Dissertation
submitted to the
Combined Faculties of the Natural Sciences and Mathematics
of the Ruperto-Carola-University of Heidelberg. Germany
for the degree of
Doctor of Natural Sciences

Put forward by
Diplom-Physiker Simon Bihr
born in: Aalen, Germany
Oral examination: 20th of April, 2016

Molecular Cloud Formation out of the Atomic Phase

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

Abstract:

This thesis is dedicated to study observationally the formation of molecular cloud formation out of the atomic phase. In the framework of 'The H I, OH, Recombination Line survey of the Milky Way' (THOR), we used the Very Large Array (VLA) to map the 21 cm H I line, 4 OH transitions, up to 19 Hα recombination lines and the continuum from 1 to 2 GHz of a significant fraction of the Milky Way. The goal of this thesis is to characterize and study the atomic and molecular content of molecular clouds.

Observations of the giant molecular cloud complex associated with the W43 star formation complex revealed large quantities of optically thick H I emission. We used strong continuum sources to measure the optical depth and used this information to correct the H I column density and hence the H I mass. Without these corrections, the mass estimate is at least a factor of 2.4 too low. In addition, we observed H I column densities up to $N_{\mathrm{H\,I}} \sim 150\,\mathrm{M}_\odot\,\mathrm{pc}^{-2}$, which is in contradiction to current cloud formation models.

Furthermore, we present a catalog of ~4400 continuum sources, extracted from the first half of the THOR survey. Due to the broad bandwidth from 1 to 2 GHz, we are able to determine a reliable spectral index for ~1800 bright sources, which allows us to distinguish between thermal and non-thermal radiation. Using this information, we can confirm four super nova remnant candidates. Beside the direct scientific studies of these evolved sources, this catalog is the basis for prospective absorption studies of the H I and OH lines.

By extracting H I self absorption (HISA) features of the molecular filament GMF38.1-32.4, we examined directly the cold and dense atomic hydrogen and compared it to the molecular counterpart. We studied the kinematics and column density probability distribution functions (PDFs) and found a log-normal shape for them, which indicates turbulent motion. Furthermore, we might observe several evolutionary stages within the filament, which will be helpful to validate theoretical models and simulations.

Zusammenfassung:
Diese Dissertation behandelt die Entstehung von Molekülwolken, welche mit der atomaren Phase beginnt. Im Rahmen der 'H I, OH, Rekombinationslinen Beobachtungskampagne der Milchstraße' (THOR) beobachteten wir mit dem Very Large Array (VLA) die 21cm H I Linie, 4 OH Linien, bis zu 19 Hα Rekombinationslinien und das Kontinuum von 1 bis 2 GHz von einem erheblichen Teil der Milchstraße. Das Ziel der Arbeit ist die Charakterisierung und Untersuchung der atomaren und molekularen Anteile von Molekülwolken.
Beobachtungen des Riesenmolekülwolken Komplex, welcher zu dem Sternentstehungskomplex W43 gehört, haben große Mengen von optisch dicker H I Strahlung offenbart. Um die optische Tiefe zu bestimmen haben wir starke Kontinuums Quellen beobachtet und konnten mit dieser Information die Säulendichte und somit auch die Masse korrekt bestimmen. Ohne diese Korrektur ist die Massenbestimmung um mindestens einen Faktor von 2.4 zu klein. Außerdem haben wir H I Säulendichten bis zu Werten von $N_{\rm H\,I} \sim 150\,{\rm M_\odot\,pc^{-2}}$ beobachtet, welche im Widerspruch zu aktuellen Wolkenenstehungsmodellen stehen.
Wir präsentieren außerdem einen Katalog mit ~4400 Kontinuums Quellen, die aus dem ersten Teil der THOR Beobachtungskampagne extrahiert wurden. Mit Hilfe des breiten Bandpasses von 1 bis 2,GHz konnten wir verlässliche spektrale Indizes für ~1800 helle Quellen bestimmen. Dadurch konnten wir zwischen thermischer und nicht-thermischer Strahlung unterscheiden und wir konnten vier Kandidaten von Supernova Überresten bestätigen. Außer der direkten wissenschaftlichen Studie dieser weit entwickelten Quellen ist der Katalog die Grundlage für zukünftige Absorptionsstudien der H I und OH Linien.
Mittels von H I Selbstabsorptions Bereichen in dem molekularen Filament GMF38.1-32.4 konnten wir kalten und dichten atomaren Wasserstoff direkt untersuchen und dies mit dem molekularen Gegenstück vergleichen. Dabei haben wir Säulendichte Wahrscheinlichkeitsverteilungsfunktionen untersucht. Diese zeigen eine Logarithmische Normalverteilungsform, was ein Hinweis auf turbulente Bewegungen ist. Außerdem beobachten wir Möglicherweise mehrere evolutionäre Stadien innerhalb des Filaments, wodurch theoretische Modelle und Simulationen überprüft werden können.

Contents

Acknowledgements

I want to thank all the people that made this thesis possible. My first big thanks goes to Henrik for not just being an amazing supervisor with an incredible amount of scientific help and tips, but also for being a positive supporter and believer, always with a smile and an open ear for anything.

I want to thank my parents for their support during my studies and PhD, not just financially, but also for their loving help that only parents can give. Without their encouragement, I would not even have passed primary school.

A big thanks goes to Tanya for not just proofreading and supporting me, but also for distracting me during the last years.

I want to thank Juergen Ott from the NRAO for co-supervising me during my time in Socorro.

I also want to thank Thomas Henning and Simon Glover for being a part of my thesis committee and helping me a lot with fruitful discussions.

A further big thanks goes to Cornelis Dullemond for being the second referee and to Ralf Klessen and Thomas Pfeifer for being in my defense committee.

I want to thank the IMPRS and the coordinator Christian Fendt for this wonderful community and also for financial support.

The last thanks goes to the great community at the MPIA, which makes life here wonderful.

The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc.

1

Introduction

1.1 Molecular cloud formation starts with atoms

Understanding the formation and evolution of molecular clouds is a key ingredient in understanding star formation in general. However, the formation of molecular clouds starts with atoms, more precise with atomic hydrogen. Cloud formation scenarios suggest the need of converging or colliding flows to form density enhancements, where the formation of molecular hydrogen out of the atomic phase can occur (e.g., Vázquez-Semadeni et al. 2006; Heitsch et al. 2008; Krumholz et al. 2009). Therefore the investigation of molecular cloud formation begins with a sophisticated characterization of the atomic HI phase.

The goal of this thesis is to characterize and study atomic hydrogen with the famous 21cm H I line. I will aim to answer a variety of general questions, including the density structure and the kinematics. The main questions are: What are the contributions and characteristic properties of the different phases of the atomic hydrogen, mainly the cold neutral medium (CNM) and the warm neutral medium (WNM)? What are the conditions under which molecular hydrogen forms out of the atomic phase? Are converging flows a main driver for this formation? These are general questions in the field of molecular cloud formation, however, with this thesis I will try to answer more specific sub-questions such as: How can we observationally disentangle the CNM and WNM and how does their interplay shape the ISM? Do we observe column density thresholds for the HI when the formation of H_2 takes place as predicted by theoretical models? What is the ratio of molecular to atomic hydrogen and does it depend on the environment? What are the shapes of atomic and molecular column density probability distribution functions (PDFs) and what is their physical interpretation? The THOR survey offers an extensive database to study these questions and in the following I will introduce the survey before I introduce basic properties of molecular clouds and the ISM.

1.2 Galactic plane surveys

In the previous decades, theoretical as well as observational projects aimed to shed light on the properties of the interstellar medium (ISM) and connect this to large scale structures, such as the Milky Way. With the advantage of powerful computers and algorithms, it is possible to simulate entire galaxies with resolutions down to a few solar masses (e.g., Dobbs et al. 2012; Smith et al. 2014). Observational projects also tackle large surveys of the Milky Way, rather than small specific regions within it. These surveys cover a large fraction of the spectral range, from the near- (UKIDSS, Lucas et al. 2008), mid- (GLIMPSE, Churchwell et al. 2009) and far-infrared (MIPSGAL, HIGAL, Carey et al. 2009; Molinari et al. 2010), to the submm (ATLASGAL, BOLOCAM, Schuller et al. 2009; Rosolowsky et al. 2010; Aguirre et al. 2011; Csengeri et al. 2014), to longer radio wavelengths studying the continuum as well as molecular lines (e.g., GRS, MAGPIS, CORNISH, HOPS, MALT90, MALT-45, Jackson et al. 2006; Helfand et al. 2006; Hoare et al. 2012; Purcell et al. 2013; Walsh et al. 2011; Purcell et al. 2012; Jackson et al. 2013).

As hydrogen is the most common element in our universe, observations of this element are a crucial ingredient to complete the picture of our Galaxy. Molecular hydrogen is difficult to observe directly as its rotational energy levels are not readily excited in the cold ISM. However, the 21 cm H I line provides a direct measurement of the atomic hydrogen. To date, the Galactic plane surveys of the 21 cm H I line have a spatial resolution of >1′ (CGPS, SGPS, VGPS, Taylor et al. 2003; McClure-Griffiths et al. 2005; Stil et al. 2006), which is not sufficient in comparison to the other Galactic plane surveys. This was the motivation to initiate a Galactic plane survey using the Karl G. Jansky Very Large Array (VLA) in C-configuration, achieving a spatial resolution of ∼20″: 'THOR - The H I, OH, Recombination line survey of the Milky Way'. The angular resolution of 20″ gives us a linear resolution of ∼0.1 to 1.5 pc at typical Galactic distances of 1 to 15 kpc. Since the new WIDAR correlator at the VLA offers a broad bandwidth, including high resolution sub-bands, we are able to observe the 21 cm H I line, 4 OH lines, 19 Hα radio recombination lines (RRL) and the continuum from 1-2 GHz simultaneously.

As we observe a large portion of the Milky Way in different lines and the continuum with the THOR survey, we can study vastly different aspects.

1.2.1 THOR survey

The Thor survey was observed using the Karl G. Jansky Very Large Array (VLA) radio interferometer, which will be introduced in detail in Sect.1.4.2. We were awarded more than 200 hours observing time in a VLA large program. The principle investigator (PI) of this project is

Henrik Beuther from the MPIA. To test our observing strategy and data reduction, we began with a pilot study of a $2^\circ \times 2^\circ$ field around W43 in 2012, which was completed in 10 hours. The H I data of this pilot study will be the basis of Chapter 2. In consecutive observing campaigns, we mapped a significant fraction of the first Galactic quadrant (Galactic longitude $l = 15-67^\circ$, Galactic latitude $|b| \leq 1^\circ$).

As explained, our goal was to achieve a high spatial resolution for the H I line. To overcome problems due to missing flux and to reconstruct the large scale structure, we aimed to combine the THOR H I data with low resolution VGPS data (Stil et al. 2006). The VGPS survey was observed using the VLA in D-array configuration, achieving a spatial resolution of $60''$. To achieve a higher resolution, we choose the next larger configuration, the VLA C-array configuration, resulting in a resolution of $\sim 20''$, depending on the frequency and the observed declination.

Since the upgrade of the VLA, which was finished in 2012, the new WIDAR correlator is capable to observe a broad bandwidth, including smaller high resolution sub-bands. This allowed us to observe the H I 21 cm line simultaneously with four OH lines at 1612, 1665, 1667, and 1720 MHz, 19 Hα radio recombination lines (RRL) and the continuum from $1 - 2$ GHz in full polarization. The velocity resolution for the H I line is $0.82\,\mathrm{km\,s^{-1}}$, for the OH lines it is $\sim 1.4\,\mathrm{km\,s^{-1}}$ and the RRL have a lower resolution of $\sim$3-5 $\mathrm{km\,s^{-1}}$. Further details about the entire survey will be given in the THOR overview paper (Beuther et al. in prep). Chapter 2 and 3 will provide more details about the used data.

Beside studies of the atomic content using the H I 21 cm line, we can study the molecular content using the OH transition. The combination of these two data sets allows us to study directly the transition from the atomic to the molecular phase. In addition, we can observe the ionized component with the RRL lines and the continuum. We can even study magnetic fields by examining the polarization of the continuum. To exploit all these different topics and data products, a large team of experts is necessary. Within the THOR team, we have these experts working on different topics. To date several papers were published covering different aspects: Bihr et al. (2015) studied the H I content of the pilot region around W43 and will be presented in Chapter 2. Walsh et al. (2016) presented the OH data of the pilot region, focusing on maser emission. Bihr et al. (2016) compiled a catalog of all continuum sources of the first half of the THOR survey and will be the topic of Chapter 3. Further papers are in preparation: Beuther et al. (in prep.) will give an overview of the entire survey, explaining the observations, calibration and imaging in detail. Rugel et al. (in prep.) will study the molecular content of the Milky Way by examining OH absorption spectra. Stil et al. (in prep.) will explore the polarization of the continuum emission and analyze the corresponding Faraday rotation. Chapter 4 presents a study of H I self absorption, which will be published in a forthcoming paper (Bihr et al. in

prep.).

1.2.2 Basic properties of the ISM and molecular clouds

It is generally known, that star formation takes place in molecular clouds (e.g., Klessen 2011; Dobbs et al. 2014). As the name already indicates, these clouds consists primarily of molecular hydrogen H_2, but also contain helium, atomic hydrogen and other molecular tracers, such as CO. The mass of these clouds can range from a few solar masses for isolated small clouds, so called Bok globules (e.g., B68, Nielbock et al. 2012), to Giant Molecular Clouds (GMCs), having masses of up to $10^7 M_\odot$ in the Galactic center (Oka et al. 2001). The surface density of these clouds is in general very high ($\sim 150\,M_\odot\,pc^{-2}$) and the molecular hydrogen is shielded from the interstellar radiation field. The internal density structure can be described using probability density functions (PDF), which reveal a log-normal shape due to turbulent motion (Kainulainen et al. 2009). More evolved clouds develop a log-normal high density tail, which is an indication of gravitational collapse (e.g. Schneider et al. 2015a,b). More details about PDFs will be presented in Chapter. 4.

The formation scenario of molecular clouds is thought to be driven by turbulent motions, rather than a coagulation model (e.g., Klessen & Glover 2014). The basic idea of these models is that molecular clouds form in post-shock regions of colliding or converging flows of lower density gas. If the density becomes high enough, molecular hydrogen can form and also 'survive' as it is shielded from the interstellar radiation field by molecules and dust (e.g., Glover & Clark 2012). The main formation mechanism of the molecular hydrogen out of the atomic state is via grain catalysis (Draine 2011; Klessen & Glover 2014). Atomic hydrogen can get bound on a grain surface and is able to 'walk' some distance on the grain. If two hydrogen atoms encounter each other, they can form molecular hydrogen. The released energy during the formation process (4.5 eV) will free the molecule from the grain and hence enrich the ISM with molecular hydrogen. In contrast to the formation, UV photons from the interstellar radiation field can dissociate the molecular hydrogen and break the bound. Therefore, molecular hydrogen will be solely present in high density regions, where it is shielded from the interstellar radiation field.

The atomic hydrogen in the ISM appears in different phases. The main phases are the cold neutral medium (CNM) and the warm neutral medium (WNM). Both phases can coexist in pressure equilibrium (Wolfire et al. 1995), even though their properties are substantially different. The CNM usually has an excitation temperature, called spin temperature T_S, of 20 to 200 K (Heiles & Troland 2003b; Kavars et al. 2005) and a particle density of $\sim 50\,cm^{-3}$ (Stahler et al. 2005). In contrast to this, the WNM is much warmer and diffuse, with temperatures up to ~ 7000 K (Murray et al. 2014) and densities of $\sim 0.5\,cm^{-3}$ (Stahler et al. 2005). Even though the mass of

both components within our Milky Way is comparable, each having a mass of ~3-4×10^9 $M_\odot$, the filling factor is considerably different with $f = 0.30$ for the WNM and $f = 0.04$ for the CNM (Stahler et al. 2005) (the largest volume is occupied by the hot ionized medium with a filling factor of 0.5). Since the atomic hydrogen coexists in different forms with different properties, it is challenging to observe and disentangle each component. Hence a profound analysis of the H I radiation, including the optical depth and spin temperature is necessary to deduce the properties of each component. This will be introduced in the next Section and the corresponding results will be given in Chapter 2 and 4.

1.3 H I radiative transfer

This section aims to summarize and explain the radiative transfer that involves the emission and absorption of the neutral hydrogen. As astronomers, we cannot perform hand on experiments with interstellar clouds to understand their properties and structure, but we can carefully observe at different wavelengths and different resolutions. To determine the corresponding properties of the observed atoms or molecules, we have to understand the radiative transfer in detail and account for optical depth effects, self absorption as well as contributions from continuum emission. Since our atmosphere is transparent in two windows, in the optical and in the radio regime, we can perform observations in these frequency ranges from the ground with large telescopes. The radio regime offers a wide range of different atomic and molecular lines, including the important 21cm line of the neutral hydrogen. In the following I will explain the origin and basic principles of this line.
The following sections are based on the text books by Wilson et al. (2010), Draine (2011) and Irwin (2007). I also used the excellent online course 'Essential Radio Astronomy' by J. J. Condon and S. M. Ransom, which can be found under www.cv.nrao.edu/course/astr534/.

1.3.1 Basic Definitions of radiation

Even though the following definitions and equations are well known text book knowledge, I will shortly summarize them, as they are the basis for the subsequent discussion. The specific intensity or specific brightness is defined as:

$$I_\nu \equiv \frac{dP}{cos(\theta)\, d\sigma\, d\nu\, d\Omega}, \tag{1.1}$$

where dP is the infinitesimal power of the observed ray, $d\nu$ is the infinitesimal observed bandwidth, θ is the angle between the normal of the surface described by $d\sigma$ and the direction toward the observed infinitesimal region of the sky, described by $d\Omega$. The units of the specific intensity I_ν is W m^{-2} Hz^{-1} sr^{-1}. I will use both names, specific intensity and specific

brightness, to describe I_ν. For simplicity, I will drop the addition 'specific' in the following. However, I want to highlight that the total intensity is the specific intensity integrated over the frequency. It is important to distinguish the intensity from the flux density (also called simply 'flux') of a source which is the intensity integrated over the area Ω_S which is covered by the source:

$$S_\nu = \int_{\Omega_S} I_\nu(\theta, \varphi)\, cos(\theta)\, d\Omega. \tag{1.2}$$

The units of the flux density is W m^{-2} Hz^{-1}. As most astronomical objects are very weak it is common to use the unit 'Jansky', which is defined as:

$$1\ \text{Jy} \equiv 10^{-26}\ \text{W}\ \text{m}^{-2}\ \text{Hz}^{-1}. \tag{1.3}$$

Hence, the unit for the intensity can be Jy $beam^{-1}$ or Jy px^{-1}. In radio astronomy it is also common to describe the intensity as a brightness temperature (see Eq. 1.12 for further details) and use the unit Kelvin. For unresolved sources the values of the flux density and intensity are the same, whereas for resolved sources the value of the flux density is always larger than the value of the intensity. This will be important to distinguish between resolved and unresolved sources.

Without further proof, I want to highlight that the intensity is constant along any ray without diffraction or extinction (see e.g. Wilson et al. (2010) for a detailed proof). Therefore the intensity is independent of the distance and a direct source property. In contrast to this, the flux of a source is distance dependent ($S_\nu \propto d^{-2}$) and depends on the location of the observer and is not a direct source property.

The next important quantity is the spectral energy density u_ν. This is the flow of power per unit area divided by the speed of the flow, which is the speed of light c:

$$u_\nu = \frac{1}{c} \int I_\nu\, d\Omega \tag{1.4}$$

1.3.2 Basic principles and definitions for emission and absorption

In this section, I will describe the interaction of radiation with matter on a macroscopic level, including emission and absorption processes. Later in this chapter, I will connect this to a microscopic level using the Einstein coefficients. As explained in the previous section, the intensity stays constant along any ray in an empty space. However, it can change via interactions with matter. For example, the frequency dependent intensity (and hence the color) of a star can change by an interstellar cloud between the star and us. This interaction is specified with the absorption coefficient κ_ν and the emission coefficient ϵ_ν. The change in intensity follows then:

$$dI_\nu = (-\kappa_\nu\, I_\nu + \epsilon_\nu)\, ds, \tag{1.5}$$

where ds describes the infinitesimal small thickness of the material. This equation is also called the 'equation of radiative transfer', or simply 'equation of transfer'. I want to highlight, that the absorption depends on the strength of the radiation field, whereas the emission is independent from the present intensity. The emission coefficient ϵ_ν can be expressend in units of Jy beam^{-1} cm^{-1}. The absorption coefficient κ_ν can be written in units of cm^{-1} and hence is the mean number of photon absorptions per unit length. The absorption coefficient can also be substituted by the optical depth τ, which is defined as:

$$d\tau \equiv -\kappa\, ds. \tag{1.6}$$

Some authors also define the optical depth positive (e.g., Draine 2011). To simplify the equation of transfer (Eq. 1.5), we define the so called source function:

$$S_\nu \equiv \frac{\epsilon_\nu}{\kappa_\nu}. \tag{1.7}$$

Using the definition of the source function and the definition of the optical depth, we can rewrite the equation of transfer in a different form:

$$dI_\nu = (I_\nu - S_\nu)\, d\tau. \tag{1.8}$$

Simple solutions for this equation can be found for the case of only-absorbing or only-emitting clouds. However, the general solution of a emitting and absorbing cloud is more complicated and can be found by multiplying Eq. 1.8 with $e^{-\tau}$ and integrating by parts (details are given in Wilson et al. 2010) and is:

$$I_\nu(s) = I_\nu(0)\, e^{-\tau_\nu(s)} + S_\nu \left(1 - e^{-\tau_\nu(s)}\right). \tag{1.9}$$

The first term describes the background radiation, which gets absorbed by the foreground cloud and therefore decreases exponentially. The second term characterizes two parts: the emitted radiation of the cloud and the absorption of this radiation by the cloud itself.
If we assume the cloud is in thermodynamic equilibrium (TE), the temperature is constant within the cloud and the intensity is described by the Planck function $B_\nu(T)$:

$$B_\nu(T) = \frac{2h\nu^3}{c^2} \frac{1}{\exp\left(\frac{h\nu}{k_B T}\right) - 1}, \tag{1.10}$$

where h and k_B describe the Planck constant and Boltzmann constant, respectively. As the intensity is constant within the cloud, $dI_\nu = 0$ and the source function S_ν becomes the Planck function $B_\nu(T)$:

$$S_\nu = B_\nu(T) = \frac{\epsilon_\nu}{\kappa_\nu}. \tag{1.11}$$

This is known as Kirchhoff's law. In radio astronomy, it is common to describe the intensity field with the brightness temperature $T_B(\nu)$, which is the temperature that a black body would have at the given intensity field. This can be calculated by using the Planck function (Eq. 1.10):

$$T_B(\nu) = \frac{h\nu}{k_B} \frac{1}{\ln\left(1 + \frac{2h\nu^3}{I_\nu c^2}\right)} \tag{1.12}$$

However, this is not a linear relation between the brightness temperature and intensity. In radio astronomy it is possible to use the Rayleigh Jeans approximation instead to get a linear relation between the brightness temperature and the intensity. Formally this is called the antenna temperature:

$$T_A(\nu) = \frac{c^2}{2k_B\nu^2} I_\nu. \tag{1.13}$$

As long as $k_B T_A(\nu) \gg h\,\nu$, we find that $T_B(\nu) \approx T_A(\nu)$. This condition is fulfilled for long wavelength radio astronomy and in the following we will talk about brightness temperature and antenna temperature synonymously. Hence, we can write the solution of the equation of transfer in terms of temperatures:

$$T_B = T_B(0)\, e^{-\tau_\nu(s)} + T\left(1 - e^{-\tau_\nu(s)}\right), \tag{1.14}$$

where T_B is the measured brightness (or antenna) temperature and T describes the excitation temperature of the medium. This solution will be the basis for further analysis of the H I line.

1.3.3 Emission and absorption for discrete energy levels

The basis for the emission and absorption processes of the 21cm H I line is the interaction of photons with matter. For discrete energy levels, such as the hyperfine splitting of the neutral hydrogen (see Sect. 1.3.4), this interaction is described by the Einstein coefficients. Figure 1.1 shows a two level system with an upper and lower energy level labeled as E_u and E_l, respectively. The system can emit or absorb photons and hence change the energy levels. This interaction can occur in three different ways. First, the spontaneous emission described with the Einstein coefficient A_{ul}, at which the system changes spontaneously the energy level from the upper to the lower level and emits a photon with frequency $\nu_{ul} = \frac{(E_u - E_l)}{h}$. Second, the stimulated emission described with the Einstein coefficient B_{ul}. This transition is triggered by an external photon with frequency ν_{ul} and the system changes from the upper to the lower energy level by releasing a second photon with the same frequency. The third and last interaction is the spontaneous absorption described with the Einstein coefficient B_{lu}, at which the system changes from the lower to the upper energy level by absorbing a photon with frequency ν_{ul}.

For a stationary system the emission and absorption of photons is in equilibrium and the level

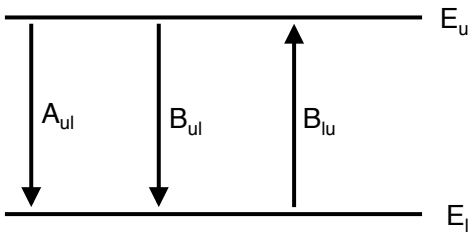

Figure 1.1: Einstein coefficient

population can be described by

$$n_u A_{ul} + n_u B_{ul} u = n_l B_{lu} u, \tag{1.15}$$

where n_u and n_l describe the density for the upper and lower level, respectively, and u describes the average energy density of the radiation field (see Eq. 1.4). If we assume that the system is in full thermodynamic equilibrium, the population of the energy levels is described by the Boltzmann distribution

$$\frac{n_u}{n_l} = \frac{g_u}{g_l} \exp\left(-\frac{h\nu}{k_B T}\right), \tag{1.16}$$

where g_u and g_l are the statistical weights of the corresponding upper and lower state. Combining Eq. 1.15 and 1.16 reveals

$$u = \frac{A_{ul}}{\frac{g_l}{g_u} \exp\left(-\frac{h\nu}{k_B T}\right) B_{lu} - B_{ul}}. \tag{1.17}$$

In thermodynamic equilibrium, the energy density of the radiation field can be described by Planck's law introduced in Eq. 1.10. Hence, Eq. 1.17 is only fulfilled if the Einstein coefficient fulfill the relations

$$g_l B_{lu} = g_u B_{ul} \tag{1.18}$$

$$A_{ul} = \frac{8\pi h\nu^3}{c^3} B_{ul}. \tag{1.19}$$

These relations will be very important to simplify the H I absorption coefficient and hence the calculation of the H I column density.

To connect the Einstein coefficients with the macroscopic emission and absorption coefficients, introduced in Sect. 1.3.2, we have to consider the contribution of the three different interaction methods to the intensity field. Similar to the discussion in Sect.1.3.2 we consider the intensity I_ν changing along a small length ds. As the energy levels E_u and E_l have a finite energy spread the energy of the emitted or absorbed photon can vary slightly. To describe this, we introduce the line profile function $\varphi(\nu)$, which is sharply peaked and normalized to one. For the emission

and absorption, this profile function can be different, however, for our consideration this is a negligible effect. The three energy contributions for the spontaneous emission, spontaneous absorption and stimulated emission are (Wilson et al. 2010):

$$\begin{aligned} dE_e &= h\nu\, n_u\, A_{ul}\, \varphi(\nu)\, d\sigma\, ds\, \frac{d\Omega}{4\pi}\, d\nu\, dt \\ dE_a &= h\nu\, n_l\, B_{lu}\, \frac{4\pi}{c} I_\nu\, \varphi(\nu)\, d\sigma\, ds\, \frac{d\Omega}{4\pi}\, d\nu\, dt \\ dE_s &= h\nu\, n_u\, B_{ul}\, \frac{4\pi}{c} I_\nu\, \varphi(\nu)\, d\sigma\, ds\, \frac{d\Omega}{4\pi}\, d\nu\, dt. \end{aligned} \tag{1.20}$$

The change of the intensity field is then:

$$dI_\nu\, d\Omega\, d\sigma\, d\nu\, dt = dE_e + dE_s - dE_a. \tag{1.21}$$

Combining Eq. 1.20 and 1.21 reveals:

$$\frac{dI_\nu}{ds} = \frac{h\,\nu}{4\pi}\, n_u\, A_{ul}\, \varphi(\nu) - \frac{h\,\nu}{c}\, (n_l\, B_{lu} - n_u\, B_{ul})\, I_\nu\, \varphi(\nu). \tag{1.22}$$

Comparing this equation with the equation of transfer (Eq. 1.5) provides the relation between the microscopic Einstein coefficients and the macroscopic emission and absorption coefficients:

$$\begin{aligned} \epsilon_\nu &= \frac{h\,\nu}{4\pi}\, n_u\, A_{ul}\, \varphi(\nu) \\ \kappa_\nu &= \frac{h\,\nu}{c}\, (n_l\, B_{lu} - n_u\, B_{ul})\, \varphi(\nu) \end{aligned} \tag{1.23}$$

The relation of the absorption coefficient with the Einstein coefficient gives us a direct handle to calculate the column density of the neutral hydrogen H I by measuring the spin temperature and the optical depth. To understand this, I will first introduce basic properties about the H I in the next section and then continue the calculation of the column density in Sect. 1.3.5.

1.3.4 Basic properties of the neutral hydrogen H I line

Neutral hydrogen is the most common and most simple atom in our universe, consisting of one proton and one electron. Important for the following discussion is the spin of this electron and proton, which can be parallel (total spin $S = 1$) or antiparallel ($S = 0$). These two states have different degeneracy $g = 2S + 1$ and a slight energy difference of $\Delta E = 5.97 \times 10^{-6}\,\mathrm{eV}$ as shown in Fig. 1.2. Even though the transition from parallel to antiparallel, called a spin flip, is a forbidden transition in quantum mechanics, the probability of this transition is nonzero. This transition releases a photon with the frequency $\nu_0 = \Delta E/h = 1.420405751766\,\mathrm{GHz}$ and a wavelength of $\lambda \sim 21.1\,\mathrm{cm}$ (Gould 1994). Hence, this transition is also called the '21cm line'. As this line is a forbidden transition, the Einstein coefficient $A_{ul} = 2.88426 \times 10^{-15}\,\mathrm{s}^{-1}$ is very small (Gould 1994) and therefore has an extremely long lifetime of ~11 million years. Since

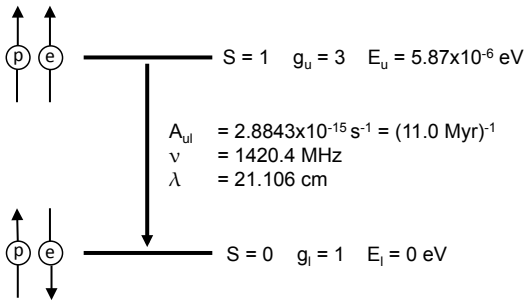

Figure 1.2: Energy levels of the hydrogen atom due to the hyper fine splitting. The parallel E_u and antiparallel E_l energy levels are shown, including the corresponding values (Draine 2011; Gould 1994).

the energy difference between the two spin states is small, low energy photons can excite the upper state and photons from the cosmic microwave background (CMB) are able to populate the upper state (Draine 2011). The excitation temperature for this transition is usually called the 'spin temperature' T_S and defined as

$$\frac{n_u}{n_l} = \frac{g_u}{g_l} \exp\left(-\frac{h\,\nu_0}{k_B\,T_S}\right). \tag{1.24}$$

Since the factor $h\nu_0/k_B = 0.068\,\mathrm{K}$ and T_S is usually above 10 K, the exponential function in Eq. 1.24 is approximately one and the population of the two spin states is $n_u/n_l \approx g_u/g_l = 3$. That means that 75% of the neutral hydrogen atoms are in the upper state and 25% are in the lower state, independent of the spin temperature. Therefore the emission coefficient ϵ_ν (Eq. 1.23) is also independent of the spin temperature. In contrast to this, the absorption coefficient κ_ν (Eq. 1.23) depends on the temperature, which will be discussed in the next section.

1.3.5 Derived properties of the neutral hydrogen H I

In this section, I will derive a formula to measure the column density of the H I, which depends solely on the spin temperature T_S and the optical depth τ. As we are usually not able to measure the extent of an H I cloud along the line of sight, we calculate the column density instead of the particle density, which is defined as

$$N_H \equiv \int n_H(s)\,ds. \tag{1.25}$$

To get the equation for the column density, I will start with the relation between the absorption coefficient κ_ν and the Einstein coefficient (Eq. 1.23) and include the relation of the two Einstein coefficients B_{ul} and B_{lu} (Eq. 1.18). This reveals:

$$\kappa_\nu = \frac{h\,\nu_0}{c}\,n_l\,B_{lu}\left(1 - \frac{g_l\,n_u}{g_u\,n_l}\right)\varphi(\nu). \tag{1.26}$$

As a next step, I include the spin temperature T_S (Eq. 1.24) and replace B_{lu} with A_{ul} (Eq.1.19). Then the absorption coefficient is:

$$\kappa_\nu = \frac{c^2}{8\pi\,\nu_0^2}\,\frac{g_u}{g_l}\,n_l\,A_{ul}\left(1-\exp\left(-\frac{h\,\nu_0}{k_B\,T_S}\right)\right)\varphi(\nu). \tag{1.27}$$

As explained in Sec. 1.3.4, the exponential function is close to one and we can perform a Taylor series. Furthermore, we plug in $g_u/g_l = 3$ and $n_l = \frac{1}{3}\,n_H$. Hence, the absorption coefficient is:

$$\kappa_\nu = \frac{3\,c^2}{32\pi\,\nu_0}\,n_H\,A_{ul}\,\frac{h}{k_B\,T_S}\,\varphi(\nu). \tag{1.28}$$

This shows that the absorption coefficient depends on the density and inversely on the spin temperature of the cloud. Therefore, we expect the strongest absorption from cold dense clouds, such as the CNM, whereas warm and more diffuse clouds, such as the WNM are optically thin. The final step is to plug in the definition of the optical depth (Eq. 1.6) and column density (Eq. 1.25) and integrate over s and ν. This reveals

$$N_H = 1.8224\times10^{18}\;T_S\int\tau(v)\,dv, \tag{1.29}$$

whereas N_H, T_S and dv are in units of cm^{-2}, K and $\mathrm{km\,s}^{-1}$. This formula is the basis to calculate the column density of any H I cloud. Even though it is a simple equation with solely two unknown, it is challenging to disentangle the optical depth and spin temperature. Different observation methods, including absorption studies are necessary to perform this task.
More details about this will be given in Chapter 2 and 4. In Chapter 2 we will use absorption studies toward strong continuum sources to determine the optical depth and use this information to correct the H I emission. Furthermore, we will use H I self absorption measurements in Chapter 4 to isolate the cold H I component.

1.4 Radio interferometry and synthesized imaging

The THOR survey is observed with the VLA in New Mexico, a state of the art radio interferometer. The data analysis of such an interferometer is challenging and can influence the results. Hence, it is an important step for this work and in the following section I will introduce the basics of a radio interferometer, the data calibration as well as the synthesized imaging. This is based on the books by Wilson et al. (2010) and Taylor et al. (1999) and the online course 'Essential Radio Astronomy' by J. J. Condon and S. M. Ransom, which can be found under www.cv.nrao.edu/course/astr534/.

1.4.1 Radio interferometry basics

For astronomical observations two aspects are important: sensitivity and resolution. A high sensitivity is important to see faint objects even at large distances and it can be increased by collecting more photons, either by long integration times or by a large collecting area. The angular resolution Θ is important to see details within the objects or disentangle different objects. For a single telescope, it is determined by the diameter of the telescope D and the observed wavelength λ:

$$\Theta \propto \frac{\lambda}{D}. \tag{1.30}$$

For radio astronomy, the wavelength is several orders of magnitude larger than in the optical regime and hence the diameter of the telescope must be accordingly larger to achieve the same resolution. Optical telescopes with diameters of ∼10 m can achieve resolutions well below $1''$. To achieve such a resolution in the radio regime at a wavelength of 21 cm, the telescope diameter must be ∼40 km. Technically, it is not possible to build a single telescope with such a large diameter. However, in radio astronomy it is possible to build interferometers and combine several smaller telescopes to a larger telescopes. The resolution of such an interferometry is not limited by the size of the single telescopes (also called 'dish' or 'antenna' in radio astronomy), but by the longest distance between two dishes, called the baseline b. Similar to Young's double slit experiment, the resolution is then:

$$\Theta \propto \frac{\lambda}{2b}. \tag{1.31}$$

The most simple interferometer is a two antenna interferometer. The signal from both antennas gets correlated, which means multiplied and integrated over a certain time, and the result is the so called spatial coherence function:

$$V_\nu(\boldsymbol{r_1}, \boldsymbol{r_2}) \approx \int I_\nu(\boldsymbol{s}) e^{-2\pi i \nu \boldsymbol{s}\cdot(\boldsymbol{r_1}-\boldsymbol{r_2})/c}\, d\Omega, \tag{1.32}$$

at which $\boldsymbol{r_1}$ and $\boldsymbol{r_2}$ describe the position of the two antennas and $\boldsymbol{s}$ describes the direction towards the observed source. To obtain the intensity I_ν of the source we have to perform a Fourier transformation of this equation. Such a simple two antenna interferometer has only a good resolution along the direction of the baseline and a poor resolution perpendicular to it. Therefore, we want to use multiple antennas at different locations to obtain as many different baselines as possible. The number of baselines n_b rises strongly with the number of antennas n_a and is:

$$n_b = \frac{\left(n_a^2 - n_a\right)}{2}. \tag{1.33}$$

Hence, the VLA with 27 antennas has 351 different baselines. As the earth rotates during the observations, these baselines also change in relation to the source. Hence, it is common to use

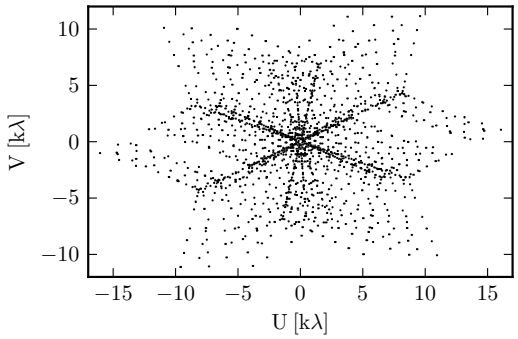

Figure 1.3: UV plane of a single pointing of the THOR observations towards galactic longitude $l = 32°$ in units of kλ at 21 cm.

a coordinate system, which describes the radio interferometer as seen by the source, called the 'UV-plane'. Fig 1.3 shows the UV-plane of one pointing of the THOR survey. It is common to show the coordinates of the baseline and the corresponding complex conjugated. The Y-shape of the VLA is clearly visible in the data. As we observed each pointing three times, this Y-shape can be seen three times, slightly shifted due to the rotation of the earth during the observation.

1.4.2 Karl G. Jansky Very Large Array

The Very Large Array (VLA) is a radio interferometer consisting of 27 antennas, each having a diameter of 25 m. All antennas are located on three arms in a Y-shape. It was build in the seventies close to the town of Socorro in New Mexico, USA. Starting around the year 2000, a major upgrade of the receivers, correlator and all electronics took place, which was finished in 2012 with the renaming of the array to Karl G. Jansky Very Large Array. As the antennas are movable on railway tracks, the array changes between four different configurations, approximately each half a year. The most compact configuration, called D-array, has a longest baseline of ∼1 km. The more extended configurations, called C-array, B-array and A-array have continuously larger baselines up to ∼36 km in the A-array. The THOR project was observed in the C-array. Different receivers (also called 'bands') allow observations from $\lambda \sim 6$ mm ('Q-band') up to $\lambda \sim 5$ m ('4-band'). Our observations were performed in the L-band between 1 and 2 GHz, corresponding to a wavelength of ∼15 to ∼30 cm.

1.4.3 Calibration

Details about the calibration can be found in Taylor et al. (1999) and the following section provides a summary. The calibration consists of several steps and was done with the

Figure 1.4: The left picture shows one arm of the VLA array including several antennas. The right picture illustrates the different receivers within one dish allowing a large coverage of different frequencies.

