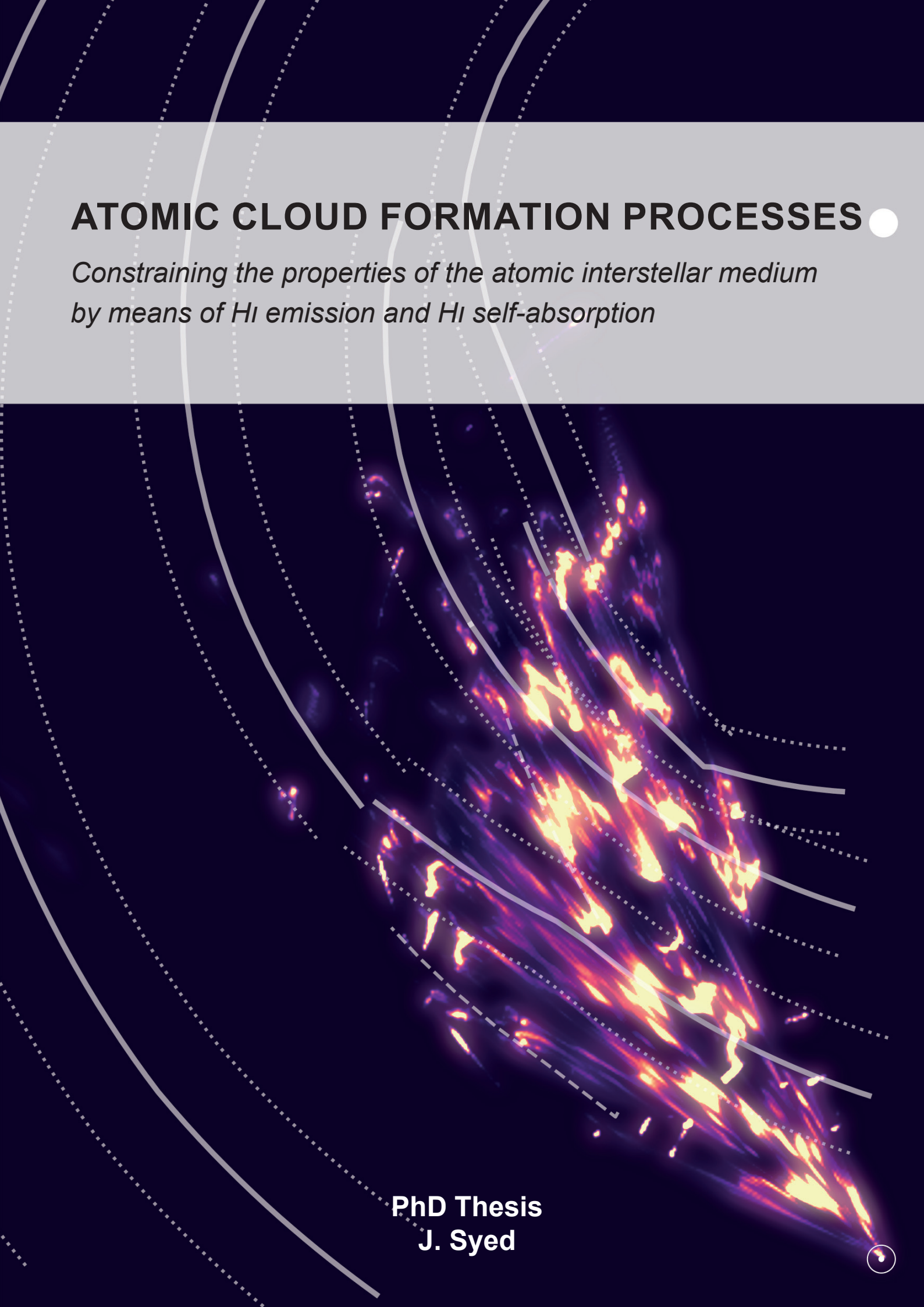
ATOMIC CLOUD FORMATION PROCESSES

Constraining the properties of the atomic interstellar medium by means of HI emission and HI self-absorption



PhD Thesis
J. Syed

Dissertation

submitted to the

Combined Faculty of Mathematics, Engineering and Natural Sciences

of Heidelberg University, Germany

for the degree of

Doctor of Natural Sciences

Put forward by

Jonas Syed

born in: Nürnberg, Germany

Oral examination: 13.07.2023

Atomic cloud formation processes –
Constraining the properties of the atomic interstellar medium
by means of H I emission and H I self-absorption

Referees

Prof. Dr. Henrik Beuther
Prof. Dr. Cornelis P. Dullemond

ABSTRACT

Atomic hydrogen gas (H I) is an integral constituent of the interstellar medium (ISM) and thus plays a critical role in the assembly of molecular clouds, the sites of star formation. An important physical agent in controlling the transition from atomic to molecular gas is cold H I. H I emission, that traces a wide range of hydrogen properties, is found throughout the Galactic plane and exhibits complex kinematic signatures that are imposed by the Galactic rotation. The unfavorable viewing geometry from our vantage point in the Galaxy requires a new set of tools that allows us to disentangle H I structures along our lines of sight. In this thesis, I constrain the properties of the interstellar hydrogen in our Milky Way galaxy and present a way to isolate cold hydrogen from H I emission using H I self-absorption (HISA). The outer Galaxy offers a less confused view on H I emission, and we exploit this circumstance with unprecedented detail using the high-angular resolution data of The H I/OH/Recombination line survey of the inner Milky Way (THOR). We discover the "Maggie" filament, one of the largest, almost purely atomic filaments in the Milky Way. Maggie has a length of 1.2 kpc and is located on the far side of the Galaxy at a distance of 12 kpc from the Galactic center. Optical depth measurements suggest that Maggie is in a largely cold H I phase and molecular gas is only found on the smallest spatial scales. When targeting molecular clouds in the inner Galactic plane, we detect pronounced HISA as a tracer of cold hydrogen. While the kinematic correlation is less surprising due to the selection bias of our sample, the spatial distribution of cold H I gas is also found to be well correlated with that of the molecular gas. The detection of HISA furthermore extends to the whole of our survey. We frequently find absorption signatures without any associated molecular counterpart. While cold atomic gas is rendered vitally important to the formation process of molecular clouds, the cold phase of atomic hydrogen marks a distinct gas phase in the ISM that exists throughout the Galaxy, independent of the occurrence of molecular gas.

ZUSAMMENFASSUNG

Atomarer Wasserstoff (H I) ist ein wesentlicher Bestandteil des interstellaren Mediums (ISM) und spielt somit eine entscheidende Rolle bei der Entstehung von Molekülwolken, den Geburtsstätten von Sternen. Ein wichtiger physikalischer Faktor bei dem Übergang von atomarem zu molekularem Gas ist kaltes H I. H I-Emission, welche ein breites Spektrum an physikalischen Eigenschaften von Wasserstoff nachweist, durchdringt das gesamte interstellare Medium und zeigt komplexe kinematische Signaturen, die nicht zuletzt durch die galaktische Rotation verursacht werden. Unsere Position innerhalb der Milchstraßenebene erfordert dedizierte Methoden, die es ermöglichen, H I-Strukturen entlang unserer Sichtlinien zu entflechten. In dieser Arbeit befasse ich mich mit den Eigenschaften des interstellaren Wasserstoffs in unserer Milchstraße und präsentiere eine neue Methode, kalten Wasserstoff mit Hilfe von H I-Selbstabsorption (HISA) zu isolieren. Dank hochauflösender Daten der THOR-Durchmusterung (The H I/OH/Recombination line survey of the inner Milky Way) untersuchen wir die entlegenen Regionen der Milchstraße auf Wasserstoffstrukturen. Wir entdecken das „Maggie“-Filament, eine der größten, fast rein atomaren Wolken in der Milchstraße. Maggie hat eine Länge von 1,2 kpc und befindet sich auf der anderen Seite der Galaxie in einer Entfernung von 12 kpc vom galaktischen Zentrum. Absorptions-Messungen legen nahe, dass sich Maggie zu einem großen Teil aus kaltem H I zusammensetzt, und molekulares Gas wird nur auf den kleinsten räumlichen Skalen sichtbar. Bei der Untersuchung von molekularen Wolken in der inneren Galaxieebene entdecken wir ausgeprägte HISA-Signaturen als Tracer für kalten Wasserstoff. Während die kinematische Korrelation aufgrund des Samples weniger überraschend ist, zeigt die räumliche Verteilung des kalten H I-Gases eine signifikante Korrelation mit der des molekularen Gases. Die Detektion von HISA erstreckt sich zudem über die gesamte Milchstraßenebene. Wir finden oft Absorptionssignaturen ohne zugehöriges molekulares Gegenstück. Während kaltes atomares Gas im Entstehungsprozess molekularer Wolken von entscheidender Bedeutung ist, repräsentiert die kalte Phase des atomaren Wasserstoffs eine deutlich ausgeprägte Gasphase, die sich unabhängig von molekularem Gas über die gesamte Galaxie erstreckt.

Contents

1

Introduction

"More light!" Those are supposed to have been Johann Wolfgang von Goethe's last words on his deathbed. What exactly he meant by this sentiment will always remain an open question. In the scientific field of astronomy, light is the essential source of information. Astronomers always strive for "more light" when proposing for observations, to make progress in unraveling the mysteries of the cosmos. Advancements in the development of new instruments and telescopes allow us to study our Universe with ever-growing precision and sensitivity. Like an astronomical newspaper does the light tell us about the physical processes acting within different astrophysical environments, and by reading this light we can push the frontiers of our scientific understanding of the Universe. However, it has become clear that new insights from experiments may not only confirm the theories we have postulated through observations and simulations but it may also unveil the limits of our understanding. It is central to the nature of scientific research to test the limitations of our theories and models. As new observations might reveal unprecedented details, astrophysical processes that have been overlooked in the past might become important to factor in, in order to fully capture a comprehensive picture of the Universe.

In this thesis, the physical mechanism and dynamical behavior of the interstellar medium (ISM) are of central interest, in particular the atomic gas phase of the ISM. The ISM is the galactic reservoir of gas and dust out of which molecular clouds form. It has been known for a long time that these molecular clouds, that mainly consist of cold molecular hydrogen (H_2) and dust, are the sites in which star formation takes place. These dense clouds are embedded in a more diffuse atomic state of the ISM. Once a star is born, it will inject energy and momentum back into the ISM, through processes that are generically called feedback, and thus shape a next generation of clouds and stars. Massive stars emit large amounts of very energetic ultraviolet photons and create an ionized gas phase in their surroundings. This complex interplay between gas, dust, and radiation is what constitutes the interstellar matter cycle, and the dynamical and physical relationship between different phases of the ISM is what sets the conditions under which star formation takes place. It is therefore

critical to have a consistent understanding of the interstellar matter cycle.

Characterizing the physical and dynamical relationship between different constituents of the ISM is a complex problem and spans a wide range of spatial and temporal scales. Local environments of star formation are ultimately linked to the large-scale dynamics of a galaxy. It follows that the highly dynamical system of a galaxy is connected to microphysical processes that allow gas in the ISM to cool and collapse locally, giving rise to the formation of stars.

How exactly does the large-scale structure in the ISM, that is dominated by the galactic dynamics of diffuse atomic hydrogen gas, influence the formation processes of molecular clouds and ultimately stars? This question poses a challenging task to tackle observationally, as will become apparent in the following sections. The atomic hydrogen (H I) of the ISM is an important ingredient in the global star formation process, as it is by far the most abundant element in the ISM (and our Universe) and it provides the raw material to give birth to stars. However, the atomic gas undergoes many intermediate steps from its large-scale diffuse nature down to the spatial scales of a single stellar system. First, atomic gas must be able to cool and form regions sufficiently dense and well-shielded against the dissociating effects of interstellar radiation coming from stars and stellar activity. This allows hydrogen atoms to bind and form molecular hydrogen clouds. The transition from the atomic gas phase of hydrogen to its molecular form is observationally still poorly constrained.

Systematic limitations in observing atomic and molecular hydrogen clouds in our own Milky Way galaxy make it difficult to constrain the properties of these gas phases, independent of the ever-growing capabilities of observatories. Determining the structure of the Milky Way from our vantage point is a generally challenging task. As we are sitting in the Galactic plane at a distance of 8.15 kpc (Reid et al., 2019) from the Galactic center, we are looking edge-on through the Milky Way plane when targeting the global structure of the ISM. Particularly the diffuse nature of H I makes it difficult to disentangle individual atomic gas structures along our lines of sight. Although angular resolution has significantly improved over the last decades, only the largest spatial scales (>100 pc) of external galaxies can be resolved by atomic hydrogen observations (e.g., Walter et al., 2008). It it thus a long-standing question what the physical and dynamical relationship between atomic and molecular hydrogen is and how the atomic gas phase drives the formation of molecular clouds.

In order to constrain the formation of molecular clouds out of the atomic phase of the Galactic ISM, it is required to have a deeper understanding of the composition of the ISM and its global structure in the Galaxy. The following subsections will briefly reiterate relevant aspects of our current understanding of the Galactic ISM.

1.1 The interstellar medium

This section discusses relevant aspects of the Galactic ISM and briefly summarizes important characteristics of different constituents of the ISM. The global structure of the atomic ISM and Galactic rotation curve are of major importance to relate observed Galactic dynamics to the 3D structure of the Milky Way. For a detailed

and more complete discussion of the Galactic ISM and atomic hydrogen, I refer the reader to the excellent reviews by Ferrière (2001), Kalberla & Kerp (2009), Draine (2011), Dickey (2013), and Klessen & Glover (2016), that many parts of this section are based on.

1.1.1 Composition of the ISM

Most of the gas in the ISM is composed of hydrogen and helium, accounting for ∼70% and ∼28% of the total gas mass, respectively. Heavier elements make up around 2% of the total gas mass in the Milky Way. Since helium is a noble gas and therefore chemically inert, it is reasonable to put the focus on the chemistry of hydrogen gas when considering the physics of molecular cloud formation in the ISM. The thermal and chemical state of hydrogen gas is commonly described in a number of distinct phases: 1) neutral atomic hydrogen (H I), 2) ionized hydrogen (H II), 3) and molecular hydrogen (H_2). Most of the gas mass is in the form of neutral atomic or molecular hydrogen. While ionized hydrogen takes up the majority of the volume in the ISM, most of the gas mass is concentrated in clouds in the Milky Way disk and occupies just a small fraction of the total volume of the ISM.

Molecular hydrogen is a symmetric molecule and has no permanent electric dipole moment. It is therefore difficult to observe it directly in the ISM since H_2 does not emit at typical ISM temperatures. Because of this limitation, astronomers often resort to observational surrogates that tend to be found in the same locations as H_2. The most commonly used tracers of molecular hydrogen include extinction and thermal emission from interstellar dust grains (see below), gamma rays produced by cosmic rays that interact with hydrogen nuclei, and most notably rotational lines of different isotopologues of carbon monoxide (CO). The most abundant isotopologue is $^{12}C^{16}O$ (or simply CO), but it is often so abundant to become optically thick, such that measurements of column densities become unreliable. The next most common isotopologue is $^{13}C^{16}O$ (usually ^{13}CO), that is often used to trace molecular clouds in the Galactic plane. Since CO has a relatively low critical density of $\sim 10^3\,\mathrm{cm}^{-3}$, it can be a suitable tracer when considering the relationship between diffuse atomic gas and molecular gas.

Neutral atomic hydrogen is widespread and more evenly distributed in the Galactic plane than molecular hydrogen. Observations of spiral galaxies reveal that H I disks have typically three times the extent of the stellar distribution. As molecular gas is observed to correlate with star formation, the distribution of the molecular gas in the Galactic plane is similar to the extent of the stellar disk. Molecular line surveys show that the radial distribution of molecular hydrogen is distinctly different than that of H I, with an excess within the central few hundred parsec of the Galaxy, known as the Central Molecular Zone (CMZ; for a recent overview see Henshaw et al., 2022). It then falls off before strongly peaking about halfway between the sun and the Galactic center, a Galactic feature that was dubbed the Molecular Ring (Stecker et al., 1975). Beyond the Molecular Ring the surface density of molecular hydrogen decreases exponentially but can be traced out to distances >10 kpc from the Galactic center (Heyer & Dame, 2015).

Optical images of our Milky Way plane reveal dark patches and obscured regions

that block the starlight coming from more distant locations in our Galaxy. These dark clouds become steadily transparent if we move our observations to lower frequencies, an effect that is due to the phenomenon known as reddening. The observed correlation of reddening with the H_2 column density points at the existence of another important constituent of the ISM, that is generically referred to as dust. Dust is responsible for absorbing light over wide ranges of frequencies and acts as a catalyst for a more efficient formation of molecular hydrogen (Gould & Salpeter, 1963). The transition from atomic to molecular hydrogen in the gas-phase is extremely inefficient. Instead, most of the formation of molecular hydrogen takes place on the surface of dust grains. Association reactions of two hydrogen atoms occur rapidly on grain surfaces, and the H_2 formation rate is therefore primarily determined by the rate at which hydrogen is adsorbed onto the dust grains.

Dust is also an important heating source in the ISM through photoelectric heating (PE), which is one of the most dominant radiative heating mechanisms in the ISM (Watson, 1972). PE heating describes the ejection of an electron from a dust grain due to the absorption of a UV photon. The chemical and thermal state of the ISM is therefore intricately coupled to the interaction between matter and the interstellar radiation field (ISRF). The so controlled heating and cooling mechanisms set the thermal balance of the ISM and hold important implications for the chemical and thermal state of atomic hydrogen.

H I phases

Much of the ISM in disk galaxies is in the form of neutral atomic hydrogen. The thermal and chemical state of atomic hydrogen can be described by a set of distinct phases that are governed by a balance of heating and cooling mechanisms. An early model of the atomic hydrogen phases was put forward by Field et al. (1969). Under the assumption that atomic hydrogen in the ISM can reach thermal pressure equilibrium, there exist two thermally stable atomic hydrogen phases over a narrow range of pressures, a cold phase known as the cold neutral medium (CNM), with temperatures $T \sim 100\,\mathrm{K}$, and a warm phase known as the warm neutral medium (WNM) at temperatures somewhat less than $T \sim 10^4\,\mathrm{K}$. A later model by McKee & Ostriker (1977) included additional consideration of ionization, primarily driven by supernova explosions. This creates ionized phases of hydrogen that are usually referred to as hot ionized medium (HIM), with temperatures $T \sim 10^6\,\mathrm{K}$, and warm ionized medium (WIM) having temperatures similar to those of the WNM ($T \sim 10^4\,\mathrm{K}$). The processes leading to the existence of ionized hydrogen also influence the cooling and heating mechanisms and thus determine the thermal balance between CNM and WNM. Adjustments in the prescription of cooling and heating rates were made by Wolfire et al. (1995), with an updated version of their model described in Wolfire et al. (2003). The Wolfire et al. (2003) model will serve as the standard reference throughout this thesis when considering the thermal state of atomic hydrogen. In the following, I will briefly summarize the results of the their model that describes the thermal state of H I.

The coexistence of thermally stable CNM and WNM is in good agreement with a wide variety of observations in the local ISM. But the two-phase model extends to the majority of the Galactic plane. Observations show that there is H I emission,

which can originate from both CNM and WNM (see Chapter 2), at all velocities allowed by the Galactic rotation (see e.g. Taylor et al., 2003; McClure-Griffiths et al., 2005; Stil et al., 2006; Wang et al., 2020a). Similarly, absorption, which originates mostly from the CNM, is also found throughout much of the Galactic disk (e.g., Dickey et al., 2009). Star formation provides an indirect test for the presence of the CNM since the gas that forms stars presumably goes through the stage of being cold H I.

The calculation of a thermal equilibrium of H I at different Galactocentric distances is controlled by radial density and chemical abundance distributions, ISRF intensity, ionization rates, and lastly the net heating and cooling of the gas. Even in the presence of interstellar turbulence, which is observed to be the main contributor to the total pressure in the ISM (Larson, 1979), Wolfire et al. (2003) find that the assumption of thermal pressure equilibrium is satisfied in most of the Galactic disk, that is the cooling time of the gas to reach an equilibrium is shorter than the typical time interval between supernova-induced shocks, that move the gas out of equilibrium.

The dominant heating process in both the CNM and WNM is PE heating. The distributions of far-ultraviolet (FUV; $6\,\mathrm{eV} < h\nu < 13.6\,\mathrm{eV}$) radiation, extreme ultraviolet (EUV; $13.6\,\mathrm{eV} < h\nu \lesssim 100\,\mathrm{eV}$) radiation, soft X-ray ($100\,\mathrm{eV} \lesssim h\nu \lesssim 1\,\mathrm{keV}$) radiation, and cosmic rays are required to estimate the ionization fraction, the charge on grains, and the PE heating rate. The PE heating rate depends on the abundance of smallest dust particles, primarily polycyclic aromatic hydrocarbons (PAHs). The PE heating rate is furthermore set by the FUV intensity, since the FUV intensity provides the total photon energy available for gas heating and it governs the grain charge and thus efficiency at which the photon energy in converted to gas heating. The FUV intensity is determined by the distribution of OB stars and the opacity provided by dust. While the heating through soft X-rays is negligible compared to the heating by photoemission from dust grains, the ionization by EUV and soft X-ray radiation helps maintain an efficient PE heating rate as it provides a source of free electrons that can neutralize positive charging of grains. At the same time, the cooling rate by electron recombination onto positively charged dust grains regulates the PE heating rate and thus plays a role in the thermal pressure balance.

The cooling rate mainly depends on the gas-phase abundances of carbon and oxygen. The dominant cooling process in the CNM is radiative line cooling through the [C II] 158 μm fine-structure transition. The main cooling mechanism of the WNM is radiative line cooling by [C II] (albeit a factor of $\sim$20 weaker per H atom than for the CNM), [O I] 63 μm, and Lyα, as well as cooling by electron recombination onto dust grains (see above).

Combining the different cooling and heating mechanisms, both the chemical equilibrium abundances and the thermal equilibrium temperature can be calculated as a function of hydrogen nucleus density $n = n_{\mathrm{H\,I}} + n_{\mathrm{H^+}} + 2n_{\mathrm{H_2}}$, where $n_{\mathrm{H\,I}}$, $n_{\mathrm{H^+}}$, and $n_{\mathrm{H_2}}$ is the neutral atomic hydrogen volume density, the ionized hydrogen volume density, and the molecular hydrogen volume density, respectively. At given n, the temperature at thermal equilibrium T can be converted to the thermal pressure using $P = \sum_i n_i k_{\mathrm{B}} T$, where i refers to the different chemical species. In this way, a phase diagram, gas thermal pressure P against the hydrogen nucleus density n, can

be constructed.

Figure 1.1 presents an example phase diagram for Galactocentric radii 3, 5, 8.5, 11, 15, 18 kpc, with the Wolfire et al. (2003) standard parameters for density, abundances, intensity, and ionization rates. Thermal stability against isobaric density perturbations exists for $\mathrm{d}P/\mathrm{d}n > 0$. If the phase diagram exhibits the characteristic S-shaped curve, as shown in Fig. 1.1, then there exists a narrow range of pressures $P_{\mathrm{min}} \leq P \leq P_{\mathrm{max}}$, at which the CNM and WNM phase can coexist in thermal equilibrium. The pressures P_{min} and P_{max} are the local minimum and local maximum, where $\mathrm{d}P/\mathrm{d}n = 0$, respectively. In the density and temperature range where $\mathrm{d}P/\mathrm{d}n < 0$, the gas is thermally unstable and would move toward a stable branch of the gas phase under isobaric perturbations. For pressures larger than P_{max} the gas can only be in the cold phase (CNM), while at pressures lower than P_{min} the gas can only be in a thermally stable warm phase (WNM). At the commonly reported thermal pressures observed in the Milky Way (e.g., Jenkins & Tripp, 2001), the model thus predicts neutral atomic hydrogen to exist in two stable phases for a large portion of the Galactic disk (3 kpc $< R <$ 18 kpc).

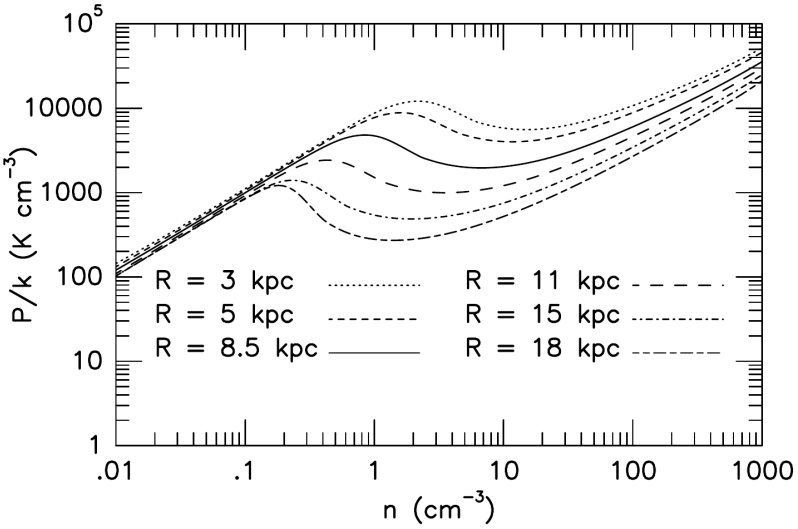


Figure 1.1: Phase diagrams showing the thermal pressure P/k_{B} against the hydrogen nucleus density n for Galactocentric radii 3, 5, 8.5, 11, 15, 18 kpc. The hydrogen gas is thermally stable against isobaric density perturbations where $\mathrm{d}P/\mathrm{d}n > 0$. Taken from Wolfire et al. (2003, see their Fig. 7)

1.1.2 Global distribution H I

The properties of the large-scale H I distribution in our Milky Way are generally derived using a Galactic rotation curve. The general consensus is that the Galactic rotation beyond a few kiloparsec from the Galactic center can be described by an almost flat rotation curve with around $v_{\mathrm{rot}} \sim 220\,\mathrm{km\,s^{-1}}$ (see e.g., Burton, 1988; McClure-Griffiths & Dickey, 2007; Kalberla et al., 2007; Reid et al., 2019). Mapping a line-of-sight velocity of the gas to a kinematic distance inside the solar circle is ambiguous, as we will see in this section. But for Galactocentric radii $R \gtrsim$

$R_\odot$, the distribution of H I surface density can be well described by an exponential distribution with a radial scale length of $R_s = 3.75\,\mathrm{kpc}$ (Kalberla & Dedes, 2008).

The Galactic disk shows a strong warp, that is commonly described as an "integral-shaped" morphology, where the disk is warped up to $\sim$5 kpc toward positive Galactic latitudes in the northern Galactic plane, and somewhat less toward negative latitudes in the southern part of the plane (Henderson et al., 1982; Burton & te Lintel Hekkert, 1986; Nakanishi & Sofue, 2003; Kalberla & Kerp, 2009). Taking into account the warp, Kalberla & Dedes (2008) have derived an average H I midplane volume density $n_0(R)$. For Galactocentric radii $7 \lesssim R \lesssim 35\,\mathrm{kpc}$, the radial distribution is approximately $n(R) = n_0 \exp(-(R - R_\odot)/R_n)$ with $n_0 = 0.9\,\mathrm{cm}^{-3}$, $R_\odot = 8.5\,\mathrm{kpc}$, and $R_n = 3.15\,\mathrm{kpc}$, which is an average of the northern and southern Galactic plane and could hold systematic differences in the two hemispheres.

The gaseous disk furthermore flares strongly. The average surface density is essentially a product of the midplane volume density and the corresponding flaring $\langle\Sigma(R)\rangle \approx \langle n_0(R)\rangle \cdot \langle h_z(R)\rangle$, so the flaring might also be approximated by an exponential distribution. The average scale height $h_z(R)$ of H I is in fact observed to increase exponentially with Galactocentric distance R and can be described by an exponential function $h_z(R) = h_0 \exp((R - R_\odot)/R_0)$, with $h_0 = 0.15\,\mathrm{kpc}$ and $R_0 = 9.8\,\mathrm{kpc}$ for $5 \lesssim R \lesssim 35\,\mathrm{kpc}$ (Kalberla & Dedes, 2008).

Kinematics of Galactic H I and rotation curve

The underlying rotation curve of the Milky Way is one of the fundamental properties to investigate the structure and dynamics of the Galaxy. The rotation curve can be measured using many different tracers. In the following, we will consider the Galactic structure that can be derived from H I observations. In order to interpret the dynamics of the large-scale structure of H I, it is common to assume a circularly rotating disk. There is evidence of regions where departures from the circular motion occur, particularly toward the central bar region and spiral arm structures (Burton & Liszt, 1993), but to first order circular motion is a reasonable assumption to characterize the global dynamics of the Milky Way disk. Under the assumption that the rotation curve is approximately flat, the angular velocity $\Omega(R)$ is a decreasing function only of the Galactocentric distance R. The radial velocity V_r, that is due to the differential rotation of the Galaxy, can then be derived using the geometry sketched in Fig. 1.2. For the sake of simplicity, we place the points of interest at Galactic latitude $b = 0°$. The observer is at a distance of R_0 from the Galactic center (GC) and is rotating about this center with angular velocity $\Omega_0 = V_0/R_0$. Observed hydrogen emission at point P_1 with a distance of d from the observer is moving with orbital velocity V about the Galactic center. For a disk in circular rotation, the observed radial velocity of P_1, V_r, can be expressed as the difference between the orbital velocity of P_1 projected along the line of sight, $V_{P_1,r}$, and the line-of-sight projection of the rotation velocity of the local standard of rest (LSR[1]), $V_{0,r}$. The radial velocity of the observed hydrogen then is

$$V_r = V_{P_1,r} - V_{0,r}\,. \tag{1.1.1}$$

[1]In the following, the subscripts r and LSR will be used interchangeably when referring to the radial velocity.

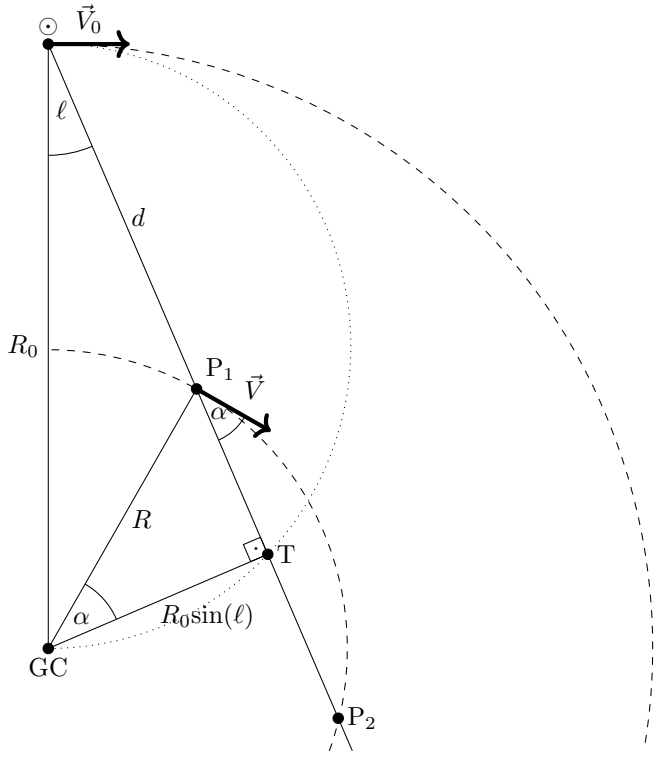


Figure 1.2: Sketch of the geometry of a circularly rotating disk. The observer ($\odot$) is located at a distance of R_0 from the Galactic center (GC) and is moving at the angular velocity $\Omega_0 = V_0/R_0$ around the center. Observed hydrogen emission at a point P_1 at Galactic longitude ℓ and at a distance of d from the sun is moving with velocity V. The dashed circles mark the circular orbits of both the sun and the point P_1. The point T marks the tangent or sub-central point, where the distance of the line of sight to the Galactic center is smallest, and thus the radial velocity component is highest. The point P_2 marks the second point along the line of sight that has the same radial velocity as P_1 with respect to the local standard of rest. The dotted circle marks the locus of sub-central points.

Since $R\cos(\alpha) = R_0\sin(\ell)$, it follows that

$$V_r = V\cos(\alpha) - V_0\sin(\ell) \tag{1.1.2}$$

$$= V\sin(\ell)\frac{R}{R_0} - V_0\sin(\ell)\,. \tag{1.1.3}$$

This gives the fundamental equation relating the radial velocity to the Galactic rotation curve $\Omega(R) = V(R)/R$

$$V_r = R_0\,[\Omega(R) - \Omega_0]\,\sin(\ell)\,, \tag{1.1.4}$$

where $\Omega_0 = V_0/R_0$ is the angular velocity of the sun given the sun's orbital velocity and Galactocentric distance. If the rotation curve $\Omega(R)$ is known, then measured radial velocities can be mapped to distances. Different methods have been employed to measure the rotation curve in different regions in the Galaxy.

The rotation curve in the inner Galactic plane can be determined from H I observations taken along directions in the first ($0° < \ell < 90°$) and fourth ($270° < \ell < 360°$) Galactic quadrant by finding the extreme velocities of the line emission,

provided there is line emission at the location of those extreme points. This is a fair assumption, given the ubiquity of atomic hydrogen in the Galactic plane. As long as the angular rotation $\Omega(R)$ does not rise with Galactocentric radius R, that is the rotation velocity V does not rise faster than $V \propto R$ (solid body), the observed radial velocity in the inner Galaxy will increase to an extremum and then decrease again. If we consider a direction in the Galactic longitude range $0° < \ell < 90°$ and increase the distance from the sun d along the line of sight, the distance to the Galactic center R first decreases. This corresponds to an increasing $\Omega(R)$ under our assumption, and thus the radial velocity V_r will increase (see Eq. 1.1.4). If we increase d further, the radial velocity will reach an extremum at the the point closest to the Galactic center (the sub-central or tangent point, see Fig. 1.2), with $R = R_{\rm min} = R_0 \sin(\ell)$. For even larger distances d, the Galactocentric distance R increases again, and $\Omega(R)$ and V_r decrease. By measuring the radial velocity at the tangent point, or "terminal" velocity, and assigning it to the minimal distance from the center $R = R_0 \sin(\ell)$, one can obtain the rotation curve $\Omega(R)$. All these points together, the so-called locus of sub-central points, make a circle centered on the point half way between the sun and the Galactic center (see Fig. 1.2). This behavior of the radial velocities can be derived directly from Eq. (1.1.4). Using the law of cosines, we obtain (see Fig. 1.2)

$$R^2 = d^2 + R_0^2 - 2\,d\,R_0 \cos(\ell)\ . \tag{1.1.5}$$

Assuming $\Omega(R)$ is known, we can insert the square root of Eq. (1.1.5) into Eq. (1.1.4) and derive the radial velocity as a function of the distance from the sun d. For illustration purposes, we assume the roation curve given by Brand & Blitz (1993), $V(R)/V_0 = a_1\,(R/R_0)^{a_2} + a_3$, where V_0, R_0, a_1, a_2, and a_3 are $220\,{\rm km\,s^{-1}}$, $8.5\,{\rm kpc}$, 1.00767, 0.0394, and 0.00712, respectively. Figure 1.3 shows the radial velocity V_r plotted against the distance d. As can be seen, the presence of a terminal velocity in the longitude range $0° < \ell < 90°$ implies that there are two distances within the solar circle corresponding to the same line-of-sight velocity. This is known as the kinematic distance ambiguity. This can also be seen in Fig. 1.2, where point $\rm P_2$ has the same radial velocity component as point $\rm P_1$ with respect to the sun. All velocities for points within the solar circle are positive in the first quadrant and negative in the fourth quadrant. Velocities outside the solar circle are negative in the first and second ($90° < \ell < 180°$) quadrant, and positive in the third ($180° < \ell < 270°$) and fourth. Note that in the outer Galaxy ($R > R_0$), the rotation curve from H I is degenerate as there is no terminal velocity, so the rotation curve in the outer Galaxy has to be constrained using other tracers, such as masers or H II regions with known distances (e.g., Reid et al., 2009, 2014, 2019). The measurement of terminal velocity is furthermore complicated as internal motion of the gas leads to line broadening effects that blur any sharp cutoff in the spectrum that marks the terminal velocity. The line profile of H I emission can furthermore be vulnerable to velocity crowding (Burton, 1972). Since the radial velocity in terms of distance is proportional to the inverse of the velocity gradient along the line of sight, equal steps in velocity (as sampled by a spectrometer, for example) will map into unequal steps of distance. If the velocity gradient of V_r is small, then a larger distance interval will be crammed into the velocity interval. This effect is therefore more pronounced around the tangent points (see Fig. 1.3).

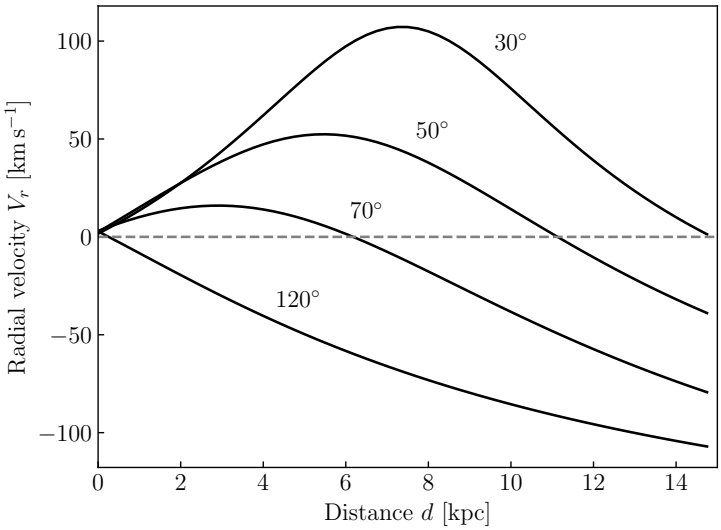


Figure 1.3: Radial velocity as a function of distance from the sun. Plotted are the radial velocities using the Galactic rotation curve by Brand & Blitz (1993) for Galactic longitudes $\ell = 30°$, 50°, 70°, and 120°, respectively. For a longitude range $0° < \ell < 90°$, there exist two distances within the solar circle (corresponding to positive velocities in the first Galactic quadrant) with the same radial velocity. For longitudes in the range $90° < \ell < 180°$, the radial velocity is negative and has no distance ambiguity. The situation in the fourth and third Galactic quadrants is similar to that illustrated above for the first and second quadrant, except for a reversal in the sign of the velocity.

There are many rotation curve models that have been proposed and revised as new observations put constraints on Galactic constants and the Milky Way rotation (see e.g., Burton, 1988; Brand & Blitz, 1993; Kalberla et al., 2007; McClure-Griffiths & Dickey, 2007; Reid et al., 2009, 2014, 2019), many of them with a more sophisticated approach and taking into account characteristics of regions of non-circular or streaming motion, like the central bar or spiral arms. These models are then usually valid over a much larger range of Galactocentric distances.

1.1.3 H I as a tracer of individual clouds

The previous subsection showed that the deduction of the Galactic H I distribution from atomic hydrogen emission can be a challenging task in the inner Galactic plane due to the kinematic distance ambiguity and the effect of velocity crowding. Investigating individual atomic cloud structures and their physical properties is usually limited to regions that lie beyond the Galactic midplane or at distances that correspond to unambiguous radial velocities. H I emission from the Galactic midplane can therefore be used as a tracer of individual atomic clouds in the outer Galaxy more easily. Figure 1.4 shows an example spectrum of H I emission taken from the THOR survey discussed in this thesis (see Sect. 3.1). Atomic hydrogen emission is observed over a wide range in velocity, and can thus be found in both spiral arm and inter-arm regions. Molecular line emission, on the other hand, such as the ^{13}CO emission spectrum shown in Fig. 1.4 is more clearly defined. A viable method to trace atomic clouds in the inner Galaxy is H I self-absorption (HISA; see Sect. 2.2), that allows to resolve H I clouds spectroscopically. HISA is produced by absorption

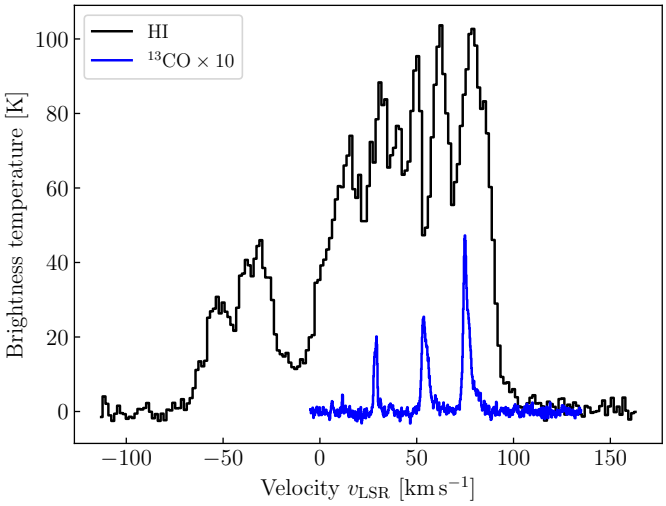


Figure 1.4: THOR-H I and ^{13}CO emission spectrum. The black spectrum taken from the THOR survey shows the beam-averaged H I emission centered at the coordinates $(\ell, b) = (36.4°, -0.2°)$. Due to the Galactic rotation, H I emission is found over almost the entire spectral range. Tracing the inner part of the Galactic disk, positive radial velocities show multiple emission components blended along the line of sight. Sharp dips in the spectrum are likely due to H I self-absorption (see Sect. 2.2) and trace cold atomic clouds in the foreground. Shown in blue is the corresponding ^{13}CO emission (multiplied by ten for better readability) taken from the Galactic Ring survey (GRS; Jackson et al., 2006).

of background H I emission at the same radial velocity. The location of atomic gas traced by HISA is thus strongly biased toward the foreground distance. Since absorption is also more sensitive to colder gas, HISA is a way to isolate cold atomic gas clouds with $T \lesssim 100\,\mathrm{K}$ from H I emission, that stems from both CNM and WNM. The H I emission dip in Fig. 1.4 corresponding to the central component of ^{13}CO is likely produced by H I self-absorption.

While the global H I distribution in the Milky Way has been studied well using surveys that cover large portions of the sky, the limited angular resolution of these surveys do not offer a dynamical view of the small-scale physics of H I. The physical and dynamical state of H I as part of the multiphase structure of the ISM requires high angular resolution data. So far, a comprehensive survey of the physical properties of the distinct gas phase traced by HISA, and its dynamical relationship with the star-forming molecular gas phase does not exist.

1.2 Motif of this thesis

In this thesis, I address the poorly constrained physical nature of cold hydrogen gas in the Galactic ISM. Using high angular resolution data and newly developed methods to identify hydrogen clouds, I provide the missing link to bridge the gap between the diffuse atomic gas and the molecular gas phase and investigate the properties of cold atomic hydrogen as a distinct gas phase during atomic and molecular cloud formation.

The spectral line of atomic hydrogen will be utilized and H I emission as well as

H I self-absorption will form the observational basis to provide an elaborate view on the atomic phase of the ISM. The central questions I will address are as follows:

1. What are the physical and kinematic properties of the atomic gas in the ISM and how is it structured?

2. How can we constrain the transition from the atomic to the molecular gas phase using observations?

3. What is the relationship between the atomic and molecular gas in the ISM?

This thesis is organized as follows. Since the analysis of H I requires a deeper understanding of spectral line diagnostics, I will give an overview of the hydrogen 21 cm spectral line and the method of H I self-absorption in Chapter 2. The observational data and methods used to derive the hydrogen properties are presented in Chapter 3. The properties of a unique, large (1 kpc) atomic hydrogen cloud are investigated in Chapter 4, using H I emission from the outer Galaxy that is less confused along the line of sight. Chapters 5 and 6 concern the detection and nature of cold atomic gas in the inner Galactic plane using H I self-absorption. In these chapters, I also develop a new approach to identify HISA in large surveys in an unbiased way. I summarize my findings and discuss conclusions drawn from these observational studies in Chapter 7, hoping to have shed "more light" on the nature of the interstellar medium.

2

Spectral line fundamentals

The derivation of the atomic hydrogen properties requires an understanding of the mechanisms governing the radiative transfer of the spectral line of H I. In the following section I will briefly discuss the most important aspects of the radiative transfer equation and the H I 21 cm line. The derivations are mainly based on the excellent online course "Essential Radio Astronomy"[1] by J. J. Condon and S. M. Ransom as well as the reviews by Draine (2011), Wilson et al. (2013), and Dickey (2013).

2.1 Radiative transfer

The specific intensity I_ν of radiation moving through free space is conserved along its path, that is the change in intensity is

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}s} = 0\,, \tag{2.1.1}$$

where s is the coordinate along the light ray between the source and the observer. The specific intensity I_ν will change only if radiation is absorbed or emitted, and this change of I_ν is described by the equation of radiative transfer. The radiative transfer equation is determined by the interaction of radiation and matter along its path. If radiation passes through an intervening medium, the change in intensity is generally nonzero and is caused by emission, absorption, and scattering processes defined by the properties of the intervening medium.

Absorption

Absorption of radiation passing through an intervening medium is the amount of energy removed from the intensity along the path element $\mathrm{d}s$

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}s} = -\kappa_\nu\, I_\nu \qquad \text{(absorption only)}\,, \tag{2.1.2}$$

[1] https://www.cv.nrao.edu/~sransom/web/xxx.html

where κ_ν is the *linear absorption coefficient* at frequency ν, with dimensions 1/length. The absorption coefficient is usually positive but can become negative under certain astrophysical conditions, in which case amplification of emission takes place (masers). Rearranging and integrating both sides of Eq. (2.1.2) gives

$$I_\nu = I_{\nu,0}\, e^{-\tau_\nu} \,, \tag{2.1.3}$$

with the definition of the *optical depth* as

$$|\tau_\nu| \equiv \kappa_\nu \, \mathrm{d}s \,. \tag{2.1.4}$$

If $\tau \ll 1$, the absorbing medium is said to be *optically thin*, if $\tau \gg 1$, it is *optically thick*.

Emission

The intervening medium might also emit photons and thus adds to the radiative transfer equation. The *emission coefficient* ϵ_ν is defined such that

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}s} = \epsilon_\nu \qquad \text{(emission only)} \,, \tag{2.1.5}$$

if there is emission only. Integration gives

$$I_\nu = I_{\nu,0} + \int_s \epsilon_\nu(s\prime)\, \mathrm{d}s\prime \,. \tag{2.1.6}$$

Radiative transfer equation

Combining the effects of absorption (Eq. 2.1.2) and emission (Eq. 2.1.5) yields the general form of the equation of radiative transfer

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}s} = -\kappa_\nu \, I_\nu + \epsilon_\nu \,. \tag{2.1.7}$$

Using the definition of the optical depth in Eq. (2.1.4) we can rewrite the radiative transfer equation as

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}\tau_\nu} = -I_\nu + S_\nu \,, \tag{2.1.8}$$

where $S_\nu = \epsilon_\nu/\kappa_\nu$ is referred to as the source function. It is important to note that Eq. (2.1.8) does not include scattering processes, such as scattering by dust grains or electrons. We have seen the simple solution for either case of only emitting and only absorbing material. In general, the intervening medium is both absorbing and emitting radiation. The formal solution of Eq. (2.1.8) can be derived by using the integrating factor e^{τ_ν}. We therefore introduce

$$\tilde{I}_\nu = I_\nu \, e^{\tau_\nu}, \qquad \tilde{S}_\nu = S_\nu \, e^{\tau_\nu} \,. \tag{2.1.9}$$

Equation (2.1.8) then becomes

$$\frac{\mathrm{d}\tilde{I}_\nu}{\mathrm{d}\tau_\nu} = \tilde{S}_\nu \,. \tag{2.1.10}$$

Writing this in terms of the specific intensity gives

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}\tau_\nu}\, e^{\tau_\nu}\, \mathrm{d}\tau_\nu + I_\nu\, e^{\tau_\nu}\, \mathrm{d}\tau_\nu = S_\nu\, e^{\tau_\nu}\, \mathrm{d}\tau_\nu\,. \tag{2.1.11}$$

After integrating Eq. (2.1.11), the second term on the left-hand side drops out leaving

$$I_\nu\, e^{\tau_\nu} - I_{\nu,\tau_\nu=0} = \int_0^{\tau_\nu} S_\nu\, e^{\tau\prime_\nu}\, \mathrm{d}\tau\prime_\nu\,. \tag{2.1.12}$$

If we assume that S_ν = const. along the line of sight (that is the temperature is constant), the integral can be computed explicitly and we arrive at the general solution to the radiative transfer equation:

$$I_\nu = I_{\nu,0}\, e^{-\tau_\nu} + S_\nu \cdot (1 - e^{-\tau_\nu})\,. \tag{2.1.13}$$

The solution of the intensity I_ν at optical depth τ_ν has a simple physical interpretation. The first term on the right-hand side describes the initial intensity $I_{\nu,0}$ attenuated by the factor $e^{-\tau_\nu}$. The second term is the emission of the medium itself, minus the effective absorption of that emission over its path from the point of emission. In full thermodynamic equilibrium (TE), matter and radiation are in equilibrium at the same temperature T. If there is complete equilibrium of the radiation with its surroundings, the brightness distribution is described by the Planck function, which depends only on the thermodynamic temperature T

$$I_\nu = B_\nu(T) = \frac{2h\nu^3}{c^2} \frac{1}{\exp\left(\frac{h\nu}{k_\mathrm{B}T}\right) - 1}\,, \tag{2.1.14}$$

where h, c, and k_B denote the Planck constant, the speed of light, and the Boltzmann constant, respectively. In equilibrium at any temperature T, the intensity does not change within the medium such that $\mathrm{d}I_\nu = 0$. This yields Kirchhoff's law

$$S_\nu = B_\nu(T) = \frac{\epsilon_\nu}{\kappa_\nu}\,. \tag{2.1.15}$$

Kirchhoff's law must be true at any point as ϵ_ν and κ_ν depend only on the local properties of the matter, that is Kirchhoff's law applies even if the matter is not in full TE with the radiation field but instead in local thermodynamic equilibrium (LTE). It is common to describe the intensity field in terms of the brightness temperature $T_B(\nu)$, which is the equivalent temperature a black body would have at a given intensity. Rearranging the Planck function in Eq. (2.1.14) gives

$$T_B(\nu) = \frac{h\nu/k_\mathrm{B}}{\ln\left[2h\nu^3/(I_\nu c^2) + 1\right]}\,. \tag{2.1.16}$$

In LTE, the brightness temperature is equal to the thermodynamic temperature of the medium. In radio astronomy, it is commonly the case that $h\nu \ll k_\mathrm{B}T$, so the brightness temperature can be expressed in the Rayleigh-Jeans limit

$$T_B(\nu) = \frac{c^2}{2k_\mathrm{B}\nu^2}\, I_\nu\,, \tag{2.1.17}$$

which is formally known as the antenna temperature and gives a linear relation between the intensity and the brightness temperature of a source. We can then write the solution to the equation of radiative transfer in terms of temperature

$$T_B = T_{B,0}\, e^{-\tau_\nu} + T \cdot (1 - e^{-\tau_\nu})\,, \tag{2.1.18}$$

where $T_{B,0}$ is the initial unimpeded brightness temperature and T is the excitation or spin temperature of the medium, which is a statistical description of the relative population levels of a system (see below). This solution will be the basis for all further analysis of the H I line.

Spectral line radiative transfer

Atomic hydrogen can be conveniently detected and studied via its famous H I 21 cm line, originating in the hyperfine splitting of the electronic ground state of atomic hydrogen. The basis for the emission and absorption processes of the H I 21 cm line is the interaction of photons with matter. On a microscopic level, atoms and molecules have discrete quantum mechanical energy states that can be populated through excitation and de-excitation. When a transition from one state to another occurs, the discrete energy difference ΔE between those states results in the emission or absorption of characteristic spectral lines with frequency $\nu_0 = \Delta E/h$, with h being the Planck constant. The probability of a transition from some upper state E_u to a lower state E_l, and vice versa, can be described by means of Einstein coefficients.

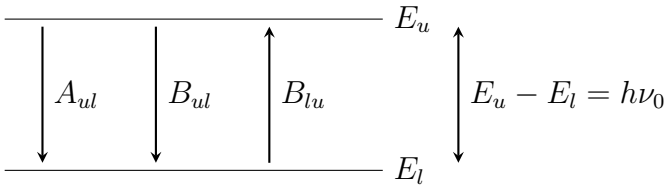


Figure 2.1: Two-level system: transitions from the upper (u) to lower (l) state and vice versa are described by Einstein coefficients.

In this section, we will connect the microscopic transition processes with the macroscopic properties of emission ϵ_ν and absorption κ_ν. Figure 2.1 shows a simple two-level system with an upper energy level E_u and a lower energy level E_l as well as the corresponding Einstein coefficients. The hyperfine splitting of the hydrogen ground state can be described by a two-level system.

The system can change its energy state by emitting or absorbing radiation. A transition can occur in three[2] different ways. The system can change its state through *spontaneous emission*, that is described by the Einstein coefficient A_{ul}. The system changes spontaneously the energy level from the upper to the lower state and emits a photon with frequency ν_0.

Stimulated emission occurs if photons of the characteristic frequency ν_0 are present. An ambient photon can then stimulate an excited hydrogen atom to radiate

[2]We are not considering collisional excitation and de-excitation here. However, a similar argument can be made considering only collisional excitation without the effects of any incident radiation field.

a photon of energy ΔE and momentum, thereby decaying to the lower level. The rate of stimulated emission is proportional to the density of photons described by the average energy density of the radiation field $\bar{u}$ and the Einstein coefficient B_{ul}. Stimulated emission (also: negative absorption) competes effectively with absorption at radio wavelengths where $h\nu \ll k_{\mathrm{B}}T$.

Absorption of a photon at frequency ν_0 is described by the Einstein coefficient B_{lu}, at which the system changes from the lower to the upper state. If the system is in equilibrium the transition rates from the lower to the upper state and vice versa are the same

$$n_u\, A_{ul} + n_u\, B_{ul}\, \bar{u} = n_l\, B_{lu}\, \bar{u}\ , \tag{2.1.19}$$

where n_u and n_l describe the density of the upper and lower states that are populated through radiative excitation and de-excitation, respectively. The average energy density of the radiation field is described by $\bar{u}$. If we assume that the system is in thermal equilibrium, the ratio of the probabilities of the two states is furthermore described by the Boltzmann factor

$$\frac{n_u}{n_l} = \frac{g_u}{g_l} \cdot \exp\left(-\frac{h\nu_0}{k_{\mathrm{B}}T}\right)\ , \tag{2.1.20}$$

where g_u and g_l are the statistical weights of the corresponding states, and T is the excitation or spin temperature. Even when collisions contribute to the excitation and de-excitation of energy levels and thermalize the gas, the relative level population would still be described by the Boltzmann factor. In that case the Boltzmann factor would be determined by the kinetic gas temperature T_{kin} in LTE. With Eq. (2.1.18), this shows that the excitation or spin temperature T is always in the range $T_B \leq T \leq T_{\mathrm{kin}}$. Using Eqs. (2.1.19) and (2.1.20), the mean spectral energy density can be expressed as

$$\bar{u} = \frac{A_{ul}}{B_{ul}\left[\frac{g_l}{g_u}\exp\left(\frac{h\nu_0}{k_{\mathrm{B}}T}\right)\frac{B_{lu}}{B_{ul}} - 1\right]}\ . \tag{2.1.21}$$

In TE, the mean spectral energy density $\bar{u}$ is given by

$$\bar{u} = \frac{4\pi}{c}B_\nu = \frac{8\pi h\nu_0^3}{c^3}\,\frac{1}{\exp\left(\frac{h\nu_0}{k_{\mathrm{B}}T}\right) - 1}\ . \tag{2.1.22}$$

Inserting this into Eq. (2.1.21) shows that the Einstein coefficients have to obey

$$A_{ul} = \frac{8\pi h\nu_0^3}{c^3}\, B_{ul}\ , \tag{2.1.23}$$

$$g_l B_{lu} = g_u B_{ul}\ . \tag{2.1.24}$$

These two equations relate A_{ul}, B_{ul}, and B_{lu}, so all three coefficients can be computed if only one is known (e.g., the spontaneous emission coefficient A_{ul}). Furthermore, the equations allow the spectral line radiative transfer problem to be solved in terms of the spontaneous emission coefficient A_{ul} alone. The energy of a photon being absorbed is $h\nu_0$. The fraction of photons absorbed per unit area per unit

length is $n_l B_{lu}/c$, and the photon energy loss per unit length at frequency ν_0 is (see Condon & Ransom, 2016)

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}s} = -\kappa_\nu\, I_\nu = -\frac{h\nu_0}{c}\, n_l\, B_{lu}\, \varphi(\nu)\, I_\nu\,, \tag{2.1.25}$$

where κ_ν is the ordinary absorption coefficient, and $\varphi(\nu)$ is the normalized line profile ($\int \varphi(\nu) = 1$) of the spectral line, determined by the velocity distribution of the H I gas. Similarly, the photon energy gain per unit length due to stimulated emission (negative absorption) is (see Condon & Ransom, 2016)

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}s} = -\kappa_\nu\, I_\nu = \frac{h\nu_0}{c}\, n_u\, B_{ul}\, \varphi(\nu)\, I_\nu\,, \tag{2.1.26}$$

giving the net absorption coefficient combing the two effects

$$\kappa_\nu = \frac{h\nu_0}{c}\,(n_l B_{lu} - n_u B_{ul})\,\varphi(\nu)\,. \tag{2.1.27}$$

The spontaneous emission coefficient is the spectral brightness added per unit volume by spontaneous transitions from the upper to lower energy levels. As above, the photon energy is $h\nu_0$. These photons are emitted isotropically over 4π sr, so

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}s} = \epsilon_\nu = \frac{h\nu_0}{4\pi}\, n_u A_{ul}\, \varphi(\nu)\,. \tag{2.1.28}$$

The equation of spectral line radiative transfer thus becomes

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}s} = -\frac{h\nu_0}{c}\,(n_l B_{lu} - n_u B_{ul})\,\varphi(\nu)\, I_\nu + \frac{h\nu_0}{4\pi}\, n_u A_{ul}\, \varphi(\nu)\,. \tag{2.1.29}$$

The H I 21 cm line

Atomic hydrogen in its simplest form consists of one electron bound to one proton in the atomic nucleus. The H I line is produced by the coupling of the electron's magnetic moment to the magnetic field induced by the magnetic moment of the hydrogen's proton. The coupling results in a hyperfine splitting of parallel and antiparallel spin states. Even in the $n = 1$ ground state a hyperfine splitting of the energy level is present.

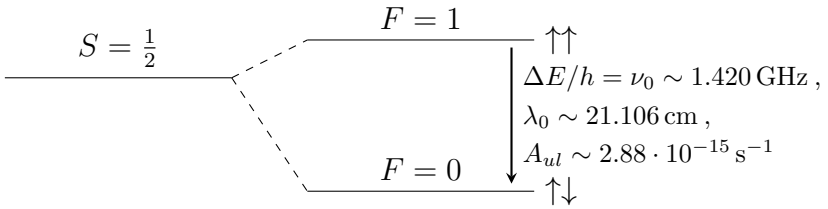


Figure 2.2: Hyperfine splitting of the hydrogen ground state.

If the spins are parallel, the quantum state has a higher energy E_u with degeneracy $g = 2F + 1 = 3$, with F being the total angular momentum of the hydrogen

atom (see Fig.2.2). The antiparallel spin state (with degeneracy $g = 1$) has a lower energy E_l. The energy difference between the $F = 1$ (parallel) state and the $F = 0$ (antiparallel) state is $\Delta E = 5.87 \cdot 10^{-6}\,\mathrm{eV}$ with a corresponding frequency of $\nu_0 = 1.420405751766\,\mathrm{GHz}$ (Gould, 1994). This translates to a wavelength of $\lambda_0 \sim 21.106\,\mathrm{cm}$.

