJ, 528, 310 11, 19
- Sheffer, Y., Rogers, M., Federman, S. R., et al. 2008, ApJ, 687, 1075 40
- Siegbahn, P. & Liu, B. 1978, Journal of Chemical Physics, 68, 2457 83
- Smith, A. M., Bruhweiler, F. C., Lambert, D. L., et al. 1991, ApJ, 377, L61 36
- Snow, T. P. & McCall, B. J. 2006, ARA&A, 44, 367 11, 15
- Sofia, U. J., Cardelli, J. A., Guerin, K. P., & Meyer, D. M. 1997, ApJ, 482, L105 19
- Sofia, U. J., Lauroesch, J. T., Meyer, D. M., & Cartledge, S. I. B. 2004, ApJ, 605, 272 193
- Spitzer, L. 1978, Physical processes in the interstellar medium 18
- Spitzer, L., Drake, J. F., Jenkins, E. B., et al. 1973, ApJ, 181, L116 42
- Spitzer, Jr., L. & Cochran, W. D. 1973, ApJ, 186, L23 42
- Spitzer, Jr., L., Cochran, W. D., & Hirshfeld, A. 1974, ApJS, 28, 373 42
- Staemmler, V. & Flower, D. R. 1991, J. Phys. B: At. Mol. Opt. Phys., 24, 2343 37, 71
- Sternberg, A. 1988, ApJ, 332, 400 33
- Sternberg, A., Le Petit, F., Roueff, E., & Le Bourlot, J. 2014, ApJ, 790, 10 33, 34, 58
- Strong, A. W., Moskalenko, I. V., & Ptuskin, V. S. 2007, Annual Review of Nuclear and Particle Science, 57, 285 13
- Sur, S., Federrath, C., Schleicher, D. R. G., Banerjee, R., & Klessen, R. S. 2012, Mon Not R Astron Soc, 423, 3148 54
- Sutherland, R. S. & Dopita, M. A. 1993, ApJS, 88, 253 21
- Tabone, B. 2018, PhD thesis 55

- Tabone, B., Godard, B., Pineau des Forêts, G., Cabrit, S., & van Dishoeck, E. F. 2020, A&A, 636, A60 56
- Teyssier, R. 2002, A&A, 385, 337 46
- Thompson, R., Nagamine, K., Jaacks, J., & Choi, J.-H. 2014, ApJ, 780, 145 34
- Valdivia, V., Godard, B., Hennebelle, P., Gerin, M., & Lesaffre, P. 2016a, 309 167
- Valdivia, V., Godard, B., Hennebelle, P., et al. 2017, A&A, 600, A114 42, 75, 167, 190
- Valdivia, V. & Hennebelle, P. 2014, A&A, 571, A46 49, 55, 58
- Valdivia, V., Hennebelle, P., Gérin, M., & Lesaffre, P. 2016b, A&A, 587, A76 34, 35, 49, 55
- van Dishoeck, E. F. & Black, J. H. 1986, ApJS, 62, 109 33
- van Dishoeck, E. F. & Black, J. H. 1988, ApJ, 334, 771 163
- Vázquez-Semadeni, E., Colín, P., Gómez, G. C., Ballesteros-Paredes, J., & Watson, A. W. 2010, ApJ, 715, 1302 54
- Wakker, B. P., Savage, B. D., Sembach, K. R., et al. 2003, ApJS, 146, 1 31
- Walch, S., Girichidis, P., Naab, T., et al. 2015, MNRAS, 454, 238 106
- Walch, S., Wünsch, R., Burkert, A., Glover, S., & Whitworth, A. 2011, ApJ, 733, 47 63, 179
- Wang, P., Li, Z.-Y., Abel, T., & Nakamura, F. 2010, ApJ, 709, 27 54
- Webber, W. R. 1998, ApJ, 506, 329 13
- Weselak, T., Galazutdinov, G. A., Musaev, F. A., & Krełowski, J. 2008, A&A, 484, 381 40, 43
- Wolcott-Green, J., Haiman, Z., & Bryan, G. L. 2011, MNRAS, 418, 838 58, 59, 126, 127, 128
- Wolfire, M. G., Hollenbach, D., McKee, C. F., Tielens, A. G. G. M., & Bakes, E. L. O. 1995, ApJ, 443, 152 14
- Wolfire, M. G., McKee, C. F., Hollenbach, D., & Tielens, A. G. G. M. 2003, ApJ, 587, 278 14, 15, 19, 22, 49, 52, 53
- Wrathmall, S. A. & Flower, D. R. 2007, J. Phys. B: At. Mol. Opt. Phys., 40, 3221 82, 83
- Xie, T., Allen, M., & Langer, W. D. 1995, ApJ, 440, 674 42
- Zari, E., Hashemi, H., Brown, A. G. A., Jardine, K., & de Zeeuw, P. T. 2018, A&A, 620, A172 29
RÉSUMÉ