CASA (Common Astronomy Software Applications) package, provided by the NRAO (National Radio Astronomical Observatory). We tested to calibrate the data manually by hand or by using the VLA pipeline. The results were similar and we decided to use the VLA pipeline as this allowed us to calibrate the large amount of data faster. For the first part of the THOR survey, which is the data base of this thesis, we used the CASA version 4.1 and the VLA pipeline version 1.2. The second part of the THOR survey will be calibrated with the newer CASA version 4.2 and VLA pipeline version 1.3, which was not available for the first part. However, the differences between the two versions are minor and do not affect the data significantly. The pipeline, including a detailed description, can be found under https://science.nrao.edu/facilities/vla/data-processing/pipeline. In principle the VLA pipeline is a collection of python and CASA commands to perform all necessary steps. First, the pipeline performs some general flagging, that means it excludes data from the following calibration and analysis. This first flagging consists of shadowed antennas, data points that are exactly zero (called 'zeros'), pointing scans and the so called 'quacking', which means to flag a certain amount of time after the slewing of the antennas to allow them to settle down. As the sensitivity drops at the edge of the spectral bands, the pipeline also flags 5% of the channels on each side of the spectral bands. The next step is to perform the flux calibration by comparing the observed flux calibrator to a known model flux. After this step, the pipeline conducts an initial phase and amplitude calibration (called 'gain') and bandpass calibration and applies this to the calibrators. Using this first calibrated data, the pipeline performs automated flagging (so called 'RFlag', see Sect. 3.3.3 for further details) on the calibrators, determines its spectral index and corrects for this. After these corrections the final calibration solutions for the bandpass and gain are determined and applied to the actual data. The last step of the pipeline is to

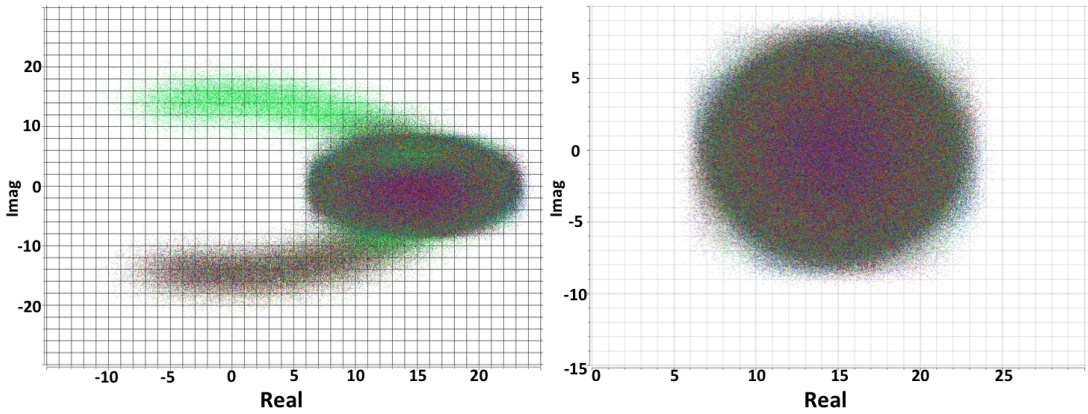

Figure 1.5: Example of diagnostic plots used for the calibration verification. Both plots show the imaginary part of the visibility as a function of the real part, which is equal to the phase as a function of the amplitude. The plots represent the flux calibrator 3C286 for the observation around galactic longitude $l = 21°$ of the H I line. The color code represents the first antenna of each baseline pair. On the left panel, a deviation from the expected round shape is visible, which is due to problems of a single antenna (in this example antenna ea17). The right panel shows the same data after flagging this 'bad antenna'.

create several diagnostic plots to check and control the pipeline. Optionally, the pipeline can perform Henning smoothing, automated flagging on the actual target data or recalculating the weights of the data. We did none of them at this stage, but performed automated flagging for some spectral windows prior to the cleaning process later.

After the pipeline run, we carefully checked the diagnostic plots and looked for any 'bad data', such as outlier antennas due to receiver or calibration problems or prominent Radio Frequency Interference (RFI). To improve this procedure, we included additional diagnostic plots. Very helpful were plots of the phase as a function of amplitude or plots of the amplitude as a function of time. Figure 1.5 shows an example of the phase as a function of the amplitude for the gain calibrator. Since the calibrators are distant quasars, we see them as point sources and the calibrated phase and amplitude should be constant and appear as a circular cloud as seen in the right panel of Fig. 1.5. In the left panel of Fig. 1.5, the phase and amplitude measured with one color-coded antenna are not constant. Such 'bad data' would influence our final results significantly and we have to flag them prior to the cleaning process. In addition, it could also influence the calibration itself. Hence, we run the pipeline again after flagging the 'bad antenna' to make sure that it is not affecting any calibration processes. If the quality checks for the calibration were appropriate, we continued with the cleaning process, which I will introduce in the next section.

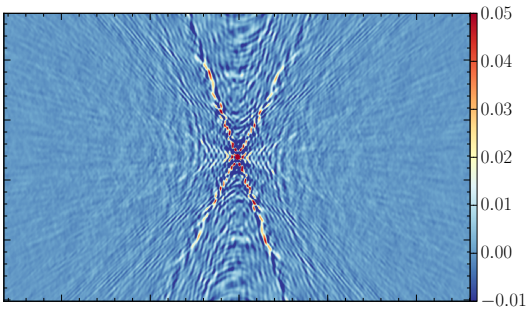

Figure 1.6: An example of the so called dirty beam. This is the the Fourier transformation of the UV-plane shown in Fig.1.3.

1.4.4 Synthesized imaging

As described in Section 1.4.1, the advantage of a radio interferometer is the high angular resolution. The observed intensity is reconstructed by a Fourier transformation of the data observed in the UV-plane. However, we cannot fill the UV-plane completely as there are gaps between the antennas. This can be seen in Fig. 1.3 as holes in the UV-plane. Hence, the Fourier transformation of this UV-coverage, which is the spatial resolution element of the array, will not be a simple 2D-Gaussian, but a more complex structure. It is called the 'dirty beam' and is shown in Fig. 1.6. The intensity distribution of our target sources contains the same structure as the dirty beam and is called the 'dirty image'. The top left panel of Fig. 1.7 shows an example of a dirty image. Around strong sources, we see artifacts which are called sidelobes. There are several different methods to remove these sidelobes and the most common is the 'clean algorithm'. This is an iterative algorithm, containing several steps:

1. Locate the highest peak in the dirty image or in the residuals for subsequent steps.
2. Record the position and intensity of this peak in the 'clean component' image (top right panel in Fig. 1.7).
3. Subtract a scaled dirty beam at this position from the dirty image. The scaling depends on the recorded intensity of the source. The result is called the 'residual image' and the final residual image is shown in the bottom left panel of Fig. 1.7.
4. Start the procedure again with the created 'residual image' and locate the next highest peak.

The loop is run multiple times until a certain threshold is reached. This threshold has to be several times larger than the expected noise in the image, otherwise the algorithm starts to clean within the noise, which could introduce artifacts. After reaching this threshold, the residual image should contain solely noise and the clean component image contains the position and intensity of the actual sources. The final image is created by convolving the final clean component image with the 'clean beam', which is a 2D-Gaussian fit to the main component of the

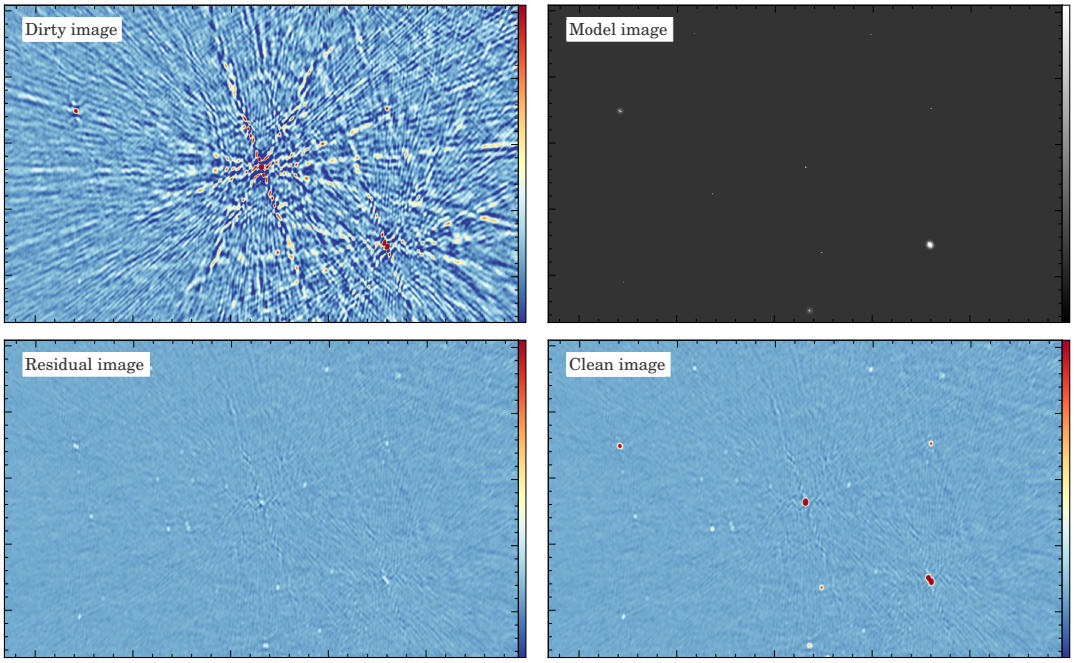

Figure 1.7: Example of the clean process. The top left panel shows the 'dirty image', prior to the clean process. Strong sidelobes are visible. The top right panel shows the 'model image', which contains the position and strength of the strongest sources, the so called 'clean components'. As we are using multiscale clean, this can be single pixels for point sources or more extended structures. The bottom left panel shows the 'residual image', which is the dirty image after subtracting all clean components convolved with the dirty beam (Fig. 1.6). The final 'clean image' is shown in the bottom right panel. All data is taken from the THOR survey and shows the continuum emission around 1.4 GHz, which is dominated by small extragalactic point sources in this region. The color scale of the dirty image, residual image and clean image is the same.

dirty beam. As a last step, the final residual image gets added to represent the noise properly and the result is called the 'clean image' (bottom right panel of Fig. 1.7), which should not contain strong sidelobes.

To avoid confusion, I want to stress that there are several different 'beams' for a radio interferometer. The 'primary beam' describes the beam of a single dish and hence is the field of view of the interferometer. The 'dirty beam', is the Fourier transformation of the UV-plane, whereas the 'clean beam' is a 2D-Gaussian fit to the main component of this dirty beam and hence the resolution element of the observations.

We used also the CASA package to perform the clean task. Within CASA there are several important extensions of the classical clean algorithm. For the THOR project, the most important is the ability to clean large mosaics, consisting of multiple pointings. For such a large mosaic, non-coplanar baselines can introduce strong artifacts. Hence, we have to consider them, which

is done in CASA with the 'w-projection' algorithm (Cornwell et al. 2008). Simply spoken, this algorithm assumes, that the VLA antennas are on different planes and calculates corrections for this effect. As the Galactic emission of the H I line and the continuum can be extended, it is also important to recover extended structures during the clean process. The classical clean algorithm searches only for point sources. However, the CASA package includes a 'multiscale algorithm' (see Rich et al. (2008), Cornwell (2008) for details). This algorithm searches not only for point sources, but also for extended structures and improves the clean result significantly.

For the H I data cubes, we had to clean ~150 channels separately, which requires an immense computer power, especially for the random access memory (RAM). To overcome this problem, we cleaned chunks of ten channels at the same time and added them at the end. The final clean task for e.g. the continuum was:

```
clean(vis=vis_name, imagename=image_name, field='*',
    spw=spw_number, mode='mfs', psfmode='hogbom',
    imagermode='mosaic', ftmachine='mosaic', imsize=image_size,
    cell=[2.5,2.5] ,weighting='briggs', robust=0, niter=100000,
    threshold='5mJy/beam', gain=0.1, phasecenter=phase_coordinate,
    wprojplanes=128, pbcor=True, multiscale=[0,4,12,25],
    cyclefactor=4,gridmode='widefield')
```

Further details about the exact cleaning process are given in each chapter.

1.4.5 Large scale structure, missing flux and noise

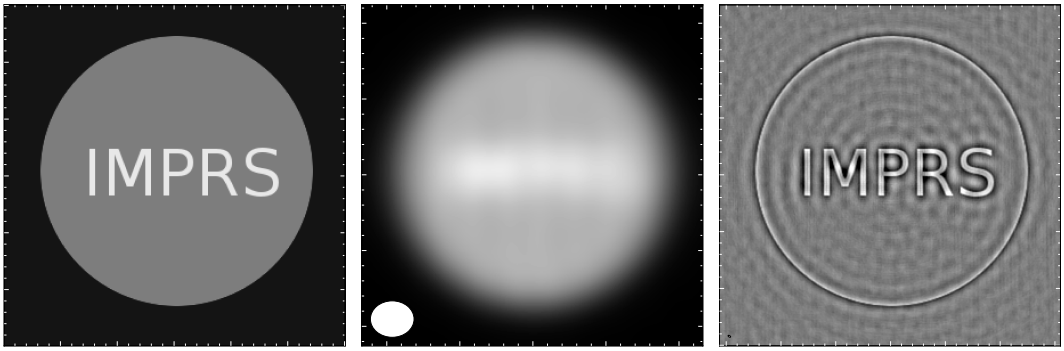

Figure 1.8: Simulated observations of the logo of the 'International Max Planck Research School' (IMPRS). The left panel shows the input, the central panel the simulated observations with the VLA in D-array and the right panel the VLA in A-array considering only large baselines.

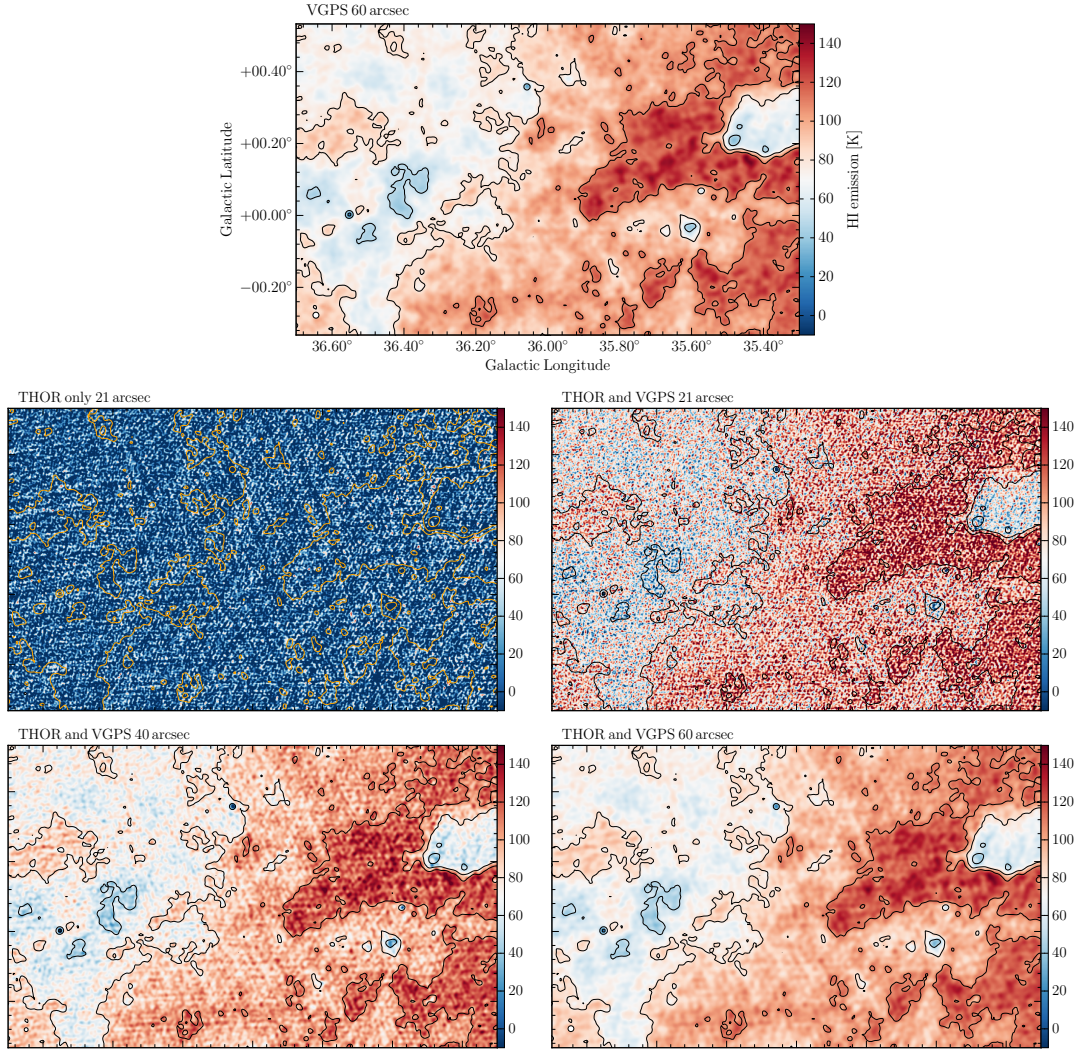

Figure 1.9: The comparison of different data sets. Each panel shows the H I emission of a single channel at $v_{\mathrm{LSR}} = 51.5\,\mathrm{km\,s^{-1}}$ in color and the contours indicate the VGPS data as reference. The top panel shows the low-resolution VGPS data at 60″. The lower panels show the THOR only or the combined data at different resolutions. The corresponding resolution is given as a title for each panel.

As explained, a radio interferometer can achieve high angular resolutions due to long baselines. However, the downside of each radio interferometer is the filtering of large scale structure due to missing UV coverage in the center. Figure 1.8 illustrates this effect for the VLA in different configurations for simulated observations of the IMPRS logo (‘International Max Planck Research School’). The left panel presents the input for the simulated observations and it consists of a black background, a uniformly filled circle and the text. The middle panel shows the logo observed with the smallest VLA configuration (D-array). For this configuration, we recover most of the extended emission of the circle, but the resolution is not good

Table 1.1: Brightness sensitivity comparison.

Data	Resolution [arcsec]	Brightness Sensitivity [K]
THOR only	21	16
Combined	21	16
Combined	40	3.9
Combined	60	1.8
VGPS only	60	1.5

Noise measured for region around $l = 35°$ in emission free channel at $v_{\rm LSR} = -85\,{\rm km\,s^{-1}}$.

enough to identify the text. The right panel represent the VLA in the most extended configuration (A-array) and only considering long baselines. The resolution for this simulation is high enough to identify the text, however, the large scale emission from the circle got filtered out.

This filtering effect is also present in the THOR data. To overcome it, we combine the H I 21cm line data with existing data from the VGPS survey (Stil et al. 2006). This survey used the VLA in the D-array, including single dish data from the Robert C. Byrd Green Bank Telescope (GBT), which has a diameter of 100-100 m. The software package CASA has several methods to combine low-resolution with high-resolution data and we will discuss them in more detail in Chapter 2.

Beside the filtering effect, another important aspect is the noise within the data. The point source sensitivity of an interferometer does not depend on the configuration of the array, but solely on the collecting area. For extended sources, the point source sensitivity is not crucial, but the brightness sensitivity which is in units of Kelvin. The brightness sensitivity becomes worse for more extended arrays, as the holes between the antennas become larger. Hence, we expect a worse brightness sensitivity for the high-resolution THOR survey in comparison with the low-resolution VGPS survey. Figure 1.9 shows a comparison of the H I data of the THOR survey only, the combined data at different resolutions and the VGPS survey only. The top panel shows the low resolution VGPS data, including single dish data. The Galactic H I emission is clearly visible for this data set, including self absorption features, which we will discuss in more detail in Chapter 4. In contrast to this, the high-resolution THOR data without the combination of the VGPS data (central left panel in Fig. 1.9) filters most of this emission. In the combined data, the H I emission appears, however, with a higher noise pattern. Smoothing this data to lower resolutions helps to improve the noise. The corresponding brightness sensitivity is given in Table 1.1. Even though we do not see the large scale emission of the H I line in the THOR data alone, this data set provides us excellent absorption spectra towards continuum background sources, which are mostly not resolved.

With the THOR data and the VGPS data at hand, we can choose which resolution and sensitiv-

ity we want for different science questions. For studies about large scale molecular clouds, the VGPS data with low noise is sufficient, whereas absorption studies toward continuum sources favor the high spatial resolution of the THOR data. Hence in the following chapters we will use the THOR data for absorption studies (Chapter 2) and the continuum emission (Chapter 3) and we will use the VGPS data for H I emission and self absorption studies (Chapter 4).

1.5 Research Projects

This thesis is split in three separate Chapters. Chapter 2 describes a study of the GMC complex around W43. We use strong continuum sources to determine the optical depth of the H I line and use this to correct the H I emission for optical depth effects and weak continuum emission. Hence, we can measure a more accurate column density and mass of this cloud and compare this result to the molecular content and test theoretical models. In Chapter 3 we extract a catalog of continuum sources from the THOR survey. Even though this is not directly related to molecular cloud studies, this is an important database for prospective absorption studies of the H I line. Due to the broad bandpass between 1 and 2 GHz, we are able to determine the spectral index for strong sources and hence identify the nature of the source. This allows us to distinguish between Galactic and extragalactic sources, which is an important information for absorption studies. Chapter 4 presents a study of H I self absorption of the long molecular filament GMF38.1-32.4. With this method, we can separate the cold H I component and compare this directly to the molecular counterpart. We determine column density PDFs for the atomic as well as the molecular phase, which are well represented by log-normal shapes, indicating turbulent motion.

2

The pilot study - H I observations of the giant molecular cloud W43

This chapter is published in A&A (Bihr et al. 2015) and slightly adapted for this thesis.

2.1 Abstract

The THOR survey started with the pilot region around the giant molecular cloud (GMC) complex W43 in 2012. In this chapter, we focus on the H I emission from this complex. Classically, the H I 21cm line is treated as optically thin with properties such as the column density calculated under this assumption. This approach might yield reasonable results for regions of low-mass star formation, however, it is not sufficient to describe GMCs. We analyzed strong continuum sources to measure the optical depth along the line of sight, and thus correct the H I 21cm emission for optical depth effects and weak diffuse continuum emission. Hence, we are able to measure the H I mass of this region more accurately and our analysis reveals a lower limit for the H I mass of $M = 6.6_{-1.8} \times 10^6\,\mathrm{M}_\odot$ ($v_{\mathrm{LSR}} = 60 - 120\,\mathrm{km\,s^{-1}}$), which is a factor of 2.4 larger than the mass estimated with the assumption of optically thin emission. The H I column densities are as high as $N_{\mathrm{H\,I}} \sim 150\,\mathrm{M}_\odot\,\mathrm{pc}^{-2} \approx 1.9 \times 10^{22}\,\mathrm{cm}^{-2}$, which is an order of magnitude higher than for low-mass star formation regions. This result challenges theoretical models that predict a threshold for the H I column density of $\sim 10\,\mathrm{M}_\odot\,\mathrm{pc}^{-2}$, at which the formation of molecular hydrogen should set in. By assuming an elliptical layered structure for W43, we estimate the particle density profile. For the atomic gas particle density, we find a linear decrease toward the center of W43 with values decreasing from $n_{\mathrm{H\,I}} = 20\,\mathrm{cm}^{-3}$ near the cloud edge to almost $0\,\mathrm{cm}^{-3}$ at its center. On the other hand, the molecular hydrogen, traced via dust observations with the Herschel Space Observatory, shows an exponential increase toward the center with densities increasing to $n_{\mathrm{H_2}} > 200\,\mathrm{cm}^{-3}$, averaged over a region of $\sim$10 pc. While atomic and molecular hydrogen are well mixed at the cloud edge, the center of the cloud is strongly dominated by H_2 emission. We do not identify a sharp transition between hydrogen

in atomic and molecular form. Our results, which challenge current theoretical models, are an important characterization of the atomic to molecular hydrogen transition in an extreme environment.

2.2 Introduction

It is well known that stars form in giant molecular clouds (GMCs; Mac Low & Klessen 2004; McKee & Ostriker 2007; Dobbs et al. 2014; Offner et al. 2014), which primarily consist of molecular hydrogen. Yet it is still under debate whether molecular hydrogen is actually necessary for star formation or whether molecular hydrogen and stars only form under the same conditions side by side (Glover & Clark 2012). The density within these clouds is high enough (particle density: $n > 1-100\,\mathrm{cm}^{-3}$, column density: $N > 1-100\,\mathrm{M}_{\odot}\,\mathrm{pc}^{-2} \sim 10^{20}-10^{22}\,\mathrm{cm}^{-2}$) for molecular hydrogen to become self-shielded from the interstellar radiation field, which would otherwise dissociate the H_2 molecules (e.g., Dobbs et al. 2014). Hence, molecular clouds form in the interior of large clouds of neutral hydrogen, which themselves are the environment of molecular clouds. Another open question is the fraction of neutral hydrogen within molecular clouds and how this affects physical conditions (Krčo & Goldsmith 2010; Goldsmith & Li 2005). Furthermore, the correlation between the transition between these two fundamental states of hydrogen and the corresponding physical conditions is also still unclear. As cold molecular hydrogen is challenging to observe directly, it is difficult to study its distribution in detail. Classically, the low-J rotational transitions of CO are used as a tracer for H_2, however, recent simulations and observations show that a large amount ($\sim 40\%$) of H_2 is CO-dark and therefore not well traced by CO (Pineda et al. 2013; Smith et al. 2014). Another approach to study molecular hydrogen is via observations of thermal dust emission or dust extinction (Lada et al. 2007; Molinari et al. 2010; Kainulainen & Tan 2013; Kainulainen et al. 2013). In contrast, the 21cm spin-flip transition of hydrogen offers a well-known method to measure the atomic gas content. Even though the 21cm line is well studied (e.g., Radhakrishnan et al. 1972; Gibson et al. 2000; Taylor et al. 2003; Gibson et al. 2005a; Heiles & Troland 2003a,b; Strasser & Taylor 2004; Goldsmith & Li 2005; Stil et al. 2006; Kalberla & Kerp 2009; McClure-Griffiths et al. 2012; Roy et al. 2013a; Fukui et al. 2015; Murray et al. 2014; Motte et al. 2014), it is difficult to disentangle the different contributions of the cold and warm H I in emission and absorption for different spin temperatures and optical depths. In addition, radio continuum emission at the frequency of the H I emission line can also suppress the intensity of the observed H I emission, an effect that is especially significant for the Galactic plane.

To address cloud formation, H I to H_2 formation as well as many other issues in ISM studies of the Milky Way, we initiated the THOR project: The H I, OH, Recombination Line Survey of the

Milky Way (Beuther et al. in prep.). The field chosen for the pilot study is around the massive star-forming complex W43 (l = 29.2 – 31.5°, $|b| \leq 1°$). This region is situated at the intersection of the Galactic bar and the first spiral arm (Scutum-Centaurus Galactic arm, Nguyen Luong et al. 2011; Carlhoff et al. 2013), leading to complex kinematic structures and possibly high star formation activity. The complex W43 is referred to as a Galactic mini-starburst region (Motte et al. 2003; Bally et al. 2010) and shows a star formation rate of $\sim 0.1\,M_\odot\,yr^{-1}$ (Nguyen Luong et al. 2011) or 5-10% of the star formation rate in the entire Milky Way. Motte et al. (2014) found velocity gradients in CO and H I 21cm emission. These velocity gradients could indicate large scale velocity flows, which could cause vast star formation activity. Motte et al. (2014) also do not find a threshold for the H I column density, which is proposed by theoretical models (Krumholz et al. 2008, 2009). They argue that we see several transition layers of H I and H_2 along the line of sight and that the theoretical models are not suited to describe complicated molecular cloud complexes such as W43. The center of W43 harbors a large H II region, which is fueled by a Wolf-Rayet and OB star cluster (Liszt et al. 1993; Lester et al. 1985; Blum et al. 1999). Furthermore, W43 exhibits several high-mass starless molecular clumps, which are still in an early stage of star formation (Beuther et al. 2012; Louvet et al. 2014). Some massive dense clumps can potentially form young massive clusters, progenitors of globular clusters (Louvet et al. 2014; Nguyen-Lu'o'ng et al. 2013). The Bar and Spiral Structure Legacy survey (BeSSeL; Brunthaler et al. 2011; Reid et al. 2014) determined the distance to W43 to be 5.5±0.4 kpc from parallax measurements of methanol and water masers (Zhang et al. 2014). This result has to be treated cautiously, as none of the four masers used for the parallax measurements are spatially directly associated with W43-Main.

2.3 Observations and data reduction

2.3.1 VLA Observations

We mapped a $2° \times 2°$ field around W43 ($l = 29.2 - 31.5°$, $|b| \leq 1°$) during the 2012A semester with the Karl G. Jansky Very Large Array (VLA) in New Mexico in C-configuration (Project 12A-161). As we used Nyquist sampling at 1.42 GHz with a primary beam size of 32', we needed 59 pointings to cover the 4 square degree mosaic. We chose a hexagonal geometry for the mosaic, similar to the VGPS survey (Stil et al. 2006). This results in a smooth areal sensitivity function with fluctuations of less than 1% in the inner region and a decreasing sensitivity toward the edges of our field. We observed each pointing for 4 × 2 min, which results in an overall observation time of ten hours, including ~2 hours overhead for flux, bandpass, and complex gain calibration. The resulting uv-coverage for one pointing, after 4 × 2 min of observing time, is shown in Fig. 2.1. We performed the observations in two blocks each with

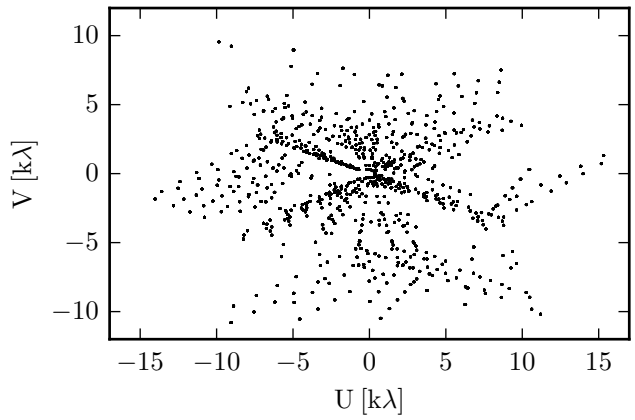

Figure 2.1: UV coverage of one pointing (centered at RA 18:47:54.125746, Dec -03:28:42.90941, J2000) after 4 × 2 min observing time.

5 hours observing time in April 2012. We chose the quasar 3C286 as a flux and bandpass calibrator and the quasar J1822-0938 as a complex gain calibrator, which was observed every ∼13 min. Observing at L-band and using the new WIDAR correlator, we were able to simultaneously observe the H I 21cm line, 4 OH lines (1612, 1665, 1667 and 1720 MHz) and 12 Hα RRL. For the pilot field, we had the spectral capability to observe 12 Hα RRL, however, for the full THOR survey we were able to observe 19 Hα RRL. The continuum, consisting of eight spectral windows between 1 and 2 GHz, was observed in full polarization. For the H I 21cm line, we used a bandwidth of 2 MHz with a channel width of 1.953 kHz. This results in a velocity range of ±200 km s^{-1} and a channel spacing of Δv ≈ 0.41 km s^{-1}.

2.3.2 Calibration

To edit and calibrate the data, we use CASA (version 4.1.0) with a modified VLA pipeline (version 1.2.0). The pipeline does automatic flagging for, e.g., zeros or shadowing of antennas. We manually performed additional flagging for radio frequency interference (RFI) and bad antennas. The pipeline also applies the bandpass, flux, and complex gain calibrator to the data. We do not use Hanning smoothing and do not recalculate the data weights ('statwt') at the end of the pipeline. The implemented modifications help to improve the quality check, and we do some flagging and editing by hand subsequently. A full description of our quality check method will be presented in our forthcoming overview paper (Beuther et al., in prep).

[a] https://science.nrao.edu/facilities/vla/data-processing/pipeline

2.3.3 Imaging and deconvolution

The H I emission and absorption covers a large range of angular scales, which is challenging for data reduction as the VLA C-Array filters out most of the large scale structure. Therefore, we combined our THOR data with the VGPS data (VLA D-Array and Green Bank Telescope data, Stil et al. 2006) to overcome missing flux problems and to reconstruct the large scale structure. We tested three different methods to combine the THOR and VGPS data. First, we tried to combine the visibility of the VGPS and THOR data and clean them together. Second, we tried the 'feather' command in CASA and third, we used the images of the VGPS survey as a starting model (parameter 'modelimage') for the clean process in CASA. After testing these three methods, we choose the third method, as this method provides the best results considering noise, side lobes, and recovery of large scale structure. We compared point source peak fluxes of the combined images to the VGPS data to check for consistency. The difference is at the 10% level. The clean process was stopped at the 5σ noise level.

We smoothed over three channels to reduce the noise, resulting in a velocity resolution of 1.24 km s^{-1}. This was the best compromise between computational time, noise, and a sufficient spectral resolution to distinguish absorption/emission features. The synthesized beam was set to 20", which is slightly larger than the best resolution achievable (~16"). The weighting parameter was set to $robust = 0.5$, which is a combination of natural and uniform weighting. These methods result in an rms of ~14 K (9mJy beam^{-1}) for emission free channels and up to a factor of 2 or 3 more in channels with strong emission due to systematic errors, such as side lobes. The dynamical range of our data set is ~100-200, depending on the region. In the following, the H I absorption and small scale structure is based on the THOR data, whereas the large scale emission is based on the VGPS data.

The 21cm continuum data are taken from the H I data cube for high and low velocities (-80 to -50 and 135 to 155 km s^{-1}), which are not affected by H I emission or absorption. Therefore the data reduction and imaging for the 21cm continuum data is the same as for the H I data, and we can avoid systematic errors due to different data reduction methods.

2.3.4 H_2 column density

As we will compare the H I and H_2 column density in Sect. 2.6.4, we need an estimate of the H_2 content. We use dust observations from the Herschel Space Observatory to assess the H_2 column density. These observations are based on the HiGal survey (Molinari et al. 2010). The H_2 column density map is taken from Fig. 9 of Nguyen-Lu'o'ng et al. (2013), which was derived by SED fitting methods described by Hill et al. (2011). As the dust observations have no velocity resolution, we see all the dust and thus gas along the line of sight. Carlhoff et al.

(2013) showed that the Herschel dust data are similar within the uncertainties to the CO data at the velocity range of W43 ($v_{\mathrm{LSR}} = 60 - 120\,\mathrm{km\,s^{-1}}$). Hence, the Herschel dust observations are dominated by the emission from W43 and the contributions of other regions along the line of sight can be neglected. We refrain from using CO data to estimate the molecular content of W43, as CO does not trace all molecular hydrogen (e.g., Smith et al. 2014) and CO becomes optically thick for the dense interior of W43 (Carlhoff et al. 2013).

2.4 H I radiative transfer

In this section, we explain the methodology we used to determine the spin temperature, optical depth and column density of the neutral hydrogen. Even though the H I 21cm line is a well-known probe of these quantities, the arrangement of different H I components with different temperatures along the line of sight can complicate these measurements. First we explain the classical methods to determine the column density via optically thin emission and H I continuum absorption (HICA). Subsequently, we outline the limitations of these methods and describe our approach to correct for optical depth and weak continuum emission. Although the basics of this approach are discussed in the literature (e.g., Wilson et al. 2010), we modified the classical approach to account for the continuum emission and optical depth. Therefore, we outline the calculations in more detail.

All following values are frequency dependent. To keep the equations and description simple, we drop the frequency dependencies in the equations. We emphasize, however, that the following emission/absorption mechanisms only works for identical frequencies and, therefore, identical velocities.

2.4.1 Column density

The classical approach to determine the H I column density is given by Wilson et al. (2010) as

$$\frac{N_{\mathrm{H}}}{\mathrm{cm^{-2}}} = 1.8224 \times 10^{18}\, \frac{T_{\mathrm{S}}}{\mathrm{K}} \int_{-\infty}^{\infty} \tau(v)\, d\left(\frac{v}{\mathrm{km\,s^{-1}}}\right), \tag{2.1}$$

where N_{H}, T_{S} and τ are the column density, the spin temperature, and the velocity dependent opacity, respectively. The spin temperature T_{S} describes the relative population of the spin states of the hydrogen atom (Wilson et al. 2010). As T_{S} is the equivalent of the excitation temperature for molecules, it is only equal to the kinetic temperature in local thermodynamic equilibrium (i.e., when there are enough collisions to thermalize the gas). We assume in the following that the spin temperature T_{S} does not vary significantly within one velocity channel. In most cases, neither T_{S} nor the optical depth are known. The simplest assumption is that the H I emission is optically thin (see Sect. 2.4.2).

2.4.2 Optically thin H I emission

Under the optically thin assumption, without background continuum emission, the expression for the brightness temperature T_B simplifies to the following:

$$T_B = T_S\,(1 - \exp(-\tau)) \approx T_S\,\tau. \tag{2.2}$$

This simplification provides a linear relation between the column density (equation 2.1) and the brightness temperature. Hence, we can estimate the column density directly from the measured brightness temperature T_B. This method is used in several studies and is well described in the literature (e.g., Lee et al. 2012; Motte et al. 2014; Wilson et al. 2010). Below, we discuss its limitations and describe a procedure to overcome them.

2.4.3 H I optical depth

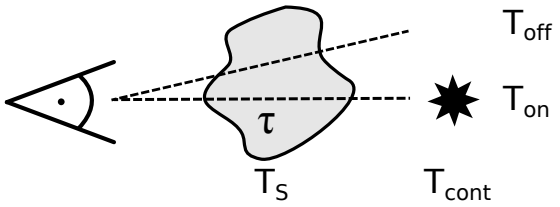

Figure 2.2: Sketch of the arrangement of the H I cloud with spin temperature T_S and optical depth τ and the continuum source in the background. The on and off positions are marked.

The H I continuum absorption method is the classical method to determine properties of the cold neutral medium (Strasser & Taylor 2004; Heiles & Troland 2003a,b; Strasser et al. 2007). This method is based on strong continuum sources ($T_B > 300$ K), such as Galactic H II regions, active galactic nuclei (AGNs), or extragalactic jets. As the brightness temperature of these continuum sources is larger than the typical spin temperature of the H I clouds ($T_S \sim 100$ K), we observe the H I cloud in absorption. In addition, the absorption spectrum is dominated by the cold neutral medium, as the H I absorption coefficient is proportional to T^{-1}.

The classical observing strategy is the 'on-off' observation (see Fig. 2.2). The 'on-source' points toward the continuum source and reveals the absorption spectrum (T_{on}), whereas the 'off-source' points slightly offset from the continuum source and reveals the emission spectrum (T_{off}). During data reduction, the continuum is not subtracted from the H I spectrum. Hence, we can measure T_{cont} for channels that are not affected by the H I line. The general assumption is that the on-source and off-source spectra originate from the same cloud with the same properties. Therefore, it is important to have these two positions as close together as possible.

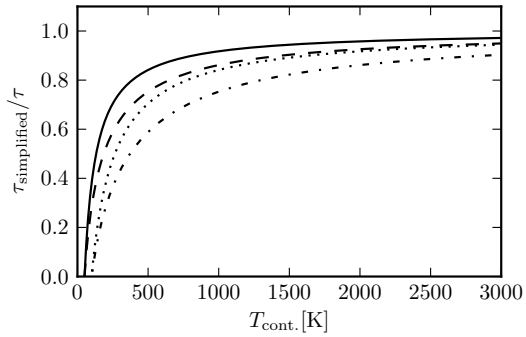

Figure 2.3: Optical depth simplification. The simplified optical depth is calculated by neglecting the spin temperature of the cloud. The solid and dashed lines represent a spin temperature of 50 K with optical depths of 1 and 2, respectively. The dotted and dash-dotted lines represent a spin temperature of 100 K with optical depths of 1 and 2, respectively.

The measured on-source and off-source brightness temperatures are

$$\begin{aligned} T_{\mathrm{on}} &= T_{\mathrm{S}}\,(1-\exp(-\tau)) + T_{\mathrm{cont}}\,\exp(-\tau), \\ T_{\mathrm{off}} &= T_{\mathrm{S}}\,(1-\exp(-\tau)). \end{aligned} \tag{2.3}$$

Hence, the optical depth is

$$\tau = -\ln\left(\frac{\mathrm{T_{on}} - \mathrm{T_{off}}}{\mathrm{T_{cont}}}\right). \tag{2.4}$$

The spin temperature can then be determined from

$$T_{\mathrm{S}} = \frac{T_{\mathrm{off}}}{1-\exp(-\tau)}. \tag{2.5}$$

The advantage of the HICA method is the direct measurement of the optical depth and the spin temperature. The challenge of the HICA method is the need for strong continuum sources. Since most of the strong continuum sources are point sources, it is not possible to map the entire Milky Way, but instead the result is an incomplete grid of measurements. This way, it is not possible to map the intrinsic structure of individual clouds. Furthermore, the spin temperature does not describe the actual temperature of the cloud, but is a mean of the cold and warm component, weighted by their column densities (Strasser & Taylor 2004). Consequently, the derived spin temperature is an upper limit for the cold component.

As some Galactic continuum sources, such as W43-Main, are extended, it is difficult to determine a proper off position. Nevertheless, it is still possible to determine the optical depth. As we use the VLA C+D array observations, including the continuum without the GBT data, we filter out most smooth large scale structure. The VLA C+D array data show H I emission of less than 30K. Therefore, we can neglect the emission of the H I cloud in equation 2.3 and

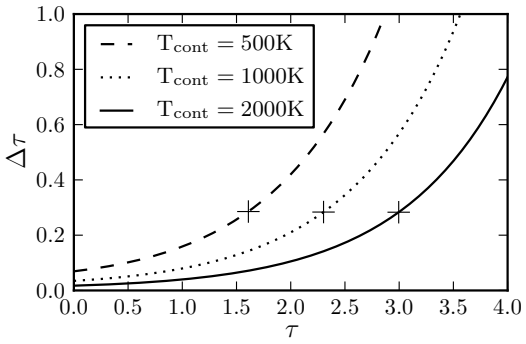

Figure 2.4: The absolute uncertainty of the optical depth as a function of the optical depth itself for three different continuum background sources. The dashed, dotted, and solid line represent continuum sources with a brightness temperature of $T_{\rm cont}$ = 500K, 1000K and 2000K, respectively. The crosses on each line show the lower limit of the optical depth that we can observe for the given continuum brightness temperature. For larger optical depths, the absorption spectra saturates (see Sec. 2.5.1 for more detail).

set $T_{\rm off} = T_{\rm S}\,(1 - \exp(-\tau)) = 0$ and calculate the optical depth without any off position as

$$\tau_{\rm simplified} = -\ln\left(\frac{\rm T_{on}}{\rm T_{cont}}\right). \tag{2.6}$$

Even without the effect of filtering out the H I emission due to the interferometer, this simplification holds for strong continuum sources ($\rm T_{cont} > 500K$). Figure 2.3 shows the relation between the simplified and actual optical depth as a function of the continuum brightness temperature, not taking the filtering of the interferometer into account. It clearly shows that even for high spin temperatures ($T_{\rm S}$ = 100 K) and high optical depth ($\tau = 2$), we miss less than 10% of the optical depth for a strong continuum source ($T_{\rm cont} \approx 2500$ K). Therefore, we are able to measure the optical depth for the H II region in the center of W43, even though we cannot determine a proper off position.