The "spin flip" of the hydrogen atom is a forbidden transition. Its probability to occur is small but nonzero. The rate at which the spin flip happens is $A_{ul} = 2.88426 \cdot 10^{-15}\,\mathrm{s}^{-1}$ (Gould, 1994), roughly once every $\sim$11 Myr. In fact, collisional excitation and de-excitation largely determine the relative populations of the two energy levels. At the typical H I densities of about $n_{\mathrm{H\,I}} \sim 0.4\,\mathrm{cm}^{-3}$ encounters result in reorientation of the spins of a hydrogen atom about once every $\sim$400 yr, so only a very small fraction of all collisions will give rise to the emission or absorption of a photon. However, since atomic hydrogen is so abundant along the large lines of sight in the ISM, H I can still be detected at a frequent rate.

Since the energy difference is small between the two spin states, the population ratio and thus an equilibrium can be maintained even at small temperatures. For all astrophysical conditions of interest, the excitation temperature is $T \gg \Delta E/k_{\mathrm{B}} = 0.0682\,\mathrm{K}$. The exponential in the Boltzmann factor can then be approximated by the zeroth order of the Taylor expansion

$$\frac{n_u}{n_l} = \frac{g_u}{g_l} \cdot \exp\left(-\frac{h\nu_0}{k_{\mathrm{B}} T_s}\right) = 3 \cdot \exp\left(-0.0682\,\mathrm{K}/T_s\right) \approx 3\,, \tag{2.1.30}$$

$$\Rightarrow \qquad n_u \approx \frac{3}{4} n_{\mathrm{H\,I}}\,, \qquad n_l \approx \frac{1}{4} n_{\mathrm{H\,I}}\,, \tag{2.1.31}$$

with $n_{\mathrm{H\,I}}$ being the hydrogen number density. In the case of the atomic hydrogen hyperfine transition, the excitation temperature T_s characterizing the Boltzmann factor is called spin temperature as it describes the relative population levels of different spin states. Because 75% of hydrogen populates the upper energy state under all astrophysical conditions, the emissivity ϵ_ν of hydrogen is effectively independent of spin temperature

$$\epsilon_\nu = \frac{h\nu_0}{4\pi}\, n_u A_{ul}\, \varphi(\nu) = \frac{3h\nu_0}{16\pi}\, n_{\mathrm{H\,I}} A_{ul}\, \varphi(\nu)\,. \tag{2.1.32}$$

As mentioned in Chapter 1, the emission of the hydrogen line is thus attributed to both cold and warm gas. In contrast to this, the absorption κ_ν depends on the spin temperature T_s. Using the Einstein coefficient relations given in Eqs. (2.1.23) and (2.1.24), and the hydrogen level population in Eq. (2.1.30) approximated to the first order, we can write the net absorption as

$$\kappa_\nu = \frac{3hc^2}{32\pi\nu_0}\, A_{ul}\, \frac{n_{\mathrm{H\,I}}}{k_{\mathrm{B}} T_s}\, \varphi(\nu)\,, \tag{2.1.33}$$

from which it can be seen that the absorption scales as $\propto n_{\mathrm{H\,I}} \cdot T_s^{-1}$. We therefore expect the strongest absorption against very cold and dense H I clouds (CNM) while the warm and diffuse medium (WNM) is expected to be optically thin.

H I column density

For spectroscopic observations, the coordinate along our line of sight is the Doppler-shifted radial velocity of the gas, V, so the physical line-of-sight extent of a source of interest is usually unknown. In order to fold in information about the density structure of an individual H I cloud along the line of sight s, we define the column density

$$N_{\mathrm{H\,I}} = \int n_{\mathrm{H\,I}}\,\mathrm{d}s\,, \tag{2.1.34}$$

which is the integral of the volume density of hydrogen $n_{\mathrm{H\,I}}$ over the line of sight. Since the line shapes in radio astronomy are usually given in terms of the corresponding Doppler velocities V

$$\frac{\nu_0-\nu}{\nu_0} = \frac{V}{c} \quad \rightarrow \quad \varphi(\nu) = \varphi(V)\frac{\mathrm{d}V}{\mathrm{d}\nu} = \varphi(V)\cdot\left(-\frac{c}{\nu_0}\right)\,, \tag{2.1.35}$$

the absorption coefficient can be transformed into

$$\mathrm{d}\tau\left(\frac{V}{\mathrm{km\,s^{-1}}}\right) = -\kappa_\nu(s)\,\mathrm{d}\left(\frac{s}{\mathrm{cm}}\right) \tag{2.1.36}$$

$$= 5.4873\times10^{-19}\left(\frac{n_{\mathrm{H\,I}}}{\mathrm{cm^{-3}}}\right)\left(\frac{T_s}{\mathrm{K}}\right)^{-1}\left(\frac{\varphi(V)}{\mathrm{km^{-1}\,s}}\right)\mathrm{d}\left(\frac{s}{\mathrm{cm}}\right)\,. \tag{2.1.37}$$

If the spin temperature T_s is constant along the line of sight, we can compute the column density by integrating both sides over ν and s and rearrange the equation

$$\boxed{N_{\mathrm{H\,I}} = 1.8224\times10^{18}\;T_s\int\tau(V)\,\mathrm{d}V} \tag{2.1.38}$$

where $N_{\mathrm{H\,I}}$, T_s, and V are in units of $\mathrm{cm^{-2}}$, K, and $\mathrm{km\,s^{-1}}$, respectively. This equation forms the basis to compute the column density of any H I cloud. Even though there are only two unknown quantities to infer a column density, it is difficult to disentangle those quantities and measure them independently. Assumptions of the spin temperatures and the optical depths are required to derive an estimation of the column density. Absorption measurements against strong continuum sources can provide optical depths that can be used to correct for optical depth effects in H I emission. Furthermore, H I self-absorption will be studied in greater detail in this thesis. This is a viable method to isolate cold H I in the Galactic plane from the warmer more diffuse H I emission, and will be discussed in the following section.

2.2 H I self-absorption – HISA

Emission maps of the H I spectral line are generally poor tracers of cold gas since the H I line intensity is proportional to excitation temperature when optically thick and is independent of temperature when optically thin. On the other hand, continuum absorption methods, that probe cold H I and provide a direct measurement of the spin temperature, are hampered by the small number of bright continuum background sources, leading to a poorly sampled census of cold H I in the ISM without

charting the gas distribution over any large area. A direct view of the spatial structure of cold atomic gas is provided by observations of H I self-absorption (HISA; Knapp, 1974; Feldt, 1993; Gibson et al., 2000) against background H I emission (see Fig. 1.4), first measured in 1954 (Heeschen, 1954, 1955).

It is common in the Galactic plane that cold hydrogen absorbs the H I emission from warmer H I gas. Self-absorption does not specifically refer to a physical process acting within a single cloud, rather it is an observable that can be induced even by H I emission originating in the far background that has the same radial velocity $V_{\rm LSR}$ as the hydrogen cloud of interest. At low Galactic latitudes the typical brightness temperatures of H I emission is 50–100 K toward the inner Galactic plane, thus the cold H I can appear as positive emission features, as discussed in the previous section, or as HISA if its spin temperature is lower than the brightness temperature of the background.

H I self-absorption is found throughout the Milky Way and has been studied in great detail (Riegel & Crutcher, 1972; Knapp, 1974; van der Werf et al., 1988; Gibson et al., 2000, 2005; Kavars et al., 2003; Dénes et al., 2018; Wang et al., 2020c; Syed et al., 2020). High-angular resolution surveys allow a detailed view of internal cloud structure but the global distribution of H I requires a broad sky coverage. The physical relationship between the CNM and other ISM constituents like dust and H_2 are difficult to investigate without unbiased, large-scale surveys.

Many studies have focused on the detection of HISA toward known sources, but statistical treatments of the kinematic properties and densities of the HISA-traced cold gas in large-scale high-resolution maps do not exist. In the context of molecular cloud formation, HISA is usually associated with molecular gas. Extensive surveys of HISA toward cold molecular clouds have shown that HISA correlates well with molecular line tracers, such as OH and ^{13}CO (Li & Goldsmith, 2003; Krčo et al., 2008). The measured spin temperatures of the cold H I are often consistent with the temperature measured with CO lines. The cold atomic gas observed through HISA in these surveys is therefore a trace constituent of these molecular clouds, likely probing the central regions where the cold atomic gas is well mixed with the molecular gas. The close correlation between the atomic and molecular gas properties can result in small HISA line widths, typically $\lesssim 1\,{\rm km\,s^{-1}}$, which is referred to as narrow H I self-absorption (HINSA; Li & Goldsmith, 2003; Goldsmith et al., 2007; Krčo et al., 2008). Beyond the central regions of dense molecular clouds, HISA is expected to probe cold atomic hydrogen toward the outer layers of a cloud as well, where HISA likely traces hydrogen gas less cold since the gas is more exposed to the photo-dissociating effects of the interstellar radiation field. These are more diffuse HISA features that can be traced over larger angular scales (see e.g., Wang et al., 2020c; Syed et al., 2020). However, it is not the case that all detected HISA features correspond to molecular lines. In fact, much of the cold atomic gas might be traced by HISA without any molecular line counterpart as the CNM is more widespread in the Galaxy than the molecular gas (Gibson et al., 2000).

HISA has been used to resolve the kinematic distance ambiguity (see Sect. 1.1.2) for molecular clouds or Galactic H II regions in the inner Galaxy. Sources of interest at the far distance are less likely to show HISA as there is less background to absorb. HISA can therefore be used to determine whether molecular clouds and H II regions

should be placed at the near or far kinematic distance (e.g., Jackson et al., 2002; Kolpak et al., 2003; Anderson & Bania, 2009; Roman-Duval et al., 2009; Duarte-Cabral et al., 2021).

For HISA to be detected, sufficient background emission along the line of sight is required. Since the warm component of atomic hydrogen is more diffuse, it fills up a larger volume than the cold component (volume filling factor $f \sim 0.4$, Kalberla & Kerp 2009). Observed H I emission can also originate from clouds that are situated in front of cold, self-absorbing clouds. In addition, we observe 1.4 GHz continuum emission that has to be taken into account. This can be a diffuse Galactic component or originate from discrete strong sources. Figure 2.3 illustrates the line-of-sight

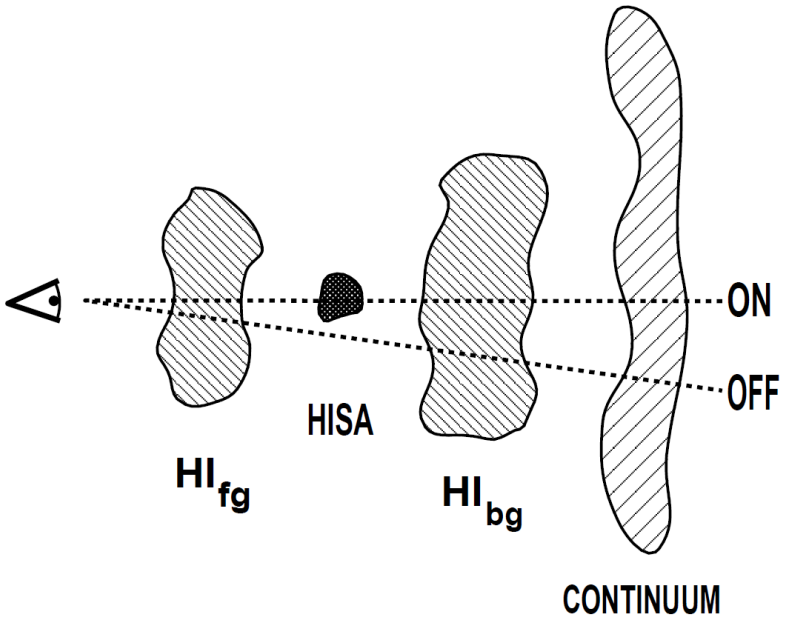


Figure 2.3: Line-of-sight geometry of H I clouds along different lines of sight. This sketch shows the presence of a self-absorbing H I cloud (HISA) between a foreground ($\mathrm{H\,I_{fg}}$) and background ($\mathrm{H\,I_{bg}}$) cloud. Diffuse continuum emission is assumed to be in the far background. Taken from Gibson et al. (2000).

geometry of H I emitting and absorbing clouds along different sight lines. For simplicity, we assume that the continuum source is located in the far background. This geometry is adopted from Gibson et al. (2000) and serves as a starting point to solve the radiative transfer equation. The radiative transfer equation, including HISA as well as foreground and background emission sources, is a sum of individual emission and absorption events along the line of sight. Assuming the viewing geometry in Fig. 2.3, we observe H I emission components located in the foreground and in the background, with H I spin temperatures of T_{fg} and T_{bg}, respectively. Cold H I with spin temperature T_{HISA} is located between the foreground and background H I emission sources, thus rendering it possible to observe the cloud in self-absorption. In addition, a diffuse continuum component with T_{cont} adding to the radiative transfer is taken into account, which is assumed to originate in the far background. A comprehensive discussion of the radiative transfer of HISA clouds is given in Gibson et al. (2000), Kavars et al. (2003), and Li & Goldsmith (2003) and I will briefly

discuss the fundamental properties of the radiative transfer equation.

Comparing the observed brightness along a line of sight toward self-absorbing gas ($T_{\rm on}$) with a line of sight that does not include HISA ($T_{\rm off}$), we can solve for the optical depth of HISA as a function of spin temperature. If HISA is observed along the line of sight, the observed brightness temperature will be (see Eq. 2.1.18)

$$\begin{aligned} T_{\rm on} = \;& T_{\rm fg}\,\left(1-e^{-\tau_{\rm fg}}\right) \\ & + T_{\rm HISA}\,\left(1-e^{-\tau_{\rm HISA}}\right)\,e^{-\tau_{\rm fg}} \\ & + T_{\rm bg}\,\left(1-e^{-\tau_{\rm bg}}\right)\,e^{-(\tau_{\rm fg}+\tau_{\rm HISA})} \\ & + T_{\rm cont}\,e^{-(\tau_{\rm fg}+\tau_{\rm HISA}+\tau_{\rm bg})} - T_{\rm cont}\;, \end{aligned} \tag{2.2.1}$$

where the subscripts "fg" and "bg" refer to foreground and background gas relative to the HISA cloud, respectively. Toward an off position, in the absence of HISA along the line of sight, the observed brightness temperature is

$$\begin{aligned} T_{\rm off} = \;& T_{\rm fg}\,\left(1-e^{-\tau_{\rm fg}}\right) \\ & + T_{\rm bg}\,\left(1-e^{-\tau_{\rm bg}}\right)\,e^{-\tau_{\rm fg}} \\ & + T_{\rm cont}\,e^{-(\tau_{\rm fg}+\tau_{\rm bg})} - T_{\rm cont}\;. \end{aligned} \tag{2.2.2}$$

In either case, τ_i denotes the optical depth of the respective component i. As is common during the reduction of the observational data, continuum emission has been subtracted from the H I data that I will utilize (see Chapter 3), which is indicated by the last term ($-T_{\rm cont}$). H I self-absorption occurs if $T_{\rm off} \gtrsim T_{\rm on}$. If we then subtract the observed off spectrum from the on spectrum, we obtain the H I self-absorption spectrum

$$\begin{aligned} T_{\rm on} - T_{\rm off} &= T_{\rm HISA}\left(1-e^{-\tau_{\rm HISA}}\right)e^{-\tau_{\rm fg}} - T_{\rm bg}\left(1-e^{-\tau_{\rm bg}}\right)\left(1-e^{-\tau_{\rm HISA}}\right)e^{-\tau_{\rm fg}} \\ &\quad - T_{\rm cont}\left(1-e^{-\tau_{\rm HISA}}\right)e^{-(\tau_{\rm fg}+\tau_{\rm bg})} \\ &= \left(T_{\rm HISA} - T_{\rm bg}\left(1-e^{-\tau_{\rm bg}}\right) - T_{\rm cont}e^{-\tau_{\rm bg}}\right) \times \left(1-e^{-\tau_{\rm HISA}}\right)e^{-\tau_{\rm fg}}\;. \end{aligned} \tag{2.2.3}$$

This equation can be simplified if we introduce the dimensionless parameter $p_{\rm bg}$ (e.g., Feldt, 1993; Gibson et al., 2000)

$$p_{\rm bg} \equiv \frac{T_{\rm bg}\,\left(1-e^{-\tau_{\rm bg}}\right)}{T_{\rm off}}\;, \tag{2.2.4}$$

where $p_{\rm bg}$ can be interpreted as the fraction of background emission in the optically thin limit. It is a fair approximation to assume that the WNM dominates the emission, at least over a larger angular scale. The foreground and background emission components are then likely to be optically thin. Given the high kinetic gas temperatures and spin temperatures of the WNM, only small optical depths can produce the observed brightness temperatures that are typically on the order of $\sim$100 K. Under this assumption, Eq. (2.2.3) becomes

$$T_{\rm on} - T_{\rm off} = \left(T_{\rm HISA} - p_{\rm bg}T_{\rm off} - T_{\rm cont}\right) \times \left(1-e^{-\tau_{\rm HISA}}\right)\;. \tag{2.2.5}$$

This equation relates the observed quantities $T_{\rm on}$, $T_{\rm off}$, and $T_{\rm cont}$ to the physical properties of a HISA cloud. The measurement of $T_{\rm off}$ is, however, a challenging task

since finding a line of sight that is not affected by self-absorption would require prior knowledge of a cloud's location. Several approaches have been proposed to estimate $T_{\rm off}$, that will be the subject of Chapter 5.

In the optically thin limit, a background fraction of $p_{\rm bg} = 1$ implies there is no foreground emission along our line of sight. The background fraction is always $0 \leq p_{\rm bg} \leq 1$, however, the fraction is strongly biased toward values closer to one. The detection of HISA is more sensitive to stronger background emission, such that the absorption depth in Eq. (2.2.5) will be larger for higher background fraction. Furthermore, the kinematic distance ambiguity favors the kinematic near distance as the location of any observed HISA cloud in the inner Galactic plane. H I emission with the same radial velocity originating in the far distance is then likely to provide the required background for HISA to be observed. The background fraction should then be > 0.5. Measuring this parameter is difficult but considerations about the viewing geometry and global H I distribution can help to constrain this value (see Chapter 5).

As can be seen in Eq. (2.2.5), the absorption depth depends on both the optical depth $\tau_{\rm HISA}$ and spin temperature $T_{\rm HISA}$ of the HISA gas. These two properties cannot be measured independently using this method. Many studies assume a fixed spin temperature to constrain the optical depth. This will be revisited in Chapter 5. To elaborate how the optical depth behaves as a function of spin temperature, we can rewrite Eq. (2.2.5) as

$$\tau_{\rm HISA} = -\ln\left(1 - \frac{T_{\rm on} - T_{\rm off}}{T_{\rm HISA} - p_{\rm bg}T_{\rm off} - T_{\rm cont}}\right) . \tag{2.2.6}$$

Figure 2.4 shows a family of curves with fixed values of $T_{\rm on} = 40\,{\rm K}$, $T_{\rm off} = 70\,{\rm K}$,

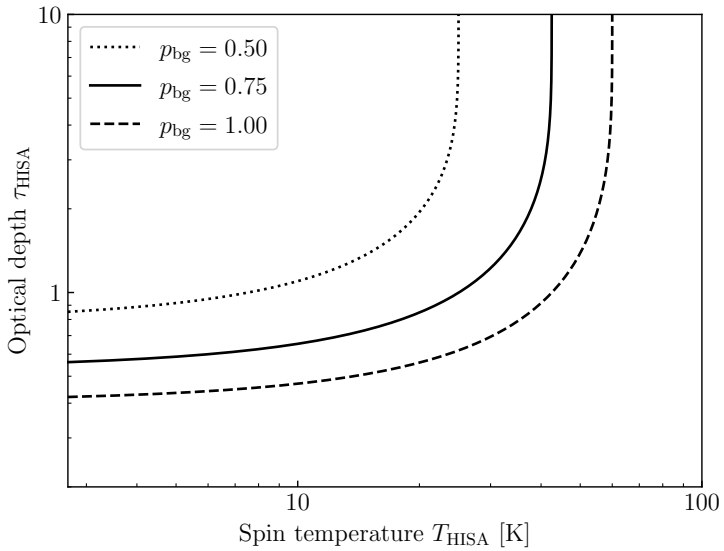


Figure 2.4: Optical depth $\tau_{\rm HISA}$ as a function of spin temperature $T_{\rm HISA}$ at the derived mean temperatures of $T_{\rm on} = 40\,{\rm K}$, $T_{\rm off} = 70\,{\rm K}$, and $T_{\rm cont} = 20\,{\rm K}$ (Eq. 2.2.5). The dotted, solid, and dashed curves represent the optical depth with background fractions 0.50, 0.75, and 1.00, respectively.

and $T_{\rm cont} = 20\,{\rm K}$, which are typical values for the H I clouds discussed in this thesis

(see Chapters 5 and 6). The dotted, solid, and dashed curve represent the optical depth at different background fractions $p_{\rm bg} = 0.50$, 0.75, 1.00, respectively. The lower limit of the spin temperature is the cosmic microwave background emission of $\sim$2.7 K. In theory, $\tau_{\rm HISA}$ is minimized for $T_{\rm HISA} = 0\,{\rm K}$ and $p_{\rm bg} = 1$. At the other extreme, optically thick clouds where $\tau_{\rm HISA} \rightarrow \infty$ put an upper limit to the spin temperature that is given by

$$T_{\rm HISA}({\rm max}) = T_{\rm on} + T_{\rm cont} - (1 - p_{\rm bg})\, T_{\rm off}\ . \tag{2.2.7}$$

It is important to note that this limit is not the maximum spin temperature of the cold gas but an observational limit of the gas temperature to which the viewing geometry allows the detection of HISA.

3

Methods and Observation

This following chapter presents the observational data that are used throughout this thesis and briefly introduces the principal methods to conduct the scientific analysis of the H I data.

3.1 THOR

Different kinds or setups of radio telescopes are useful to measure H I emission and absorption at different angular scales. As for all observations, angular resolution is an important consideration, particularly for the detection of the less diffuse CNM. Interferometers are standard for achieving high resolution at radio frequencies as their angular resolution is approximated by

$$\Theta \approx \frac{\lambda_0}{B_{\mathrm{max}}} \,, \tag{3.1.1}$$

where λ_0 is the wavelength of the spectral line at rest and B_{max} is the maximum baseline, that is the largest distance between two antennas of the interferometer. However, even though interferometers can achieve high angular resolution, they have one fundamental weakness in not measuring uv spacings[1] shorter than the minimum distance B_{min} between antennas. By design, interferometers only provide images of the sky brightness that is high-pass filtered. The images are convolved with a spatial filter that removes Fourier components with low spatial frequencies. The brightness distribution at spatial frequencies corresponding to the angular resolution $\Theta \approx \lambda_0/B_{\mathrm{min}}$ and lower is irretrievably lost by the interferometer. For a detailed overview of radio interferometers and aperture synthesis I refer the reader to Wilson et al. (2013). These lost low spatial frequencies contain much or even most of the brightness of H I line so single dish observations have to be included to complement the H I brightness.

[1] The uv space refers to the Fourier domain corresponding to the image.

The H I data used in this thesis, that the following analyses on the Galactic properties of H I are based on, are taken from The H I/OH/Recombination line survey of the inner Milky Way (THOR; Beuther et al., 2016; Wang et al., 2020a). The THOR survey is a Galactic plane survey covering Galactic longitudes in the range $14.0° < \ell < 67.4°$ and latitudes $|b| < 1.25°$. The survey has been carried out with the Very Large Array (VLA) in C-configuration, observing in L band between 1–2 GHz. An overview of THOR is given in the two data release papers by Beuther et al. (2016) and Wang et al. (2020a). The survey includes observations of the H I 21 cm line, four OH transition lines, 19 radio recombination lines (RRLs), as well as the continuum between 1–2 GHz. For the H I observations, the THOR C-

Figure 3.1: The Milky Way seen over the Karl G. Jansky Very Large Array west of Socorro, New Mexico. Credit: NRAO/AUI/NSF, Jeff Hellerman

configuration data have been combined with data from the H I Very Large Array Galactic Plane Survey (VGPS; Stil et al., 2006), that consist of VLA D-configuration data combined with single-dish observations taken with the Green Bank Telescope (GBT), to account for the missing flux on short uv spacings. In order to image the data, the THOR C-configuration data were first combined with the VGPS D-configuration data in the visibility domain. The continuum has been subtracted from the visibility data set before imaging the data using the multiscale CLEAN algorithm provided by CASA[2]. The images were then combined with the images of the VGPS survey (D-configuration+GBT) using the task "feather" in CASA. This procedure for the image reconstruction has shown to be less affected by sidelobe artifacts. Since the D-configuration data of VGPS cover a slightly smaller longitude range $17.6° < \ell < 67.0°$, the combined H I data are restricted to that coverage. The combined H I data (THOR+VGPS), in the following "THOR-H I" data, have been smoothed to an angular resolution of $\Theta = 40''$ to improve brightness sensitivity. The data have an rms noise of $\sim$4 K at the spectral resolution of 1.5 km s^{-1}, measured in emission-free channels. The radial velocity ranges from -113 to $+163$ km s^{-1}.

The THOR survey also provides continuum data both at the native angular resolution of $\sim 14''$ and at $25''$. The THOR 1.4 GHz continuum data have been combined with the continuum data of the VGPS survey that include VLA D-configuration data and single dish data taken with the Effelsberg 100 m telescope. The continuum data

[2]Common Astronomy Software Applications; Version 5.1.1.

are used to account for continuum emission when investigating the HISA properties in Chapter 5. In addition to the combined H I data, C-configuration-only data at the native angular resolution of $\sim$14″ are used, where the continuum has not been subtracted. This can provide measurements of the H I optical depth against bright continuum background sources (Bihr et al., 2016; Wang et al., 2018).

H I emission spectra toward the inner Galactic midplane typically show multiple spectral components due to atomic hydrogen's ubiquity in the Galaxy and pose a difficult challenge in interpreting Galactic structure. Figure 1.4 has illustrated this complexity. The emission can be associated with a wide range of physical properties and the positive velocity gas shows additional confusion along the line of sight due to the kinematic distance ambiguity. The presence of self-absorption can exacerbate the difficulties analyzing the spectrum (Fig. 1.4). The emission from the outer Galaxy is associated with less line-of-sight confusion and allows that H I clouds can be discerned in the emission more easily.

Figure 3.3 presents an overview of the THOR-H I data showing moment 0 maps, integrated over velocity intervals of 15 km s^{-1} in the range $[-71, +124]$ km s^{-1}. In the first Galactic quadrant, the negative velocities are tracing the outer Galaxy with less line-of-sight confusion, while at positive velocities the kinematic distance ambiguity blends the emission at different distances. The midplane of the H I disk at negative velocities is located at higher latitudes due to the H I disk being strongly warped at large Galactocentric distances (Henderson et al., 1982; Burton & te Lintel Hekkert, 1986; Nakanishi & Sofue, 2003).

3.2 Spectral decomposition

The kinematic properties of individual H I clouds along the line of sight are generally confused and it is a complex task to decompose emission or absorption spectra into separate velocity components. We therefore make use of the fully automated spectral decomposition tool GaussPy+ (Riener et al., 2019) to derive the kinematic properties of H I in an unbiased way. The GaussPy+ package is based on the previously developed GaussPy algorithm (Lindner et al., 2015) and comes with physically-motivated improvements specifically designed to analyze spectral line data of Galactic plane surveys. The GaussPy+ tool includes an automated noise estimation routine and a spatial coherence check that takes into account neighboring pixel spectra. The underlying principle of GaussPy is an autonomous Gaussian decomposition technique that decomposes emission or absorption spectra into Gaussian components.

GaussPy automatically determines initial guesses of Gaussian fit components for each spectrum using derivative spectroscopy. This technique calculates second and higher order derivatives to find maxima and minima in spectra in order to locate signal peaks. However, noise fluctuations in observational spectra are greatly amplified in higher order derivatives. It is therefore integral to smooth the spectra prior to the decomposition to remove noise fluctuations while retaining real signal (see Fig. 2 in Lindner et al., 2015). The optimal smoothing kernels are found employing a supervised machine learning routine that trains on a few hundred well-fit spectra taken from a subset of the original data. For a detailed description of GaussPy

and GaussPy+, I refer the reader to Lindner et al. (2015) and Riener et al. (2019).

3.3 Baseline extraction of HISA features

In Chapters 5 and 6 we obtain the baselines of H I self-absorption features $T_{\rm off}$ (see Sect. 2.2) by employing the newly developed tool astroSaber[3]. The algorithm astroSaber makes use of asymmetric least squares smoothing (Eilers, 2004) to average out dips in H I emission spectra, such that the smooth spectra can be used as baselines $z(= T_{\rm off})$ to extract HISA features. The averaging, or smoothing, is done by a least squares method with asymmetric weights w_i and a regularization term $\mathcal{R}(z)$ defined by the second derivative representation of the spectrum. The final baseline is then found by minimizing the least squares function

$$F(z) = \sum_{i}^{n} w_i (y_i - z_i)^2 + \lambda \mathcal{R}(z) \,, \tag{3.3.1}$$

where w_i are the weights of the squared residuals (see the details given in Chapter 5), y_i and z_i are the input H I emission spectrum and baseline z at velocity channel i, respectively, and $\mathcal{R}(z)$ is the regularization that imposes a penalty on the fluctuation of z. The parameter λ that controls the amount of smoothing is found by a gradient descent that minimizes the loss between some input data (an assumed H I emission spectrum free of absorption) and the output data (the HISA baseline) to be obtained. The input and output data are generated from the H I emission data set itself. This method allows us to retrieve baselines of HISA features independent of the velocities of H I clouds. A detailed introduction and discussion of astroSaber is presented in Chapter 5.

Figure 3.2: astroSaber logo. Rendered with blender 3.1.0. Image: J. Syed, CC BY-NC-SA 3.0

[3]The code is publicly available at https://github.com/astrojoni89/astrosaber.

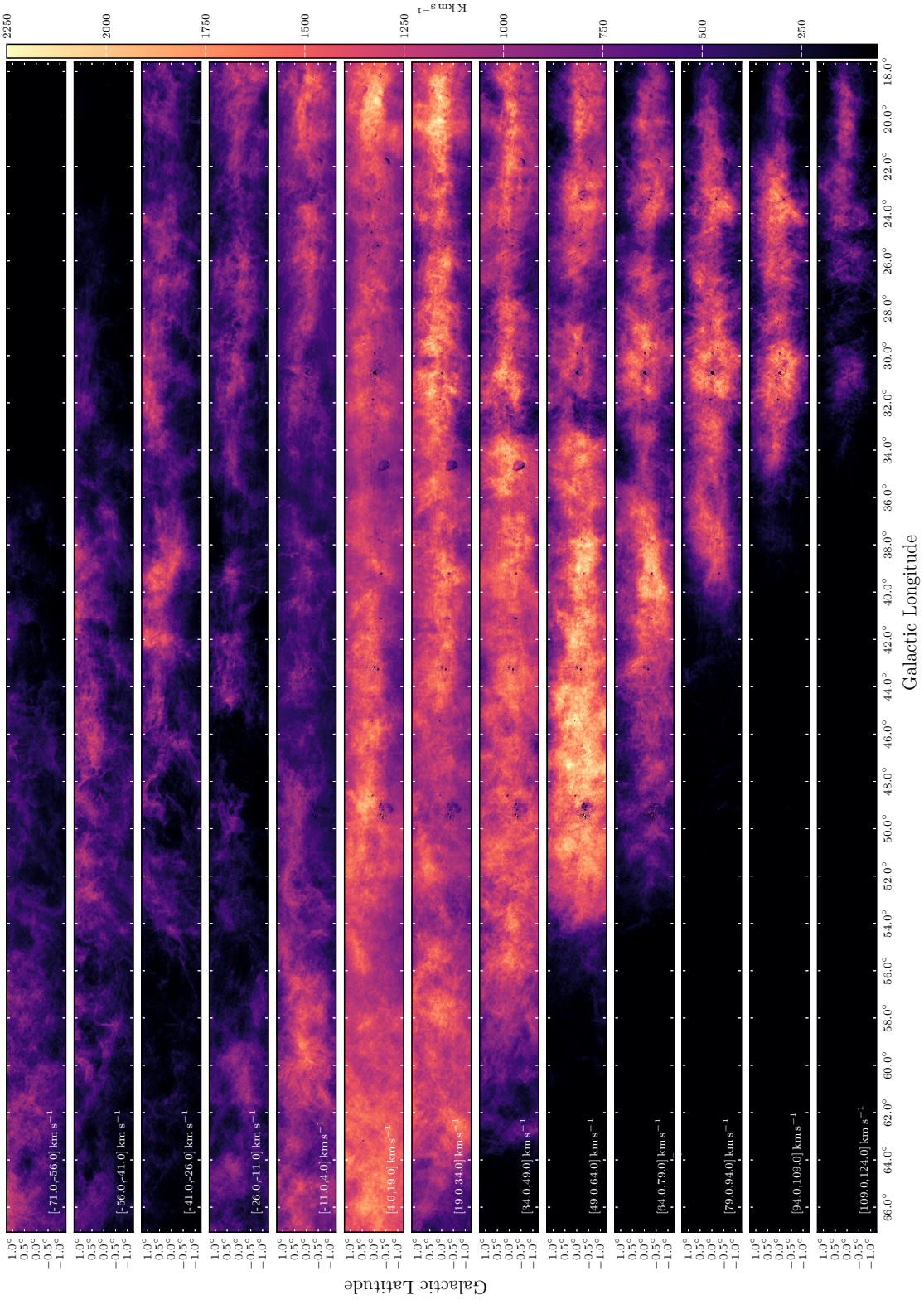


Figure 3.3: THOR-H I moment 0 maps obtained by integrating the H I emission over the velocity ranges indicated in each panel.

4

The "Maggie" filament

Physical properties of a giant atomic cloud

J. Syed, *J. D. Soler, H. Beuther, Y. Wang, S. Suri, J. D. Henshaw, M. Riener, S. Bialy, S. Rezaei Kh., J. M. Stil, P. F. Goldsmith, M. R. Rugel, S. C. O. Glover, R. S. Klessen, J. Kerp, J. S. Urquhart, J. Ott, N. Roy, N. Schneider, R. J. Smith, S. N. Longmore, H. Linz*

This following Chapter is based on the work *The "Maggie" filament: Physical properties of a giant atomic cloud* published in the January, 2022 issue of Astronomy&Astrophysics (Syed et al., 2022).

The published paper has been reformatted to match the format of this thesis. The work on the paper was distributed as follows: I conducted the scientific analyses presented in this Chapter. In Sects. 4.3.1 and 4.4.2 co-authors provided pieces of code to analyze the data. Extensive discussions with S. Bialy led to the discussion section 4.4.4.

I drafted a first version of the entire manuscript. Co-authors contributed by commenting on at least one version of the draft and giving suggestions on additional analysis to be included.

ABSTRACT

The atomic phase of the interstellar medium plays a key role in the formation process of molecular clouds. Due to the line-of-sight confusion in the Galactic plane that is associated with its ubiquity, atomic hydrogen emission has been challenging to study. We investigate the physical properties of the "Maggie" filament, a large-scale filament identified in H I emission at line-of-sight velocities, $v_{\rm LSR} \sim -54\,{\rm km\,s^{-1}}$. Employing the high-angular resolution data from The H I/OH Recombination line survey of the inner Milky Way (THOR), we have been able to study H I emission features at negative $v_{\rm LSR}$ velocities without any line-of-sight confusion due to the kinematic distance ambiguity in the first Galactic quadrant. In order to investigate the kinematic structure, we decomposed the emission spectra using the automated Gaussian fitting algorithm GaussPy+. We identify one of the largest, coherent, mostly atomic H I filaments in the Milky Way. The giant atomic filament Maggie, with a total length of $1.2 \pm 0.1\,{\rm kpc}$, is not detected in most other tracers, and it does not show signs of active star formation. At a kinematic distance of 17 kpc, Maggie is situated below (by $\approx$500 pc), but parallel to, the Galactic H I disk and is trailing the predicted location of the Outer Arm by $5 - 10\,{\rm km\,s^{-1}}$ in longitude-velocity space. The centroid velocity exhibits a smooth gradient of less than $\pm 3\,{\rm km\,s^{-1}}\,(10\,{\rm pc})^{-1}$ and a coherent structure to within $\pm 6\,{\rm km\,s^{-1}}$. The line widths of $\sim$10 km s^{-1} along the spine of the filament are dominated by nonthermal effects. After correcting for optical depth effects, the mass of Maggie's dense spine is estimated to be $7.2^{+2.5}_{-1.9} \times 10^5\,M_\odot$. The mean number density of the filament is $\sim 4\,{\rm cm^{-3}}$, which is best explained by the filament being a mix of cold and warm neutral gas. In contrast to molecular filaments, the turbulent Mach number and velocity structure function suggest that Maggie is driven by transonic to moderately supersonic velocities that are likely associated with the Galactic potential rather than being subject to the effects of self-gravity or stellar feedback. The probability density function of the column density displays a log-normal shape around a mean of $\langle N_{\rm H\,I} \rangle = 4.8 \times 10^{20}\,{\rm cm^{-2}}$, thus reflecting the absence of dominating effects of gravitational contraction. While Maggie's origin remains unclear, we hypothesize that Maggie could be the first in a class of atomic clouds that are the precursors of giant molecular filaments.

4.1 Introduction

Stars form in the cold, dense interiors of molecular clouds. The physical properties of these clouds therefore set the initial conditions under which star formation takes place. A key question in understanding the star formation process as part of the global interstellar matter cycle addresses the formation of large-scale molecular clouds out of the diffuse atomic phase of the interstellar medium (ISM; for a review see Ferrière, 2001; Draine, 2011; Klessen & Glover, 2016). However, the atomic ISM and its dynamical properties are still observationally poorly constrained.

The ISM has a hierarchical structure and facilitates the formation of filaments that are governed by the Galactic potential on a large scale. Soler et al. (2020) present a network of H I filaments already evident in the diffuse atomic phase of the ISM that is structured mostly parallel to the Galactic plane. Only in a few cases is the orientation of the filaments locally no longer dictated by the overall drag of the Galactic disk, but rather it is determined by the effects of stellar feedback and strong magnetic fields. Soler et al. (2020) identified a unique H I filament that they have named "Maggie", after the largest river in Colombia, the Río Magdalena. Maggie is shown to be a highly elongated filamentary cloud (see Fig. 4.1) extending over $\sim 4°$

on the sky in Galactic longitude. Given its central velocity of $v_{\rm LSR} \approx -54\,{\rm km\,s^{-1}}$ and assuming circular motion, Maggie is located approximately 17 kpc away from us and has a length of more than 1 kpc.

In this paper, we present a detailed study of the Maggie filament and aim to understand its physical nature. Moreover, we investigate the physical and kinematic properties in comparison with large-scale molecular filaments. In this context, it is an open question if the properties of filamentary molecular clouds are generally inherited from an atomic counterpart in the diffuse ISM. Giant molecular filaments are to date the largest coherent entities identified in the Milky Way and have been subject of many recent studies investigating the evolution of large filaments with respect to the global dynamics of the Milky Way on the one hand and local stellar feedback on the other hand (Jackson et al., 2010; Ragan et al., 2014; Goodman et al., 2014; Wang et al., 2015, 2016; Abreu-Vicente et al., 2016; Zucker et al., 2015, 2018; Wang et al., 2020b).

The highly filamentary infrared dark cloud (IRDC) "Nessie", first identified in Jackson et al. (2010), is argued to be one of the first in a class of filaments whose morphology is likely governed by the structural dynamics of the Galaxy. Goodman et al. (2014) dubbed this type of filaments "bones" of the Milky Way – highly filamentary molecular clouds whose formation, evolution, and shape could be closely linked to the global spiral structure of the Galaxy.

Zucker et al. (2018) utilized a standardized approach to re-analyze a set of Galactic filaments identified in the literature. In doing that, it is possible to draw meaningful conclusions and reliably compare physical properties between filaments rather than being subject to the systematics of selection criteria and different methodology. They found in their sample of filaments that the giant molecular filaments (GMFs), first identified in Ragan et al. (2014) and Abreu-Vicente et al. (2016), have the highest masses ($\sim 10^5\,M_\odot$) while exhibiting the lowest column densities and star-forming activity, making them good candidates to be the immediate descendants of atomic ISM structures.

The GMFs are first identified as near and mid-infrared extinction features with a spatial extent of $\approx$100 pc. Ragan et al. (2014) and Abreu-Vicente et al. (2016) then confirm velocity contiguity of the filaments via ^{13}CO emission. Giant molecular filaments are not only associated with spiral arms but are also located in inter-arm regions.

The recently discovered Radcliffe wave (Alves et al., 2020) is a coherent 2.7 kpc long association of local molecular cloud complexes. It appears to be undulating above and below the Galactic midplane, and its three-dimensional shape (in position-position-position space) is well described by a damped sinusoidal wave. The Radcliffe wave provides a framework for future studies of molecular cloud formation and evolution with respect to the Galactic dynamics of the Milky Way.

Simulations find that giant molecular clouds often form as large filaments, and their formation can be related to the dynamics of the galaxy and position with respect to the spiral arm potentials. Smith et al. (2014a) use simulations of a four-armed spiral galaxy to investigate the formation of molecular gas structures. They find that high-density filaments tend to form in spiral arms while lower-density gas resides in long inter-arm spurs that are stretched by galactic shear. While the

simulations account for the chemical evolution of the gas, they do not include the effects of self-gravity of the gas or stellar feedback.

Including both stellar feedback and self-gravity, smoothed particle hydrodynamics simulations of a high-resolution section of a spiral galaxy indicate that galactic shear between spiral arms plays a critical role in the formation of highly filamentary structures (Dobbs, 2015; Duarte-Cabral & Dobbs, 2016). Duarte-Cabral & Dobbs (2017) expand this analysis to the time evolution of these filaments, finding that giant molecular clouds tend to sustain their large filamentary shape before entering the spiral arms, where they are prone to being disrupted by local events of star formation and stellar feedback.

The CloudFactory simulations presented in Smith et al. (2020) strongly suggest that spiral arms and differential rotation tend to arrange molecular clouds in long filaments that are likely to fragment while regions of clustered feedback randomizes the orientation of them. Molecular filaments that are dominated by galactic-scale forces also exhibit low internal velocity gradients and are tightly confined to the galactic plane. Filaments formed in regions of higher turbulence due to supernova (SN) feedback show a range of orientations with respect to the galactic plane and are more widely distributed.

The systematic search for atomic counterparts to molecular clouds poses a challenging task. By means of the 21cm-line of H I emission it is generally possible to probe large atomic clouds in the ISM but it has proven difficult to identify them as clearly defined objects in the inner Galactic plane. Traditionally, H I clouds are then either observed at high Galactic latitudes (e.g., Kalberla et al., 2016) or they must have velocities significantly different than those imposed by the Galactic rotation (e.g., Wakker & van Woerden, 1997; Westmeier, 2018).

As the analysis of H I emission in the Galactic plane within the solar circle suffers from the kinematic distance ambiguity, any H I cloud with $v_{\mathrm{LSR}} \gtrsim 0\,\mathrm{km\,s^{-1}}$ (in the first Galactic quadrant) might be the product of blending foreground and background components that correspond to the same line-of-sight velocity. However, at negative v_{LSR} velocities any H I structure can be identified with less line-of-sight confusion[1] (see e.g., Brand & Blitz, 1993).

Additionally, the physical properties of atomic hydrogen are not straightforward to derive from emission studies alone. In thermal pressure equilibrium, theoretical considerations based on ISM heating and cooling processes predict two stable phases of atomic hydrogen at the observed pressures in the ISM, namely the cold neutral medium (CNM) and warm neutral medium (WNM; Field et al., 1969; McKee & Ostriker, 1977; Wolfire et al., 2003; Bialy & Sternberg, 2019). Observations of H I emission are thus generally attributed to both CNM and WNM, which have significantly different physical properties. In an attempt to observationally isolate the CNM from the bistable emission, H I self-absorption (HISA; Gibson et al., 2000; Li & Goldsmith, 2003; Wang et al., 2020c; Syed et al., 2020) is a viable method to study H I clouds in the inner Milky Way but it heavily depends on the presence of sufficient background emission.

The CNM is observed to have temperatures $\lesssim 300\,\mathrm{K}$ and number densities of $\gtrsim n_{\mathrm{min,CNM}} = 10\,\mathrm{cm^{-3}}$ while the thermally stable WNM exceeds temperatures of

[1] vice versa in the fourth Galactic quadrant.

$\sim$5000 K with number densities $\lesssim n_{\rm max,WNM} = 0.1\,{\rm cm}^{-3}$ (Heiles & Troland, 2003; Kalberla & Kerp, 2009). For densities between $n_{\rm min,CNM}$ and $n_{\rm max,WNM}$, the gas is thermally unstable (denoted by UNM: unstable neutral medium) and it will move toward a stable CNM or WNM branch under isobaric density perturbations (Field, 1965).

Throughout the scope of this paper, we use the analytic model presented in Wolfire et al. (2003) as a standard reference. According to this, the two stable atomic hydrogen phases can coexist in thermal equilibrium over a narrow range of pressures, $P_{\rm min} \leq P \leq P_{\rm max}$, that is governed by the total heating and cooling rates of the interstellar gas (see e.g., the phase diagrams shown in Fig. 7 in Wolfire et al. 2003). The dominant heating process under typical ISM conditions is photoelectric (PE) heating from dust grains and polycyclic aromatic hydrocarbons (PAHs). The WNM is mainly cooled by Lyα emission while the most efficient cooling mechanism of the CNM is metal-line fine-structure emission primarily excited by collisions with electrons and neutral hydrogen atoms. Assuming a constant ratio of the cosmic-ray (or X-ray) ionization rate, ζ, to the intensity of the UV interstellar radiation field, $I_{\rm UV}$, PE heating is proportional to $I_{\rm UV}$. If $I_{\rm UV}$ is then increased (decreased), the density at which metal-line cooling balances heating must increase (decrease) in the same manner. As a result, the pressure range at which both WNM and CNM occur shifts to higher or lower values in proportion to $I_{\rm UV}$.

Furthermore, the metallicity Z' and dust abundance are crucial parameters for the thermal balance. If the metallicity is moderately lowered, the equilibrium pressure range dictating the WNM and CNM properties decreases since the dust-to-gas ratio and associated PE heating rate decreases more rapidly with decreasing Z' than does the metal-line cooling[2] (Bialy & Sternberg, 2019, see their Fig. 4).

The Maggie filament has been discovered in Soler et al. (2020) via H I emission in The H I/OH Recombination line survey of the inner Milky Way (THOR; Beuther et al., 2016; Wang et al., 2020a). It is identified as a large coherent filament using a Hessian matrix approach. This method allows to systematically identify filamentary structures by their spatial curvature that emerge as second derivative signatures in the H I emission maps of the THOR survey.

We follow up on this discovery and derive the physical properties of the Maggie filament in the subsequent analysis. This paper is organized as follows. We introduce the observations and data in Sect. 4.2 and outline the method we use for the spectral decomposition of the H I emission. We present the fundamental properties of Maggie and show the results of the kinematics and column density in Sect. 4.3. In Sect. 4.4 we examine Maggie for molecular counterparts and explore the possible formation process of Maggie. We also discuss the implications of the velocity and column density structure. Finally, we review the relationship between Maggie and giant molecular filaments and draw our conclusions in Sect. 4.5.

[2]Here we do not consider the effect of cosmic ray heating that becomes the dominant heating mechanism at metallicities $Z' \lesssim 0.1$ (see Fig. 2 in Bialy & Sternberg, 2019).

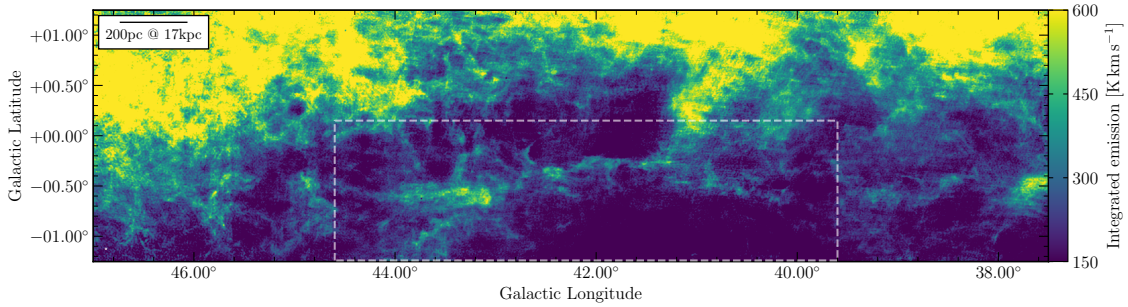


Figure 4.1: Overview of Maggie. The map shows the integrated H I emission in the velocity interval between -57.5 and $-48.5\,\mathrm{km\,s^{-1}}$. Maggie is located below the Galactic midplane. The white dashed box containing Maggie marks the region that is considered in the subsequent analysis.

4.2 Observation and methods

4.2.1 H I 21 cm line and continuum

In the following analysis we employ the H I and 1.4 GHz continuum data from the THOR survey. A complete overview of the THOR survey and the data products is given in Beuther et al. (2016) and Wang et al. (2020a). We investigate the kinematics based on the H I emission data, where the 1.4 GHz continuum emission has been subtracted from the THOR H I observation during data reduction. The final H I emission data (THOR-H I) have been obtained from observations with the Very Large Array (VLA) in C- and D-configuration, as well as single dish observations from the Greenbank Telescope. The data have an angular resolution of 40″ and the noise in emission-free channels is ∼4 K at the spectral resolution of $1.5\,\mathrm{km\,s^{-1}}$. The covered velocities range from -113 to $+163\,\mathrm{km\,s^{-1}}$.

In Sect. 4.3.3 we correct for the optical depth that we measured against discrete continuum sources. We therefore select the THOR H I+continuum data that consist of VLA C-array configuration data only (THOR-only). THOR-only data have a higher angular resolution of ∼14″, making them suitable to study absorption against discrete continuum sources. Since this data set consists of interferometric observations only, large-scale H I emission is effectively filtered out. This large-scale filtering due to the interferometer can be exploited when only discrete continuum sources are of interest.

4.2.2 Gaussian decomposition

The THOR H I emission spectra show Gaussian-like structures at negative velocities, with usually three or fewer components superposed. We therefore use the fully automated Gaussian decomposition algorithm GaussPy+[3] (Riener et al., 2019) to study the kinematics of the filament.

GaussPy+ is a multicomponent Gaussian decomposition tool based on the earlier GaussPy algorithm (Lindner et al., 2015) and provides an improved fitting

[3] https://github.com/mriener/gausspyplus

routine and a fully automated means to decompose emission spectra using machine-learning algorithms. GaussPy+ automatically determines initial guesses for Gaussian fit components using derivative spectroscopy. To decompose the spectra, the spectra require smoothing to remove noise peaks while retaining real signal. The optimal smoothing parameters are found by employing a machine-learning algorithm that is trained on a subsection of the data set.

As Maggie is found at negative velocities and the H I emission toward the inner Galactic plane with $v_{\rm LSR} \gtrsim 0\,{\rm km\,s^{-1}}$ is ubiquitous in the spectra, it would not be sensible to fit the whole spectra with GaussPy+. Instead, to save computational resources and achieve a better decomposition performance, we masked all spectral channels at velocities $\leq -71\,{\rm km\,s^{-1}}$ and $\geq -23\,{\rm km\,s^{-1}}$ (see Fig. 4.2). Maggie has velocities around $-54\,{\rm km\,s^{-1}}$ (see Fig. 4.3 and Sect. 4.3.2) and we aim to disentangle components that might blend in with Maggie.

However, it is essential to reliably estimate the noise in the spectra to obtain good fit results. GaussPy+ comes with an automated noise estimation routine as a preparatory step for the decomposition. For this step, we supplied the full spectra to GaussPy+ as the masked spectra might not contain enough noise channels. The noise map derived in that way was then used for the decomposition of Maggie.

We ran the GaussPy+ training step with 500 randomly selected spectra from the H I data to find the optimal smoothing parameters for the fitting, as recommended in Riener et al. (2019). For H I observations, we would expect both narrow and broad line widths owing to the multiphase (i.e., WNM-CNM-UNM) nature of H I emission. To account for that, we chose not to refit broad or blended components, parameters that can easily be adjusted in the GaussPy+ routine.

After the initial fitting, the algorithm applies a two-phase spatial coherence check that can optimize the fit by refitting the components based on the fit results of neighboring pixels. Figure 4.2 shows example spectra along the Maggie filament marked in Fig. 4.3, and the final fit results from the GaussPy+ decomposition. Between two and three components were typically fit by GaussPy+ in the given velocity range. In every pixel spectrum, we selected the component with the lowest centroid velocity as the "Maggie" component (see Sect. 4.3.2) since no significant portion of H I emission is found at $v_{\rm LSR} \lesssim -60\,{\rm km\,s^{-1}}$. Due to blended components and strong emission at higher velocities, a selection based on amplitude of the components fails to recover the structure identified in Soler et al. (2020) as the Maggie filament.

4.3 Results

4.3.1 Location, distance, and morphology

Figure 4.1 shows the integrated H I emission covering the Galactic plane at longitudes and latitudes $37.5° < \ell < 47.0°$ and $|b| < 1.25°$, respectively. The filament Maggie subtends an area from $(\ell, b) = (40.4, -0.3)°$ to $(44.2, -0.9)°$ in Galactic longitude and latitude, and has a velocity around $-54\,{\rm km\,s^{-1}}$ with respect to the local standard of rest (LSR). In this region in position-position-velocity (p-p-v) space, the Galactic H I disk shows a warp toward higher latitudes (see e.g., Burton, 1988; Dickey & Lockman, 1990; Sparke, 1993; Dickey et al., 2009), which places Maggie

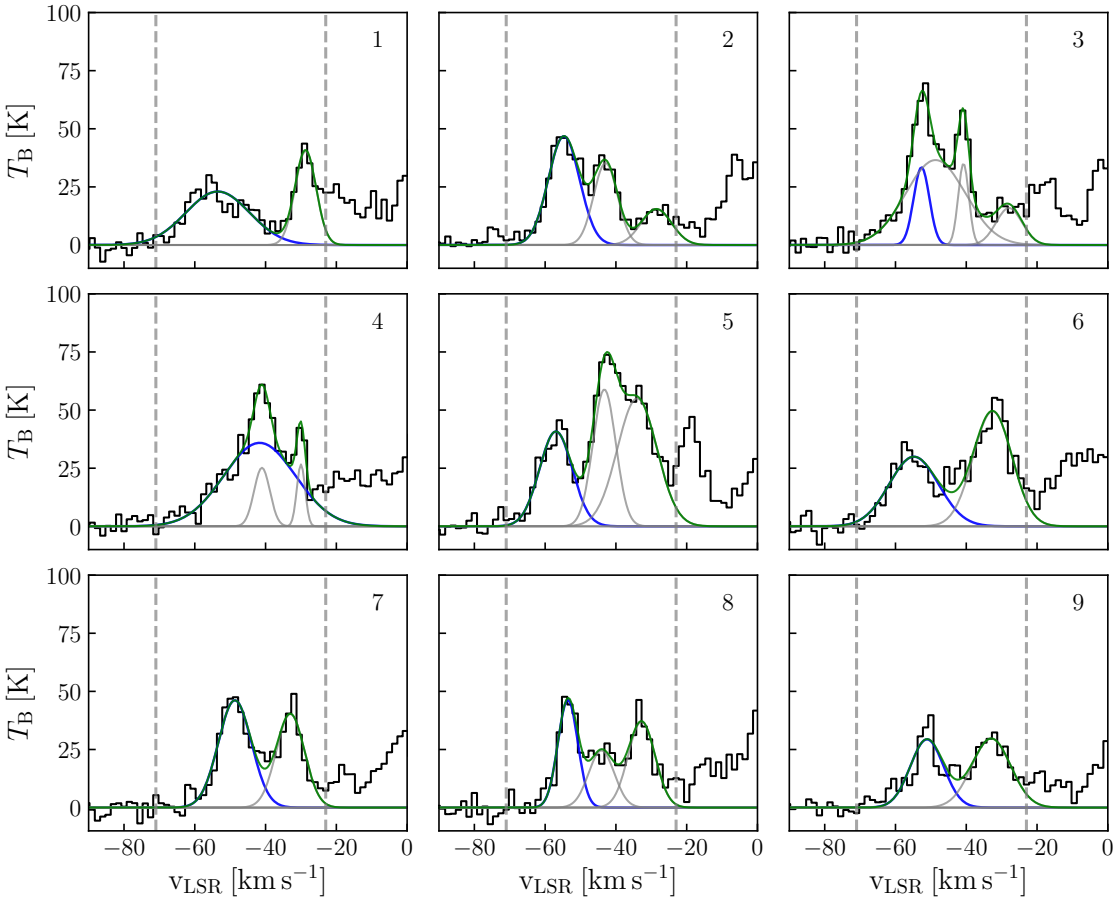


Figure 4.2: H I spectra and GaussPy+ decomposition results along the filament. The black curves show H I spectra corresponding to the positions marked in the middle panel of Fig. 4.3. The vertical dashed lines at -71 and $-23\,\mathrm{km\,s^{-1}}$ mark the velocity range taken into account for the GaussPy+ decomposition. The blue curve in each panel shows the first (i.e., lowest-centroid-velocity) component, which we attribute to Maggie. The gray curves show all additional components that were fit. The green curve represents the total spectrum fit in the considered velocity range.

at a location significantly displaced from the Galactic midplane. The H I midplane lies even beyond the coverage of the THOR survey at $b \geq +1.25°$. Therefore, the midplane and Maggie are separated by at least $\Delta b \geq 1.6°$ on the plane of the sky.

In the forthcoming analysis we assumed the Galactic rotation model by Reid et al. (2019) and used the Bayesian distance calculator of the BeSSeL survey (Reid et al., 2016) to translate the location in *p*-*p*-*v* space into a physical distance. Maggie's estimated distance away from us is $d = 17 \pm 1\,\mathrm{kpc}$ and its distance to the Galactic center is $R_{\mathrm{GC}} = 12 \pm 1\,\mathrm{kpc}$. This physical scale locates the filament $\geq 470\,\mathrm{pc}$ below the Galactic midplane, which is greater than the average H I scale height ($\sim$200 pc) at this Galactocentric radius (Kalberla & Kerp, 2009).

The Maggie filament discloses a hub-like feature in the east, on which smaller-scale filaments appear to converge, and a tail that thins out toward the west. The northwestern part shows a connection to the midplane, potentially feeding off the H I material located at higher latitudes. Most of the filament, however, appears to

be disconnected from the Galactic midplane material.

We define a backbone that runs through Maggie by selecting a spline based on visual inspection of the integrated emission (see top panel in Fig. 4.3). The length and width of Maggie are estimated using the Filament Characterization Package[4] (FilChaP; Suri et al., 2019). These properties are not heavily affected by a potential misplacement of Maggie's spine.