Le milieu interstellaire local (MIS) est un système ouvert et complexe, gouverné par les effets combinés de la turbulence, du champ magnétique, de la gravité, et des photons ultraviolets. La quantité croissante de données collectées par les spectromètres de longueurs d'onde radio à l'ultraviolet (UV) ouvre une nouvelle ère où les informations statistiques et chimiques contenues dans les observations peuvent être utilisées de manière concomitante afin de comprendre les rôles spécifiques de ces différents réservoirs d'énergie.

Cette thèse vise à comprendre les processus physiques régissant l'évolution du MIS diffus local et sa composition chimique 3D. Pour quantifier les impacts de la densité moyenne, du champ de rayonnement UV, de l'échelle intégrale, de la résolution, du forçage turbulent, du champ magnétique et de la gravité sur les observables, nous avons utilisé des simulations MHD de pointe. L'abundance de l'hydrogène moléculaire est calculée hors équilibre dans la simulation, et un solveur chimique est appliqué en post-traitement pour calculer les abondances chimiques d'autres espèces. Pour comparer les résultats des simulations aux observations, nous proposons une nouvelle version du test de Kolmogorov-Smirnov qui est un outil précieux pour estimer la distance entre deux distributions de probabilité. L'originalité de ce travail est de se concentrer non seulement sur la simulation de grands échantillons de densités de colonne individuelles mais aussi sur leur statistique, c'est-à-dire les probabilités d'occurrence de ces densités de colonne le long de lignes de visée aléatoires. Dans ce travail, nous nous concentrons sur trois observables chimiques: H₂, CH⁺ et CI. Nous constatons que la transition de l'hydrogène atomique à l'hydrogène moléculaire dépend fortement de la densité du plan médian galactique, de la densité des étoiles OB et de l'échelle des nuages diffus neutres. Ces paramètres affectent grandement la distribution des densités de colonne de CH⁺ qui dépendent également de l'amplitude du forçage turbulent. Si l'action conjointe de l'instabilité thermique et de la turbulence peut reproduire les densités de colonne observées de H, H₂ et CH⁺, elle n'est pas en mesure de reproduire les observations des populations des niveaux de structure fine du carbone neutre.

L'approche pour comparer les simulations et les observations présentée dans cette thèse utilise deux observables mais peut être étendue à N dimensions, en utilisant plusieurs observables simultanément.

MOTS CLÉS

Milieu interstellaire diffus, phases, turbulence, simulations, hydrogen moleculaire, RAMSES, astrochimie

ABSTRACT

The local interstellar medium (ISM) is an open and complex system, driven by the combined effects of turbulence, magnetic field, gravity, and ultraviolet photons. The rising amount of data collected by spectrometers from radio to ultraviolet (UV) wavelengths opens a new era where the statistical and chemical information contained in the observations can be used concomitantly to understand the specific role of each of these energy inputs.

This Ph.D. aims to understand the physical processes governing the evolution of the local diffuse ISM and its 3D chemical composition. To quantify the impacts of the mean density, the UV radiation field, the integral scale, the resolution, the turbulent forcing, the magnetic field, and the gravity on the observables, we used state-of-the-art MHD simulations. The molecular hydrogen abundance is computed on-the-fly in the simulation, and a chemical solver is applied in post-processing to compute the chemical abundances of other species. To compare the results of the simulations to the observational sample, we propose a new version of the Kolmogorov-Smirnov test which is a valuable tool for estimating the distance between two probability distribution functions. The original feature of this work is to not only focus on the simulation of large samples of individual column densities but also on their statistics, meaning the probabilities of occurrence of these column densities along random lines of sight.

In this work, we focus on three chemical observables: H_2 , CH^+ , and CI. We find that the transition from atomic to molecular hydrogen strongly depends on the Galactic midplane density, the density of OB stars, and the scale of neutral diffuse clouds. These parameters greatly affect the distribution of column densities of CH^+ which also depends on the strength of the turbulent forcing. While the joint action of the thermal instability and the turbulence can reproduce the observed column densities of H, H₂, and CH⁺, it is unable to reproduce the observations of the fine-structure level populations of neutral carbon. The approach to compare simulations and observables presented in this Ph.D. uses two observables but it can be extended to N-dimensions, using multiple observables simultaneously.

KEYWORDS

Diffuse interstellar medium, phases, turbulence, simulations, molecular hydrogen, RAMSES, astrochemistry