We also consider the effect of optically thick clouds. For these optically thick clouds, the absorption spectra approaches zero. Due to the rms of the spectra, it is also possible that the absorption spectra become negative, which is physically not meaningful. We set the optical depth to a lower limit for all absorption values that are close to zero. If the absorption spectrum $T_{\rm on}$ is smaller than 5 times the rms, we set the optical depth to

$$\tau_{\rm lower-limit} = -\ln\left(\frac{5\cdot\sigma(T_{\rm on})}{T_{\rm cont}}\right). \tag{2.7}$$

We find that the optical depth saturates for the main velocity range of W43-Main, which has consequences for our interpretation and conclusions. The uncertainty of the optical depth (equation 2.4) depends on $T_{\rm on}$, $T_{\rm off}$ and $T_{\rm cont}$. To estimate the uncertainty, we assume an uncertainty of 20K for all three quantities. The uncertainty of the optical depth is shown in Fig.

2.4 as a function of the optical depth itself for three different continuum brightness values. It shows that the uncertainty increases significantly for increasing optical depths. However, we are not able to measure these high optical depths, as the absorption spectra saturates for optically thick clouds, and we report lower limits (see equation 2.7). These lower limits are marked with crosses in Fig. 2.4. Up to the lower limit of the optical depth the corresponding uncertainty is ~0.3 for the three different continuum brightness values. For strong continuum sources, such as W43-Main, the uncertainty of the optical depth is ~10% up to the lower limit of ~3.

2.4.4 Column density corrections

In addition to distinct, small continuum sources, such as H II regions or extragalactic jets, we also find weak diffuse continuum emission in the Galactic plane. This component has a strength between 10 and 50 K. Therefore, it is not strong enough to induce absorption features (HICA), but nevertheless it can influence the H I emission. To overcome this problem, the classical approach is to subtract this weak continuum emission during data reduction. Hence, we subtract $T_{\rm cont}$ in equation 2.3, and we find

$$\begin{aligned} T_{\rm B} &= T_{\rm S}(1 - \exp(-\tau)) + T_{\rm cont}\ \exp(-\tau) - T_{\rm cont} \\ &= (T_S - T_{\rm cont})\ (1 - \exp(-\tau)). \end{aligned} \tag{2.8}$$

This shows that even if we subtract the continuum from our data, it can still influence the measured brightness temperature. If we neglect the weak continuum emission, our measured H I emission is suppressed and thus the calculated column density is underestimated. Therefore it is important to investigate the influence of the weak diffuse continuum emission.

In the following, we assess the assumption of optically thin emission and the influence of weak diffuse continuum emission on the determination of the measured column density for appropriate model clouds. We assume two model clouds, which have a spin temperature of 50 K and 100 K, respectively. We use equation 2.1 to calculate the expected column density $N_{\rm H\,I}$ for different optical depths in one velocity channel ($dv = 1.24\,{\rm km\,s^{-1}}$). Furthermore, we assume the cloud to be optically thin and use Eq. 2.2 to calculate the brightness temperature of the H I emission. Using this result, we can calculate the column density of the cloud, but this time with the simplification of optically thin emission. Hence, we call it the observed column density $N_{\rm H\,I}(observed)$. The solid lines in Fig. 2.5 show the ratio of the expected and observed column density for a range of optical depths. Obviously, for small optical depths ($\tau < 0.1$), our assumption of optically thin emission is sufficient and we observe more than 90% of the expected column density. For larger optical depths ($\tau > 0.1$), however, we miss a significant fraction of the column density (>40%).

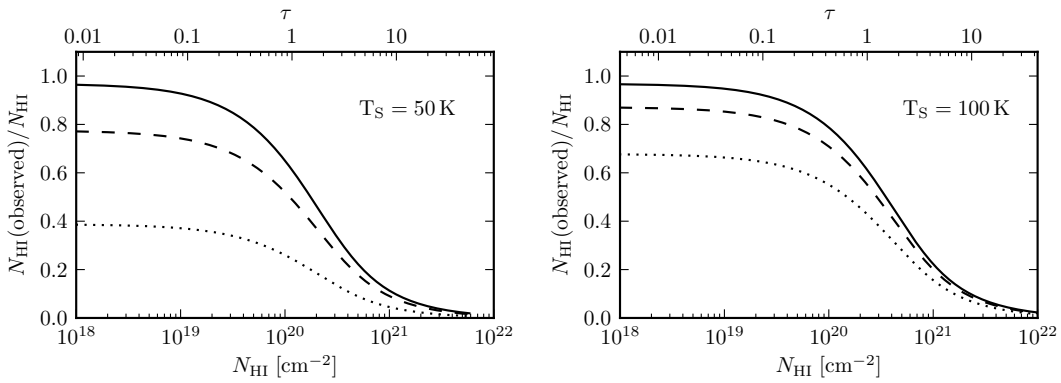

Figure 2.5: Ratio of observed and expected column density for a model cloud with spin temperatures of $T_S = 50\,K$ and $100\,K$ for the left and right panel, respectively. The solid lines show the ratio for the assumption of optically thin emission. The dashed and dotted lines show the ratio for the assumption of optically thin emission and a diffuse continuum source with brightness $10\,K$ and $30\,K$, respectively.

In addition, we add weak continuum emission in the background, which changes the brightness temperature according to Eq. 2.8 and therefore suppresses the H I emission. Nevertheless, if we still assume optically thin emission and do not consider the weak continuum emission, we can calculate the observed column density. The dashed and dotted lines in Fig. 2.5 show this case for continuum emission of $10\,K$ and $30\,K$, respectively. Even for small optical depths ($\tau < 0.1$) we miss a significant fraction of the column density, which depends on the ratio of the spin temperature and the continuum emission brightness temperature. In the worst case ($T_S = 50\,K$ and $T_{\rm cont} = 30\,K$), we observe only 40% of the expected column density, even for small optical depths.

To summarize, we measure the brightness temperature of the H I emission T_B, as well as the brightness temperature for the continuum $T_{\rm cont}$ and combine this information with the optical depth τ, which we measure toward strong continuum sources. This allows us to calculate the corrected column density, by combining Eqs. 2.1 and 2.8, which yields

$$N_{\rm H} = 1.8224 \times 10^{18} \left(\frac{T_{\rm B}}{1 - e^{-\tau}} + T_{\rm cont} \right) \int_{-\infty}^{\infty} \tau(v)\, dv, \tag{2.9}$$

where the column density N_H is given in units cm^{-2}, the brightness temperature T_B and the continuum brightness temperature $T_{\rm cont}$ are measured in K, and the velocity dv is given in km s^{-1}.

2.4.5 Continuum correction for strong point sources

For the H I continuum absorption (HICA) toward strong continuum sources, we also have to consider weak diffuse continuum emission, which contributes to the on and off positions.

Therefore we have to modify the equation for the on and off positions (Eq. 2.3) by adding another term for the weak diffuse continuum emission, which we call $T_{\rm cont,dif}$. This does not change our result for the optical depth (equation 2.4), as the optical depth depends solely on the difference between $T_{\rm on}$ and $T_{\rm off}$. For the calculated spin temperature, however, we have to modify Eq. 2.5 to the following:

$$T_{\rm S} = \frac{T_{\rm off} - T_{\rm cont,dif}\ e^{-\tau}}{1 - e^{-\tau}}. \tag{2.10}$$

The effect of the diffuse continuum correction is discussed in Sect. 2.6.2.

2.5 Results

2.5.1 Optical depths determined using compact continuum sources

As described in Sect. 2.4.3, we can use strong continuum sources in the background as lighthouses that shine through foreground H I clouds and create absorption spectra. Below we characterize the continuum sources.

To determine the optical depth of the H I accurately, we need continuum sources in the background which are brighter than the typical H I spin temperature (see Sect. 2.4.3). We extract all continuum sources in our field that have a brightness temperature $T_{\rm cont} > 400\,{\rm K}$ and a point-like structure, which yields eight point sources. The analysis of extended sources follows in Sect. 2.5.4. We use a two-dimensional Gaussian to fit the position and size of the continuum source. To measure the continuum brightness temperature, we average over high- and low-velocity channels in our H I data cubes (-80 to -50 and 135 to 155 $\rm km\,s^{-1}$) that are not affected by the H I line. The results are shown in Table 2.1.

To determine the off spectrum ($T_{\rm off}$), we average the emission spectrum around the continuum source within an annulus with inner and outer radius of 60"and 120", respectively. The upper two panels of Fig. 2.6 show a typical emission and absorption spectrum. We use Eq. 2.4 to calculate the optical depth for each channel. To avoid unrealistic optical depths, we only calculate the optical depth for those channels in which the emission/absorption is five times larger than the corresponding noise. The gray shaded areas in the upper three panels in Fig. 2.6 and 2.7 show the 5 sigma level. The optical depth is set to zero for channels, where the emission/absorption is below five times the corresponding noise.

As explained in Sec. 2.4.3, we have to consider the effect of optically thick clouds for which we can determine only lower limits for the optical depth. The gray shaded areas in the third panels of Fig. 2.6 and 2.7 show these lower limits of the optical depth.

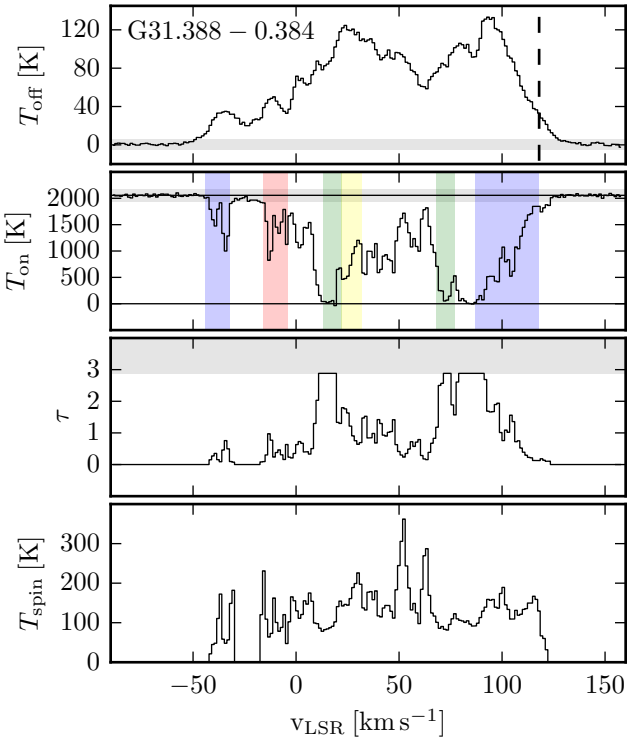

Figure 2.6: H I emission and absorption spectrum of the extragalactic continuum point source G31.388−0.384. The emission spectrum is shown in the first panel and is measured in an annulus around the point source with inner and outer radii of 60"and 120", respectively (corresponds to 3 and 6 times the restoring beam). The second panel presents the absorption spectrum toward the point source. The color shaded areas represent the approximate velocities of the Milky Way spiral arms (Vallée 2008) in blue, red, green, and yellow, the Scutum-Centaurus, Cygnus, Sagittarius, and Perseus arm, respectively; the dashed black line in the first panel indicates the tangential velocity. In the first two panels, the gray shaded area indicates the 5σ noise level. The third panel shows the optical depth computed using Eq. 2.4 and the gray shaded area indicates the saturated optical depth limit, computed using Eq. 2.7. In the fourth panel, the spin temperature is presented, which is computed using Eq. 2.10.

2.5.2 H I spin temperature toward compact continuum sources

For each channel that allows for an optical depth measurement, we compute the spin temperature using Eq. 2.10. The spin temperature is shown in the fourth panel of Fig. 2.6 and 2.7. This method reveals absorption features in 655 channels for all eight continuum sources. The median of the spin temperature is 97.5 K and the distribution of all absorption features is shown in Fig. 2.8. For large optical depths, the spin temperature approaches the brightness temperature of the off position, however, for large optical depths ($\tau \gtrsim 2$) we can only report lower limits for τ. Hence, we overestimate the spin temperature in the optically thick regime. Nevertheless, this overestimation is small, as for optical depths of $\tau = 2$ the measured spin temperature is

Table 2.1: Continuum point sources that were used for H I absorption studies.

Name	RA.(2000) (h : m : s)	Dec.(J2000) (°:′:″)	MAJ x MIN (″ x ″)	$T_{\rm cont}$ K	$T_{\rm cont,dif}$ K	far Scutum arm	spectral index	location
G31.388−0.384	18:49:59.1	-1.32.55.9	21.3 x 20.3	2055±24	20.2	yes	-0.82	ext. gal
G31.411+0.307	18:47:34.1	-1.12.45.5	24.0 x 23.0	914±26	19.4	no	0.63	gal.
G30.234−0.138	18:47:00.4	-2.27.52.9	21.7 x 20.7	853±23	42.1	no	0.00	gal.
G30.534+0.021	18:46:59.3	-2.07.27.9	28.0 x 21.5[a]	653±21	35.9	no	0.21	gal.
G31.242−0.110	18:48:44.7	-1.33.13.7	25.1 x 21.2	573±22	30.1	no	0.58	gal.
G29.090+0.512	18:42:35.8	-3.11.04.7	20.0 x 10.0	571±44	20.5	yes	-1.03	ext. gal
G30.744+1.008	18:43:51.4	-1.29.14.8	23.1 x 20.1	472±36	16.9	yes	-0.90	ext. gal
G30.699−0.630	18:49:36.3	-2.16.27.2	22.3 x 19.6	403±18	23.4	yes	-1.11	ext. gal

Continuum point sources in the observed field with brightness temperatures $T_{\rm cont} > 400$ K. The names and spectral indexes are based on work presented in Johnston et al. (in prep) and are preliminary results. Coordinates, major, and minor axes of the 2D fit, maximum brightness temperature $T_{\rm cont}$ of the continuum point source and the diffuse weak continuum temperature around the point source are given. The column 'far Scutum arm' indicates whether absorption features at $v_{\rm LSR} \approx -40\,{\rm km\,s^{-1}}$ are present in the H I spectra. The spectral index and the absorption of the far Scutum arm are used to determine the location of the point sources, which is given in the last column.
[a] This source consists of two blended sources, which explains the large eccentricity.

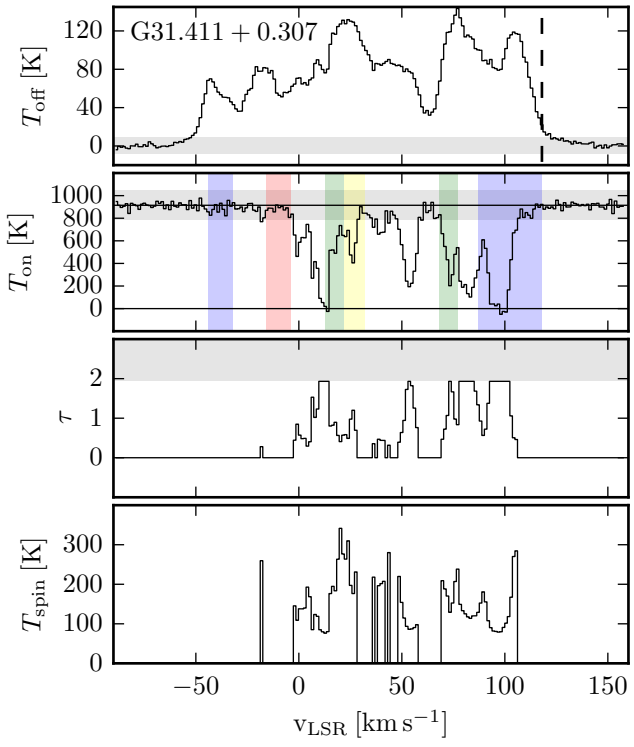

Figure 2.7: Same layout as Fig. 2.6, except for the Galactic continuum source G31.411+0.307

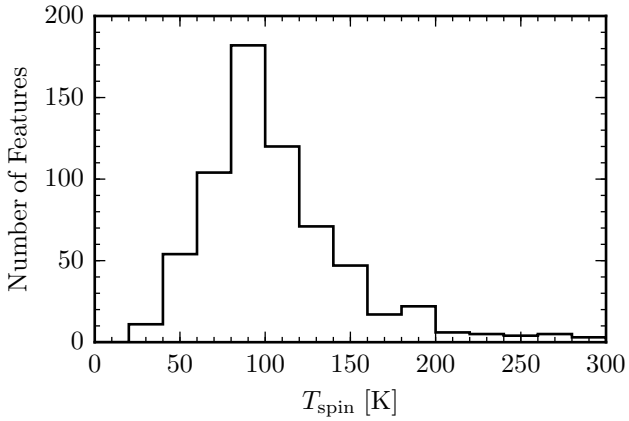

Figure 2.8: Distribution of the spin temperature for all absorption features (655) toward point-like continuum sources. The median of the spin temperature is 97.5 K.

$T_{\rm spin} \approx 1.16 \cdot T_{\rm off}$ (see equation 2.5). Therefore we overestimate $T_{\rm spin}$ by at most $\sim$ 15%. We discuss these results in detail in Sect. 2.6.2.

2.5.3 Location of continuum sources

To characterize and understand the H I absorption spectrum toward the continuum point sources, it is important to know the location of the continuum source. We can distinguish between extragalactic and Galactic point sources. The Galactic sources are mostly H II regions, whereas the extragalactic point sources are radio lobes from extragalactic jets or AGN. To distinguish between them, we have two criteria: the spectral index and H I absorption of the far Scutum-Centaurus arm. As we have to consider, e.g., primary beam effects or different spatial filtering of the interferometer for different frequencies, it is very difficult to determine accurate spectral indexes (e.g., Rau et al. 2014; Bhatnagar et al. 2011, 2013). For this analysis, we calculated the spectral indices, using flux measurements in the two most separated spectral windows at 1.05 and 1.95 GHz and are based on work presented in Johnston et al. (in prep.).

Four sources in our sample have negative spectral indices (Table 2.1), which is typical of synchrotron radiation from extragalactic jets or AGNs. The remaining four sources show flat or positive spectral indices, which indicates free-free emission from potentially optically thick H II regions. These four free-free emission sources also show no H I absorption feature for the far Scutum-Centaurus arm. Therefore, these sources reside in the Milky Way. Furthermore, we can study the H I absorption spectra and molecular emission spectra of the Galactic point sources to estimate their distance and overcome the near-far distance ambiguity. This is done by Anderson et al. (2014) and they find that, e.g., the source G31.411+0.307 has a near distance of 6.6 kpc. For this source, we also see a sharp cutoff in the absorption spectra at

$v_{\rm lsr} \approx 100\,{\rm km\,s^{-1}}$ (see Fig.2.7), which corresponds to the molecular line velocity reported in Anderson et al. (2014).

Figure 2.6 shows an example of an extragalactic continuum source. The characteristic absorption of the far Scutum-Centaurus arm at $v_{\rm lsr} \approx -40\,{\rm km\,s^{-1}}$ is clearly observable. Furthermore, the absorption and emission spectra show a similar cutoff for high velocities, when approaching the tangential point of the near Scutum-Centaurus arm. On the other hand, the absorption spectrum of Fig. 2.7 neither shows the absorption of the far Scutum-Centaurus arm, nor approaches zero at the velocity of the tangential point. This is typical for Galactic continuum sources.

2.5.4 Extended continuum sources - W43-Main

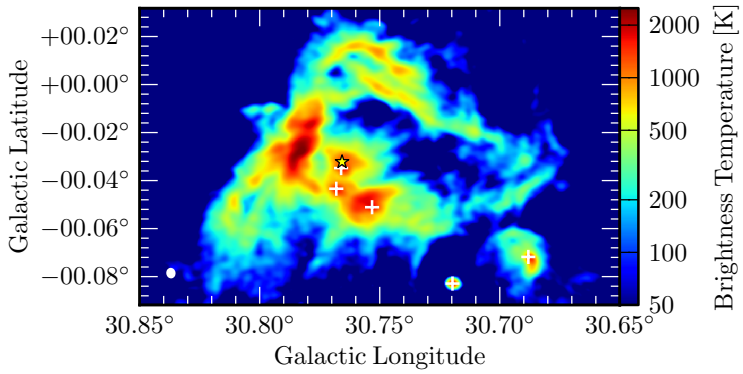

Figure 2.9: Continuum emission of 1.4 GHz of W43-Main. The yellow star indicates the position of the OB cluster (Lester et al. 1985) and the white crosses marks the position of UCH II regions observed by the CORNISH survey (Hoare et al. 2012; Purcell et al. 2013)

Besides point-like continuum sources, our observed field contains three strong ($T > 800\,{\rm K}$) extended continuum sources. The most prominent source is the well-known H II region around W43-Main (Lester et al. 1985), which is shown in Fig. 2.9. This region has an angular extent of $\sim 300''$, which corresponds to $\sim 8\,{\rm pc}$ at a distance of 5.5 kpc. As our resolution is 20", we are able to resolve the internal structure well. It is known that the nebula is ionized by an OB cluster and the observed continuum signal is the result of free-free emission. Figure 2.9 shows the continuum emission of W43-Main and the yellow star marks the position of the OB cluster (J2000, 18h47m36s, -1° 56′ 33″, Lester et al. 1985). Furthermore, several UCH II regions can be found in the CORNISH survey, which are marked with white crosses (Hoare et al. 2012; Purcell et al. 2013). The maximum brightness temperature for this region is $T_{\rm cont} \sim 2200\,{\rm K}$ and, therefore, we are able to calculate the optical depth, even though we

cannot determine a proper off position (see Sect. 2.4.3 for details). The H I absorption spectra shows a cutoff at $v_{\rm lsr} \sim 100\,{\rm km\,s^{-1}}$ (see Fig. 2.10), which marks the velocity of the continuum source. We measure radio recombination line emission at the same velocity (Johnston et al. in prep, Anderson et al. 2011). Therefore, the continuum source W43-Main is situated at $v_{\rm lsr} \approx 100\,{\rm km\,s^{-1}}$, which is close to the tangent point velocity of the Scutum-Centaurus arm (Nguyen Luong et al. 2011).

Furthermore, the observed field contains two other extended continuum sources: W43-South and the supernova remnant SNR G029.7-00.2 with continuum brightness temperatures of at most $T_{\rm cont} \approx 1840\,{\rm K}$ and $T_{\rm cont} \approx 850\,{\rm K}$, respectively.

2.5.5 Optical depth of W43-Main

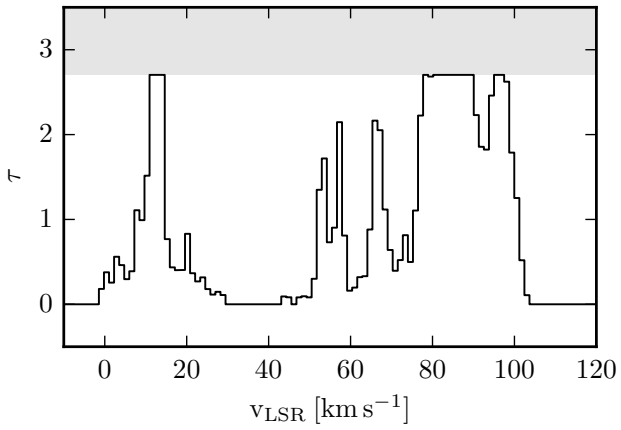

Figure 2.10: Optical depth of W43-Main, calculated using equation 2.6. The gray shaded area indicates the maximum measurable optical depth of $\tau_{\rm lower-limit} = 2.7$ calculated using Eq. 2.7.

In Sect. 2.4.3, we described that we can determine the optical depth for strong continuum sources using the absorption spectrum. As the brightness of the continuum source W43-Main is $T_{\rm cont} \sim 2200\,{\rm K}$, the uncertainty for the optical depth measurement is $\sim 10\%$ (see Fig. 2.3). Figure 2.10 shows the optical depth of W43-Main and the gray shaded area indicates the lower limit of our measurement with $\tau_{\rm lower-limit} = 2.7$. The optical depth peak at $v_{\rm lsr} \approx 10\,{\rm km\,s^{-1}}$ can be allocated to the near Sagittarius arm and therefore is not connected to the actual star-forming region W43. In contrast, it is difficult to allocate the distinct absorption features between 50 and $80\,{\rm km\,s^{-1}}$ and it is not clear whether they are spatially connected to W43. The prominent absorption feature of W43 is situated between 80 and $100\,{\rm km\,s^{-1}}$. In this region, our measurement is saturated and, therefore, we can only report lower limits for the optical depth.

As we resolve the strong continuum source W43-Main spatially, we can determine the optical

depth along different lines of sight and thus investigate the spatial distribution of absorption features as done in detail by Liszt et al. (1993). We refrain from this kind of study as we are mostly interested in the velocity range of W43 ($v_{\rm lsr} = 80 - 110\,{\rm km\,s^{-1}}$) at which the absorption features are mostly saturated preventing a detailed study of the spatial distribution. We instead measured the optical depth toward the strongest continuum peak to estimate the maximum optical depth possible (see Fig. 2.10), which is nevertheless a lower limit.

2.5.6 H I column density with optically thin assumption

In this section, we assess the H I column density and the H I mass of the GMC associated with W43. These two quantities depend strongly on the chosen velocity range. Nguyen Luong et al. (2011) defined for the 'main' and 'complete' velocity range of W43 values of $v_{\rm lsr}(main) = 80 - 110\,{\rm km\,s^{-1}}$ and $v_{\rm lsr}(complete) = 60 - 120\,{\rm km\,s^{-1}}$. In the following, we use the 'complete' velocity range, but we stress that a different velocity range can significantly change the column density and mass.

As described in Sect. 2.4.2, we can determine the column density by assuming optically thin H I emission. For this method, the column density is proportional to the observed H I brightness temperature. The column density map assuming optically thin emission is shown in the top right panel of Fig. 2.11. To estimate the mass, we integrate this column density over the main region of W43 ($l = 29.0 - 31.5°$, $|b| \leq 1°$). We mask all regions, where the emission spectrum has negative values, and exclude them. Given a distance of 5.5 kpc (Zhang et al. 2014), our analysis finds an H I mass of $M = 2.7 \times 10^6\,{\rm M_\odot}$ ($l = 29.0 - 31.5°$, $|b| \leq 1°$, $v_{\rm lsr} = 60 - 120\,{\rm km\,s^{-1}}$).

Weak diffuse continuum emission can influence the H I emission spectrum and therefore the H I column density calculation needs to be modified according to Eq. 2.8 (Sect. 2.4.4). The top right panel of Fig. 2.11 shows this effect clearly. The color represents the column density determined by the optically thin assumption and the black contours indicate the weak diffuse continuum emission. These two components show a clear anti-correlation. However, this anti-correlation is the result of the expected observational effect in which H I emission is suppressed by weak continuum emission (see Sec. 2.4.4). To overcome this problem, we have to consider the optical depth, which we discuss in the next section.

2.5.7 H I column density with optical depth correction

With the measurement of the optical depth and the weak continuum emission, we can correct the H I emission as explained in Sect. 2.4.4. This allows for a more accurate determination

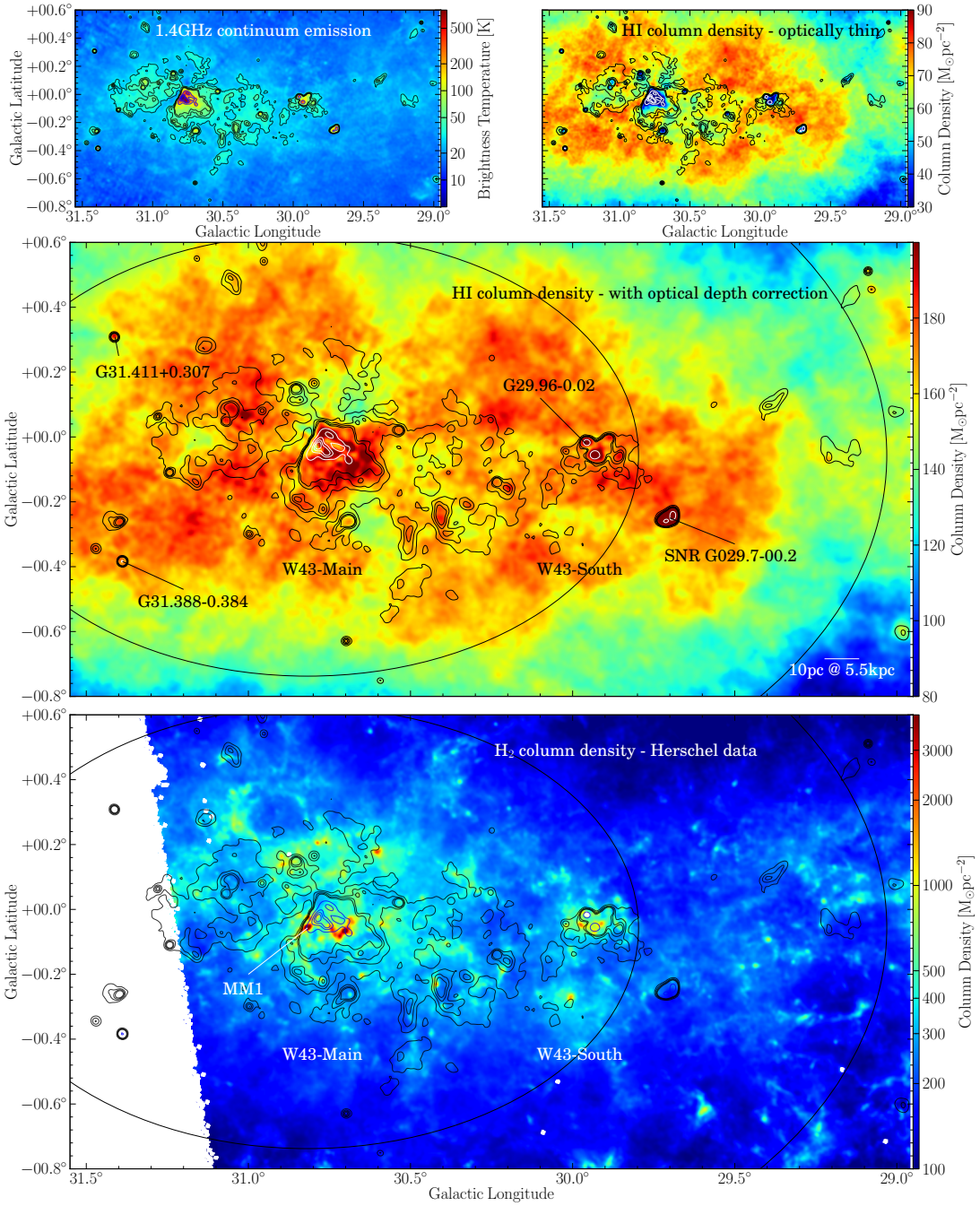

Figure 2.11: The top left panel shows the continuum emission at 21cm in Kelvin. The top right and middle panels presents the H I column density for optically thin assumption and optical depth corrections, respectively. The continuum and H I emission data are based on VGPS, whereas the optical depth correction used in the middle panel is based on the THOR data. The bottom panel shows the H_2 column density from the HiGAL data (Nguyen-Lu'o'ng et al. 2013). In all panels, the black and white/blue contours present the continuum emission at 21cm (black contours show levels of 10, 30, and 70 K; white/blue contours show levels of 200, 400, 600, and 800 K). The black ellipses in the middle and lower panel have an equivalent radius of $r = 80$ and 140 pc. Several important objects are marked.

of the column density, which is shown in the middle panel of Fig. 2.11. We chose the 'complete' velocity range of $v_{\rm lsr} = 60 - 120\,{\rm km\,s^{-1}}$. As we correct for the optical depth, we are able to observe a larger H I column density. We also correct for the weak continuum emission around W43-Main. This correction removes the anti-correlation between the continuum emission and the H I column density. Hence, we determine the H I mass to be $M = 6.1 \times 10^6\,{\rm M_\odot}$ ($l = 29.0 - 31.5°$, $|b| \leq 1°$, $v_{\rm lsr} = 60 - 104\,{\rm km\,s^{-1}}$).

The optical depth spectrum of W43-Main (Fig. 2.10) shows, that the absorption ends abruptly at $100\,{\rm km\,s^{-1}}$, but the H I emission as well as CO emission (Nguyen Luong et al. 2011; Carlhoff et al. 2013) reveals features up to $v_{\rm lsr} = 110\,{\rm km\,s^{-1}}$. The reason for the abrupt drop in the absorption spectrum is not the absence of H I, but the location along the line of sight of the continuum source at this velocity (see Sect. 2.5.3 for further details). As we do not see H I in absorption for velocities larger than $100\,{\rm km\,s^{-1}}$, we are also not able to determine the corresponding optical depth. Therefore we cannot apply our corrections to the H I column density for velocities larger than $100\,{\rm km\,s^{-1}}$. Hence, the velocity range for the previously mentioned H I mass is only $v_{\rm lsr} = 60 - 104\,{\rm km\,s^{-1}}$. Nevertheless we can determine the H I mass for the velocity range from $v_{\rm lsr} = 104 - 120\,{\rm km\,s^{-1}}$ using the optically thin assumption. This reveals a H I mass of $M = 0.5 \times 10^6\,{\rm M_\odot}$. Hence, the total H I mass for W43 in the velocity range $v_{\rm lsr} = 60 - 120\,{\rm km\,s^{-1}}$ is $M = 6.6 \times 10^6\,{\rm M_\odot}$ with the optical depth correction for the velocity range $v_{\rm lsr} = 60 - 104\,{\rm km\,s^{-1}}$ and optically thin assumption for the velocity range $v_{\rm lsr} = 104 - 120\,{\rm km\,s^{-1}}$. This is 2.4 times larger than the H I mass determined with the optically thin assumption.

The limitations and uncertainties of our determined H I column density and H I mass with the optical depth corrections are discussed further in Sect. 2.6.3.

2.5.8 H I self absorption

The H I self absorption (HISA) method uses the diffuse broad H I emission background of the Milky Way as illumination for colder foreground clouds (e.g., Gibson et al. 2000, 2005a,b; Li & Goldsmith 2003; McClure-Griffiths et al. 2006). Dark regions on maps and narrow absorption features in spectra reveal these HISA features. The terminology 'self absorption' can be misleading: the H I emission and absorption can occur in the same cloud or at the same position, but does not have to. The advantage of this method is that it is possible to map entire absorption clouds and study their intrinsic structure. In contrast, the disadvantage is that a sufficient background emission with the same velocity as the absorbing cloud is necessary to detect HISA features. Therefore this method misses a large portion of the cold H I clouds. The differentiation between actual HISA features and the lack of H I emission can also be challenging. Gibson et al. (2005b), however, present an efficient method to detect HISA features

automatically. Another disadvantage is that the optical depth and spin temperature can only be measured together and further assumptions are needed to disentangle these two values.
As described in the previous section, we correct for the optical depth effects and weak continuum emission. This correction does not account for locally confined HISA features, as we assume a uniform optical depth for the entire W43 region. The HISA features could have a higher and spatially varying optical depth that we cannot measure. Furthermore, the weak and diffuse continuum emission around W43 makes the search for HISA features even more complicated. Hence, we refrain from searching and analyzing the possible HISA features around W43. Therefore, we are likely missing some cold neutral hydrogen in our analysis. We will discuss HISA features in more detail in Chapter 4 and present an example of a long filament and study the corresponding HISA feature.

2.6 Discussion

2.6.1 Phases of the neutral atomic hydrogen

It is well known that the neutral atomic ISM has several phases that coexist side by side with very different properties (e.g., Clark 1965; Wolfire et al. 1995; Heiles & Troland 2003b; Wolfire et al. 2003). The main constituents are the cold neutral medium (CNM) and the warm neutral medium (WNM) with spin temperatures on the order of <100K (Strasser et al. 2007) and $\sim 10^4$K (Murray et al. 2014; Roy et al. 2013a), respectively. Furthermore, their density differs by two order of magnitude (CNM: $n_H \sim 50\text{cm}^{-3}$, WNM: $n_H \sim 0.5\text{cm}^{-3}$, e.g., Stahler et al. 2005). Because of the different spin temperature and density of the CNM and WNM, their corresponding optical depths are significantly different with typical values of $\tau_{\text{WNM}} \sim 10^{-3} - 10^{-4}$ (Murray et al. 2014) and $\tau_{\text{CNM}} \gtrsim 0.1$ (Strasser & Taylor 2004). This is important for our interpretation.
Because of the low optical depth of the WNM, we see in absorption merely the CNM. Hence, the optical depth shown in Fig. 2.10 is the optical depth of the CNM. As the absorption spectra are strongly dominated by the CNM, we are not able to measure the optical depth of the WNM and we assume the WNM is optically thin. In contrast, we see a combination of the CNM and WNM for the H I emission. For the correction of the column density (see Sec. 2.5.7), we use the optical depth information from the absorption study to correct the H I emission data. As we do not distinguish between the two phases, we might combine two different quantities, namely the optical depth of the CNM with the emission of the CNM and the WNM. This could lead to an overestimation of the column density. In the following, we assess this effect.
If we assume a CNM cloud with varying T_{spin}(CNM)~20-80K and varying optical depth surrounded by a WNM with T_{spin}(WNM)~7000K (Murray et al. 2014) and optical depth

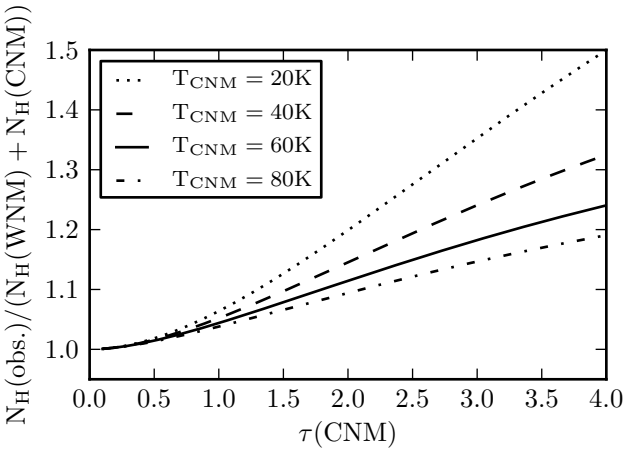

Figure 2.12: The ratio of the observed and actual column density as a function of the CNM optical depth τ(CNM) for different CNM spin temperatures T_{spin}(CNM).

τ(WNM)~5×10^{-3} , we can calculate the column density of each component separately (CNM and WNM). Furthermore, we can calculate the brightness temperature that we would observe and apply our correction method described in Sec. 2.4.4. Finally, we can compare the actual column density of the CNM and WNM with the column density we would observe with our correction method and investigate the overestimation of the actual column density. The result is shown in Fig. 2.12, which shows the ratio of the observed and actual column density as a function of the optical depth of the CNM τ(CNM) for different CNM spin temperatures T_{spin}(CNM). Fig. 2.12 shows that we measure the column density accurately for the optically thin case and we overestimate the column density for larger optical depths. However, even for the extreme case (T_{spin}(CNM)~20K, τ(CNM)~3), we overestimate the column density by at most 1.35. This effect is smaller than the underestimation of the column density due to saturated optical depths and therefore we only consider a single cold component in the following. Similar results in simple simulations for different combination of CNM and WNM fraction and a wide range of N_H and τ were found by Chengalur et al. (2013) and Roy et al. (2013a,b).

2.6.2 H I spin temperature measurements toward point sources

The spin temperature is the best measure of the kinetic temperature of the H I gas along the line of sight. Hence, it is a principal determinant of the physical processes that occur in the H I gas. As described in Sect. 2.4.5, we corrected the H I spin temperature for diffuse weak continuum emission. If we neglect this correction we would systematically overestimate the H I spin temperature and our sample would have a median H I spin temperature of 110 K. That means the

median value would be ∼ 12 K higher compared to our corrected value of 97.5 K (see Fig. 2.8). Similar studies of the spin temperature with similar methods can be found in the literature. For example, Strasser & Taylor (2004) report a median spin temperature of 120 K. This is ∼ 20 K higher than our value of 97.5 K, but they do not consider the diffuse weak continuum emission and therefore probably overestimate the spin temperature.