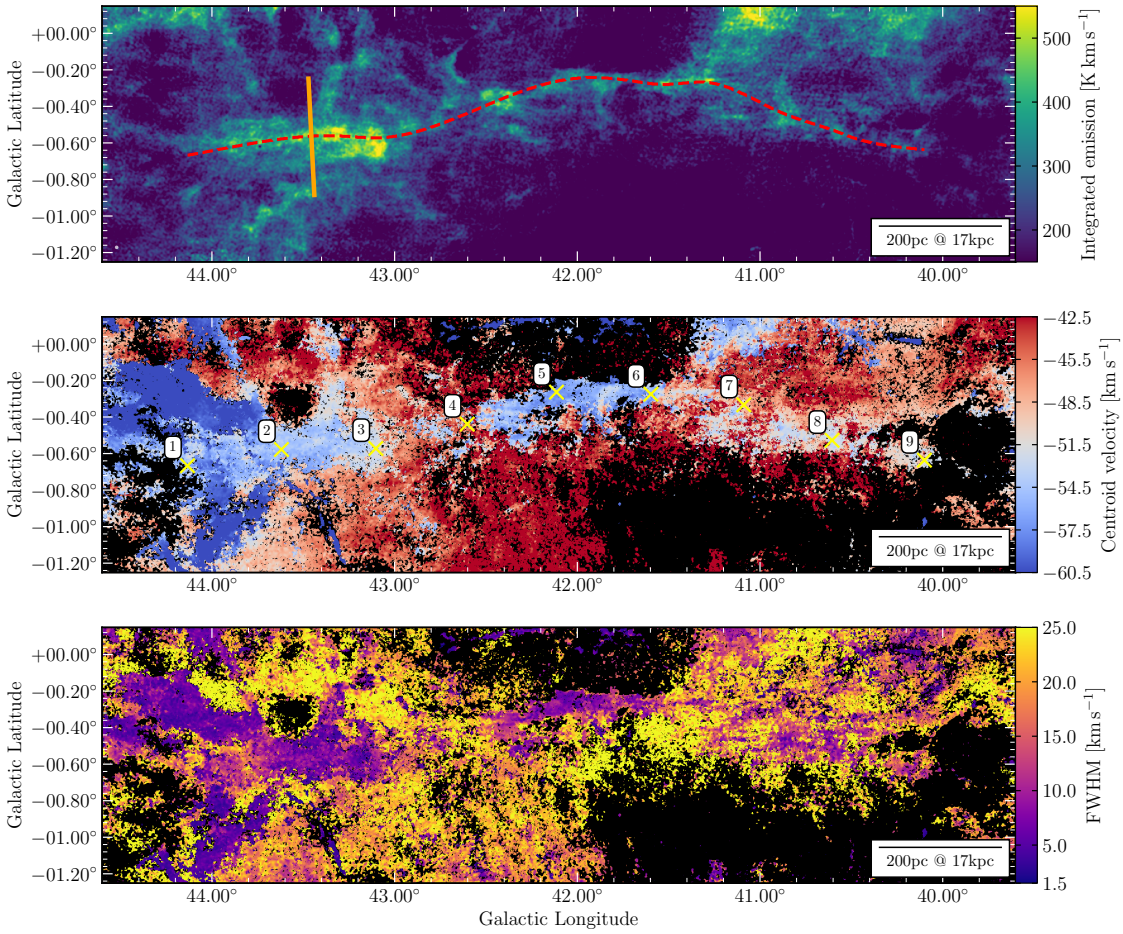


Figure 4.3: Kinematic overview of the Maggie filament. Top panel: Integrated H I emission in the velocity interval between -57.5 and $-48.5\,\mathrm{km\,s^{-1}}$. The red dashed curve marks the spine of the filament. The orange line perpendicular to the spine marks the cut of the average emission profile shown in black in the upper panel of Fig. 4.4. Middle panel: Centroid velocities of the first (or single) component derived from the GaussPy+ decomposition of the emission spectra. The fits are restricted to the velocity range $-71 \leq v_{\mathrm{LSR}} \leq -23\,\mathrm{km\,s^{-1}}$. The yellow crosses mark the positions of the spectra shown in Fig. 4.2. Bottom panel: Corresponding line widths in terms of the full width at half maximum (FWHM). Both maps are clipped at 5σ ($\sim$20 K).

The length of the filament as marked by the red spine in the top panel of Fig. 4.3 is $\ell_{\mathrm{tot}} = 1.2 \pm 0.1\,\mathrm{kpc}$ assuming a distance of 17 kpc. The uncertainty in length is dominated by the uncertainty in distance. The width of the filament was estimated using the integrated emission map shown in Fig. 4.3. We took perpendicular slices of 200 pc width with a step size of 10 pc ($\sim$3$\times$beam) along the filament. We then

[4] https://github.com/astrosuri/filchap

averaged the radial emission profiles over three neighboring slices and fit Gaussian and Plummer-like functions to the averaged emission profile. Plummer-like functions (as functions of cylindrical radius r) have the form

$$I(r) \propto \frac{\rho_c R_{\rm flat}}{\left[1 + (r/R_{\rm flat})^2\right]^{\frac{p-1}{2}}} , \tag{4.3.1}$$

where p is the power-law index, ρ_c is the density at the center of the filament, and $R_{\rm flat}$ is the characteristic radius of the flat inner portion of the density profile. More details about the fitting can be found in Suri et al. (2019).

Plummer-like functions are used to describe the radial column density (i.e., integrated emission in the optically thin limit) profile of a cylindrically shaped filament that has a flat density distribution up to $R_{\rm flat}$ and a power-law fall-off beyond (Arzoumanian et al., 2011). We used constant power-law indices $p = 2$ and 4 for the Plummer-like functions due to the degeneracy between ρ_c and $R_{\rm flat}$ (and therefore p) in finding the best fit solution (see Smith et al., 2014b). The power law index $p = 4$ describes the Ostriker (1964) model of an isothermal filament in hydrostatic equilibrium. The power-law index $p = 2$ is derived by Arzoumanian et al. (2011) as a free fit parameter from small-scale filamentary structures seen in *Herschel* data and might be attributed to nonisothermal or magnetized filaments (see Arzoumanian et al., 2011, and references therein).

The fit widths along the filament are shown in Fig. 4.4, with zero distance at the outermost point in the east. The widths derived from Gaussian fits are systematically larger than the widths estimated by Plummer-like functions. While the Plummer-like function with an index of $p = 2$ is able to reliably reproduce individual peaks (upper panel of Fig. 4.4), the fits on average underestimate the width of the emission profiles. Gaussian fits do not fully recover the amplitude of the profiles but give a more accurate representation of the wings. As the noise and blending of different emission features make it difficult to determine which function represents the emission profiles best, we consider the fits equally good. The reduced chi-squared distributions of the fits are similar and do not allow any preference of a fitting function. Three positions along the filament (vertical gray bars in Fig. 4.4) are discarded as the fits gave no reliable results, regardless of the fitting function. For simplicity, we report the average width taken over the full length of the filament. Taking into account all fitting methods, the average width is $w = 40 \pm 6$ pc, which is well resolved by the beam of the THOR observations corresponding to a spatial scale of ≈ 3.3 pc. The uncertainty is dominated by the variance of the measured widths along the filament. The filament width is only weakly affected by the velocity integration range. If we integrate the emission over the velocity range –57.5 to –42.5 km s^{-1}, the average filament width becomes 44 pc, which is well within the uncertainty of our derived width. However, due to blending emission components toward higher velocities, the derivation of the filament width becomes increasingly difficult with increasing velocity integration range. We therefore constrain the integration range to roughly the lower and upper quartile of the velocity distribution (see Sect. 4.3.2).

One of the characteristics that is usually used to define filaments is the projected aspect ratio (see e.g., Zucker et al., 2015, 2018). Taking the average width, Maggie

has an aspect ratio of (30:1). Although Maggie is identified through atomic line emission, as opposed to the selection of molecular filaments investigated by Zucker et al. (2018), Maggie roughly agrees with the overall aspect ratios of other large-scale filaments (Wang et al., 2015) and the identified Milky Way bones (Goodman et al., 2014; Zucker et al., 2015).

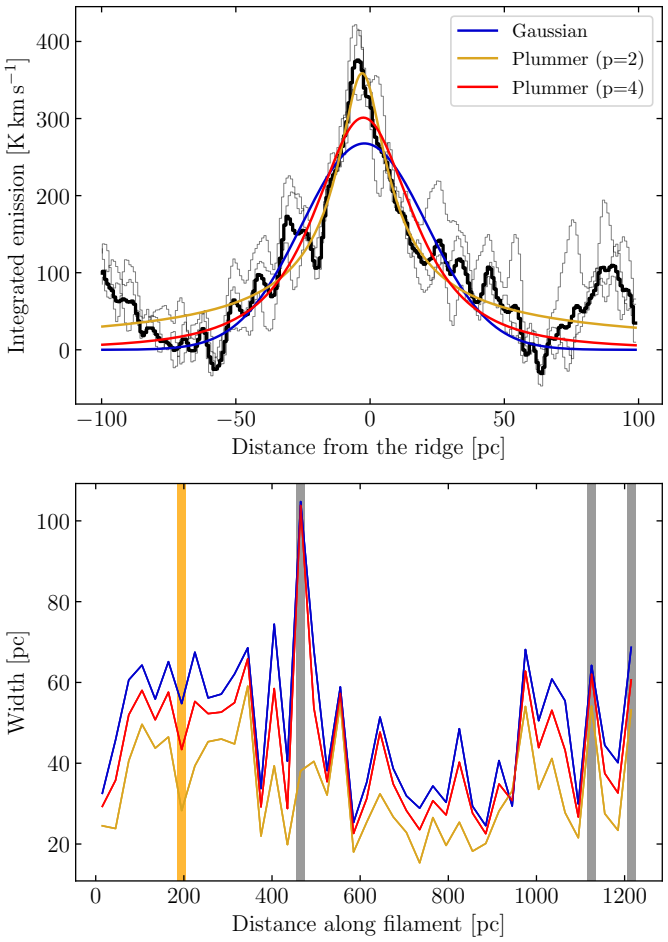


Figure 4.4: Width of the Maggie filament. Top panel: Example emission profile perpendicular to the filament. At each step along the filament, three neighboring emission profiles (gray) are taken to estimate an average profile (black). A fit baseline has been subtracted from each profile. The cut along which the average emission profile is measured is marked in the upper panel of Fig. 4.3. To derive the widths, we fit the average emission profiles with Gaussian (blue), and Plummer-like functions with index $p = 2$ (yellow) and $p = 4$ (red), respectively. Bottom panel: The fit widths of Maggie are plotted against the distance along the filament. The colors correspond to the fits shown in the upper panel. The gray bars show the widths that are discarded as these profiles could not be fit properly. The orange bar marks the position of the example profile shown in the upper panel.

4.3.2 Kinematics

In the following section, we focus in detail on the kinematic properties of Maggie. We show in the middle and bottom panel of Fig. 4.3 the maps of the fit centroid velocities and line widths of the first fit component (corresponding to the lowest-

velocity component of the GaussPy+ decomposition), to which we refer as the "Maggie component" in the following. Accordingly, the second fit component refers to the next component at higher centroid velocity. Overall, the peak velocities along the filament vary between ~ -60 and $\sim -45\,\mathrm{km\,s^{-1}}$ and exhibit an undulating pattern (see also Fig. 4.9). For the most part, the velocity map clearly reflects the spine of the filament, as components off the spine have significantly larger velocities but are picked up as the first component in the absence of Maggie. Given the large spatial scales, the velocities along the spine show a smooth gradient of less than $\pm 3\,\mathrm{km\,s^{-1}}\,(10\,\mathrm{pc})^{-1}$, a kinematic property that satisfies the bone criteria in Zucker et al. (2015). The line widths as well as centroid velocities shown in Fig. 4.3 indi-

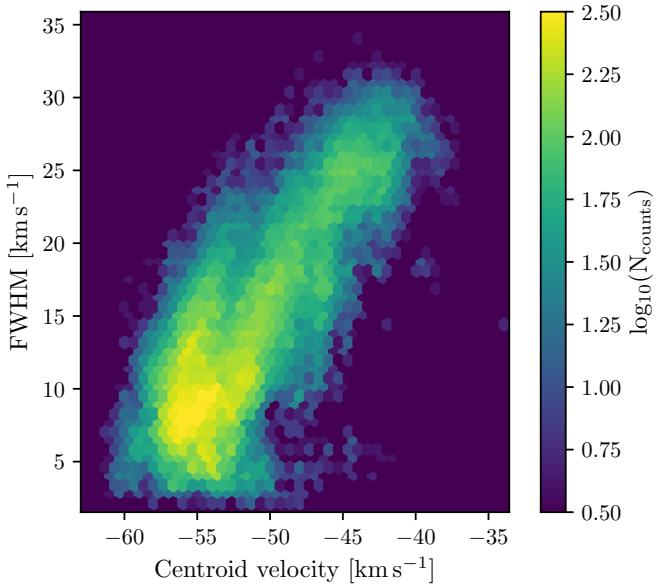


Figure 4.5: Line width–centroid velocity correlation. The scatter plot shows the line width (FWHM) as a function of the centroid velocity at all pixel positions along the 40 pc-wide spine of Maggie.

cate a stark contrast between the filament and the background. Furthermore, the kinematics in terms of the centroid velocity and line width along the filament spine are moderately correlated, where narrow line width corresponds to lower centroid velocity. Figure 4.5 presents the relationship between the centroid velocity and line width along the 40 pc-wide spine of the filament. The Pearson correlation coefficient between the velocity and line width distribution is 0.76, implying that under the assumption of a linear dependence ∼60% of the variance in the line width distribution can be explained by the scatter present in the centroid velocity distribution. We note that this does not necessarily imply a causal relation between the line width and velocity. Rather, this correlation is likely to be due to contamination from background gas at positions where the filament emission is less well defined and difficult to disentangle from blending components. Figure 4.6 shows the histogram of centroid velocities of Maggie, where all individual pixel positions along the 40 pc-wide spine of the filament are being sampled. The centroid velocities along the filament exhibit a strong peak at $-54\,\mathrm{km\,s^{-1}}$, with some skewness toward $-45\,\mathrm{km\,s^{-1}}$ that might be attributed to velocity blending of broad WNM components at higher ve-

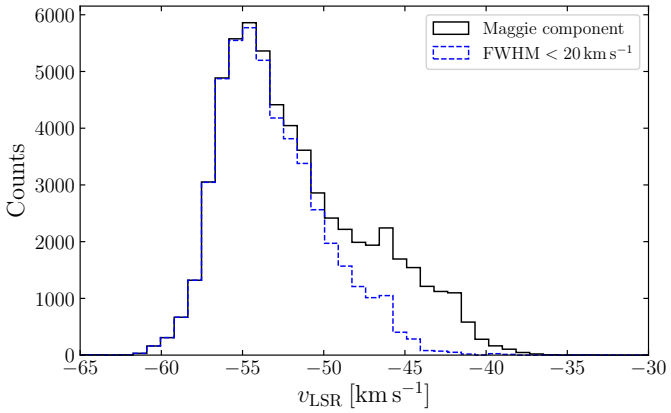


Figure 4.6: Histograms of fit peak velocities of Maggie. The black histogram shows the distribution of the centroid velocities at all pixel positions inferred from the GaussPy+ fits along the 40 pc-wide spine of the filament. The blue histogram shows the peak velocity distribution corresponding to components with line widths $\Delta v < 20\,\mathrm{km\,s^{-1}}$.

locities. Figure 4.7 offers a position-velocity view of Maggie and its location with

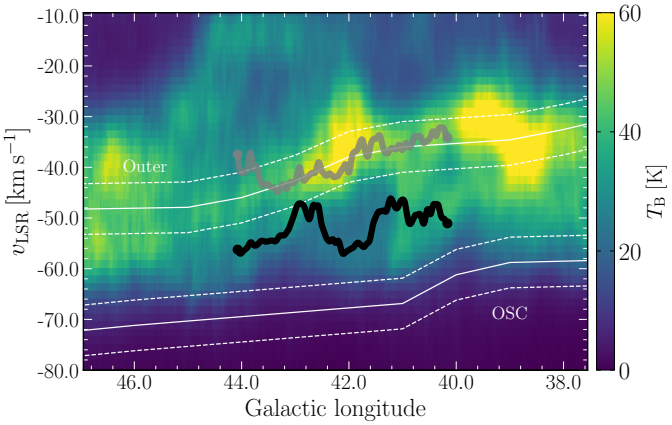


Figure 4.7: Position-velocity diagram of H I. The colored background shows the H I emission averaged over $-1.25° < b < +1.25°$. The black and gray curve indicate the intensity-weighted average velocity of Maggie and the second fit component at each longitude as given by the GaussPy+ decomposition, respectively. The white line segments give the predicted spiral arm centroids of the Outer arm and Outer-Scutum-Centaurus (OSC) arm along with their uncertainties (dashed lines) in velocity (Reid et al., 2019).

respect to the average H I emission. We also overplot the predicted locations of the spiral arms as given by Reid et al. (2019). The bulk of H I material and the second fit component of our spectral decomposition trace out the spine of the Outer arm remarkably well while Maggie exhibits an offset of 5–15 km s^{-1}. Given the distance from the midplane, neither spatially nor kinematically can Maggie be associated with any spiral arm structure. We note, however, that different spiral arm models can vary by 10 km s^{-1} or more in position-velocity space, a concern raised by Zucker et al. (2015) when classifying bones of the Milky Way.

The histogram of line widths in Fig. 4.8 shows a broad distribution, with line widths between 4 and 30 km s^{-1}. The histogram has a peak around 10 km s^{-1} and

a shoulder that extends to 25 km s^{-1}. We additionally show in Figs. 4.6 and 4.8 the histograms corresponding to velocity components with a line width $\Delta v < 20$ km s^{-1}, discarding the shoulder in the line width distribution. The median values of the total centroid velocity distribution is -52.7 km s^{-1}, and only taking into account line widths $\Delta v < 20$ km s^{-1}, the median shifts to -53.8 km s^{-1}.

To study the kinematic structure with respect to the position along the filament we show in Fig. 4.9 the average velocity of each 40 pc-wide slice weighted by the amplitude at each pixel position. The uncertainties are estimated by the standard deviation of each slice. The wave-like velocity structure of Maggie is now more clearly visible. The Maggie velocities beyond > -48.5 km s^{-1} around 400 pc distance (corresponding to the region around spectrum 4 in Fig. 4.2) are difficult to clearly separate from the second fit component. We note that the mean velocity and particularly line width have to be treated with some caution as the velocity blending makes it difficult to distinctively identify Maggie in this part of the region. We revisit the velocity structure in Sect. 4.4.2 to investigate signatures of characteristic spatial scales.

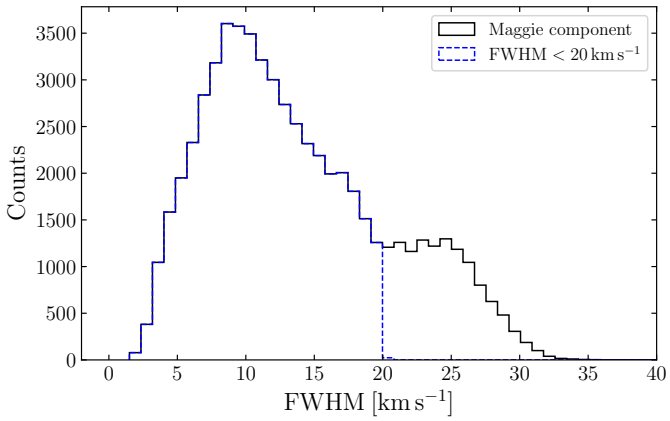


Figure 4.8: Histograms of fit line widths of Maggie. The black histogram shows the distribution of the line widths at all pixel positions in terms of the FWHM inferred from the GaussPy+ fits along the 40 pc-wide spine of the filament. The blue histogram shows the line width distribution corresponding to components with line widths $\Delta v < 20$ km s^{-1}.

4.3.3 Column density and mass

Measuring H I absorption against strong continuum sources allows us a direct derivation of the optical depth. We therefore use the high angular resolution H I+continuum data, that consist of VLA C-array observations only, to filter out large-scale emission and measure H I absorption against discrete continuum sources. We measured H I+continuum spectra against seven continuum sources taken from Wang et al. (2018) that have brightness temperatures ≥ 200 K for a sufficient signal-to-noise ratio. The sources are located both within and outside the area of Maggie (i.e., within and outside the contour at 4.2 $M_\odot$ pc^{-2}, see later in this section). Since the synthesized beam of the native C-configuration data is $\sim$14″, we extracted an average H I+continuum spectrum from a 16″ × 16″ (4 × 4 pixels) area. To the first order, we

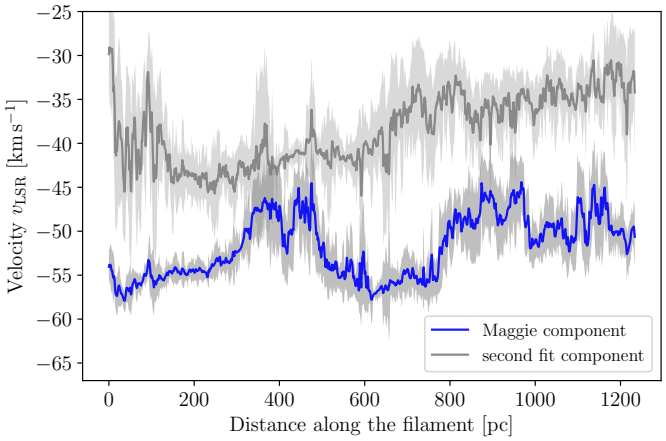


Figure 4.9: Position-velocity plot along the filament. This plot shows the centroid velocity over distance along the filament marked by the red spine in the upper panel of Fig. 4.3. The centroid velocity is an intensity-weighted average over the 40 pc-wide slice perpendicular to every filament position. The blue curve represents the designated Maggie component, that is the first fit component of the GaussPy+ results. The gray curve shows the second fit component for comparison. The gray shaded area around both curves indicates the uncertainty estimated by the standard deviation over each slice.

can then estimate the H I optical depth by (see Bihr et al., 2015)

$$\tau_{\mathrm{simplified}} = -\ln\left(\frac{T_{\mathrm{on}}}{T_{\mathrm{cont}}}\right) , \tag{4.3.2}$$

where T_{on} is the H I brightness temperature against the continuum source and T_{cont} is the brightness of the continuum source. Using the optical depth, we compute the H I spin temperature with

$$T_{\mathrm{s}} = \frac{T_{\mathrm{off}}}{1 - e^{-\tau}} . \tag{4.3.3}$$

We measured the H I brightness at an offset position T_{off} using the combined THOR H I data (see Sect. 4.2.1). We therefore selected an annulus around each source with inner and outer radii of 60″ and 120″, respectively, to measure an averaged T_{off}.

All spectra measured against Galactic H II regions do not exhibit any absorption features in the velocity range of Maggie (Tab. 4.1). This gives additional support to the argument that the Maggie filament is in fact on the far side of the Galaxy as Galactic H II regions are located in the foreground and can therefore not be detected in absorption. Furthermore, we find that no absorption is evident in the spectra of both Galactic and extragalactic continuum sources offset from the line of sight of Maggie.

The extragalactic continuum sources G43.738-0.620 and G43.921-0.479[5] are the only sources along Maggie against which we detect H I in absorption. We show in Fig. 4.10 an overview of the optical depth and spin temperature measurement toward G43.738-0.620.

[5]The optical depth measurement against G43.921-0.479 is close to the detection limit and will not be discussed further here. However, the mean optical depth $\tau_{\mathrm{mean}} = 0.30$ and velocity-integrated optical depth $\int \tau(v)\mathrm{d}v = 2.2\,\mathrm{km\,s^{-1}}$ are consistent with the other source G43.738-0.620.

Table 4.1: Continuum sources.

Source ID	Glon [°]	Glat [°]	$T_{\rm cont}$ [K]	Type	L.o.s.	Absorption
G43.921-0.479	43.92	-0.48	304	extragal.	✓	✓
G43.890-0.783	43.89	-0.78	481	H II	✗	✗
G43.738-0.620	43.74	-0.62	503	extragal.	✓	✓
G43.177-0.519	43.18	-0.52	296	H II	✓	✗
G42.434-0.260	42.43	-0.26	277	H II	✗	✗
G42.028-0.605	42.03	-0.60	519	extragal.	✗	✗
G41.513-0.141	41.51	-0.14	205	H II	✗	✗

Notes. Column 1 gives the source name as listed in the continuum catalog by Wang et al. (2018). Columns 2 and 3 are the Galactic longitude and latitude, respectively. Column 4 denotes the average brightness temperature of the continuum source over an area of $16'' \times 16''$. Column 5 describes the physical nature of the continuum source. Column 6 marks if the continuum source is located along the line of sight (L.o.s), that is within the contour at $4.2\,M_\odot\,{\rm pc}^{-2}$ (see Sect. 4.3.3). Column 7 indicates if absorption above 3σ is detected in the velocity range of Maggie.

The top panel illustrates the averaged H I emission profile. In pressure equilibrium, the two stable phases of atomic hydrogen (CNM and WNM) both contribute to the emission. It is difficult to disentangle the properties of either phase from the emission alone, so the information from the absorption profile is critical to learn about the nature of the atomic gas. As the optical depth is proportional to $T_{\rm s}^{-1}$, we assume that any absorption measured toward the continuum is due to the CNM.

For simplicity we fit both the emission and absorption profile in this specific case (Fig. 4.10) with a single Gaussian component. The line width of the emission profile is $14.1\,{\rm km\,s^{-1}}$ and therefore significantly wider than the absorption line ($8.8\,{\rm km\,s^{-1}}$), suggesting that there is contribution to the emission from a broader component that is not evident in the absorption spectrum. However, the line width contributions are not able to pin down the column density fractions of either the CNM or WNM as the (kinetic) gas temperatures are unknown and the line widths can be significantly broadened by nonthermal effects. The total line width, however, can therefore set an upper limit on the kinetic temperature (Heiles & Troland, 2003). We note that the nonthermal contribution to the CNM component, due to effects such as turbulent motion, should be dominant as the thermal line width of the CNM even with the loose constraint $T_{\rm k} < 500\,{\rm K}$ (e.g., Heiles, 2001; Heiles & Troland, 2003) would account for a line width $< 4.8\,{\rm km\,s^{-1}}$. The optical depth above the 3σ limit, that we ascribe to the CNM, is plotted in the third panel of Fig. 4.10. The mean and velocity-integrated optical depth over the velocity range of Maggie is $\tau_{\rm mean} = 0.33$ and $\int \tau(v){\rm d}v = 2.5\,{\rm km\,s^{-1}}$, respectively. We note that the reported values in Wang et al. (2020a) are inferred from integrating over the whole velocity range of the THOR survey and are thus significantly larger. We use the optical depth and brightness temperature from the emission profile to compute the spin temperature (Eq. (4.3.3)) in the bottom panel of Fig. 4.10. The mean spin temperature is $\langle T_{\rm s} \rangle = 140\,{\rm K}$. Owing to the mixing of warm and cold H I along the line of

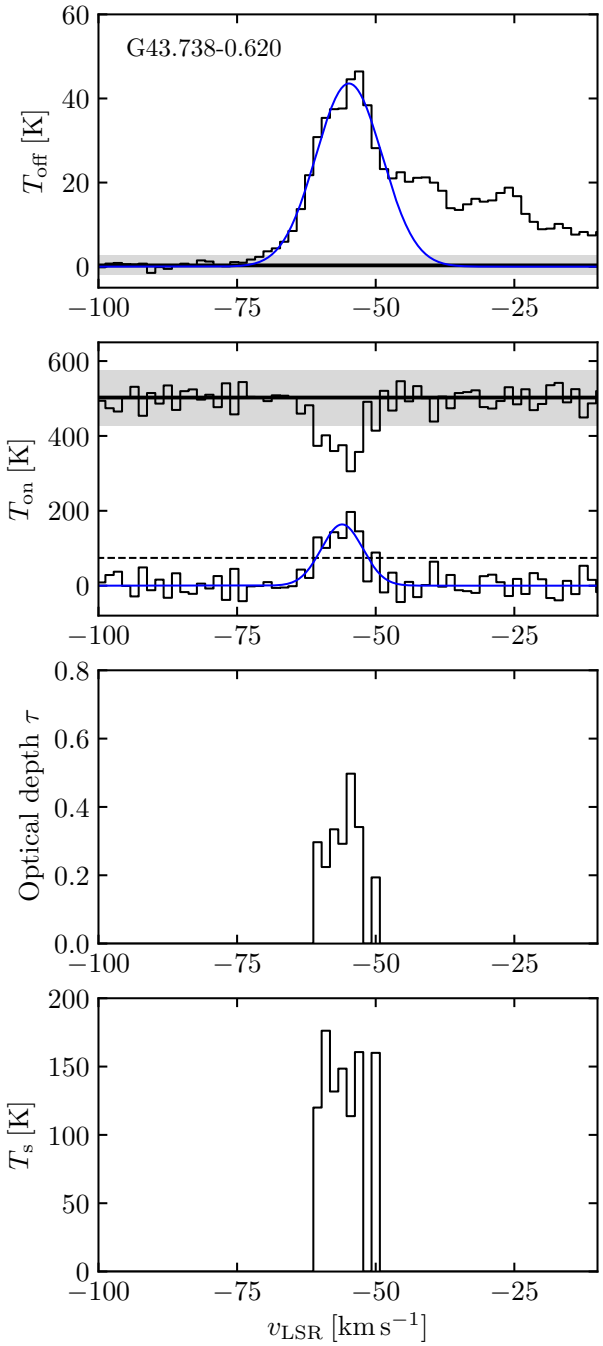


Figure 4.10: H I emission and absorption spectrum of the extragalactic continuum source G43.738-0.620. The top panel shows the emission spectrum and is measured in an annulus around the source with inner and outer radii of 60″ and 120″, respectively (corresponding to $3-6$ beams of the H I data). The second panel presents the absorption spectrum toward the point source. The lower spectrum indicates the inverted absorption spectrum along with the 3σ noise (dashed line). In the first two panels, the gray shaded area and blue curve indicate the 3σ noise and a single-component Gaussian fit, respectively. The third panel shows the optical depth computed using Eq. (4.3.2). In the bottom panel we present the spin temperature, which is computed using Eq. (4.3.3).

sight, this spin temperature should not be interpreted as a gas kinetic temperature. Rather, it is a density-weighted harmonic mean of the kinetic temperature under the assumption that $T_\mathrm{s} = T_\mathrm{k}$ (see Dickey et al., 2000, 2003, 2009). If we assume that the warm H I does not contribute significantly to the optical depth, that the correction for H I self-absorption is negligible and that the CNM has a roughly constant temperature, then we can derive the CNM fraction via

$$f_\mathrm{CNM} \equiv \frac{N_\mathrm{CNM}}{N_\mathrm{CNM} + N_\mathrm{WNM}} \simeq \frac{T_\mathrm{CNM}}{\langle T_\mathrm{s} \rangle} \,, \tag{4.3.4}$$

where N_CNM and N_WNM are the column density of the CNM and WNM, respectively, and T_CNM is the CNM spin temperature. A representative value for the CNM spin temperature should lie within $40 < T_\mathrm{CNM} < 100\,\mathrm{K}$ (Heiles & Troland, 2003; Strasser & Taylor, 2004; Dickey et al., 2009), which is also in good agreement with the theoretical models by Wolfire et al. (2003) at the Galactocentric distance $R_\mathrm{GC} = 11\,\mathrm{kpc}$. This conservative estimate gives a CNM fraction in the range $30\% < f_\mathrm{CNM} < 70\%$. Applying this finding to the whole cloud, Maggie should consequently be composed of a significant fraction of CNM.

We estimate the column density of the atomic hydrogen using (e.g., Wilson et al., 2013)

$$N_\mathrm{H\,I} = 1.8224 \times 10^{18} \int T_\mathrm{s}(v)\, \tau(v)\, \mathrm{d}v \,, \tag{4.3.5}$$

where $N_\mathrm{H\,I}$ is the total column density of H I in units of cm^{-2} as a function of the H I spin temperature $T_\mathrm{s}(v)$ and optical depth $\tau(v)$ integrated over the velocity v. The spin temperature is $T_\mathrm{s}(v) = T_\mathrm{B}/(1 - e^{-\tau(v)})$, where T_B is the brightness temperature of the H I emission. The optical depth correction of the column density scales as $\tau/[1-\exp(-\tau)]$, resulting in a correction factor of 1.17 for $\tau_\mathrm{mean} = 0.33$. We integrated the column density from -57.5 to $-48.5\,\mathrm{km\,s^{-1}}$, taking into account the predominant velocities of Maggie. The column density then has values in the range $3 - 10 \times 10^{20}\,\mathrm{cm}^{-2}$. We used the observed THOR-H I emission data instead of the individual Maggie component of the decomposition to compute the column density although there might be additional contributions from components blending in at a similar velocity. We show in Appendix A.1 that the integrated Maggie component is not spatially coherent in amplitude and shows patches of lower amplitude, likely owing to the existence of blended components that are difficult for GaussPy+ to disentangle. To estimate a maximum fraction of unrelated blended emission contributing to the column density, we compare the two column density maps. We find that the column density (and ultimately mass) inferred from the single Maggie component (App. A.1) is smaller by approximately 30%.

We note that choosing a single fit component as the "filament" can be problematic as the multiphase nature of H I emission may have multiple components at similar velocities contributing to Maggie. Here, we define Maggie by the centroid velocity of a single component and do not impose any restrictions on the H I phase of Maggie, which is reasonable in order to study the velocity structure. Yet, by choosing a single component we might include only one H I phase. GaussPy+ does not explicitly check for spatial coherence in amplitude or line width but for coherence in number of components and their centroid velocity. As we argue above, in a CNM-WNM

mixture of the gas it is then difficult to characterize Maggie based on the centroid velocity of a single component alone. One way to fit and identify multiple phases in H I emission spectra, is to introduce spatial regularization terms to the loss function of the fitting that are able to cluster different phases (here: CNM and WNM) even if close in velocity (Marchal et al., 2019). This approach, however, is computationally expensive and will not be followed here.

The picture of a well-defined CNM phase layer within a WNM envelope assuming thermal pressure equilibrium might be overly simplified. Instead, we would expect a clumpy CNM embedded in a 'sea' of WNM, implying that the CNM has a lower volume and area filling factor than the WNM and a strong density contrast (Kalberla & Kerp, 2009). The patchy structure seen in Fig. A.1 could be a reflection of the clumpy structure of the CNM. Lower intensity in small-scale structures could also hint at CNM that has already converted to molecular gas (see Sect. 4.4.4). If we want to infer the global multiphase column density of Maggie, however, it is beneficial to examine the properties of the filament by means of the integrated emission rather than using a single GaussPy+ fit component, in order to avoid missing mass.

Figure 4.11 shows the mass surface density map of Maggie derived from the observed emission data. The mass surface density reaches values of more than $\sim$8.0 $M_\odot\,\mathrm{pc}^{-2}$ ($= 1.0 \times 10^{21}\,\mathrm{cm}^{-2}$, $A_V \sim 1\,\mathrm{mag}$)[6] toward the eastern hub and its distribution is constrained to within one order of magnitude (see Sect. 4.4.3). Wolfire et al. (2003) predict an average midplane thermal pressure at the Galactocentric radius $R_\mathrm{GC} = 12\,\mathrm{kpc}$ of $P_\mathrm{th,ave}/k \sim 1580\,\mathrm{K\,cm}^{-3}$. We can estimate the expected length scale of Maggie along the line of sight by $l_\mathrm{los} \sim N_\mathrm{H\,I}\,T_\mathrm{k}/(P_\mathrm{th,ave}/k)$, where we set the kinetic temperature to be close to the density-weighted mean spin temperature $T_\mathrm{k} \approx \langle T_\mathrm{s} \rangle$. Since Maggie is significantly below the midplane, the pressure is likely to be smaller than $1580\,\mathrm{K\,cm}^{-3}$. Thus, we assume a canonical pressure of $P_\mathrm{th,ave}/k \sim 1000\,\mathrm{K\,cm}^{-3}$. At the mean column density of $\langle N_\mathrm{H\,I} \rangle = 4.8 \times 10^{20}\,\mathrm{cm}^{-2}$ (see also Sect. 4.4.3), the line-of-sight length is estimated to be $\sim$22 pc, which is roughly consistent with the on-sky width of 40 pc given the uncertainty in temperature and pressure.

Because of the diffuse nature of the WNM that is significantly contributing to the observed H I emission it is difficult to define a clear edge of the filament. We therefore estimate two different masses: a "diffuse mass" in which the dense filament is embedded (mass above 5σ), and a "dense filament mass" which resides within an approximately closed contour at the 9σ level ($= 4.2\,M_\odot\,\mathrm{pc}^{-2}$) that roughly corresponds to the filament width obtained in Sect. 4.3.1.

The total diffuse mass including the dense filament is $1.7 \times 10^6\,M_\odot$. As mentioned above, due to its diffuse nature the mass derivation of the H I has a high uncertainty as it is dependent on the selection of filament regions taken into account. Varying the contour level by $\pm 1\sigma$ to estimate the uncertainty, the dense filament mass is $7.2^{+2.5}_{-1.9} \times 10^5\,M_\odot$. The atomic mass of Maggie therefore is comparable to the highest masses of large-scale filaments identified in the Milky Way (Ragan et al., 2014; Abreu-Vicente et al., 2016; Wang et al., 2015, 2016; Zucker et al., 2015, 2018),

[6] We relate the total hydrogen column density $N_\mathrm{H} = N_\mathrm{HI} + 2\,N_{\mathrm{H}_2}$ to the visual extinction using $N_\mathrm{H} = 2.2 \times 10^{21}\,\mathrm{cm}^{-2}\,\mathrm{mag}^{-1} \times A_V$ (Güver & Özel, 2009). We take into account the molecular column density using the upper limit estimated in Sect. 4.4.1.

providing a large atomic gas reservoir to potentially host molecular cloud formation.

If we take the column density we use to define the edge of the "dense" filament and assume that the width of the filament is the same along the line of sight as in the plane of the sky, we obtain a density of $4.2/40\,M_\odot\,\mathrm{pc}^{-3} \sim 7 \times 10^{-24}\,\mathrm{g\,cm}^{-3}$, which equates to an H nucleus number density of roughly $4\,\mathrm{cm}^{-3}$. This is intermediate between the average densities we expect for the WNM and CNM at this Galactocentric radius (Wolfire et al., 2003), which is best explained by the filament being a mix of both components.

However, given a mean density of $\bar{n} = 4\,\mathrm{cm}^{-3}$, we point out that either the CNM mass fraction f_{CNM} must be higher or that a substantial fraction of the WNM must be in an unstable phase to be in agreement with the theoretical model given in Wolfire et al. (2003). This follows from the mass conservation:

$$\bar{n} = f_{\mathrm{V,CNM}}\, n_{\mathrm{CNM}} + (1 - f_{\mathrm{V,CNM}})\, \gamma\, n_{\mathrm{CNM}}\,, \tag{4.3.6}$$

where $f_{\mathrm{V,CNM}}$ is the volume filling fraction of the CNM, n_{CNM} is the CNM number density, and $\gamma = n_{\mathrm{WNM}}/n_{\mathrm{CNM}} = T_{\mathrm{CNM}}/T_{\mathrm{WNM}} \sim 0.01$ is the ratio of the CNM and WNM temperature in pressure equilibrium. The CNM mass fraction $f_{\mathrm{CNM}} = f_{\mathrm{V,CNM}}\, n_{\mathrm{CNM}}/\bar{n}$ can then be expressed as

$$f_{\mathrm{CNM}} = \frac{1}{1-\gamma}\left(1 - \gamma\,\frac{n_{\mathrm{CNM}}}{\bar{n}}\right)\,. \tag{4.3.7}$$

Assuming the Wolfire et al. (2003) CNM-WNM model, with $\bar{n} = 4\,\mathrm{cm}^{-3}$ the CNM mass fraction is higher than our assumed values. For example, even for a CNM density $n_{\mathrm{CNM}} = 40\,\mathrm{cm}^{-3}$ (the maximum CNM density at $R_{\mathrm{GC}} = 11\,\mathrm{kpc}$; Wolfire et al., 2003), the CNM fraction is still $f_{\mathrm{CNM}} = 0.91$. The mean density can be brought into agreement with the previously derived CNM fraction of ∼0.5 if the temperature ratio becomes $\gamma \sim 0.05$, which corresponds to a mean temperature in the unstable regime ∼1000–2000 K, compared to the higher temperatures of the classical WNM around ∼6000 K. Observational works have found the fraction of H I in the unstable phase to be at least 20–50% (Heiles & Troland, 2003; Roy et al., 2013b; Murray et al., 2015, 2018; Nguyen et al., 2019), possibly even higher (Koley & Roy, 2019).

4.4 Discussion

4.4.1 Molecular gas tracers, continuum, and dust

We investigate the molecular gas fraction of Maggie and possible signatures of molecular cloud formation by examining the molecular lines ^{12}CO, ^{13}CO, and C^{18}O ($J = 1$–0) using the Milky Way Imaging Scroll Painting (MWISP; Su et al., 2019) survey. The rms noise of the data is 0.5 K for ^{12}CO at the velocity resolution of $0.16\,\mathrm{km\,s}^{-1}$ and 0.3 K for ^{13}CO and C^{18}O at $0.17\,\mathrm{km\,s}^{-1}$ resolution. The spatial resolution in all three tracers is ∼50″.

In all three tracers, we detect no molecular gas emission in the velocity range between −57.5 and $-48.5\,\mathrm{km\,s}^{-1}$. We show in Appendix A.2 a map of the integrated

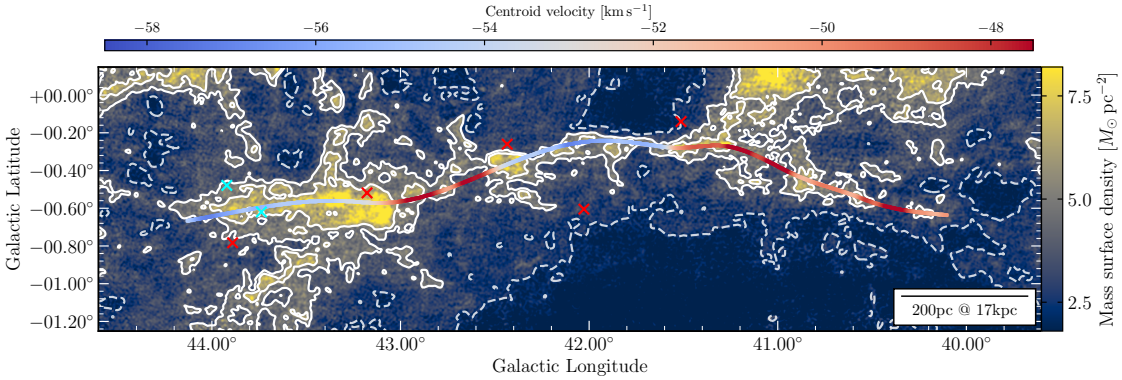


Figure 4.11: Mass surface density map. The map shows the H I mass surface density integrated over the velocity interval $-57.5 < v_{\rm LSR} < -48.5\,{\rm km\,s^{-1}}$. The dashed contour gives the 5σ level at $2.3\,M_\odot\,{\rm pc^{-2}}$ by which the diffuse mass is calculated. The white contours show the mass surface densities at the levels 4.2 and $5.6\,M_\odot\,{\rm pc^{-2}}$, respectively. The spine of Maggie is color-coded by the centroid velocity shown in Fig. 4.9. The crosses mark the positions of the continuum sources listed in Tab. 4.1. The cyan crosses indicate the positions of absorption detections.

^{12}CO emission as an example. To estimate the molecular hydrogen content, we used the CO-H_2 conversion factor $X = 1.8 \times 10^{20}\,{\rm cm^{-2}\,K^{-1}\,km^{-1}\,s}$ (Dame et al., 2001). We determined an upper limit at 5σ for the molecular surface density of $\Sigma_{\rm H_2} \sim 13\,M_\odot\,{\rm pc^{-2}}$. Since the sensitivity limit is close to the observed H I column densities, we cannot entirely dismiss the presence of CO-bright H_2 gas.

Indeed, Nakanishi et al. (2020) identify molecular gas clumps toward Maggie within small-scale H I clouds that they infer from a dendrogram analysis of the VGPS H I data (Stil et al., 2006). The average size of the H I clouds is 5 pc. They derive the molecular mass by summing the brightness temperature of the FUGIN ^{12}CO data (Umemoto et al., 2017) within each H I cloud, including voxels with a brightness temperature lower than the noise. In doing that, diffuse emission can be recovered in the same manner as stacking emission spectra. The total H_2 mass of all 353 identified cloudlets that are within the dense filament contour at $4.2\,M_\odot\,{\rm pc^{-2}}$ (area filling factor ∼0.21) is $\sim 6\times10^4\,M_\odot$, only a small fraction of the H I mass (0.08), with the cloudlets' average surface density being $\sim 2.5\,M_\odot\,{\rm pc^{-2}}$ that is consistent with our upper limit. The mean molecular mass fraction of the H I clouds is ∼0.2 of the total gas mass (Nakanishi et al., 2020). Naturally, the H I "leaves" of the dendrogram analysis sample the highest-density parcels of the medium, hence a portion of the total molecular mass outside the H I clouds might be missed.

In addition, CO might not always be a good tracer of molecular hydrogen as "CO-dark" H_2 could account for a significant fraction of the total H_2 (Pineda et al., 2008; Goodman et al., 2009; Tang et al., 2016), particularly at low column densities (Goldsmith et al., 2008; Planck Collaboration et al., 2011). The CO-dark fraction has furthermore been observed to increase with Galactocentric distance (Pineda et al., 2013). Simulations of molecular gas in Milky Way type galaxies indicate that the CO-dark fraction is indeed higher in regions of lower surface density (Smith et al., 2014a). CO observations at visual extinctions $A_{\rm V} \leq 3.5$ could underestimate the molecular hydrogen content by a significant fraction (Glover & Mac Low, 2011).

As OH may be a better molecular gas tracer than CO in atomic-molecular tran-

sition regions at low column densities (e.g., Tang et al., 2017), we inspected the strongest OH line at 1667 MHz in the THOR OH data (see Rugel et al., 2018). The THOR OH data consist of VLA C-array observations only, so large-scale emission cannot be recovered due to the missing flux on short *uv* spacings. However, we can derive an OH abundance from OH absorption measurements against continuum sources. At the velocity range of Maggie, no absorption is found against any of the continuum sources listed in Tab. 4.1. We note that depending on the peak flux of the continuum source and opacity of the OH line, the detection limit could be above the expected OH column density. The presence of molecular gas traced by OH can therefore not be ruled out. Follow-up OH observations at higher sensitivity will address this in a future study.

We investigated the possibility of Maggie being a foreground structure by examining the extinction of stars up to a distance of 5 kpc using the early installment of the third *Gaia* data release (Gaia Collaboration et al., 2021). The K_s-band extinctions are obtained using the Rayleigh-Jeans Colour Excess (RJCE, Majewski et al., 2011) method which combines near- and mid- infrared data from the Two Micron All-Sky Survey (2MASS; Skrutskie et al., 2006) and the Wide-Field Infrared Survey Explorer (WISE; Wright et al., 2010), respectively. The *Gaia* Archive[7] provides the crossmatch of *Gaia* sources with both the 2MASS and WISE catalogs.

In the presence of an obscuring medium, for example an atomic cloud, high extinction in the densest parts would prevent the detection of stars in the *Gaia* sample, yet extinction toward the outer portions of the cloud can still be detected in order to trace its silhouette. Figure A.3 presents the obtained K_s-band extinctions toward Maggie for sources up to a distance of 5 kpc from us. If Maggie were located within 5 kpc from the Sun, higher extinction stars would represent its shape in the ℓ-b plane. However, since no signature of Maggie emerges from the extinction map, this indicates that Maggie is not a foreground cloud with large deviation from the Galactic rotation velocity.

We furthermore detect neither extinction features in the Spitzer Galactic Legacy Infrared Mid-Plane Survey Extraordinaire (GLIMPSE; Churchwell et al., 2009) that might hint at the presence of IRDCs nor any excess infrared emission or signatures of stellar activity as observed in all bands of GLIMPSE and the Improved Reprocessing of the IRAS Survey (IRIS; Miville-Deschênes & Lagache, 2005). There are two possible reasons for not detecting IRDC features: 1) Maggie has not formed cold, dense molecular regions on a large scale such that there exist no IRDCs within Maggie or 2) Maggie is too distant to be observed as an IRDC due to the lack of emission background.

We have also examined the high-sensitivity maps of *Planck* at 353, 545, and 857 GHz emission (Planck Collaboration et al., 2014) and the higher angular resolution *Herschel* Hi-GAL data (Marsh et al., 2017) to further search for signatures of Maggie. We do not detect Maggie in any of the continuum bands while noting that we are most likely dominated by foreground emission in the Galactic plane.

After taking a variety of surveys into consideration, Maggie appears to show no signs of stellar activity as it is mostly atomic, while molecular gas has formed only in high-density clumps on the smallest spatial scales (Nakanishi et al., 2020).

[7] https://gea.esac.esa.int/archive/.

4.4.2 Kinematic signatures

Mach number distribution

In the following, we adopt a stable two-phase CNM-WNM model to derive the turbulent Mach number distributions. We estimate the three-dimensional scale-dependent Mach number of the filament assuming isotropic turbulence with $\mathcal{M} = \sqrt{3}\,\sigma_{\rm turb}/c_s$, where $\sigma_{\rm turb}$ and c_s are the turbulent one-dimensional velocity dispersion and sound speed, respectively. The turbulent line width is calculated by subtracting the thermal line width contribution from the observed line width as

$$\sigma_{\rm turb} = \sqrt{\sigma^2_{\rm obs} - \sigma^2_{\rm th}}\,, \tag{4.4.1}$$

where $\sigma_{\rm obs}$, and $\sigma_{\rm th}$ are the observed, and thermal velocity dispersion, respectively. Since we expect the observed line width to be a combination of both CNM and WNM, and do not know the turbulent contribution to the observed line width of either H I phase, we compute a single Mach number for the two-phase medium using a weighted mean of the thermal line width and sound speed. The thermal velocity dispersion can be decomposed into a CNM and WNM component as

$$\sigma^2_{\rm th} = f_{\rm CNM}\,\sigma^2_{\rm th,CNM} + (1 - f_{\rm CNM})\,\sigma^2_{\rm th,WNM}\,, \tag{4.4.2}$$

where $\sigma_{\rm th,CNM}$ and $\sigma_{\rm th,WNM}$ refer to the thermal velocity dispersion of the CNM and WNM, respectively. The combined velocity dispersion and sound speed can then be computed using the weighted mean kinetic temperature

$$\langle T_{\rm k} \rangle = f_{\rm CNM}\,T_{\rm k,CNM} + (1 - f_{\rm CNM})\,T_{\rm k,WNM}\,, \tag{4.4.3}$$

where $T_{\rm k,CNM}$ and $T_{\rm k,WNM}$ are the kinetic temperatures of the CNM and WNM, respectively.

We assume that the kinetic temperature of the CNM is close to the spin temperature and set $T_{\rm s,CNM} = T_{\rm k,CNM}$. In pressure equilibrium, Wolfire et al. (2003) predict a range of WNM kinetic temperatures between 4100–8800 K. As the densities of the WNM are low, the gas can usually not be thermalized by collisional excitation only, such that $T_{\rm s,WNM} < T_{\rm k,WNM}$. Through excitation by resonant Lyα photon scattering, sufficient Lyα radiation can allow $T_{\rm s,WNM} \rightarrow T_{\rm k,WNM}$. The WNM spin temperature $T_{\rm s}$ has been observed to be as high as 7200 K (Murray et al., 2014). We therefore assume a mean WNM kinetic temperature of $T_{\rm k,WNM} = 6000$ K.

Since the thermal line width and sound speed scale as $T_{\rm k}^{1/2}$, the variation is moderate and does not change the Mach number significantly. We calculate three different Mach number distributions for $f_{\rm CNM} = 30,\,50,\,70\%$, corresponding to $T_{\rm k,CNM} = 40,\,70,\,100$ K.

Figure 4.12 shows the Mach number distributions of Maggie under different assumptions of the CNM fraction. The median Mach number is 1.5, 1.7, and 2.0 for the CNM fractions 30, 50, and 70%, respectively, which is generally in contrast to molecular filaments being driven by highly supersonic turbulence (e.g., Elmegreen & Scalo, 2004). However, since we use a single mean kinetic temperature in each case, the Mach number distributions also reflect a slight bimodality that is already

evident in the line width distribution (Fig. 4.8). While each Mach number distribution represents a different mixture of CNM-WNM gas, we note that the bump toward higher Mach numbers reflected in each distribution is due to the increased line widths associated with higher $v_{\rm LSR}$ velocities, likely contaminated by blending background gas (see Sect. 4.3.2).

Numerical studies show that the condensation process toward a stable CNM phase depends on the properties of turbulence in the WNM (Seifried et al., 2011; Saury et al., 2014; Bellomi et al., 2020). These studies suggest that highly supersonic turbulence that is driven by compressive modes regulates the formation of a stable CNM phase as turbulent compression and decompression of the gas will be much more rapid than the cooling timescale. Once the gas has moved toward an unstable phase, that is the cooling timescale $t_{\rm cool}$ is shorter than the dynamical time $t_{\rm dyn} = \ell/c_s$, the gas can cool down fast enough to reach a stable branch of the CNM before being decompressed again (Hennebelle & Pérault, 1999). On the other hand, a too low degree of turbulence cannot produce sufficient density fluctuations to sustain the formation of a thermally stable CNM. The WNM is mostly found in a subsonic and transonic regime (e.g., Marchal & Miville-Deschênes, 2020), therefore providing the dynamical conditions for thermal instability to grow efficiently to form a bistable H I phase.

Absorption studies show that a substantial fraction of H I in fact lies in a thermally unstable or transient phase (Heiles, 2001; Heiles & Troland, 2003; Roy et al., 2013a,b). Shifting the mean temperature of the WNM to an unstable regime around $\sim$1500 K leads to Mach numbers in the range $2-4$.

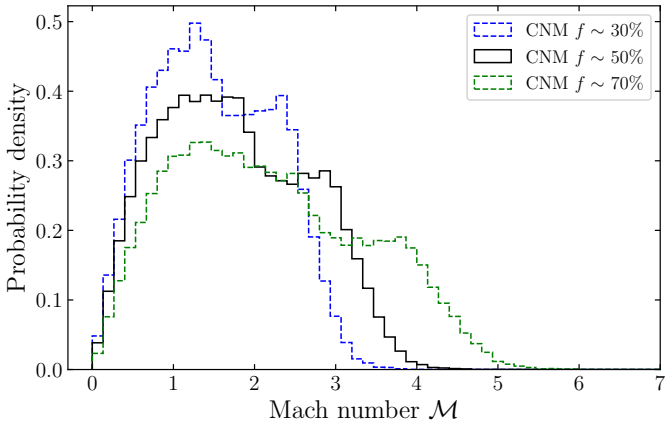


Figure 4.12: Mach number distribution of Maggie. The blue, black, and green histograms show the Mach number distributions assuming a CNM fraction of 30, 50, and 70%, respectively. Each distribution is computed using a single thermal line width and sound speed inferred from a weighted mean kinetic temperature.

Structure function

In order to investigate the velocity and column density structure of Maggie in a statistical sense, we compute the 1-D structure function (see Henshaw et al., 2020)

of the data using

$$S_p(\ell) = \langle \delta x(r,\ell)^p \rangle = \langle |x(r) - x(r+\ell)|^p \rangle \,, \tag{4.4.4}$$

where $\delta x(r,\ell)$ is the absolute difference in the quantity x measured between two locations r and $r+\ell$ separated by ℓ. The brackets indicate the average of the quantity over all locations. The order p of the structure function can be directly related to a physical quantity. We use the second order structure function, which in the case of the velocity structure is proportional to the kinetic energy of the line-of-sight component of the flow. We show in Fig. 4.13 the second order structure function derived from the weighted centroid velocity (Fig. 4.9), after having subtracted a 2-D plane fit to the velocity field in order to better examine the wave-like oscillation, and the column density along the 40 pc-wide spine of Maggie. A significant local minimum is evident at the spatial scale of $\ell_{2,vel} \approx 500$ pc. Since the structure function is evaluated for pairs of positions that are separated by the lag ℓ, the maximum spatial scale at which the structure function can be fully sampled is given by half the total length of the filament. Therefore, the structure function is only computed over spatial scales $\ell \lesssim 600$ pc. Consequently, the full velocity 'dip' cannot be recovered and we estimate the uncertainty of the scale by the difference between $\ell_{\rm max}$ and $\ell_{2,vel}$, giving $\ell_{2,vel} = 502 \pm 113$ pc. This result confirms the wave-like pattern seen in Figs. 4.3 and 4.9.

The column density structure of Maggie does not show any characteristic length scale at which the structure function has a local minimum but roughly represents a power-law scaling, indicative of the scale-invariant behavior of a gas structure generated by turbulence (Elmegreen & Scalo, 2004).

The phenomenological Kolmogorov relation for incompressible turbulence predicts a dissipationless cascade of energy in the inertial subrange, such that the kinetic energy transfer rate $\mathcal{E} \approx \delta v^3/\ell$ stays constant (Kolmogorov, 1941). Due to the nature of the turbulent cascade, the velocity dispersion then increases with scale as $\delta v \propto \ell^{q_{\rm theo}}$, where $q_{\rm theo} = 1/3$, which we can relate to the power-law scaling of the second order structure function $\langle \delta v^2 \rangle \propto \ell^{\gamma}$, where $\gamma = 2q_{\rm theo}$ (see e.g., Elmegreen & Scalo, 2004; Henshaw et al., 2020). We therefore fit the velocity structure function with a power-law function of the form $S_2(\ell) \propto \ell^{\gamma}$. We set the lower limit of the fitting range to three beam sizes (10 pc), below which no reliable structure functions can be recovered, and the upper limit at the approximate scale above which the structure function begins to turn over (100 pc). For Maggie we find $q = 0.36 \pm 0.02$ in the case where we subtracted the 2-D plane fit, which is in good agreement with the expected power-law scaling $q_{\rm theo} = 1/3$ in the subsonic and transonic regime (see also Wolfire et al., 2003; Elmegreen & Scalo, 2004; McKee & Ostriker, 2007; Saury et al., 2014; Marchal & Miville-Deschênes, 2020). While the dip in the velocity structure function is less pronounced if we do not subtract a 2-D plane fit, the power-law scaling remains the same with $q = 0.37 \pm 0.02$. These values are in agreement with the observational scaling exponents originally found by Larson (1979, 1981), who investigated the velocity dispersion–size relation in a sample of molecular clouds and clumps using different tracers. He utilized the radial velocity dispersion mainly in terms of the line width of optically thin lines, such that the relation between line width and individual cloud size can be interpreted as a correlation between velocity

fluctuation and scale size (see also Scalo, 1984). Larson (1979, 1981) found a velocity dispersion–size relation following a power law with an index $q_L = 0.38$, which is, although slightly larger, similar to the Kolmogorov relation. Systematic studies of more homogeneous samples of molecular clouds traced in CO found velocity scaling exponents in a range between 0.24 and 0.79 (see Table 2 in Izquierdo et al. (2021), which summarizes more than 40 years of observational work).

Izquierdo et al. (2021) find in their simulations that the scaling exponents lean toward lower values when molecular clouds are dominated by the influence of the Galactic potential rather than the effects of clustered SN feedback and self-gravity (see also Chira et al., 2019). This suggests that low velocity scaling exponents are associated with diffuse regions of low star forming activity (e.g., Federrath et al., 2010), which also favors the development of long filamentary structures (Smith et al., 2020).

Maggie's velocity scaling is in better agreement with the lower end of the values found for molecular clouds (see above). This regime around 1/3 is indicative of Kolmogorov turbulence, and following Izquierdo et al. (2021), this turbulence may be caused by the Galactic potential rather than being driven by the effects of self-gravity or stellar feedback.