Heiles & Troland (2003a,b) developed an extensive method to fit Gaussian components to the absorption and emission spectra. Spectral features shown in absorption and emission are assigned to the cold neutral medium (CNM), whereas emission-only features are assigned to the warm neutral medium (WNM). Using this method, they are able to distinguish these two phases and measure their properties, such as as the spin temperature, individually. However, they report that it is difficult to use their method for sources close to the Galactic plane ($|b| < 10$ deg), as multiple components can overlap and the Galactic rotation can broaden their profiles. Figure 2.6 and 2.7 illustrate this problem for our region and for these spectra it is impossible to find unique Gaussian components. As Heiles & Troland (2003a,b) are able to fit individual components, they find on average lower spin temperatures for the CNM in the range of $\sim 40-70$ K. Given that we can not distinguish between the CNM and the WNM, the measured spin temperature is a column density weighted mean between these two components (Strasser & Taylor 2004). This could explain the increased spin temperature reported in Sec. 2.5.2. Another explanation for the increased spin temperature could be the strong radiation field of W43.

Further complications in our sample are the Galactic continuum sources. As these only trace the optical depth up to their location within the Milky Way, we miss the optical depth of the hydrogen behind the continuum source. On the other hand, the off-position measurement traces all the hydrogen along the line of sight. As the kinematic distance is uncertain due to near and far ambiguities, we cannot use the velocity to distinguish the distance of the emitting and absorbing hydrogen. Hence we underestimate the optical depth for the Galactic continuum sources and therefore overestimate the spin temperature. However, we do not see a significant difference in the mean spin temperature for Galactic and extragalactic continuum sources and we can neglect this effect.

2.6.3 Column density and mass estimate

In Sect. 2.5.7 we explained that we correct the H I column density for optical depth effects as well as for the diffuse continuum emission. This leads to a more accurate estimate of the column density and a more accurate mass estimate. Nevertheless these corrections have limitations which we discuss in the following section.

As we need a strong continuum source in the background to determine the optical depth, we

can measure the optical depth only toward certain locations. For our H I column density correction, we used the strong continuum source W43-Main. Furthermore we assumed that the optical depth is the same for the entire cloud. This assumption might not hold, especially for the outer parts of the cloud. To investigate this effect, we compared the optical depth measurements for other sources at other positions. However, in the range of $v_{\rm lsr} = 80 - 110\,{\rm km\,s^{-1}}$, we find that the optical depth is mostly saturated and determined by the corresponding lower limit. Two examples are given in the third panel of Fig. 2.6 and 2.7, which reveal a lower limit of $\tau_{\rm lower-limit} = 2.9$ and $\tau_{\rm lower-limit} = 1.9$, respectively. Other continuum sources that have a larger separation from the Galactic mid-plane show similar results. For example, the continuum source G30.699−0.630 has a Galactic latitude of $b \approx 0.6°$ and still shows a saturated optical depth for the velocity range of W43 with a lower limit of $\tau_{\rm lower-limit} = 1.5$. If we use this continuum source to correct for optical depth effects and the weak diffuse continuum emission, we determine a mass of $M = 4.8 \times 10^6\,{\rm M_\odot}$ for the whole cloud ($l = 29.0 - 31.5°$, $|b| \leq 1°$, $v_{\rm lsr} = 60 - 120\,{\rm km\,s^{-1}}$). Hence, by using W43-Main to correct the optical depth for the entire cloud, we do not overestimate the mass in the outer parts significantly, but at most by a factor of 1.4. We also use this mass estimate as a lower limit for the mass of W43.

The saturation of the optical depth is the second limitation we have to consider, especially for the inner part of the cloud. In this region, we underestimate the opacity and, therefore, the column density which would lead to a further underestimation of the mass. Estimating this effect is difficult and, therefore, we report only lower limits for the column density and the mass in the inner part of W43.

The third limitation is the distance of the continuum source that we use to determine the optical depth. As explained in Sect. 2.5.3, the continuum source W43-Main is located close to the tangential point of the Scutum-Centaurus arm at a distance of 5.5 kpc. Hence, we only see H I absorption as far as this distance, and miss all the H I that is located behind the continuum source but still within the cloud. If we assume that the continuum source is at the center of the cloud, we underestimate the column density by a factor of two. Another approach to overcome this limitation is to look at other more distant continuum sources. The continuum source G31.388−0.384 is extragalactic and has a comparable brightness temperature to W43-Main. Hence, it is an ideal candidate for this test. We used the optical depth shown in the third panel of Fig. 2.6 to correct the H I column density and measured the corrected mass for the same area and velocity range, as in Sect. 2.5.7. The absoprtion spectrum of W43-Main and G31.388−0.384 are similar, except for the velocity range $v_{\rm lsr} = 100 - 120\,{\rm km\,s^{-1}}$, which is missing in the W43-Main spectrum. We can also use the optical depth measurements of G31.388−0.384 to correct the H I emission. However, as these absorption spectra are similar, the corrected masses are the same within the uncertainties ($M = 6.9 \times 10^6\,{\rm M_\odot}$ for the correc-

tion with the optical depth of G31.388−0.384).

Summing up, we might overestimate the mass in the outer part of W43, but underestimate the mass in the inner part of W43. Because of the position of the continuum source, we might underestimate the mass by a factor of two. As explained in Sect. 2.5.8, we also miss some H I due to self-absorption. Hence we report the H I mass of W43 to be a lower limit of $M = 6.6_{-1.8} \times 10^6\,\mathrm{M}_\odot$.

Several H I mass estimates are given in the literature (Nguyen Luong et al. 2011; Motte et al. 2014). All of them are calculated with the assumption of optically thin emission.

Motte et al. (2014) measured the H I mass in the inner part of W43 ($l = 29°.6 \sim 31°.4$ and $b = -0°.5 \sim 0°.3$) and for the velocity range ($v_{\mathrm{lsr}} = 60 - 120\,\mathrm{km\,s^{-1}}$). The assumption of optically thin emission reveals an H I mass of $M = 0.9 \times 10^6\,\mathrm{M}_\odot$. If we use our corrected H I column density to measure the mass in this area, we determine a H I mass of $M = 2.2 \times 10^6\,\mathrm{M}_\odot$ (this value is smaller than the value given in Sect. 2.5.7 as we only consider the inner part of W43). As outlined above, we claim that this mass estimate is a lower limit and, therefore, the mass determined with the optically thin emission is at least a factor of 2.4 too small. This has implications for theoretical models, which we discuss in Sect. 2.6.6.

2.6.4 Spatial distribution of hydrogen

Beside the total H I mass, knowing the spatial distribution of the H I is crucial to understand the formation of W43. As we have corrected the column density map for optical depth effects and the weak diffuse continuum emission, we can use this corrected column density map (see Fig. 2.11) to investigate the H I spatial distribution in more detail, especially in the center.

Similar work was done by Motte et al. (2014). They measured the H I column density in rectangular annuli around the center of W43 ($l = 30.5°$, $b = 0°$) with an aspect ratio of 3:2 and find an increasing H I column density inward from $N_{\mathrm{H\,I}} \sim 40\,\mathrm{M}_\odot\,\mathrm{pc}^{-2}$ at a distance of 170 pc to $N_{\mathrm{H\,I}} \sim 80\,\mathrm{M}_\odot\,\mathrm{pc}^{-2}$ at a distance of 50 pc from the center (velocity range $v_{\mathrm{lsr}} = 60{-}120\,\mathrm{km\,s^{-1}}$). Since they assume optically thin H I emission, they underestimate the H I column density, especially in the central region (as discussed in Sect. 2.6.3).

To compensate for the approximate elliptical structure of W43, we choose elliptical annuli with an aspect ratio of 3:2 for major and minor axis, which fits the large scale structure of W43 well (see Fig. 2.11). As we focus on W43-Main, we choose the most massive submillimeter peak W43-MM1 ($l = 30.8175°$ and $b = -0.0571°$, Motte et al. 2003) as the center for the ellipses. The width of each annulus is 10 pc for the major axis and 6.6 pc for the minor axis. For each elliptical annulus, we calculated the equivalent radius $r = \sqrt{major \cdot minor}$ and assigned these values as the distance to the center shown in Fig. 2.13.

The black squares in Fig. 2.13 represent the H I column density mean value and the corres-

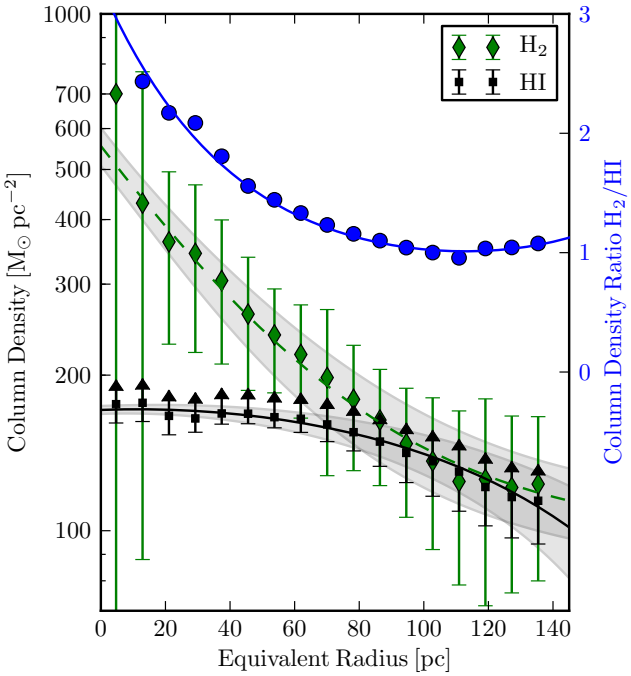

Figure 2.13: The column density of H I and H_2 measured in elliptical annuli around W43 MM1. The x-axis presents the equivalent radius of these elliptical annuli. The black and green lines represent the fitted curves with the corresponding 1σ uncertainties shown as a gray shaded area. The larger symbols (diamonds for H_2, squares for H I) present the averaged value of the elliptical annuli and their corresponding variations within the annuli. As the optical depth spectra saturates, we can only determine lower limits for the H I column density. The blue dots and line show the H_2 and H I column density ratio of the data and the fitted curves, respectively.

ponding standard deviation of each elliptical annulus. We confirm the result of Motte et al. (2014), which indicates that the H I column density rises inward. Our corrections allow us for the first time to study the central region of W43 ($r < 50$ pc) and, within the uncertainties, we report a flat column density distribution in this inner region. However, we mention that this flatness could also be due to the underestimation of the column density in the central part, as the optical depth saturates and therefore the column density is a lower limit.

Furthermore, the diamond symbols in Fig. 2.13 present the column density of the molecular hydrogen. The H_2 distribution is centrally concentrated and the column density rises steeply toward the center, which is different from the H I profile. The large uncertainties in the center are due to the clumpiness of the molecular hydrogen as the most prominent clumps, such as W43-MM1, are located in the first two elliptical annuli. Beside W43-Main, the second most prominent molecular clump is W43-South (see Fig. 2.11). However, in this analysis we focus on W43-Main and, therefore, we choose W43-MM1 as the center for the ellipses. Further-

more, we masked W43-South for this analysis, as it would introduce a large uncertainty at an equivalent radius of $r \sim 50\,\mathrm{pc}$.

We use a nonlinear least square method ('curvefit' in the scipy package) to fit the mean values of the thin elliptical annuli with their corresponding uncertainties. For the H I column density we assume a quadratic radial distribution, i.e.,

$$N_{\mathrm{H\,I}}(r) = a \cdot r^2 + b \cdot r + c, \tag{2.11}$$

where $N_{\mathrm{H\,I}}(r)$ describes the H I column density and r describes the equivalent radius. The free parameters a, b and c have the fitted values of $a = -3.9 \pm 0.7 \times 10^{-3}\,\mathrm{M_\odot\,pc^{-4}}$, $b = 0.089 \pm 0.099\,\mathrm{M_\odot\,pc^{-3}}$ and $c = 171 \pm 3\,\mathrm{M_\odot\,pc^{-2}}$. For the H_2 distribution we assume an exponential function to fit the data as follows:

$$N_{\mathrm{H_2}}(r) = d \cdot \exp(-e \cdot r) + f, \tag{2.12}$$

where $N_{\mathrm{H_2}}(r)$ describes the H_2 column density and r describes the equivalent radius. The free parameters d, e, and f have the fitted values of $d = 458 \pm 44\,\mathrm{M_\odot\,pc^{-2}}$, $e = 0.022 \pm 0.004\,\mathrm{pc^{-1}}$ and $f = 97 \pm 15\,\mathrm{M_\odot\,pc^{-2}}$.

The black and green lines in Fig. 2.13 present the column density fits and the gray shaded areas represent their corresponding uncertainties. Both, the H I and H_2 distributions are well fit by the assumed functions. As the uncertainties for the innermost part of the H_2 distribution ($r < 20\,\mathrm{pc}$) are large, the fitted curve deviates from the data points and this area has to be treated cautiously. We also tested different functions to fit the data, such as functions with only two free parameters, or a Gaussian distribution for the H I distribution, but the results were similar within the uncertainties. Finally, we chose the aforementioned functions as they could reproduce well the ratio of the H_2 and H I column density. This ratio is shown in blue in Fig. 2.13. The circles represent the ratio of the data points, whereas the solid line represents the ratio of the fitted curves. The ratio stays fairly constant at $N_{\mathrm{H_2}}/N_{\mathrm{H\,I}} \approx 1$ for $140\,\mathrm{pc} > r > 60\,\mathrm{pc}$. For this region, the H I column density also rises to its maximum value of $N_{\mathrm{H\,I}} \approx 170\,\mathrm{M_\odot\,pc^{-2}}$. Further inward ($r < 60\,\mathrm{pc}$), the H I column density stays constant at this maximum value, whereas the H_2 column density rises sharply. Hence, the H_2/H I ratio also rises sharply to values above three. Summing up, the column density measurements imply that we have a mixture of H I and H_2 in the outskirt of the cloud and a molecular dominated region in the center. In the following, we investigate this structure for the particle density.

As mentioned before, we find a flat column density for the H I distribution toward the center of W43-Main, but what does this imply for the actual density in the center? If we assume that W43 has an elliptical shape, we can decompose the cloud into different layers, similar to an onion. Furthermore, we assign the appropriate column density to each layer with the information given

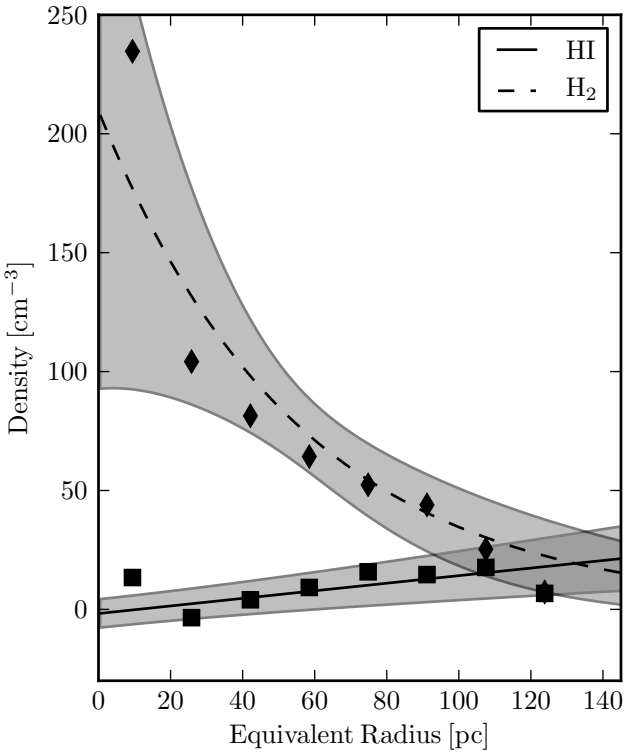

Figure 2.14: The density of H I and H_2 as a function of the distance toward the center of W43 (equivalent radius). The diamonds and the dashed line represent the molecular hydrogen, whereas squares and the solid line represent the atomic hydrogen. The corresponding 3σ uncertainties from the fit introduced in equation 2.14 are shown as gray shaded areas.

in Fig. 2.13. As the column density is additive, the appropriate column density of each layer is the measured column density at the position of the considered layer minus the column density of all layers further outside. Hence, a flat H I column density distribution toward the center (see Fig. 2.13) means that the layers in the center have no or a very small column density and therefore also a very small particle density. In the following section, we use the concept of a elliptical layered structure to determine the actual particle density and show that indeed the H I particle density drops toward the center of W43 within this model.

2.6.5 Linking column density to particle density

While the column density neglects the third dimension, it does not necessarily reflect the actual particle density. Hence, we have to estimate the third dimension. Using the elliptical, layered structure for W43 previously described and the results presented in Fig. 2.13, we estimate the

particle density in each layer, i.e.,

$$n_{\mathrm{i}} = \frac{(N_{\mathrm{i}} - N_{\mathrm{i+1}})}{2\ (r_{\mathrm{i+1}} - r_{\mathrm{i}})}, \tag{2.13}$$

where n_{i} describes the particle density of layer i; N_{i} and $N_{\mathrm{i+1}}$ describes the column density of layer i and layer i+1, respectively; and r_{i} and $r_{\mathrm{i+1}}$ describes the equivalent radius of layer i and i+1. The factor of two accounts for the two layers of the elliptical annuli in the front and back of the cloud. The result for this calculation is shown in Fig. 2.14 as squares and diamond data points for the H I and H_2 density, respectively. To increase the signal to noise ratio, we used larger elliptical annuli with a major axis of 20 pc and a minor axis of 13.3 pc.

In addition, we can use the fitted curves of the column density to estimate the density as well, by inserting equation 2.11 and 2.12 into equation 2.13. Hence, the radial distribution of the H I and H_2 densities are

$$\begin{aligned} n_{\mathrm{H\,I}} &= -0.5\,(a\ (r_{\mathrm{i}} + r_{\mathrm{i+1}}) - b), \\ n_{\mathrm{H_2}} &= \frac{d}{2\ (r_{\mathrm{i+1}} - r_{\mathrm{i}})}\ \left(\exp(-e \cdot r_{\mathrm{i}}) - \exp(-e \cdot r_{\mathrm{i+1}})\right), \end{aligned} \tag{2.14}$$

where $n_{\mathrm{H\,I}}$ and $n_{\mathrm{H_2}}$ describes the particle density of H I and H_2, respectively; r_{i} describes the equivalent radius of layer i; and a, b, c, and d are the free fit parameters introduced in equations 2.11 and 2.12. In Fig. 2.14, the calculated density distributions are shown as black lines with the corresponding uncertainties as a gray shaded areas. We note that the density distributions of H I and H_2 are very distinct. While the H I distribution follows a simple linear relation, the H_2 distribution shows an exponential increase toward the center. The particle density shows a mixture of the atomic and molecular hydrogen in the outskirts of W43 ($r > 100$ pc) and a molecular dominated interior similar to the measurements of the column density. However, the particle density of the atomic hydrogen drops to almost zero toward the center, which results in the observed constant column density.

Another way to present our results is shown in Fig. 2.15. This figure shows the particle density of H I and H_2 introduced in Eq. 2.14 as a function of the total column density ($N_{\mathrm{H\,I}} + N_{\mathrm{H_2}}$). High column density regions, i.e., the central region of W43, are dominated by molecular hydrogen. On the other hand, we find an equivalent mixture of atomic and molecular hydrogen for low column density regions, which represent the envelope of W43.

In addition, we study the correlation between the H I and H_2 density. To do so, we replace the equivalent radius in Fig. 2.14 and plot the H_2 density as a function of the H I density. Figure 2.16 shows the corresponding plot with the uncertainties for the H_2 density. The statistical and systematical uncertainties for this plot are relatively large, but nevertheless we see an anti-correlation between the H I and H_2 density. This can be explained with a simple model that H_2 forms out of H I in the innermost part of W43. However, we do not detect a sharp transition

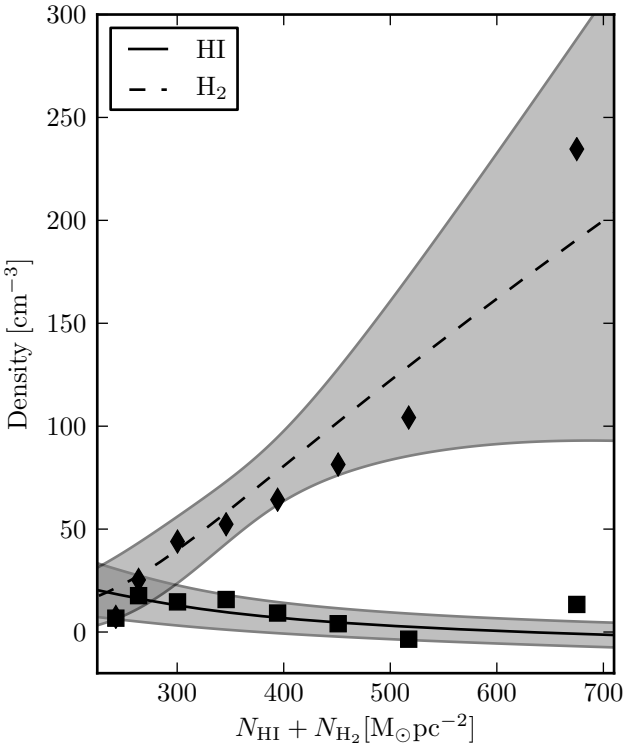

Figure 2.15: The particle density of H I and H_2 as a function of the total column density ($N_{H\,I}$ + N_{H_2}). The diamonds and the dashed line represents the molecular hydrogen, whereas squares the solid line represents the atomic hydrogen. The corresponding 3σ uncertainties from the fit introduced in equations 2.11 and 2.12 are shown as gray shaded areas.

between H_2 and H I predicted by Krumholz et al. (2008, 2009). In the following section, we discuss possible implications.

2.6.6 Column density threshold for atomic hydrogen

A fundamental question of molecular cloud formation is how does molecular hydrogen form out of neutral atomic hydrogen and what are the corresponding conditions (Dobbs et al. 2014). It is well known that the density must be high enough to shield the molecular hydrogen from the interstellar radiation field, to avoid dissociation back to its atomic form (Hollenbach & Tielens 1997; McKee & Ostriker 2007; Krumholz et al. 2008, 2009; Mac Low & Glover 2012).
To describe this transition, Krumholz et al. (2008, 2009) suggest an analytic model, which is based on the assumption of a spherical cloud that is embedded in an isotropic external radiation field. They approximate that the transition between H I in the envelope and H_2 in the center occurs in an infinitely thin shell. An important result is that an H I column density of $\sim$ 10 M$_\odot$ pc^{-2} is necessary to shield the molecular hydrogen from the interstellar radiation

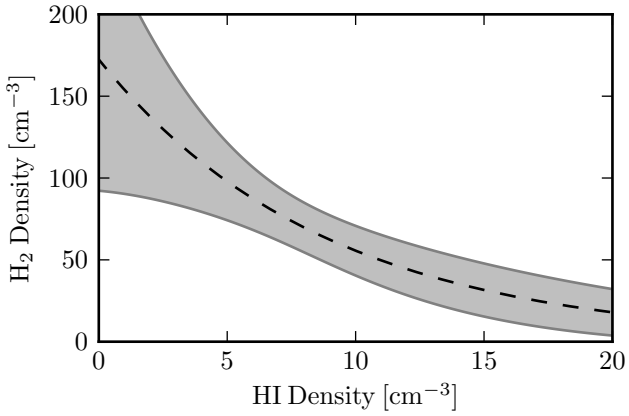

Figure 2.16: The particle density of molecular hydrogen as a function of the atomic hydrogen. The black curve represents the best-fit curves from Fig. 2.14 and the gray shaded area shows the 3σ uncertainty for the H_2 fit.

field. If the cloud reaches this critical H I column density the formation of molecular hydrogen is efficient enough that most of the hydrogen goes into molecular form and the column density of H I remains relatively constant at this level. Therefore we should not expect to observe H I column densities larger than $\sim 10\,M_\odot\,pc^{-2}$, contradicting our results presented in Fig. 2.13. Furthermore, they show that this H I column density threshold is independent of the external radiation field, but has a weak dependence on the metallicity of the gas.
This model has three free parameters. First, the ratio of the measured CNM density n_{CNM} to the minimal CNM density n_{min} , i.e.,

$$n_{CNM} = \phi_{CNM} \cdot n_{min}. \tag{2.15}$$

The minimal CNM density is determined by the pressure balance with the warm neutral medium (WNM) and has a typical value of $n_{min} \approx 7 - 8 cm^{-3}$. As the range of pressure balance between the CNM and the WNM is limited, the maximum CNM density can be at most $\sim 10 \cdot n_{min}$. Hence, ϕ_{CNM} can vary between 1 and 10. Krumholz et al. (2009) assume $\phi_{CNM} \approx 3$ for their fiducial value. The second free parameter is the ratio of the CNM density n_{CNM} to the molecular density n_{mol} , i.e.,

$$n_{mol} = \phi_{mol} \cdot n_{CNM}. \tag{2.16}$$

Krumholz et al. (2009) suggests that this ratio is $\phi_{mol} \approx 10$ and should not vary significantly. The last free parameter is the metallicity Z. This model is supported by extragalactic observations (Krumholz et al. 2009) as well as observations in the Perseus molecular cloud by Lee et al. (2012). They find a uniform H I column density of $N_{H\,I} \sim 6 - 8\,M_\odot\,pc^{-2}$ for H_2 column densities up to $N_{H_2} \sim 80\,M_\odot\,pc^{-2}$. To fit their data to the model of Krumholz et al. (2009),

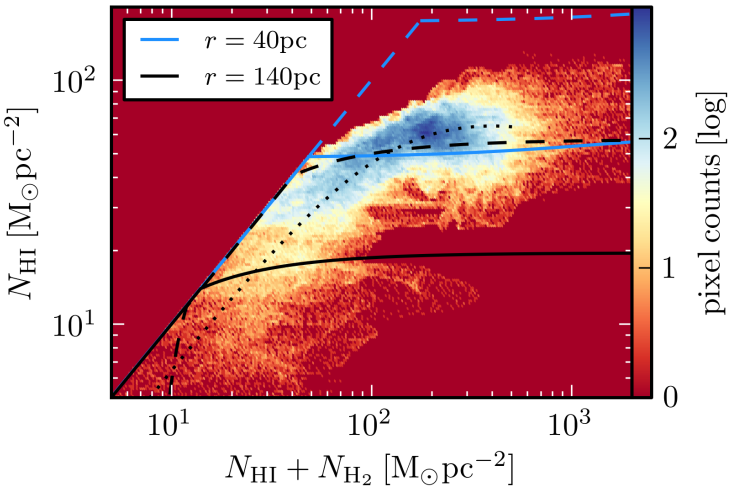

Figure 2.17: The column density of the atomic hydrogen ($v_{\rm lsr}$ = 60 − 120 km s^{-1}) as a function of the total hydrogen column density ($N_{\rm H\,I}$ + $N_{\rm H_2}$). The column density for the envelope ($r > 140pc$, $N_{\rm H\,I}$ = 107 ± 20 $M_\odot$ pc^{-2}, $N_{\rm H_2}$ = 116 ± 18 $M_\odot$ pc^{-2}) is subtracted. The solid and dashed lines describe the theoretical model of Krumholz et al. (2008, 2009) for solar values (G_0 = 1, Z = 1) and for more realistic values of W43 (G_0 = 5, Z = 1.4), respectively. The blue and black colors represent the model parameters at different equivalent radii of r = 40 and 140 pc, respectively. The black dotted line shows the fitted curve of the elliptical annuli analysis introduced in Sect. 2.6.3.

they fixed $\phi_{\rm mol}$ = 10 and Z = 1.0 $Z_\odot$ and fitted $\phi_{\rm CNM}$, which reveals values of $\phi_{\rm CNM} \approx 6 - 10$. The complex W43 is a more extreme test case for this theory as we have a much larger mass reservoir as well as an OB cluster that acts as a strong radiation source in the center. For this analysis, we extracted the column density of H I and H_2 for each pixel. As a basis for the H I column density, we chose the optical depth corrected version with a velocity range of $v_{\rm lsr}$ = 60 − 120 km s^{-1}. For the H_2 column density, we again use the HiGal data (see Sect. 2.3.4). Even though the data base is the same as that used in Sect. 2.6.5, we stress that the method is different. Here we conduct a pixel by pixel comparison, whereas in Sect. 2.6.5 we averaged the column density over elliptical annuli. In addition, we focus merely on the inner part of W43 with r < 140 pc and, therefore, we have to subtract the column density of the envelope. We assess the column density at r = 140 pc as the envelope column density and subtract this value. The column density for the H I and H_2 envelopes are $N_{\rm H\,I}$ = 107±20 $M_\odot$ pc^{-2} = 1.3 ± 0.3 × 10^{22} cm^{-2} and $N_{\rm H_2}$ = 116 ± 18 $M_\odot$ pc^{-2} = 7.3 ± 1.1 × 10^{21} cm^{-2}, respectively.

Figure 2.17 shows the result of the pixel by pixel comparison of the H I and H_2 column density. Since we subtracted the column density of the envelope for this analysis, the values are smaller than in Fig. 2.13. For better readability, we do not show each single pixel comparison, but we performed a pixel binning. We do not observe the predicted threshold for the H I column dens-

ity of $\sim 10\,M_\odot\,pc^{-2}$, instead our data show high H I column density values, peaking between $N_{H\,I} = 50 - 80\,M_\odot\,pc^{-2}$. Nevertheless we will try to fit the analytic model of Krumholz et al. (2009) to our data.

As described in Sec. 2.6.5, we use elliptical annuli to estimate the particle density. We can use this information (Fig. 2.14) to determine the model parameters ϕ_{CNM} and ϕ_{mol} using equations 2.15 and 2.16, respectively. Given that the H I and H_2 densities vary as a function of equivalent radius, the model parameters ϕ_{CNM} and ϕ_{mol} also vary. We calculate the model parameters for three different distances $r = 40$, 80, and 140 pc, which characterize the molecular dominated interior, the transition region, and the well-mixed outer area. To calculate ϕ_{CNM} , we have to know n_{min} , which is given by Krumholz et al. (2009) as

$$n_{min} \approx 31 \frac{G_0}{1 + 3.1\,Z^{0.365}} cm^{-3}, \tag{2.17}$$

where G_0 and Z are the far-ultraviolet radiation intensity (in units of the Habing field) and the total metallicity, both normalized to their values in the solar neighborhood. For W43, we have a slightly higher metallicity of $Z = 1.4 Z_\odot$ (Motte et al. 2014) and a large radiation field, which could be up to $> 100\,G_0$ (Beuther et al. 2014) in the central region. We assume a more moderate radiation field for the outer regions of W43 with values around $5 - 10\,G_0$. In Table 2.2, we present the H I and H_2 density as well as the model parameters for different distances. The parameter ϕ_{CNM} is calculated for solar values ($G_0 = 1$, $Z = 1$) and for more realistic values of W43 ($G_0 = 5$, $Z = 1.4$) that results in $n_{min} = 7.5 cm^{-3}$ and $n_{min} = 34.4 cm^{-3}$, respectively. The solid blue and black lines in Fig. 2.17 show the theoretical column density for solar values and for the equivalent radius of $r = 40$ and 140 pc, respectively, whereas the dashed lines represent the more realistic values of W43 ($G_0 = 5$, $Z = 1.4$). The dashed black line, which represents the outer area of W43 ($r = 140$ pc) with a moderate radiation field ($G_0 = 5$, $Z = 1.4$) might fit the data. However, the model parameters might have unrealistically low values of $\phi_{CNM} = 0.58$ and $\phi_{mol} = 0.85$. Using a stronger radiation field increases n_{min} and therefore decreases ϕ_{CNM} to even lower values. Hence, values with smaller distances and/or higher radiation fields predict column densities that are too high.

We do not have a conclusive answer as to why the model suggested by Krumholz et al. (2008, 2009) does not describe W43, but we suggest that the strong UV radiation produced by the central OB cluster (Blum et al. 1999) and several further clusters in the environment are responsible for the dissociation of the molecular hydrogen. Another explanation was given by Motte et al. (2014), They performed a similar analysis and their conclusion was that the analytical model by Krumholz et al. (2008, 2009) cannot describe a complicated molecular cloud complex, as we see several transition layers between H I and H_2 along a single line of sight and the assumption of a simple spherical cloud, without internal radiation sources breaks

Table 2.2: Densities and model parameter.

r [pc]	n_{HI} [cm^{-3}]	n_{H_2} [cm^{-3}]	ϕ_{mol}	ϕ_{CNM} ($G_0 = 1$)	ϕ_{CNM} ($G_0 = 5$)
140	20±4	17±5	0.85	2.7	0.58
80	11±3	49±5	4.4	1.5	0.32
40	5±2	101±8	20.2	0.7	0.14

H I and H_2 particle density for different equivalent radii with corresponding 1σ uncertainty, extracted using the information given in Fig. 2.14 and using equation 2.14. The given uncertainties are the statistical uncertainties, but do not take the systematical uncertainties due to the saturation of the optical depth spectra into account. The model parameters ϕ_{mol} and ϕ_{CNM} are calculated using Eq. 2.16 and 2.15, respectively. The parameter ϕ_{CNM} is calculated for solar values ($G_0 = 1$, $Z = 1$) and for more realistic values of W43 ($G_0 = 5$, $Z = 1.4$).

down.

2.7 Conclusions

The H I, OH, Recombination Line Survey of the Milky Way (THOR) is a Galactic plane survey covering a large portion of the Galactic disk ($l = 15 - 67°$, $|b| \leq 1°$). We use the VLA to observe the 21cm H I line, 4 OH lines, 19 Hα recombination lines and the continuum from 1-2 GHz. We present the H I data of the pilot field centered on the GMC associated with W43 ($l = 29.2 - 31.5°$, $|b| \leq 1°$). The main results can be summarized as:

1. We measured the average spin temperature of the neutral hydrogen gas along the line of sight toward eight strong continuum sources. Half of them are Galactic and half of them are extragalactic. We find a median value of $T_S = 95.7$ K, which is in agreement with other studies.
2. We can estimate the optical depth for the H I line toward strong continuum sources at various locations in W43. The measured optical depth saturates at the main velocity component of W43 ($v_{lsr} = 80 - 100\,km\,s^{-1}$) with lower limits of $\tau \sim 2.7$ in the center. Hence, the derived H I masses based on optically thin emission strongly underestimates the hydrogen content. Employing further corrections for the weak and diffuse continuum emission, we obtain a lower limit for the H I mass of $M = 6.6_{-1.8} \times 10^6\,M_\odot$ for a velocity range of $v_{lsr} = 60 - 120\,km\,s^{-1}$ and an area of $l = 29.0 - 31.5°$ and $|b| \leq 1°$. This is a factor of ~2.4 larger than the H I mass estimates with the assumption of optically thin emission.
3. The measured H I column density exceeds values of $N_{H\,I} \sim 150\,M_\odot\,pc^{-2}$ over much of the inner region with $r < 80$ pc. This is an order of magnitude larger than for low-mass star-forming regions such as Perseus.

4. As we corrected the H I column density for optical depth effects and the weak continuum emission, we are able to study the H I distribution spatially even in the innermost part of W43. We assumed an elliptical layered structure for the GMC associated with W43. This allows us to reconstruct the particle density of H I and we find a linearly decreasing H I density toward the center of W43 with values from $n_{\mathrm{H\,I}} = 20$ to almost $0\,\mathrm{cm}^{-3}$. Furthermore, we compared our results to the density of molecular hydrogen based on Herschel continuum data. The density of molecular hydrogen shows an exponential increase toward the center of W43 with values rising from $n_{\mathrm{H_2}} = 15$ to $200\,\mathrm{cm}^{-3}$. For smaller clumps, the density of H_2 can rise to even higher values.
5. We compared our measurements to the analytic model suggested by Krumholz et al. (2008, 2009). Our data does not show a sharp transition between H I and H_2, and nor do we find the predicted threshold for the H I column density of $\sim 10\,\mathrm{M}_\odot\,\mathrm{pc}^{-2}$. Based on these models, larger H I column densities should not exist, as molecular hydrogen should form for such high H I column densities. To fit the model, we have to assume low model parameters, which may indicate that the model is not applicable in a region with such a high radiation field. We suggest that the addition of an internal radiation field from a central cluster may be required to describe the observations. Thus, this work has shown that the H I content of W43 and its relation to H_2 challenges current models of H_2 formation.

3

Continuum sources from the THOR survey between 1 and 2 GHz

This chapter is published in A&A (Bihr et al. 2016) and slightly adapted for this thesis.

3.1 Abstract

In this chapter we present a catalog of the continuum sources in the first half of the survey ($l = 14.0-37.9°$ and $l = 47.1-51.2°$, $|b| \leq 1.1°$) at a spatial resolution of $10-25''$, depending on the frequency and sky position with a spatially varying noise level of $\sim 0.3 - 1$ mJy beam^{-1}. The catalog contains ~4400 sources. Around 1200 of these are spatially resolved, and ~1000 are possible artifacts, given their low signal-to-noise ratios. Since the spatial distribution of the unresolved objects is evenly distributed and not confined to the Galactic plane, most of them are extragalactic. Thanks to the broad bandwidth of the observations from 1 to 2 GHz, we are able to determine a reliable spectral index for ~1800 sources. The spectral index distribution reveals a double-peaked profile with maxima at spectral indices of $\alpha \approx -1$ and $\alpha \approx 0$, corresponding to steep declining and flat spectra, respectively. This allows us to distinguish between thermal and non-thermal emission, which can be used to determine the nature of each source. We examine the spectral index of ~300 known H II regions, for which we find thermal emission with spectral indices around $\alpha \approx 0$. In contrast, supernova remnants (SNR) show non-thermal emission with $\alpha \approx -0.5$ and extragalactic objects generally have a steeper spectral index of $\alpha \approx -1$. Using the spectral index information of the THOR survey, we investigate potential SNR candidates. We classify the radiation of four SNR candidates as non-thermal, and for the first time, we provide strong evidence for the SNR origin of these candidates.

3.2 Introduction

At present, high resolution (<20") Galactic plane surveys are available for studying different questions concerning star formation and the interstellar medium (ISM). Starting in 2012 with a pilot study around the giant molecular cloud (GMC) associated with the W43 star formation complex (Bihr et al. 2015; Walsh et al. 2016), we observed a large fraction of the Galactic plane in the first quadrant of the Milky Way ($l = 14 - 65°$, $|b| \leq 1.1°$) in consecutive semesters. In this paper, we present the results of the continuum observations of the first half of the survey ($l = 14.0 - 37.9°$ and $l = 47.1 - 51.2°$, $|b| \leq 1.1°$), covering ~56 deg^2, including a catalog of the extracted sources. The full survey will be presented in a forthcoming paper by Beuther et al. (in prep.).

The discrete continuum sources between 1 and 2 GHz are dominated by two distinct emission classifications: thermal and non-thermal emission (Wilson et al. 2010). The thermal emission is mostly due to free-free emission from electrons, whereas the non-thermal emission is due to the synchrotron emission of relativistic electrons in magnetic fields. These different emission mechanisms can be distinguished by the spectral index α, which is defined as $I(\nu) \propto \nu^{\alpha}$, where $I(\nu)$ is the frequency dependent intensity. The thermal free-free emission shows a flat or positive spectral index, depending on the optical depth. The values can vary between 2 and -0.1 for the optically thick and thin regime, respectively (e.g., Mezger & Henderson 1967; Keto 2003; Wilson et al. 2010). In contrast to this, synchrotron emission shows a negative spectral index depending on the particle energy distribution. One usually finds spectral indices below -0.5 (e.g., Rybicki & Lightman 1979; Meisenheimer 1999). Supernova remnants (SNR) show a spatially varying spectral index around $\alpha = -0.5$ (e.g., Bhatnagar et al. 2011; Green 2014; Reynoso & Walsh 2015; Dubner & Giacani 2015). The broad bandpass of our VLA observations allows us to determine the spectral index for bright sources and therefore distinguish between the two radiation mechanisms. However, knowing the kind of radiation does not directly disclose the source type. Thermal free-free emission can emerge from H II regions or planetary nebulae. Non-thermal synchrotron radiation can be produced by extragalactic jets powered by an active galactic nucleus (AGN) or from Galactic SNR. Thermal radiation from extragalactic sources is possible, but might be too weak to be detected in our observations. As a result, thermal emission is most likely of Galactic origin, and the non-thermally emitting sources could be extragalactic AGN or Galactic SNR. The ability to characterize continuum sources and thus distinguish between Galactic and extragalactic emission is crucial for prospective THOR H I and OH absorption studies.

3.3 Observations and data reduction

3.3.1 VLA observations

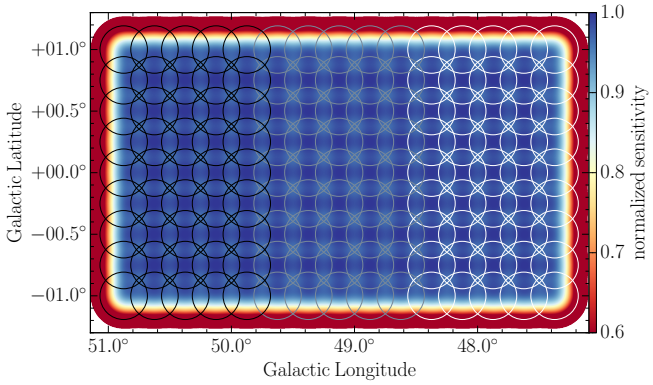

Figure 3.1: Normalized sensitivity pattern in color of the observed mosaic for the spectral window with the highest frequency at 1.95 GHz. The sensitivity drops toward the edge, and the variations within the mosaic are smaller than 4%. The circles represent the primary beam at this frequency, which is ∼ 23′. The different colors of the circles represent three different observing blocks.

We used the VLA in New Mexico in C configuration to map the continuum in the L band from 1 to 2 GHz simultaneously with the H I 21 cm line, 4 OH lines, and 19 Hα recombination lines. For the VLA in C-configuration, the baselines range from 35 to 3400 m. The corresponding primary beam changes with frequency from ∼45′ at 1 GHz to ∼23′ at 2 GHz and therefore the actual size of the mosaics changes as well. The data presented in this chapter were observed in two campaigns. The first campaign was the THOR pilot observations (l = 29.2 − 31.5°, $|b| \leq 1.1°$) during the 2012A semester (Project 12A-161, see also Bihr et al. 2015). We used a hexagonal geometry for the mosaic for this 2° × 2° field at 17.9'spacing, which results in 59 pointings. Each pointing was observed 4 × 2 min, which results in an overall integration time of ten hours for the pilot field, including around two hours overhead for flux, bandpass, and complex gain calibration. The second campaign covered a large section of the first quadrant of the Milky Way (l = 14.0 − 29.2° and l = 31.5 − 37.9° and l = 47.1 − 51.2°, $|b| \leq 1.1°$) and was observed during the 2013A semester (Project 13A-120). In contrast to the pilot field, we used a rectangular grid for the mosaic (see Fig. 3.1) with a smaller spacing of 15'. The close spacing meant that the sensitivity variations are at most 4% for the spectral window around 1.95 GHz and less for smaller frequencies. The second campaign was split into 20 observing blocks, each covering a field of Δl = 1.25° and $|b| \leq 1.1°$ with 45 pointings each. Each pointing was observed 3 × ∼2 min, which results in a total integration time of five hours for each observing block, including ∼50 min overhead for flux, bandpass, and complex gain calibration. We chose

Table 3.1: Summary of spectral windows

Frequency [MHz]	Lowest resolution	Highest resolution
989 – 1117	$24.4'' \times 15.1''$	$16.5'' \times 15.7''$
1244 – 1372	$19.7'' \times 12.5''$	$13.1'' \times 12.3''$
1372 – 1500	$18.1'' \times 11.1''$	$12.6'' \times 11.9''$
1628 – 1756	$15.4'' \times 9.1''$	$10.5'' \times 9.9''$
1756 – 1884	$14.5'' \times 8.9''$	$10.0'' \times 9.7''$
1884 – 2012	$13.1'' \times 8.1''$	$9.0'' \times 8.3''$

Owing to the varying declination of different observing blocks, we obtain different resolution elements.

the quasar 3C286 as a flux and bandpass calibrator for all fields. As complex gain calibrator, we used the quasar J1822-0938 for all observing blocks between $l = 14.0 - 37.9°$ (including the pilot field) and the quasar J1925+2106 for all observing blocks between $l = 47.1 - 51.2°$. The achieved resolution depends on the frequency and the sky position and varies between 10 and $25''$ (see Table 3.1 for further details). By the date of publication of this thesis, the other half of the survey will have been observed. However, since the calibration and imaging is an enormous computing and person power effort, the data reduction of that second half is still going on. The full survey will be presented in a future article.

We used the new WIDAR correlator and observed the continuum between 1 and 2 GHz using eight sub-bands, so-called spectral windows, each with a bandwidth of 128 MHz. Owing to strong contamination of radio frequency interference (RFI), we could not use two spectral windows. The frequencies of the six remaining spectral windows are given in Table 3.1. We split each spectral window further into 64 channels with a channel width of 2 MHz. This setup allows us to flag individual channels that might be contaminated by, for instance, RFI without significantly losing sensitivity.

3.3.2 Calibration

We used the CASA package (version 4.1.0) in combination with a modified VLA pipeline[a] (version 1.2.0) to edit and calibrate the data. Prior to the calibration, we manually flagged strong RFI and bad antennas. The pipeline uses automated flagging algorithms such as RFlag on the calibrator observations to improve the calibration solutions, but does not flag the target fields. Subsequently, the pipeline applies the bandpass, flux, and gain calibration. At this point, we neither used Hanning smoothing nor recalculated the data weights (CASA command 'statwt'), since this could influence very bright continuum sources. We implemented some modifications to the pipeline to improve the quality checking and performed further flagging

[a] https://science.nrao.edu/facilities/vla/data-processing/pipeline

on the target fields with automated flagging algorithms (see Sect. 3.3.3 for further details) and by hand after the pipeline run. A detailed description of our calibration procedure will be given in the THOR survey overview paper (Beuther et al., in prep).

3.3.3 Automated flagging algorithm RFlag

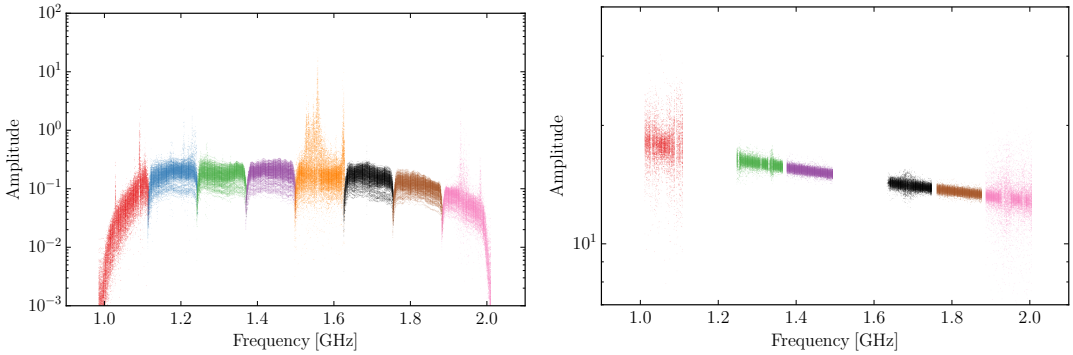

Figure 3.2: Observed amplitude of the bandpass/flux calibrator 3C286 as a function of frequency prior to the calibration on the left side and after the calibration on the right side. The color coding represents the eight spectral windows. The spectral windows at 1.2 and 1.6 GHz were flagged because of RFI (see Sect. 3.3.3).

As shown in the lefthand panel of Fig. 3.2, some spectral windows in our data are affected by RFI. The spectral windows around 1.2 and 1.6 GHz have the strongest contamination, and we cannot use them. The spectral window around 1.6 GHz is severely affected by the GPS satellites, which can be seen as outliers from the normal bandpass shape in the lefthand panel of Fig. 3.2, and we are not even able to calibrate the data. The spectral window around 1.2 GHz can be calibrated. However, the images show a consistently strong RFI contamination, which cannot be removed by the automated flagging algorithm discussed below. Within the other spectral windows, we found RFI contamination varying in frequency, sky position, and time. Therefore it is very difficult and time-consuming to flag all data manually, so we explored the possibility of automated flagging algorithms. CASA provides the so-called RFlag algorithm, which was introduced previously to AIPS by E. Greisen in 2011. The RFlag algorithm iterates the data in chunks of time and performs a time analysis for each channel, as well as a spectral analysis, for each time step and flags outliers (see the CASA manual[b] for further details). Using the standard threshold greatly improved the results as shown in Fig. 3.3. The RFlag algorithm flags almost all RFI reliably. However, a useful automated flagging algorithm must not only flag the RFI reliably, but also keep the actual scientific signal unchanged. We therefore tested the effects of the RFlag algorithm on the thermal noise in our data, as well as the flux

[b] available on the CASA webpage: http://casa.nrao.edu

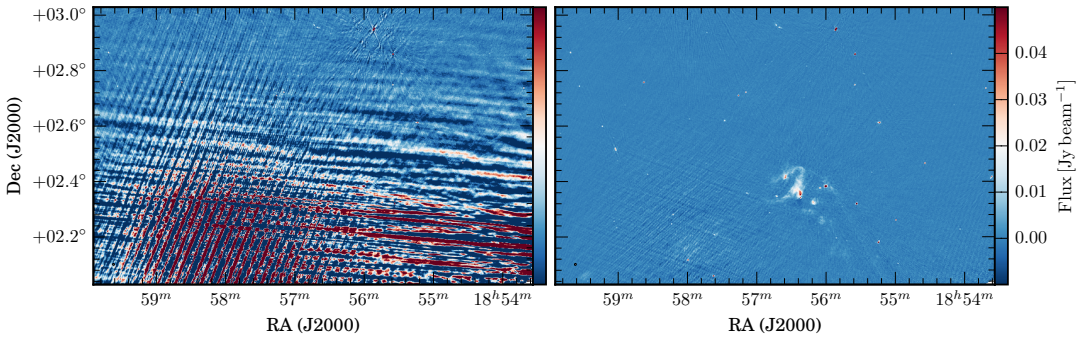

Figure 3.3: Left panel: the imaged data for one spectral window around 1.3 GHz after the calibration, without automated flagging. Strong RFI features are present. Right panel: the same region after applying the automated flagging method RFlag.

densities of our sources. For these two tests, we investigated the spectral window at ~1.4 GHz of the field around $l = 22°$. The frequency range around the H I 21 cm line is a protected band and is indeed almost free of terrestrial RFI. Applying the RFlag algorithm on this spectral window should not affect the thermal noise and the flux densities of our sources. We calibrated and cleaned the data exactly the same way, but on one data set, we applied the RFlag algorithm before cleaning, whereas we cleaned the other data set without automated flagging and used this as a reference.

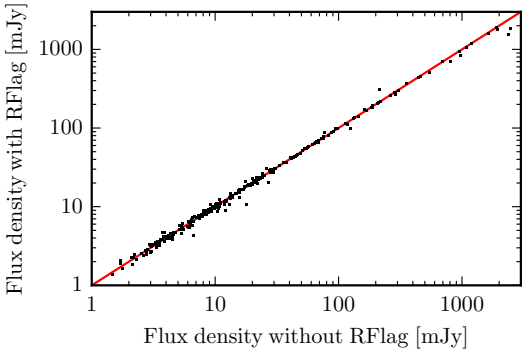

Figure 3.4: Extracted flux density of sources with the RFlag method applied as a function of the extracted flux density for the same sources without the RFlag method applied. For this comparison we used the spectral window around 1.4 GHz and the field around $l = 22°$. The red line represents a one-to-one relation. Over a wide range of flux densities, the RFlag method does not influence the actual source flux densities.

As a first test we compared the noise between the two data sets. Because the spectral window around 1.4 GHz is mostly free of RFI, we did not find different noise levels for the two data sets. This shows that the RFlag algorithm does not flag good data, which would increase the

noise level. For both data sets we extracted the continuum sources using the method described in Sect. 3.4.3, cross-matched the two data sets, and compared the flux densities for each source. Figure 3.4 shows the result of this comparison of the flux densities with and without the RFlag algorithm applied. Over the full range of flux density values, we see no significant deviation for unresolved and small sources (smaller than ∼100"). However, more extended sources might be affected by the RFlag algorithm, and this effect will be discussed in Sect. 3.3.5.
In summary, the RFlag algorithm provides a reliable tool for removing RFI from the continuum data. While the noise level and unresolved and small sources are not affected significantly by the RFlag algorithm, large extended sources have to be treated more carefully. We discuss this in more detail in Sect. 3.3.5.

3.3.4 Imaging and deconvolution

For the imaging and deconvolution, we used the task *clean* in the CASA package. Since we cover a large area on the sky, we created mosaics consisting of several pointings. This is an algorithmic, as well as a computational challenge, and we extensively tested different versions of the mosaicking algorithm in the CASA package, including versions 4.2.2, 4.3, 4.4, and a test version of 4.5. Our main focus was to obtain consistent flux density measurements, so we compared flux density and intensity values of point sources in mosaics created with the above-mentioned CASA versions with their corresponding values in individually cleaned pointings. The clean and deconvolution algorithms for single pointings are simpler and well tested and were therefore used as reference. In collaboration with the CASA developer team, we could identify several problems in the mosaic algorithm of the CASA versions 4.3, 4.4, and a test version of 4.5. Therefore we decided to use version 4.2.2 because this version provides flux density values in the final mosaics that are within ten percent of the flux density values measured in single pointings.
Since we cover a wide range of frequency from 1 to 2 GHz, we cannot clean all spectral windows together without considering the frequency dependence of the sources, as well as primary beam effects. While the CASA package is able to clean wide-band images for single pointings (using the parameter *nterm*), to date (up to version 4.4) this is not available for mosaics. We could clean each observed channel separately, but this would reduce the signal-to-noise significantly and requires immense computational resources. As a compromise, we cleaned each 128 MHz-wide spectral window separately, thus neglecting the frequency dependence inside each spectral window. Thereafter, we compared the peak intensity between the spectral windows to determine the spectral index (see Sect. 3.4.7 for further details). To suppress the sidelobes and increase the resolution, we chose $robust = 0$ as a weighting parameter, which is a compromise between uniform and natural weighting. As a pixel size, we chose 2.5", which is

sufficient to sample the smallest possible resolution element (synthesized beam width) of ~8". To achieve a uniform noise between the separate observing blocks, we included the neighboring observing blocks in the clean process. Because the clean command in CASA works with equatorial coordinates, we have to choose a large image size of 4600×4600 pixels to cover one field, consisting of three observing blocks. We applied primary beam corrections to obtain reliable flux densities. Because the continuum emission covers a wide range of spatial scales, we used the multiscale clean in CASA to recover the large scale structure. In this cleaning method we selected four different scales: besides the point source, also 1, 3, and 6× of the resolution element. We stopped the cleaning process at a threshold of 5 mJy beam^{-1} or 10^5 iterations, whichever was reached first. As the noise in our data is dominated by the sidelobe noise, the cleaning threshold is higher than the thermal noise level. The final resolution depends on the frequency of each spectral window and the declination of the observed field. Table 3.1 provides an overview of the highest and lowest resolution for each spectral window. The noise level of the images are discussed in Sec.3.4.5.

3.3.5 Extended sources

Extended sources suffer from different filtering effects owing to both the interferometer and the applied RFlag method. The first effect is due to the incomplete sampling in the uv plane. Each interferometer suffers from this effect, and it depends to first order on the shortest available baseline. Theoretically the VLA in C-configuration can observe all spatial scales up to 970"for the L band (VLA-manual). However, this value is for a 12-hour observation near the zenith and snapshot observations may recover scales diminished by a factor of two. However, this rule-of-thumb estimate might be too optimistic, and more realistic observations are not able to reach this value. To examine the insensitivity of the large spatial scales of the interferometer in a more realistic environment, we performed simulated observations of artificial sources with the THOR observation setup. We tested sources with a Gaussian intensity profile and varying sizes. These tests showed that we are able to recover sources with sizes up to ~120"reasonably well (80% flux recovery) for all frequencies between 1 and 2 GHz. To achieve this result, the use of the multiscale clean was crucial. However, Galactic sources do not show simple 2D Gaussian profiles so that to quantify the filtering effect in detail is difficult. Nevertheless, these simulations show that sources up to ~120"are not severely affected by the insensitivity to large spatial scales of the interferometer. Since this insensitivity depends on the frequency, this can artificially change the spectral index of extended sources. However, our simulated observations revealed that this only affects sources larger than ~120".

The second filtering effect for extended sources is due to the applied RFlag algorithm. Extended sources show high amplitudes for short uv distances (see Fig. 3.5). As the RFlag algorithm

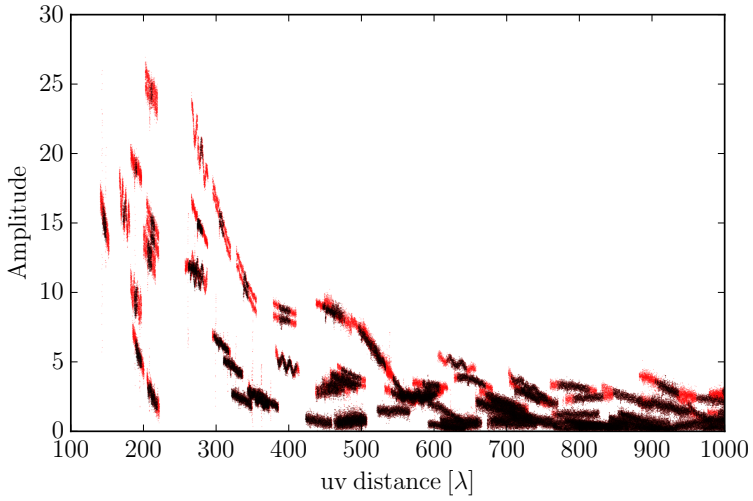

Figure 3.5: Amplitude as a function of uv distance for a single pointing close to the spatially very extended SNR G021.8-00.6 (pointing center at J2000 18:32:52.2, −10:03:32.3). The black points represent the data we used for the imaging whereas the red points represent the data points flagged by the RFlag method.

searches for outliers in a frequency and time domain, it recognizes some of these high values as outliers and flags them accordingly. Quantifying this effect is complicated because the flagging depends on the source size, its intensity, and the internal intensity structure. However, similar to our tests in Sect. 3.3.3, we used the spectral window around ∼1.4 GHz with and without applied RFlag to examine the effects of the algorithm on the large scale structure. Figure 3.5 shows the amplitude as a function of uv distance for one pointing close to the extended SNR G021.8-00.6. The red data points show the data points flagged by the RFlag algorithm. For a uv distance smaller than 300λ the RFlag algorithm flags significantly more data (∼70%) in comparison to larger uv distances (∼25%). In the simple approximation of $\theta = \lambda/D$, where θ, λ, and D are the angular scale, the wavelength, and the diameter or baseline length of the telescope, respectively, the uv distance of 300λ describes an angular scale of ∼600″. The flagging of the data points for short uv distances removes part of the large scale structure of the source. However, only large and bright sources are affected by this filtering. The SNR G021.8-00.6 has a spatial extent of ∼1200×400″ and the RFlag algorithm flags ∼40% of the flux density. Smaller sources on the order of ∼100 to 300″ show lower values of ∼5 to 10% flux density removal. Owing to these two filtering effects for largely extended (>400”) and bright (>1 Jy) sources, we refrained from analyzing these sources in detail and the corresponding flux values have to be treated cautiously. For the spectral index determination (see Sect. 3.4.7), we use the peak intensity rather than the integrated flux density since the former is less affected by the explained filtering effects.

3.4 Source extraction

In this section, we explain the source extraction method using the BLOBCAT software as well as the method of determining the spectral index. To achieve a higher signal-to-noise ratio for the source extraction, we use the average of two spectral windows to detect the sources. Thereafter we extract the peak intensity for each source in each spectral window separately to subsequently fit the spectral index.

3.4.1 Averaging spectral windows

The extraction algorithm and method can influence the resulting catalog, and several different methods are common (e.g., Williams et al. 1994; Hancock et al. 2012; Berry 2015). For our data we must solve several challenges: we want to achieve the best signal-to-noise ratio, but avoid picking up artifacts in the images caused by RFI or sidelobes. To get the best signal-to-noise ratio, mosaicking the entire spectral range from 1 to 2 GHz would be preferable; however, CASA is currently (up to version 4.4) not able to perform wide-band mosaics (see Sect. 3.3.4), and several spectral windows are severely affected by RFI (see Sect. 3.3.3). We therefore cleaned each spectral window separately. To achieve a higher signal-to-noise ratio, we averaged the two spectral windows around 1.4 and 1.8 GHz, because these spectral windows contain no significant RFI. Prior to the averaging process, we smoothed the spectral window around 1.8 GHz to the lower resolution of the spectral window around 1.4 GHz. Averaging over more than the two mentioned spectral windows does not increase the detection of sources significantly, but increases the detection of artifacts due to RFI contamination in the other spectral windows.

3.4.2 Noise estimate

Since our data are limited by the sidelobe noise, we have to consider strongly varying noise within our observed region. Close to strong emission sources, the noise is dominated by the sidelobes and can be an order of magnitude higher than in emission free regions (see Figs. 3.6 and 3.7). As a result, the main challenge is to consider this varying noise during the process of source extraction. To create a reliable noise map, we followed the instructions given in Hales et al. (2012). The described method determines the rms value for each pixel by determining the median in a specified area (50×50 px) around the pixel in the residual image from the clean process. Prior to the median estimate, the algorithm clips all peak values in the specified area until all values are within $\pm 3\sigma$, where σ is the median in the specified area. This method ensures that most real emission, which might still be present in the residual image is removed

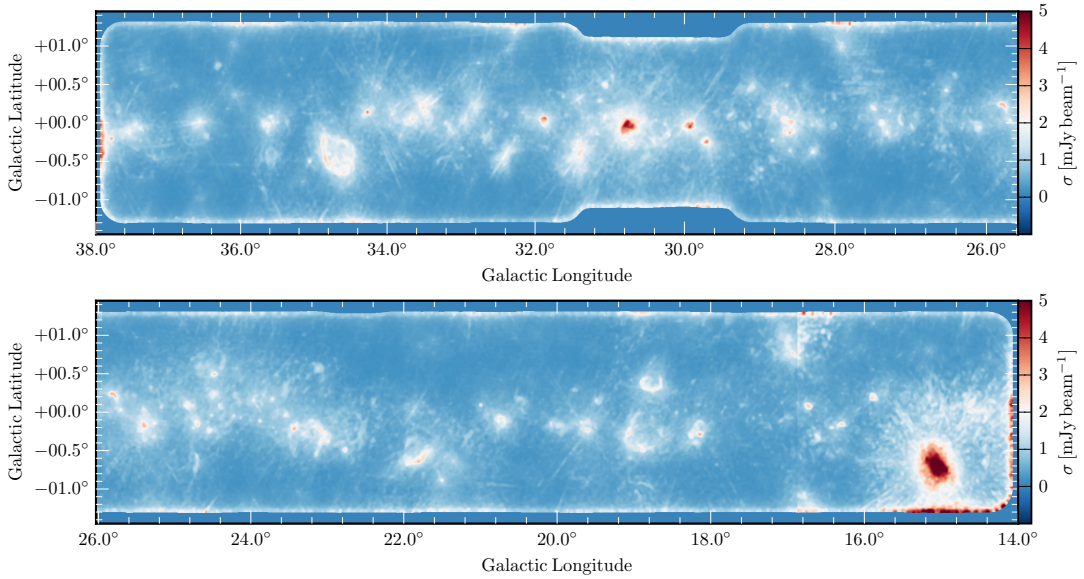

Figure 3.6: Noise map of the first part of the THOR survey using the average of two spectral windows 1.4 and 1.8 GHz (see Sect. 3.4.1).

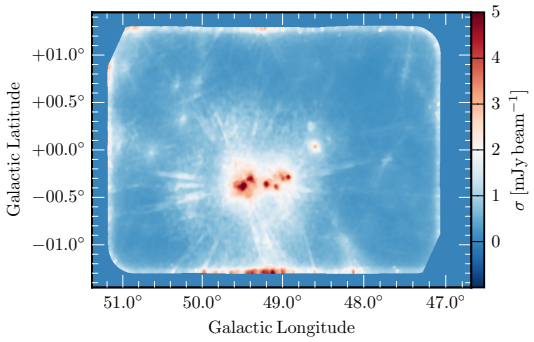

Figure 3.7: Noise map of the region around l=49° of the THOR survey using the average of two spectral windows 1.4 and 1.8 GHz (see Sect. 3.4.1).

from the noise image and the determined noise map consists of the thermal and sidelobe noise. The noise maps are given in Figs. 3.6 and 3.7.

3.4.3 BLOBCAT

We used the BLOBCAT software (Hales et al. 2012) to extract the sources from the averaged continuum images. This software is a flood-fill algorithm that considers locally varying noise. BLOBCAT creates a signal-to-noise ratio map by dividing the actual input image by the given noise map. This dimensionless map is used for the source extraction by searching for all pixels above a given detection threshold, which we set to 5σ. Thereafter, BLOBCAT identifies all

neighboring pixels around the peak pixel down to a given flooding threshold, which we set to the standard value of 2.6σ (Hales et al. 2012). These "islands" of pixels are labeled and written to a table. BLOBCAT also performs several corrections for pixellation errors, peak, and integrated surface brightness biases (see Hales et al. (2012) for further details). Using BLOBCAT, we extracted in total 4772 sources, however, this includes artifacts that we subsequently removed by hand (see Sect. 3.4.4).

3.4.4 Visual inspection

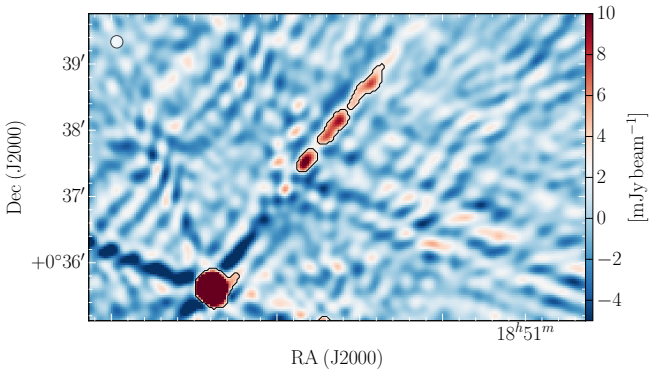

Figure 3.8: Example of an obvious artifact. A strong emission source, which is located in the bottom left part, creates sidelobes that were identified by the extraction software as actual sources. The black contours show the area of extracted sources identified by the BLOBCAT software.

Even though we have considered the spatially varying noise during the source extraction process, strong artifacts, especially sidelobes, can be picked up by the BLOBCAT extraction software. Especially problematic is sidelobe contamination from strong sources located just outside our survey boundaries, which cannot be removed by the algorithm. We therefore inspected each source visually and removed obvious artifacts by hand. Figure 3.8 shows an example of an obvious sidelobe, which was picked up by the extraction software. We identified 349 sources as obvious artifacts and removed them from the catalog. This leaves 4422 sources in the catalog. Besides the obvious artifacts, it can be difficult to distinguish between artifacts and actual sources for certain extracted sources. We classified these sources as "possible artifacts" and labeled them accordingly in the catalog. Besides visually inspected possible artifacts, we classified and labeled all sources with a signal-to-noise ratio lower than 7σ as "possible artifacts". Out of the 4772 extracted sources, we classified 1057 as "possible artifacts", 349 as artifacts, and therefore 3366 sources remain as reliable detections. The following analysis is based on the reliable detections; however, in the catalog, we also present the "possible artifacts".

3.4.5 Completeness

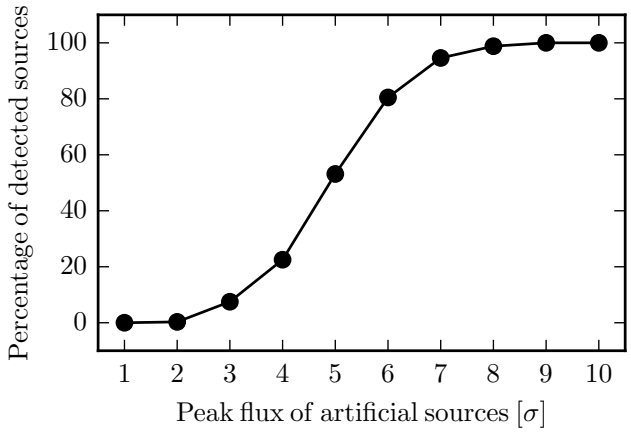

Figure 3.9: Percentage of detected sources as a function of peak intensity of the added artificial sources in units of the noise level σ.

As our noise is spatially varying, it is difficult to estimate the completeness of our catalog. In the vicinity of strong extended Galactic sources, it is not possible to detect weak extragalactic sources. Our survey is therefore incomplete in these regions. However, we performed several tests to verify our source extraction method. We chose a region of $0.5^\circ \times 0.5^\circ$ with a constant noise level and added artificial 2D Gaussian sources that have the size of the resolution element and different peak intensities. Using the source extraction method described in Sect. 3.4.3, we extracted these artificial sources and estimated the completeness. The res-

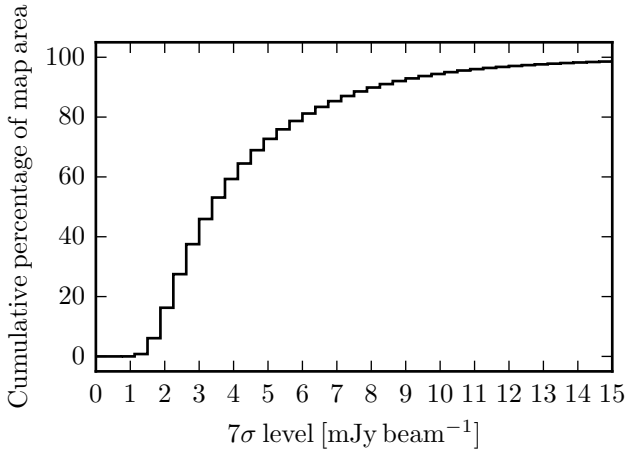

Figure 3.10: Cumulative percentage of the map area as a function of the corresponding noise level 7σ in mJy beam^{-1}. 50% of the survey area has a noise level of $7\sigma \sim 3$ mJy beam^{-1} or better.

ult is shown in Fig. 3.9. Above the chosen threshold of 7σ for reliable sources, we detected 95% of all sources. Furthermore, we determined the fraction of the area that covers a certain noise level, which is shown in Fig. 3.10. The lowest noise level in our survey of $7\sigma = 1-2\,\mathrm{mJy\,beam^{-1}}$, which is dominated by the thermal noise, is achieved in only a small fraction (~10%) of the survey area. About half the survey area has a noise level of $7\sigma < 3\,\mathrm{mJy\,beam^{-1}}$, and only 10% of the survey area shows a noise level of $7\sigma > 8\,\mathrm{mJy\,beam^{-1}}$. Using this information, we can create completeness maps for different sources intensities, which are shown in the appendix in Figs. A.3 to A.10.

3.4.6 Resolved and unresolved sources

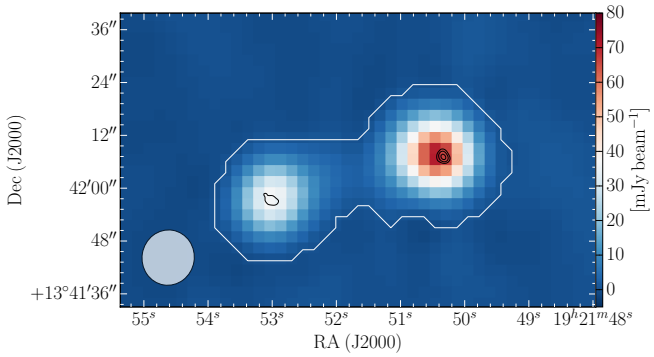

Figure 3.11: Example of an hourglass-shaped source (G48.561-0.364) that consists of two unresolved sources close together. The white contours represent the area of the source extracted by BLOBCAT. The black contours show observations from CORNISH (see Sect. 3.6.4) at 5 GHz with a resolution of 1.5"at levels of 2, 5, and 10 $\mathrm{mJy\,beam^{-1}}$.

As a first classification of the sources, we divide them in two groups: resolved and unresolved sources. The BLOBCAT software provides the number of pixels as an output, but this is not a good measurement to distinguish between resolved and unresolved sources. Because the BLOBCAT software uses a fixed threshold of 2.6σ to flood-fill the neighboring pixels around the peak pixel, the number of pixels of a source depends on the corresponding peak intensity. If we use a simple cut based on the number of pixels, we would, on the one hand, misclassify strong unresolved sources as resolved, and on the other, we would misclassify weak but closely resolved sources as unresolved. Therefore we use a comparison of the peak intensity and the flux density to distinguish between resolved and unresolved sources. The peak intensity and flux density have the same value for unresolved sources, whereas resolved sources show a higher flux density value in comparison to the peak intensity value. However, we have to consider the uncertainties in the peak intensity as well as in the flux density, so we use a less strict

condition and classify all sources as unresolved sources that have $S_\nu < 1.2 \times I_\nu$, where S_ν and I_ν are the flux density in Jy and peak intensity in Jy beam^{-1}, respectively. In our full catalog we classify 3184 sources in total as unresolved and 1238 sources (28%) as resolved. For the sources with the flag "possible artifacts" (see Sect. 3.4.4), the ratio of unresolved and resolved sources is similar with 76% of the sources being unresolved.

This classification scheme classifies two overlapping, but unresolved sources as resolved. For unresolved sources that are randomly distributed in the sky, this arrangement is unlikely, however for extragalactic radio lobes, this overlap can occur frequently. Figure 3.11 shows an example of two radio lobes that are close together. Even though each radio lobe is unresolved, we extracted them as one source and hence the flux density is larger than the peak intensity and we classify them as resolved. This affects the classification of extragalactic and Galactic sources. However, in many cases (e.g., Fig 3.11), the spectral index helps to resolve this problem.

The flux density of unresolved sources can be affected in several ways and therefore has to be treated cautiously. We find for the ratio of S_ν/I_ν values less than one, which means that the flux density is lower than the peak intensity. For unresolved sources, this ratio should be one. We could identify three reasons for this low ratio. First, the source extraction software BLOBCAT does not fit enough pixels for weak sources, which lowers the flux density. In extreme cases, the fitted area of BLOBCAT can be smaller than the resolution element. Second, unresolved sources can be situated in slightly negative sidelobes from nearby strong extended sources, which affects the flux density, as well as the peak intensity, and this can change the ratio. Third, weak sources ($I_\nu \lesssim 5$ mJy beam^{-1}) are not cleaned properly because these sources are below our cleaning threshold, which lowers the measured flux density and changes the ratio of S_ν/I_ν to values below one. We therefore suggest using peak intensities for unresolved sources for further analysis, and we indicate the corresponding flux densities within our catalog with brackets.

3.4.7 Spectral index determination

As our observations cover a wide bandwidth from 1 to 2 GHz, we are able to determine spectral indices by extracting the peak intensity of each source within each spectral window and perform a fit of the spectral index α with the form $I(\nu) \propto \nu^\alpha$. As explained in Sect. 3.3.5, we use the peak intensity instead of the integrated flux density to determine the spectral index, since the peak intensity is less affected by filtering effects for extended sources. For unresolved sources, both quantities reveal the same result. To overcome problems due to different resolutions, we smooth all spectral windows to a common resolution of 25″ prior to extracting the peak intensity. Furthermore, we use the same technique as described in Sect. 3.4.2 to determ-

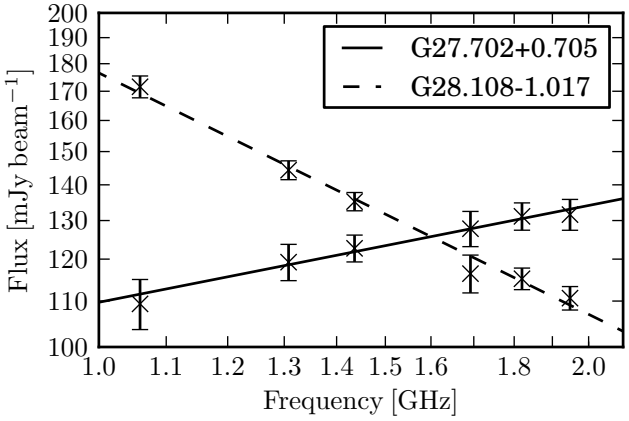

Figure 3.12: Example of the extracted peak intensity as a function of frequency. Each spectral window is represented by one data point including the 3σ uncertainty. G27.702+0.705 is represented by the solid line and has a spectral index of $\alpha = 0.29 \pm 0.03$, whereas G28.108-1.017 is represented by the dashed line and has a spectral index of $\alpha = -0.72 \pm 0.02$.

ine the spatially varying noise and to estimate the noise within each spectral window. Because we smooth two spectral windows to perform the source extraction, the signal-to-noise ratio is higher for the source extraction in comparison to the intensity extraction within each spectral window separately. We therefore use a less rigid threshold for the intensity extraction of 3 sigma in comparison to 5 sigma for the source extraction. The extracted peak intensities for each spectral window are given in the catalog presented in this paper. Figure 3.12 shows an example for the extracted intensities, including the fit of the spectral index for two different sources. Fig. 3.13 presents an example source showing the images of all spectral windows that are included in the spectral index fit. In the appendix in Figs. A.1 to A.2, two more examples are given. We use the scipy function "curve_fit" to fit the data points and use the uncertainty of the fit as the uncertainty for the spectral index. With this method, we can determine a spectral index for 3625 sources.

For some sources, we are not able to extract the peak intensity for all six spectral windows, owing to higher noise or contamination by RFI, for example. In such cases, we determine the spectral slope from the remaining data points. Naturally, this leads to larger uncertainties. As a result, we introduce the label "reliable spectral index" for all sources that have a reliable intensity for all six spectral windows, hence a reliable spectral index fit. The catalog contains 1840 sources that fulfill this criterion, which is about 50% of the sources where it is possible to determine a spectral index. Figure 3.14 shows the distribution of the uncertainty of the determined spectral index for all sources and for the sources with the label "reliable spectral index". The labeled sources show a significantly smaller uncertainty with a mean of

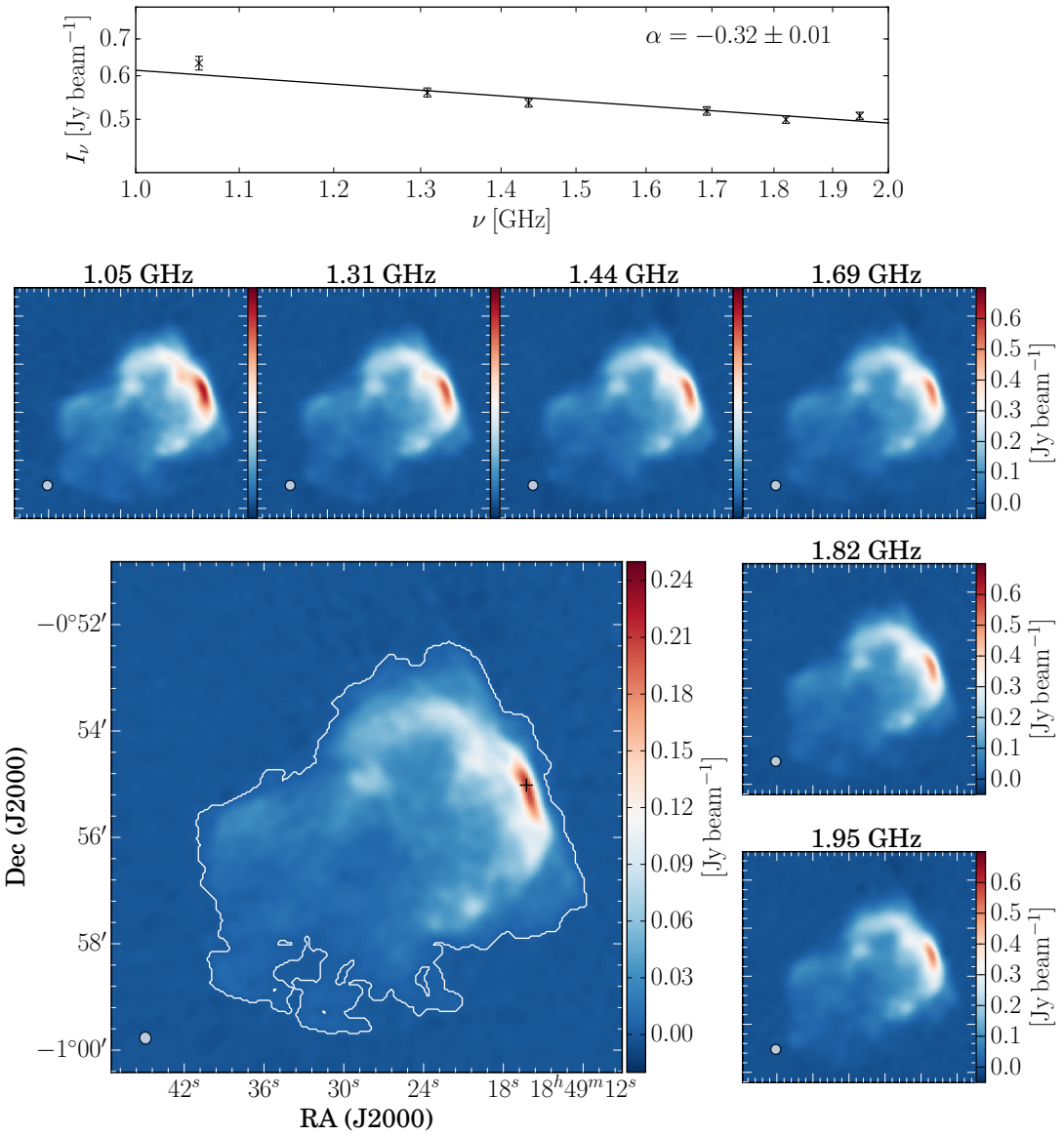

Figure 3.13: Example image of the THOR source G31.869+0.064, which corresponds to the known SNR G31.9+00.0 (Green 2014). The large image represents an averaged image of the two spectral windows around 1.4 and 1.8 GHz, which we used for the source extraction (see Sects. 3.4.1 and 3.4.3). The white contours show the extent of the source determined by the BLOBCAT algorithm. The black cross marks the peak position, which we used to determine the spectral index. The small images show each spectral window separately, and the top panel presents the peak intensity for each spectral window and the corresponding spectral index fit. In each image the synthesized beam is given in the lower left corner.

$\Delta\alpha = 0.18$, whereas all sources show a mean uncertainty of the spectral index of $\Delta\alpha = 0.62$. In the following, we concentrate our analysis of the spectral index on the sources with reliable spectral indices.