The existence of density enhancements that are spaced at a characteristic scale within a highly filamentary cloud would be in broad agreement with theoretical predictions of the fragmentation of a self-gravitating fluid cylinder due to the "sausage" instability (e.g., Chandrasekhar & Fermi, 1953). As we find no characteristic scale in the column density structure, we argue that the effects of self-gravity have not yet become important and it is difficult to correlate the observed velocity signature with possible mechanisms driving the column density structure formation.

We caution, as pointed out in Sect. 4.3.3, that the column density is a product of both the CNM and WNM and is inferred from the integration over velocity. This could lead to a structure that smooths over an underlying higher-density CNM structure due to the diffuse nature of the WNM. The column density structure can furthermore not be described by single power-law alone. This could be due to the filament being a mix of both CNM and WNM and that the two phases have different turbulence scalings. With the WNM being more widespread, the structure function as probed by H I emission could be dominated by the warm diffuse gas on large scales while the CNM dominates the power-law on small scales (see e.g., Choudhuri & Roy, 2019). The integrated emission might therefore not be the best estimator of the spatial distribution of the column density.

4.4.3 Column density PDF

To elaborate on the dominant physical processes acting within the cloud, we employed the commonly used probability density function (PDF) of the column density. The shape of column or volume density PDFs are used as a means to describe the underlying physical mechanisms of a cloud (e.g., Federrath & Klessen, 2013; Kainulainen et al., 2014; Schneider et al., 2015). A log-normal shape of a PDF is considered a signature of turbulent motion dominating a cloud's structure (see e.g., Klessen, 2000). The magnitude of turbulence can furthermore be reflected in the width of a

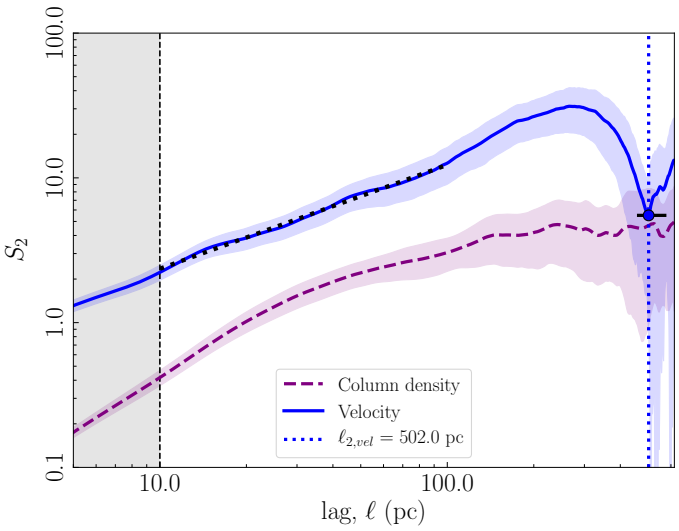


Figure 4.13: Second order structure function. This plot shows the second order structure function of column density (dashed purple) and velocity (blue) computed from the respective average quantity along the Maggie filament. A 2-D plane fit to the velocity field has been subtracted prior to computing the velocity structure function. The vertical blue dotted line indicates the characteristic spatial scale of the velocity. The black dotted line is a power-law fit to the velocity structure function. The gray shaded area indicates the range of scales below three beams for which the recovered structure function is unreliable.

log-normal PDF and can then be associated with the Mach number (e.g., Padoan et al., 1997; Passot & Vázquez-Semadeni, 1998; Kritsuk et al., 2007; Federrath et al., 2008; Konstandin et al., 2012; Molina et al., 2012), while noting that the turbulence driving scale and CNM-WNM mass ratio also affect the width of the PDF (Bialy et al., 2017b).

Molecular clouds that are subject to the increasing effect of self-gravity develop high-density regions, producing a power-law tail in their PDF (e.g., Girichidis et al., 2014; Burkhart et al., 2017). Many star-forming molecular clouds have been confirmed to show such power-law tails (Kainulainen et al., 2009; Schneider et al., 2013, 2016). Even before the effects of gravity become dominant, gravitationally unbound clumps can exhibit power-law tails due to pressure confinement from the surrounding medium (Kainulainen et al., 2011).

We show in Fig. 4.14 the normalized column density PDF (N-PDF) of Maggie taken above the 5σ limit at $\sim 3\times10^{20}\,\mathrm{cm}^{-2}$ ($= 2.3\,M_\odot\,\mathrm{pc}^{-2}$). The mean of the N-PDF is $4.8\times10^{20}\,\mathrm{cm}^{-2}$. The distribution varies within one order of magnitude and can be well described by a log-normal function of width $\sigma = 0.28$. We caution that the shape and width of the resulting N-PDFs are sensitive to the region taken into account (Schneider et al., 2015; Lombardi et al., 2015; Ossenkopf-Okada et al., 2016; Körtgen et al., 2019). If we only consider column densities within an approximately last closed contour at $\sim 4\times10^{20}\,\mathrm{cm}^{-2}$, the N-PDF shows a narrower log-normal shape with a width smaller by $\Delta\sigma = 0.05$.

Figure 4.14 presents a comparison between the N-PDF of Maggie and the optical depth corrected N-PDFs derived from H I emission and self-absorption toward the giant molecular filament GMF20.0-17.9 (Syed et al., 2020). The width of Maggie's

N-PDF comes close to the value found for the CNM gas traced by HISA and is larger than that of the atomic gas traced by H I emission toward the midplane (see also Wang et al., 2020c). However, the width of the CNM distribution is likely to be higher as the HISA N-PDF is limited by the column density range at which H I self-absorption can be detected.

Since we use the same data set and follow the same methodology for deriving the N-PDF as Wang et al. (2020c) and Syed et al. (2020), systematic differences should be minimized. Despite that, we note a slight cutoff in the PDF at $N_{\rm H\,I} \sim 1 \times 10^{21}\,{\rm cm}^{-2}$. As we use a uniform optical depth correction for the whole cloud (Sect. 4.3.3), the optical depth could be underestimated in regions of high column density. If we assume a column density $\sim 10^{21}\,{\rm cm}^{-2}$ and a FWHM line width $\sim 10\,{\rm km\,s}^{-1}$, the peak optical depth becomes $\tau \sim 1$. The cutoff in the PDF could therefore be due to the cloud becoming optically thick and the width could then be underestimated.

Narrow log-normal shaped N-PDFs are commonly observed in the diffuse H I emission toward well-known molecular clouds (Burkhart et al., 2015; Imara & Burkhart, 2016; Rebolledo et al., 2017). On the other hand, the HISA N-PDFs that trace the CNM show similar or somewhat broader distributions, indicative of the clumpy structure and higher degree of turbulence. The CNM-WNM mixture of Maggie, that is spatially more clearly defined than the material usually traced by H I emission, agrees with our previous findings and suggests that Maggie might be at the verge of becoming a dense filament forming out of the diffuse atomic ISM.

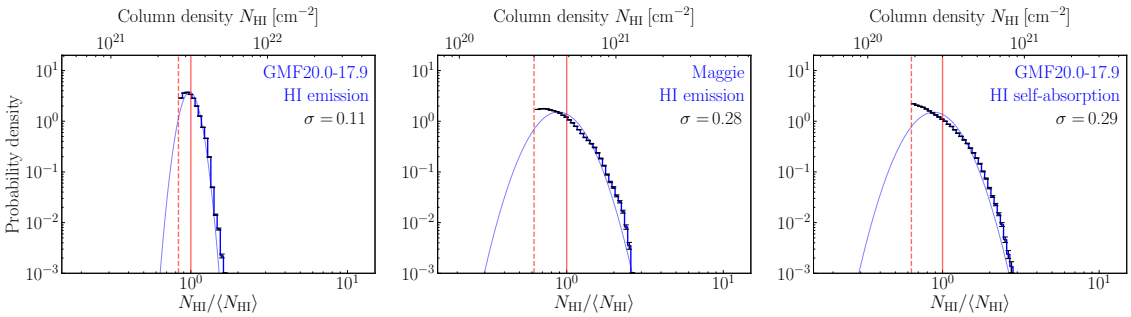


Figure 4.14: Comparison of N-PDFs. The blue histograms show the column density distributions of Maggie (middle panel) in comparison with the N-PDFs taken from a different GMF region analyzed in Syed et al. (2020) that trace the midplane H I emission (left) and CNM by means of H I self-absorption (right). The blue curve in each panel indicates a log-normal fit to the distribution. The red vertical dashed and solid lines mark the column density thresholds taken into account and the mean column density, respectively.

4.4.4 Molecular gas formation and timescale

We address Maggie's current evolutionary stage and evaluate the surface density threshold of H I required to efficiently form molecular hydrogen, based on the analytical models by Sternberg et al. (2014), and Bialy & Sternberg (2016). These authors derived a steady-state H I-to-H_2 transition model assuming a slab irradiated isotropically on two sides by FUV flux. The expected H I surface density at the

transition then is

$$\Sigma_{\mathrm{H\,I}} = \frac{6.71}{\tilde{\sigma}_g} \ln\left(\frac{\alpha G}{3.2} + 1\right) M_\odot\,\mathrm{pc}^{-2} \tag{4.4.5}$$

(see Eq. (1) in Bialy et al., 2017a), where $\tilde{\sigma}_g$ is the dust absorption cross section per hydrogen nucleus in the Lyman-Werner (LW) dissociation band (11.2–13.6 eV) relative to the standard solar neighborhood value, α is the ratio of the unshielded H_2 dissociation rate to formation rate, and G is the average H_2 self-shielding factor in dusty clouds. The product αG is the H_2 formation-to-destruction rate ratio, accounting for H_2 shielding by H_2 line absorption and dust, and may be written as

$$\alpha G = 2.0\, I_{\mathrm{UV}} \left(\frac{30\,\mathrm{cm}^{-3}}{n}\right) \left(\frac{9.9}{1 + 8.9\,\tilde{\sigma}_g}\right)^{0.37} \tag{4.4.6}$$

(Sternberg et al., 2014; Bialy & Sternberg, 2016), where I_{UV} is the free-space (unshielded) interstellar radiation intensity relative to the Draine (1978) field, and n is the number density. The UV intensity exponentially decreases with Galactocentric distance (Wolfire et al., 2003). For $R_{\mathrm{GC}} = 12\,\mathrm{kpc}$, Wolfire et al. (2003) suggest $I_{\mathrm{UV}} \sim 0.39$. We note that this approximation is derived for the midplane and should be roughly constant up to the scale height of OB stars and could be significantly lower at larger distances from the plane. Since Maggie's distance from the plane is an order of magnitude greater than the OB scale height in the solar neighborhood (Reed, 2000), the UV intensity should be viewed as an upper limit with a relative uncertainty of at least 40% (see also Elias et al., 2006)

The dust absorption cross-section $\tilde{\sigma}_g$ can be estimated as the inverse of the gas-to-dust ratio at the position of Maggie relative to the value in the local ISM if we assume that the dust grain properties are the same in the outer Galaxy. Giannetti et al. (2017) derive a gas-to-dust ratio γ as a function of galactocentric distance and metallicity based on $C^{18}O$ observations. Their reported gas-to-dust gradient predicts $\gamma \sim 305$ for Maggie and ∼141 as a local value for $R_{\mathrm{GC}} = 8.15\,\mathrm{kpc}$ (Reid et al., 2019). The relative dust absorption cross-section then becomes $\tilde{\sigma}_g = 0.46$. With the estimated number density of $\bar{n} = 4\,\mathrm{cm}^{-3}$ (see Sect. 4.3.3), the H I threshold for molecular hydrogen to form efficiently is expected at $\Sigma_{\mathrm{H\,I}} \approx 17.5\,M_\odot\,\mathrm{pc}^{-2}$. This exceeds the observed column densities by a factor of more than 2.

However, we note that the model by Sternberg et al. (2014) is highly idealized and the derived gas-to-dust relation in Giannetti et al. (2017) has large systematic uncertainties due to variations in the CO abundance and poorly constrained dust properties. In addition with the uncertainties in distance and UV intensity, the H I threshold is uncertain by at least ∼50%.

Furthermore, observational effects might produce a bias against the measurement of a clumpy multiphase medium. Small pockets of high-density CNM clumps could be hidden below the resolution of the telescope beam (≈3.3 pc). Even if we assume that those unresolved CNM clumps have a small volume filling fraction $f_{\mathrm{V,CNM}}$, a substantial fraction of the mass $f_{\mathrm{CNM}} = (n_{\mathrm{CNM}} \cdot f_{\mathrm{V,CNM}})/\bar{n}$ could be allocated in the CNM phase although the observed mean density is low compared to that of the CNM. High-density CNM clumps are much more efficient at shielding and allow the onset of H_2 formation at lower surface densities. As we argue in Sect. 4.3.3, the

CNM fraction should be considerable and could allow H_2 formation on a scale that is inaccessible to our observations.

In fact, Nakanishi et al. (2020) report a total molecular gas mass of $\sim 6 \times 10^4\, M_\odot$ that resides within small CNM cloudlets with an average size of 5 pc (see Sect. 4.4.1). These cloudlets have a mean H I density of $\sim 50\,\mathrm{cm}^{-3}$. Inserting this density into Eqs. (4.4.6) and (4.4.5) gives a molecule formation threshold at $\Sigma_{\mathrm{H\,I}} \approx 2.5\, M_\odot\,\mathrm{pc}^{-2}$, which is below their observed H I surface density of $\sim 7\, M_\odot\,\mathrm{pc}^{-2}$. Hence, we would expect molecule formation to take place within these high-density CNM cloudlets. Clearly, employing segmentation methods such as the dendrogram algorithm rather than using integrated quantities is important when aiming at identifying small-scale CNM structures in a clumpy multiphase medium.

We estimate the timescale of H_2 formation by the time after which the balance between H_2 destruction and formation rate has reached a steady state. If we consider the dissociation rate to be negligible compared to the formation rate, the H_2 formation timescale is given by

$$t_{\mathrm{H_2}} \sim \frac{1}{2\,R\,n} \tag{4.4.7}$$

(see Eq. (23) in Bialy & Sternberg, 2016), where $R = 3 \times 10^{-17}\,\tilde{\sigma}_g\,\mathrm{cm}^3\,\mathrm{s}^{-1}$ is the H_2 formation rate coefficient and n is the total hydrogen number density that is essentially given by our estimated H I number density in the absence of H_2. For a mean density of $\bar{n} = 4\,\mathrm{cm}^{-3}$ the formation timescale is $\sim$300 Myr, about an order of magnitude larger than those predicted by models of nearby molecular clouds (Goldsmith et al., 2007) and numerical simulations (Clark et al., 2012). As the formation of molecular hydrogen is most effective on the surface of dust grains (Gould & Salpeter, 1963), the formation timescale at a given density is primarily limited by the dust-to-gas ratio $\propto \tilde{\sigma}_g$, rendering the formation of molecular hydrogen less likely in the outer Galaxy. However, the effects of small-scale density fluctuations, either due to a clumpy CNM-WNM medium or due to transient structures produced by supersonic turbulence (Glover & Mac Low, 2007), can shorten the timescale. For CNM clumps at several times the observed density, the timescale could decrease significantly. Again using the H I density reported by Nakanishi et al. (2020), the molecule formation timescale within the CNM-rich structures becomes $\sim$25 Myr.

In conclusion it is difficult to tightly constrain the formation of molecules within Maggie based on integrated H I and CO observations alone. Because of observational effects small-scale density fluctuations might be missed that allow the formation of molecular hydrogen even at the current stage. On the smallest scales within CNM-rich structures, the presence of molecular hydrogen traced by CO emission has been confirmed by Nakanishi et al. (2020). In addition, time delays between the formation of H_2 and CO could lead to Maggie not being identified as a molecular cloud, even though the formation of molecular hydrogen is already taking place on larger scale (Clark et al., 2012).

4.4.5 Comparison with literature

Maggie's origin remains unclear. While there are morphological and kinematic similarities, it is difficult to argue in favor of Maggie being an immediate counterpart

to large-scale molecular filaments known in the literature.

The aspect ratio of Maggie is in rough agreement with large-scale filaments and Milky Way bones (Wang et al., 2015; Zucker et al., 2015), even though we point out that the determination of filament widths is in many cases subject to the tracer being used to identify the filament. Furthermore, Maggie has a smooth velocity gradient of less than $\pm 3\,\mathrm{km\,s^{-1}}\,(10\,\mathrm{pc})^{-1}$ (Sect. 4.3.2) and therefore satisfies the Milky Way bone criteria applied in Zucker et al. (2015).

Maggie does not display any close association with a spiral arm structure although the predicted location of the spiral arms is highly model-dependent (see e.g., Zucker et al., 2015). It is observed that all types of large-scale filaments identified in the literature commonly show a range of orientations with respect to spiral arm features in position-velocity (ℓ-v) space, both tracing out the spine of spiral arms as well as following ℓ-v tracks that are inclined to spiral arms (Zucker et al., 2018). The latter could potentially trace a spurious feature trailing off a spiral arm.

The orientation of Maggie's major axis is remarkably well aligned with the Galactic midplane, a fact that the majority of large-scale filaments have in common (Zucker et al., 2018). However, Maggie is observed at a large distance of $\approx$500 pc from the Galactic midplane while most filaments are found in spatial proximity to it. The fact that Maggie shows both a spatial (position-position) and kinematic (position-velocity) displacement from any spiral structure suggests a different formation path than those of known molecular filaments in the disk.

The Radcliffe wave (Alves et al., 2020) that consists of an association of molecular cloud complexes in the solar neighborhood has an aspect ratio of 20:1 and exhibits an undulating structure (in p-p-p space) below and above the midplane. At its peak, it reaches an offset of $\sim$200 pc from the plane. It is argued that the accretion of a tidally stretched gas cloud settling into the Galactic disk could in principle mimic the shape of the observed structure, but seems unlikely given that it would require the Radcliffe wave to be synchronized with the Galactic rotation over large scales.

Infalling gas, possibly of extragalactic origin, would interact with the gas in the disk and become compressed or even shocked along the edge closest to the disk if the relative velocities are large enough. In the case of Maggie, we then consider the chances of infalling gas becoming plane-parallel to be small. Additionally, in this scenario it seems unlikely for the gas to keep a coherent velocity structure.

Conversely, gas that is being expelled out of the plane as a result of SN feedback tends to have a randomized orientation with respect to the Galactic disk (Smith et al., 2020). Observations of the preferentially vertical H I filaments seen in the Galactic plane can be associated with enhanced SN feedback (Soler et al., 2020) and indicate that Maggie is unlikely to have originated in SN feedback events.

As Maggie could be linked to the global spiral arm structure in ℓ-v space, we speculate that the filament originated in the midplane and might have been brought to large distance by vertical oscillations. Levine et al. (2006) decomposed the structure of the Galactic H I disk into Fourier modes. They found that the Galactic warp (e.g., Burton, 1988) can be described sufficiently well by the first three Fourier modes. Locally, significant high-frequency modes at large Galactocentric distances are also evident within the disk. This suggests that deviations from the midplane generally do occur but do not pervade the Galactic disk as they could be triggered

by local perturbations.

Wobbles and disk corrugations could be induced by merger events or satellite galaxies plunging through the Galactic plane as suggested by simulations of the stellar disk of galaxies (e.g., Edelsohn & Elmegreen, 1997; Widrow et al., 2014; Chequers & Widrow, 2017). This hypothesis is supported by stellar populations that have formed in the stellar disk but are found in the Galactic halo at kiloparsec distances from the plane (Bergemann et al., 2018). Unsupervised machine learning measurements of stellar populations in the solar neighborhood reveal filamentary or string-like clusters that are oriented parallel to the Galactic disk, partly tens of parsecs above and below the plane (Kounkel & Covey, 2019). Their filamentary shape is argued to be primordial and not an effect of tidal interaction, suggesting that the structure might be inherited from parental molecular clouds.

4.5 Conclusions

We have studied the atomic gas properties within the giant atomic filament Maggie. Kinematic information was obtained from the spectral decomposition of H I emission spectra using the automated Gaussian decomposition routine GaussPy+. The main results are summarized as follows:

1. Maggie is one of the largest coherent filaments identified in the Milky Way, detected solely in atomic gas. At a kinematic distance of ∼17 kpc, the projected length of 1.2 ± 0.1 kpc exceeds those of the largest molecular filaments identified to date. Maggie has an aspect ratio of 30:1, thus making its filamentary shape comparable to smaller-scale molecular filaments. Despite being parallel to the Galactic disk, Maggie has a distance of ≈500 pc from the midplane.

2. The centroid velocity information obtained from the spectral decomposition reveals an undulating velocity structure with a smooth gradient of less than $3\,\mathrm{km\,s^{-1}}\,(10\,\mathrm{pc})^{-1}$. Based on the kinematic information, Maggie could be remotely linked to the global spiral structure of the Milky Way as it is trailing the Outer Arm by $5-10\,\mathrm{km\,s^{-1}}$. There is a slight bimodality in the line width distribution that is attributed to the phase degeneracy of the H I emission and broad components blending in at similar v_{LSR} velocities.

3. Optical depth measurements against strong continuum sources indicate that Maggie is composed of a significant fraction of CNM. Detections and nondetections of absorption toward extragalactic sources and Galactic H II regions, respectively, suggest that the distance inference based on the kinematic information is reasonable and that Maggie is located on the far side of the Galaxy. K_{S}-band extinction obtained from 2MASS and WISE combined with Gaia parallaxes confirm that Maggie is not located within a distance of 5 kpc from us. After correcting for optical depth effects, the column density has a mean of $\langle N_{\mathrm{H\,I}} \rangle = 4.8 \times 10^{20}\,\mathrm{cm^{-2}}$. We estimate a "dense" filament mass of $M = 7.2^{+2.5}_{-1.9} \times 10^5\,M_\odot$, providing a large atomic gas reservoir for molecular gas to form. If we assume radial symmetry about the major axis, the column densities equate to an average hydrogen number density of $\sim 4\,\mathrm{cm^{-3}}$.

4. We find no molecular counterpart to Maggie as traced by integrated CO emission or dust. However, by summing diffuse CO emission voxels within H I cloudlets obtained with a dendrogram analysis, Nakanishi et al. (2020) confirm the presence of molecular gas on the smallest spatial scales. But as the molecular gas mass is only a few percent of the total mass budget of the filament, Maggie still appears to be in a predominantly atomic phase. Yet we note that CO might not be a good probe at low column densities and early evolutionary stages as the onset of molecular filaments becoming CO-bright could lag behind the formation of H_2. Although the formation timescale of molecular hydrogen in CNM-rich structures can be on the order of a few tens of Myr and Maggie exposes the onset of molecule formation in the highest-density pockets of H I, the formation of H_2 is not yet likely to be efficient in the more diffuse parts of the filament.

5. Kinematic signatures are investigated by means of the Mach number distribution and reveal transonic and moderately supersonic velocities. The velocity structure function exhibits a modest power-law scaling that is still consistent with the Kolmogorov scaling for subsonic turbulence and is at the lower end of commonly observed scaling exponents in molecular clouds. The column density PDF can be represented by a log-normal distribution with a moderate width of $\sigma = 0.28$, indicative of an intermediate CNM-WNM mix that is not dominated by the effects of gravitational contraction.

6. Given the large distance from the Galactic midplane while remaining well aligned with the disk, we speculate that Maggie could be the signature of Galactic disk oscillations in the vertical direction. Future simulations of atomic and molecular structure formation in spiral galaxies under the influence of vertical perturbations could help constrain this possible formation pathway.

Clearly, further work is needed to better understand the origins of Maggie. Higher-sensitivity CO observations as well as CO-dark molecular gas tracers will facilitate our understanding of this unique filament. While Maggie so far constitutes a case study, future studies of the entire THOR survey as well as other H I surveys will investigate the more general nature of such large atomic filamentary structures.

5

Cold atomic gas identified by H I self-absorption

Cold atomic clouds toward giant molecular filaments

J. Syed, *H. Beuther, P. F. Goldsmith, Th. Henning, M. Heyer, R. S. Klessen, J. M. Stil, J. D. Soler, L. D. Anderson, J. S. Urquhart, M. R. Rugel, K. G. Johnston, A. Brunthaler*

This following Chapter is based on the work *Cold atomic gas identified by* H I *self-absorption – Cold atomic clouds toward giant molecular filaments* that is submitted to Astronomy&Astrophysics and is currently under review.

The paper has been reformatted to match the format of this thesis. The work on the paper was distributed as follows: I developed the code the following analysis is based on and conducted the scientific analyses presented in this Chapter.

I drafted a first version of the entire manuscript. Co-authors contributed by commenting on at least one version of the draft and giving suggestions to compare the code to traditional approaches and to test the limitations of the algorithm.

ABSTRACT

Stars form in the dense interiors of molecular clouds. The dynamics and physical properties of the atomic interstellar medium (ISM) set the conditions under which molecular clouds and eventually stars will form. It is therefore critical to investigate the relationship between the atomic and molecular gas phase to understand the global star formation process. Using the high angular resolution data from The H I/OH/Recombination line survey of the Milky Way (THOR), we aim to constrain the kinematic and physical properties of the cold atomic hydrogen gas phase toward the inner Galactic plane. H I self-absorption (HISA) has proven to be a viable method to detect cold atomic hydrogen clouds in the Galactic plane. With the help of a newly developed self-absorption extraction routine (astroSaber), we extend upon previous case studies to identify H I self-absorption toward a sample of Giant Molecular Filaments (GMFs). We find the cold atomic gas to be spatially correlated with the molecular gas on a global scale. The column densities of the cold atomic gas traced by HISA are usually of the order of $10^{20}\,\mathrm{cm}^{-2}$ whereas those of molecular hydrogen traced by ^{13}CO are at least an order of magnitude higher. The HISA column densities are attributed to a cold gas component that accounts for a fraction of ∼5% of the total atomic gas budget within the clouds. The HISA column density distributions show pronounced log-normal shapes and clearly bridge the gap between the diffuse atomic interstellar medium as traced by H I emission and the molecular gas. The cold atomic gas is found to be moderately supersonic with Mach numbers of a ∼few, while the molecular gas within the majority of the filaments is driven by highly supersonic dynamics. While H I self-absorption is likely to trace just a small fraction of the total cold neutral medium within a cloud, probing the cold atomic ISM by the means of self-absorption significantly improves our understanding of the dynamical and physical interaction between the atomic and molecular gas phase during cloud formation.

5.1 Introduction

Atomic hydrogen provides the raw material to form molecular clouds, the sites of star formation. The dynamical and physical conditions under which molecular clouds form are therefore critical to understand the global star formation process. On a large scale ($\gtrsim 10^2\,\mathrm{pc}$), molecular clouds form out of the diffuse atomic interstellar medium (ISM; for a review see Ferrière, 2001; Draine, 2011; Klessen & Glover, 2016) and are shaped by galactic dynamics and turbulence, stellar feedback, and magnetic fields. One major constituent of the ISM is the neutral atomic gas that provides the raw material to molecular clouds out of which stars eventually form and this gas phase takes in most of the energy and momentum feedback from its environment. The kinematic and physical relationship between the atomic and molecular gas phase of the ISM is then of central interest in the understanding of cloud formation processes.

According to the classical photodissociation region (PDR) model, layers of cold atomic hydrogen can effectively shield the cloud from photo-dissociating UV radiation at sufficiently high densities, allowing a transition of atomic hydrogen to its molecular form. In this idealized picture, pockets of high-density molecular hydrogen are embedded in an envelope of cold atomic hydrogen.

Using high angular resolution data of H I emission, the structure of the atomic ISM can be studied in great detail. However, probing the physical properties of the atomic gas from H I emission studies alone is not straightforward. In thermal

pressure equilibrium, theoretical considerations based on ISM heating and cooling processes predict two stable phases of atomic hydrogen at the observed pressures in the ISM, namely the cold neutral medium (CNM) and warm neutral medium (WNM; Field et al., 1969; McKee & Ostriker, 1977; Wolfire et al., 2003; Bialy & Sternberg, 2019). Observations of H I emission are thus generally attributed to both CNM and WNM, which have significantly different physical properties (see below). In an attempt to observationally isolate the CNM from the bistable emission, H I self-absorption (HISA; Gibson et al., 2000; Li & Goldsmith, 2003; Wang et al., 2020c; Syed et al., 2020) is a viable method to identify cold atomic gas and study H I clouds in the inner Milky Way but it heavily depends on the presence of sufficient background emission. In the following, we refer to cold atomic gas traced by HISA as "CoAt" gas, to make a distinction between the CNM as a whole and HISA-traced cold gas, which is a subset of the CNM. Due to the Galactic rotation, any positive (negative) line-of-sight velocity in the first (fourth) Galactic quadrant generally corresponds to a near kinematic and far kinematic distance within the solar circle (see e.g., Burton, 1988). HISA has been used to resolve this kinematic distance ambiguity for molecular clouds or Galactic H II regions. Sources of interest at the far distance are less likely to show HISA as there is less background to absorb. Any detection of corresponding HISA would then place molecular clouds or H II regions at the near kinematic distance (e.g., Jackson et al., 2002; Anderson & Bania, 2009; Duarte-Cabral et al., 2021).

Since the warm component of atomic hydrogen is more diffuse and has a lower density, it fills up a larger volume than the cold component (McKee & Ostriker, 1977; Stahler & Palla, 2005; Kalberla & Kerp, 2009). H I self-absorption occurs when a cold H I cloud is located in front of a warmer H I emitting cloud. Self-absorption can occur within the same cloud but can also be induced by an emitting cloud in the far background that has the same velocity as the absorbing medium with respect to the local standard of rest $v_{\rm LSR}$. Therefore, the clouds do not have to be physically associated for HISA to be observable.

The CNM is observed to have temperatures $\lesssim$300 K and number densities of $\gtrsim n_{\rm min,CNM}$=10 cm^{-3} while the thermally stable WNM exceeds temperatures of $\sim$5000 K with number densities $\lesssim n_{\rm max,WNM}$=0.1 cm^{-3} (Heiles & Troland, 2003; Kalberla & Kerp, 2009). In contrast to the properties of the CNM, the atomic gas traced by HISA has typical spin temperatures below 100 K, in most cases even below 50 K (e.g., Gibson et al., 2000; Li & Goldsmith, 2003; Krčo et al., 2008; Wang et al., 2020c), thus highlighting the limited sensitivity of the HISA method for higher-temperature gas. For densities between $n_{\rm min,CNM}$ and $n_{\rm max,WNM}$, the gas is thermally unstable (denoted by UNM: unstable neutral medium) and it will move toward a stable CNM or WNM branch under isobaric density perturbations (Field, 1965).

H I self-absorption is found throughout the Milky Way in various environments. Many studies have focused on the detection of HISA, first measured in 1954 (Heeschen, 1954, 1955), in known sources, but statistical treatments of the kinematic properties and densities of the HISA-traced cold gas in large-scale high-resolution maps are still scarce.

Extensive investigations of the HISA properties are limited to individual observational case studies (e.g., Gibson et al., 2000; Li & Goldsmith, 2003; Kavars et al.,

2003; Krčo et al., 2008; Wang et al., 2020c; Syed et al., 2020) and few simulations (e.g., Seifried et al., 2022). In this paper, we aim to extend upon these case studies and investigate the HISA properties toward a sample of Giant Molecular Filaments (GMFs; Ragan et al., 2014) as they are likely to be at an early evolutionary stage of giant molecular clouds forming out of the atomic phase of the interstellar medium (Zucker et al., 2018). These giant filaments potentially trace the Galactic structure, such as spiral arms and spurs, and large concentrations of molecular gas.

Seifried et al. (2022) present synthetic H I observations including H I self-absorption toward molecular clouds and investigate the observational effects and limitations of HISA. Their synthetic observations are based on 3D-MHD simulations within the scope of the SILCC-Zoom project that include out-of-equilibrium H I-to-H_2 chemistry, detailed radiative transfer calculations, as well as observational effects like noise and limited spatial and spectral resolution that are similar to those of the THOR survey (see Sect. 5.2.1). Using commonly employed methods to derive the HISA properties, their results show that the H I column densities inferred from self-absorption tend to underestimate the real column densities of cold H I in a systematic way.

Traditionally, HISA features are obtained through various methods. To quantify the absorption depth, the strength of the warm emission background inducing H I self-absorption has to be determined first. Since the warm atomic gas component is often assumed to have less spatial variation in intensity, the emission background is commonly estimated by taking an "off" position spectrum at a different line of sight close to the location where HISA is expected to occur (e.g., Gibson et al., 2000; Kavars et al., 2003). However, it is challenging to select a position that is close enough, such that the off position can serve as a good proxy for the true HISA background, but far enough not to be interfered with by (partially) self-absorbing medium. Therefore, the location and spatial distribution of self-absorbing gas has to be known prior to estimating the background, which may work for single isolated cases. But particularly toward the Galactic plane, where multiple emission components and the Galactic rotation can add to the confusion along the line of sight, finding a clean off position has proven to be difficult to accomplish since the off spectrum can vary significantly over the angular size of the background cloud (see Wang et al., 2020c).

Another approach is to recover self-absorption in the spectral domain of H I observations. If the location of HISA in the H I emission spectrum is known, the baseline can be determined by fitting the emission range around a HISA feature with a simple polynomial or Gaussian function (Li & Goldsmith, 2003; Kavars et al., 2003; Wang et al., 2020c; Syed et al., 2020). However, the assumption of a velocity range where HISA is located introduces an additional source of bias, together with the specific fitting function that is used to derive the background. To address this issue, Krčo et al. (2008) and Dénes et al. (2018) have employed second and higher derivatives of the emission spectra to search for narrow HISA features (HINSA; Li & Goldsmith, 2003; Goldsmith & Li, 2005; Goldsmith et al., 2007) over the entire spectral range in a more unbiased way. Sharp kinks and dips in the spectra that are due to self-absorption are therefore expected to become readily apparent when investigating the derivatives. This technique allows HISA features to be filtered out

without prior knowledge of their central velocities but it relies on high sensitivity, a well-sampled HISA line width, and HISA features that are much narrower than the average emission component. However, the baselines of these identified absorption features would then still need to be obtained using, for example, polynomial fits or making other physical assumptions of the HISA properties (e.g., Krčo et al., 2008).

In this paper, we address the lack of a versatile self-absorption reconstruction algorithm that can be applied to any data set, at any spectral resolution, and self-absorption line width, and without the prior assumption that the cold H I gas is tightly correlated with molecular gas. We present the algorithm astroSABER (**S**elf-**A**bsorption **B**aseline **E**xtracto**R**) that operates by smoothing emission spectra in an asymmetric way, such that it not only identifies signal dips in the spectrum but directly provides a baseline of potential self-absorption features. It works in multiple iterations, such that both narrow and broad absorption components can be recovered. An optimization step has been implemented that is designed to tune the amount of smoothing that is required to recover self-absorption features, irrespective of spectral resolution and line width. To test the performance and applicability of the algorithm, we apply astroSABER to the known sample of GMFs (Ragan et al., 2014) since they serve as a good laboratory to investigate the presence of CoAt gas. The properties of H I self-absorption toward two of these molecular filaments have already been investigated in dedicated case studies employing previous HISA extraction methods (GMF20.0-17.9 in Syed et al., 2020 and GMF38.1-32.4 in Wang et al., 2020c).

The paper is organized as follows: In Sect. 5.2 we briefly introduce the data used in this analysis and outline the methods of our newly developed H I self-absorption extraction routine and Gaussian decomposition. In Sect. 5.3 we present the kinematic and column density properties derived from the HISA extraction and spectral decomposition. We discuss the kinematic and spatial relationship between the CoAt gas and molecular gas as well as the column density properties in Sect. 5.4. We furthermore elaborate on some of the limitations of our HISA extraction method before concluding with our summary in Section 5.5.

5.2 Methods and observations

5.2.1 H I, CO, and continuum observations

The following analysis of the HISA properties toward molecular clouds is based on the H I and 1.4 GHz continuum observations as part of The H I/OH Recombination line survey of the inner Milky Way (THOR; Beuther et al., 2016; Wang et al., 2020a). The final THOR-H I and 1.4 GHz continuum data include observations taken with the Karl G. Jansky Very Large Array (VLA) in C-configuration that were combined with the H I Very Large Array Galactic Plane Survey (VGPS; Stil et al., 2006), which consists of VLA D-configuration data. To account for missing flux on short *uv* spacings, the VGPS also includes single-dish observations of H I and 1.4 GHz continuum taken with the Green Bank and Effelsberg 100m telescope, respectively. The final H I emission data, from which the continuum has been subtracted during the data reduction, have an angular and spectral resolution of $\Delta\Theta = 40''$ and

1.5 km s^{-1}, respectively. The rms noise in emission-free channels is ∼4 K.

We selected six GMF regions to investigate the presence of CoAt gas. Our selection is based on the findings of Ragan et al. (2014) and Zucker et al. (2018). Ragan et al. (2014) identified seven mid-infrared extinction features as giant filaments that exhibit corresponding ^{13}CO emission and velocity coherence over their full length. Of these seven GMFs, six of the fields are covered by the THOR survey. We present an overview of the six fields covering the filament regions in Table 5.1. The indices of the source names refer to the approximate range in Galactic longitude the giant filaments cover. The selected filament regions are in close proximity to the Galactic midplane and are located in the inner disk of the Milky Way, a site where HISA is more likely to occur. These GMFs serve as a good laboratory to investigate the relationship between the atomic and molecular gas as they are molecular concentrations of lengths >50 pc and likely to be at an early evolutionary stage having formed out of the large-scale diffuse ISM (Zucker et al., 2018). More details about each region can be found in Ragan et al. (2014). In Sect. 5.3.2, we correct for optical depth

Table 5.1: Properties of studied filament regions.

(1)	(2)	(3)	(4)	(5)	(6)
Source name$^{(a)}$	Glon	Glat	$v_{\rm LSR}$	$d_{\rm near}{}^{(b)}$	$D_{\rm GC}{}^{(b)}$
	[°.°]	[°.°]	[km s^{-1}]	[kpc]	[kpc]
GMF20.0-17.9	17.80 – 20.60	−1.00 – +0.30	37 – 50	3.2	5.2
GMF26.7-25.4	25.10 – 26.90	+0.40 – +1.20	41 – 51	2.9	5.7
GMF38.1-32.4a	33.30 – 37.30	−1.00 – +0.60	50 – 60	3.2	5.9
GMF38.1-32.4b	33.30 – 37.30	−1.00 – +0.60	43 – 46	2.6	6.2
GMF41.0-41.3	40.80 – 41.50	−0.70 – +0.50	34 – 42	2.2	6.7
GMF54.0-52.0	52.30 – 54.20	−0.50 – +0.40	20 – 26	1.4	7.4

Notes. Columns (2) and (3) give the Galactic longitude range and latitude range of the filament regions, respectively. Column (4) gives the line-of-sight velocity range of each GMF as defined in Ragan et al. (2014). Columns (5) and (6) give the kinematic near distance from us and the Galactocentric distance, respectively.
$^{(a)}$ as in Ragan et al. (2014).
$^{(b)}$ The distances are not taken from Ragan et al. (2014) but have been recalculated using the more recent spiral arm model by Reid et al. (2019).

effects to compute the atomic hydrogen column densities from H I emission. The optical depths are taken from the measurements provided by Wang et al. (2020a) and have been obtained from VLA C-configuration data only that have an angular resolution of ∼15″ and effectively filter out large-scale emission, such that H I absorption against discrete continuum sources can yield a direct measurement of the optical depth of atomic hydrogen.

In order to provide a comprehensive description of the kinematic and spatial relationship between the atomic gas and the molecular gas, we investigate the molecular gas properties toward the GMF regions using two different data sets. The kinematic information is based on the ^{13}CO(1–0) data of the Galactic Ring Survey (GRS;

Jackson et al., 2006), with an angular and spectral resolution of 46″ and 0.21 km s^{-1}, respectively. Riener et al. (2020a) present an overview of a Gaussian decomposition of the entire GRS using the fully automated GaussPy+ algorithm (Riener et al., 2019). Since the decomposition results are publicly available, we use these data to investigate the kinematic properties of the clouds.

In Sect. 5.3.2 we compute the ^{13}CO column densities from the ^{12}CO(1–0) and ^{13}CO(1–0) emission line data taken from the Milky Way Imaging Scroll Painting survey (MWISP; Su et al., 2019). The GRS does not include ^{12}CO observations, that are required to estimate the CO excitation temperatures and ultimately ^{13}CO column densities. We therefore use both the ^{13}CO and ^{12}CO from the MWISP data to derive the column density properties in a consistent way, and to reduce systematic errors arising from observational biases. The MWISP ^{12}CO and ^{13}CO data have an angular resolution of ∼55″ and an rms noise of 0.5 K and 0.3 K at a spectral resolution of 0.16 km s^{-1} and 0.17 km s^{-1}, respectively. The ^{12}CO data have been reprojected onto the same spectral grid as the ^{13}CO data to infer the excitation temperatures on a voxel-by-voxel basis. The rms noise of the ^{12}CO data is then reduced to 0.4 K.

5.2.2 Absorption baseline reconstruction

In this section we describe the astroSaber method that we used to obtain self-absorption baselines to recover HISA features. The basic workflow of astroSaber is the following: 1) Generating mock H I spectra to use as "training data"[1] (described in Sect. 5.2.2), 2) finding optimal smoothing parameters using gradient descent (described in Sect. 5.2.2 and Appendix B.1), 3) applying baseline extraction with optimal smoothing parameters found in step 2) (described in Sect. 5.2.2).

Asymmetric Least Squares Smoothing

The publicly available python-based astroSaber algorithm (`https://github.com/astrojoni89/astroSABER`) is an automated baseline extraction routine that is designed to recover baselines of absorption features that are superposed with H I emission spectra. It utilizes asymmetric least squares smoothing first proposed by Eilers (2004) in the context of Raman spectroscopy. The algorithm progresses iteratively in two cycles to obtain a smoothed baseline, the major (outer) cycle and the minor (inner) cycle executed at each iteration of the major cycle. The basis of the minor cycle is to find a solution that minimizes the penalized least squares function

$$F(\mathbf{z}) = (\mathbf{y} - \mathbf{z})^\top \mathbf{W} (\mathbf{y} - \mathbf{z}) + \lambda \, \mathbf{z}^\top \mathbf{D}^\top \mathbf{D} \mathbf{z} \,, \tag{5.2.1}$$

where $\mathbf{y}$ is the input signal (e.g., the observed H I spectrum) and $\mathbf{z}$ is the asymmetrically smoothed baseline to be found. The first and second term on the right-hand side describe the fitness of the data and the smoothness of $\mathbf{z}$ defined by the second order differential matrix $\mathbf{D}$, respectively. The parameter λ, which is a two-dimensional

[1] While the terms "test data" and "training data" are commonly used in the context of machine learning algorithms, we note that the accuracy of astroSaber is not tested on unseen data but the underlying concepts are the same, such that these concepts can be used to integrate them in a machine learning algorithm.

vector by default (see below), adjusts the weight of the smoothing term. In order to correct the baseline with respect to peaks and dips in the spectrum, the asymmetry weighting matrix $\mathbf{W} = \mathrm{diag}(\mathbf{w})$ is introduced. The asymmetry weights are initialized to be $w_i = 1$. After a first iteration of the minor cycle with equal weights, the weights for channels containing signal are then assigned as follows:

$$w_i = \begin{cases} p, & y_i > z_i \\ 1-p, & y_i \leq z_i \end{cases} . \tag{5.2.2}$$

The asymmetry parameter $p \in [0, 1]$ is set to favor either peaks or dips while smoothing the spectra. Given both the parameters λ and p, a smoothed baseline $\mathbf{z}$ is updated iteratively. Depending on p and the deviation of $\mathbf{z}$ from $\mathbf{y}$ after each iteration, peaks (dips) in the spectrum will be retained while dips (peaks) will be given less weight during the smoothing. Since we only aim to asymmetrically smooth real signals, spectral channels containing only noise will be given equal weights of 0.5, hence the baseline will be within the noise in emission-free channels. The signal range estimation is described in Sect. 5.2.2. As can be seen in Eq. (5.2.1), there is a degeneracy in the solution of the least squares function introduced by the weighting factors $\mathbf{W}(p)$ and λ. It is then sensible to keep one of these parameters fixed while finding the best-fit solution for the other parameter in order to optimize the smoothing (see Sect. 5.2.2). In the case of self-absorption features, we therefore chose to fix the asymmetry parameter at $p = 0.9$.

After n_{minor} iterations, the minor cycle converges, such that the iteratively updated baseline $\mathbf{z}$ will not change anymore given the input spectrum $\mathbf{y}$. However, in order to effectively smooth out dips while still retaining real signal peaks in the spectra, the smoothed baseline $\mathbf{z}$ is then passed to the next iteration of the major cycle as an input (i.e. now $\mathbf{y}$) for its minor cycle smoothing.

After evaluating the THOR-H I data, the minor cycle has shown to already converge after three iterations. Hence, the number of minor cycle iterations has been fixed at $n_{\mathrm{minor}} = 3$ in the algorithm. This parameter affects the output of astroSABER only mildly since the final smoothed baseline is mostly dependent on the number of iterations in the major cycle and on the λ parameter that tunes the smoothing (see Sect. 5.2.2).

The algorithm stops as soon as a convergence criterion in the major cycle is met, or if the maximum number of iterations (`niters`) is reached. The convergence criterion is met if the change in baseline from one major cycle iteration to the next is below a threshold set by `thresh`=`sig * noise` for at least some number of iterations (given by `iterations_for_convergence`). The default values set by astroSABER are `niters`=20, `sig`=1, and `iterations_for_convergence`=3. There is a slight degeneracy between the actual number of iterations needed to make the baseline converge and a fixed smoothing parameter λ used for every smoothing iteration. For λ sufficiently high, fewer iterations are needed to smooth out sharp kinks and dips in the spectrum. In the case of the THOR-H I data where the continuum has been subtracted during data reduction, the maximum number of iterations can be reached for emission spectra that are contaminated by imperfect continuum subtraction toward very strong continuum sources. Inspecting the number of iterations can therefore serve as an additional quality check of the spectra. For high-sensitivity

data at a spectral resolution that is much smaller than the HISA line width, the optimal smoothing parameter λ might be too large to make the algorithm converge since the change in baseline will be significant after every major cycle iteration. It can then be sensible to decrease the convergence threshold or to reduce the maximum number of iterations to force the algorithm to terminate and thus break down the aforementioned degeneracy.

In order to recover both narrow and broad features and to account for the possibility of an absorption baseline that exceeds the intensity of that in adjacent velocity channels, the astroSABER routine can be set to add a residual (`add_residual=True`), which is the absolute difference between the first and last iteration of the major cycle. An example of this is an isolated emission feature with a Gaussian shape that has an absorption dip at the line center, or the "flat-top" spectrum observed in [C II] emission toward the H II region RCW120 (Kabanovic et al., 2022, see their Fig. 6). To add flexibility to the baseline reconstruction, the very first iteration can be set to operate with its own individual smoothing parameter λ_1 while all following iterations use a smoothing parameter λ_2. A λ_2 smoothing parameter close to zero is then effectively equal to a spectral smoothing without adding the residual. In Sect. 5.2.2 we investigate how to optimize the smoothing parameters using mock-H I data.

Figure 5.1 shows a step-by-step baseline extraction of a mock spectrum to illustrate the major cycle workflow. The mock-H I contains three emission components where two absorption features of different line widths have been added. Given the observed spectrum (black spectrum in Fig. 5.1), astroSABER is run with optimal smoothing parameters (λ_1, λ_2) (see Sect. 5.2.2). The left panel in Fig. 5.1 shows the baseline after the first major cycle iteration, that is after the minor cycle smoothing converged given the input spectrum (i.e. after Eq. (5.2.1) has been solved for $\mathbf{z}$). The middle panel then shows the converged baseline after the last major cycle iteration before adding the residual. The right panel presents the final baseline obtained by astroSABER after adding the residual. The baseline so obtained is able to recover the pure emission spectrum well within the uncertainties. We note that if the `add_residual` setting is turned off, the smoothing parameters obtained during the optimization (Sect. 5.2.2) will be adjusted to have larger values in order to recover the baseline. The differences in baseline between these settings are likely to be small at the velocities of the absorption signals. However, real signal is then also more likely to be smoothed out by the higher smoothing weight.

Test data and optimization

We implemented a gradient descent method (see e.g., Ruder, 2016) that uses generated mock data to find the optimal smoothing parameters λ_1 and λ_2. The $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$ parameter generally depends on the spectral resolution, noise level, line width of the absorption features, and on variations in the emission background. In a preparatory step astroSABER generates "pure emission" and "observed" mock data that are based on the actual observational data. The pure emission data represent emission line spectra that do not contain any absorption features and those are used as the test data that are to be recovered by the astroSABER algorithm. The observed data contain spectra where randomly generated (but known) absorption features were added to the pure emission data, and those are used as the input

data for astroSABER. The mock data are generated by randomly selecting a defined number of spectra (set by `training_set_size`) taken from the real observational data on which the baseline extraction is to be applied later. The algorithm then uses asymmetric least squares smoothing with predefined and fixed parameters to smooth the spectra in the training set. The algorithm then adds a user-defined noise level to the spectra, thus creating spectra that will be used as pure emission data to be recovered. The reason for the smoothing in this preparatory step is to remove any dips that are present in the real data, such that generated absorption features can be added anywhere in the spectrum. The exact setting of the parameters p and $\boldsymbol{\lambda}$ in the preparation step does not heavily affect the optimization as these parameters are only used to generate data that show similarity to the overall structure of the real spectra. We have taken samples of 200 spectra for each filament region to use as test data to find the optimal smoothing parameters. The randomly gener-

Table 5.2: Optimal smoothing parameters.

Source	λ_1	λ_2
GMF20.0-17.9	3.10	0.56
GMF26.7-25.4	3.40	0.46
GMF38.1-32.4a	2.76	0.50
GMF38.1-32.4b	$2.76^{(a)}$	$0.50^{(a)}$
GMF41.0-41.3	3.45	0.43
GMF54.0-52.0	3.70	0.39

Notes. The second column and third column give the best-fit λ_1 and λ_2 smoothing parameters obtained during the optimization step of astroSABER, respectively. Minor differences in these optimal parameters are expected due to different noise and fluctuations in the emission spectra.
$^{(a)}$ Since the optimal smoothing parameters are obtained from the same data cube as GMF38.1-32.4a, the $\boldsymbol{\lambda}$ values are the same.

ated absorption spectra are created using Gaussian functions whose 1) amplitude, 2) mean, and 3) standard deviation parameters are drawn from normal distributions with the following mean and standard deviation (μ and σ) by default: 1) the amplitude values follow a normal distribution with $\mu_{\rm amp} = 7\sigma_{\rm rms}$, and $\sigma_{\rm amp} = 1\sigma_{\rm rms}$, where $\sigma_{\rm rms}$ is the noise of the observational data, 2) the mean velocity values follow a normal distribution where the mean $\mu_{\rm mean}$ is set by the central velocity at which there is signal in each spectrum, its standard deviation is set accordingly, such that $3\sigma_{\rm mean}$ is at the edge of the signal range of the spectrum, 3) the magnitude of the $\boldsymbol{\lambda}$ parameters that is required for smoothing is crucially dependent on the width of the absorption features, so the standard deviation values of the absorption features drawn from a normal distribution have to be defined by the user. In the case of the THOR-H I data, Wang et al. (2020c) and Syed et al. (2020) report HISA FWHM values of $\sim$4 km s^{-1}. We have therefore set the mean and standard deviation of the line width distribution to $\mu_{\rm lw} = 4\,{\rm km\,s^{-1}}$ and $\sigma_{\rm lw} = 1\,{\rm km\,s^{-1}}$, respectively. The number of self-absorption components that are added to each spectrum are drawn from a normal distribution with $\mu_{\rm n} = 2.0$ and $\sigma_{\rm n} = 0.5$ by default, where all samples

below 0.5 are set to 1.0 to add at least one self-absorption feature to each spectrum.

Once the mock spectra have been generated, a gradient descent algorithm is run to find the optimal smoothing parameters λ_1 and λ_2. The gradient descent is designed to minimize the residual between the actual test data (pure emission mock spectra) and the baselines obtained by the astroSABER smoothing routine. Since we do not expect large variations in the emission spectra and absorption baselines within single HISA regions, we aim to find single λ_1 and λ_2 values that we then apply to the whole filament region in each case. We have found that the median value of all reduced chi square results of the entire training data serves as a good metric for the cost function and it has shown to be quite robust in recovering good baselines for the entire training set. The reduced chi square is only evaluated in channels where artificial absorption features have been added. More details about the gradient descent method applied in this paper can be found in Appendix B.1.

With increasingly complex emission spectra containing multiple broad and narrow emission peaks as well as absorption features, adding a residual to a moderately smoothed spectrum has shown to give the best results for all the mock data that we have tested. In the subsequent analysis, we have therefore used the two-phase smoothing with two $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$ parameters and added a residual that is the difference between the very first major cycle iteration (using λ_1) and the last major cycle iteration (using λ_2). We note, however, that a simpler one-phase smoothing without adding a residual might give generally good results, depending on the signal-to-noise and complexity in the spectrum. The final smoothing parameters obtained in the optimization step of astroSABER that were used for the final baseline reconstruction are listed in Table 5.2. The inferred smoothing parameters are similar toward all filament regions and compare well to each other. We expect small differences between the samples because of the training data sets containing different emission spectra and noise. These minor differences in the smoothing parameters only have a limited impact on the extraction results (see also Fig. B.1). In particular, the accuracy of the fit results does not heavily depend on λ_1. Figure B.1 shows similar accuracy in the fits for a range between 3–6. An example of the final output of the extraction step is shown in Fig. 5.2. The figure shows maps of an example region toward a (100×100) pixel subsection of GMF20.0-17.9 (see Ragan et al., 2014; Syed et al., 2020) that is also made publicly available with the astroSABER code. The maps present the H I emission data, the baselines obtained with optimized smoothing parameters, and the resulting H I self-absorption data, respectively.

Noise and signal range estimation

The signal ranges of the spectra are determined by borrowing parts of the noise estimation routine included in the GAUSSPY+ tool described in Sect. 5.2.3. For a detailed description, we refer the reader to Sect. 3.1.1 in Riener et al. (2019). The underlying assumptions to determine signal or noise ranges in the spectra are as follows: 1) the noise distribution is Gaussian, 2) the spectral channels are uncorrelated, 3) the noise has a mean around zero. The assumption is that a spectrum containing white noise has on average an equal number of negative and positive channels. We can then estimate the probability of a given positive or negative feature observed in consecutive spectral channels to be caused by white noise (as opposed to be due

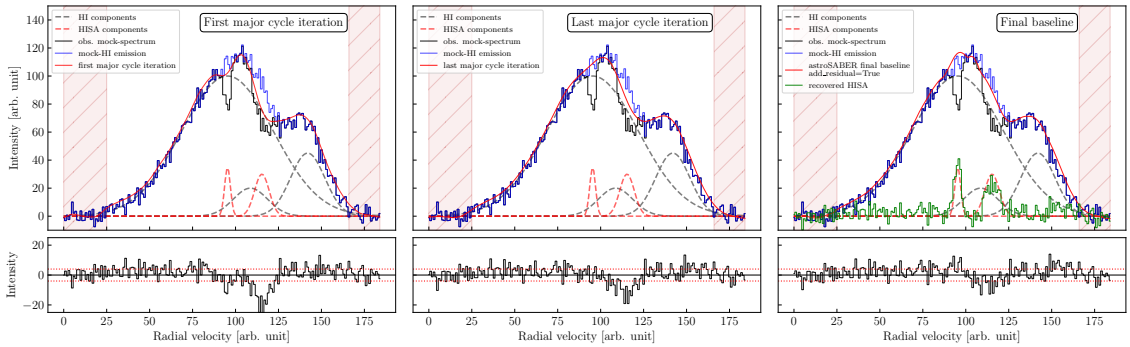


Figure 5.1: Baseline extraction workflow of astroSaber. In each panel, the black mock spectrum represents the observed H I emission spectrum, which is the sum of the three gray dashed components, with self-absorption features (two red dashed components) superposed. The blue spectrum shows the "pure emission" spectrum that is to be recovered by the astroSaber algorithm. The algorithm is then applied to the observed spectrum using the optimal smoothing parameters (λ_1, λ_2). Hatched red areas indicate spectral channels that are masked out due to missing signal. *Left panel:* The astroSaber baseline (red) after the first major cycle iteration, that is, after the minor cycle smoothing converged given the input mock spectrum (i.e. after Eq. (5.2.1) has been solved for $\mathbf{z}$). *Middle panel:* The astroSaber baseline (red) after the last major cycle iteration, that is, after the major cycle smoothing converged and before adding the residual, which is the absolute difference between the first and last major cycle iteration. *Right panel:* The final astroSaber baseline (red) after adding the residual. The baseline so obtained reproduces the pure emission spectrum (blue) well. The resulting HISA features expressed as equivalent emission features are shown in green, and show a good match with the the real HISA absorption features. The smaller subpanels in each column show the residual, which is the difference between the red baseline and the blue emission spectrum, with the horizontal dotted red lines marking values of $\pm\sigma_{\rm rms}$.

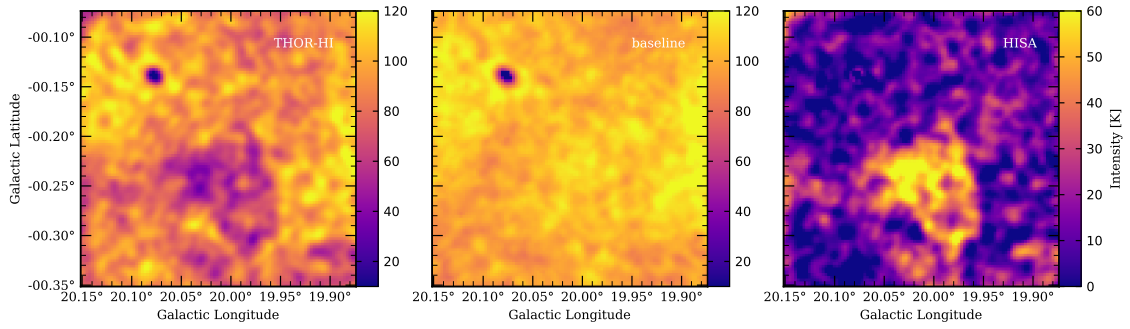


Figure 5.2: Example H I self-absorption extraction. The *left panel* shows the observed THOR-H I emission channel map toward a (100×100) pixel subsection of the giant filament GMF20.0-17.9 at the velocity $44.5\,\mathrm{km\,s^{-1}}$. The *middle panel* shows the map of the self-absorption baseline obtained with optimized smoothing parameters. The *right panel* gives the resulting HISA map, which is the difference between the baseline map and the H I emission map. The HISA feature in the bottom half of the map could be successfully recovered by astroSaber, while the strong continuum source in the top left was masked during a quality check of the spectra.

to real signal) using a Markov chain. The routine to determine signal ranges then selects all features in spectral channels that have a probability to be caused by white noise of less than a user-defined threshold. We set this probability limit value to $P_{\rm limit} = 1\%$. In the case of the THOR-H I data with 185 channels per spectrum, we find that all features with more than 15 consecutive positive channels have a

probability to be caused by noise of less than $P_{\rm limit} = 1\%$. To set the mean velocity in the velocity distribution of the absorption features, we additionally clipped the determined signal ranges by five channels on either side to ensure that the signal has sufficient intensity from which absorption features can be subtracted.