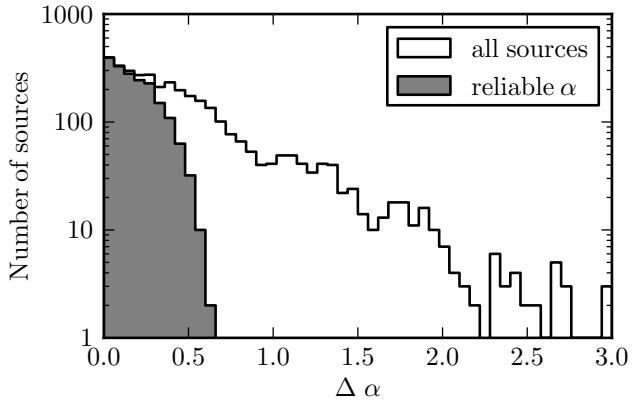

Figure 3.14: Histogram of the uncertainty of the determined spectral index. The black line includes all sources for which we are able to determine a spectral index, whereas the gray shaded area represents the sources for which we have an intensity measurement in all six spectral windows and therefore a reliable spectral index measurement.

3.5 Catalog

The published catalog contains 27 entries for each source. Table 3.3 describes each entry in detail. As explained, we use a smaller beam for the source extraction than for the intensity extraction. This makes the published values for the corresponding peak intensities different. Table A.1 shows an example, and Fig. 3.13 presents an example sources showing all the different data we used, including the spectral index fit. Further examples can be found in Figs. A.1 - A.2.

Table 3.2: Statistics of the catalog

Description	Number	Percentage
All	4422	100%
Unresolved sources	3184	72%
Resolved sources	1238	28%
Possible artifacts	1057	24%
Reliable alpha	1840	41%

Table 3.2 summarizes the number of extracted sources, including the introduced labels. The exact numbers have to be treated cautiously. Compact sources superimposed on large regions of extended emission are missed in the catalog. In contrast to this, large, extended sources, such as SNRs, can be split up in different sources and therefore create multiple entries in our catalog, even though the emission occurs most likely from the same object. The majority (72%) of the extracted sources are not resolved. Most of them might be extragalactic in origin

Table 3.3: Description of the catalog entries.

Col.	Name	Unit	Description
1	Gal. ID		Name of the source the form G'Gal. long'±'Gal. latitude'[1]
2	RA	deg	RA in J2000 of the peak position.
3	Dec	deg	Dec in J2000 of the peak position.
4	S_p [2]	Jy beam^{-1}	Peak intensity of aver. image used for source extraction (see Sect. 3.4.1).
5	SNR		Signal-to-noise ratio in the averaged image.
6	S_int	Jy	Integrated flux density of the averaged image (see Sect. 3.4.3).
7	BMAJ	arcsec	Major axis of the resolution element used for the source extraction.
8	BMIN	arcsec	Minor axis of the resolution element used for the source extraction.
9	BPA	deg	Rotation angle of the resolution element used for the source extraction.
10	n_pix		Number of pixels flooded by BLOBCAT (see Sect. 3.4.3).
11	resolved_source		Resolved source label (see Sec 3.4.6). 1 = True, 0 = False.
12	possible_artifact		Label for possible artifacts and/or SNR<7. 1 = True, 0 = False.
13	S_p(spw-1060) [3]	Jy beam^{-1}	Peak intensity around 1.06 GHz used for spectral index (see Sect. 3.4.7).
14	delta_S_p(spw-1060) [3]	Jy beam^{-1}	Uncertainty of peak intensity around 1.06 GHz.
15	S_p(spw-1310) [3]	Jy beam^{-1}	Peak intensity around 1.31 GHz used for spectral index (see Sect. 3.4.7).
16	delta_S_p(spw-1310) [3]	Jy beam^{-1}	Uncertainty of peak intensity around 1.31 GHz.
17	S_p(spw-1440) [3]	Jy beam^{-1}	Peak intensity around 1.44 GHz used for spectral index (see Sect. 3.4.7).
18	delta_S_p(spw-1440) [3]	Jy beam^{-1}	Uncertainty of peak intensity around 1.44 GHz.
19	S_p(spw-1690) [3]	Jy beam^{-1}	Peak intensity around 1.69 GHz used for spectral index (see Sect. 3.4.7).
20	delta_S_p(spw-1690) [3]	Jy beam^{-1}	Uncertainty of peak intensity around 1.69 GHz.
21	S_p(spw-1820) [3]	Jy beam^{-1}	Peak intensity around 1.82 GHz used for spectral index (see Sect. 3.4.7).
22	delta_S_p(spw-1820) [3]	Jy beam^{-1}	Uncertainty of peak intensity around 1.82 GHz.
23	S_p(spw-1950) [3]	Jy beam^{-1}	Peak intensity around 1.95 GHz used for spectral index (see Sect. 3.4.7).
24	delta_S_p(spw-1950) [3]	Jy beam^{-1}	Uncertainty of peak intensity around 1.95 GHz.
25	alpha		Spectral index of source using all available data points (see Sect.3.4.7).
26	delta_alpha		Uncertainty of spectral index.
27	reliable_alpha		Label for reliable spectral index (see Sect. 3.4.7). 1 = True, 0 = False.

[1] Indicating the peak position.
[2] Synthesized beam is different for different fields and is given in rows 7-9.
[3] Synthesized beam is smoothed to 25"×25".

since their spectral indices are negative (see Sect. 3.6.5). About 28% of the extracted sources are classified as resolved, but as explained in Sect. 3.4.6, some of them might be two closely separated sources. The distribution of resolved and unresolved sources as a function of Galactic latitude is shown in Fig. 3.15. This reveals an over-density of resolved sources close to the

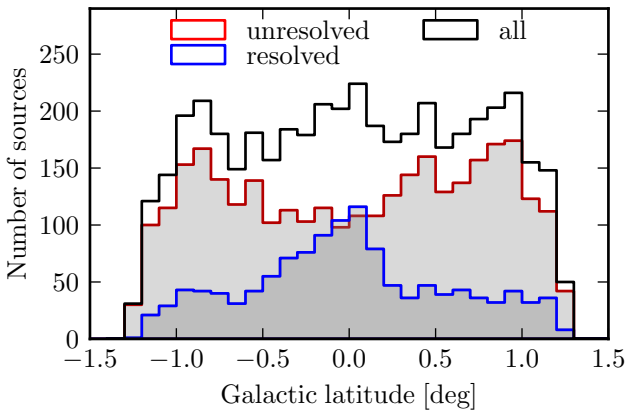

Figure 3.15: Histogram for the number of sources as a function of Galactic latitude.

Galactic midplane, whereas unresolved sources are equally distributed. The distribution drops for $|b| > 1^\circ$ as the noise increases at the survey edges. The distribution of the unresolved sources also indicates a slight drop toward the Galactic plane ($b \sim 0^\circ$) because we miss weak extragalactic sources in the close vicinity of strong Galactic sources, which are mostly located along the Galactic plane. Similar results can be found in Helfand et al. (2006). This shows that a large number of the sources in our catalog are not confined to the Galactic plane and therefore have an extragalactic origin.

3.6 Discussion

3.6.1 Comparison with other surveys

Since the THOR survey is not the first cm-continuum survey in the Galactic plane, we compare our results to previous surveys to check for consistency in the flux density, intensity, and position. We focus our comparison on three major surveys: the Multi-Array Galactic Plane Imaging Survey (MAGPIS, Helfand et al. 2006), The NRAO VLA Sky Survey (NVSS, Condon et al. 1998), and the Co-Ordinated Radio 'N' Infrared Survey for High-mass star formation (CORNISH, Hoare et al. 2012; Purcell et al. 2013).

3.6.2 MAGPIS

The MAGPIS survey (Helfand et al. 2006) used the VLA in D, C, and B configurations to map the Galactic plane in the region $5^\circ < l < 48^\circ$ and $|b| < 0.8^\circ$ with two continuum bands at 1365 and 1435 MHz, achieving a resolution of $\sim 6''$ and a sensitivity limit of 1-2 mJy, depending on neighboring bright extended emission. They cataloged 3000 discrete sources in the region $5^\circ < l < 32^\circ$ with diameters less than $30''$ and 400 diffuse sources. Within the overlap region of the THOR survey ($14.2^\circ < l < 32^\circ$, $|b| < 0.8^\circ$), the MAGPIS catalog contains 2256 discrete and 290 extended sources. The THOR continuum catalog contains 1848 sources in the same area, including possible artifacts and therefore fewer sources than the MAGPIS catalog. Using a best match method and a circular matching threshold of 20", we match 1568 sources in total. Choosing a smaller matching threshold of $5''$ does not change the result significantly. Owing to different spatial filtering of the THOR and MAGPIS data, the determined area for extended sources is different within the two surveys. This effect accounts for the majority of the non-matches. Merely matching the point sources of the THOR survey reveals a matching rate of ~92%, including possible artifacts. If we do not consider the possible artifacts, the matching rate is even higher with ~97% and the matching rate considering only the possible artifacts is ~78%. This shows that almost all reliable sources within the THOR catalog have a

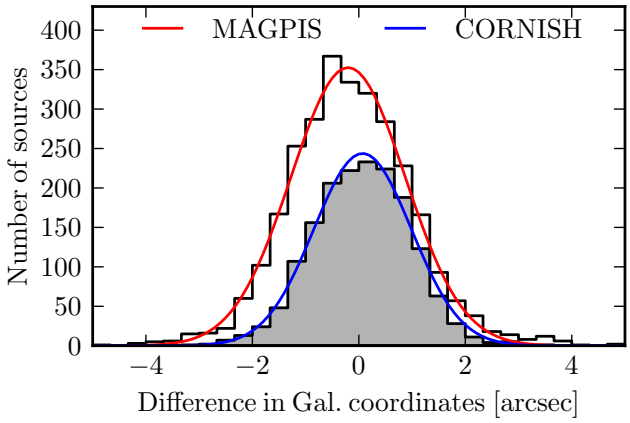

Figure 3.16: Histogram for the number of sources as a function of the difference in Galactic coordinates of the peak position for the matched sources with the MAGPIS and CORNISH catalogs in red and blue, respectively. The black histogram represents the actual data, whereas the colored lines show the corresponding fits.

counterpart in the MAGPIS catalog, and therefore the number of false positives due to artifacts or sidelobes is low within our catalog. Since the matching rate for possible artifacts is still high, the majority of these sources will also be real detections.

Because the matching with the MAGPIS survey worked well, we used the matched sources to verify the positions, as well as the flux density. For these comparisons we employed the MAGPIS discrete source catalog and neglected the diffuse sources because they suffer from different

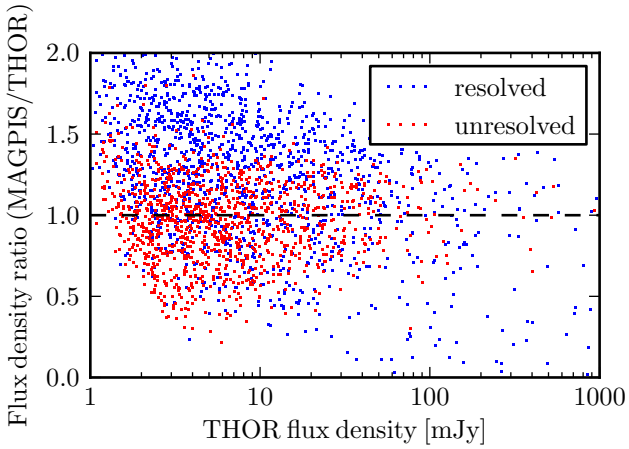

Figure 3.17: Ratio of the MAGPIS and THOR flux density as a function of the THOR flux density. The red and blue points represent the unresolved and resolved sources, respectively, as defined in Sect. 3.4.6. The dashed black line represents a one-to-one relation.

spatial filtering, which makes the comparison inaccurate. Figure 3.16 shows the histogram of the difference in Galactic coordinates for the peak position, along with the corresponding fits. We used a Gaussian function to fit the distribution and find a shift of -0.2″ and a FWHM for the distributions of 2.5″, which is the size of one pixel. The comparison of the flux density is shown in Fig. 3.17. Similar to the NVSS sources, the unresolved sources show a tight correlation. In contrast to this, the resolved sources show higher flux density values in the MAGPIS data, owing to less filtering. These tests show that our observation, calibration, and imaging processes work well, and our work is consistent with previous observations.

3.6.3 NVSS

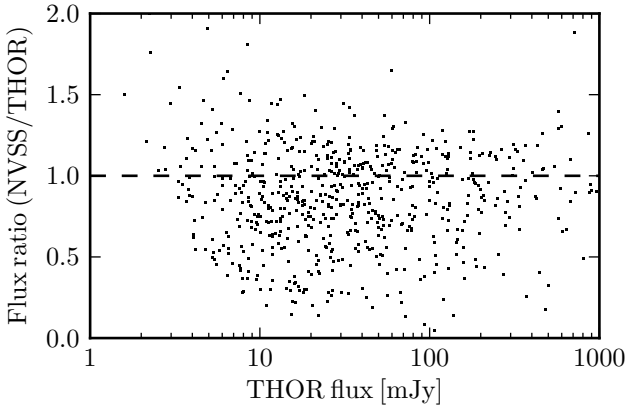

Figure 3.18: Flux density comparison with the NVSS data. The ratio of NVSS and THOR flux density is shown as a function of the THOR flux density for all matched sources with a matching threshold of 5"for the peak positions. The black dashed line represents a one-to-one relation.

The NVSS (Condon et al. 1998) is a continuum survey at 1.4 GHz with the VLA in D and DnC configuration covering the northern sky for J2000 $\delta > -40°$. The catalog contains $\sim 2 \times 10^6$ sources with a sensitivity limit of ~2.5 mJy and a resolution of 45". Within the region of the THOR continuum catalog, the NVSS catalog contains 7587 sources and therefore almost twice as many sources as our catalog. We find a match of 1351 sources for a circular matching threshold of 20″ and only 657 for a circular matching threshold of 5"for the peak position. Further analysis of the NVSS images showed that the NVSS catalog is severely contaminated with obvious false detections due to strong sidelobes from sources close to the Galactic plane or due to ghost artifacts (Grobler et al. 2014). Therefore the matching process is not reliable for large matching radius as we match THOR sources with false positives in the NVSS catalog. To overcome this problem, we only compare the measured flux densities for all matched sources

with a tight matching radius of 5". The result is shown in Fig. 3.18. For these tightly matched sources, the flux density comparison shows a good correlation over three orders of magnitude, with a slight bias. As shown in Fig. 3.18, the THOR flux density values are slightly higher than the NVSS flux density values. This bias is visible for resolved and unresolved sources so is not a filtering effect. We do not have a good explanation for this bias. However, the THOR and MAGPIS flux densities are consistent (see Sect. 3.6.2), and we report a slight inconsistency with the NVSS flux densities. Since the matching with the NVSS catalog is difficult due to artifacts in the NVSS images, we refrain from comparing the peak positions of the sources, but we perform this comparison with MAGPIS and CORNISH.

3.6.4 CORNISH

CORNISH (Hoare et al. 2012; Purcell et al. 2013) is a Galactic plane survey from $10° < l < 65°$ and $|b| < 1°$ using the VLA in B and BnA configuration at a frequency of 5 GHz. Therefore, the resulting resolution of 1.5″ is higher than the THOR observations, but objects larger than ∼14″ are filtered out. The mean noise level is ∼0.4 mJy beam^{-1} and 3062 sources are detected above a 7σ limit. Within the THOR region, CORNISH includes 1367 reliable sources. We used all THOR sources, which we classify as unresolved to match the CORNISH sources, and we find a best match of 834 sources using a circular matching threshold of 20″ for the peak position. As the frequency and filtering is different, we refrained from comparing the flux densities, but we verified the peak positions. Figure 3.16 shows the difference in Galactic coordinates for the matched sources. Similar to the comparison with the MAGPIS survey (see Sect. 3.6.2), we do not detect a significant shift or offset in the distribution, and the corresponding Gaussian fit has a shift of 0.07″ and a FWHM of 2.3″. The small position offset between the sources in the MAGPIS and CORNISH surveys with the THOR survey show that our data do not suffer significantly from systematical uncertainties for the position and the uncertainty of the position depends on the synthesized beam and the signal-to-noise ratio for each source and is better than 2″.

3.6.5 Spectral index

As outlined in the introduction, the spectral index allows us to distinguish various physical processes. In Sect. 3.4.7, we determine a reliable spectral index for ∼1800 sources. This information helps to distinguish between thermal and non-thermal radiation, showing positive and negative spectral indices, respectively. Figure 3.19 shows the distribution for the spectral index. Considering all sources (black line in Fig. 3.19), we find a prominent peak around $\alpha \sim -1$ and a second weaker peak around $\alpha \sim 0$. Considering only the unresolved sources

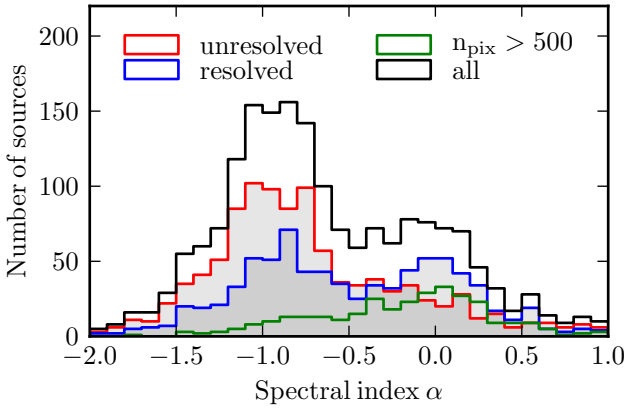

Figure 3.19: Histogram of the spectral index for all sources with a reliable spectral index measurement (~1800 sources). The black histogram represents all sources, whereas the red and blue histogram represents unresolved and resolved sources. The green histogram shows the spectral index of all sources that are larger than 500 pixels.

(red line in Fig. 3.19), we recover the strong peak around $\alpha \sim -1$, whereas the second peak around $\alpha \sim 0$ decreases. As a result, most of the unresolved sources show a negative spectral index that indicates non-thermal synchrotron radiation. Therefore, we classify the unresolved sources with a negative spectral index of $\alpha < -0.2$ as extragalactic sources. In contrast to the unresolved sources, the resolved sources (blue line in Fig. 3.19) show two peaks at $\alpha \sim -1$ and $\alpha \sim 0$. Most of the resolved sources with a flat spectral index can be matched with Galactic H II regions (see Sect. 3.6.6). The resolved sources with negative spectral index are mainly radio lobes from extragalactic jets that were classified as resolved sources, but might be two overlapping unresolved sources (see Sect. 3.4.6). If we consider only the largest sources in our sample with more than 500 pixels, which corresponds to an effective radius of $\sim 32''$ (green line in Fig. 3.19), we find a broad distribution from $\alpha \sim -1$ to 0.5. The sources with flat spectral index can be classified as H II region, and the sources with negative spectral index can be SNR. We explore the H II regions and SNR in more detail in Sects. 3.6.6 and 3.6.7, respectively.

3.6.6 H II regions

Since H II regions are formed by OB stars, they are ideal objects to locate high-mass star formation. Anderson et al. (2014) present the most complete catalog of these objects, using mid-infrared observations from the Wide-Field Infrared Survey Explorer (WISE) satellite (Wright et al. 2010). They detected ~8000 sources within the Milky Way. Approximately 2000 of these sources are H II regions with measured velocities from ionized gas spectroscopy, whereas

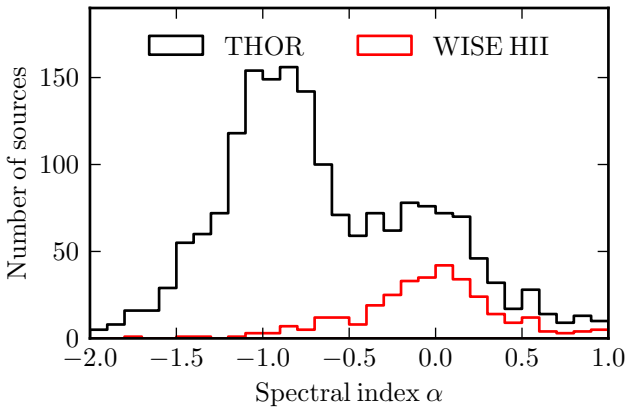

Figure 3.20: Histogram of the spectral index determined with the THOR data as explained in Sect.3.4.7. The red and black lines represent the matched WISE H II sources and all THOR continuum sources that reveal a reliable spectral index, respectively.

the remaining 6000 are H II region candidates that either show (~2000 objects) or do not show (~4000 objects) radio continuum emission. The mid-infrared sizes range from 10″ to more than 20′ with a mean of ~100″ for all previously known H II regions. The wide range of sizes for the H II regions makes it challenging to match them with our continuum catalog. Within a single large H II region, we usually detect several extragalactic background sources, which are not related to the H II region. A visual inspection of all sources is very time consuming. However, we visually inspected 6 deg^2 (~10%) of the THOR region and used this result to test several automated matching methods. For the automated matching, the best result was achieved if we exclude large H II regions from WISE with $r > 150''$ and use only the reliable THOR continuum sources. As a matching threshold, we used the size of the H II region. This method could recover over 90% of the visually inspected sources with less than 10% false detections. Within the THOR region, the WISE H II region catalog (Anderson et al. 2014) contains 791 sources that show radio emission and are smaller than $r < 150''$, including known H II regions, as well as H II region candidates. Using the described matching threshold, we match 388 sources.

As H II regions exhibit thermal radio emission, we expect a flat or positive spectral index, depending on the optical depth. Out of the 388 matched sources, 326 show a reliable spectral index. Figure 3.20 presents the distribution of the reliable spectral index for all matched sources (red histogram) in comparison to the entire THOR continuum catalog (black histogram). As expected, we find a single peak around zero, which confirms the thermal origin of the radiation for these sources. About 80% of the matched WISE H II regions are resolved, which is a significantly larger percentage than for the entire set of THOR continuum sources (~30%). For

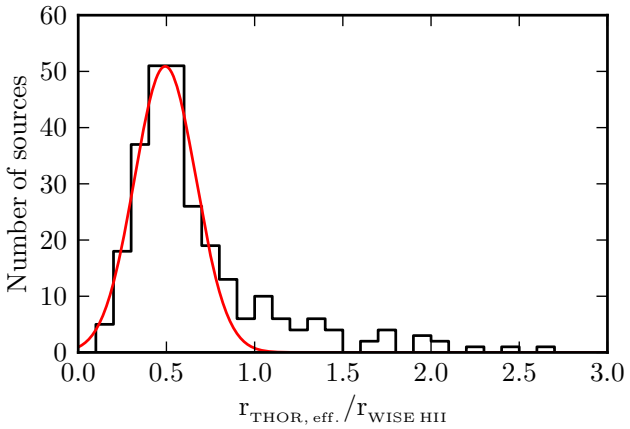

Figure 3.21: Radius of the H II regions measured by the THOR data compared to the mid-infrared WISE data given by Anderson et al. (2014). The red line represents a Gaussian fit.

the resolved sources, we can compare the sizes of the H II regions measured in mid-infrared emission by the size measured in the radio emission. For the THOR sources we can estimate an effective radius for the area determined with the BLOBCAT extraction algorithm. The mid-infrared emission at 12 —m traces the photo-dissociation region at the outer edge of the H II regions, whereas the radio emission traces the enclosed ionized gas in the interior of the H II regions. This can be seen for several H II regions presented by Watson et al. (2008). We therefore expect that the ratio of the THOR radius divided by the WISE radius is less than one. The result of this comparison is shown in Fig. 3.21. We find a close correlation around 0.5. This close correlation has to be treated cautiously, as the comparison suffers from systematic uncertainties due to the different methods of the size determination. The radius for the WISE H II regions is determined visually with circles, whereas the radius of the THOR sources is an effective radius of the extracted source area.

3.6.7 Supernova remnants

To date the most complete catalog of Galactic SNR contains 294 sources (Green 2014). Most of them are discovered or confirmed in the radio continuum. The size distribution and intensity of these SNR varies over several orders of magnitude, making them difficult to observe. Within the THOR region, the SNR catalog by Green (2014) contains 43 sources. Out of these 43 sources, we identify 26 sources within our catalog. Since the SNR can be very clumpy, several sources within the THOR continuum catalog can be matched with a single SNR from the catalog by Green (2014). Table A.2 shows the sources that are matched visually. Seventeen SNR from the catalog by Green (2014) are below the threshold for our THOR continuum cata-

Table 3.4: MAGPIS SNR candidates in comparison with THOR sources, Green SNR and WISE H II regions.

SNR candidate	THOR source	Green SNR	WISE H II	diam. [']	distance [kpc]	α	$\Delta\alpha$
G16.358−0.183			G016.352-00.179	2.8		–	–
G17.017-0.033	G17.030-0.069	G017.0-00.0	G017.032-00.052[2]	4.0		-0.19[3]	0.64
G17.336-0.139	G17.335-0.139		G017.336-00.146	1.8		0.09	0.29
G18.150-0.172[1]	G18.193-0.174	G018.1-00.1	G018.195-00.171[2]	7.0	6.3±0.5	-0.37	0.10
	G18.171-0.213			–		-0.68	0.25
	G18.107-0.134			–		-0.72	0.25
G18.254-0.308[1]	G18.270-0.289		G018.253-00.298	3.5	4.3±0.6	-0.46	0.05
G18.638-0.292	G18.610-0.316	G018.6-00.2		4.0	4.6±0.6	0.17	0.22
G18.758-0.074[1]	G18.760-0.072			1.6	4.9±0.6	-0.35	0.08
G19.461+0.144[1]	G19.492+0.135		G019.489+00.135[2]	6.0	6.8-17.5	0.15	0.02
	G19.475+0.173			–		-0.30	0.12
G19.580-0.240[1]	G19.610-0.235		G019.554-00.248[2]	3.2	6.3±0.5	0.95	0.01
	G19.555-0.230			–		-0.14	0.28
G19.592+0.025[1]	G19.592+0.028		G019.594+00.024	0.8		-0.41	0.19
G19.610-0.120[1]	G19.614-0.133		G019.629-00.095	4.5	11.6±0.5	-0.45	0.11
G19.660-0.220[1]	G19.610-0.235		G019.675-00.226[2]	4.5		0.95	0.01
	G19.691-0.204			–		-0.41	0.28
G20.467+0.150	G20.502+0.155	G020.4+00.1		5.5		-0.54	0.33
G21.557-0.103		G021.5-00.1	G021.560-00.108	4.0		–	–
G21.642+0.000	G21.632-0.007		G021.634-00.003	2.8		-0.28	0.49
G22.383+0.100	G22.360+0.064		G022.357+00.064[2]	7.0		-0.72	0.18
G22.758-0.492[1]	G22.760-0.478		G022.761-00.492	3.8	5.1±0.6	-0.04	0.04
G22.992-0.358	G22.980-0.370		G022.988-00.360[2]	3.8	5.0±0.5	-0.51	0.18
	G22.974-0.345			–		-0.12[3]	0.60
G23.567-0.033	G23.541-0.039		G023.572-00.020[2]	9.5	6.4±0.7	-0.03[3]	0.14
	G23.585+0.030			–		0.36	0.14
	G23.645-0.028			–		0.63[3]	0.46
G24.180+0.217	G24.200+0.192		G024.185+00.211[2]	5.2		0.13[3]	0.41
	G24.197+0.243			–		-0.09[3]	0.15
	G24.166+0.251			–		-0.03	0.22
G25.222+0.292	G25.220+0.286		G025.220+00.289	2.0		-0.00	0.24
G27.133+0.033[1]	G27.158+0.063			11.0	6.1-16.2	-0.63	0.28
	G27.119-0.027			–		-0.39	0.21
G28.375+0.203			G028.376+00.208[2]	10.0		–	–
G28.517+0.133				14.0	6.2-15.9	–	–
G28.558-0.008[1]	G28.569+0.020			3.0	6.5-15.9	-0.15	0.09
G28.767-0.425				9.5		–	–
G29.067-0.675			G029.088-00.675	8.0		–	–
G29.078+0.454[1]	G29.079+0.458			0.7		-0.19	0.08
G29.367+0.100[1]	G29.372+0.104			9.0	5.8-15.8	-0.40	0.14
G30.849+0.133[1]	G30.854+0.151		G030.847+00.140	2.2	6.7-15.6	-0.11	0.04
	G30.866+0.114			–		0.71	0.06
	G30.839+0.117			–		-1.09	0.19
G31.058+0.483[1]	G31.057+0.497		G031.054+00.491[2]	4.5	6.6-15.5	-0.08	0.10
	G31.034+0.459			–		-0.34	0.19
G31.610+0.335[1]	G31.598+0.330		G031.610+00.335	3.1	6.6-15.5	0.19	0.19
G31.821-0.122	G31.823-0.117		G031.806-00.115	1.8		-0.10	0.24

The SNR candidates are given in Helfand et al. (2006). Since the SNR candidates can be clumpy, several THOR sources can be found within one MAGPIS SNR candidate so we list all corresponding THOR sources. The associated SNR names taken from Green (2014) are given, as well as associated WISE H II regions given in Anderson et al. (2014). The given H II regions can be associated with the SNR, but they can also be foreground or background objects. Figure 3.23 shows the size for each H II region. The diameter of the MAGPIS SNR candidates is taken from Helfand et al. (2006), and the distance is determined via H I absorption and taken from Johanson & Kerton (2009). The spectral index α is measured for the peak position of the THOR source (see Sect. 3.4.7) and can vary significantly within the source (see Fig. 3.23).

[1] Spectral index map is shown in Fig. 3.23.

[2] Multiple H II regions can be found within the region.

[3] Determination of the spectral index is uncertain, since the source is not detected in all spectral windows.

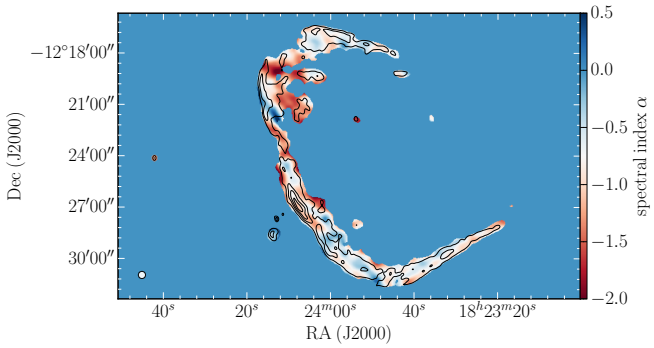

Figure 3.22: Spectral index α of the SNR G018.8+00.3 (THOR source G18.761+0.287). The black contours represent the continuum emission in steps of 10, 20, 30, and 40 mJy beam^{-1}.

log and are not extracted. These SNR are either too weak, not visible in the radio continuum, or too diffuse, and we filter them out with the VLA C-Array configuration. However, we visually inspected the missing 17 SNR and could find traces of at least nine SNR, below the used extraction threshold of 5σ.

The spectral index for SNR peaks around $\alpha \sim -0.5$ (Green 2014; Dubner & Giacani 2015). The spectral index can vary spatially for the same SNR from $\alpha \sim 0$ to -1 (e.g., Bhatnagar et al. 2011; Reynoso & Walsh 2015). Here we highlight one example to demonstrate the capability of this data set and show the spectral index map of the well known SNR G018.8+00.3 in Fig. 3.22. Furthermore, we use this technique to examine several SNR candidates proposed in the MAGPIS survey.

MAGPIS (Helfand et al. 2006) provides 49 new SNR candidates, which are only partly included in the SNR catalog by Green (2014). Their criteria for a SNR candidate is that they have: 1.) a high ratio between 20 cm continuum and 21 μm flux density, 2.) a counterpart at 90 cm continuum emission, and 3.) a distinctive SNR morphology, e.g., shell-type or filled-center (Dubner & Giacani 2015). Followup observations for these candidates to determine the distance using H I absorption are presented in Johanson & Kerton (2009). Since MAGPIS has higher spatial resolution, it is more sensitive at detecting the mentioned SNR morphology. However, the THOR survey can help to characterize the radiation and distinguish between thermal and non-thermal radiation. Within the THOR region, we find 33 MAGPIS SNR candidates, which are listed in Table 3.4. Only five of them are included in the SNR catalog presented by Green (2014). In contrast to this, 24 of these MAGPIS SNR candidates have at least one counterpart in the WISE H II region catalog presented by Anderson et al. (2014). However, thanks to the high density of H II regions within the Galactic plane, these can be chance alignments along the line of sight.

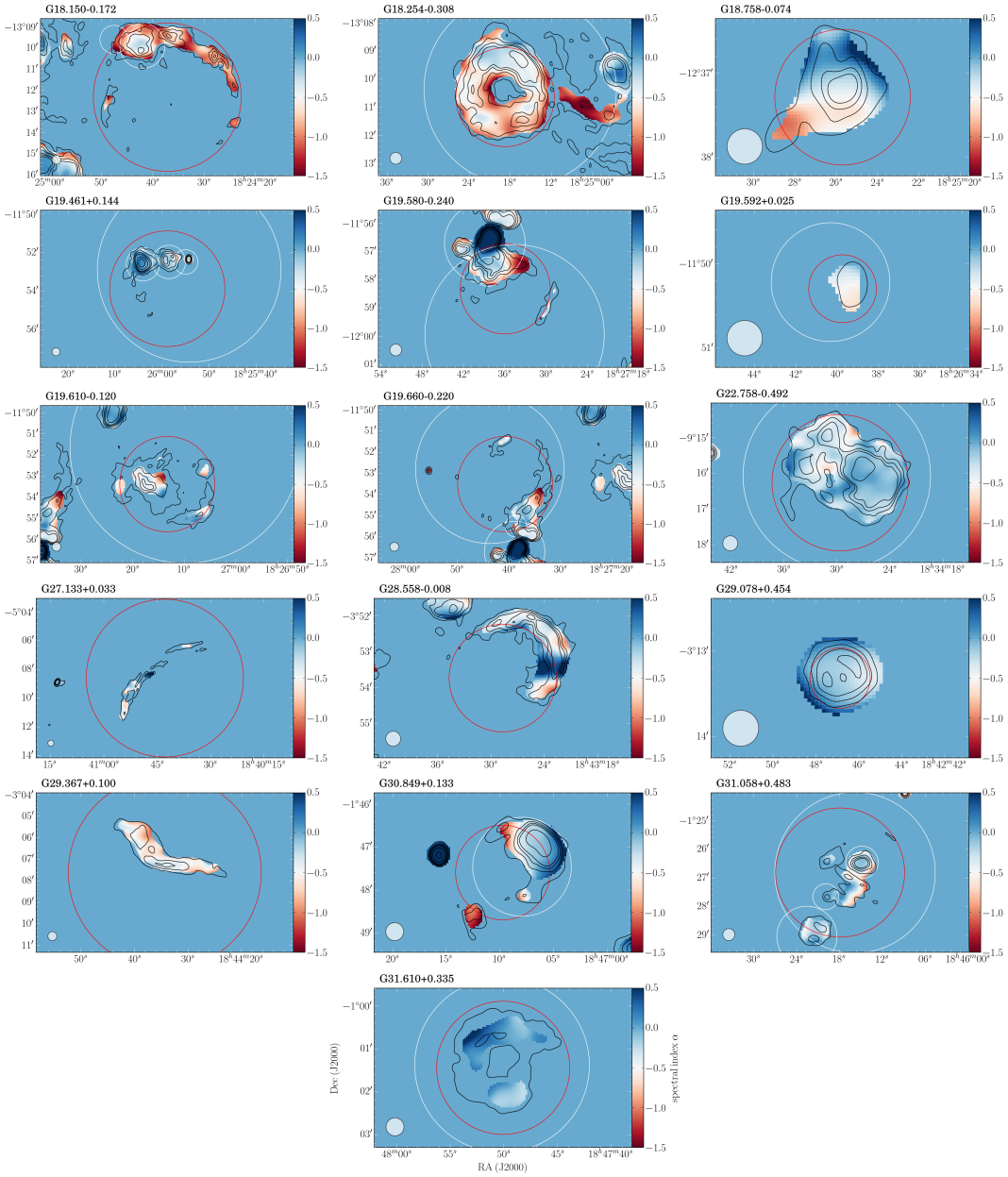

Figure 3.23: Spectral index maps of extended MAGPIS SNR candidates. The color scale represents the spectral index α from -1.5 to 0.5, the black contours show the continuum emission in steps of 5, 10, 15, 20, 40, and 100 mJy beam^{-1}. The red and white circles indicate the sizes of the SNR candidates given in Helfand et al. (2006) and the sizes of the H II regions given in Anderson et al. (2014), respectively. In each panel the synthesized beam is shown in the bottom left corner.

As explained, we can use the spectral index to distinguish between thermal and non-thermal radiation. We present spectral index maps for 16 extended MAGPIS SNR candidates in Fig. 3.23. Similar to the example of the well known SNR G018.8+00.3 (Fig. 3.22), we find strongly varying spectral index maps. On the one hand, several sources clearly indicate a flat spectral index (e.g., G31.610+0.335), which is characteristic for an H II region. On the other hand, several sources (e.g., G18.150-0.172) show clear signatures of a negative spectral index around -0.5, which is typical of SNR. However, for some sources the classification as thermal or non-thermal radiation is difficult as the spectral index shows both positive and negative values. The source G19.580-0.240 is a good example for such a behavior. This can be explained by several different sources along the line of sight. Our goal is to use this information to classify sources as potential SNR or as H II regions. As explained in Sect. 3.3.5, we have to be cautious with the spectral index for extended sources due to different filtering at different wavelengths. The spectral index maps of the MAGPIS SNR candidates G18.150-0.172, G18.758-0.074, G27.133+0.033, G28.558-0.008, G29.367+0.100 show negative values, and they are not directly correlated with any known H II region. Therefore, these five sources are excellent candidates for SNR. However, only one of them (G18.150-0.172) is listed in the SNR catalog by Green (2014). The morphology of the five proposed SNR differs widely. We find two examples of shell-type SNR (G18.150-0.172, G28.558-0.008), one small filled-center (G18.758-0.074) and two that may be part of a larger shell or a more filamentary SNR (G27.133+0.033, G29.367+0.100).

Beside these proposed SNR without any correlation to known H II regions, we find one source (G18.254-0.308) that is a well known H II region (Anderson et al. 2014) showing the same morphology in the infrared, but the spectral index map shows mostly negative values down to $\alpha = -1$. This is an indicator of non-thermal radiation, which contradicts the expected radiation coming from an H II region. Owing to the different spatial filtering in our data (see Sect. 3.4.7), we cannot exclude a systematic shift of the spectral index. However, the source is strong ($\sim 30\sigma$) and not very extended, which minimizes the filtering effects. Further analysis will be needed to reveal the nature of this source.

3.6.8 Special source G48.384+0.789

In this section, we introduce a special source, which has a remarkably high positive spectral index. The THOR source G48.384+0.789 is unresolved and bright (30-100 mJy beam^{-1}) and shows a positive spectral index of $\alpha = 1.70 \pm 0.02$. Since this source is unresolved, the spectral index determination does not suffer from spatial filtering due to the VLA C-array configuration and is well constrained. Figure 3.24 shows the flux density measurements for each spectral window, and the corresponding spectral index fit. This source has a counterpart

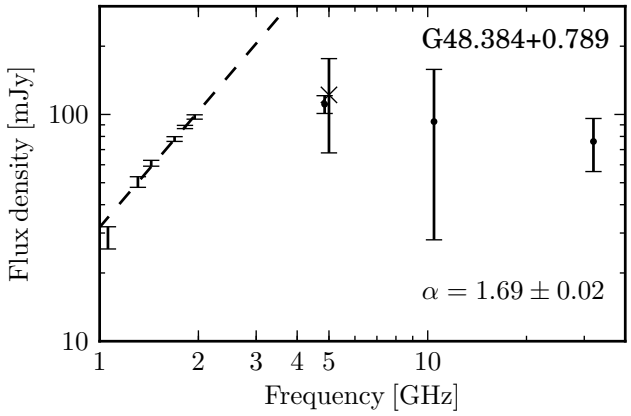

Figure 3.24: Spectrum of the THOR source G48.384+0.789. The data points between 1 and 2 GHz are taken from the THOR survey, and the dashed line represents the fitted spectral index to these data points of $\alpha = 1.69 \pm 0.02$. The data point at 5 GHz (x-shape) is taken from the CORNISH survey (Purcell et al. 2013), and the data points at 4.85, 10.45, and 32 GHz (circles) are taken from Vollmer et al. (2008). For each data point, a 5σ uncertainty is shown.

in CORNISH (G048.3841+00.7889, Purcell et al. 2013) at 5 GHz. The corresponding flux density measurement at 5 GHz is also given in Fig. 3.24, but we do not consider this data point for the spectral index determination. Within CORNISH, this source is classified as "IR-quiet" and even with the high resolution of CORNISH of 1.5", this source is unresolved. Further observations at 4.85, 10.45, and 32 GHz using the Effelsberg 100m telescope presented by Vollmer et al. (2008) show a flat spectrum for higher frequencies (see Fig. 3.24). We do not find any counterpart in sub-mm emission (ATLASGAL) or in CO emission (GRS). However, Very Long Baseline Array (VLBA) observations presented by Immer et al. (2011) reveal a detection, and they classified this source as compact ("compactness B"). This does not translate trivially to an actual size of the object because VLBA observations suffer from severe filtering effects. But this detection shows that the object has a very compact component typical of extragalactic sources, such as AGNs. The spectral index could also be explained by an AGN as similar spectral shapes are found in the literature (e.g., Brunthaler et al. 2005).

3.7 Conclusions

The THOR survey, which is the H I, OH, recombination line survey of the Milky Way is a Galactic plane survey covering a large portion of the first Galactic quadrant ($l = 14 - 67°$, $|b| \leq 1.1°$) using the VLA in the C-Array configuration, achieving a spatial resolution of ~10-25". In this paper, we present a catalog of continuum sources within the first half of the survey

($l = 14.0 - 37.9°$ and $l = 47.1 - 51.2°$, $|b| \leq 1.1°$). We summarize our work and the main results below.

1. With the BLOBCAT extraction software, we extracted 4422 sources. We used a spatially varying noise map, as well as automated RFI flagging methods (RFlag) to decrease the number of false detections. Furthermore, we visually inspected each source for obvious artifacts. About 75% (3366 sources) of the extracted sources are reliable detections above 7σ. The catalog is complete up to 95% above the 7σ detection limit, whereas the noise is spatially varying. Half of the observed area has a noise level of $7\sigma < 3\,\mathrm{mJy\,beam^{-1}}$.
2. We cross-matched the THOR catalog with the NVSS, MAGPIS, and CORNISH catalogs to validate the position and flux density. The position comparison with MAGPIS and CORNISH reveals no significant shift, and we reported a position uncertainty that depends on the strength of the source but is smaller than 2". The flux density and peak intensity comparison with MAGPIS shows a one-to-one relation; however, we find a slight bias in comparison with the NVSS survey.
3. Thanks to the broad bandpass between 1 and 2 GHz, we were able to determine reliable spectral indices for 1840 sources. We extracted the peak intensity of six different spectral windows and used a linear fit to describe the spectral index α with the form $I(\nu) \propto \nu^{\alpha}$. The spectral index distributions reveals two peaks at $\alpha = -1$ and $\alpha = 0$, highlighting two groups of sources, which are dominated by thermal and non-thermal radiation, respectively.
4. We used the spectral index information to investigate the spectrum of H II regions. We cross-matched the THOR catalog with the WISE H II region catalog and found an overlap of 388 sources. For about 326 of these sources, we were able to determine a reliable spectral index. The distribution reveals a single peak around $\alpha = 0$, indicating thermal free-free emission.
5. The spectral index can also be used to confirm potential SNR candidates because they exhibit a typical spectral index of $\alpha = -0.5$. We investigated the MAGPIS SNR candidate catalog and determined spectral index maps for 16 SNR candidates. Owing to potential line-of-sight contamination with H II regions, the differentiation between thermal and non-thermal radiation is difficult. However, we confirmed five SNR candidates, showing non-thermal radiation and no correlation with H II regions. Four of them are not listed in the SNR catalog presented by Green (2014).

4

Studying atomic hydrogen during cloud formation by means of HISAs: Kinematics and probability distribution functions

4.1 Abstract

Probability density functions (PDFs) of the column density or volume density of hydrogen are a common tool to examine molecular clouds. Due to turbulent motion, the initial PDFs have a log-normal shape and evolve into a power-law tail at high column densities due to collapse and gravitational forces. To date, these studies are mostly limited to the molecular content of the cloud. In this chapter, we will study the cold atomic content of the giant molecular filament GMF38.1-32.4, calculate column density PDFs and study the corresponding kinematics. We extracted a long H I self absorption (HISA) feature, which correlates partly with CO emission. The peak velocity of the HISA and CO shows a close correlation on one side of the filament, whereas a velocity step is visible on the other side. The column density of the cold absorbing H I is on the order of 10^{20}-10^{21} cm^{-2}. In contrast to this, the column density of the molecular hydrogen, traced with CO, is an order of magnitude higher. The shape of the atomic and molecular column density PDF reveal mostly log-normal shapes, indicating turbulent motion as the main driver. This interpretation is supported by the measured linewidth of Δv_{FWHM} = 6-8 km s^{-1} for the HISA. We speculate that we observe different evolutionary stages within the filament. The left sub-region seems to be forming a molecular cloud out of the atmoic environment, whereas the right sub-region already shows high column density peaks and active star formation. Such studies are an important characterization of the transition between the atomic and molecular phase and influence simulations as well as theoretical studies.

4.2 Introduction

With the high sensitivity and resolution of the Herschel space telescope, filamentary structures at different scales were found everywhere within our Milky Way (e.g., André et al. 2014). These filaments play an important role for star formation as they provide a channel to gather mass. In numerical simulations, filaments are created by many processes, such as turbulent motions, converging flows or colliding sheets. The longest of these structures are molecular filaments, such as the famous 'Nessie' cloud (Jackson et al. 2010). This cloud is situated parallel to the Galactic plane within a spiral arm. Hence, it is also called a 'bone of the Milky Way' (Goodman et al. 2014). The length of Nessie is still under debate and between 80 and ~400 pc long (Goodman et al. 2014).

To study large filaments similar to 'Nessie' in a statistical manner, Ragan et al. (2014) compiled a catalog of long molecular filaments. They used near-infrared (NIR) and mid-infrared (MIR) observations from the Spitzer Galactic Plane survey (GLIMPSE, Benjamin et al. 2003) to detect long (>1°) elongated absorption features. This 'by-eye' search revealed a sample of 12 long absorbing filaments. To confirm that these absorbing features belong to the same structure, they examined the velocity structure, using CO emission from the Galactic Ring Survey (GRS, Jackson et al. 2006). This resulted in a sample of seven velocity coherent filaments called 'giant molecular filament' (GMF). Their length and mass ranges from 60 to 200 pc and 0.5 to 7×10^5 $M_\odot$, respectively. In contrast to the 'Nessie' cloud, the giant molecular filaments presented by Ragan et al. (2014) are not just confined to Galactic spiral arms, but are also located in the inter-arm region. Hence, some of them might be so called spurs, which are observed in nearby galaxies (e.g., in M51, Schinnerer et al. 2013). However, this explanation has to be treated cautiously due to large uncertainties for the model of the Galactic spiral arms. Another survey of large filaments is presented by Wang et al. (2015), employing Herschel data.

A common tool to study molecular clouds are the probability distribution functions (PDFs) of the column density or the volume density (e.g., Kainulainen et al. 2009, 2014; Sadavoy et al. 2014; Stutz & Kainulainen 2015; Schneider et al. 2015a). The observed shape of these PDFs depends on the physical processes acting within the cloud. In the early evolution of a molecular cloud, the PDF reveals a log-normal shape which is due to turbulent motions within the cloud and determines the width of the log-normal shape. More evolved clouds develop a high-density power-law tail, which indicates that gravity becomes the dominant driver (Kainulainen et al. 2009). These evolved clouds can be related to evolutionary stages by measuring the slope of their power-law, whereas shallower slopes might indicate earlier evolutionary stages (Stutz & Kainulainen 2015). High-mass star-forming regions reveal even multiple power-laws, having a shallower slope for the highest density regions. This indicates a slower collapse for such re-

gions (Schneider et al. 2015a). Lombardi et al. (2015) presents a controversial work, reporting that all PDFs have a power-law shape and the log-normal shape could be an observational bias. Theoretical work and simulations of molecular clouds also reproduce PDFs in different forms (e.g., Vazquez-Semadeni 1994; Federrath et al. 2010). Numerical simulations and analytic calculations also show the evolution of the column density PDF from a pure log-normal shape for early stages and the development of a power-law tail for more evolved regions (Ballesteros-Paredes et al. 2011; Girichidis et al. 2014). These studies are mainly focused on the molecular hydrogen for the star-forming clouds and to our knowledge only one publication determined column density PDFs for the atomic content. Burkhart et al. (2015) presents a study of H I observations of the Perseus molecular cloud. They used data from the GALFA-H I survey (Peek et al. 2011), which was observed with the Arecibo Observatory. The H I column density was calculated by correcting the H I emission for optical depth effects (Lee et al. 2012, 2015), similar to the data presented in Chapter 2. The presented H I column density PDF shows a pure log-normal shape, without any power-law tail (Burkhart et al. 2015).

Even though the 21cm H I line offers a direct measurement to study the atomic hydrogen, it is difficult to measure its properties. The main challenge is the coexistence of the warm neutral medium (WNM) and the cold neutral medium (CNM) in pressure equilibrium side by side (Wolfire et al. 2003). However, studies of H I self absorption (HISA) overcome this problem, as they solely trace the CNM. HISA features occur if cold, dense H I is in front of an emission background. Within the Milky Way, HISA features are found basically everywhere (Gibson et al. 2005a; Kavars et al. 2005). Their spin temperature ranges from ~10-50 K, (e.g. Gibson et al. 2000; Kavars et al. 2005; McClure-Griffiths et al. 2006). A special case of the HISA features are so called H I narrow self absorption (HINSA) features, revealing small linewidths on the order of $\sim 1\,\mathrm{km\,s^{-1}}$ (Li & Goldsmith 2003; Krčo & Goldsmith 2010).

To investigate the transition of atomic to molecular hydrogen in more detail, we initiated a project to examine the hydrogen content in detail of one giant molecular filament found by Ragan et al. (2014). The goal is to study the kinematics of one long filament in the molecular and atomic hydrogen traced by CO emission and H I self absorption, respectively. Furthermore, we will analyze column density PDFs for the atomic and molecular hydrogen and compare their properties.

4.3 Observational Methods

4.3.1 H I self absorption

In Chapter 2 we examined the H I emission of the star forming complex W43 and corrected for optical depth effects by measuring the H I optical depth towards strong continuum sources.

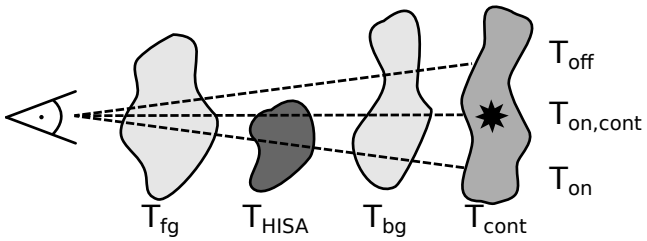

Figure 4.1: A sketch of the arrangement of the different components. The cold absorbing cloud (HISA) is surrounded by emitting clouds with temperature $T_{\rm fg}$ and $T_{\rm bg}$. Behind the H I clouds, several continuum sources can be situated, either diffuse or discrete.

As explained, this is the H I continuum absorption (HICA) method. Beside the absorption of background continuum sources, we can also measure the absorption of cold H I in front of an emitting H I cloud. This is called the H I self absorption (HISA). The terminology 'H I self absorption' can be misleading, as the emission and absorption processes can occur in the same cloud, but must not. It is possible, that the H I emission stems from a distant background cloud, which has the same velocity with respect to the local standard of rest $v_{\rm LSR}$ than the absorbing foreground cloud. This is feasible due to the rotation curve of the Milky Way.

Since we observe several different components along the line of sight, the equations become more complicated than the simple HICA description given in Chapter 2 and shown in Fig. 2.2. A comprehensive discussion about the radiative transfer of HISA features can be found in Gibson et al. (2000), Kavars et al. (2003) and Li & Goldsmith (2003). Figure 4.1 illustrates the discussed arrangement of clouds. In general, we observe an emitting foreground and background H I cloud, which have spin temperatures $T_{\rm fg}$ and $T_{\rm bg}$, respectively. Between these two emitting clouds, a cold absorbing H I cloud can be situated, having the spin temperature $T_{\rm HISA}$. Furthermore, we observe 1.4 GHz continuum emission, which can be a diffuse Galactic component or discrete strong sources. These continuum sources were discussed in detail in Chapter 3. For simplicity, we assume that the continuum emission is situated in the background. However, this assumption will not be fulfilled generally. In the following section, we will exclude the presence of strong, discrete continuum sources and consider solely a weak diffuse continuum background. In Sec. 4.5.2, we will utilize strong continuum sources to determine the optical depth. Following Eq. 1.9, we measure the on and off position as:

$$\begin{aligned} T_{\rm off} &= T_{\rm fg}(1-e^{-\tau_{\rm fg}}) + T_{\rm bg}(1-e^{-\tau_{\rm bg}})e^{-\tau_{\rm fg}} + T_{\rm cont}e^{-(\tau_{\rm fg}+\tau_{\rm bg})} - T_{\rm cont} \\ T_{\rm on} &= T_{\rm fg}(1-e^{-\tau_{\rm fg}}) + T_{\rm HISA}(1-e^{-\tau_{\rm HISA}})e^{-\tau_{\rm fg}} + \\ &\quad T_{\rm bg}(1-e^{-\tau_{\rm bg}})e^{-(\tau_{\rm fg}+\tau_{\rm HISA})} + T_{\rm cont}e^{-(\tau_{\rm fg}+\tau_{\rm HISA}+\tau_{\rm bg})} - T_{\rm cont}, \end{aligned} \tag{4.1}$$

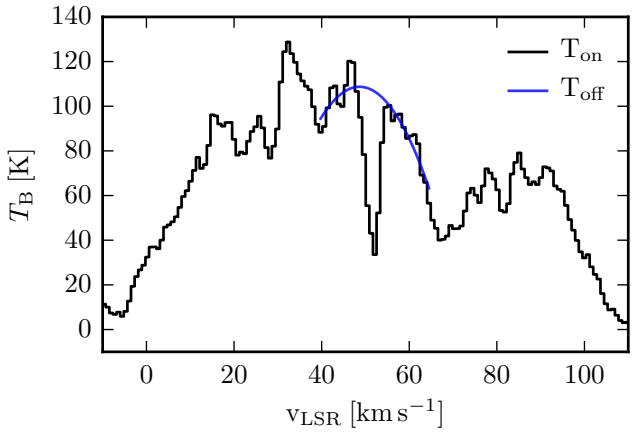

Figure 4.2: Example spectrum showing a prominent HISA around $v_{\mathrm{LSR}} \sim 55\,\mathrm{km\,s^{-1}}$. The actual H I spectra is shown in black (T_{on}) and the estimated background emission using a second order polynomial (see Sect. 4.4.2) is shown in blue (T_{off}).

where τ_{fg}, τ_{bg}, τ_{HISA} are the corresponding optical depths of each component shown in Fig. 4.1 and T_{cont} is the continuum brightness temperature. During the data reduction, we subtract the continuum emission from the H I data, which is indicated by the last factor ($-T_{\mathrm{cont}}$). An example spectrum illustrating T_{on} and T_{off} is shown in Fig.4.2. Calculating the difference between the on and off position, we find:

$$\begin{aligned} T_{\mathrm{on}} - T_{\mathrm{off}} &= T_{\mathrm{HISA}}(1 - e^{-\tau_{\mathrm{HISA}}})e^{-\tau_{\mathrm{fg}}} - T_{\mathrm{bg}}(1 - e^{-\tau_{\mathrm{bg}}})e^{-\tau_{\mathrm{fg}}}(1 - e^{-\tau_{\mathrm{HISA}}}) \\ &\quad - T_{\mathrm{cont}}(1 - e^{-\tau_{\mathrm{HISA}}})e^{-(\tau_{\mathrm{fg}}+\tau_{\mathrm{bg}})} \\ &= \left(T_{\mathrm{HISA}} - T_{\mathrm{bg}}(1 - e^{-\tau_{\mathrm{bg}}}) - T_{\mathrm{cont}}e^{-\tau_{\mathrm{bg}}}\right) \times (1 - e^{-\tau_{\mathrm{HISA}}})e^{-\tau_{\mathrm{fg}}} \end{aligned} \tag{4.2}$$

This equation can be further simplified by introducing the dimensionless parameter p (e.g., Feldt 1993; Gibson et al. 2000):

$$p \equiv \frac{T_{\mathrm{bg}}(1 - e^{-\tau_{\mathrm{bg}}})}{T_{\mathrm{off}}}. \tag{4.3}$$

As a last simplification, we assume that the foreground and background clouds are optically thin and therefore τ_{fg} and τ_{bg} are small. This results in:

$$T_{\mathrm{on}} - T_{\mathrm{off}} = (T_{\mathrm{HISA}} - p\,T_{\mathrm{off}} - T_{\mathrm{cont}}) \times (1 - e^{-\tau_{\mathrm{HISA}}}). \tag{4.4}$$

For the simplification of small τ_{fg} and τ_{bg}, the dimensionless parameter p (Eq. 4.3) describes the fraction of the H I background emission. That means for $p = 1$, there is no foreground emission and for $p = 0.5$, the amount of foreground and background emission is equal. Measuring this parameter is difficult and we usually have to assume a certain value. The corresponding uncertainty will be discussed in more detail in Sect.4.5.5.

Equation 4.4 connects the properties of the HISA, such as the cloud spin temperature T_{HISA}

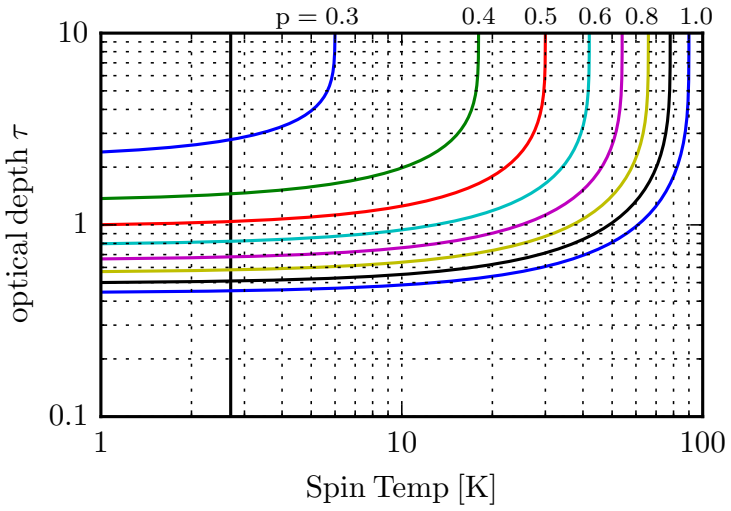

Figure 4.3: The optical depth τ as a function of the spin temperature calculated using Eq. 4.4 and assuming typical values for an HISA feature ($T_{\rm off}$ = 120 K, $T_{\rm on}$ = 70 K, $T_{\rm cont}$ = 20 K). The different color represent different values of p (Eq. 4.3) from 0.3 to 1. The black vertical line indicates the temperature of the cosmic microwave background of 2.7 K, which is a lower limit for the spin temperature.

and the optical depth $\tau_{\rm HISA}$, with the observable quantities $T_{\rm on}$, $T_{\rm off}$ and $T_{\rm cont}$. However, this equation alone does not allow us to disentangle the spin temperature and the optical depth and further assumptions or measurements are necessary.

To study this relation in more detail, we can plot the optical depth as a function of the spin temperature which is shown in Figure 4.3 for assumed values of $T_{\rm off}$ = 120 K, $T_{\rm on}$ = 70 K and $T_{\rm cont}$ = 20 K. These are realistic values, which can be found for H I self absorption features (e.g., Gibson et al. 2000). The different colors represent different values of p from 0.3 to 1. The black vertical line indicates a lower limit of the spin temperature of T = 2.7 K, which is the temperature of the cosmic microwave background. The general interpretation of the curves is that a higher optical depth is necessary to induce the assumed absorption feature for higher spin temperatures. This dependence becomes very steep at a certain point, depending on p.

As we need a high brightness sensitivity to extract the HISA feature, we use the VGPS data Stil et al. (2006) in the following (see Sect. 1.4.5 for further details). This data has an angular resolution of $\Theta = 60''$ and a velocity resolution of $\Delta v = 0.82\,{\rm km\,s^{-1}}$.

4.3.2 H_2 mass estimate via ^{13}CO observations

We want to compare the cold H I component with the molecular content of the filament. However, cold molecular hydrogen is not directly observable and we have to use tracers or indirect measurements to estimate it. In Chap. 2 we used dust emission measured with the Herschel

telescope to estimate the molecular content. In this chapter, we are using the CO molecule to trace the molecular content, because we are not only interested in the column density, but also in the kinematics of the molecular filament. Furthermore, with the CO data we can directly associated the molecular and atomic hydrogen kinematically and hence avoid contributions from other spiral arms. Dust emission measurements do not provide these kinematical information. As a molecular tracer, we use the ^{13}CO, $J = 1 \rightarrow 0$ data from the Galactic Ring Survey (GRS, Jackson et al. 2006). This data has an angular resolution of $\Theta = 44''$ and a velocity resolution of $\Delta v = 0.21\,\mathrm{km\,s^{-1}}$, which is superior to the used H I data and we have to smooth the CO data accordingly to $\Theta = 60''$ and $\Delta v = 0.82\,\mathrm{km\,s^{-1}}$.
To estimate the column density of the CO molecule, we assume that the ^{13}CO, $J = 1 \rightarrow 0$ emission is optically thin and use the equation (Wilson et al. 2010):

$$N(^{13}\mathrm{CO}) = 3.0 \times 10^{14}\ \frac{\int \mathrm{T_{MB}(v)\,dv}}{1 - \exp(-5.3/\mathrm{T_{ex}})}, \tag{4.5}$$

where $N(^{13}\mathrm{CO})$ is the column density of the ^{13}CO in units of $\mathrm{cm^{-2}}$, dv is the velocity in $\mathrm{km\,s^{-1}}$ and T_{MB} is the main beam brightness temperature and T_{ex} is the excitation temperature, both in K. We do not have a direct measurement for the excitation temperature, but we can assume that excitation temperature from the ^{12}CO and ^{13}CO emission is equal. Roman-Duval et al. (2010) derived the ^{12}CO excitation temperature for 580 molecular clouds within the GRS survey assuming that the ^{12}CO line is optically thick. The determined mean value for the excitation temperature is ~6 K. We will use this value to calculate the column density. Using larger values of T_{ex} will increase the column density uniformly, but not change the structure. To obtain finally the H_2 column density, we have to follow two steps: First estimate the CO column density from the ^{13}CO column density and then estimate the H_2 column density from the CO column density. To estimate the ratio of CO to ^{13}CO we use Eq. 3 in Milam et al. (2005), which depends on the distance from the Galactic center R_{gal}. For the Galactocentric distance of 6 kpc for GMF38.1-32.4, this results in a ratio of $N(\mathrm{CO})/\mathrm{N}(^{13}\mathrm{CO}) \sim 51$. Following the original calculations about the GMF38.1-32.4 presented in Ragan et al. (2014), we use a constant ratio of $N(\mathrm{CO})/\mathrm{N}(\mathrm{H_2}) = 1.1 \times 10^{-4}$ (Pineda et al. 2010). This results in a final ratio of $N(\mathrm{H_2})/\mathrm{N}(^{13}\mathrm{CO}) \sim 4.6 \times 10^5$. Since these ratios have large uncertainties, the calculated molecular column density has an uncertainty of at least a factor of two. On top of that we have uncertainties due to the assumption of optical thin emission and the assumption of the excitation temperature. It is also known that CO does not trace all H_2, which is called 'CO-dark-H_2'. Simulations and observations showed that this can make up a significant fraction up to 42% (Pineda et al. 2013; Smith et al. 2014).

4.4 Results

4.4.1 Giant molecular filament GMF38.1-32.4

The giant molecular filament GMF38.1-32.4 is the longest and most massive filament in the sample presented by Ragan et al. (2014). Its angular length is 3.8°, which reveals a physical length of ~230 pc at a distance of ~3.5 kpc (Ragan et al. 2014). The coherent velocity range is from $v_{\mathrm{LSR}} = 50 - 60\,\mathrm{km\,s^{-1}}$. The top panel of Fig. 4.4 shows the integrated ^{13}CO emission from the GRS survey (Jackson et al. 2006). The CO emission covers almost the entire length of the filament continuously, whereas the main emission peaks are situated on the right side. The integrated H I emission is shown in the middle panel of Fig. 4.4 and reveals a more diffuse emission covering a larger area than the CO emission. The strongest H I emission does not coincide with CO emission, but almost an anti correlation of the H I and CO emission can be observed. This is due to the H I self absorption, which we will study in detail in Sect. 4.4.3. To understand the H I emission and absorption, it is crucial to consider the 1.4 GHz continuum emission, which is shown in the lower panel of Sect. 4.4.3. Besides diffuse continuum emission around the Galactic plane, several strong sources are apparent. The largest source is the SNR G34.7-0.4, also known as W44. Another strong continuum source is close to the CO emission peak and is an UCH II region. Within the THOR catalog it has the name G34.256+0.146 (see Chapter 3, Bihr et al. 2016). We will use this source to measure the H I optical depth in Sect. 4.5.2. In the central part of the filament around $l = 35.6°$ the diffuse continuum emission is high ($T_{\mathrm{cont}} \sim 30$ K), which makes the analysis of the H I line difficult.

4.4.2 Background estimate to measure T_{off}

To extract a reliable HISA feature, it is crucial to know the background emission. Different methods can be found in the literature to perform this task. The first method is to use absorption-free H I emission spectra, located close to the absorption feature (e.g., Gibson et al. 2000), called 'off-position'. For this method, we have to assume that the H I background emission stays spatially constant over the absorption feature, which might be fulfilled for spatially small HISA features. Even though we are studying a large HISA feature, spanning several degrees on the sky, we tested this method and extracted H I spectra from five different regions, which are labeled as 'Off 1' to 'Off 5' in Fig. 4.5. These off-positions were chosen as regions without significant 1.4 GHz continuum emission and without CO emission. Furthermore, these regions did not show significant self absorption features at the velocity range of $v_{\mathrm{LSR}} = 50 - 60\,\mathrm{km\,s^{-1}}$. The corresponding spectra are presented in Fig. 4.6. These spectra reveal large variations, which makes them difficult to use as a general off position.

The second method utilizes a fit of H I spectra to get T_{off}, excluding the HISA feature for the

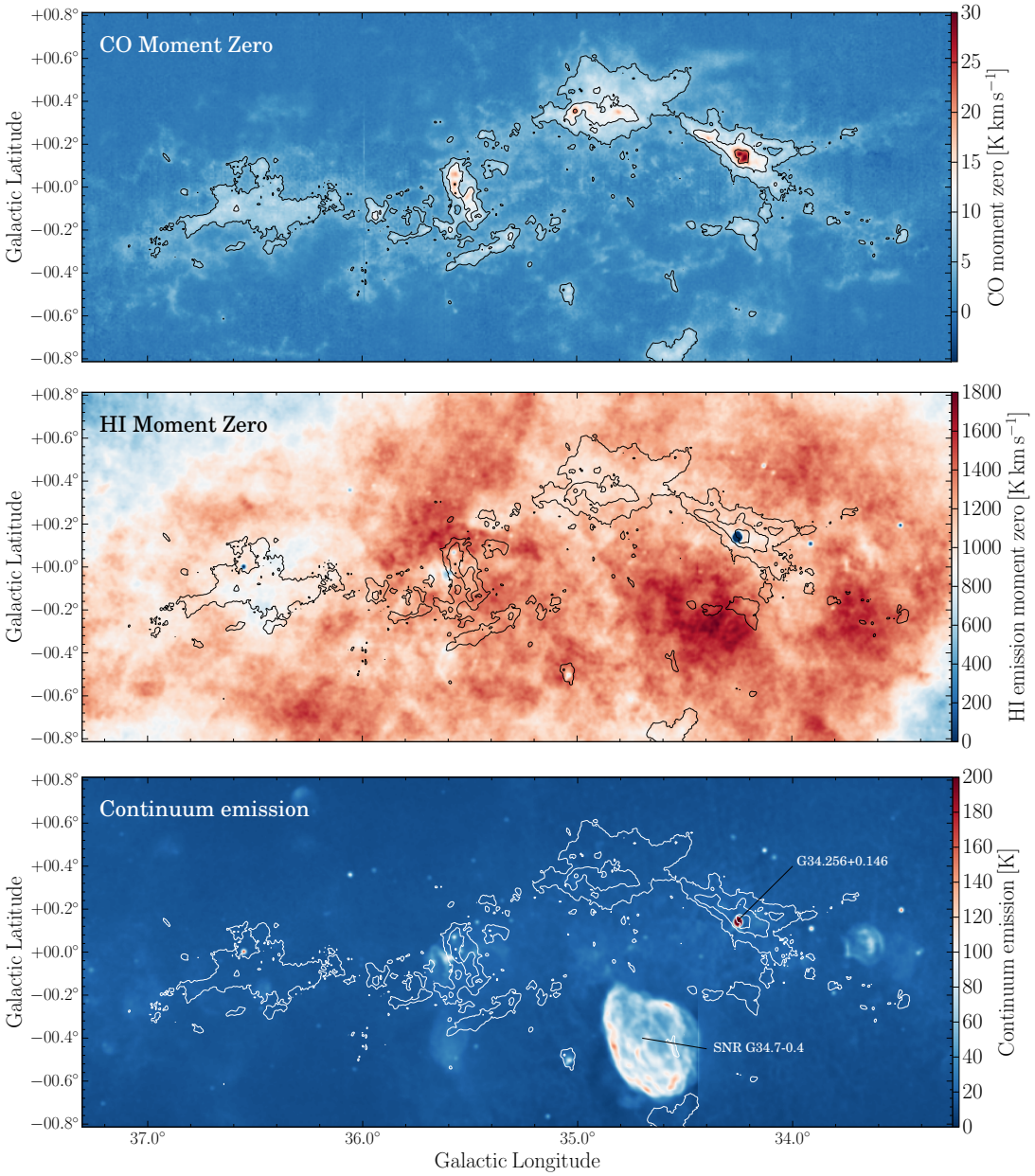

Figure 4.4: The overview of the giant molecular filament GMF38.1-32.4. The top panel shows the integrated ^{13}CO emission from the GRS survey (Jackson et al. 2006) from $v_{\mathrm{LSR}} = 50 - 60\,\mathrm{km\,s^{-1}}$. The integrated H I emission for the same velocity range is shown in the middle panel and taken from the VGPS survey (Stil et al. 2006). The corresponding 1.4 GHz continuum emission from the VGPS survey is shown in the bottom panel. The contours in each panel indicate the integrated CO emission at levels of 5, 10 and 20 K km s^{-1}.

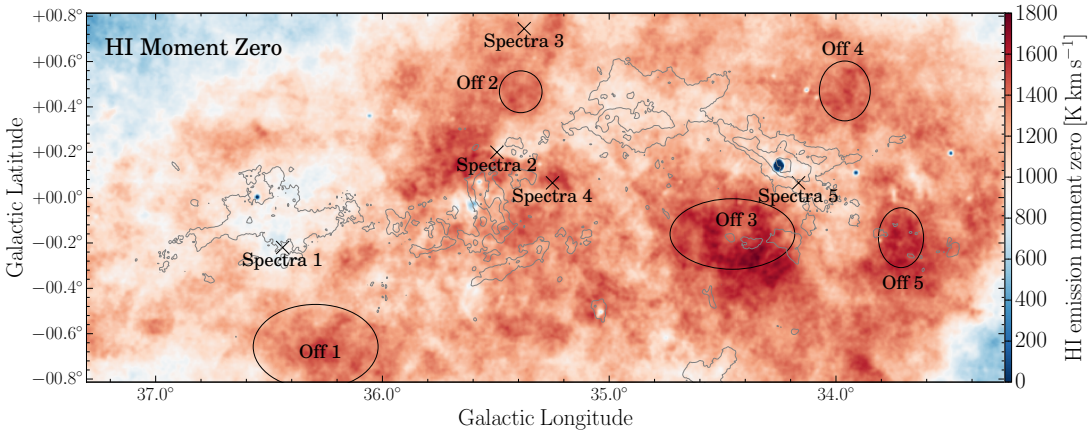

Figure 4.5: The H I integrated emission (moment zero) is presented in color and the gray contours indicate the integrated CO emission at levels of 5, 10 and 20 K km s^{-1}. The black ellipses show the position of the 'off-position' shown in Fig. 4.6 and the black 'X' show the position of the spectra shown in Fig. 4.7.

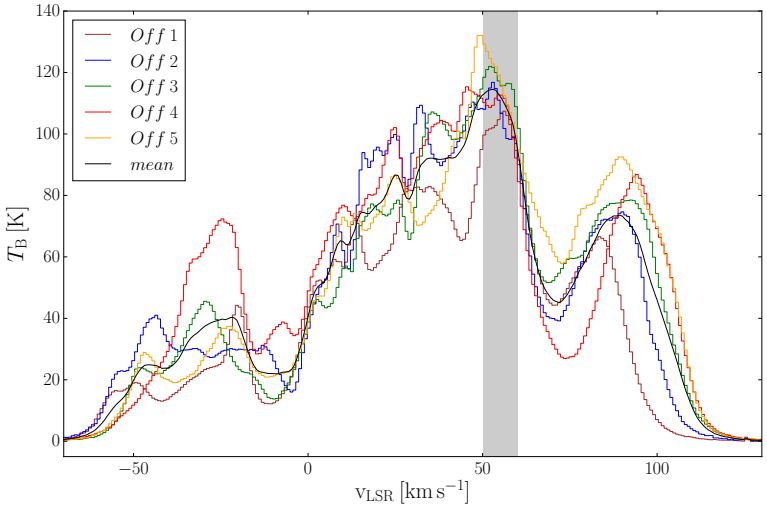

Figure 4.6: Different H I emission spectra around the GMF38.1-32.4, which can be used as 'off-positions'. The area, which we used for the extraction is shown in Fig. 4.5. The black line shows the mean spectra of all five off-spectra and the gray shaded area indicates the velocity range of the HISA feature (v_{LSR} = 50 – 60 km s^{-1}).

actual fit and interpolate for the absorption feature. This method is used frequently, using different functions to fit the H I emission, such as linear fits (e.g., Kavars et al. 2003; McClure-Griffiths et al. 2006) or polynomials with different order (e.g. Li & Goldsmith 2003). We use this method as well, testing different fit functions. Fig. 4.7 presents five different spectra from different positions indicated in Fig. 4.5. We used a polynomial with second and fourth order and fitted the spectra for the velocity range around the HISA (v_{LSR} = 40 – 50 and 60 – 70 km s^{-1}).

A polynomial of third order gave very similar results as the polynomial of second order and for clarity we do not show this here.

It is difficult to estimate which function is more suitable to fit the H I spectra. For regions without absorption, we expect that the fitted spectra represents the actual spectra. Spectrum 3 in Fig. 4.7 shows such a region and both functions represent the H I spectra well. Spectra 1, 2 and 5 in Fig. 4.7 represents H I absorption features and the difference between the second and forth order polynomial is small. In contrast to this, Spectrum 4 in Fig. 4.7 reveals a large difference between the fit functions. The fourth order polynomial fit is much higher (~50 K) than the second order polynomial. It is not obvious which function describes the H I spectra more accurate. However, the fourth order polynomial might overestimate the actual spectra as steep slopes within the fitted velocity range 'push' the fitted spectra to high values for the velocity range of the absorption feature. These fitted values reach values of $T_{\rm off} \sim 175$ K, which is about 20 K higher than the maximum values we observe for the original H I emission. In contrast to this, the second order polynomial might underestimate the H I emission for this spectra. Hence, the fourth order polynomial might be an upper limit and the second order polynomial might be a lower limit. We will use both functions in the following analysis to estimate the uncertainty of $T_{\rm off}$.

The mean spectrum of the five off-positions shown in Fig. 4.6 is shown in gray in Fig. 4.7 as well. Depending on the location of the absorption spectra, this mean off-position represents the H I spectra well (e.g., Spectrum 2 in Fig. 4.7), but in general it does not (e.g., Spectrum 1 or 3 in Fig. 4.7). Hence, we will use the fit method to extract the HISA feature rather than using a mean off-spectrum.

4.4.3 HISA extraction

In Sect. 4.4.2, we discussed different methods to estimate $T_{\rm off}$ in Eq. 4.4. Using this information, we can measure the depth of the absorption feature ($T_{\rm off} - T_{\rm on}$). This is shown as a blue and red solid curve in Fig. 4.7. To analyze the absorption features, we use a Gaussian curve to fit them. This allows us to study the exact depth of the absorption features and their kinematics. The peak values of the fitted Gaussian curves are shown in Fig. 4.8 for different $T_{\rm off}$. The absorption depth of the HISA shows values between 30 and 70 K. Several absorption features are visible for all three methods to estimate $T_{\rm off}$, especially a prominent absorption feature peaking at the position of spectrum 2. However, the different methods to estimate $T_{\rm off}$ reveal also large difference for the peak value. The peak values using the spatially averaged off positions (top panel in Fig. 4.8) show unrealistically large values at the edges. As the Galactic H I emission drops at the edge, this method identifies the edge regions as absorption features. Fitting the spectra for each pixel separately solves this problem (middle and bottom panel of

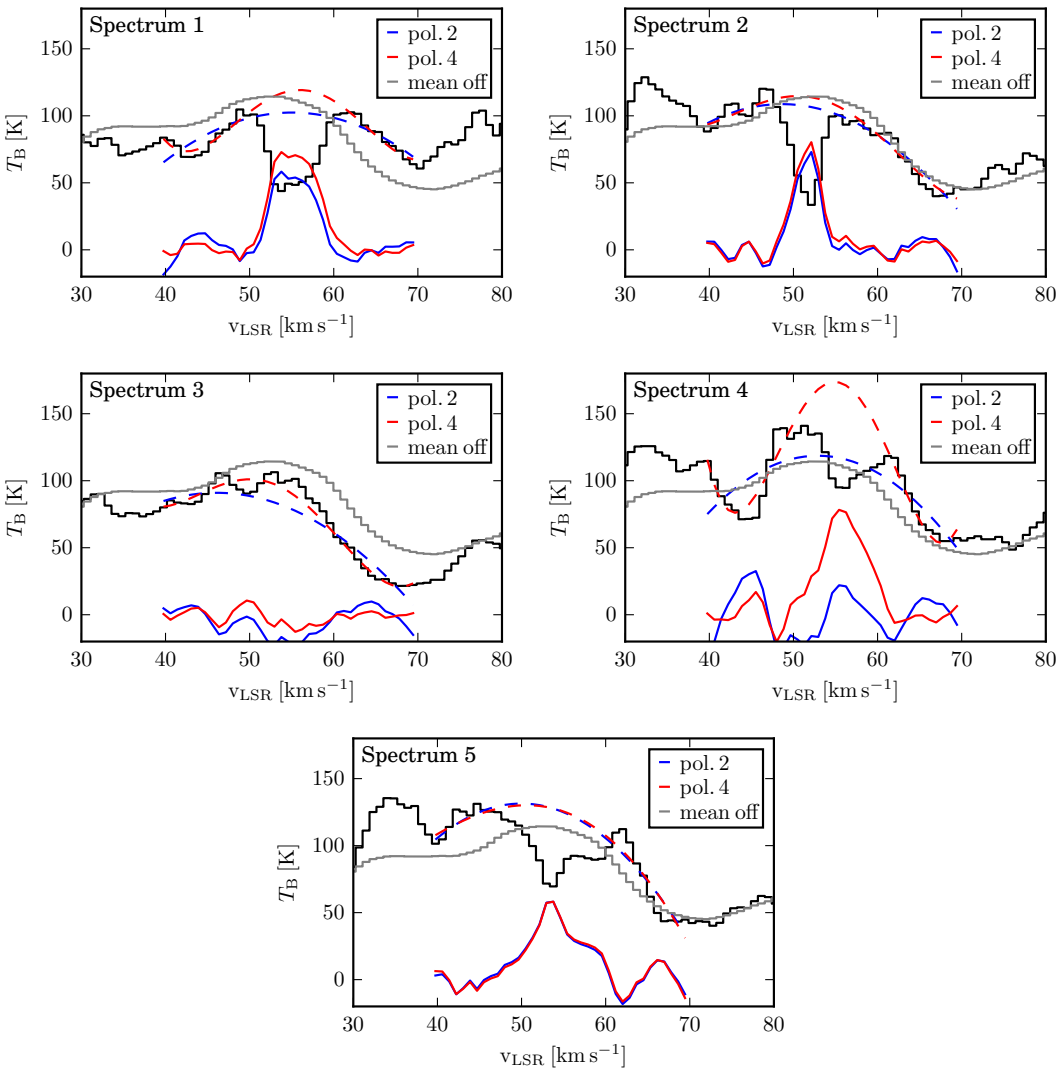

Figure 4.7: Different H I emission spectra are shown in black. The position of each spectrum is shown in Fig. 4.5. The gray spectra indicates the mean of the five off-positions presented in Fig. 4.6. The blue and red dashed lines represent the fits of the HISA spectrum for a polynomial of second and fourth order, respectively, using the velocity range of $v_{LSR} = 40 - 50$ and $60 - 70\,\mathrm{km\,s^{-1}}$ for the fit. The blue and red solid line show the difference between fitted H I spectra and the measured H I spectra ($T_{on} - T_{off}$ in Eq. 4.4) for a polynomial of second and fourth order, respectively.