5.2.3 Gaussian decomposition

After astroSABER has been applied to all six giant filament regions with optimized smoothing parameters, in each case the resulting output gives four data cubes containing the reconstructed baseline spectra, the self-absorption features (i.e. the H I emission spectra subtracted from the baselines), a map of the number of iterations that were required for the baselines to converge, and a map with flags for spectra that did not meet the convergence criteria, either due to missing signal in the spectra or having reached the maximum number of iterations set by the user. Spectra are flagged with "missing signal" if there is no significant emission (defined by `check_signal_sigma * noise`) in more than a specified number of consecutive spectral channels (set by `velo_range` [$\mathrm{km\,s^{-1}}$]), and these spectra are removed from the final data cubes by default, as is done for the strong continuum source in Fig. 5.2. For the THOR-H I data, we applied the default settings `check_signal_sigma=6` and `velo_range=15` $\mathrm{km\,s^{-1}}$.

The final self-absorption data cubes obtained from astroSABER contain what we refer to as HISA "candidates" since all signal dips have been extracted from the emission spectra. In the following steps, we decomposed the HISA candidate data cubes into their spectral components using the fully automated Gaussian decomposition algorithm GAUSSPY+[2] (Riener et al., 2019) and identified "real" H I self-absorption by cross-matching the centroid velocities with the molecular kinematics of the GMF regions given in Ragan et al. (2014).

GAUSSPY+ is a multicomponent Gaussian decomposition tool based on the earlier GAUSSPY algorithm (Lindner et al., 2015) and it provides an improved fitting routine and a fully automated means to decompose spectra using machine-learning algorithms. GAUSSPY+ automatically determines initial guesses for Gaussian fit components using derivative spectroscopy. To decompose the spectra, the spectra require smoothing to remove noise peaks while retaining real signal. The optimal smoothing parameters are found by employing a machine-learning algorithm that is trained on a subsection of each data set.

It is essential to reliably estimate the noise in the spectra to obtain good fit results. As we described above, GAUSSPY+ comes with an automated noise estimation routine as a preparatory step for the decomposition that also considers the median absolute deviation (MAD) of negative spectral channels to identify narrow emission peaks in the spectra that are masked before estimating the rms noise.[3]

For each data cube, we ran the GAUSSPY+ training step with 300 randomly selected spectra from the HISA candidate data to find the optimal parameters for the

[2] https://github.com/mriener/gausspyplus

[3] This step in the noise estimation of GAUSSPY+ is not included in astroSABER since we only want to identify signal ranges that are broad enough such that they can be used for generating mock self-absorption spectra.

fitting, as recommended in Riener et al. (2019). Owing to the absorption properties of the H I gas, we would naturally expect HISA to probe very cold gas so we set a line width limit in the GaussPy+ routine such that broad components are refit, a setting that can easily be adjusted in the GaussPy+ routine.

After the initial fitting, the algorithm applies a two-phase spatial coherence check that can optimize the fit by refitting the components based on the fit results of neighboring pixels. Mostly one velocity component was fit by GaussPy+ in the given velocity ranges of the filament regions. Only for some small isolated regions and single pixels more than one component was fit to the HISA spectra.

Spectra where the maximum number of iterations (`niters`) is reached during the baseline extraction are flagged but not removed from the astroSaber routine. The affected spectra are usually toward positions where continuum emission contaminates the detection of self-absorption. We removed these spectra manually by masking pixels where there is strong continuum emission $T_{\rm cont} \geq 100$K. Due to the systematic uncertainty in the baselines and to ensure we only report reliable HISA features that are well detected, we additionally masked all pixels of the fit result maps where the corresponding fit amplitude is below $\sqrt{2} \cdot 5\sigma_{\rm rms}$, with $\sigma_{\rm rms}$ being the rms noise of the THOR-H I emission data. The factor $\sqrt{2}$ accounts for the uncertainty in HISA amplitude that is due to the difference between the extracted HISA baseline and the H I emission.

5.3 Results

We show in Table 5.1 an overview of the filament regions analyzed in this paper, which are motivated by the results of Ragan et al. (2014). We will use their designated names (and shortened versions thereof) to refer to these regions thoughout this paper.

We detect HISA toward all six filament regions. However, toward GMF26 and GMF41 only a small amount of CoAt gas could be recovered as HISA. The HISA-traced gas toward GMF26 does not appear to trace the distribution of the molecular gas well. Toward GMF20, GMF38a, and GMF38b we recovered a large cold atomic counterpart to the molecular gas within the filaments.

5.3.1 Kinematics

In this section, we discuss the kinematic properties of both HISA and their molecular counterpart as traced by ^{13}CO emission. As an example, we show the detected HISA and corresponding ^{13}CO emission map toward GMF20 in terms of their centroid velocities in Fig. 5.3. The kinematic maps of the remaining filament regions can be found in Appendix B.4. As we show in Appendix B.3, the centroid velocities and line widths are not heavily affected by our astroSaber routine and have an uncertainty of $0.4\,\mathrm{km\,s^{-1}}$ and $1.0\,\mathrm{km\,s^{-1}}$ (FWHM), respectively. Since the beam size of the THOR-H I data is similar to the one of the GRS survey (40″ and 46″, respectively), we chose to keep the original resolution (both spatially and spectrally) when comparing the kinematic maps. We tested smoothing the H I maps to the common beam size of 46″, which had a negligible effect.

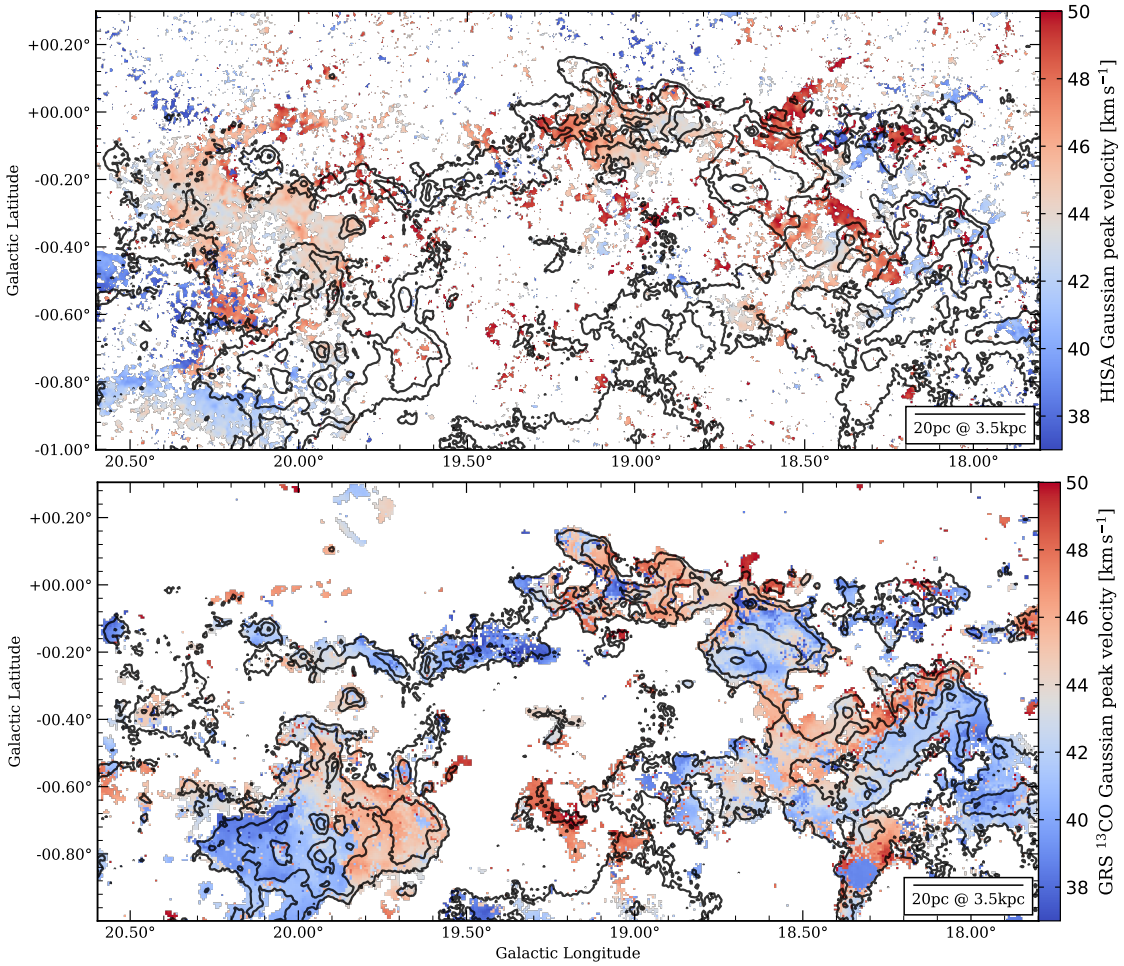


Figure 5.3: Fit peak velocity toward GMF20. These maps show the peak velocities of fit components derived from the GaussPy+ decomposition of the spectra. If multiple components are present in a single pixel spectrum within the velocity range of the filament region, the component with the lowest peak velocity is shown. The black contours in both panels show the integrated GRS ^{13}CO emission at the levels 8.0, 16.0, 32.0, and 42.0 K km s^{-1}. *Top panel:* Fit HISA peak velocity. *Bottom panel:* Fit ^{13}CO peak velocity.

For each of the kinematic histograms, we show every fit component along each line of sight within the velocity range of each filament region, thus taking into account multiple components if present. We furthermore only report fit components with an amplitude $> 5\sigma$ of the respective data cube. The histograms of the centroid velocities of HISA and ^{13}CO show correlation for most of the filament regions (Fig. 5.4). The median peak velocity toward GMF20 is $v_{\rm LSR} = 44.7\,{\rm km\,s^{-1}}$ for HISA and 44.0 km s^{-1} for ^{13}CO, which is in very good agreement with the results obtained in Syed et al. (2020). Particularly in the case of HISA, the histogram is mildly affected by components at higher velocities that might not be associated with the giant filament region or that might be troughs between two emission features erroneously picked up by the astroSaber routine. This effect is also evident in the histogram of GMF26. However, the median peak velocities also do agree toward GMF26, with $v_{\rm LSR} = 44.9\,{\rm km\,s^{-1}}$ and 45.4 km s^{-1} for HISA and ^{13}CO, respectively. Toward GMF38a the histogram of peak velocities obtained with both astroSaber

and the automated spectral line decomposition GaussPy+ reproduces the results presented in Wang et al. (2020c), with the median peak velocities agreeing to within 0.5 km s^{-1} ($v_{\rm LSR}$ = 54.3 km s^{-1} and 54.8 km s^{-1} for HISA and ^{13}CO, respectively). The median HISA peak velocity of $v_{\rm LSR}$ = 44.6 km s^{-1} toward GMF38b agrees with the ^{13}CO velocity of $v_{\rm LSR}$ = 44.4 km s^{-1} within the uncertainty of our HISA extraction method. However, we caution that this agreement might be the result of a selection bias that only takes into account velocities in a rather narrow range. Since the GMF38b filament region is identified in the narrow velocity range between 43.0 km s^{-1} and 46.0 km s^{-1}, it is clear that the selection of velocities will show a smaller deviation between the two tracers. Toward GMF41 and GMF54 there is a more pronounced difference in median peak velocity. Within the GMF41 filament region, the median velocity traced by HISA is $v_{\rm LSR}$ = 38.0 km s^{-1} while the median ^{13}CO velocity is $v_{\rm LSR}$ = 39.0 km s^{-1}. The median peak velocities toward GMF54 are $v_{\rm LSR}$ = 24.2 km s^{-1} and 23.1 km s^{-1} for HISA and ^{13}CO, respectively. We show the

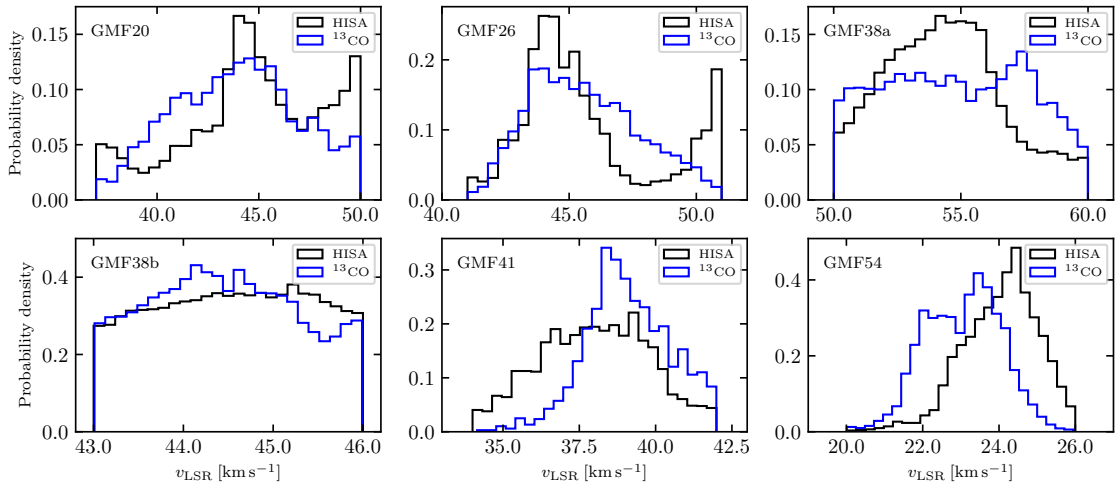


Figure 5.4: Histograms of peak velocities. The panels show for each of the six giant filament regions the normalized histogram of peak velocities of HISA and ^{13}CO in black and blue, respectively.

histograms of line width in Fig. 5.5. The peaks of the line width distributions are well resolved and above the spectral resolution limit, so the spectral resolution does not heavily affect the statistics of the kinematics. We find in general higher observed line widths in HISA than in ^{13}CO. The ^{13}CO line widths are 1.3–3.0 km s^{-1} while the HISA line widths are between 3.1 km s^{-1} and 5.2 km s^{-1}. The kinematic properties of the clouds are also summarized in Table 5.3. Assuming a kinetic temperature, we can estimate the expected thermal line width. In local thermodynamic equilibrium (LTE), the thermal line width (FWHM) is given by $\Delta v_{\rm th} = \sqrt{8 \ln 2 \, k_{\rm B} T_{\rm k} / (\mu m_{\rm H})}$, where $k_{\rm B}$, $T_{\rm k}$, and μ are the Boltzmann constant, kinetic temperature, and the mean molecular weight of H I ($\mu_{\rm H\,I}$ = 1.27) and the CO molecule ($\mu_{\rm CO}$ = 2.34; Allen, 1973; Cox, 2000) in terms of the mass of a hydrogen atom $m_{\rm H}$, respectively. The kinetic temperature can be well approximated by the estimated spin or excitation temperature of HISA and ^{13}CO, given the low temperatures and high densities of the cold gas (see Sects. 5.3.2 and 5.3.2). If different line broadening effects are

Table 5.3: Kinematic properties of the giant filament regions.

	HISA			^{13}CO		
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Source name	$\langle v \rangle$ [km s^{-1}]	$\langle \Delta v \rangle$ [km s^{-1}]	$\langle \mathcal{M} \rangle$	$\langle v \rangle$ [km s^{-1}]	$\langle \Delta v \rangle$ [km s^{-1}]	$\langle \mathcal{M} \rangle$
GMF20.0-17.9	44.7	3.9	4.8	44.0	3.0	9.5
GMF26.7-25.4	44.9	3.2	3.7	45.4	2.1	7.9
GMF38.1-32.4a	54.3	3.8	4.1	54.8	2.5	9.0
GMF38.1-32.4b	44.6	5.2	6.0	44.4	2.4	7.0
GMF41.0-41.3	38.0	3.6	3.7	39.0	2.5	8.2
GMF54.0-52.0	24.2	3.1	3.7	23.1	1.3	4.8

Notes. Columns (2) and (5) give the median peak velocity as traced by HISA and ^{13}CO for all six filament regions, respectively. Similarly, columns (3) and (6) present the median line width as traced by HISA and ^{13}CO, respectively. Columns (4) and (7) give the median sonic Mach number of HISA and ^{13}CO, respectively, which are computed in Sect. 5.4.2 using the sound speed at the temperatures estimated in Sects. 5.3.2 and 5.3.2.

uncorrelated, the total observed line width will be

$$\Delta v_{\rm obs} = \sqrt{\Delta v_{\rm th}^2 + \Delta v_{\rm nth}^2 + \Delta v_{\rm res}^2}\,, \tag{5.3.1}$$

where $\Delta v_{\rm nth}$ is the line width due to nonthermal effects and $\Delta v_{\rm res}$ is the line width introduced by the spectral resolution of the data. The thermal line widths are on the order of $\sim$0.5 km s^{-1} for ^{13}CO and $\sim$1.0 km s^{-1} for HISA at the given temperatures. The observed line widths of both HISA and ^{13}CO therefore show that the line widths cannot be explained by thermal broadening alone. Nonthermal effects such as turbulent motions have a significant effect on the observed line widths and are most likely the dominant driver for the broadening of the lines. We investigate the turbulent Mach number of the gas in Sect. 5.4.2.

5.3.2 Column density and mass

In this section, we compute the column densities toward each filament region using HISA, ^{13}CO, and H I as a tracer for the CNM, molecular gas, and bulk atomic gas, respectively. For the derivation of the column density maps we integrated each filament region over the velocity range given in Table 5.1. The column density maps of each tracer can be found in Appendix B.5.

CNM column density traced by HISA

Following the derivation given in Gibson et al. (2000), we compute the optical depth of HISA as

$$\tau_{\rm HISA} = -\ln\left(1 - \frac{T_{\rm on} - T_{\rm off}}{T_{\rm HISA} - p_{\rm bg} T_{\rm off} - T_{\rm cont}}\right)\,, \tag{5.3.2}$$

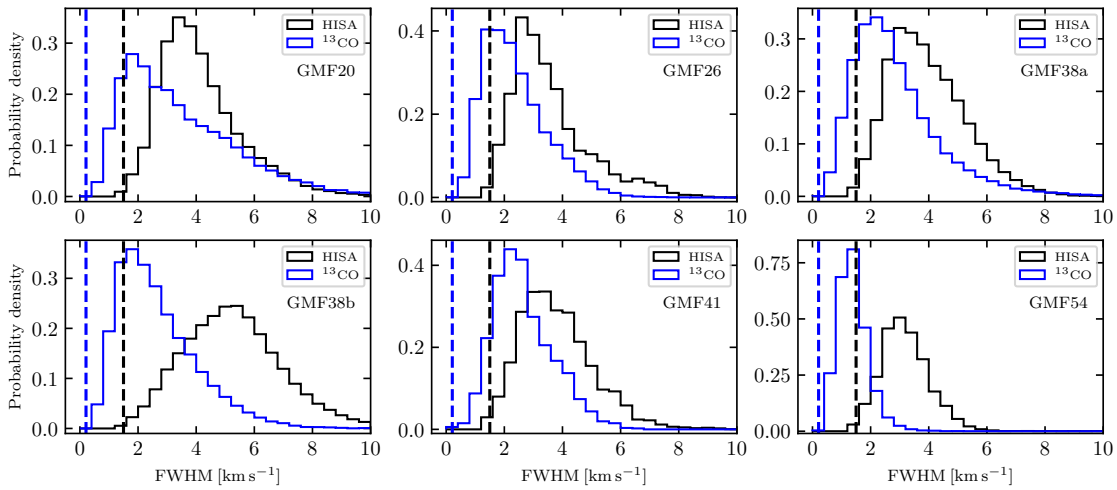


Figure 5.5: Histograms of line widths. The panels show for each of the six giant filament regions the normalized histogram of FWHM line widths of HISA and ^{13}CO in black and blue, respectively. The black and blue vertical dashed lines mark the spectral resolution limit of the HISA and ^{13}CO data, respectively.

with the dimensionless parameter $p_{\rm bg} \equiv T_{\rm bg}\,(1 - e^{-\tau_{\rm bg}})\,/T_{\rm off}$ describing the fraction of background emission in the optically thin limit (Feldt, 1993). Assuming a HISA spin temperature $T_{\rm s}$ ($= T_{\rm HISA}$), we can then calculate the H I column density of the cold H I gas using the general form (Wilson et al., 2013)

$$\frac{N_{\rm H}}{\rm cm^{-2}} = 1.8224 \times 10^{18}\,\frac{T_{\rm s}}{\rm K}\int \tau\,(T_{\rm s}, v)\,\left(\frac{{\rm d}v}{\rm km\,s^{-1}}\right)\,, \tag{5.3.3}$$

where $T_{\rm s}$ is the spin temperature of atomic hydrogen and $\tau\,(T_{\rm s}, v)$ describes the optical depth. We estimate the column density uncertainty by setting $T_{\rm on} = T_{\rm off} - \Delta T$ in Eq. (2.2.5) as the limit at which we can detect H I self-absorption, where ΔT is the rms noise in emission-free channels.

We can estimate the amount of background emission from the radial H I volume density distribution in the Galactic plane. For Galactocentric radii $7 \lesssim D_{\rm GC} \lesssim 35\,{\rm kpc}$ Kalberla & Dedes (2008) report an average mid-plane volume density distribution of $n(R) \sim n_0\,e^{-(D_{\rm GC} - D_\odot)/D_n}$ with $n_0 = 0.9\,{\rm cm^{-3}}$, $D_\odot = 8.5\,{\rm kpc}$, and with the radial scale length $D_n = 3.15\,{\rm kpc}$ (IAU recommendations). We assume a constant volume density distribution in the inner Galaxy of $n(D_{\rm GC} < 7\,{\rm kpc}) = n(D_{\rm GC} = 7\,{\rm kpc})$ as the volume density distribution is flattening off at lower Galactocentric distances (see Fig. 5 in Kalberla & Dedes, 2008). This relation gives the averaged distribution of the northern and southern Galactic plane and could hold systematic differences in some regions (Kalberla & Kerp, 2009).

In principle, only atomic gas located at a distance that corresponds to the radial velocity of HISA is relevant to the estimation of the background. Since we usually place any observed HISA at the near kinematic distance, most of the emission must stem from the background and thus move $p_{\rm bg}$ close to 1 at any given finite spectral resolution element. However, due to the velocity dispersion of H I, in particular the WNM, atomic hydrogen emission that has radial velocities slightly offset from the HISA velocity can blend into the velocity channels under consideration. Atomic

hydrogen emission in the foreground or background that corresponds to radial velocities around HISA can therefore contribute to the observed feature and affect the optical depth computation. We estimate the background fraction using the volume densities in the kinematic near and far distance intervals Δd corresponding to radial velocity intervals around the mean velocities of the clouds. The length of the distance interval is estimated from an average velocity dispersion of $\sim$10 km s^{-1} that falls between typical CNM and WNM velocity dispersions found in the Milky Way (Heiles & Troland, 2003; Haud & Kalberla, 2007). Equal steps in radial velocity $v_{\rm LSR}$ will map into unequal steps in distance that are proportional to the inverse of the velocity gradient along the line of sight. This gives distance intervals between $\sim$0.6 kpc and $\sim$0.9 kpc for the considered clouds. The gas density is then integrated from $d_{\rm near} - \Delta d$ to $d_{\rm near}$ to obtain a foreground fraction of the emission at the velocity of HISA. The background gas fraction is inferred by adding the integrated gas density in the interval $[d_{\rm near}, d_{\rm near} + \Delta d]$ to the gas density integrated on the kinematic far side interval $[d_{\rm far} - \Delta d, d_{\rm far} + \Delta d]$. The derived background fractions are between 0.75 and 0.77. We do note that while considering many factors in the treatment of the background fraction, the uncertainties are substantial due to non-circular and streaming motions superposed with the Galactic rotation, or systematic differences in the density distribution of H I. This adds a considerable source of uncertainty to the column density derivation and will be revisited in Sect. 5.3.3. Since we expect most of the emission background to originate in more diffuse H I gas, we assume a constant $p_{\rm bg}$ for each filament region (see Table 5.4).

Depending on the assumed spin temperature $T_{\rm s}$ and background fraction $p_{\rm bg}$, there might be no solution to Eq. (2.2.5) in some velocity channels of the spectra if the spin temperature is too high. Disregarding these channels in the line-of-sight integration, the column density computed in Eq. (5.3.3) underestimates the true column density toward some regions. To resolve this, we derived the maximum spin temperature limit by $T_{\rm max} = T_{\rm on} + T_{\rm cont} - (1 - p_{\rm bg}) T_{\rm off}$ for $\tau \to \infty$ (see Eq. 5.3.2), and set the 0.1-percentile of the maximum spin temperature to be the spin temperature of the whole cloud, such that Eq. (2.2.5) gives a solution for 99.9% of pixels in the integrated column density map. The assumed spin temperatures $T_{\rm s}$ and estimated background fractions $p_{\rm bg}$ are shown in Table 5.4 for each GMF source. Since a lower assumed spin temperature (at constant background fraction) producing the same observed HISA feature results in a lower column density (see Eqs. 5.3.2 and 5.3.3), the total column density will yet again be underestimated. This shortcoming of the HISA column density computation is addressed in the HISA simulations conducted by Seifried et al. (2022). However, assuming a constant spin temperature for the entire cloud appears to be the best approach to qualitatively recover the true column density structure of the cloud (Seifried et al., 2022).

H_2 column density traced by ^{13}CO emission

Assuming that ^{12}CO is becoming optically thick toward these GMFs, we estimated the column density of molecular hydrogen from ^{13}CO emission. In the optically thin

Table 5.4: Assumed spin temperatures and background fractions.

Source	T_s [K]	background fraction p_{bg}
GMF20	26	0.75
GMF26	27	0.75
GMF38a	32	0.75
GMF38b	33	0.75
GMF41	37	0.76
GMF54	24	0.77

Notes. The second column gives the spin temperature T_s assumed toward each GMF region. The third column gives the background fraction p_{bg} that is estimated from the ratio of foreground and background column density along the line of sight (see Sect. 5.3.2).

limit, the ^{13}CO column density is computed by (Wilson et al., 2013)

$$N(^{13}\mathrm{CO}) = 3.0 \times 10^{14} \frac{\int T_\mathrm{B}(v)\,\mathrm{d}v}{1 - \exp(-5.3/T_\mathrm{ex})}\,, \tag{5.3.4}$$

where $N(^{13}\mathrm{CO})$ is the column density of ^{13}CO molecules in cm^{-2}, dv is in units of km s^{-1}, T_B and T_ex are the brightness temperature and excitation temperature of the ^{13}CO line in units of Kelvin, respectively. Under the assumption that the excitation temperature T_ex of ^{12}CO and ^{13}CO are the same in LTE, we computed the ^{13}CO excitation temperature from ^{12}CO line emission. Both the ^{12}CO and ^{13}CO data are taken from the high-resolution survey MWISP (Su et al., 2019) to compute the excitation temperatures and column densities, as described in Sect. 5.2.1. The excitation temperature is computed as (Wilson et al., 2013)

$$T_\mathrm{ex} = 5.5 \cdot \left[\ln\left(1 + \frac{5.5}{T_\mathrm{B}^{12} + 0.82}\right)\right]^{-1}\,, \tag{5.3.5}$$

where T_B^{12} is the brightness temperature of the ^{12}CO line in units of Kelvin. To calculate the excitation temperature for each voxel, we reprojected the ^{12}CO data cubes onto the same spectral grid as the ^{13}CO data.

We set a lower limit to the excitation temperatures for regions where the ^{12}CO brightness temperatures reach the 5σ noise level. We can then derive the optical depth of the ^{13}CO line from the excitation and brightness temperature, using (see e.g., Wilson et al., 2013; Schneider et al., 2016)

$$\tau = -\ln\left[1 - \frac{T_\mathrm{B}}{5.3} \cdot \left(\left[\exp\left(\frac{5.3}{T_\mathrm{ex}}\right) - 1\right]^{-1} - 0.16\right)^{-1}\right]\,. \tag{5.3.6}$$

We then estimated a lower limit of the optical depth for ^{13}CO brightness temperatures at the 5σ noise level and the highest excitation temperatures we find toward each GMF region. The lower and upper limits to the excitation temperatures as well

Table 5.5: Limits of CO excitation temperatures and optical depths.

Source	$T_{\rm ex,low}$ [K]	$T_{\rm ex,up}$ [K]	$\tau_{\rm low}$	$X([{\rm H_2}]/[^{13}{\rm CO}])$
GMF20	5	29	0.05	3.1×10^5
GMF26	5	16	0.11	3.6×10^5
GMF38a	5	21	0.08	3.9×10^5
GMF38b	5	18	0.10	4.1×10^5
GMF41	5	12	0.26	4.7×10^5
GMF54	5	36	0.04	5.6×10^5

Notes. The second and third column gives the lower limit and upper limit of the CO excitation temperature, respectively. The fourth column shows the lower limit of the optical depth estimated from the 5σ ^{13}CO noise and the highest excitation temperature found toward each source. The last column gives the ^{13}CO-to-H_2 conversion factor that we have used for each source.

as the lower limits of the optical depth are listed in Table 5.5 for each source. To account for high optical depth effects, we employ a correction factor by replacing the integral in Eq. (5.3.4) with (Frerking et al., 1982; Goldsmith & Langer, 1999)

$$\int T_B(v)\,{\rm d}v \rightarrow \frac{\tau}{1-e^{-\tau}} \int T_B(v)\,{\rm d}v\,. \tag{5.3.7}$$

This correction factor is accurate to 15% for $\tau < 2$. We used Galactocentric distance-dependent abundance relations to estimate an $[{\rm H_2}]/[^{13}{\rm CO}]$ conversion factor for each source. Giannetti et al. (2014) give a ^{12}CO-to-^{13}CO abundance relation of $[^{12}{\rm CO}]/[^{13}{\rm CO}] = 6.2 D_{\rm GC} + 9.0$, and the H_2-to-^{12}CO abundance given in Fontani et al. (2012) is $[{\rm H_2}]/[^{12}{\rm CO}] = [8.5 \times 10^{-5} \exp(1.105 - 0.13 D_{\rm GC})]^{-1}$, where $D_{\rm GC}$ is the Galactocentric distance in units of kpc. We estimate the uncertainty in H_2 column density to be at least 50% due to the large uncertainties in these relations. Furthermore, CO might not always be a good tracer of H_2 as "CO-dark H_2" could account for a significant fraction of the total H_2 (Pineda et al., 2008; Goodman et al., 2009; Pineda et al., 2013; Smith et al., 2014a; Tang et al., 2016), particularly at low column densities and early evolutionary stages as molecular clouds might not have become CO-bright yet (Goldsmith et al., 2008; Planck Collaboration et al., 2011).

Atomic gas column density seen in H I emission

In addition to the cold atomic gas traced by HISA, we investigated the properties of the total atomic hydrogen gas budget (WNM+CNM) by measuring the column density from H I emission and correcting for optical depth effects and diffuse continuum. As the optically thin assumption might not hold for some regions, we can utilize strong continuum emission sources to directly measure the optical depth. H I continuum absorption (HICA) is a classical method to derive the properties of the CNM (e.g., Strasser & Taylor, 2004; Heiles & Troland, 2003). This method uses strong continuum sources, such as Galactic H II regions or active galactic nuclei (AGNs), to measure the optical depth of H I. As these sources have brightness

temperatures that are larger than typical spin temperatures of cold H I clouds, we observe the H I cloud in absorption. The absorption feature is furthermore dominated by the CNM since the absorption is proportional to T_s^{-1} (e.g., Wilson et al., 2013).

The advantage of this method is the direct measurement of the optical depth. However, the HICA method requires strong continuum emission sources. As most strong continuum sources are discrete point sources, this method results in an incomplete census of optical depth measurements. However, Wang et al. (2020a) derived a velocity-resolved optical depth map computed from 228 continuum sources within the THOR survey that are above a 6σ noise threshold and interpolated the measurements using a nearest-neighbor method. For more details about the optical depth measurement we refer to Wang et al. (2020a). To the first approximation, we can use this optical depth map to correct the H I column density as confirmed in Syed et al. (2020), in spite of potential kinematic distance ambiguities and the location of a continuum source along each line of sight that might add or miss optical depth for each line-of-sight velocity, respectively. For each velocity channel, we take the spatial average of the optical depth map measured toward each filament region in order to avoid artifacts introduced by the interpolation.

Besides strong continuum sources we observe weak continuum emission throughout the Galactic plane. This component has brightness temperatures between 10 and 50 K. The continuum emission has been subtracted during data reduction as described in Sect. 5.2.1. As even weak continuum emission might suppress H I emission and therefore lead to an underestimate of the column density, we account for the weak emission component when computing the H I column density (see Bihr et al., 2015, Eq. 9). We estimate the column density and mass uncertainty by varying the optical depth by 10%, which roughly corresponds to the 1σ brightness variation of our weakest continuum sources.

5.3.3 Masses

Based on the column density estimates in the previous sections, we can directly estimate the (cold) atomic and molecular mass toward the filament regions (see Table 5.6). We compute the masses by summing up the mass pixels above a column density threshold corresponding to significant emission or H I self-absorption. These thresholds are then also used to derive column density PDFs (see Sect. 5.4.1). The change in mass that comes with varying thresholds is relatively small compared to the uncertainties of the column density derivation itself.

The CNM mass traced by HISA corresponds to 3–9% of the total atomic gas mass, depending on the region and assumed spin temperature. The HISA mass fraction toward GMF38b is 0.09 and exceeds that found in all other filament regions. We recovered column density regions off the main cloud that is defined as GMF38.1-32.4b (see Fig. B.18). These regions might not be tightly associated with the molecular gas that defines the GMF. The HISA mass fraction reduces to 3–4% if we only take into account the gas in close proximity to the main molecular feature of the cloud (gas beyond the lowest contour in Fig. B.18 to within 0.2°), thus being comparable to other filament regions. However, this example also illustrates that

we recover cold atomic gas structures that do not have a molecular counterpart. The cold phase of the atomic ISM appears to be much more widespread than the molecular gas in Fig. B.18. The masses of both GMF20 and GMF38a are similar to the masses found by Syed et al. (2020) and Wang et al. (2020c). Given that we assume a spin temperature of 26 K for GMF20 (instead of 20 K and 40 K in Syed et al. 2020), the derived mass falls within the mass range 4.6×10^3–$1.3 \times 10^4\,M_\odot$ obtained in Syed et al. (2020).

In general, the uncertainty in column density directly translates to an uncertainty in mass. If the spin temperature is varied by 10 K, the column density and mass traced by HISA changes by a factor of ∼2. Hence, the largest uncertainty arises from the assumption of a spin temperature and the background fraction that is coupled to the optical depth computation. Independent of the assumed spin temperature, the inferred column density will in most cases be underestimated, as shown in Seifried et al. (2022). Even for an assumed spin temperature that comes close to the limit at which the optical depth computation gives an analytic solution (see Eq. 5.3.2), the column density will still be underestimated due to line-of-sight variations in spin temperature and observational noise. By assuming an optically thick HISA cloud with $\tau \to \infty$, we are able to determine the spin temperature limit above which the line-of-sight geometry does not allow the computation of the column density. The uncertainty in column density and mass is further amplified by the background fraction $p_{\rm bg}$ in Eq. (2.2.5). If the background fraction is lowered, the column density will increase as the cold H I cloud would be more efficient in producing the same observed HISA feature given the weaker background. A moderate variation in the background fraction of 10% at fixed spin temperature results in a ∼30% change in column density and therefore mass. We derive a HISA mass uncertainty by varying the background fraction by 10% and adjusting the spin temperature accordingly, such that we have again a solution for most pixels in the map. We report these uncertainties in Table 5.6 as well. For a more detailed discussion of these uncertainties we refer to Wang et al. (2020c) and Syed et al. (2020).

These uncertainties add to the intrinsic systematic effects of the HISA method. We are generally limited by the background emission that enables the observation of HISA. Furthermore, HISA is only sensitive to gas that is colder than the gas that contributes to the emission background. The CNM is reported to have spin temperatures up to ∼300 K (e.g., Heiles & Troland, 2003; Kalberla & Kerp, 2009), rendering the HISA detection of the CNM in many cases impossible given the observed brightness temperatures. According to the simulations conducted by Seifried et al. (2022), the HISA-traced mass underestimates the mass of the CNM that could in principle be observed through HISA by a factor of 3–10. This underestimation is generally attributed to two effects. The proper estimation of the spin temperature that is required to compute the HISA properties is a challenging task because of its variation within a cold H I cloud. Due to the line-of-sight geometry, an assumed H I spin temperature that is too low will result in an underestimate of the optical depth and the true column density (see Eqs. 5.3.2 and 5.3.3). An H I spin temperature that is too high will cause the HISA-traced cloud to have no solution to the optical depth at least for some part of the spectrum (Eq. 5.3.2). This will again underestimate the integrated column density as individual spectral channels are omitted. Varying

the spin temperature along the line-of-sight or spatially can lead to an even larger deviation and might recover a column density structure that does not reflect the true distribution qualitatively. The challenges of unknown spin temperature consequently give rise to a large uncertainty in the determination of the column density and mass.

The molecular hydrogen mass is on the order of 10^4–10^5 $M_\odot$. The total atomic gas fraction shows large differences among the filament regions. The atomic gas mass is generally comparable to the molecular gas mass. However, for GMF54 the atomic gas seen in H I emission and HISA accounts to a total that is just one quarter of the total hydrogen mass. With respect to the molecular gas phase, the total atomic gas fraction is found to increase with Galactocentric distance on average (e.g., Nakanishi & Sofue, 2016; Miville-Deschênes et al., 2017). In spite of having the largest Galactocentric distance in our sample, GMF54 appears to have used up much of the atomic gas in which it was embedded to transition into a more complete molecular gas phase.

Table 5.6: Derived masses of the filament regions.

(1)	(2)	(3)	(4)	(5)	(6)
	M(HISA) [$M_\odot$]	M(H I) [$M_\odot$]	$M(H_2)$ [$M_\odot$]	$f_{\rm HISA}$	$f_{\rm atomic}$
GMF20.0-17.9	$7.5^{+1.3}_{-1.9} \times 10^3$	$2.7^{+0.2}_{-0.3} \times 10^5$	$2.3 \pm 1.2 \times 10^5$	0.03	0.55
GMF26.7-25.4	$1.1^{+0.3}_{-0.2} \times 10^3$	$3.0^{+0.3}_{-0.3} \times 10^4$	$5.5 \pm 2.8 \times 10^4$	0.04	0.36
GMF38.1-32.4a	$1.1^{+0.2}_{-0.3} \times 10^4$	$3.6^{+0.5}_{-0.7} \times 10^5$	$3.0 \pm 1.5 \times 10^5$	0.03	0.55
GMF38.1-32.4b	$6.2^{+1.3}_{-1.6} \times 10^3$	$6.2^{+0.7}_{-0.7} \times 10^4$	$5.5 \pm 2.8 \times 10^4$	0.09	0.55
GMF41.0-41.3	$5.3^{+0.7}_{-1.1} \times 10^2$	$1.4^{+0.2}_{-0.3} \times 10^4$	$1.7 \pm 0.9 \times 10^4$	0.04	0.46
GMF54.0-52.0	$3.3^{+0.7}_{-1.0} \times 10^2$	$5.2^{+1.1}_{-0.9} \times 10^3$	$1.6 \pm 0.8 \times 10^4$	0.06	0.26

Notes. The masses were calculated from the column density maps shown in Appendix B.5. Column (2) presents the mass of the cold atomic hydrogen traced by HISA. The uncertainties are statistical errors arising from the uncertainties in background fraction and spin temperature and do not include the systematic uncertainties due to the detection method. Column (3) shows the atomic hydrogen mass inferred from the optical depth and continuum corrected H I emission. The uncertainties are estimated from variations in the optical depth measurement. Column (4) gives the molecular hydrogen mass as traced by ^{13}CO emission along with a conservative 50% uncertainty owing to the large uncertainties in the CO-H_2 conversion. Column (5) gives the mass traced by HISA as a fraction of the total atomic gas mass M(HISA) + M(H I). Column (6) is the fraction of the total atomic gas mass traced by HISA and H I emission with respect to the total gas mass M(HISA) + M(H I) + $M(H_2)$.

5.4 Discussion

5.4.1 Column density PDF

We employ the probability density function (PDF) of the column density to investigate the physical processes acting within the filament regions. The shape of column

or volume density PDFs are commonly used as a means to describe the underlying physical mechanisms of a cloud (e.g., Federrath & Klessen, 2013; Padoan et al., 2014; Kainulainen et al., 2014; Schneider et al., 2015, 2022). Turbulence is considered to be the dominant driver of a cloud's structure if its PDF shows a log-normal shape. Furthermore, the width of a log-normal PDF is linked to the Mach number as it changes with the magnitude of the turbulence driving the cloud's structure (e.g., Padoan et al., 1997; Passot & Vázquez-Semadeni, 1998; Padoan & Nordlund, 2002; Kritsuk et al., 2007; Federrath et al., 2008; Konstandin et al., 2012; Molina et al., 2012; Kainulainen et al., 2014; Beattie et al., 2022), while noting that the turbulence driving scale and CNM-WNM mass ratio also affect the width of the PDF (Bialy et al., 2017b).

Molecular clouds that are subject to the increasing effect of self-gravity develop high-density regions, producing a power-law tail in their PDF (e.g., Klessen, 2000; Girichidis et al., 2014; Burkhart et al., 2017). Many star-forming molecular clouds have been confirmed to show such power-law tails (Kainulainen et al., 2009; Schneider et al., 2013, 2016, 2022). Even before the effects of gravity become dominant, gravitationally unbound clumps can exhibit power-law tails due to pressure confinement from the surrounding medium (Kainulainen et al., 2011).

We show in Fig. 5.6 the column density PDFs (N-PDFs) of all filament regions as traced by H I emission, HISA, and ^{13}CO. We take into account only column densities above the noise threshold and find that the widths of each N-PDF do not change significantly when considering higher thresholds. We fit all column density PDFs with a log-normal function and report their widths in Fig. 5.6. Since we use a consistent way in deriving the PDFs, systematic differences between the distributions should be small, such that they can be well compared in relative terms.

Toward most filaments, the HISA-traced cold atomic gas shows a column density distribution that clearly bridges the gap between the narrow distribution of the diffuse atomic gas (left panels of Fig. 5.6) and the broad molecular gas distribution (right panels) in terms of their width. The mean column densities of molecular hydrogen are at least an order of magnitude higher than the column densities traced by HISA. While the N-PDFs of the molecular gas can be described well by broad log-normals, we additionally fit power-law functions ($p(x) \propto x^{-\alpha}$) to the high column density tail of the H_2 N-PDFs for comparison. The best minimal column density for the power-law fit is obtained from the minimal Kolmogorov-Smirnov distance between the fit and the N-PDF. The fits were performed using the python package *Powerlaw* (Alstott et al., 2014). The fit power-law exponents are also shown in Fig. 5.6. Toward GMF38b the column density threshold for the power-law fit was set to high column densities where there is a drop-off in the distribution, possibly due to optical depth effects. We note the narrow distributions in the molecular gas phase toward GMF26 and GMF41. This is related to the low excitation temperatures we find toward these clouds. This might be an indication of an early evolutionary stage where gravity has not yet become dynamically important. This is further supported by the low number of YSOs identified toward GMF41 (see Zhang et al., 2019).

The narrow log-normal shaped N-PDFs are commonly observed in the diffuse H I emission toward well-known molecular clouds (Burkhart et al., 2015; Imara & Burkhart, 2016; Rebolledo et al., 2017; Schneider et al., 2022). The HISA N-PDFs

that trace the CNM show broader distributions, indicative of the clumpy structure and higher degree of turbulence. Considering the column density PDFs, HISA appears to trace the cold atomic gas phase that connects the diffuse state of the atomic ISM with the transition of a cloud becoming molecular.

5.4.2 Mach number distribution

In the following, we derive the turbulent Mach number distributions using the constant HISA spin temperatures given in Table 5.4 and the excitation temperatures of ^{13}CO derived in Sect. 5.3.2. Given the low temperature regime, we approximate the kinetic gas temperatures of the CNM and the molecular gas with the spin temperature and mean excitation temperature, respectively. We then estimate the three-dimensional scale-dependent Mach number of the filaments assuming isotropic turbulence with $\mathcal{M} = \sqrt{3}\,\sigma_{\rm turb}/c_s$, where $\sigma_{\rm turb}$ and c_s are the turbulent one-dimensional velocity dispersion and sound speed, respectively. The turbulent line width is calculated by subtracting the thermal line width contribution from the observed line width as

$$\sigma_{\rm turb} = \sqrt{\sigma^2_{\rm obs} - \sigma^2_{\rm th} - \sigma^2_{\rm res}}\,, \tag{5.4.1}$$

where $\sigma_{\rm obs}$, and $\sigma_{\rm th}$ are the observed, and thermal velocity dispersion, respectively. For completeness, we also account for the broadening introduced by the spectral resolution $\sigma_{\rm res}$. Since the thermal line width and sound speed scale as $T_{\rm k}^{1/2}$, the variation with spin temperature is moderate and does not change the Mach number significantly. Seifried et al. (2022) showed that the Mach number estimate inferred through HISA is robust and can be determined with an accuracy within a factor of $\sim$2.

Five of the six filament regions show very similar Mach number distributions (Fig. 5.7). The Mach number distributions traced by HISA are generally much narrower than those traced by ^{13}CO emission, and peak round $\mathcal{M} \sim 3-6$, with few values as high as $\sim$10. Our findings are in very good agreement with recent HISA observations (Burkhart et al., 2015; Nguyen et al., 2019; Wang et al., 2020c; Syed et al., 2020) and the simulations conducted by Seifried et al. (2022).

With the exception of GMF54, the molecular gas is highly supersonic, and has median Mach numbers between $\mathcal{M} \sim 7-10$. The molecular gas toward GMF54 is moderately supersonic and has a median Mach number around $\sim$5. The total observed line widths are generally small with a few $\sim$km s^{-1} (see Fig. B.14) and we do find the highest excitation temperatures up to $\sim$35 K in GMF54. As the HISA Mach numbers are also smallest toward GMF54, we consider this an imprint of a different physical mechanism dominating the dynamics of the cloud. In combination with the high excitation temperatures, low atomic mass fraction, and the most pronounced power-law tail in its column density distribution that we find in our sample, GMF54 appears to be at a much more advanced stage in its evolution, at which gravity seems to be the dominant driver of the cloud's dynamics. Zhang et al. (2019) also find a star formation rate surface density that is among the highest in their sample of giant molecular filaments.

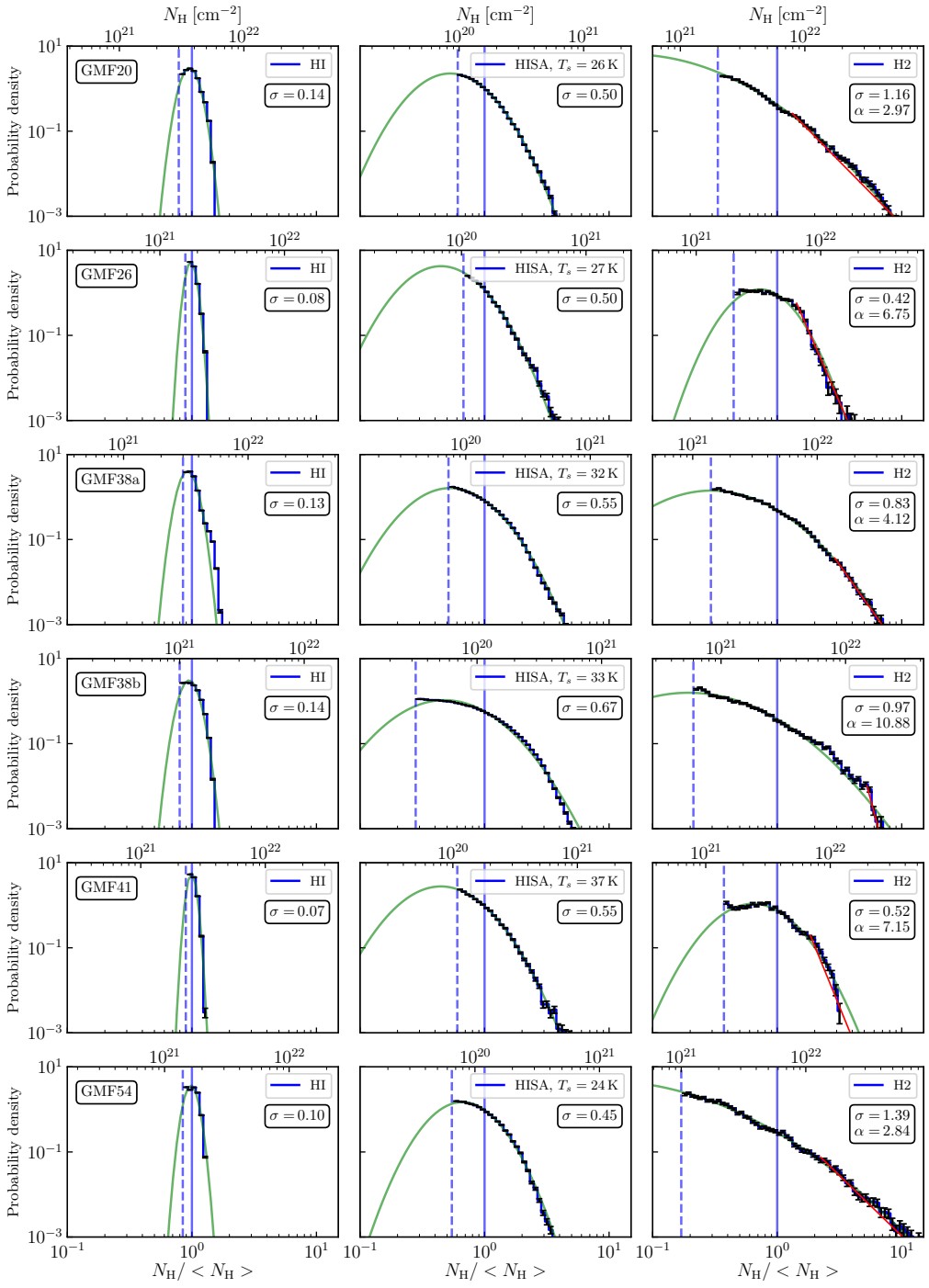


Figure 5.6: Normalized column density PDFs of H I, HISA, and H_2 toward the giant filament regions. Each row shows the H I, HISA, and H_2 N-PDF toward one GMF region. *Left panels:* N-PDFs traced by H I emission. The distributions are derived from the H I column densities that have been corrected for optical depth and continuum emission. *Middle panels:* The N-PDFs of the gas traced by HISA. *Right panels:* H_2 N-PDFs traced by ^{13}CO in units of hydrogen atoms per cm^{-2}. The green curves indicate a log-normal fit to the distributions. The blue vertical dashed and solid lines mark the column density threshold and mean column density, respectively. The red solid lines in the right column panels indicate a fit to the tail of the distributions.

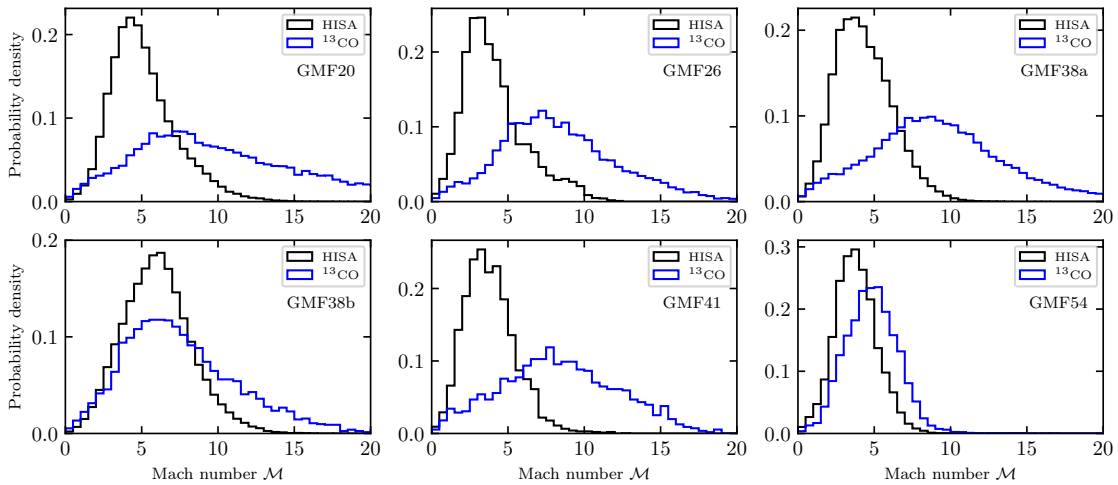


Figure 5.7: Histograms of Mach numbers. The panels show for each of the six filament regions the normalized histograms of the Mach numbers of HISA and ^{13}CO in black and blue, respectively.

5.4.3 Spatial correlation between CoAt and molecular gas

According to the classical idealized photodissociation region (PDR) picture, we would expect cold atomic gas to be spatially associated with its molecular counterpart (e.g., van Dishoeck & Black, 1988; Andersson et al., 1991). We therefore employed the Histogram of Oriented Gradients tool[4] (HOG; Soler et al., 2019) to investigate the spatial correlation between HISA and ^{13}CO emission. The HOG method is based on machine vision to examine the spatial correlation between two spectral line tracers across their spectral domain. A detailed description of the HOG is given in Soler et al. (2019).

The underlying principle is the computation of intensity gradients in each velocity channel map of the respective line tracer. The relative angles between the intensity gradients of the line tracers (here HISA and ^{13}CO) are then computed for each pair of velocity channel maps. To statistically evaluate the significance of spatial correlation in terms of relative orientation between intensity gradients, the HOG uses the projected Rayleigh statistic V as a metric, which is a test to determine if the distribution is non-uniform and centered around 0°. It is tuned such that the sign of V is indicative of the angle distribution having a peak around $\theta = 0°$ ($V > 0$) or $\theta = 90°$ ($V < 0$) (Jow et al., 2018). The absolute value of V indicates the significance of that preferred orientation in the angle distribution. The projected Rayleigh statistic is therefore

$$V = \frac{\sum_{ij}^{m,n} w_{ij} \cos(2\theta_{ij})}{\sqrt{\sum_{ij}^{m,n} w_{ij}/2}} \,, \tag{5.4.2}$$

where the indices i and j run over the pixel locations in the two spatial dimensions for a given velocity channel and w_{ij} is the statistical weight of each angle θ_{ij}. We account for the spatial correlation between pixels introduced by the telescope beam and set the statistical weights to $w_{ij} = (\delta x/\Delta)^2$, where δx is the pixel size and Δ

[4] https://github.com/solerjuan/astroHOG

is the diameter of the derivative kernel that we used to calculate the gradients. We set the derivative kernel to $\Delta = 92''$, which is twice the beam size of the GRS.

We smoothed the extracted HISA cubes to a common beam size of $46''$ and reprojected them onto the same spatial grid as the ^{13}CO data to run the HOG. Furthermore, we restricted the radial velocity range to $v_{\rm LSR,low} - 25\,{\rm km\,s^{-1}}$ and $v_{\rm LSR,up} + 25\,{\rm km\,s^{-1}}$ to save computational cost, where $v_{\rm LSR,low}$ and $v_{\rm LSR,up}$ are the lower and upper velocity limits given in Table 5.1, respectively. The extension of the velocity range $\pm 25\,{\rm km\,s^{-1}}$ provides a baseline measure of V (assuming there are signal-free channels over this velocity range). The projected Rayleigh statistic V should be $\sim$0 for these channels.

We use Monte Carlo sampling to propagate the errors introduced by the uncertainties in the flux measurement in each velocity channel (see e.g. Soler et al., 2020). For each velocity channel map, we generated 10 random realizations per tracer with the same mean intensity and observational noise. Using this sampling, the uncertainty of the correlation can be determined by the variance of the correlation of different Monte Carlo realizations. Since we expect a contribution from non-gaussian noise introduced by the observation, we report only $\geq 5\sigma$ confidence levels. We show in Fig. 5.8 the computed spatial correlation in terms of the pro-

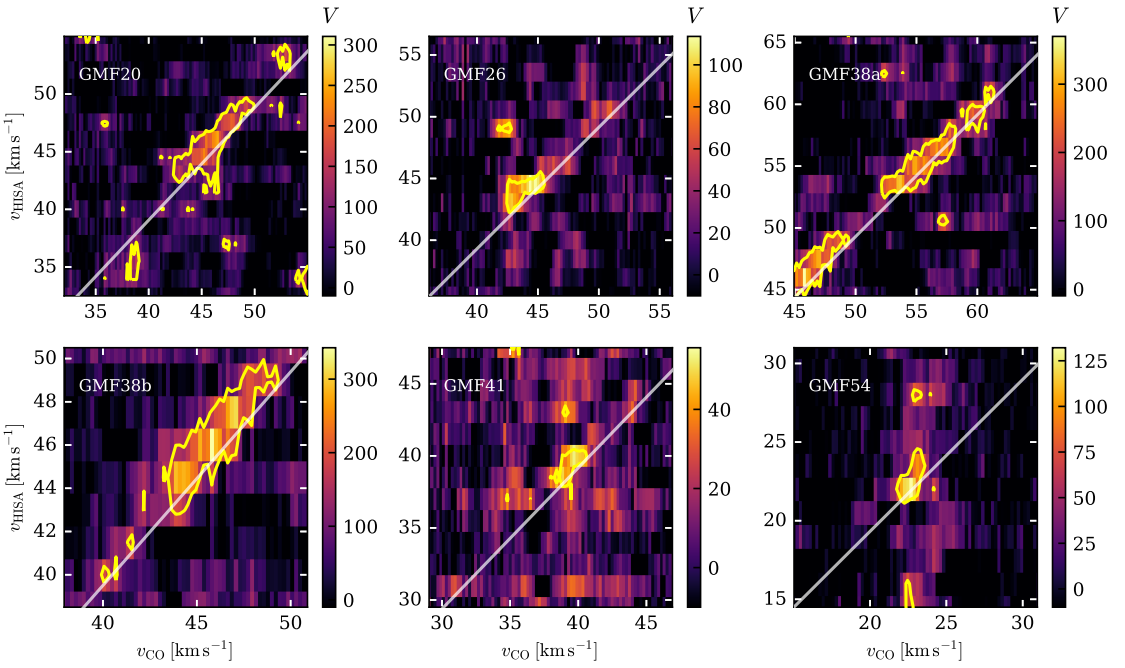


Figure 5.8: Correlation in the distribution of HISA and ^{13}CO emission toward the six GMF regions as quantified by the projected Rayleigh statistic (V) in the HOG method (Soler et al., 2019). The panels present the computed spatial correlation between HISA and ^{13}CO across velocities in terms of the projected Rayleigh statistic V for each filament region. The values of V are indicated by the color bar to the right of each panel. The white line in each panel shows the 1-to-1 correlation. The yellow contours show the 5σ threshold estimated from the Monte Carlo sampling. Large values of V indicate a high spatial correlation. Values of V close to zero indicate a random orientation of the HISA structures with respect to ^{13}CO emission.

jected Rayleigh statistic V for each filament region. We observe a strong spatial correlation between HISA and ^{13}CO toward GMF20, GMF38a, and GMF38b across

multiple velocity channels. Even toward GMF26, GMF41, and GMF54 we detect significant spatial correlation in few velocity channels, despite little HISA detection. We were able to reproduce the result obtained by Syed et al. (2020) toward GMF20, showing a significant spatial correlation within the entire filament region. Although we detect significant spatial correlation between HISA and ^{13}CO toward all filament regions in our sample, we note that there might be little to no correlation when investigating subsections of filament regions, as observed in Syed et al. (2020) where the western part of the filament showed a strong agreement between the spatial distribution of HISA and ^{13}CO while the eastern part entirely lacks correlation.

We conclude that the CNM traced by HISA generally appears to be associated with molecular gas in the giant filament regions on a large spatial scale. However, toward specific sub-regions within each filament systematic differences in spatial correlation can be evident that could be indicative of local events of star forming activity (see e.g. Soler et al., 2020, 2021).

5.4.4 The astroSaber method and physical implications

When dealing with finite spectral resolution, one of the shortcomings of classical approaches using finite-difference derivatives is the strong dependence on sensitivity and line width. Noise fluctuations are greatly amplified in second (or higher order) derivatives of a spectrum. Only HINSA with line widths $\lesssim$1 km s^{-1} might be identified using this approach. It is then often assumed that there is a tight physical correlation in temperature between the cold H I gas traced by self-absorption and the molecular gas within a cloud. This correlation is then used to constrain the baselines of the self-absorption features (see Krčo et al., 2008), which is a reasonable approximation given the projected spatial correlation and small line widths often observed toward the central regions of molecular clouds. However, the tight correlation observed through HINSA is likely to trace only the cold ($\sim$10 K) H I gas that is well mixed with the molecular gas in well-shielded regions (Li & Goldsmith, 2003; Goldsmith & Li, 2005), where the UV photo-dissociation rate of H_2 might become comparable to the cosmic ray dissociation in the central region of a cloud. By construction of the detection method, the CNM traced by HINSA likely results in atomic gas not being detected far beyond the inner regions of a molecular cloud (Goldsmith et al., 2007). However, once it is shown the HINSA-traced gas is coincident with ^{13}CO emission, the uncertainty in kinetic temperature should be considerably less than with our method.

With the newly developed algorithm astroSaber we identify H I self-absorption in an unbiased way, independent of the occurrence of molecular gas. The astroSaber algorithm can therefore complement the detection of CNM in the outer layers of molecular clouds or even the detection of CoAt gas that has no CO-bright molecular counterpart, which is likely to have larger line widths and would otherwise be missed by a second derivative approach, as we show in Appendix B.2.

In the following, we discuss some of the limitations and ways that might boost the performance of the astroSaber routine. Since the observed spectra also contain noise where there is signal, the baseline smoothing slightly overestimates the baselines within signal ranges as it also weights the noise asymmetrically that is superposed

with the emission. One way to take the noise within signal ranges into account is to adjust the weightings in Eq. (5.2.2) according to the mean and standard deviation of the positive and negative difference values between the spectrum and the baseline after each iteration (see e.g., Baek et al., 2015; Liu et al., 2022). However, we are only interested in ranges where we expect self-absorption to be present. As we tune the smoothing parameter such that significant dips in the spectra will be smoothed out, any variation of the obtained baselines within emission ranges without absorption should be limited to the noise. Any features in those ranges will therefore not be identified as significant absorption anyway.

As we show in Appendix B.3, the centroid velocities recovered by astroSABER and Gaussian fitting show little deviation from the input velocities within our test environment. The distribution of centroid velocity differences has a mean and standard deviation of $-0.01\,\mathrm{km\,s^{-1}}$ and $0.35\,\mathrm{km\,s^{-1}}$, respectively. Based on the findings by Wang et al. (2020c) and Syed et al. (2020), the input line widths of $\sim$4 $\mathrm{km\,s^{-1}}$ (FWHM) could be recovered with a standard deviation of $\sim$1 $\mathrm{km\,s^{-1}}$. The larger scatter in line widths is likely due to employing a constant smoothing parameter for both narrow and broad absorption components. The difference in amplitude shows the largest scatter around the mean as a single smoothing parameter is used for the entire region.

Since we set a constant $\boldsymbol{\lambda}$ value for all spectra in each field, we account for significant broad absorption features by performing multiple iterations to obtain their baselines. However, depending on the number of iterations, the final baseline might not reflect the original spectrum within emission ranges accurately as in each iteration an updated baseline is used as an input for the next major cycle iteration. One way to address this is to not have multiple major cycle iterations but instead adjust the smoothing parameter channel by channel as broader absorption features would require more iterations than narrow ones at constant $\boldsymbol{\lambda}$.

With a single iteration, broader absorption features require a larger smoothing parameter $\boldsymbol{\lambda}$ if the asymmetry weighting for negative differences (i.e. absorption dips) is constant but nonzero. This baseline "drag" because of nonzero weighting could be corrected for if we introduced another coefficient vector $\boldsymbol{\alpha}$ that adjusts the smoothing parameter in Eq. (5.2.1) for each channel in the spectrum, with its components being defined as

$$\alpha_i = \frac{\mathrm{abs}(y_i - z_i)}{\mathrm{max}(\mathrm{abs}(\mathbf{y} - \mathbf{z}))}\,, \tag{5.4.3}$$

where the numerator is the absolute difference of the spectrum and baseline at channel i, and the denominator is the maximum of the absolute differences in the spectrum (see e.g., Zhang et al., 2020). Equation (5.2.1) would then change to

$$F(\mathbf{z}) = (\mathbf{y} - \mathbf{z})^\top \mathbf{W} (\mathbf{y} - \mathbf{z}) + (\lambda\,\boldsymbol{\alpha})\,\mathbf{z}^\top \mathbf{D}^\top \mathbf{D}\mathbf{z}\,. \tag{5.4.4}$$

This could be a way to tune the smoothing parameter to an optimum without the need of having to perform multiple iterations. Since the weight curve and smoothing coefficients would also be fixed in that case, the smoothing parameter $\boldsymbol{\lambda}$ would still be the only parameter to be optimized.

In summary, the results could be improved by utilizing parameterized smoothing and asymmetry weights. Ultimately, these training and test data could then be used to feed a machine learning algorithm that sets an optimized smoothing parameter for each spectrum. However, as we have achieved good results with astroSaber in its current state, that already outperforms our traditional approach of using polynomial fits to specific ranges of the emission spectra, we leave the optimization of performance and efficiency to future investigations.

5.5 Conclusions

We have investigated the properties of the cold atomic gas and molecular gas toward a sample of six giant molecular filament regions. We traced the cold atomic gas phase by H I self-absorption and obtained these features using the newly developed baseline extraction algorithm astroSaber. The kinematic properties of both the cold atomic gas and molecular gas were obtained using the spectral decomposition tool GaussPy+ (Riener et al., 2019). The main results are summarized as follows:

1. We detect HISA toward all giant filament regions. The mass traced by HISA accounts to a few percent of the total atomic hydrogen mass traced by H I self-absorption and emission. The total atomic mass is in most cases comparable to the molecular mass. Deviations from these mass fractions can be linked to different evolutionary stages of the clouds.

2. The median centroid velocities of HISA and ^{13}CO are similar, and the well-resolved observed line widths of HISA are systematically larger than those of ^{13}CO. The CoAt gas traced by HISA is found to be moderately supersonic with Mach numbers of a $\sim$few, while the molecular gas within the majority of the filaments is driven by highly supersonic dynamics.

3. The derived column densities of the CoAt gas traced by HISA are on the order of $\sim 10^{20}$ cm^{-2} and the column density distributions of the CoAt gas can be well described by a log-normal. In terms of their width, the distributions clearly bridge the gap between the N-PDFs of the diffuse atomic gas traced by H I emission and the molecular gas.

4. The recovered HISA features show a spatial correlation with the molecular gas toward many regions within the filaments. The Histogram of Oriented Gradients analysis (Soler et al., 2019) confirms a significant spatial correlation between HISA and ^{13}CO toward all filament regions at similar velocities.

Probing the cold atomic gas by means of H I self-absorption toward molecular clouds is a powerful tool to investigate the dynamical and physical interplay between the atomic and molecular gas during cloud formation. While molecular clouds are ideal targets to investigate the properties of HISA, we are looking to extend upon our findings and identify HISA without the bias of corresponding molecular line emission. We will investigate the global distribution of HISA in the inner Galactic plane in an upcoming paper.

6

Cold atomic gas identified by H I self-absorption

Cold atomic clouds in the inner Galactic plane

***J. Syed**, H. Beuther, and the THOR collaboration*

This following Chapter is based on the work I have done so far for the paper *Cold atomic gas identified by* H I *self-absorption – Cold atomic clouds in the inner Galactic plane* that is currently in preparation and will be submitted to Astronomy&Astrophysics.

The paper has been reformatted to match the format of this thesis. I conducted the analysis of the data and applied the HISA extraction and Gaussian decomposition on the THOR survey. The entire draft was written by me.

ABSTRACT

Cold atomic hydrogen is an important constituent of the interstellar medium. In the context of molecular cloud formation, atomic hydrogen (H I) is expected to go through a phase of being cold H I before becoming molecular. Observational studies tend to have a strong bias in the detection of cold H I since observational targets typically involve regions that are bright in molecular line emission. We identify H I self-absorption (HISA) toward the inner Galactic plane and investigate the global distribution of the cold atomic gas in an unbiased way. We extend upon our previous findings and employ the newly developed tool astroSaber to obtain HISA features toward the inner Galactic plane within the scope of The H I/OH/Recombination line survey of the inner Milky Way. We derive the kinematic properties of the extracted HISA features using the spectral decomposition tool GaussPy+. We identify features that can be due to true self-absorption throughout the Galactic plane. The distribution of these absorption-like features is less diffuse than what is observed in atomic hydrogen emission. The kinematics of the extracted features show a median velocity dispersion of $\sigma_v \sim 3.6\,\mathrm{km\,s^{-1}}$ and a higher prevalence of features at positive radial velocities, owing to the kinematic distance ambiguity providing sufficient background for absorption to occur. With this work, we also make our entire decomposition results available.

6.1 Introduction

Cold atomic hydrogen is an integral constituent of the interstellar medium (ISM). Atomic hydrogen (H I) is predicted to exist in two distinct stable phases under typical Milky Way conditions (Field et al., 1969; McKee & Ostriker, 1977; Wolfire et al., 1995, 2003). The stable warm neutral medium of atomic hydrogen (WNM) has temperatures close to $\sim 10^4$ K and number densities $\sim 0.1\,\mathrm{cm^{-3}}$, while the the cold neutral medium (CNM) has temperatures up to a few $\sim 10^2$ K and number densities on the order of $\sim 10\,\mathrm{cm^{-3}}$. As atomic hydrogen is expected to undergo a state of being cold H I before becoming molecular, this atomic gas phase is an essential step in the formation of molecular clouds. Tracing cold atomic hydrogen in the ISM, however, is a challenging task. The emission of H I is effectively independent of spin temperature, and the direct measurement of optical depths to constrain the spin temperature of H I is restricted to the relatively sparse population of sufficiently bright continuum sources toward the Galactic plane. H I self-absorption (e.g., Knapp, 1974; Gibson et al., 2000; Li & Goldsmith, 2003; Krčo et al., 2008; Wang et al., 2020c) can act as an additional agent to trace the cold phase of atomic hydrogen, and it can do so more extensively as the widespread H I emission itself acts as a background (hence the term 'self'-absorption). The background does not have to be physically associated with the cold cloud for HISA to occur. Rather, the background can be H I emission originating in the far distance that has the same line-of-sight velocity with respect to the cold cloud.

In the context of molecular cloud formation, it is reasonable to target regions in the Galaxy that are bright in the commonly used molecular gas tracer CO, and search for cold atomic counterparts. However, this approach puts a strong bias on the detection of cold atomic hydrogen through self-absorption. Not only does the formation of molecular hydrogen (H_2) precede the formation of CO (Clark et al., 2012), which would lead to a cloud not being identified as molecular at an early stage,

cold atomic hydrogen furthermore exists throughout much of the Galactic disk, both predicted by theory (McKee & Ostriker, 1977; Wolfire et al., 1995, 2003) and confirmed by observations (Heiles & Troland, 2003). Consequently, while molecular line emission can argue in favor of true H I self-absorption, if the emission in H I decreases as molecular line emission increases, cold atomic gas as traced by HISA is frequently found without any molecular counterpart (Gibson et al., 2000).

In this paper, we recover HISA in an unbiased way using the tool astroSaber (see Sect. 5), that identifies signal dips based on the structure of H I emission profiles. In order to characterize the kinematics of the extracted features, we make use of the Gaussian decomposition tool GaussPy+ (Riener et al., 2019). The paper is organized as follows. In Sect. 6.2 we describe the survey data on which the HISA extraction is based, and present the method setup for the HISA identification and decomposition. Section 6.3 presents first results of the HISA extraction and the Gaussian decomposition. We discuss features that are falsely identified as absorption and draw a comparison with the molecular gas distribution that can serve as a validation check of the extracted features. We conclude in Sect. 6.5.

6.2 Observational data and methods

The subsequent analysis is based on data taken from The H I/OH/Recombination line survey of the inner Milky Way (THOR; Beuther et al., 2016; Wang et al., 2020a). The THOR survey is a Galactic plane survey covering a large portion of the first Galactic quadrant, at Galactic longitudes $17.6° \leq \ell \leq 67.0°$ and latitudes $|b| \leq 1.25°$. The survey includes observations of the H I 21 cm line, four OH lines, 19 Radio Recombination lines (RRLs), as well as the continuum between 1–2 GHz, obtained with the Very Large Array (VLA) in C-configuration. The final H I data were combined with VLA D-configuration and single dish observations that are part of the VLA Galactic Plane Survey (VGPS; Stil et al., 2006). The final H I data used in the following analysis have an angular and spectral resolution of $40''$ and $1.5\,\mathrm{km\,s^{-1}}$, respectively. The rms noise in emission-free channels is $\sigma_{\mathrm{rms}} \sim 4\,\mathrm{K}$.

6.2.1 HISA extraction

We employ the newly developed baseline extraction tool astroSaber to obtain baselines of H I self-absorption features. The astroSaber package uses a least squares method with asymmetric weights and a regularization term to smooth out dips in H I emission spectra. The regularization term is defined as the second order derivative of the spectrum and imposes a smoothness of the baseline. The balance between the amount of smoothing and the data fidelity is adjusted by two smoothing parameters, λ_1 and λ_2. For a detailed overview of the astroSaber algorithm, we refer the reader to Sect. 5.

The optimal smoothing parameters are found using training and test data that are generated from the original H I emission data. In the training step of astroSaber, the algorithm runs a gradient descent and iterates in a controlled manner through different values of the two smoothing parameters, such that the final smoothing parameters minimize the squared residuals between the resulting baselines and the

provided best-fit solutions of the test data. The amount of smoothing required to successfully recover expected HISA features depends on the noise of the data, the spectral resolution, as well as on the expected line width of HISA. The user therefore has to supply an expected line width, from which the training data are generated. We set the expected line width to $4\,\mathrm{km\,s^{-1}}$ (FWHM), as found in case studies by Wang et al. (2020c) and Syed et al. (2020) with an independent HISA fitting approach.

Since the H I emission spectra and noise show large variations across the Galactic longitude coverage of the THOR survey, we generated nine different training and test data sets that were randomly sampled from the entire THOR survey. Each training and test data set contained 500 spectra. We list the optimal smoothing parameters for astroSaber in Table 6.1. The optimal smoothing parameters found for each training data set compare well to each other. Since the final baseline extraction is less sensitive to variations in λ_1 than it is to variations in λ_2 (see Fig. B.1 in Appendix B.1), even the slightly larger variation in λ_1 makes a marginal difference in the final baselines. Given the similarity of the smoothing parameters across all training sets, we chose the median values $(\lambda_1, \lambda_2) = (3.16, 0.47)$ as our smoothing parameters for the baseline extraction. Using fixed smoothing values ensures that the extracted features are equivalent throughout the entire survey.

Table 6.1: Optimal smoothing parameters obtained for the astroSaber training sets.

Sample	λ_1	λ_2
1	3.09	0.32
2	3.45	0.42
3	3.16	0.43
4	2.64	0.62
5	3.19	0.41
6	3.15	0.52
7	3.71	0.47
8	3.18	0.49
9	3.12	0.47

Notes. The second column and third column give the best-fit λ_1 and λ_2 smoothing parameters obtained during the optimization step of astroSaber, respectively.

We then performed the baseline extraction using the median smoothing values and the default settings of astroSaber. The HISA features are then obtained by subtracting the H I emission from the resulting baselines. The algorithm also provides flags for which the baseline extraction did not converge given a maximum number of iterations. This is commonly the case where imperfect continuum subtraction toward strong continuum sources contaminate the spectra. We therefore removed all flagged spectra and additionally masked out regions where the continuum emission is $T_{\mathrm{cont}} > 100\,\mathrm{K}$ to ensure a clean data set.

6.2.2 Gaussian decomposition

For the spectral decomposition of the HISA data we used the fully automated Gaussian decomposition tool GaussPy+ (Riener et al., 2019), that extends upon the capabilities of the previously developed GaussPy (Lindner et al., 2015) and offers an improved fitting routine. The GaussPy algorithm uses derivative spectroscopy to make initial guesses on the number of Gaussian components and their parameters. The extension GaussPy+ includes an automated noise estimation routine and supplementary quality checks of the fit results. Moreover, GaussPy+ offers additional steps in the fitting routine to refit poorly-fit spectra and check for spatial consistency in the decomposition results. For further details about the GaussPy+ algorithm and its performance on challenging test fields or synthetic data, including extensive parameter studies, we refer the reader to Riener et al. (2019, 2020a).

Our goal is to provide a homogeneous kinematic decomposition and analysis of the entire THOR-HISA data set. We used the same parameter settings for the decomposition in the entire survey regions to ensure comparable results and avoid systematic effects introduced by non-uniform noise. The raw HISA cube extracted by astroSaber contains 16,007,253 spectra. Due to limited computational resources, the decomposition of the entire survey at once was not feasible. We therefore split the survey along the Galactic longitude axis into 25 individual, non-overlapping subcubes covering $\sim 2^\circ$ in longitude each.

For the decomposition of the HISA data, GaussPy+ uses a two-phase smoothing approach to remove noise that will have undesirable effects on the derivatives and thus the determination of signal peaks. The optimal smoothing parameters α_1 and α_2 (corresponding to the size of two different Gaussian smoothing kernels) are found using a machine learning algorithm on a training set consisting of typically a few hundred well-fit spectra. Similar to the HISA extraction using astroSaber, we created nine different training sets, each containing 500 spectra that are randomly selected from the entire HISA survey.

In Table 6.2 we list the optimal smoothing parameters α_1 and α_2 obtained with GaussPy+ for the nine training sets. The values do not show large differences. Small variations are generally expected due to variation in the noise of the different training samples but should have a limited effect on the decomposition results, as discussed by Riener et al. (2019, 2020a). We again decided to use the median values $(\alpha_1, \alpha_2) = (1.38, 2.02)$ as our chosen parameters for the decomposition of all individual subcubes.

We then performed all decomposition steps on the individual subcubes, including the spatial refitting that takes into account fit results of neighboring spectra and performs a refit if necessary. Because of the spatial refitting, the final decomposition results might show discontinuities at the edges of the individual subcubes. This can be investigated using the number of fit components that GaussPy+ evaluates for spatial consistency during the refitting phase. In general, the map of the number of Gaussian components fit per spectrum in Fig. 6.1 shows a spatially consistent distribution over almost the entire THOR coverage. There are slight inconsistencies at the Galactic longitudes around $\ell \sim 63.5^\circ$ and $\ell \sim 47.0^\circ$. These longitudes are, however, not at the edges of the subcubes and the features are due to artifacts evident in both the THOR-H I emission and HISA data (see also Fig. 6.2). We

Table 6.2: Smoothing parameters (in units of spectral channels) obtained for the GaussPy+ training sets.

Sample	α_1	α_2
1	1.38	2.02
2	1.39	2.03
3	1.32	1.95
4	1.19	1.92
5	1.29	1.96
6	1.42	2.07
7	1.43	2.09
8	1.28	1.97
9	1.42	2.08

Notes. The second column and third column give the best-fit α_1 and α_2 smoothing parameters obtained during the training step of GaussPy+, respectively.

conclude that the edges do not introduce systematic differences and that the fit results at the edges of the subcubes are overall consistent with neighboring spectra.

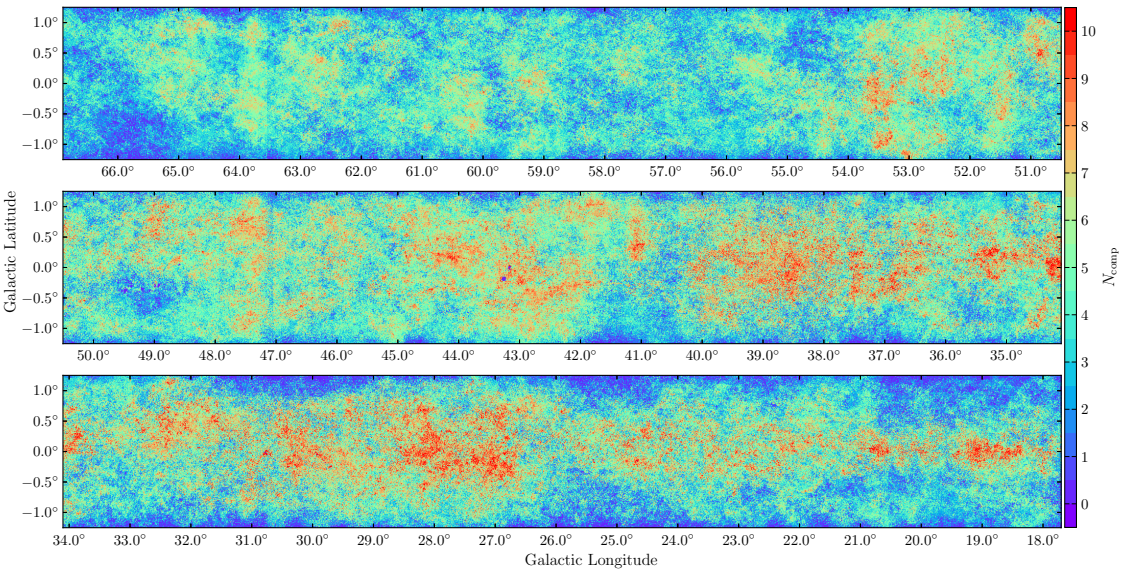


Figure 6.1: Map of the number of Gaussian components fit per HISA spectrum. The discontinuities at longitudes $\ell \sim 47.0°$ and $\ell \sim 63.5°$ arise from observational artifacts in the THOR data and are not a result of splitting the survey into subsections prior to the spectral decomposition.

6.3 Results

6.3.1 Global properties of HISA gas

We show in Fig. 6.2 a map of the resulting absorption features obtained with astroSABER, expressed as equivalent emission features and integrated over 15 km s^{-1} velocity bins. We identify tentative HISA features across the entire survey and at all velocities. We will refer to the extracted data as HISA or absorption, but we emphasize that the extracted features are likely not all due to true self-absorption as the algorithm identifies any signal dips comparable to the HISA line widths in the emission profiles of H I. We extracted less absorption at negative velocities. Since the distances corresponding to negative velocities are not ambiguous and probe the outer Galaxy, negative velocity channels are less crowded with H I emission that could provide a background for HISA. We would consequently expect less HISA at negative velocities. At low positive velocities, we extracted features over larger angular scales, possibly owing to self-absorption occurring in nearby clouds with distances less than 2 kpc. If we take the extracted baseline data as the total H I emission (H I without self-absorption), then HISA reduces the total emission by a fraction of 12.9%. We note the artifact seen across multiple velocity channels at Galactic longitude $\ell \sim 47°$. This is an artifact in the THOR-H I data originating at the edge of an observational tile.

In the following we focus on the positive velocities of the extracted features as we would expect a strong bias toward HISA features that are more likely to have sufficient background because of the kinematic distance ambiguity in the inner Galactic plane. We construct in Fig. 6.3 a longitude-velocity (ℓ-v) diagram based on the extracted HISA features by computing an intensity-weighted mean along the latitude axis. We plot the predicted locations of spiral arms (Reid et al., 2016, 2019) to compare the global distribution of the features with the Galactic structure. Most strikingly, we do not find a tight association of the extracted HISA features with the locations of spiral arms predicted in the Reid et al. (2016, 2019) model. Toward the Local Spur and the Aquila Spur, we observe a clear anti-correlation between their location in the ℓ-v diagram and HISA. The far side of the Sagittarius arm (dashed curve) also shows weaker absorption features, which is expected given the likely absence of emission background. The strongest and most pronounced absorption feature that we find is coincident with the Aquila Rift between ∼18° and ∼30° of Galactic longitude and at a velocity of ∼5 km s^{-1}, which is known to have strong absorption in H I (Riegel & Jennings, 1969; Riegel & Crutcher, 1972; Crutcher & Riegel, 1974). This absorption feature is also clearly seen in H I emission (see Fig. 4 in Wang et al., 2020a).

Beyond the velocities of the spiral arms and spurs, we detect significant absorption signals. We inspected spectra toward three positions at $(\ell, b) = (60.2°, -0.85°)$, $(45.9°, 0.00°)$, and $(38.6°, -0.45°)$, respectively, that approximately correspond to lines of sight that cut through the Local Spur, the Sagittarius arm and the Aquila Spur, respectively. We found that the astroSABER algorithm extracted features that are at the terminal velocities of the spectra (see Fig. 6.4). This is due to the algorithm smoothing over the sharp edges of the spectra, that mark the terminal

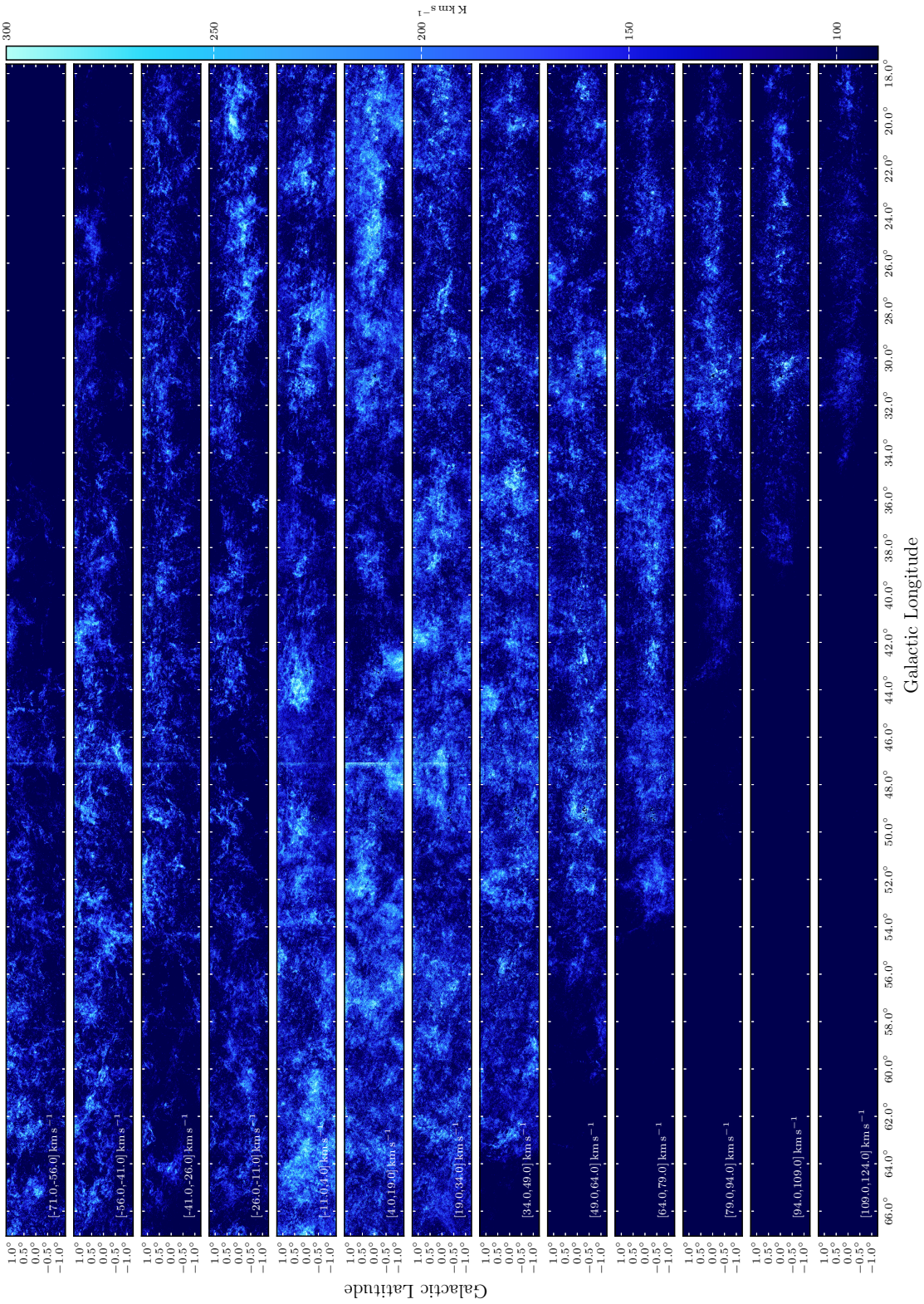


Figure 6.2: THOR-HISA moment 0 maps obtained by integrating the H I self-absorption features obtained with astroSaber over the velocity ranges indicated in each panel. The vertical feature recovered in some velocity channels at $\ell \sim 47°$ is an observational artifact in the THOR-H I emission survey.

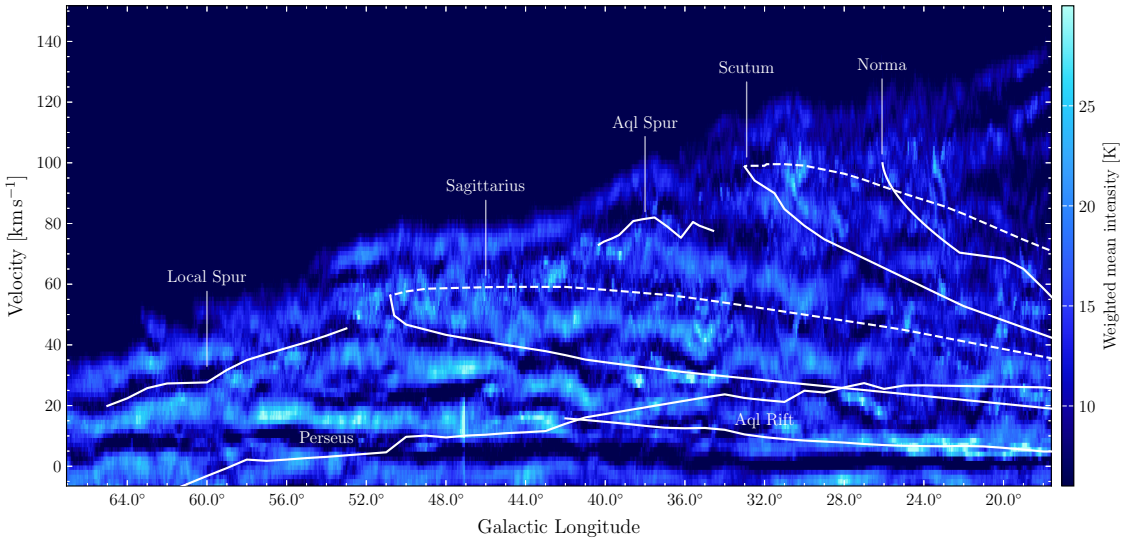


Figure 6.3: HISA longitude-velocity (ℓ-v) diagram constructed by calculating the intensity-weighted mean over the latitude range $|b| \leq 1.25°$. The overlaid curves trace the Sagittarius, Scutum, Norma, and Perseus spiral arms, as well as smaller features like the Local Spur, Aquila Spur (Aql Spur), and the Aquila Rift (Aql Rift) taken from (Reid et al., 2016, 2019). The near and far sides of the spiral arms are plotted with solid and dashed lines, respectively. The feature visible at $\ell \sim 47°$ and velocities between 5–20 km s^{-1} corresponds to the artifact seen in Fig. 6.2.

velocities. These features are therefore artifacts that are not true self-absorption features and we apply a post-processing step in Sect. 6.4.1 to remove these artifacts.

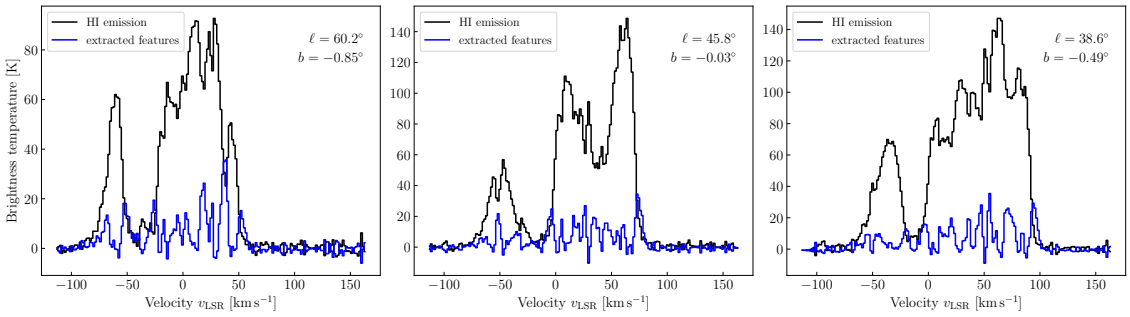


Figure 6.4: H I emission spectrum and extracted absorption features. The H I emission (black) and HISA spectrum (blue) in each panel are taken at the locations $(\ell, b) = (60.2°, -0.85°)$, $(45.9°, -0.03°)$, and $(38.6°, -0.49°)$, respectively, and are averaged over one beam size. Features at the terminal velocities of the spectra have erroneously been identified as signal dips.

6.3.2 Gaussian decomposition and statistics of components

We performed the decomposition on all extracted features, including the features at the terminal velocities discussed in the previous subsection. The final decomposition contains 67,383,250 Gaussian fit components. The fit solutions include fit components with S/N as low as 1.5, which is the default setting of GaussPy+. As pointed out by Riener et al. (2019), it can be beneficial to include such low S/N

as broad low-amplitude peaks can still add to the overall emission or absorption that is to be recovered by the fit. However, depending on the application of these data, it can still be sensible to impose a higher noise threshold. About ∼93% of all fit components have $S/N > 3$, and ∼42% have $S/N > 5$. The number of fit components is associated with the complexity in the spectrum and it can be a proxy for how many structures we identify along the line of sight. We find that lines of sight near the Galactic midplane require multiple Gaussian components to fit the spectra (see Fig. 6.1). We would naturally expect most CNM features near the Galactic midplane, either as true absorption features or as narrow emission components that could mimic absorption dips by a superposition and thus enter the astroSaber detection. Figure 6.5 shows the normalized histograms of the number of fit components. For ∼4% of the spectra, GaussPy+ could not fit any Gaussian component and about 8% are fit with one component. The median number of components fit per HISA spectrum is 4. One percent of the spectra are fit with ten or more components. We show in Fig. 6.6 statistics of the fit parameters of all

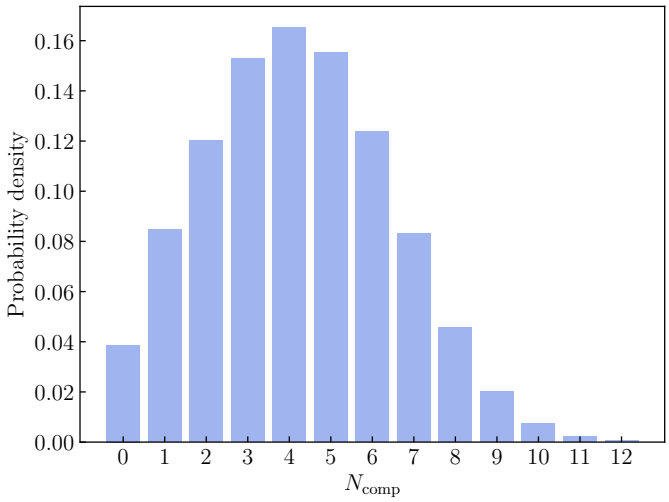


Figure 6.5: Normalized histogram of the number of Gaussian fit components per spectrum across the whole extracted HISA data set.

fit components within the extracted HISA data. We find a median amplitude or equivalent absorption depth of the extracted HISA features of ∼23 K, which corresponds to a $\sim 4.4\sigma_{\mathrm{rms,hisa}}$ detection, with $\sigma_{\mathrm{rms,hisa}}$ being the HISA rms noise. The middle panel in Fig. 6.6 shows that we fit HISA features across a wide range of velocities. The distribution has a median of $\sim 17\,\mathrm{km\,s^{-1}}$ and shows a peak around $v_{\mathrm{LSR}} \sim 0\,\mathrm{km\,s^{-1}}$. This is likely an effect of lines of sight that are confused at velocities close to the reference velocity, where fluctuations in the dynamics of the local H I emission might produce the observed features. Toward negative velocities there is a clear drop in the number of components, likely owing to the fact that there is statistically less background because of the missing kinematic distance ambiguity. Although we trained our extraction routine on HISA line widths of $\sim 4\,\mathrm{km\,s^{-1}}$, we identify most components with fit line widths of $\sim 6\,\mathrm{km\,s^{-1}}$, also highlighting the algorithm's ability to detect larger line widths. With this paper, we publish our full GaussPy+ decomposition results of the HISA cube obtained with astroSaber,

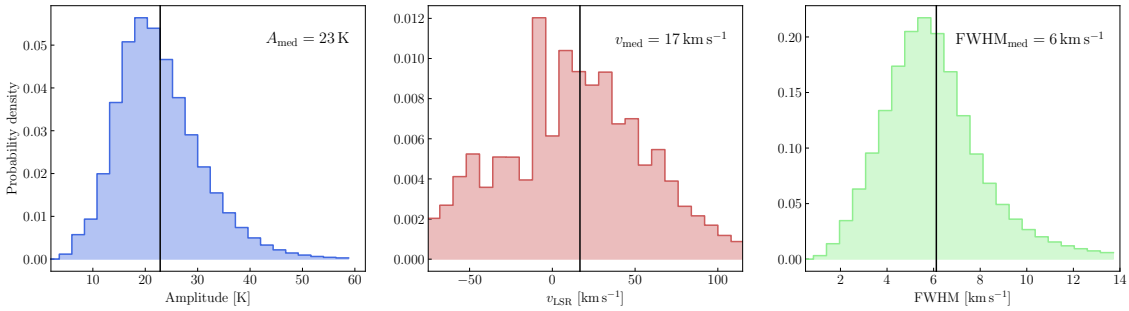


Figure 6.6: Normalized histograms of Gaussian fit component parameters. The left panel shows the distribution of the fit amplitude values of the Gaussian decomposition. The middle panel shows the fit mean velocity of the components. The right panel shows the fit line widths, expressed as full width at half maximum (FWHM) of the Gaussian components. The vertical black line in each panel marks the median value of the distribution.

including the components that are identified toward the terminal velocities of the spectra (see Sect. 6.4.1).

6.4 Discussion

6.4.1 Removal of terminal velocity artifacts

Our spectral decomposition allows us to isolate undesired features in the extracted HISA spectra. As described in Sect. 6.3.1, the astroSABER algorithm created artifacts at the edges of the H I emission spectra by smoothing over the sharp edges that mark the terminal velocities. In a post-processing step, we selected the highest-velocity component in each fit spectrum and checked the intensity of H I emission in five neighboring velocity channels on either side of the component. If the median emission is below the $3\sigma_{\rm rms}$ noise on one or both sides of the component, we flagged the component in our decomposition as an "edge" component.

Figure 6.7 shows the reconstructed spectra from the Gaussian decomposition, after removing the flagged components, toward the same locations as in Fig. 6.4. We could successfully remove the edge components, that accounted for ∼19% of all fit components. For our final decomposition catalog we opted to keep the edge components but mark them with an additional flag if the baseline around the component is below the noise. We note that the extracted features, that we consider artifacts in our self-absorption detection, could serve as a reference to constrain the terminal velocities of H I (see e.g., McClure-Griffiths & Dickey, 2007). The fits provide a mean velocity of edge components, that lie between H I intensity peaks and the noise, and can be a potential proxy for the terminal velocity. We make an ℓ-v diagram of the HISA cube that is reconstructed from the GAUSSPY+ decomposition results, after removing edge components. Figure 6.8 clearly shows that the features beyond the velocities of the Galactic structures remain after removing edge components. We conclude that the extracted features evident in the ℓ-v diagram beyond the spiral arm structures are significant and reliable.

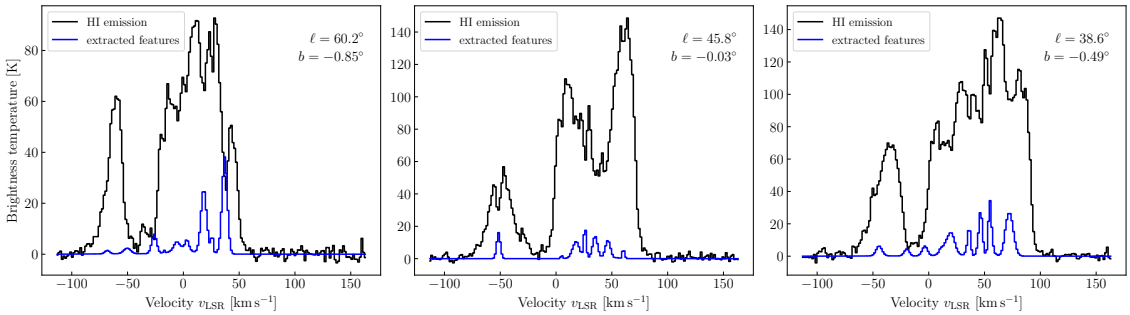


Figure 6.7: H I emission spectrum and extracted absorption features fit with GaussPy+. The H I emission (black) and HISA spectrum (blue), reconstructed from the GaussPy+ decomposition, are taken at the locations $(\ell, b) = (60.2°, -0.85°)$, $(45.9°, -0.03°)$, and $(38.6°, -0.49°)$, respectively, and are averaged over one beam size. Fit components at the terminal velocities have been removed from the spectra in a post-processing step.

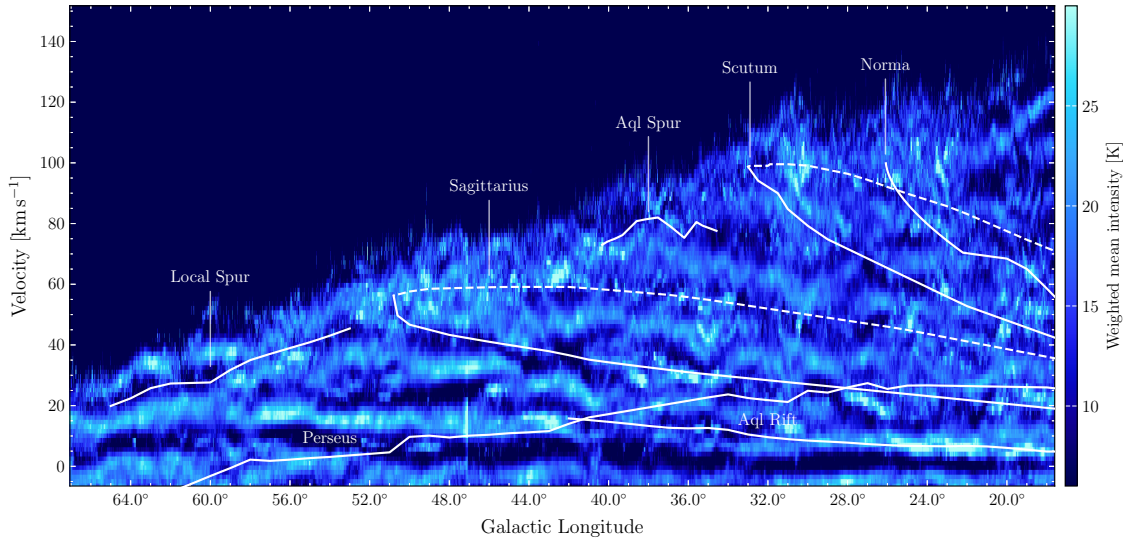


Figure 6.8: Longitude-velocity (ℓ-v) diagram of the HISA GaussPy+ decomposition constructed by calculating the intensity-weighted mean over the latitude range $|b| \leq 1.25°$. The overlaid curves are the same as in Fig. 6.3. We note that the weighted mean is slightly enhanced compared to Fig. 6.3, since weak absorption is not fit by GaussPy+ and does not enter the calculation.

6.4.2 Validation of extracted features

Since we have extracted tentative absorption features based solely on the emission profiles of H I spectra, we do not impose any correlation of HISA with molecular line emission. However, corresponding molecular line emission can be a strong indication of the presence of true H I self-absorption. In Fig. 6.9 we plot a ratio ℓ-v diagram of our extracted features over ^{13}CO emission taken from the Exeter FCRAO CO Galactic Plane Survey (Mottram & Brunt, 2010), and Galactic Ring Survey (GRS; Jackson et al., 2006). We smoothed and reprojected the HISA and ^{13}CO data onto the same spatial and spectral grid (pixel size of 22″, beam size of 46″, and velocity resolution of 1.5 km s^{-1}). We constructed this diagram by dividing the HISA ℓ-v diagram with the ^{13}CO diagram. For the ^{13}CO diagram, we used a lower limit of 3σ where the average emission is below 3σ, with $\sigma \sim 0.05$ K in the ^{13}CO

ℓ-v diagram. The absence of HISA toward the Local Spur and the Aquila Spur is even more pronounced due to enhanced ^{13}CO emission. While HISA does not appear to be completely absent toward the Sagittarius and Scutum spiral arms, the ratio is dominated by ^{13}CO emission that closely traces the predicted spiral arm locations. Since many locations show distinct features in the ratio ℓ-v diagram, there does not appear to be a tight positive correlation between HISA and ^{13}CO. While corresponding molecular line emission can make a strong case in favor of features being due to self-absorption, we stress, however, that a lack of any molecular counterpart does not disqualify any of our extracted features.

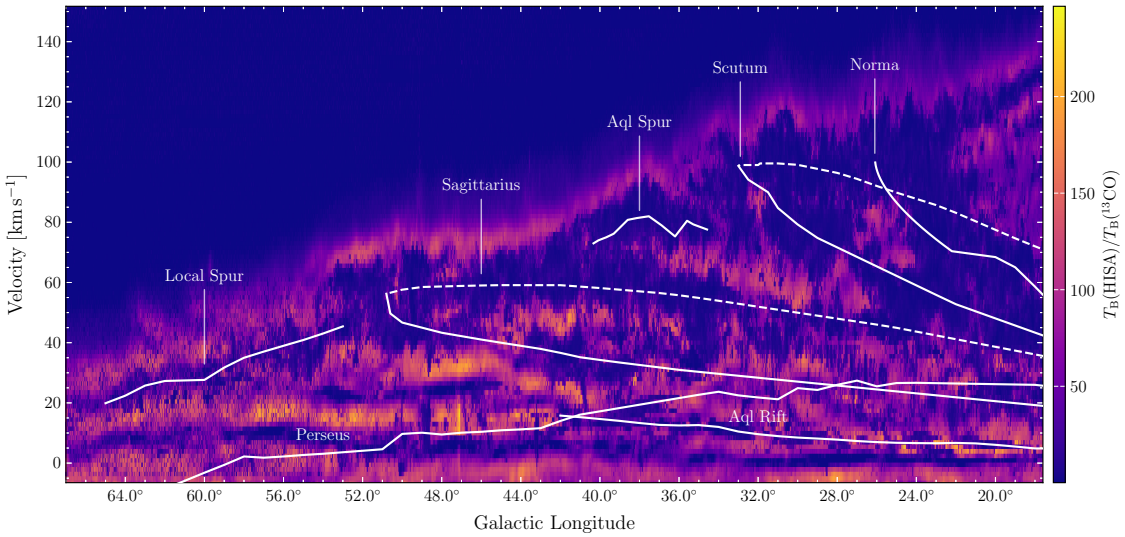


Figure 6.9: Ratio ℓ-v diagram of HISA, expressed as equivalent emission, over ^{13}CO emission. The respective ℓ-v diagrams are derived by computing the intensity-weighted mean over the latitude range $|b| \leq 1.0°$. The overlaid curves are the same as in Fig. 6.3.

6.4.3 Face-on view of H I self-absorption

We estimated the distances of our fit components using the Bayesian distance calculator (BDC; Reid et al., 2016, 2019) that is designed to calculate distances of spiral arm sources. Given the coordinates $(\ell, b, v_{\rm LSR})$, the BDC computes the probability density function (PDF) based on multiple prior probabilities that include: 1) a kinematic distance prior (KD), 2) the Galactic latitude (GL), 3) the proximity to parallax sources (PS), that are high-mass star forming regions whose parallaxes have been measured as part of the Bar and Spiral Structure Legacy (BeSSeL) Survey[1] and the Japanese VLBI Exploration of Radio Astrometry (VERA)[2], 4) the proximity to features of an assumed spiral arm model (SA), 5) and the proper motion of the source (PM). The prior probabilities have a range between $[0, 1]$. In the default setting of the BDC, the priors of KD, GL, and SA, are set to 0.85. The prior of the parallax sources is given less weight of 0.15 and the proper motion is set to 0. In addition, the BDC takes a prior probability of the kinematic far distance solution into

[1] http://bessel.vlbi-astrometry.org
[2] http://veraserver.mtk.nao.ac.jp

account, that can be adjusted by the user. By default, the kinematic near and far distance are given an equal probability of 0.5. Since the BDC is specifically designed to assign reliable distances to sources that are expected to be associated with spiral arms, the distance estimator has a strong bias in placing any input source near spiral arm features (see Fig. 6 in Reid et al., 2016). We have therefore performed two different distance calculations, in the same fashion as done in Riener et al. (2020b). For both cases, we kept the PM prior at zero since we have no information about the proper motion of the gas. In the first set of the distance calculation, we included all other distance priors KD, GL, PS, and SA. As Riener et al. (2020b) have shown, setting the spiral arm prior to the default value of 0.85 leads to a very strong bias in the distance calculation, such that other prior become negligible. In order to achieve a more balanced ratio between the priors, we have set the SA prior to 0.5. As the Galactic latitude prior is coupled to the spiral arm prior, we adjusted the prior GL and set it to 0.5 as well, to keep the same ratio between the two priors. The KD and PS priors were set to the default of 0.85.

In the second run, we removed the spiral arm prior from our distance calculation by setting its weight to 0. As the GL prior is combined with the SA prior, setting SA to zero effectively removes the GL from the distance estimation as well. To keep the prior of the latitude information, we modified the BDC source code, such that we could use it without having to include the spiral arm prior. Like in our first run, we set the GL prior to 0.5. Since we expect HISA to have a strong bias toward near distances, we set the kinematic far distance probability to 0.1 in both runs. In the following, we only consider positive radial velocities and focus on the Galactic plane within the solar circle.

Given the input coordinates and priors, the BDC computes a distance PDF and reports up to two distances that have the highest integrated probability. For every solution in the distance PDF, we selected the component corresponding to the kinematic near distance and imposed an integrated probability threshold of 50%. If the component has a probability below 50%, we removed it from our distance catalog.

We show in Fig. 6.10 the face-on view of HISA, expressed as equivalent integrated emission, in the Galactic plane. The left panel shows the distance calculation including the spiral arm prior, that clearly moves most of the HISA toward the position of spiral arms. Without the spiral arm prior (right panel in Fig. 6.10), the HISA gas is more spread out and does not strictly follow the spiral arm locations. However, we can still discern the spiral arm pattern in the distribution of HISA. The positions of Galactic features in the spiral arm model by Reid et al. (2016, 2019) are based on maser sources and archival H I and CO data. The observed gas is then still likely to be moved near prominent features that characterize the spiral structure of our Galaxy. We would furthermore expect most HISA where we observe overdensities in H I.

6.5 Conclusions

We have investigated the presence of cold atomic hydrogen in the Galactic plane by means H I self-absorption. In order to characterize absorption features independent

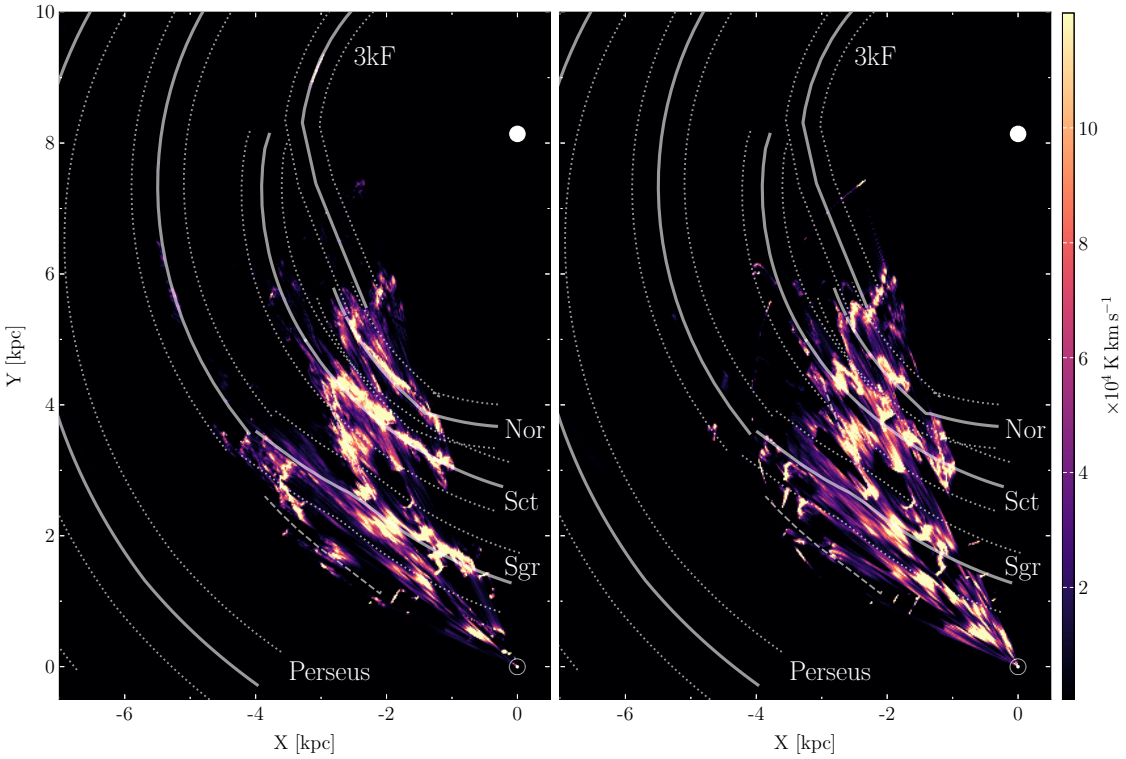


Figure 6.10: Face-on view of the integrated HISA, expressed as equivalent emission for the BDC results obtained with (left) and without (right) the spiral arm prior. The values are binned in (20×20) pc cells and are summed up along the z_{gal} axis. The position of the sun and Galactic centre are indicated by the sun symbol and white dot, respectively. The predicted locations (solid line segments) and widths (dotted) of the 3 kpc far arm (3kF), and the Perseus, Sagittarius (Sgr), Scutum (Sct), Norma (Nor) spiral arms are overplotted for reference (Reid et al., 2019). The Local Spur and Aquila Spur are plotted with dashed lines (without labels).

of the occurrence of molecular gas emission, we made use of the astroSABER algorithm that identifies signal dips in H I emission spectra. We decomposed the spectral features obtained with astroSABER using the decomposition tool GAUSSPY+ (Riener et al., 2019).

The derived features are observed throughout the Galactic plane, with a higher prevalence at $v_{\mathrm{LSR}} > 0\,\mathrm{km\,s^{-1}}$. The median absorption depth is 23 K and the observed line widths (FWHM) of the features are $\sim$6 km s^{-1}. There is no clear correlation between HISA and ^{13}CO emission, or the predicted locations of spiral arms and spurs, that are in many cases well traced by molecular gas. In a following paper, we will identify coherent associations of HISA features using the SCIMES algorithm (Colombo et al., 2015) and publish a catalog of HISA clusters, together with validity checks concerning molecular line emission and H I optical depth.

7

Summary and outlook

We revisit the questions posed in the beginning of this thesis.

1. What are the physical and kinematic properties of the atomic gas in the ISM and how is it structured?

2. How can we constrain the transition from the atomic to the molecular gas phase using observations?

3. What is the relationship between the atomic and molecular gas in the ISM?

7.1 The Maggie filament

In Chapter 4 we have seen that atomic gas assembles on a large scale to form filaments that are preferentially aligned with the Galactic disk (Soler et al., 2020). The Maggie filament is a unique condensation of atomic gas that contains little molecular gas (molecular gas mass fraction $\sim$8%) and has a length of more than one kiloparsec. Given its location in the outer Galaxy, at a distance of 12 kpc from the Galactic center and $\sim$500 pc below the Galactic midplane, the formation of molecular gas is deemed unlikely. The kinematics of Maggie are governed by the large-scale dynamics of the Milky Way disk. The centroid velocities suggest that the filament follows the Galactic rotation close to the predicted location of the Outer Arm with an offset of $\sim$5–10 km s^{-1}, possibly due to its distance from the plane at which there can be a lag in the corotation (Kalberla & Kerp, 2009). It is important to note, however, that the predicted location of the Outer Arm at longitudes $\ell \lesssim 70°$ is an extrapolation of measured parallaxes in the Galactic longitude range $70° \lesssim \ell \lesssim 230°$ (Reid et al., 2016, 2019). Spiral arms have kinks and considerable changes in pitch angle over characteristic lengths of 5–8 kpc (Honig & Reid, 2015). The location of the Outer Arm in the longitude range of Maggie has therefore a large uncertainty and the discrepancy between Maggie and any Galactic feature can be due to the highly uncertain spiral arm model for the far side of the Galaxy.

Maggie shows a spatial offset from the plane and illustrates in a good way the complex vertical structure of the Galactic H I disk. Plane-parallel features like Maggie can be explained by corrugations in the Milky Way plane. The Galactic H I disk can be described as an oscillating plane (Levine et al., 2006), with low-frequency modes accounting for the global asymmetries observed as the Galactic warp (e.g., Nakanishi & Sofue, 2003), and significant higher-order components related to local perturbations of the disk. Local perturbations in the Galactic potential of the disk can result in different bending modes and can be triggered by the gravitational interaction of satellites with the disk (Widrow et al., 2014).

The favorable viewing geometry renders Maggie a unique filament, in the sense that it allows us to discern it in H I emission because of its offset from the midplane. A statistical view of the global structure of H I emission shows that plane-parallel features are fairly common throughout the Milky Way, particularly in the outer Galaxy (Soler et al., 2022). This can be linked to a decrease in the average star formation rate surface density with Galactocentric distance (Lee et al., 2016; Elia et al., 2022). The structure of atomic hydrogen is then less affected by the dynamical signatures of stellar activity, rather it is dominated by the large-scale kinematics of the Galactic rotation. While the Maggie filament is a unique observational discovery, this case study could be the first in what constitutes a new class of atomic filaments shaped by the Galaxy.

7.2 H I self-absorption and cold atomic gas

Atomic hydrogen must concentrate in clumps and undergo a state of being cold H I before it can drive the formation of cold molecular clouds. For that reason, molecular clouds are ideal targets in the search of cold atomic gas that can be traced by H I self-absorption. The presence of molecular gas is often used to identify a decrease in H I intensity as HISA. The statistical properties of HISA are then a priori related to the properties of molecular clouds. With the newly developed tool astroSABER, we are able to identify signatures of absorption features that can serve as a baseline for HISA, independent of the spatial or kinematic correlation with molecular gas. In Chapter 5, we have tested our extraction on an extended sample of six giant molecular filament regions that are bright in CO emission and have been known to be sources of self-absorption (Wang et al., 2020c; Syed et al., 2020). By restricting the velocity range of our extracted features to that of the molecular clouds, we implicitly impose a kinematic bias that associates any detected absorption feature with the molecular gas. We detect HISA toward all investigated filament regions, of which three show strong HISA counterparts to the molecular gas. Since the astroSABER algorithm is sensitive to a certain range of absorption line widths, based on the smoothing weight that is optimized for specific line widths in the training data, we find similar observed line widths toward all filament regions. Independent of the imposed correlation in velocity between identified HISA and ^{13}CO emission, we detect significant spatial correlation in all regions between both gas tracers (Sect. 5.4.3). The column densities traced by HISA likely account for only a fraction of the total column density of cold atomic gas (see Seifried et al., 2022). The column density structure of the HISA-traced gas clearly connects the

diffuse atomic gas observed in H I emission with the much more concentrated spatial distribution of the molecular gas (see Fig. 5.6).

In Chapter 6 we have applied the HISA extraction on the entire Galactic plane covered by the THOR survey. The detection of absorption signatures extends to the whole of the survey, with a preference of features at positive radial velocities that correspond to the inner Galactic plane. This is a detection bias owing to the kinematic distance ambiguity, providing sufficient background for HISA to be observed. A comparison of all extracted absorption features with the molecular gas traced by ^{13}CO emission does not reflect a tight correlation. This suggests that a considerable fraction of HISA might not have a molecular counterpart. We emphasize that all extracted absorption features, with or without molecular emission counterpart, trace similar signatures in that they are significant in absorption depth and have line widths comparable to what is observed toward molecular clouds. The kinematic properties of all extracted features have been derived using the automated spectral decomposition tool GaussPy+ (Riener et al., 2019). We will make our full decomposition results available with an upcoming paper.

7.2.1 Cluster identification

In a future publication we aim to add validation checks to our extracted features with additional post-processing steps. To do so, we identified clusters that are coherent in *p-p-v* (position-position-velocity) space. In order to decompose our extracted HISA data into discrete clusters, we have employed the Spectral Clustering for Interstellar Molecular Emission Segmentation (SCIMES[1]) algorithm (Colombo et al., 2015) that builds upon hierarchical structures identified in a dendrogram analysis (Rosolowsky et al., 2008). The SCIMES algorithm is an automated tool originally designed for molecular cloud identification that uses spectral clustering and graph theory to analyze the hierarchical structure of emission data. The algorithm works well on high-resolution data and finds discrete regions with similar emission properties and groups them into 'clusters.' When expressed as equivalent emission (the inverted absorption spectrum), we can use SCIMES for the segmentation of discrete HISA clouds.

In our cluster identification and the choice of our input parameters, we adopted the procedures fully described in Colombo et al. (2019) and Duarte-Cabral et al. (2021). We performed the cluster identification on subsets of the full THOR-HISA survey, as the clustering algorithm is computationally expensive. This requires additional post-processing steps to combine the identified clusters back into a single coherent catalog, which is described in Duarte-Cabral et al. (2021). We split our survey into 20 tiles of $2.5° \times 2.5°$. To account for noise fluctuations and to achieve a more reliable clustering, we masked the data by setting it to zero where the equivalent HISA 'emission' is below $2\sigma_{\rm hisa}$ of the local noise. As described in Sects. 4.2.2 and 5.2.3, GaussPy+ comes with a pixel-by-pixel noise estimation that automatically identifies emission-free channels to estimate the noise.

We then performed a dendrogram analysis which forms the first step of the SCIMES clustering algorithm. We selected the median noise of $\sigma_{\rm med} = 5.1\,{\rm K}$ as a

[1] https://github.com/Astroua/SCIMES

uniform noise value for all tiles, a $4\sigma_{\rm med}$ value as the minimum difference between two peaks for them to be considered as separate structures, and a lower threshold for detection of $2\sigma_{\rm med}$, to maximize the connections between different structures at contiguous lower intensity levels. These uniform thresholds ensure consistent and homogeneous clustering based on the same criteria. In addition, we set a minimum number of voxels for a structure to be considered as real to be six times the number of pixels per beam ($N_{\rm ppbeam} = 4$).

With the dendrograms as an input, we ran SCIMES using the default 'volume' and 'flux' as the clustering properties to be considered (see Colombo et al., 2015). Our extraction of HISA clusters resulted in 1962 clusters from the 20 tiles. Many of these clusters are potentially duplicated as they overlap between neighboring tiles. To resolve that, we performed a second run with tiles that have a longitude range that is shifted by half the tile size (1.25°) compared to the first run. If an identified cluster touches an edge of a tile, it should be fully recovered in the other run of the clustering (unless the cluster stretches across more than 2.5° in longitude). The complete cluster catalog will become available in the near future.

We show an example of identified clusters toward the GMF38a region (see Chapter 5) in Fig. 7.1. Single clusters are represented by contours with the same color. The algorithm confirms our 'by-eye' inspection of the extracted HISA cloud and agrees well with the self-absorption features investigated in Wang et al. (2020c).

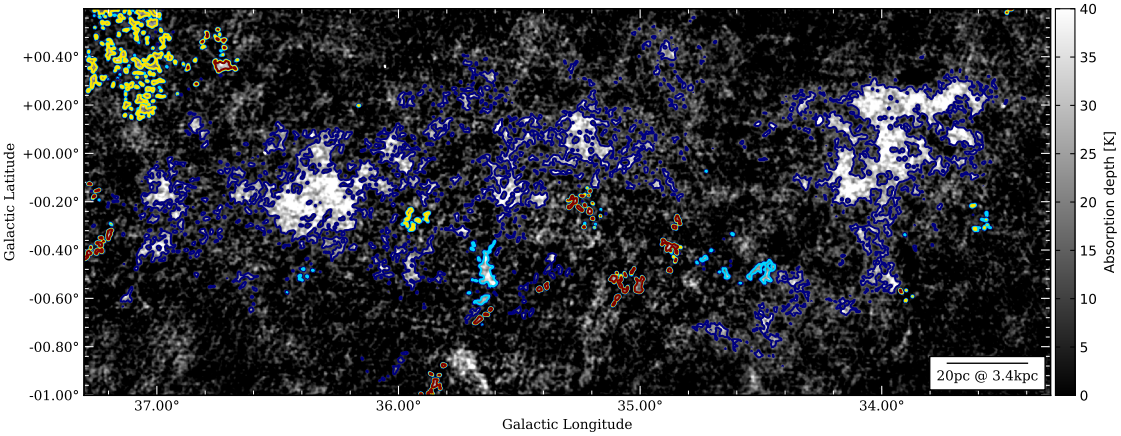


Figure 7.1: HISA channel map of GMF38a. This map shows the HISA intensity toward GMF38a (see Chap. 5) at the velocity channel $v_{\rm LSR} = 55\,{\rm km\,s^{-1}}$. The colored contours represent clusters identified with SCIMES.

The identified clusters can then be matched with the H I optical depth measurements in Wang et al. (2020a), to confirm if the HISA clouds correspond to high optical depth. In this way, we can construct a comprehensive catalog of HISA features that can be compared to molecular line surveys on a cloud-by-cloud basis. Eventually, we aim to provide a full catalog of HISA clouds, with additional flags indicating associated molecular line emission and high optical depth.

A

Additional material to Chapter 4

A.1 Column density of Maggie component

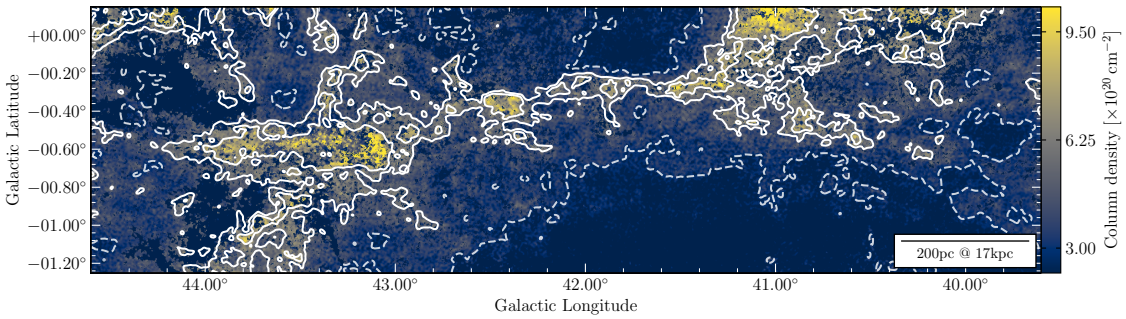


Figure A.1: Integrated column density of the fit Maggie component. In a thermally bistable medium, two components (i.e., CNM and WNM), which are close in velocity, are encoded in the emission profile. We identify the Maggie component based on the mean position in the velocity domain, thus picking up both CNM and WNM features. The contours are the same as in Fig. 4.11.

A.2 MWISP CO observations

A.3 Stellar extinctions using Gaia, 2MASS, and WISE

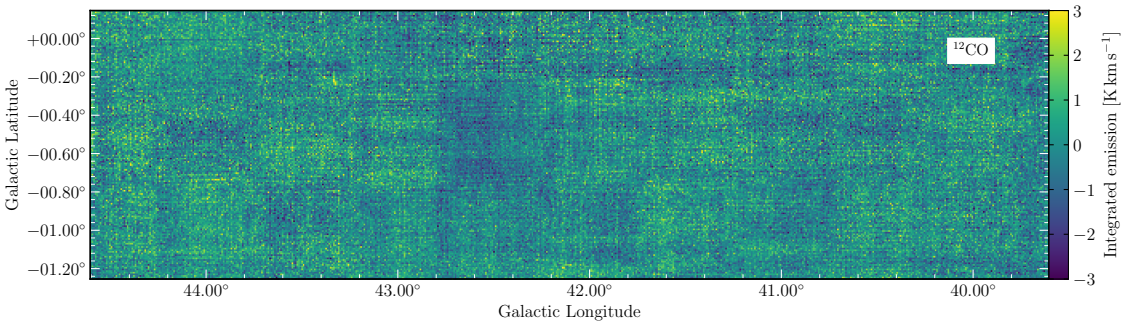


Figure A.2: MWISP ^{12}CO integrated emission. The ^{12}CO (J =1–0) emission data taken from the MWISP survey are integrated over the velocity interval between -57.5 and $-48.5\,\mathrm{km\,s^{-1}}$.

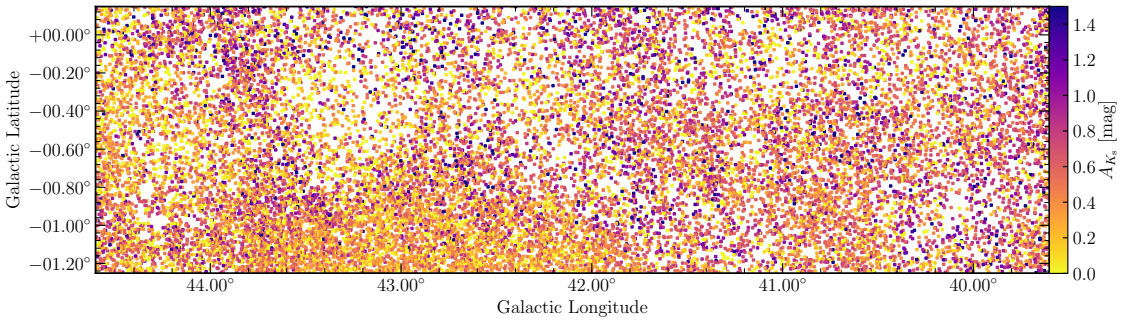


Figure A.3: Extinctions up to 5 kpc distance using *Gaia* in combination with near- and mid-infrared photometry data from 2MASS and WISE. The blank spaces are regions containing obscuring material that has blocked out stars from our *Gaia* sample and left some highly extinguished stars in their surroundings.

B

Additional material to Chapter 5

B.1 Momentum-driven gradient descent

The smoothing parameter $\boldsymbol{\lambda}$ (which is in our case a two-component vector by default) is tuned to maximize the fitness of the self-absorption baselines using a batch gradient descent with momentum (Ruder, 2016). We define the median reduced chi square $\langle \chi^2_{\rm red} \rangle$ as the cost function $\mathcal{C}$ that we wish to minimize in order to achieve the highest goodness of fit result:

$$\mathcal{C}(\boldsymbol{\lambda}) = \langle \chi^2_{\rm red} \rangle = \left\langle \frac{\sum_{i=1}^{N} \frac{(y_i - z_i(\boldsymbol{\lambda}))^2}{\sigma^2_{\rm rms}}}{N-k} \right\rangle , \tag{B.1.1}$$

with y_i and $z_i(\boldsymbol{\lambda})$ denoting the data and baseline value at channel position i, respectively, N is the sample size (in this case the number of spectral channels containing the self-absorption feature), k denotes the degrees of freedom, which is in our case $k = 1$ for one-phase smoothing or $k = 2$ for two-phase smoothing, and $\sigma_{\rm rms}$ is the rms noise of the data.

In a classical gradient descent, updates to the smoothing weight $\boldsymbol{\lambda}$ are made by moving in the direction of greatest decrease in the cost function, that is $\Delta\boldsymbol{\lambda} = -\ell\, \nabla\mathcal{C}(\boldsymbol{\lambda})$, where the learning rate ℓ controls the step size. Since the cost function is usually highly non-convex, we implemented a gradient descent with added momentum to overcome local minima that might be due to noise or fluctuations in the spectra. Therefore, at the n-th iteration, the change in $\boldsymbol{\lambda}$ is given by

$$\Delta\boldsymbol{\lambda}^{(n)} = -\ell\, \nabla\mathcal{C}(\boldsymbol{\lambda}) + \phi\, \Delta\boldsymbol{\lambda}^{(n-1)} , \tag{B.1.2}$$

where the momentum ϕ controls the degree to which the previous step influences the current one. The gradient $\nabla\mathcal{C}(\boldsymbol{\lambda})$ in Eq. (B.1.2) is defined as

$$\nabla\mathcal{C}(\boldsymbol{\lambda}) = \begin{pmatrix} \frac{\mathcal{C}(\lambda_1+\epsilon,\lambda_2) - \mathcal{C}(\lambda_1-\epsilon,\lambda_2)}{2\epsilon} \\ \frac{\mathcal{C}(\lambda_1,\lambda_2+\epsilon) - \mathcal{C}(\lambda_1,\lambda_2-\epsilon)}{2\epsilon} \end{pmatrix} , \tag{B.1.3}$$

where we set the finite-difference step $\epsilon = 0.1$. Figure B.1 shows example tracks of $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$ when using the gradient descent with different initial values for λ_1 and λ_2 during the two-phase optimization on THOR-H I data. We find that small-scale local optima are ignored effectively during the search for large-scale optima.

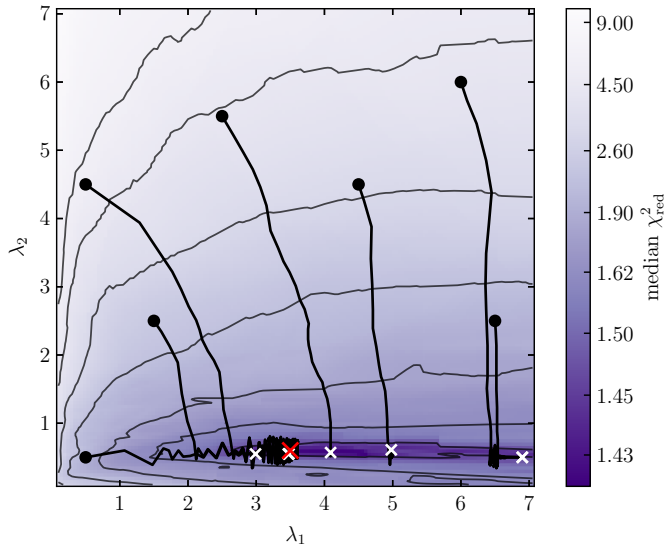


Figure B.1: Smoothing parameter optimization using gradient descent. The map shows a sampled representation of the underlying $\boldsymbol{\lambda}$ parameter space in terms of the median value of the reduced chi square results. Initial values, tracks, and convergence locations of the (λ_1, λ_2) parameters during the optimization are represented by black circles, black lines, and white crosses, respectively. The red cross marks the global minimum in the sampled parameter space. Initial locations that start off too far from the global best solution ($\lambda_1 = 3.5, \lambda_2 = 0.6$) might converge to local minima with less accurate fit results.

B.2 Classical second derivative approach

Another way to identify HISA features uses the second derivative of the observed H I spectrum as described in Krčo et al. (2008). Pronounced self-absorption features would therefore become readily apparent as signatures in the second derivative representation of the spectrum. In the following, we discuss the limitations of this method and how astroSABER overcomes the issues imposed by finite spectral resolution and noise.

Calculating the second (or higher) derivatives using finite-difference techniques might not always give reliable results as noise in the observational data will be greatly amplified. This is illustrated in Fig. B.2. While the top panel shows a mock-H I spectrum including two self-absorption components that does not contain noise, the bottom panel presents the same spectrum with added noise that is comparable to the THOR data (same spectrum as in Fig. 5.1). The green spectrum in each panel shows the finite-difference second derivative of the spectrum. For the noise-less data, the narrow HISA component emerges as a signature in the second derivative. Although less pronounced, even the broader absorption feature can be identified

through an enhancement in its second derivative. On the other hand, given the observed spectrum that contains noise (lower panel in Fig. B.2) the HISA components do not become visible as noise fluctuations dominate the second derivative of the spectrum. To overcome this, regularized differentiation can be used to mitigate the effect of noise fluctuations. It is a method of regularization of ill-posed problems that commonly occur in models with large numbers of parameters or inverse-solving during optimization. For example, this so-called Tikhonov regularization (Tikhonov, 1963) may be used to enforce smoothness of a given vector, giving preference to solutions that minimize the second derivative.

In a similar way, astroSABER uses this type of regularization when it introduces a penalty term to the (asymmetric) least squares function that minimizes the second derivative (see Eq. 5.2.1). This is demonstrated in a simplified way in the lower panel of Fig. B.2. The dashed blue spectrum shows a (in this case symmetric) least squares solution to the mock-H I spectrum with a regularization term as in Eq. (5.2.1). The second derivative of the smooth representation of the spectrum now responds to the narrow absorption feature (blue spectrum) and shows a peak. However, the broader feature cannot be identified in the second derivative.

For conceptual purposes, if we assume that a self-absorption feature is Gaussian $g(v_{\rm LSR})$, the second derivative of the feature will be

$$\frac{\mathrm{d}^2 g(v_{\rm LSR})}{\mathrm{d}v_{\rm LSR}^2} = \left(-\frac{1}{\sigma^2} + \frac{v_{\rm LSR}^2}{\sigma^4}\right) g(v_{\rm LSR})\,, \tag{B.2.1}$$

where σ is the standard deviation of the Gaussian, thus showing a strong dependence on line width. Narrow self-absorption features can then be identified through their second derivative more easily. In conclusion, the second derivative alone only works reliably well for high sensitivity, sufficient spectral resolution, and HISA line widths that are much smaller than the average emission component. Furthermore, even if the spectral ranges of HINSA (Li & Goldsmith, 2003; Goldsmith & Li, 2005; Goldsmith et al., 2007) were determined with second derivatives, the baselines would still need to be inferred using, for example, polynomial fits or making physical assumptions of the HINSA properties (see Krčo et al., 2008).

By introducing an asymmetry weighting and an optimized regularization term, that simultaneously mitigates the undesirable effect of noise fluctuations, astroSABER is able to recover baselines while identifying absorption dips without the necessity of assuming a fitting function or a tight physical correlation between the cold H I gas and the molecular gas.

B.3 Robustness of kinematics

To test how well the kinematics of the recovered absorption features match the input data, we ran astroSABER on an example data cube taken from a subsection of GMF20.0-17.9 (see Ragan et al., 2014; Syed et al., 2020). This example cube is also made available along with astroSABER source code. We created mock data as described in Sect. 5.2.2 containing 100 test spectra where known self-absorption have been added. We then ran astroSABER to extract the self-absorption baselines

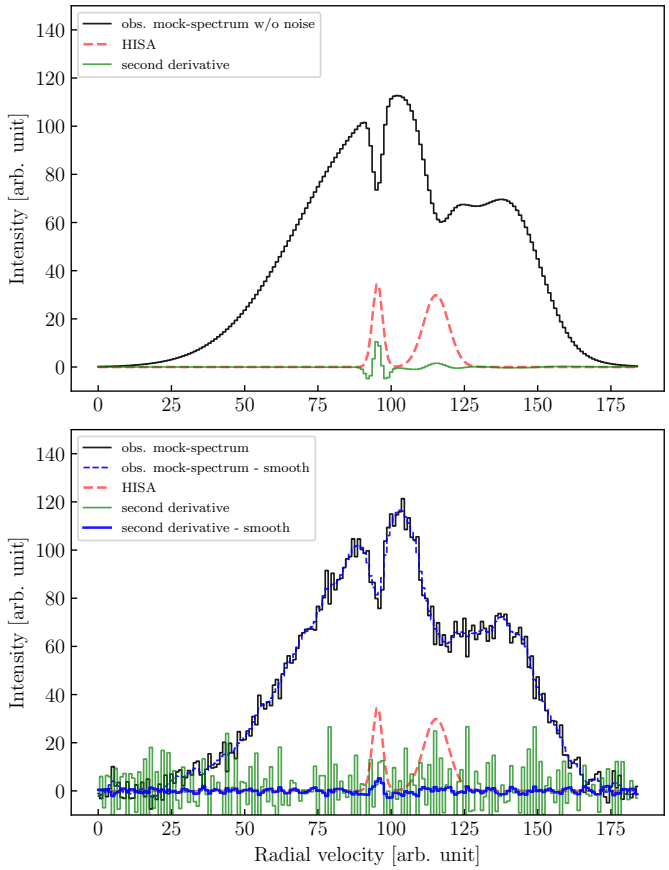


Figure B.2: Second derivative representation as a means to identify self-absorption. *Top panel:* The black mock spectrum represents the H I emission spectrum, with two self-absorption features superposed (red dashed components) and without any observational noise. The green spectrum shows the second derivative of the black mock spectrum, obtained from the finite differences between spectral channels. *Bottom panel:* The black mock spectrum represents the H I emission spectrum, with two self-absorption features superposed (red dashed components) and with added noise that is comparable to the noise of the THOR-H I observations (same spectrum as in Fig. 5.1). The green spectrum shows the second derivative of the black mock spectrum, obtained from the finite differences between spectral channels. The dashed blue spectrum represents a regularized least squares solution to the H I spectrum, which minimizes the second derivative. The corresponding second derivative is shown in blue, which is now less affected by noise fluctuations.

and spectra after finding the optimal smoothing parameters. To obtain the kinematic properties of the extracted self-absorption features, we fit several Gaussian components to the self-absorption spectra, depending on the number of components that were added. In total, 207 self-absorption components have been added while generating the mock spectra.

In Fig. B.3 we present histograms showing the residuals between the true amplitudes, centroid velocities, line widths (FWHM) and their respective fit results. All distributions show a mean around zero. The line widths show a weak systematic trend as the mean of the residuals is $0.25\,\mathrm{km\,s^{-1}}$, implying that the line width fits slightly underestimate the true line width. As expected, the amplitude distribution shows the largest dispersion as we use only one set of smoothing parameters for the entire region. The recovered centroid velocities are very robust as the histogram

shows a mean around zero and a standard deviation of 0.35 km s^{-1}.

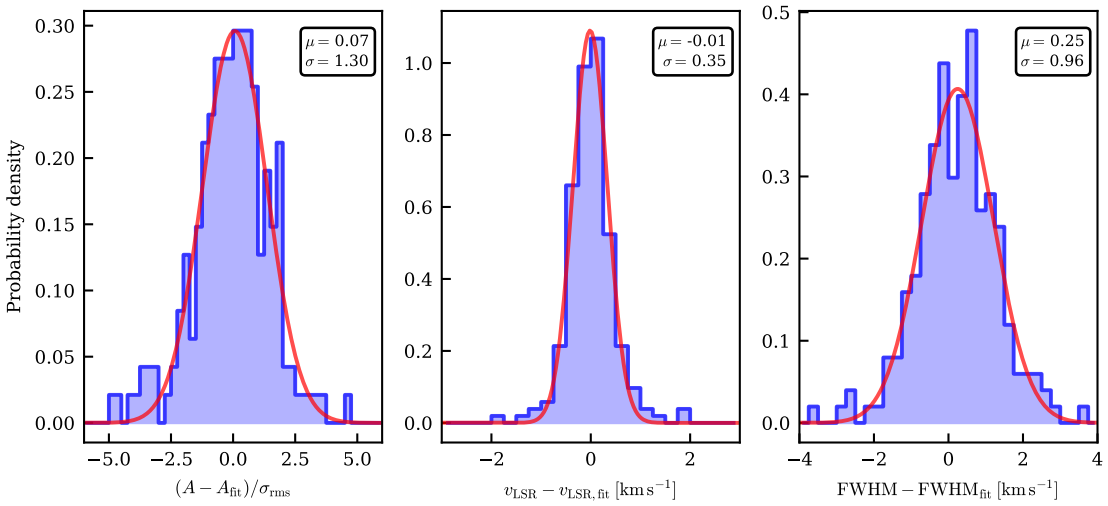


Figure B.3: Histograms of residuals between input features and their respective fit results. *Left panel:* The distribution shows the residuals between the amplitudes that were used to generate self-absorption features and the fit results (in units of the observational noise) after running astroSaber. *Middle panel:* The distribution shows the residuals between the input velocities of self-absorption features and the recovered fit velocities. *Right panel:* Similarly, the distribution in the right panel shows the residuals of the line widths. The red curve in each panel shows a Gaussian fit to the distribution.

B.4 Kinematics maps

The kinematic properties are presented in this section. The following maps show the fit peak velocities and line widths (FWHM) obtained with GaussPy+ for both HISA and ^{13}CO emission toward all remaining filament regions. If multiple components are identified within the velocity range of the filament, we only show the component with the lowest peak velocity.

B.5 Column density maps

The column density maps are presented in this section. The following maps show the column density maps for both HISA and H_2 as traced by ^{13}CO emission integrated over the velocity range of the respective filament region. Details about the column density derivation of each tracer can be found in Sect. 5.3.2.

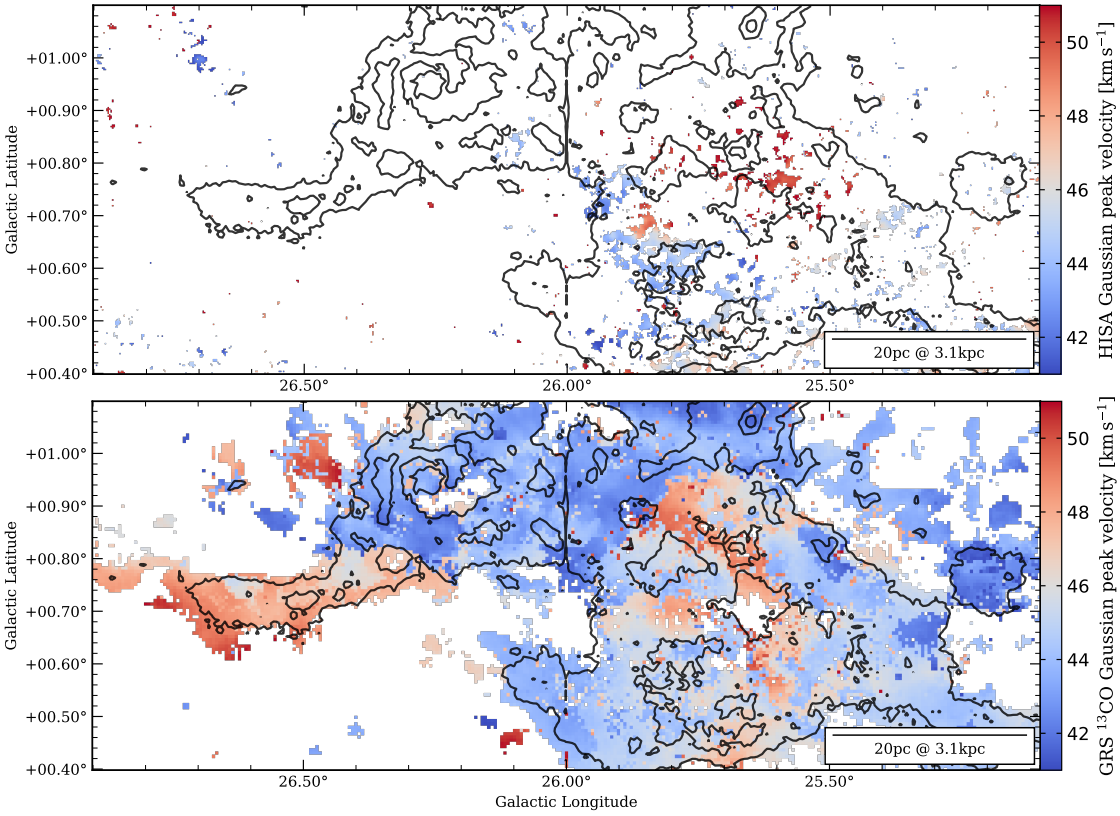


Figure B.4: Fit peak velocity toward GMF26. These maps show the peak velocities of fit components derived from the GaussPy+ decomposition of the spectra. If multiple components are present in a single pixel spectrum within the velocity range of the filament region, the component with the lowest peak velocity is shown. The black contours in both panels show the integrated GRS ^{13}CO emission at the levels 6.0, 12.0, 24.0, and 34.0 K km s^{-1}. The contour feature at longitude $\ell = 26°$ is an artifact in the observational data. *Top panel:* Fit HISA peak velocity. *Bottom panel:* Fit ^{13}CO peak velocity.

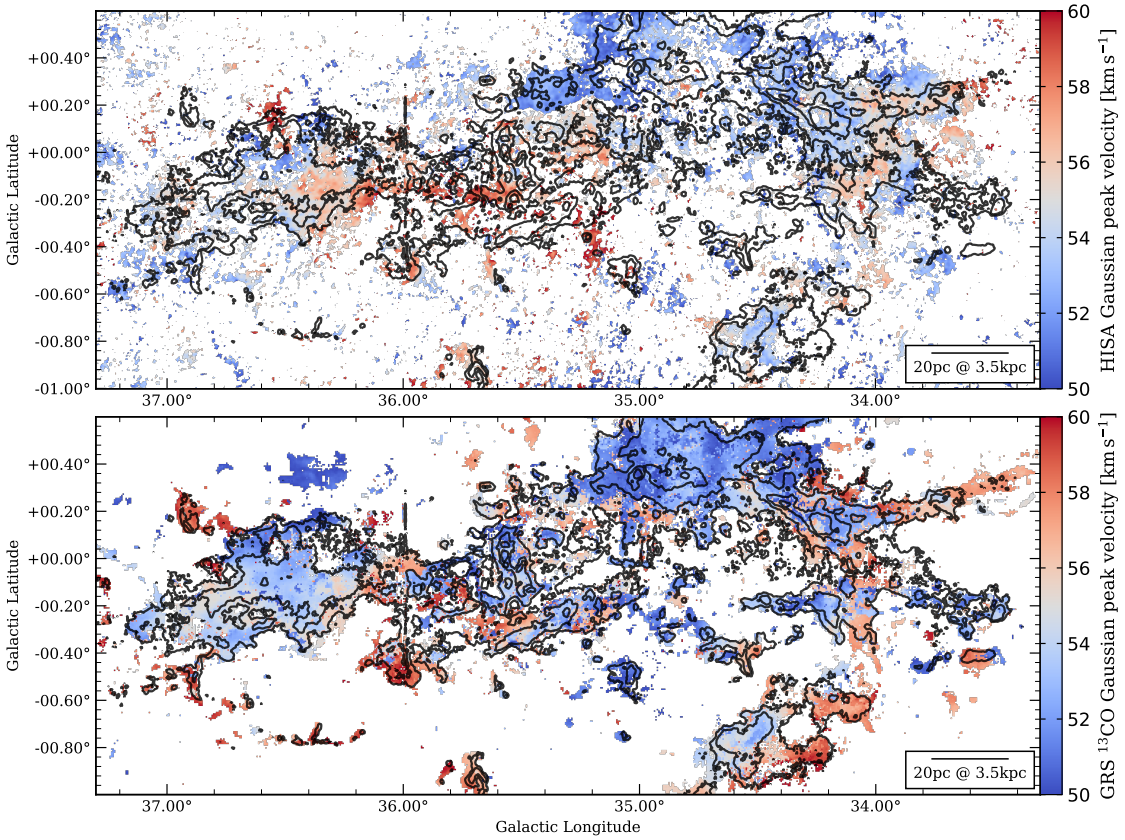


Figure B.5: Fit peak velocity toward GMF38a. These maps show the peak velocities of fit components derived from the GaussPy+ decomposition of the spectra. If multiple components are present in a single pixel spectrum within the velocity range of the filament region, the component with the lowest peak velocity is shown. The black contours in both panels show the integrated GRS ^{13}CO emission at the levels 5.0, 10.0, 20.0, and 30.0 K km s^{-1}. The contour feature at longitude $\ell = 36°$ is an artifact in the observational data. *Top panel:* Fit HISA peak velocity. *Bottom panel:* Fit ^{13}CO peak velocity.

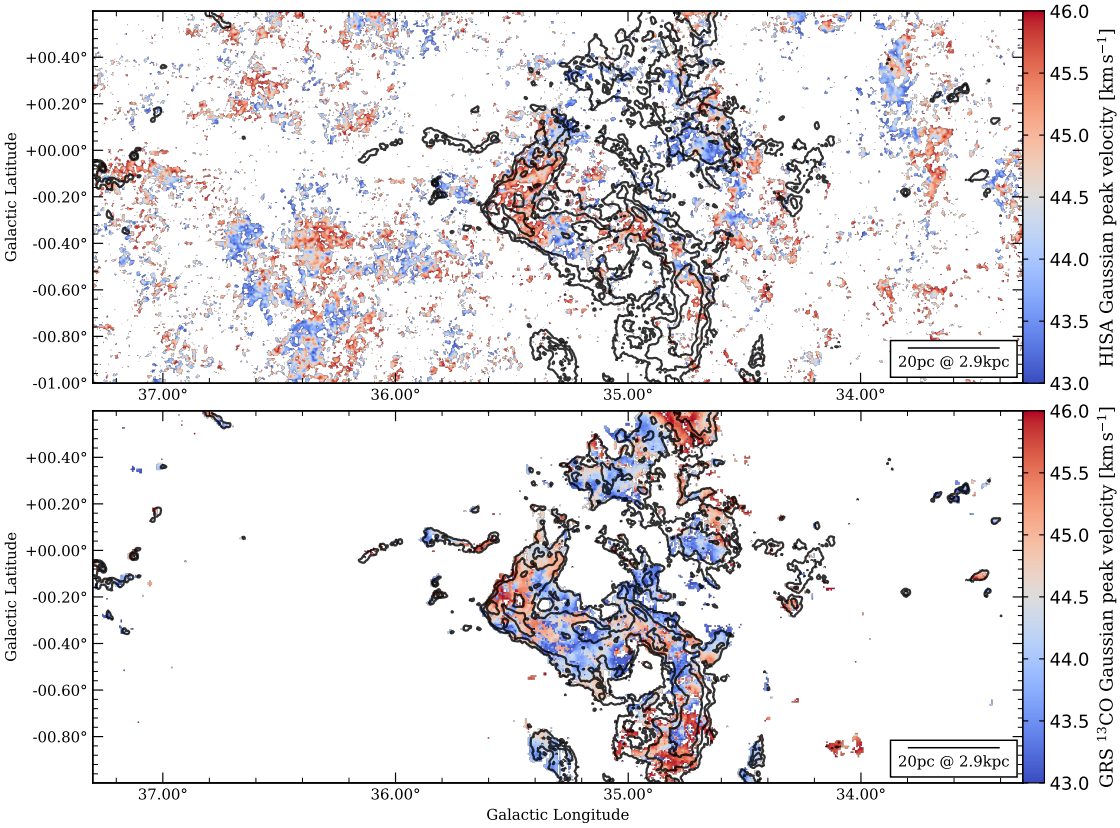


Figure B.6: Fit peak velocity toward GMF38b. These maps show the peak velocities of fit components derived from the GaussPy+ decomposition of the spectra. If multiple components are present in a single pixel spectrum within the velocity range of the filament region, the component with the lowest peak velocity is shown. The black contours in both panels show the integrated GRS ^{13}CO emission at the levels 2.5, 5.0, 10.0, and 20.0 K km s^{-1}. *Top panel:* Fit HISA peak velocity. *Bottom panel:* Fit ^{13}CO peak velocity.

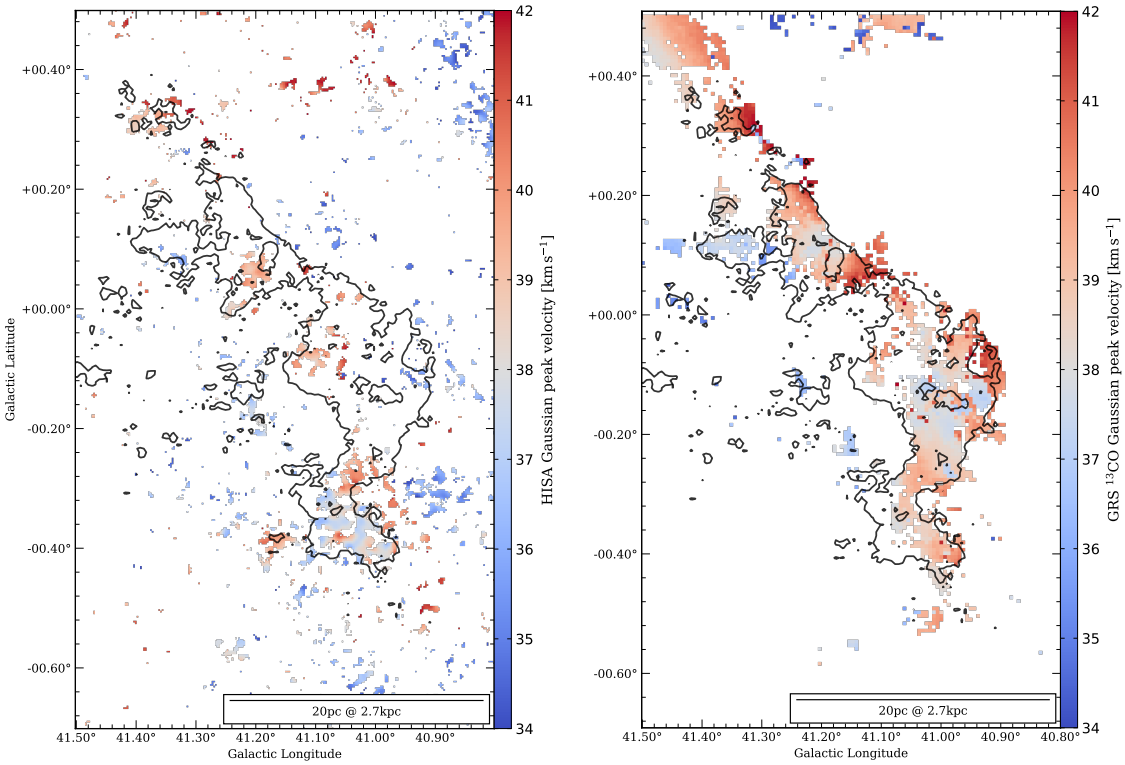


Figure B.7: Fit peak velocity toward GMF41. These maps show the peak velocities of fit components derived from the GaussPy+ decomposition of the spectra. If multiple components are present in a single pixel spectrum within the velocity range of the filament region, the component with the lowest peak velocity is shown. The black contours in both panels show the integrated GRS ^{13}CO emission at the levels 6.0, 12.0, 24.0, and 34.0 K km s^{-1}. *Top panel:* Fit HISA peak velocity. *Bottom panel:* Fit ^{13}CO peak velocity.

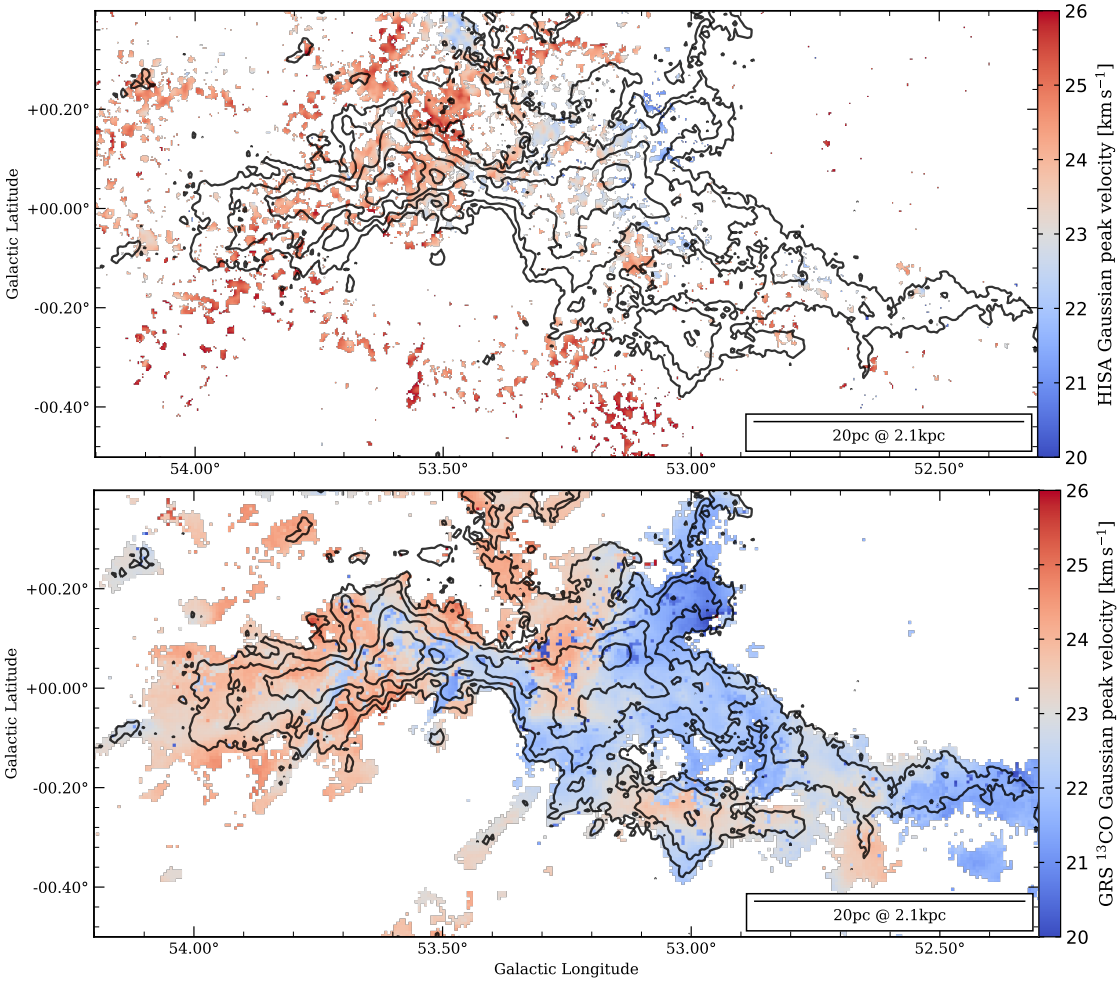


Figure B.8: Fit peak velocity toward GMF54. These maps show the peak velocities of fit components derived from the GaussPy+ decomposition of the spectra. If multiple components are present in a single pixel spectrum within the velocity range of the filament region, the component with the lowest peak velocity is shown. The black contours in both panels show the integrated GRS ^{13}CO emission at the levels 2.5, 5.0, 10.0, and 20.0 K km s^{-1}. *Top panel:* Fit HISA peak velocity. *Bottom panel:* Fit ^{13}CO peak velocity.

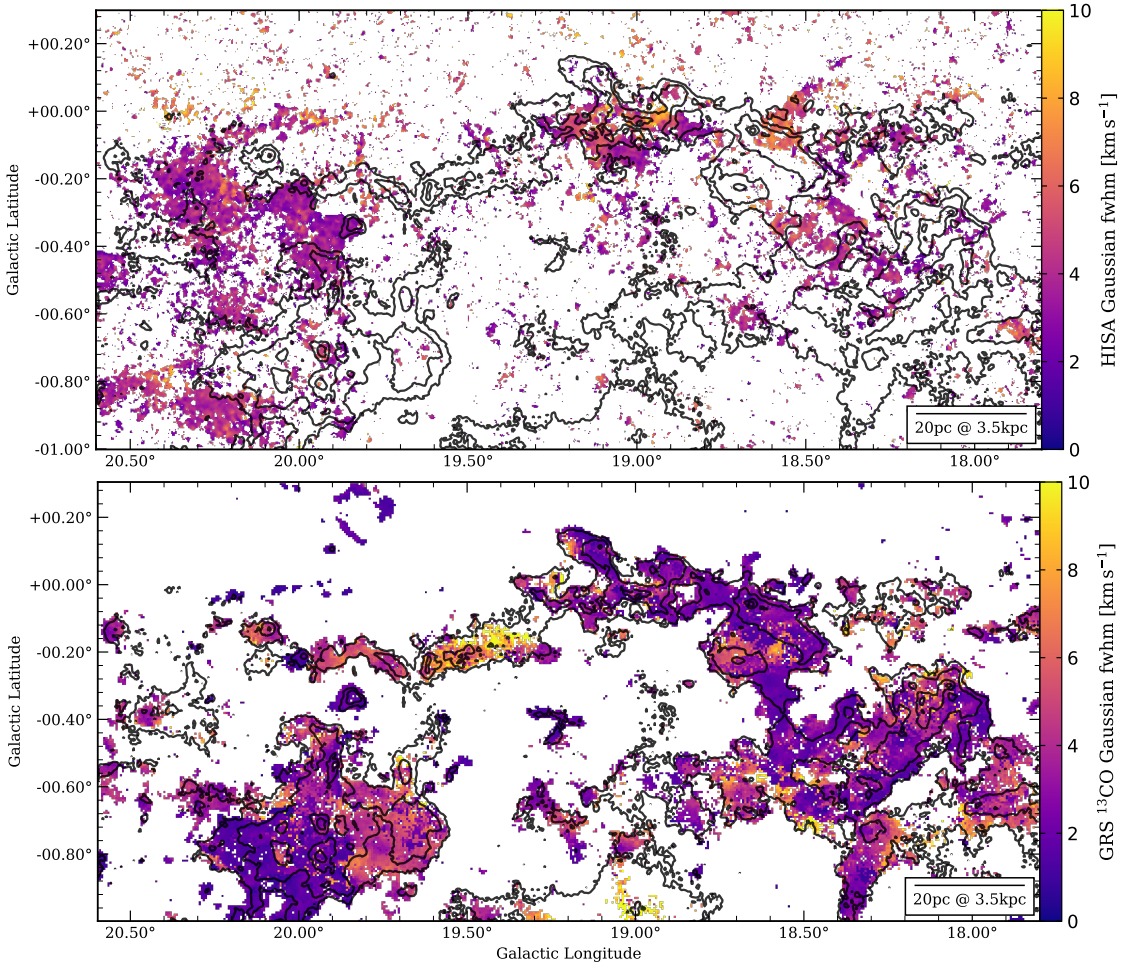


Figure B.9: Fit line width (FWHM) toward GMF20. These maps show the line widths of fit components derived from the GaussPy+ decomposition of the spectra. If multiple components are present in a single pixel spectrum within the velocity range of the filament region, the component with the lowest peak velocity is shown. The black contours in both panels show the integrated GRS ^{13}CO emission at the levels 8.0, 16.0, 32.0, and 42.0 K km s^{-1}. *Top panel:* Fit HISA line width. *Bottom panel:* Fit ^{13}CO line width.

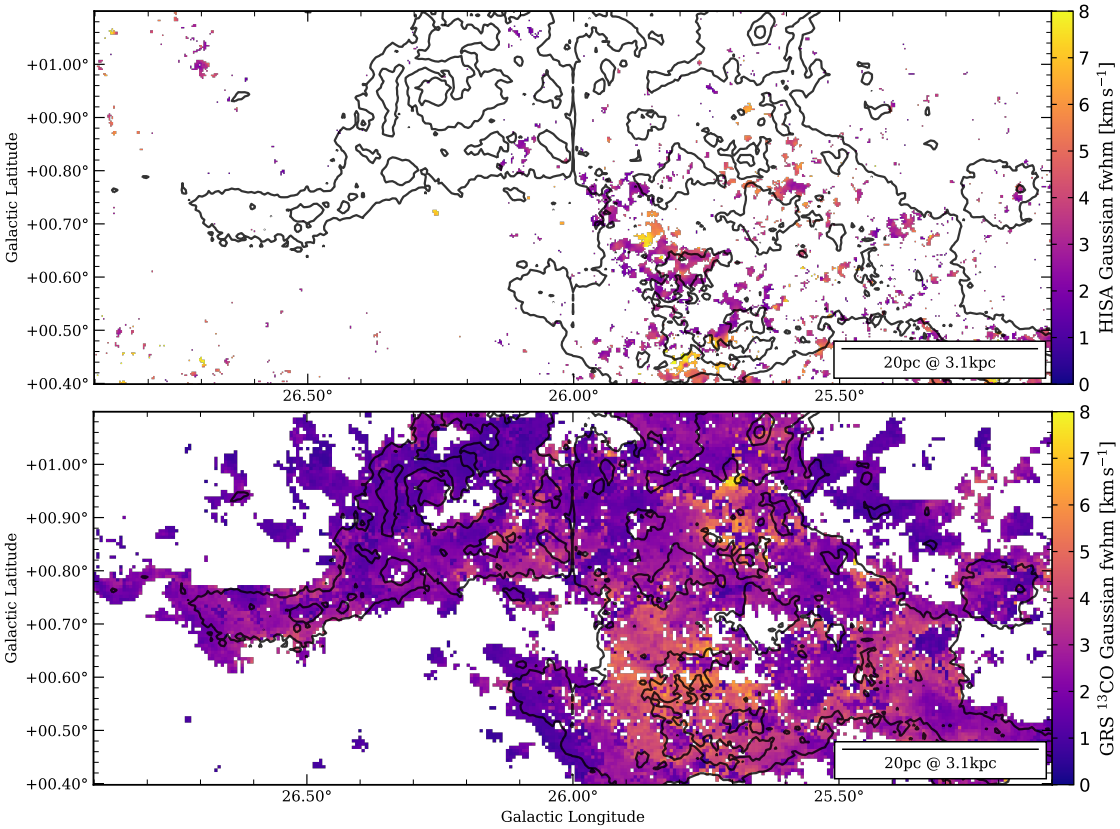


Figure B.10: Fit line width (FWHM) toward GMF26. These maps show the line widths of fit components derived from the GaussPy+ decomposition of the spectra. If multiple components are present in a single pixel spectrum within the velocity range of the filament region, the component with the lowest peak velocity is shown. The black contours in both panels show the integrated GRS ^{13}CO emission at the levels 6.0, 12.0, 24.0, and 34.0 K km s^{-1}. The contour feature at longitude $\ell = 26°$ is an artifact in the observational data. *Top panel:* Fit HISA line width. *Bottom panel:* Fit ^{13}CO line width.

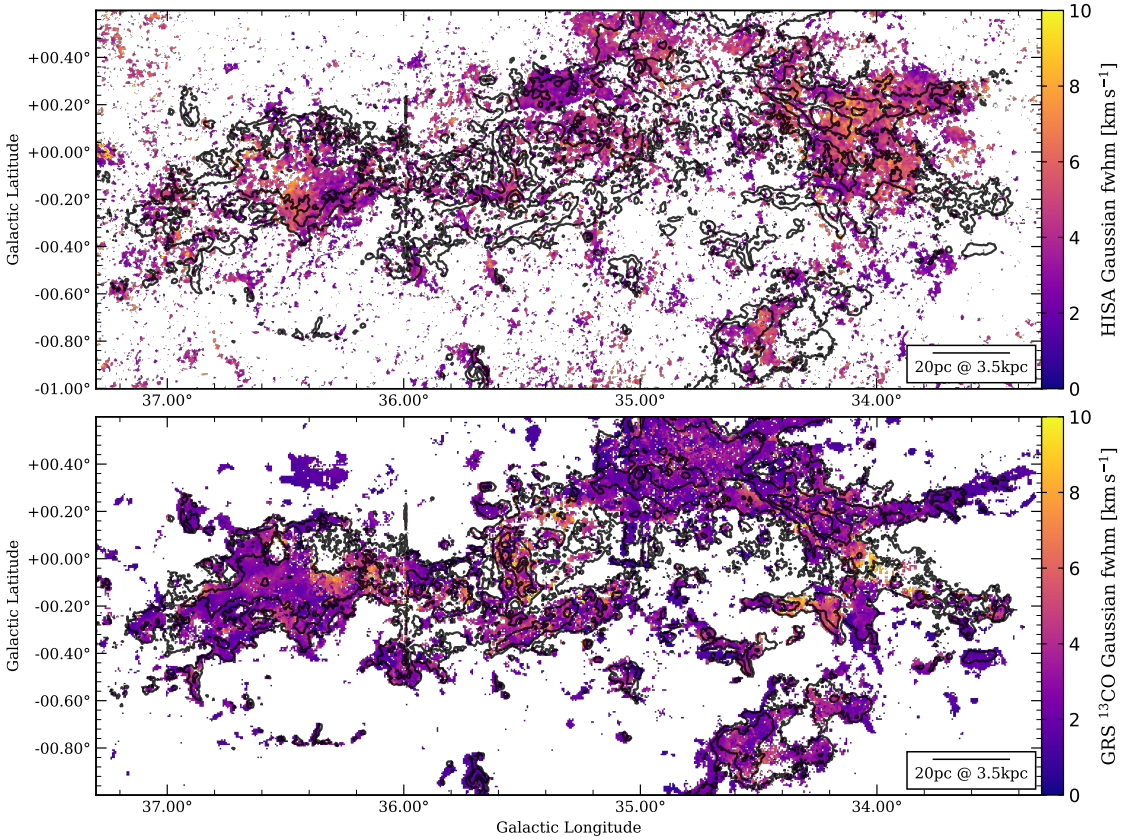


Figure B.11: Fit line width (FWHM) toward GMF38a. These maps show the line widths of fit components derived from the GaussPy+ decomposition of the spectra. If multiple components are present in a single pixel spectrum within the velocity range of the filament region, the component with the lowest peak velocity is shown. The black contours in both panels show the integrated GRS ^{13}CO emission at the levels 5.0, 10.0, 20.0, and 30.0 K km s^{-1}. The contour feature at longitude $\ell = 36°$ is an artifact in the observational data. *Top panel:* Fit HISA line width. *Bottom panel:* Fit ^{13}CO line width.

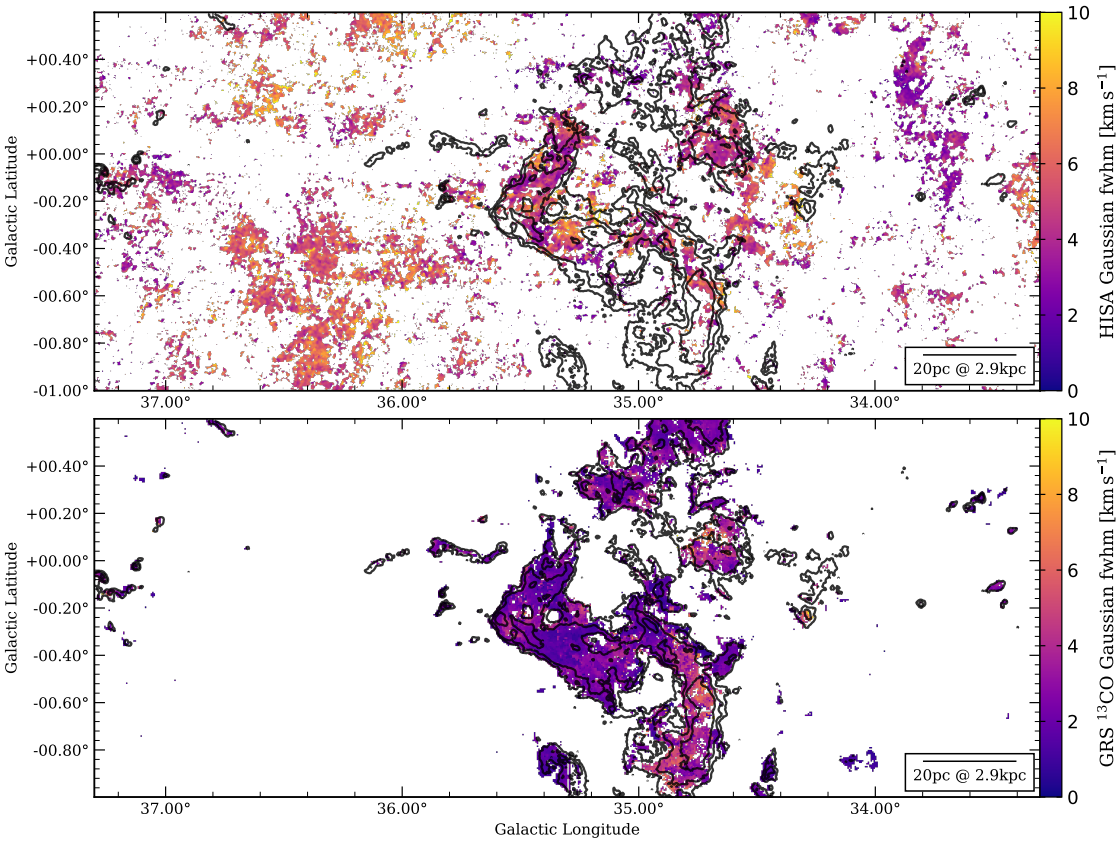


Figure B.12: Fit line width (FWHM) toward GMF38b. These maps show the line widths of fit components derived from the GaussPy+ decomposition of the spectra. If multiple components are present in a single pixel spectrum within the velocity range of the filament region, the component with the lowest peak velocity is shown. The black contours in both panels show the integrated GRS ^{13}CO emission at the levels 2.5, 5.0, 10.0, and 20.0 K km s^{-1}. *Top panel:* Fit HISA line width. *Bottom panel:* Fit ^{13}CO line width.

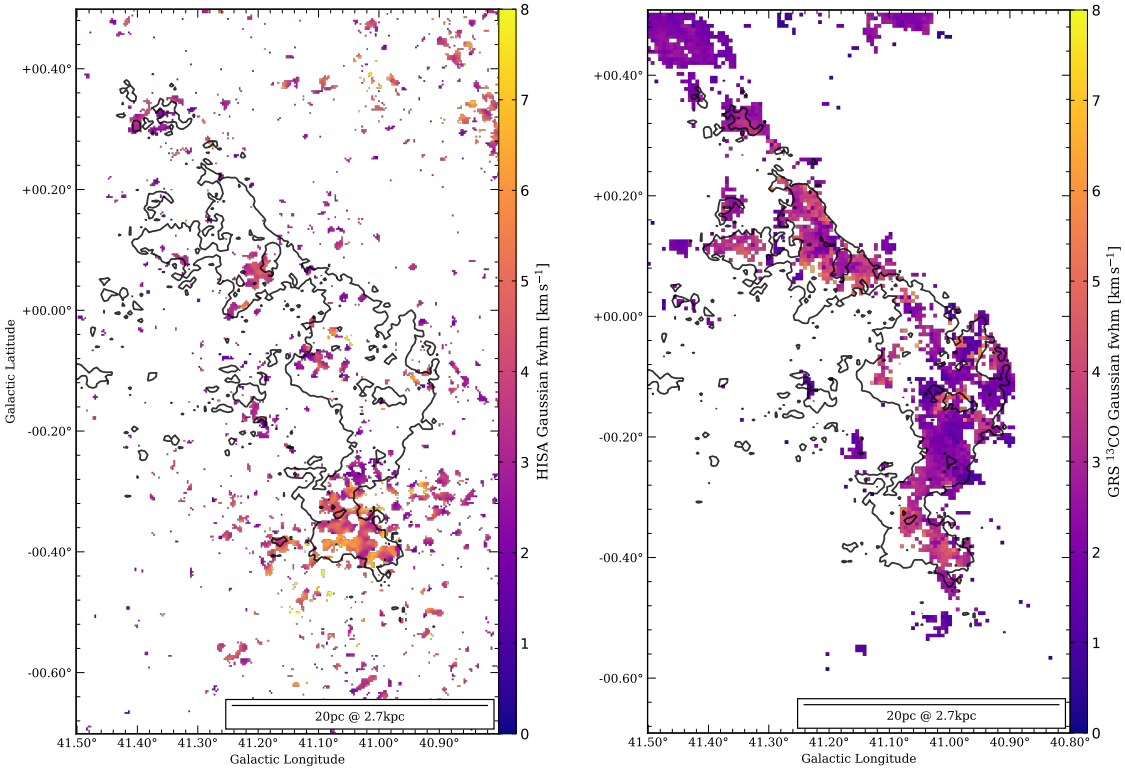


Figure B.13: Fit line width (FWHM) toward GMF41. These maps show the line widths of fit components derived from the GaussPy+ decomposition of the spectra. If multiple components are present in a single pixel spectrum within the velocity range of the filament region, the component with the lowest peak velocity is shown. The black contours in both panels show the integrated GRS ^{13}CO emission at the levels 6.0, 12.0, 24.0, and 34.0 K km s^{-1}. *Top panel:* Fit HISA line width. *Bottom panel:* Fit ^{13}CO line width.

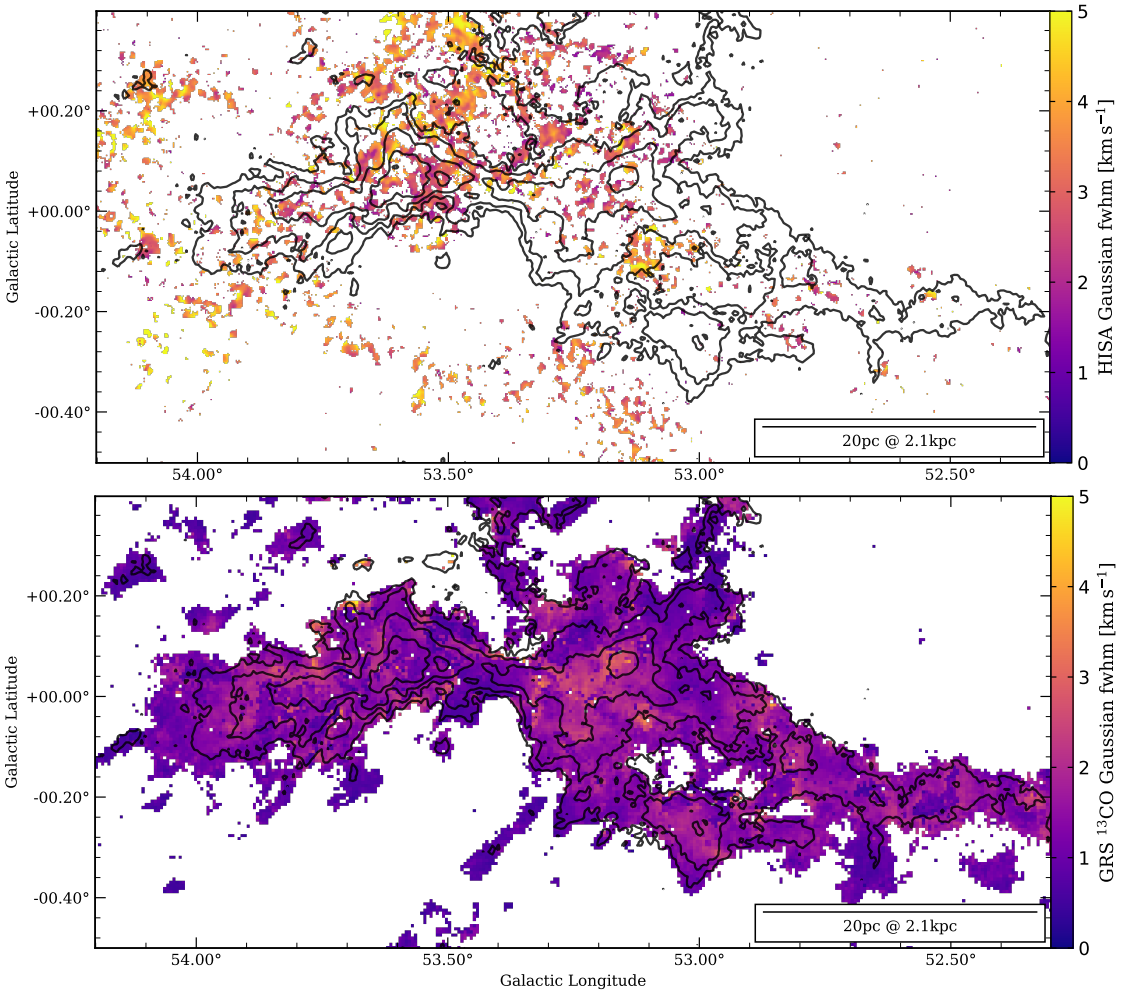


Figure B.14: Fit line width (FWHM) toward GMF54. These maps show the line widths of fit components derived from the GaussPy+ decomposition of the spectra. If multiple components are present in a single pixel spectrum within the velocity range of the filament region, the component with the lowest peak velocity is shown. The black contours in both panels show the integrated GRS ^{13}CO emission at the levels 2.5, 5.0, 10.0, and 20.0 K km s^{-1}. *Top panel:* Fit HISA line width. *Bottom panel:* Fit ^{13}CO line width.

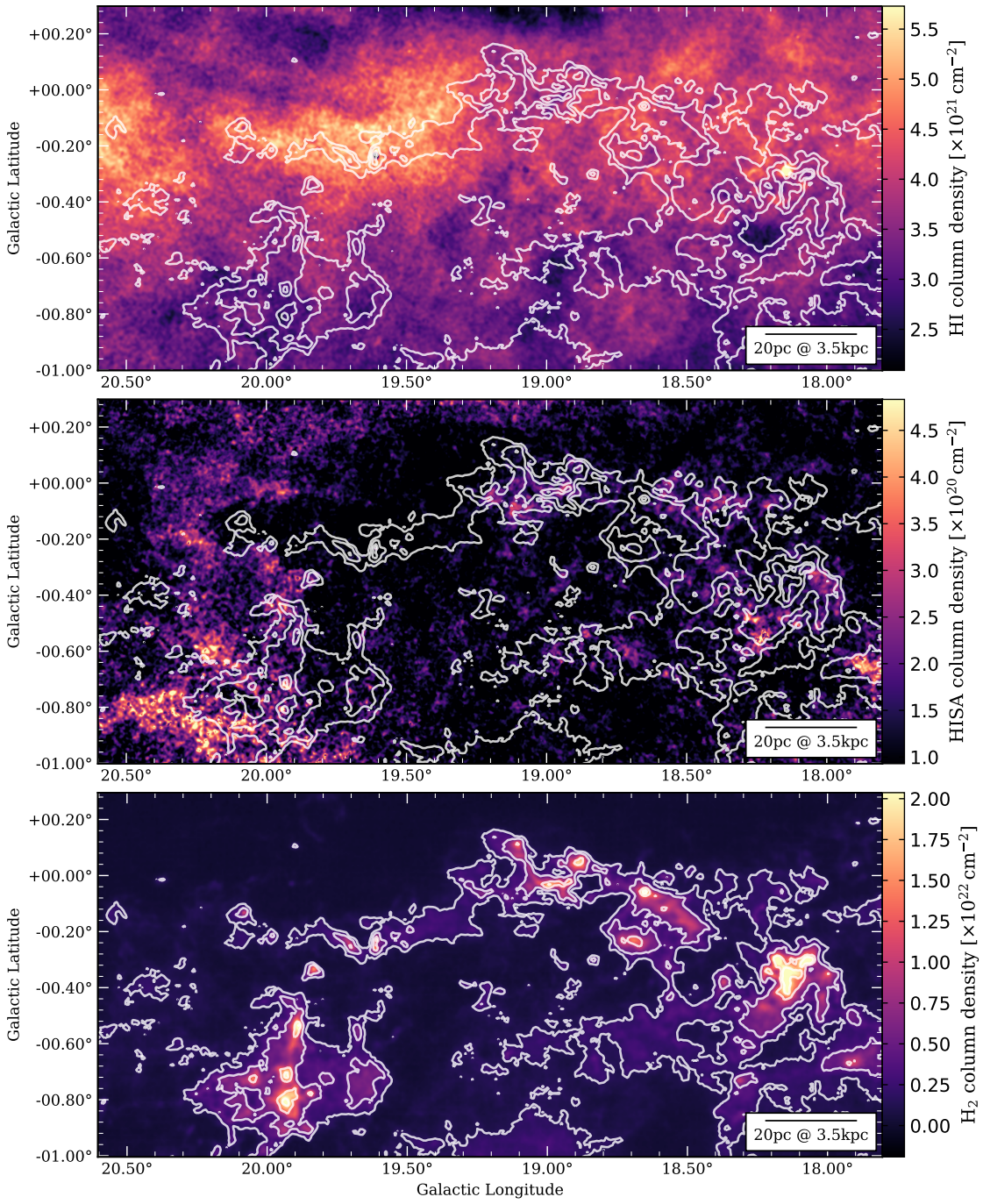


Figure B.15: Column density toward GMF20. These maps show the column densities of atomic hydrogen traced by H I emission, the cold hydrogen gas traced by HISA, and molecular hydrogen traced by ^{13}CO emission, respectively. The column densities are integrated over the velocity range of the filament region given in Table 5.1. The white contours in both panels show the integrated MWISP ^{13}CO emission at the levels 8.0, 16.0, 32.0, and 42.0 K km s^{-1}. *Top panel:* H I column density traced by H I emission. *Middle panel:* HISA column density. *Bottom panel:* H_2 column density traced by ^{13}CO.

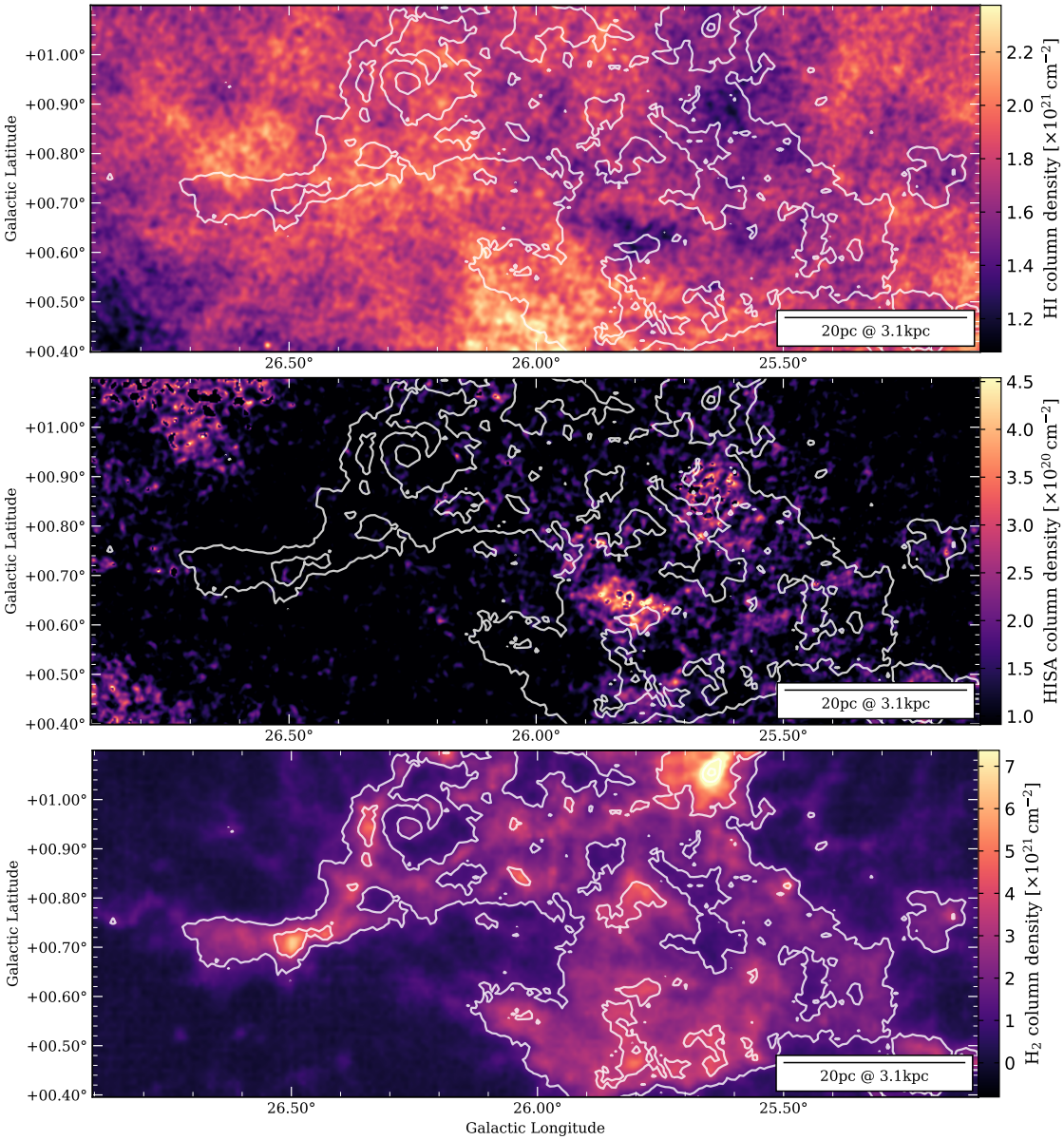


Figure B.16: Column density toward GMF26. These maps show the column densities of atomic hydrogen traced by H I emission, the cold hydrogen gas traced by HISA, and molecular hydrogen traced by ^{13}CO emission, respectively. The column densities are integrated over the velocity range of the filament region given in Table 5.1. The white contours in both panels show the integrated MWISP ^{13}CO emission at the levels 6.0, 12.0, 24.0, and 34.0 K km s^{-1}. *Top panel:* H I column density traced by H I emission. *Middle panel:* HISA column density. *Bottom panel:* H_2 column density traced by ^{13}CO.

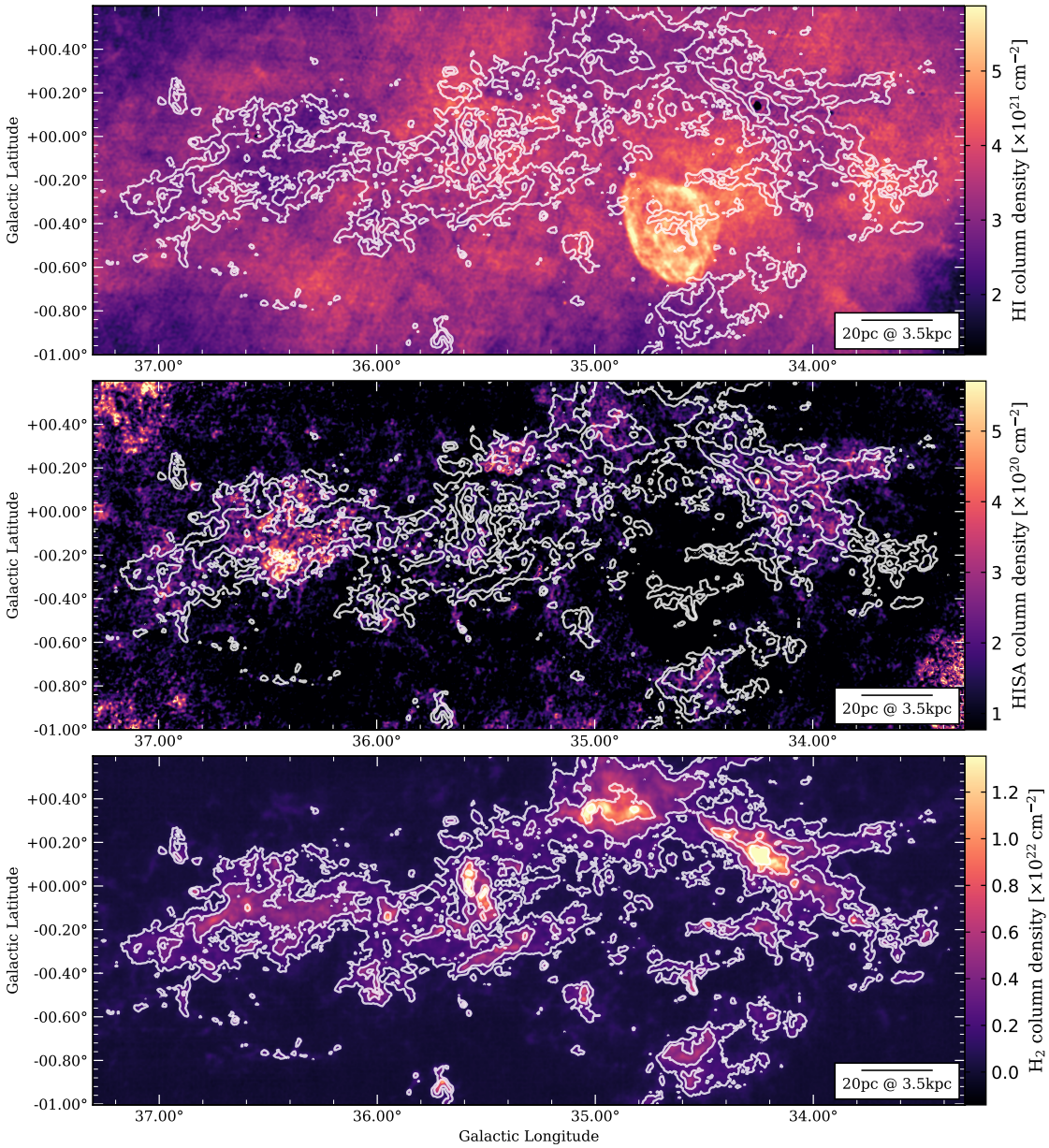


Figure B.17: Column density toward GMF38a. These maps show the column densities of atomic hydrogen traced by H I emission, the cold hydrogen gas traced by HISA, and molecular hydrogen traced by ^{13}CO emission, respectively. The column densities are integrated over the velocity range of the filament region given in Table 5.1. The white contours in both panels show the integrated MWISP ^{13}CO emission at the levels 5.0, 10.0, 20.0, and 30.0 K km s^{-1}. *Top panel:* H I column density traced by H I emission. *Middle panel:* HISA column density. *Bottom panel:* H_2 column density traced by ^{13}CO.

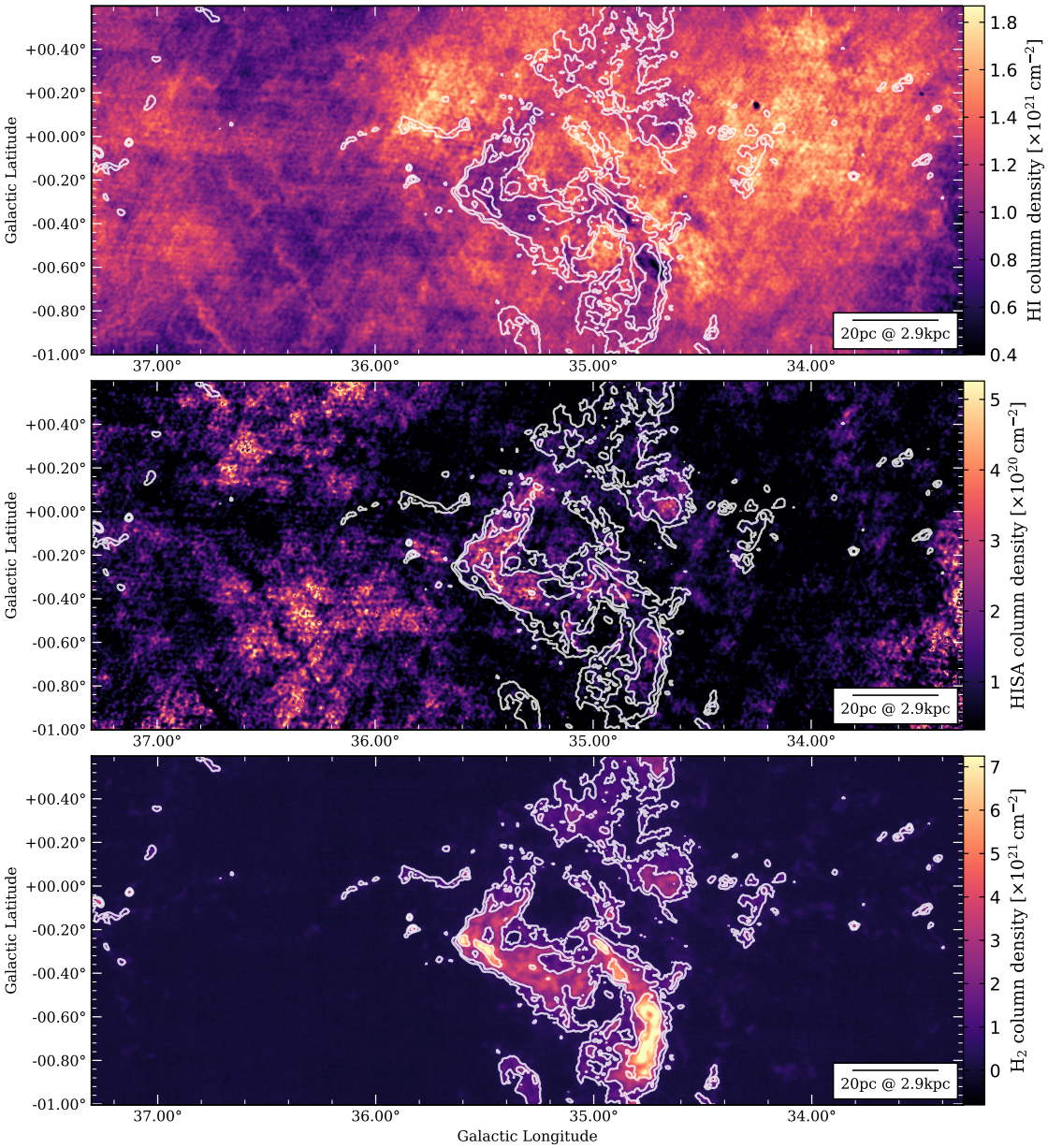


Figure B.18: Column density toward GMF38b. These maps show the column densities of atomic hydrogen traced by H I emission, the cold hydrogen gas traced by HISA, and molecular hydrogen traced by ^{13}CO emission, respectively. The column densities are integrated over the velocity range of the filament region given in Table 5.1. The white contours in both panels show the integrated MWISP ^{13}CO emission at the levels 2.5, 5.0, 10.0, and 20.0 K km s^{-1}. *Top panel:* H I column density traced by H I emission. *Middle panel:* HISA column density. *Bottom panel:* H_2 column density traced by ^{13}CO.

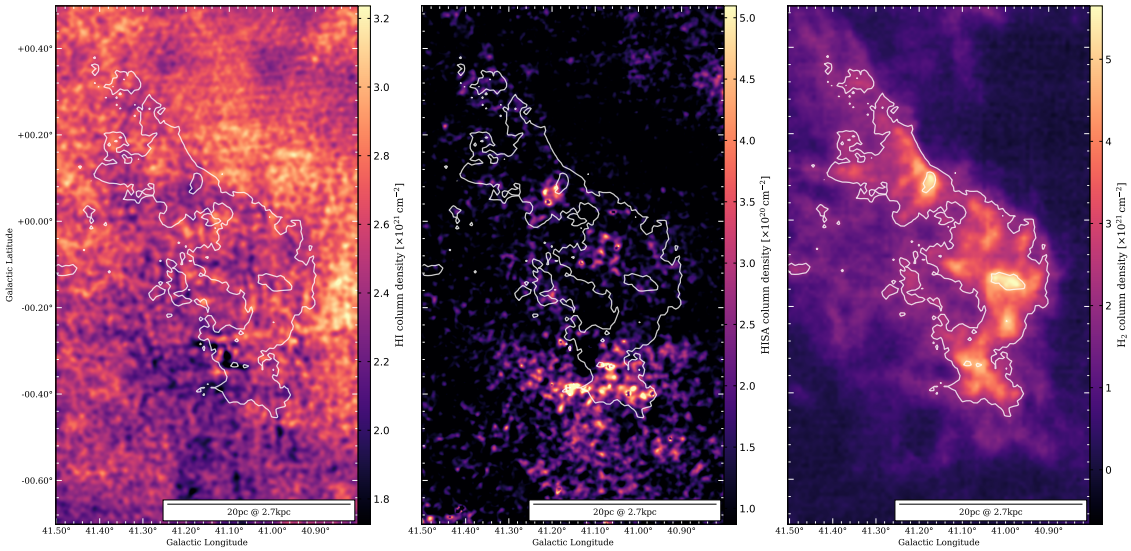


Figure B.19: Column density toward GMF41. These maps show the column densities of atomic hydrogen traced by H I emission, the cold hydrogen gas traced by HISA, and molecular hydrogen traced by ^{13}CO emission, respectively. The column densities are integrated over the velocity range of the filament region given in Table 5.1. The white contours in both panels show the integrated MWISP ^{13}CO emission at the levels 6.0, 12.0, 24.0, and 34.0 K km s^{-1}. *Left panel:* H I column density traced by H I emission. *Middle panel:* HISA column density. *Right panel:* H_2 column density traced by ^{13}CO.

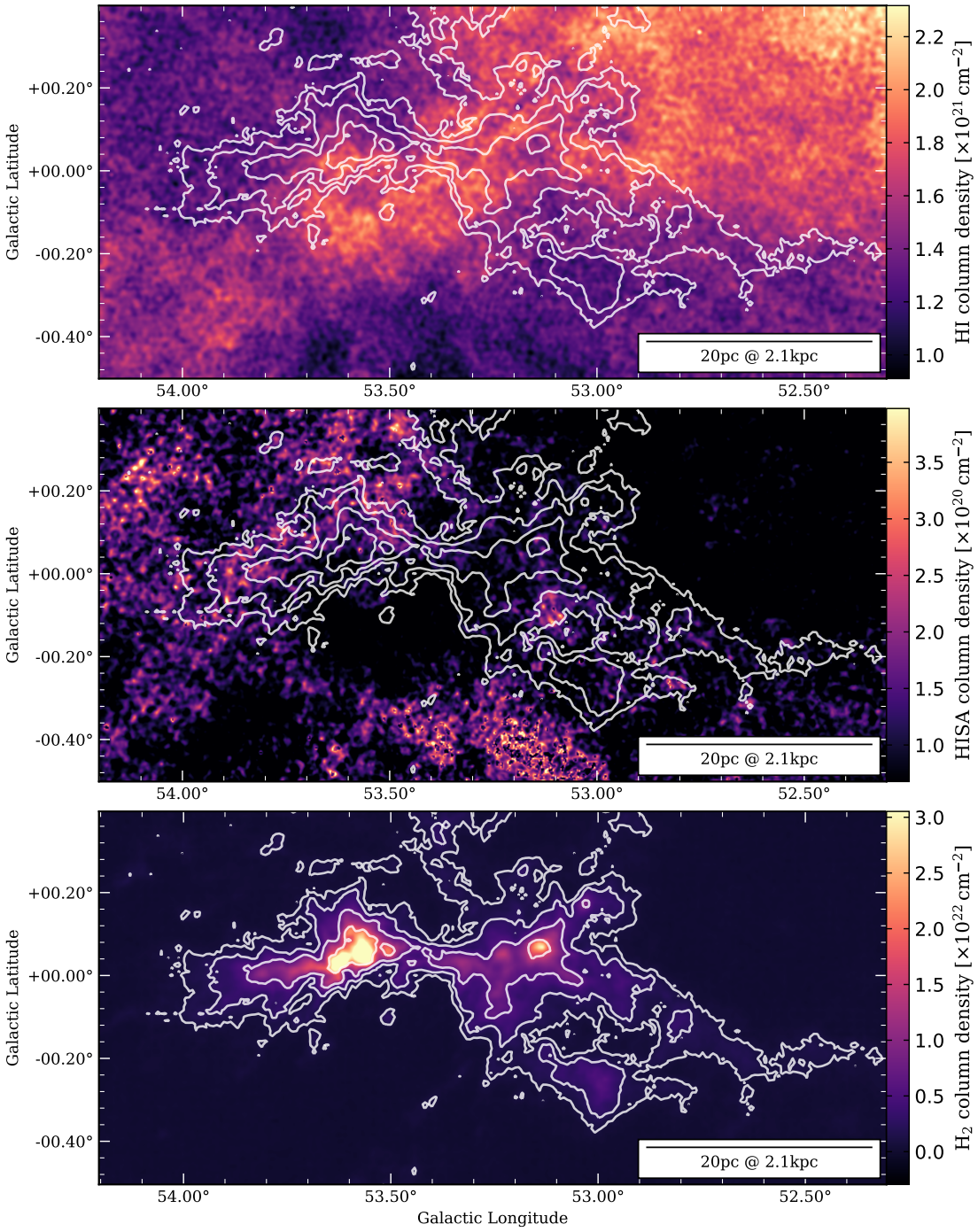


Figure B.20: Column density toward GMF54. These maps show the column densities of atomic hydrogen traced by H I emission, the cold hydrogen gas traced by HISA, and molecular hydrogen traced by ^{13}CO emission, respectively. The column densities are integrated over the velocity range of the filament region given in Table 5.1. The white contours in both panels show the integrated MWISP ^{13}CO emission at the levels 2.5, 5.0, 10.0, and 20.0 K km s^{-1}. *Top panel:* H I column density traced by H I emission. *Middle panel:* HISA column density. *Bottom panel:* H_2 column density traced by ^{13}CO.

List of Figures

List of Tables

Publication list

The following published paper is part of this thesis (Chapter 4):

- **Syed, J.**, Soler, J. D., Beuther, H., et al. 2022, A&A, 657, A1:
  The "Maggie" filament: Physical properties of a giant atomic cloud

The following paper is also part of this thesis (Chapter 5) and is currently under review:

- **Syed, J.**, Beuther, H., Goldsmith, P. F., et al. 2023, submitted to A&A:
  Cold atomic gas identified by H I *self-absorption – Cold atomic clouds toward giant molecular filaments*

The following paper is also part of this thesis (Chapter 6) and is in preparation:

- **Syed, J.** & Beuther, H. 2023 in prep.:
  Cold atomic gas identified by H I *self-absorption – Cold atomic clouds in the inner Galactic plane*

The following published paper is **not** part of this thesis:

- **Syed, J.**, Wang, Y., Beuther, H., et al. 2020, A&A, 642, A68:
  Atomic and molecular gas properties during cloud formation

The following co-authored publications are all **not** part of this thesis:

- Dietrich, J., Apai, D., Schlecker, M. et al., **incl. Syed, J.** 2023, AJ, 165, 149:
  EDEN Survey: Small Transiting Planet Detection Limits and Constraints on the Occurrence Rates of Planets around Late-M Dwarfs within 15 pc

- Basu, A., Roy, N., Beuther, H., et al., **incl. Syed, J.** 2022, MNRAS, 517, 5063:
  Properties of atomic hydrogen gas in the Galactic plane from THOR 21-cm absorption spectra: a comparison with the high latitude gas

- Shanahan, R., Stil, J. M., Anderson, L., et al., **incl. Syed, J.** 2022, ApJ, 939, 92:
  Polarized Emission from Four Supernova Remnants in the THOR Survey

- Seifried, D, Beuther, H., Walch, S., et al., **incl. Syed, J.** 2022, MNRAS, 512, 4765:
  On the accuracy of H I observations in molecular clouds - More cold H I than thought?

- Gieser, C., Beuther, H., Semenov, D., et al., **incl. Syed, J.** 2022, A&A, 657, A3:
  Clustered star formation at early evolutionary stages. Physical and chemical analysis of the young star-forming regions ISOSS J22478+6357 and ISOSS J23053+5953

- Soler, J. D., Beuther, H., **Syed, J.**, et al. 2021, A&A, 651, L4:
  The filamentary structures in the CO emission toward the Milky Way disk

- Beuther, H., Gieser, C., Suri, S., et al., **incl. Syed, J.** 2021, A&A, 649, A113:
  Fragmentation and kinematics in high-mass star formation. CORE-extension targeting two very young high-mass star-forming regions

- Soler, J. D., Beuther, H., **Syed, J.**, et al. 2020, A&A, 642, A163:
  The history of dynamics and stellar feedback revealed by the H I filamentary structure in the disk of the Milky Way

Bibliography

Abreu-Vicente, J., Ragan, S., Kainulainen, J., et al. 2016, A&A, 590, A131
Allen, C. W. 1973, Astrophysical quantities
Alstott, J., Bullmore, E., & Plenz, D. 2014, PLOS ONE, 9, 1
Alves, J., Zucker, C., Goodman, A. A., et al. 2020, Nature, 578, 237
Anderson, L. D. & Bania, T. M. 2009, ApJ, 690, 706
Andersson, B. G., Wannier, P. G., & Morris, M. 1991, ApJ, 366, 464
Arzoumanian, D., André, P., Didelon, P., et al. 2011, A&A, 529, L6
Baek, S.-J., Park, A., Ahn, Y.-J., & Choo, J. 2015, Analyst, 140, 250
Beattie, J. R., Mocz, P., Federrath, C., & Klessen, R. S. 2022, MNRAS, 517, 5003
Bellomi, E., Godard, B., Hennebelle, P., et al. 2020, A&A, 643, A36
Bergemann, M., Sesar, B., Cohen, J. G., et al. 2018, Nature, 555, 334
Beuther, H., Bihr, S., Rugel, M., et al. 2016, A&A, 595, A32
Bialy, S., Bihr, S., Beuther, H., Henning, T., & Sternberg, A. 2017a, ApJ, 835, 126
Bialy, S., Burkhart, B., & Sternberg, A. 2017b, ApJ, 843, 92
Bialy, S. & Sternberg, A. 2016, ApJ, 822, 83
Bialy, S. & Sternberg, A. 2019, ApJ, 881, 160
Bihr, S., Beuther, H., Ott, J., et al. 2015, A&A, 580, A112
Bihr, S., Johnston, K. G., Beuther, H., et al. 2016, A&A, 588, A97
Brand, J. & Blitz, L. 1993, A&A, 275, 67
Burkhart, B., Lee, M.-Y., Murray, C. E., & Stanimirović, S. 2015, ApJ, 811, L28
Burkhart, B., Stalpes, K., & Collins, D. C. 2017, ApJ, 834, L1
Burton, W. B. 1972, A&A, 19, 51
Burton, W. B. 1988, The structure of our Galaxy derived from observations of neutral hydrogen., ed. K. I. Kellermann & G. L. Verschuur (Springer New York), 295–358
Burton, W. B. & Liszt, H. S. 1993, A&A, 274, 765
Burton, W. B. & te Lintel Hekkert, P. 1986, A&AS, 65, 427
Chandrasekhar, S. & Fermi, E. 1953, ApJ, 118, 116
Chequers, M. H. & Widrow, L. M. 2017, MNRAS, 472, 2751
Chira, R. A., Ibáñez-Mejía, J. C., Mac Low, M. M., & Henning, T. 2019, A&A, 630, A97
Choudhuri, S. & Roy, N. 2019, MNRAS, 483, 3437
Churchwell, E., Babler, B. L., Meade, M. R., et al. 2009, PASP, 121, 213
Clark, P. C., Glover, S. C. O., Klessen, R. S., & Bonnell, I. A. 2012, MNRAS, 424, 2599

Colombo, D., Rosolowsky, E., Duarte-Cabral, A., et al. 2019, MNRAS, 483, 4291
Colombo, D., Rosolowsky, E., Ginsburg, A., Duarte-Cabral, A., & Hughes, A. 2015, MNRAS, 454, 2067
Condon, J. J. & Ransom, S. M. 2016, Essential Radio Astronomy (Kassel: Princeton University Press)
Cox, A. N. 2000, Allen's astrophysical quantities
Crutcher, R. M. & Riegel, K. W. 1974, ApJ, 188, 481
Dame, T. M., Hartmann, D., & Thaddeus, P. 2001, ApJ, 547, 792
Dénes, H., McClure-Griffiths, N. M., Dickey, J. M., Dawson, J. R., & Murray, C. E. 2018, MNRAS, 479, 1465
Dickey, J. M. 2013, in Planets, Stars and Stellar Systems. Volume 5: Galactic Structure and Stellar Populations, ed. T. D. Oswalt & G. Gilmore, Vol. 5, 549
Dickey, J. M. & Lockman, F. J. 1990, ARA&A, 28, 215
Dickey, J. M., McClure-Griffiths, N. M., Gaensler, B. M., & Green, A. J. 2003, ApJ, 585, 801
Dickey, J. M., Mebold, U., Stanimirovic, S., & Staveley-Smith, L. 2000, ApJ, 536, 756
Dickey, J. M., Strasser, S., Gaensler, B. M., et al. 2009, ApJ, 693, 1250
Dobbs, C. L. 2015, MNRAS, 447, 3390
Draine, B. T. 1978, ApJS, 36, 595
Draine, B. T. 2011, Physics of the Interstellar and Intergalactic Medium - (Kassel: Princeton University Press)
Duarte-Cabral, A., Colombo, D., Urquhart, J. S., et al. 2021, The SEDIGISM survey: molecular clouds in the inner Galaxy, Monthly Notices of the Royal Astronomical Society, Volume 500, Issue 3, pp.3027-3049
Duarte-Cabral, A. & Dobbs, C. L. 2016, MNRAS, 458, 3667
Duarte-Cabral, A. & Dobbs, C. L. 2017, MNRAS, 470, 4261
Edelsohn, D. J. & Elmegreen, B. G. 1997, MNRAS, 287, 947
Eilers, P. H. C. 2004, Analytical Chemistry, 76, 404, pMID: 14719890
Elia, D., Molinari, S., Schisano, E., et al. 2022, ApJ, 941, 162
Elias, F., Cabrera-Caño, J., & Alfaro, E. J. 2006, AJ, 131, 2700
Elmegreen, B. G. & Scalo, J. 2004, ARA&A, 42, 211
Federrath, C., Glover, S. C. O., Klessen, R. S., & Schmidt, W. 2008, Physica Scripta Volume T, 132, 014025
Federrath, C. & Klessen, R. S. 2013, ApJ, 763, 51
Federrath, C., Roman-Duval, J., Klessen, R. S., Schmidt, W., & Mac Low, M. M. 2010, A&A, 512, A81
Feldt, C. 1993, A&A, 276, 531
Ferrière, K. M. 2001, Reviews of Modern Physics, 73, 1031
Field, G. B. 1965, ApJ, 142, 531
Field, G. B., Goldsmith, D. W., & Habing, H. J. 1969, ApJ, 155, L149
Fontani, F., Giannetti, A., Beltrán, M. T., et al. 2012, MNRAS, 423, 2342
Frerking, M. A., Langer, W. D., & Wilson, R. W. 1982, ApJ, 262, 590
Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2021, A&A, 649, A1
Giannetti, A., Leurini, S., König, C., et al. 2017, A&A, 606, L12
Giannetti, A., Wyrowski, F., Brand, J., et al. 2014, A&A, 570, A65

Gibson, S. J., Taylor, A. R., Higgs, L. A., Brunt, C. M., & Dewdney, P. E. 2005, ApJ, 626, 195
Gibson, S. J., Taylor, A. R., Higgs, L. A., & Dewdney, P. E. 2000, ApJ, 540, 851
Girichidis, P., Konstandin, L., Whitworth, A. P., & Klessen, R. S. 2014, ApJ, 781, 91
Glover, S. C. O. & Mac Low, M.-M. 2007, ApJ, 659, 1317
Glover, S. C. O. & Mac Low, M. M. 2011, MNRAS, 412, 337
Goldsmith, P. F., Heyer, M., Narayanan, G., et al. 2008, ApJ, 680, 428
Goldsmith, P. F. & Langer, W. D. 1999, ApJ, 517, 209
Goldsmith, P. F. & Li, D. 2005, ApJ, 622, 938
Goldsmith, P. F., Li, D., & Krčo, M. 2007, ApJ, 654, 273
Goodman, A. A., Alves, J., Beaumont, C. N., et al. 2014, ApJ, 797, 53
Goodman, A. A., Pineda, J. E., & Schnee, S. L. 2009, ApJ, 692, 91
Gould, R. J. 1994, ApJ, 423, 522
Gould, R. J. & Salpeter, E. E. 1963, ApJ, 138, 393
Güver, T. & Özel, F. 2009, MNRAS, 400, 2050
Haud, U. & Kalberla, P. M. W. 2007, A&A, 466, 555
Heeschen, D. S. 1954, AJ, 59, 324
Heeschen, D. S. 1955, ApJ, 121, 569
Heiles, C. 2001, ApJ, 551, L105
Heiles, C. & Troland, T. H. 2003, ApJ, 586, 1067
Henderson, A. P., Jackson, P. D., & Kerr, F. J. 1982, ApJ, 263, 116
Hennebelle, P. & Pérault, M. 1999, A&A, 351, 309
Henshaw, J. D., Barnes, A. T., Battersby, C., et al. 2022, arXiv e-prints, arXiv:2203.11223
Henshaw, J. D., Kruijssen, J. M. D., Longmore, S. N., et al. 2020, Nature Astronomy, 4, 1064
Heyer, M. & Dame, T. M. 2015, ARA&A, 53, 583
Honig, Z. N. & Reid, M. J. 2015, ApJ, 800, 53
Imara, N. & Burkhart, B. 2016, ApJ, 829, 102
Izquierdo, A. F., Smith, R. J., Glover, S. C. O., et al. 2021, MNRAS, 500, 5268
Jackson, J. M., Bania, T. M., Simon, R., et al. 2002, ApJ, 566, L81
Jackson, J. M., Finn, S. C., Chambers, E. T., Rathborne, J. M., & Simon, R. 2010, ApJ, 719, L185
Jackson, J. M., Rathborne, J. M., Shah, R. Y., et al. 2006, ApJS, 163, 145
Jenkins, E. B. & Tripp, T. M. 2001, ApJS, 137, 297
Jow, D. L., Hill, R., Scott, D., et al. 2018, MNRAS, 474, 1018
Kabanovic, S., Schneider, N., Ossenkopf-Okada, V., et al. 2022, A&A, 659, A36
Kainulainen, J., Beuther, H., Banerjee, R., Federrath, C., & Henning, T. 2011, A&A, 530, A64
Kainulainen, J., Beuther, H., Henning, T., & Plume, R. 2009, A&A, 508, L35
Kainulainen, J., Federrath, C., & Henning, T. 2014, Science, 344, 183
Kalberla, P. M. W. & Dedes, L. 2008, A&A, 487, 951
Kalberla, P. M. W., Dedes, L., Kerp, J., & Haud, U. 2007, A&A, 469, 511
Kalberla, P. M. W. & Kerp, J. 2009, ARA&A, 47, 27
Kalberla, P. M. W., Kerp, J., Haud, U., et al. 2016, ApJ, 821, 117

Kavars, D. W., Dickey, J. M., McClure-Griffiths, N. M., Gaensler, B. M., & Green, A. J. 2003, ApJ, 598, 1048
Klessen, R. S. 2000, ApJ, 535, 869
Klessen, R. S. & Glover, S. C. O. 2016, Saas-Fee Advanced Course, 43, 85
Knapp, G. R. 1974, AJ, 79, 527
Koley, A. & Roy, N. 2019, MNRAS, 483, 593
Kolmogorov, A. 1941, Akademiia Nauk SSSR Doklady, 30, 301
Kolpak, M. A., Jackson, J. M., Bania, T. M., Clemens, D. P., & Dickey, J. M. 2003, ApJ, 582, 756
Konstandin, L., Girichidis, P., Federrath, C., & Klessen, R. S. 2012, ApJ, 761, 149
Körtgen, B., Federrath, C., & Banerjee, R. 2019, MNRAS, 482, 5233
Kounkel, M. & Covey, K. 2019, AJ, 158, 122
Kritsuk, A. G., Norman, M. L., Padoan, P., & Wagner, R. 2007, ApJ, 665, 416
Krčo, M., Goldsmith, P. F., Brown, R. L., & Li, D. 2008, ApJ, 689, 276
Larson, R. B. 1979, MNRAS, 186, 479
Larson, R. B. 1981, MNRAS, 194, 809
Lee, E. J., Miville-Deschênes, M.-A., & Murray, N. W. 2016, ApJ, 833, 229
Levine, E. S., Blitz, L., & Heiles, C. 2006, ApJ, 643, 881
Li, D. & Goldsmith, P. F. 2003, ApJ, 585, 823
Lindner, R. R., Vera-Ciro, C., Murray, C. E., et al. 2015, AJ, 149, 138
Liu, B., Wang, L., Wang, J., Peng, B., & Wang, H. 2022, PASA, 39, e050
Lombardi, M., Alves, J., & Lada, C. J. 2015, A&A, 576, L1
Majewski, S. R., Zasowski, G., & Nidever, D. L. 2011, ApJ, 739, 25
Marchal, A. & Miville-Deschênes, M.-A. 2020, arXiv e-prints, arXiv:2012.03160
Marchal, A., Miville-Deschênes, M.-A., Orieux, F., et al. 2019, A&A, 626, A101
Marsh, K. A., Whitworth, A. P., Lomax, O., et al. 2017, MNRAS, 471, 2730
McClure-Griffiths, N. M. & Dickey, J. M. 2007, ApJ, 671, 427
McClure-Griffiths, N. M., Dickey, J. M., Gaensler, B. M., et al. 2005, ApJS, 158, 178
McKee, C. F. & Ostriker, E. C. 2007, ARA&A, 45, 565
McKee, C. F. & Ostriker, J. P. 1977, ApJ, 218, 148
Miville-Deschênes, M.-A. & Lagache, G. 2005, ApJS, 157, 302
Miville-Deschênes, M.-A., Murray, N., & Lee, E. J. 2017, ApJ, 834, 57
Molina, F. Z., Glover, S. C. O., Federrath, C., & Klessen, R. S. 2012, MNRAS, 423, 2680
Mottram, J. C. & Brunt, C. M. 2010, ASP Conf. Ser., 438, 98
Murray, C. E., Lindner, R. R., Stanimirović, S., et al. 2014, ApJ, 781, L41
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
Nakanishi, H., Fujita, S., Tachihara, K., et al. 2020, PASJ, 72, 43
Nakanishi, H. & Sofue, Y. 2003, PASJ, 55, 191
Nakanishi, H. & Sofue, Y. 2016, PASJ, 68, 5
Nguyen, H., Dawson, J. R., Lee, M.-Y., et al. 2019, ApJ, 880, 141
Ossenkopf-Okada, V., Csengeri, T., Schneider, N., Federrath, C., & Klessen, R. S. 2016, A&A, 590, A104
Ostriker, J. 1964, ApJ, 140, 1056

Padoan, P., Federrath, C., Chabrier, G., et al. 2014, in Protostars and Planets VI, ed. H. Beuther, R. S. Klessen, C. P. Dullemond, & T. Henning, 77
Padoan, P. & Nordlund, Å. 2002, ApJ, 576, 870
Padoan, P., Nordlund, A., & Jones, B. J. T. 1997, MNRAS, 288, 145
Passot, T. & Vázquez-Semadeni, E. 1998, Phys. Rev. E, 58, 4501
Pineda, J. E., Caselli, P., & Goodman, A. A. 2008, ApJ, 679, 481
Pineda, J. L., Langer, W. D., Velusamy, T., & Goldsmith, P. F. 2013, A&A, 554, A103
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014, A&A, 571, A1
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2011, A&A, 536, A19
Ragan, S. E., Henning, T., Tackenberg, J., et al. 2014, A&A, 568, A73
Rebolledo, D., Green, A. J., Burton, M., et al. 2017, MNRAS, 472, 1685
Reed, B. C. 2000, AJ, 120, 314
Reid, M. J., Dame, T. M., Menten, K. M., & Brunthaler, A. 2016, ApJ, 823, 77
Reid, M. J., Menten, K. M., Brunthaler, A., et al. 2019, ApJ, 885, 131
Reid, M. J., Menten, K. M., Brunthaler, A., et al. 2014, ApJ, 783, 130
Reid, M. J., Menten, K. M., Zheng, X. W., et al. 2009, ApJ, 700, 137
Riegel, K. W. & Crutcher, R. M. 1972, A&A, 18, 55
Riegel, K. W. & Jennings, M. C. 1969, ApJ, 157, 563
Riener, M., Kainulainen, J., Beuther, H., et al. 2020a, A&A, 633, A14
Riener, M., Kainulainen, J., Henshaw, J. D., & Beuther, H. 2020b, A&A, 640, A72
Riener, M., Kainulainen, J., Henshaw, J. D., et al. 2019, A&A, 628, A78
Roman-Duval, J., Jackson, J. M., Heyer, M., et al. 2009, ApJ, 699, 1153
Rosolowsky, E. W., Pineda, J. E., Kauffmann, J., & Goodman, A. A. 2008, ApJ, 679, 1338
Roy, N., Kanekar, N., Braun, R., & Chengalur, J. N. 2013a, MNRAS, 436, 2352
Roy, N., Kanekar, N., & Chengalur, J. N. 2013b, MNRAS, 436, 2366
Ruder, S. 2016, arXiv e-prints, arXiv:1609.04747
Rugel, M. R., Beuther, H., Bihr, S., et al. 2018, A&A, 618, A159
Saury, E., Miville-Deschênes, M. A., Hennebelle, P., Audit, E., & Schmidt, W. 2014, A&A, 567, A16
Scalo, J. M. 1984, ApJ, 277, 556
Schneider, N., André, P., Könyves, V., et al. 2013, ApJ, 766, L17
Schneider, N., Bontemps, S., Motte, F., et al. 2016, A&A, 587, A74
Schneider, N., Ossenkopf, V., Csengeri, T., et al. 2015, A&A, 575, A79
Schneider, N., Ossenkopf-Okada, V., Clarke, S., et al. 2022, A&A, 666, A165
Seifried, D., Beuther, H., Walch, S., et al. 2022, MNRAS, 512, 4765
Seifried, D., Schmidt, W., & Niemeyer, J. C. 2011, A&A, 526, A14
Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163
Smith, R. J., Glover, S. C. O., Clark, P. C., Klessen, R. S., & Springel, V. 2014a, MNRAS, 441, 1628
Smith, R. J., Glover, S. C. O., & Klessen, R. S. 2014b, MNRAS, 445, 2900
Smith, R. J., Treß, R. G., Sormani, M. C., et al. 2020, MNRAS, 492, 1594
Soler, J. D., Beuther, H., Rugel, M., et al. 2019, A&A, 622, A166
Soler, J. D., Beuther, H., Syed, J., et al. 2020, A&A, 642, A163
Soler, J. D., Beuther, H., Syed, J., et al. 2021, A&A, 651, L4

Soler, J. D., Miville-Deschênes, M. A., Molinari, S., et al. 2022, A&A, 662, A96
Sparke, L. S. 1993, in American Institute of Physics Conference Series, Vol. 278, Back to the Galaxy, ed. S. S. Holt & F. Verter, 447–456
Stahler, S. W. & Palla, F. 2005, The Formation of Stars (New York: John Wiley & Sons)
Stecker, F. W., Solomon, P. M., Scoville, N. Z., & Ryter, C. E. 1975, ApJ, 201, 90
Sternberg, A., Le Petit, F., Roueff, E., & Le Bourlot, J. 2014, ApJ, 790, 10
Stil, J. M., Taylor, A. R., Dickey, J. M., et al. 2006, AJ, 132, 1158
Strasser, S. & Taylor, A. R. 2004, ApJ, 603, 560
Su, Y., Yang, J., Zhang, S., et al. 2019, ApJS, 240, 9
Suri, S., Sánchez-Monge, Á., Schilke, P., et al. 2019, A&A, 623, A142
Syed, J., Soler, J. D., Beuther, H., et al. 2022, A&A, 657, A1
Syed, J., Wang, Y., Beuther, H., et al. 2020, A&A, 642, A68
Tang, N., Li, D., Heiles, C., et al. 2016, A&A, 593, A42
Tang, N., Li, D., Heiles, C., et al. 2017, ApJ, 839, 8
Taylor, A. R., Gibson, S. J., Peracaula, M., et al. 2003, AJ, 125, 3145
Tikhonov, A. N. 1963, Soviet Math. Dokl., 4, 1035
Umemoto, T., Minamidani, T., Kuno, N., et al. 2017, PASJ, 69, 78
van der Werf, P. P., Goss, W. M., & Vanden Bout, P. A. 1988, A&A, 201, 311
van Dishoeck, E. F. & Black, J. H. 1988, ApJ, 334, 771
Wakker, B. P. & van Woerden, H. 1997, ARA&A, 35, 217
Walter, F., Brinks, E., de Blok, W. J. G., et al. 2008, AJ, 136, 2563
Wang, K., Testi, L., Burkert, A., et al. 2016, ApJS, 226, 9
Wang, K., Testi, L., Ginsburg, A., et al. 2015, MNRAS, 450, 4043
Wang, Y., Beuther, H., Rugel, M. R., et al. 2020a, A&A, 634, A83
Wang, Y., Beuther, H., Schneider, N., et al. 2020b, A&A, 641, A53
Wang, Y., Bihr, S., Beuther, H., et al. 2020c, A&A, 634, A139
Wang, Y., Bihr, S., Rugel, M., et al. 2018, A&A, 619, A124
Watson, W. D. 1972, ApJ, 176, 103
Westmeier, T. 2018, MNRAS, 474, 289
Widrow, L. M., Barber, J., Chequers, M. H., & Cheng, E. 2014, MNRAS, 440, 1971
Wilson, T. L., Rohlfs, K., & Hüttemeister, S. 2013, Tools of Radio Astronomy (Berlin Heidelberg: Springer Science & Business Media)
Wolfire, M. G., Hollenbach, D., McKee, C. F., Tielens, A. G. G. M., & Bakes, E. L. O. 1995, ApJ, 443, 152
Wolfire, M. G., McKee, C. F., Hollenbach, D., & Tielens, A. G. G. M. 2003, ApJ, 587, 278
Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868
Zhang, F., Tang, X., Tong, A., et al. 2020, Spectroscopy Letters, 53, 222
Zhang, M., Kainulainen, J., Mattern, M., Fang, M., & Henning, T. 2019, A&A, 622, A52
Zucker, C., Battersby, C., & Goodman, A. 2015, ApJ, 815, 23
Zucker, C., Battersby, C., & Goodman, A. 2018, ApJ, 864, 153

Acknowledgments

Heidelberg, May 2023

I want to express my deepest gratitude to Henrik Beuther. I might be biased here, but I think I can say that he is objectively one of the best PhD advisers one can have. He has always been an excellent teacher, adviser, and a good mentor. Thanks to him, I am almost comfortable to say that I have become a somewhat decent scientist. The Franconian in me would say: „Bassd scho!“[1]

His guidance helped me make progress ever since I came to MPIA during my bachelor's. I remember the first seminar of his that I attended, where I learned how to handle and give proper feedback on scientific talks and papers. Building upon this, I soon learned how to navigate the academic work environment and how to conduct scientific research. I admire Henrik for his personal and professional integrity. I could always turn to him for advice, even on things that go beyond my academic career. And this is what makes a great adviser. To not only care about the scientific results but also the person behind it.

I have met so many great people along the way. I want to thank Simon Bihr who made sure that my first weeks in this scientific environment were easy, fun, and stimulating. I want to thank Martin Kürster who was extremely understanding after I quickly felt overwhelmed by the tasks he had given me during my HiWi job. His encouragement made me feel more confident in my skills, such that I wanted to pursue a scientific career in the first place. There is too many people that I would like to give a shoutout, but can't because I'm an idiot and I like to forget important things. Like my first office mates Asmita, Camille, and Christos. It was so much fun with you! All my current and past group members who helped me improve the quality of my research and made working at MPIA so much more fun. I should probably mention my good "friend" Sara, thanks for being such an awesome colleague...and a friend (for real!) ;-) To my bestie Molly, thank you for being who you are. I know it can feel so incredibly overwhelming sometimes, but you have been crushing it from the start. You are and will be doing amazing for your PhD. Thank you for caring, this means so much to me. And also, "I think you are very cool!"

A big chunk of my PhD was defined by the pandemic, which led to some of us being moved to temporary offices in a separate building. Camille, thanks for being an incredible human being. It was tough but we made it! Thanks for all

[1] An expression of highest praise and approval. Or not.

those refreshing conversations, while we should have been working. Thank you for listening when I was whining about my research. Thanks for being this awesome scientist that I look up to. Thanks for helping me (still today!) with all the questions I have about my PhD.

Thanks, Marten, for all those fun Tischtennis matches (and pull requests, sigh!) that distracted me from work. Arv, it is awesome to have a friend that I can hang out with, without talking too much about work but instead talk about cricket, or football, or literally anything else. Or just to get drunk (maybe not Mel's drunk). Knowing you, it also brings me so much more joy to see Chelsea lose. Thanks, mate!

I also want to thank my Glubb[2] for teaching me how to deal with setbacks on a weekly basis. This made my PhD feel like a walk in the park.

Sadly, I had to say goodbye to friends and family members during the last four years. This is for my uncle Johannes Hauff, my grandparents Margarete&Martin Scheytt, my grandfather Jamil Hasan Syed, and my good friend Mathias Schlicht. I miss you dearly.

I want to thank my family, particularly my parents, for their endless support. You went to great lengths to make sure I have everything I need to pursue the things I want to pursue in life. Thank you to my family in Bangladesh and in Rottweil. It is encouraging to feel the support from you. Lastly, I want to thank my partner Svenja. You are the love of my life, and there is no way I could have done this without you. Sorry for being such a pain in the neck sometimes. It can be a little intimidating not knowing what lies ahead of us, but I am actually pretty confident about the future.

[2] 1. FC Nürnberg

Atomic hydrogen gas (HI) is an integral constituent of the interstellar medium (ISM) and thus plays a critical role in the assembly of molecular clouds, the sites of star formation. An important physical agent in controlling the transition from atomic to molecular gas is cold HI. HI emission, that traces a wide range of hydrogen properties, is found throughout the Galactic plane and exhibits complex kinematic signatures that are imposed by the Galactic rotation. The unfavorable viewing geometry from our vantage point in the Galaxy requires a new set of tools that allows us to disentangle HI structures along our lines of sight. In this thesis, I constrain the properties of the interstellar hydrogen in our Milky Way galaxy and present a way to isolate cold hydrogen from HI emission using HI self-absorption (HISA). The outer Galaxy offers a less confused view on HI emission, and we exploit this circumstance with unprecedented detail using the high-angular resolution data of The HI/OH/Recombination line survey of the inner Milky Way (THOR). We discover the "Maggie" filament, one of the largest, almost purely atomic filaments in the Milky Way. Maggie has a length of 1.2 kpc and is located on the far side of the Galaxy at a distance of 12 kpc from the Galactic center. Optical depth measurements suggest that Maggie is in a largely cold HI phase and molecular gas is only found on the smallest spatial scales. When targeting molecular clouds in the inner Galactic plane, we detect pronounced HISA as a tracer of cold hydrogen. While the kinematic correlation is less surprising due to the selection bias of our sample, the spatial distribution of cold HI gas is also found to be well correlated with that of the molecular gas. The detection of HISA furthermore extends to the whole of our survey. We frequently find absorption signatures without any associated molecular counterpart. While cold atomic gas is rendered vitally important to the formation process of molecular clouds, the cold phase of atomic hydrogen marks a distinct gas phase in the ISM that exists throughout the Galaxy, independent of the occurrence of molecular gas.