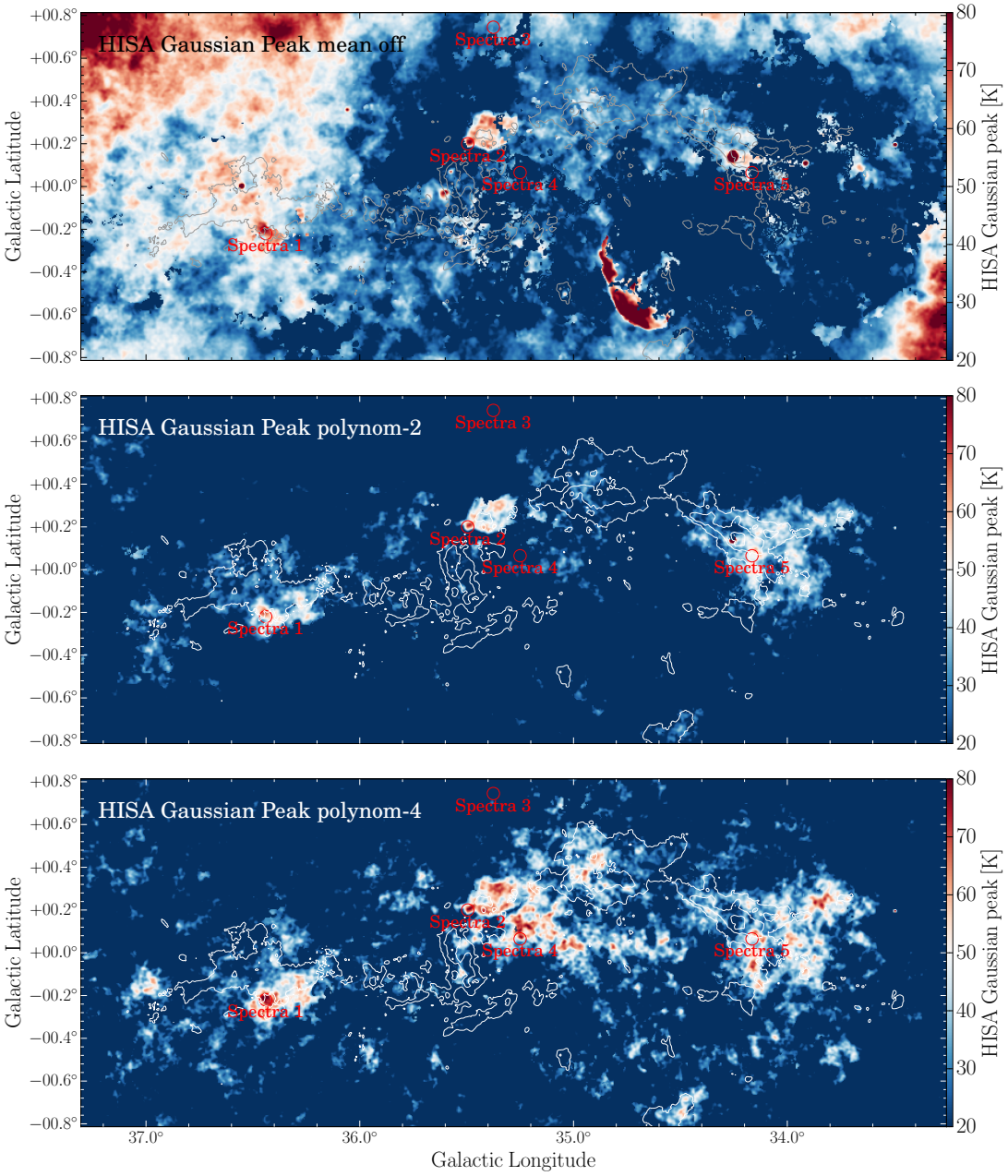

Figure 4.8: In color is shown the peak value of a Gaussian fit for the absorption depth $T_{\rm off} - T_{\rm on}$ for different methods of estimating $T_{\rm off}$. The top panel represents the method using a spatially averaged off spectrum shown in Fig. 4.6, whereas the middle and bottom panel show the method using a polynomial fit with second and forth order, respectively. The white contours represent the integrated CO emission at levels of 5, 10 and 20 K km s^{-1} for reference. The spectra shown in Fig. 4.7 are labeled and marked with red circles.

Fig. 4.8). Another problem is the continuum emission in the background. For the method using the spatially averaged off-positions, the structure of the continuum emission is clearly visible (top panel in Fig. 4.8). Since the continuum emission creates strong absorption in the H I spectra, this is also identified as absorption features using an averaged off-position. This problem is also solved by fitting each H I spectra individually. Hence, the fitting method is more appropriate for our analysis and in the following we will focus on that.

As already mentioned in Sect. 4.4.2, we find for some regions large differences for $T_{\rm off}$ using a polynomial of second or forth order. Fig. 4.8 reveals these regions and we observe the largest differences close to the center of the filament around $l = 35°$. Using a forth order polynomial for the background estimate, we find signifacntly more absorption in this region. However, as already explained, the forth order polynomial might overestimate the background emission and hence also overestimate the absorption depth. In contrast to this, other regions are not affected significantly by the choice of the fit function, e.g., the region around spectra 1 or 2. This will also influence the column density of the absorption feature, which we discuss in Sect. 4.4.5.

4.4.4 Kinematics

In this section, we discuss the kinematical properties of the absorption feature. In contrast to the absorption depth, the peak velocity is not affected significantly by the chosen fit function. Therefore we present here only the velocity structure using a second order polynomial for the determination of $T_{\rm off}$. The velocity structure revealed by a fourth order polynomial is similar. To compare the kinematic of the H I absorption feature with the CO data we use a Gaussian fit for both data sets. The peak velocity map is presented in Fig. 4.9. The CO peak velocity shows a coherent structure along the filament, which was a selection criteria for the catalog by Ragan et al. (2014). A slight velocity step of $\sim$4 km s^{-1} can be seen for the right part of the CO filament around $l = 34°$ having a high velocity of $\sim$57-58 km s^{-1}. For this region, the peak velocity revealed by the HISA feature shows a different velocity of $\sim$54-55 km s^{-1}, which is about 3-4 km s^{-1} lower than the CO velocity. This can also be seen in the right panel of Fig. 4.10, where we show a histogram of the H I and CO peak velocities. In contrast to this, the left side of the filament around $l = 36.5°$ shows a close correlation of the peak velocities, which can also be seen in the left panel of Fig. 4.10. We will discuss this effect in Sect. 4.5.1.

Beside the peak velocity, the Gaussian fit also reveals the linewidth. All linewidth measurements in the following will be given as a full width half maximum (FWHM). The result is shown in Fig. 4.11. The linewidth for the ^{13}CO emission shows extremely high values above 10 km s^{-1} for the central region of the filament around $l = 35°$. However, these values have to be treated cautiously as the CO emission exhibits multiple lines in this region and we use solely a single Gaussian function to fit them. This increases the linewidth and we will discuss

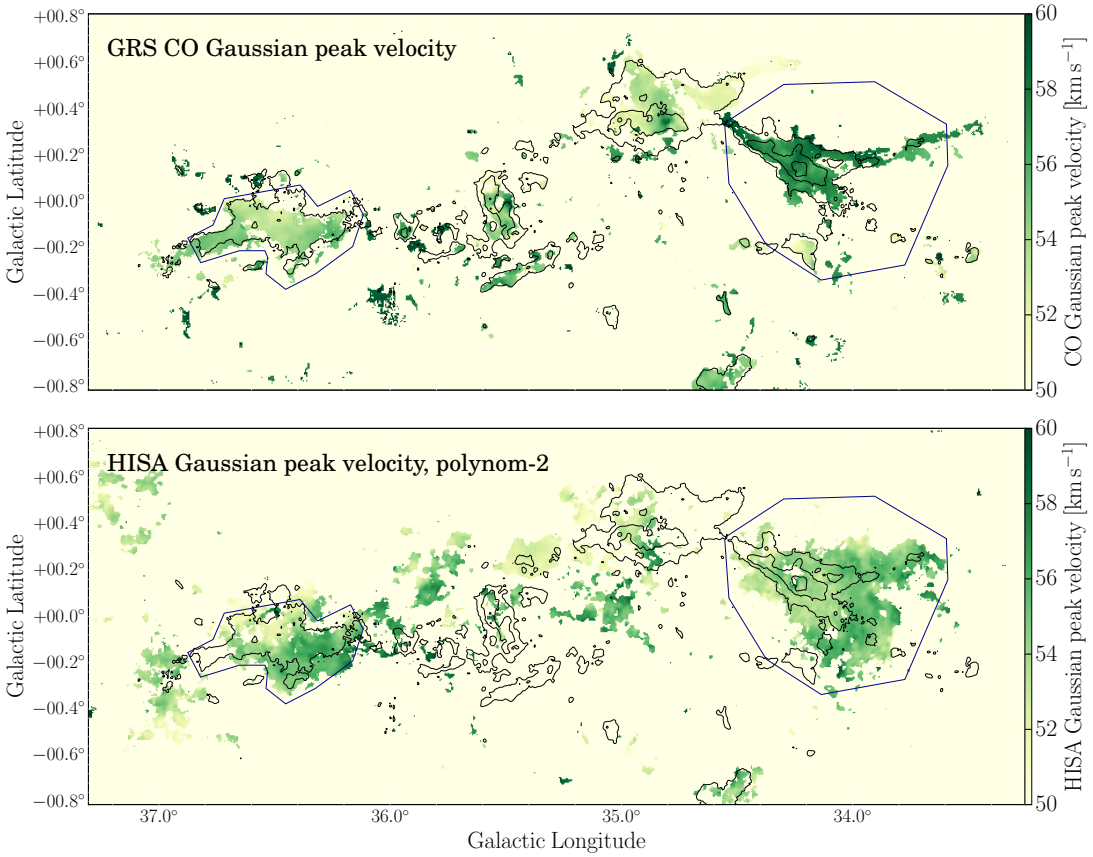

Figure 4.9: In color is shown the peak velocity of the Gaussian fit of the ^{13}CO data and the HISA data in the top and bottom panel, respectively. The H I absorption feature is extracted by using a second order polynomial to estimate the background emission. The black contours indicate the integrated ^{13}CO emission at levels of 5, 10 and 20 K km s^{-1} for reference.

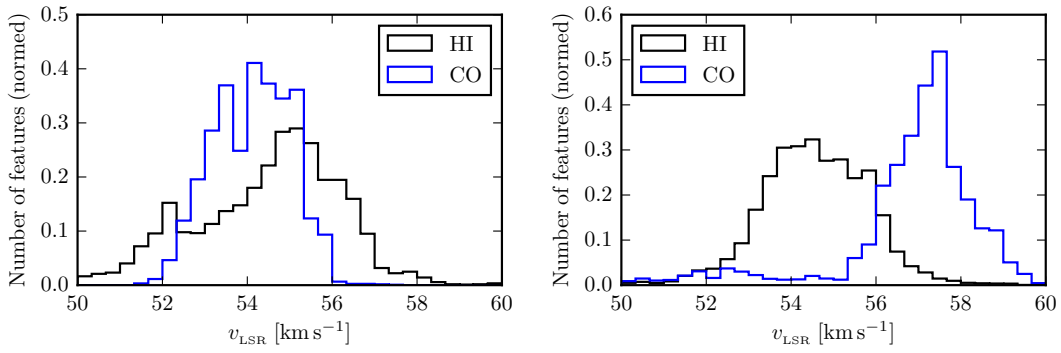

Figure 4.10: The histogram of the peak velocity of the H I and CO emission in black and blue, respectively. The left and right panel show all extracted pixel within the marked left (l = 36.5°) and right (l = 34°) polygon in Fig. 4.9.

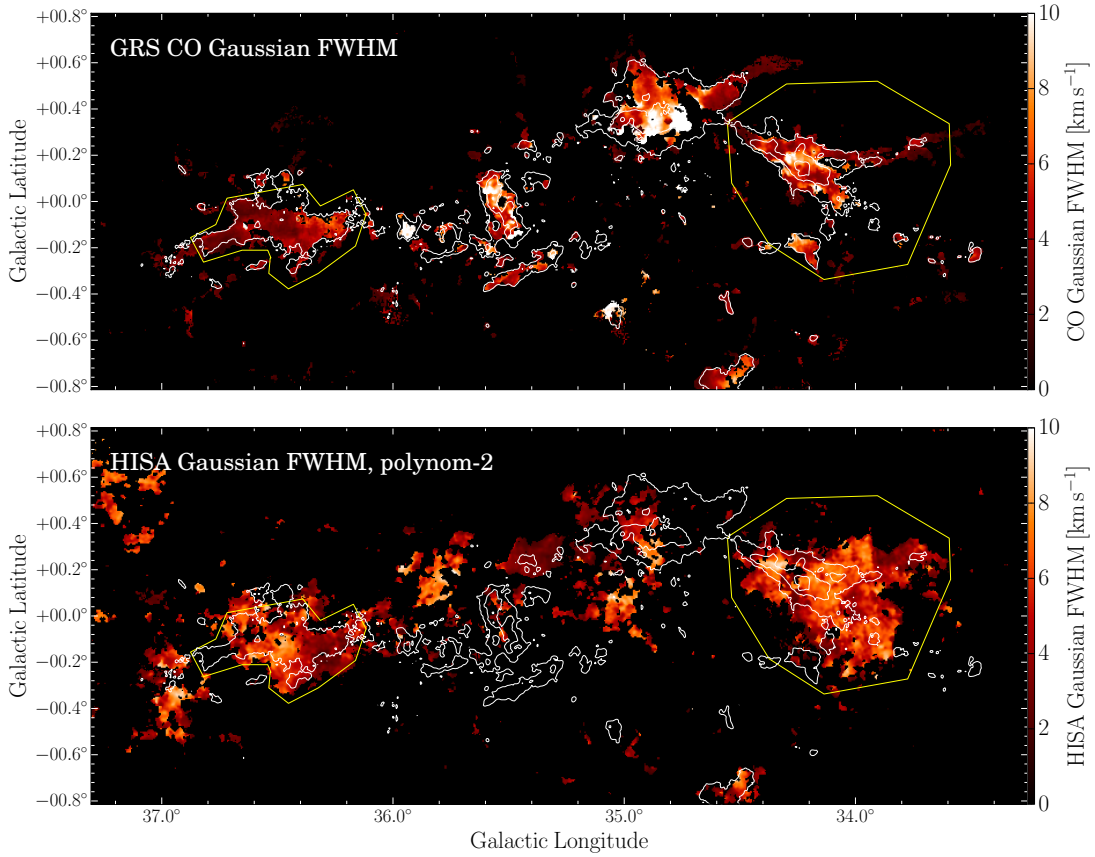

Figure 4.11: In color is shown the full width half maximum linewidth of the ^{13}CO emission and the H I absorption feature in the top and bottom panel, respectively. The linewidth is determined by using a single Gaussian fit. The contours indicate the integrated CO emission at levels of 5, 10 and 20 K km s^{-1} for reference.

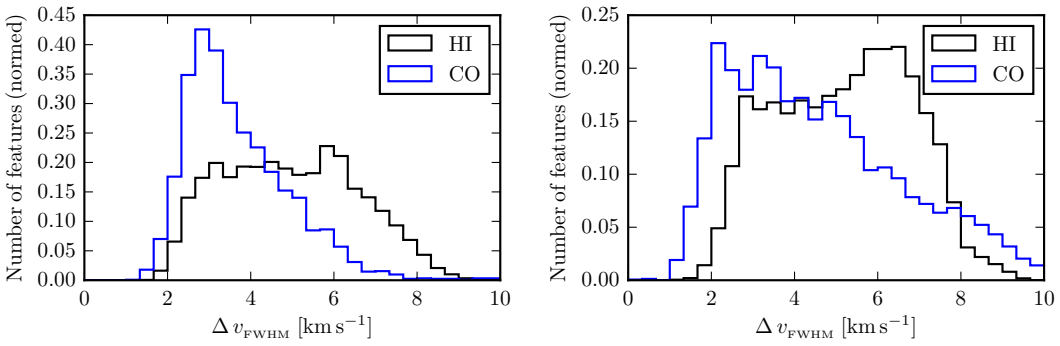

Figure 4.12: The histogram of the linewidth is shown in the left and right panel for all pixels within the left and right polygon indicated in Fig. 4.11, respectively. The black and blue curve represents the linewidth of the H I and CO emission, respectively.

this effect in Sect. 4.5.1. On the left side of the filament around $l = 36.5°$ we find mostly single components for the CO emission and the linewidth is $\Delta v \sim 3\,\mathrm{km\,s^{-1}}$. The linewidth of the H I absorption feature shows values between $\Delta v \sim 4 - 7\,\mathrm{km\,s^{-1}}$. This can also be seen in the left panel of Fig. 4.12, where a histogram of the line widths is shown. The linewidth of the CO emission is systematically higher in the right region of the filament, whereas the linewidth of the HISA feature is similar to the left region. This result has to be treated cautiously as we see multiple components for the CO line within the right region, which increases the linewidth. We will discuss the relation between the linewidth in detail in Sect. 4.5.1.

Assuming a certain spin or excitation temperature, we can estimate the thermal linewidth for the H I absorption feature and the CO emission, respectively. The thermal linewidth is given as (Stahler et al. 2005):

$$\Delta v_{\mathrm{FWHM}}(therm.) = \sqrt{\left(\frac{8 \ln 2\, k_B\, T}{m}\right)}, \tag{4.6}$$

where T describes the spin or excitation temperature and m describes the mass of the neutral hydrogen or the ^{13}CO molecule. For the H I absorption feature, this reveals a thermal linewidth between $\Delta v(therm.) = 1 - 1.5\,\mathrm{km\,s^{-1}}$ assuming a spin temperature of $T_S = 20 - 50\,\mathrm{K}$. As the measured linewidth is significantly larger ($\Delta v \gtrsim 4\,\mathrm{km\,s^{-1}}$), we can conclude that the linewidth broadening is not due to thermal motion, but other mechanism such as turbulent motion must drive the linewidth broadening. To determine the thermal linewidth of the CO, we have to consider the kinetic temperature, which might be higher than the excitation temperature. Assuming a kinetic temperature of $T_{\mathrm{kin.}} = 6 - 50\,\mathrm{K}$, the thermal linewidth of the ^{13}CO emission line is $\Delta v(therm.) = 0.1 - 0.3\,\mathrm{km\,s^{-1}}$. Hence, the linewidth broadening of the ^{13}CO emission line is also not due to thermal broadening, even for high kinetic temperatures.

4.4.5 Column density

Beside the kinematics, the column density of the H I and H_2 are a crucial point to study. As explained in Sect. 4.3.2 we can use the ^{13}CO emission to estimate the column density and hence the mass. We assume that the ^{13}CO emission is optically thin and $T_{\mathrm{ex}} = 6\,\mathrm{K}$ (Roman-Duval et al. 2010). To estimate the column density of the HISA feature we can use Eq. 1.29, which calculates the column density of the H I using the spin temperature and the optical depth. However, as mentionied in Sect. 4.3.1 we can solely measure the spin temperature and the optical depth together and disentangling them is difficult. Hence we assume that the spin temperature is constant over the cloud and calculate the optical depth using Eq. 4.4. Furthermore, we have to assume the ratio between background and foreground emission, parameterized with the factor p. As can be seen in Fig. 4.3, a larger p value reveals a smaller optical depth for the same spin temperature and hence a smaller column density. As we have no measurement for p,

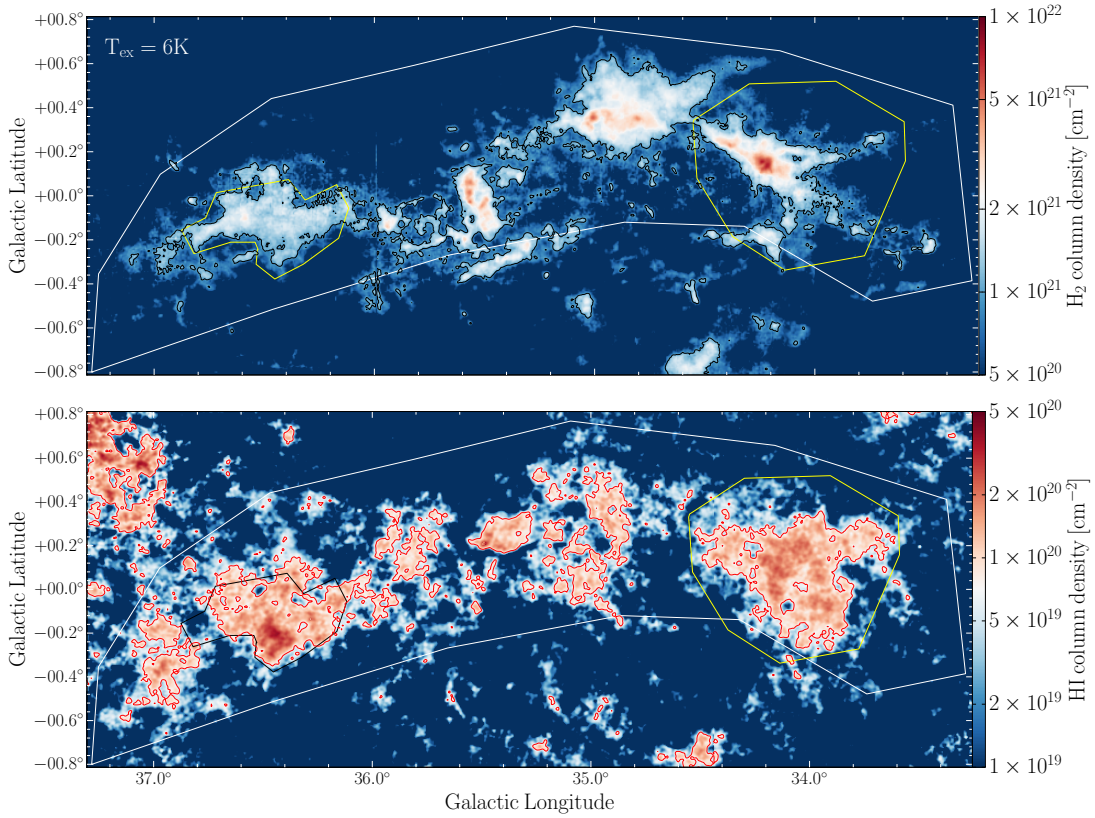

Figure 4.13: The column density of the molecular and atomic hydrogen is shown in the top and bottom panel, respectively. For the H_2 column density we used ^{13}CO as a tracer and assumed optically thin emission and T_{ex} = 6 K (see Sect. 4.3.2). The column density for the atomic hydrogen is estimated by extracting a HISA feature and assuming T_S = 20 K, p = 0.7 and using a second order polynomial to estimate the background temperature T_{off}. The contours indicate 'closed contours' (see Sect. 4.5.4). For the H_2 and H I they have levels of 1×10^{21} cm^{-2} and 7×10^{19} cm^{-2}, respectively. The white, yellow and black polygons indicate the regions for the mass estimates and the column density PDF measurements shown in Figs. 4.19 and 4.20.

this value is difficult to estimate. Figure 4.3 shows that values of p below ~0.3 are not possible as the spin temperature would become smaller than the temperature of the cosmic microwave background. In the following we will assume a value of p = 0.7 and discuss the corresponding uncertainties in Sect. 4.5.5. The chosen method to estimate the background temperature T_{off} influences the column density as well and this effect will also be discussed in Sect. 4.5.5

The result of the column density calculation is given in Fig. 4.13, assuming for the HISA feature a spin temperature T_S = 20 K, p = 0.7 and using a second order polynomial to estimate the background temperature T_{off}. Larger spin temperatures do not change the structure of the column density map significantly, but will increase it uniformly. It is striking that the column density peaks in the H_2 do not coincide with the column density peaks for the atomic hydro-

gen. As shown in Fig. 4.4, the highest peak in the CO (around $l = 34°$) coincides with a strong continuum source and hence makes the determination of the HISA feature difficult. However, we can use this continuum source to measure the optical depth, which we present in Sect. 4.5.2. The highest column density peak for the atomic hydrogen can be found in the left area of the filament ($l = 36.5°$). In this region, the H_2 column density is diffuse and low. Another H I column density peak can be found in the center of the filament around $l = 35.5°$, $b = +0.3°$. This H I feature has almost a round shape and only a very weak counterpart in the CO.

4.4.6 Mass estimate

Table 4.1: Mass estimates of the cloud.

Region	M(H_2) [$M_\odot$]	M(H I) (T_S=20 K) [$M_\odot$]	M(H I) (T_S=40 K) [$M_\odot$]	M(H_2) / M(H I) (T_S=20 K)	M(H_2) / M(H I) (T_S=40 K)
Full filament	2.7×10^5	3.7×10^3	1.2×10^4	73	23
Left region	2.9×10^4	6.4×10^2	2.3×10^3	45	13
Right region	6.3×10^4	1.2×10^3	3.7×10^3	53	17

As we know the column density and the distance to the cloud (~3.5 kpc, Ragan et al. 2014), we can directly estimate the mass. We estimated the mass for three different regions. The 'full filament', which is marked as a white box in Fig. 4.13. In addition, we measure the mass for two smaller regions, called 'left region' and 'right region', which are also marked in Fig. 4.13. Table 4.1 summarizes the mass measurements. The molecular mass for the entire filament is $\sim2.7\times10^5\,M_\odot$ and the H I mass is significantly less, showing values of $3.7\text{-}12\times10^3\,M_\odot$, depending on the spin temperature.

Furthermore, we studied the molecular to atomic mass ratio M(H_2) / M(H I). For the entire filament, this value is between 23 and 73. However this value has to be treated cautiously. As explained in Sect. 4.4.1 the HISA extraction method does not work reliably in the center of the filament due to strong continuum emission and we might miss some H I mass. The mass ratio for the smaller regions show slightly smaller values (13-53). However, the left and right region have similar values, which was not directly expected from the column density map presented in Fig. 4.13. The H_2 column density shows significantly higher values for the right region in comparison to the left region. In contrast to this, the H I column density reveals a prominent peak on the left side and hence we would expect different ratios for the right and left region. Since the H I column density covers also a larger area in the right region, the ratio stays similar. Since we determine the mass of the HISA feature, we solely measure the mass of the cold

H I. However, it is very likely that this component is surrounded or even mixed with a warm component. Estimating the mass of the warm component can be done by assuming that the fitted background emission ($T_{\rm off}$) is optically thin. Assuming a polynomial of second order to fit the background reveals a mass of $\sim 1.8 \times 10^5\,M_\odot$ for the full filament. This is about a factor of ten larger than the mass estimated for the cold HISA feature. However, this value has to be treated cautiously as we use the diffuse emission to measure the mass of the warm component, which must not be correlated with the cold component or the H_2 filament, but could also occur from background or foreground clouds. On the scale of the entire Milky Way, it is known that the mass of the CNM and the WNM is approximately the same (Stahler et al. 2005).

4.5 Discussion

4.5.1 Kinematics

For nearby galaxies, the ratios of the CO to H I linewidth reveals values around $\sigma_{\rm H\,I}/\sigma_{CO} = 1 - 1.4$ (Caldú-Primo et al. 2013; Mogotsi et al. 2016). The linewidth values found in these studies are approximately $\sigma \sim 6 - 12\,{\rm km\,s^{-1}}$, which corresponds to $\Delta v_{\rm FWHM} \sim 14 - 28\,{\rm km\,s^{-1}}$ for the H I and CO lines. These measurement are done for the CO and H I emission over large regions (~0.5 kpc), which can increase the linewidth due to superposition of different velocity components. Our Galactic measurements of the GMF38.1-32.4 show significantly smaller values for the linewidth of $\Delta v_{\rm FWHM} \sim 2 - 8\,{\rm km\,s^{-1}}$. The reason is that we observe a much smaller region and we are able to observe multiple components. Furthermore, we observe cold H I absorption features instead of warm emission features. However, we already discussed that the observed linewidth broadening is not thermal (see Sect. 4.4.4).

To study the linewidth ratio in detail, we determine this ratio for the left region, which is indicated in Fig. 4.11. We use solely this region as it is not affected significantly by multiple line spectra. The result is shown in Fig. 4.14. A histogram of the H I/CO ratio is shown in the left panel and the right panel presents the CO linewidth as a function of the H I linewidth. Similar to the mentioned extra galactic measurements, we find a H I/CO linewidth ratio around 1, even though the linewidth values are significantly smaller. The mean value for the linewidths are $\Delta v_{\rm FWHM}(CO) = 4.0\,{\rm km\,s^{-1}}$ and $\Delta v_{\rm FWHM}({\rm H\,I}) = 5.1\,{\rm km\,s^{-1}}$.

To examine the effect of averaging over a larger area, we average all spectra of the CO emission and H I absorption in the left region. A Gaussian function fits this averaged spectra well and is shown in Fig. 4.15. The corresponding linewidth is $\Delta v_{\rm FWHM}(CO) = 4.8\,{\rm km\,s^{-1}}$ and $\Delta v_{\rm FWHM}({\rm H\,I}) = 5.9\,{\rm km\,s^{-1}}$. This results in a H I/CO linewidth ratio of 1.2. Even though the linewidth ratio is not affected by the averaging process, the actual linewidth values are slightly affected by the spectrum averaging process. Since we average over slightly shifted velocity

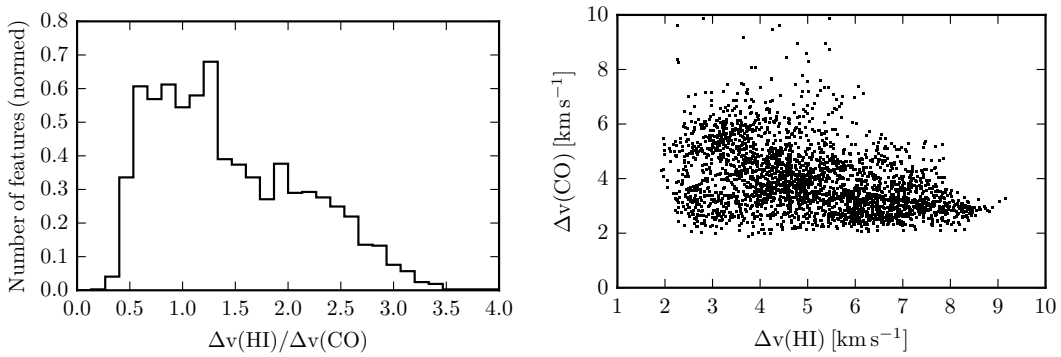

Figure 4.14: The linewidth ratio of the left region indicated in Fig. 4.11. The left panel shows a histogram of the ratio, whereas the left panel presents the CO linewidth as a function of the H I linewidth.

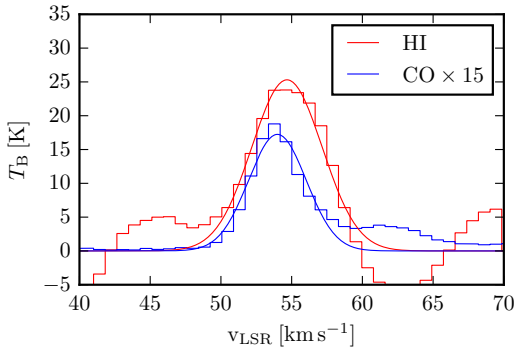

Figure 4.15: The averaged spectra of the left region indicated in Fig. 4.11 for the CO and H I line in blue and red, respectively. The CO line is multiplied by 15 for better readability.

components, the averaged linewidth is larger. This can also explain the large values observed in extra galactic systems as these observations have a poor linear resolution and hence are averaged over large areas (~0.5 kpc).

4.5.2 Optical depth measurement toward a strong continuum source

Similar to the method used in Chapter 2, we can use strong continuum sources to estimate the optical depth. The UCH II region G34.256+0.146 (see Fig. 4.4) is an ideal candidate to perform this task as it is very bright ($T_{\rm cont}(max) \sim 2100$ K) and slightly extended ($d \sim 70''$). We use the THOR data to determine the absorption spectra. To increase the signal-to-noise ratio, we create an averaged absorption spectra for this source and determine the optical depth using Eq. 2.6. In the velocity range of the HISA feature, the absorption spectra saturates and the determined

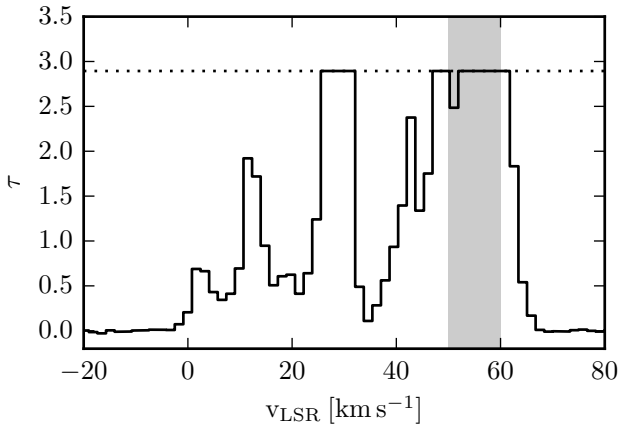

Figure 4.16: The H I optical depth spectra towards the UCH II region G34.256+0.146 using the THOR data. For some channels, the absorption spectra saturates and the measured optical depth is a lower limit of $\tau = 2.9$, which is indicated by the dotted line. The gray shaded area indicates the velocity range of the HISA feature ($v_{\rm LSR} = 50 - 60\,{\rm km\,s^{-1}}$).

optical depth $\tau = 2.9$ is a lower limit. The optical depth spectrum is shown in Fig. 4.16.

A problem for HISA studies is that there is a possibility of interpreting missing background emission as an absorption feature. To avoid this, we can solely study very narrow absorption features, so called 'H I narrow self absorption' (HINSA) (e.g., Li & Goldsmith 2003). The steep absorption profile seen for these HINSA features can not be induced by two broad emission profiles on each side of the absorption feature. However, we see broad HISA features, which could be caused by two emission profiles. Fortunately, the optical depth information provided by the strong continuum source helps to solve this problem. Since the optical depth is high ($\tau > 2.9$) for the velocity range of the HISA feature, we are confident that we actually observe a cold absorbing HISA feature rather than missing H I emission. Furthermore, the correlation of the HISA feature with the CO emission is another indicator for cold, dense H I.

As explained in Sect. 4.3.1, the general HISA extraction method measures the optical depth together with the spin temperature and we are not able to disentangle them in general. However, the additional information from the absorption spectra towards the strong continuum source and the corresponding optical depth measurement allows us to overcome this problem. Figure 4.17 presents the optical depth as a function of the spin temperature for different values of p. The lower limit of the optical depth measurement is shown at $\tau = 2.9$ using a black horizontal line. Assuming $p = 0.7$ reveals a spin temperature of $T_{\rm S} \sim 40\,{\rm K}$. This is two times higher than the assumed spin temperature for the column density determination presented in Fig.4.13 and on the upper end of the assumed spin temperature for the mass estimate presented in Sect. 4.4.6. Since we measure the spin temperature close to an UCH II region, we expect rather high values.

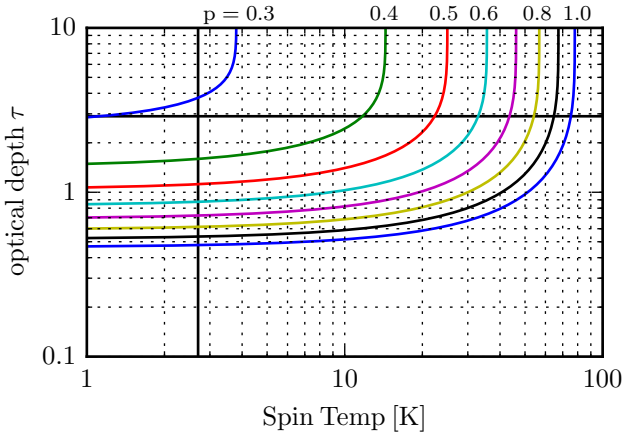

Figure 4.17: The optical depth as a function of the spin temperature following Eq. 4.4 for the region close to the strong continuum source G34.256+0.146. The optical depth of $\tau = 2.9$ is indicated with a black horizontal line. The Temperature of the CMB is indicated at $T = 2.7$ K with a black vertical line. This Figure is similar to Fig. 4.3, but here we represent actual measurements close to the continuum source G34.256+0.146 with $T_{\rm off} = 106$ K, $T_{\rm on} = 60$ K and $T_{\rm cont} = 18$ K.

However, we will show in Sect. 4.5.3 that such a high spin temperature cannot be achieved in other HISA regions.

4.5.3 Maximum spin temperature

As explained in Sec. 4.5.2, we can use strong continuum sources to measure the optical depth and therefore disentangle the spin temperature and the optical depth. However, this is a special case in the vicinity of a strong continuum source. In general we can solely give the spin temperature as a function of the optical depth. Fig. 4.17 shows the relation and it is striking that the function becomes very steep for a certain spin temperature. Hence, this Figure shows that the maximum spin temperature will be reached for the case of large optical depth. This can also be seen by solving Eq. 4.4 for $T_{\rm S}$:

$$T_{\rm S} = \frac{T_{\rm on} - T_{\rm off}}{1 - e^{-\tau}} + p\,T_{\rm off} + T_{\rm cont}. \tag{4.7}$$

Since $T_{\rm on} - T_{\rm off}$ is always negative, $T_{\rm S}$ reaches an upper limit if $\frac{T_{\rm on}-T_{\rm off}}{1-e^{-\tau}}$ becomes minimal, which is the case for $\tau \to \infty$. This means the maximal $T_{\rm S}$ is:

$$\begin{aligned} T_{\rm S}(max.) &= T_{\rm on} - T_{\rm off} + p\,T_{\rm off} + T_{\rm cont} \\ &= T_{\rm on} + T_{\rm cont} - (1-p)\,T_{\rm off}. \end{aligned} \tag{4.8}$$

This equation depends on the assumption of the ratio of foreground and background emission, which is described with the factor p. For $p = 1$, the upper limit of the spin temperature reaches

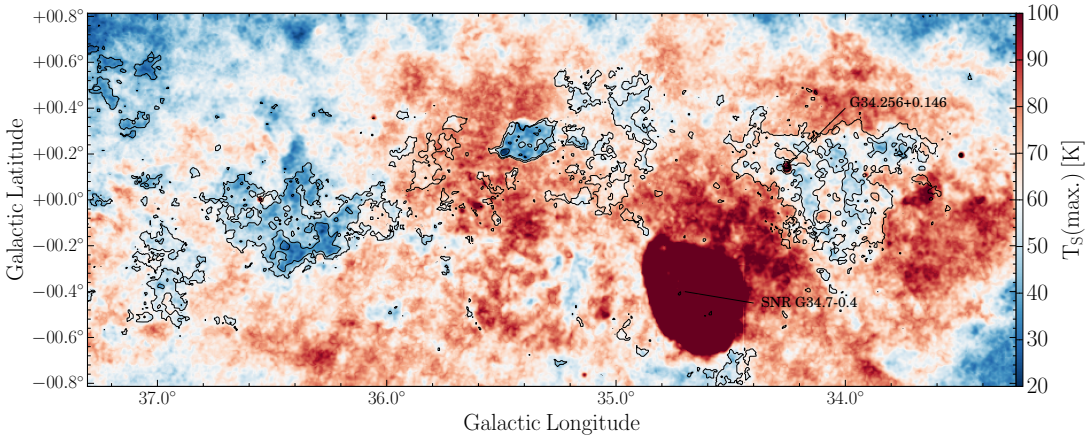

Figure 4.18: In color is presented the maximum spin temperature (Eq. 4.8) of the absorption features assuming $p = 0.7$. The black contours indicate the peak value of the Gaussian fit of $T_{\rm off} - T_{\rm on}$ and hence the depth of the H I self absorption. Regions with strong continuum emission are not reliable, such as SNR G34.7-0.4 and G34.256+0.146 (see Fig. 4.4).

a maximum. We can use this information to calculate the upper limit of the spin temperature for each pixel in our map. However, we do not assume $p = 1$, but rather a more realistic value of $p = 0.7$. The result is given in Fig.4.18. Regions with strong continuum emission are not reliable for this calculation as we would observe H I continuum absorption (HICA) towards these continuum sources, rather than H I self absorption (HISA). As expected, we see a clear correlation of the upper limit for the spin temperature and the absorption depth of the HISA feature. The lowest values are found for the compact HISA feature in the center around $l = 35.5°$ with values around $T_S(max.) \sim 10$ K. Similar values can be found for the left region of the filament, whereas the right side of the filament shows in general higher values around $T_S(max.) \sim 10$ K. As this is solely an upper limit for the spin temperature, we cannot conclude directly the actual temperature. However, it is plausible that the H I spin temperature is higher on the right side of the cloud due to star formation activity and feedback mechanisms, such as the prominent UCH II region. We will discuss this aspect further in Sect. 4.5.6.

For the H I column density determination in Sect. 4.4.5, we assumed a spin temperature of $T_S = 20$ K. As seen in Fig. 4.18, this is higher than the upper limit of the spin temperature for certain cold regions. Hence, the calculated optical depth would go to infinity and we cannot determine a column density for these regions. To overcome this problem, we exclude these pixels from our column density calculation. Since we observe only small regions with $T_S(max.) < 20$ K in only a few velocity channels, the column density calculation is not affected significantly. However, assuming a larger value for the spin temperature increases this effect and larger regions are affected, which makes the determined column density unreliable.

4.5.4 Column density probability density function

The column density maps derived in Sect. 4.4.5 can be utilized to determine column density probability distribution functions (PDFs). Fig. 4.19 presents the column density PDF for the entire filament. For the HISA feature, we assumed two different spin temperatures of T_S = 20 and 40 K and used a second order polynomial to estimate the background emission, which are the same assumptions used to produce Fig. 4.13. For the H_2 column density we assumed an excitation temperature of T_{ex} = 6 K (Roman-Duval et al. 2010). The shape of the PDF can also be influenced by the chosen region, especially in the low column density regime (Lombardi et al. 2015). To minimize this effect, we estimated the lowest contour for the column density values, which is still fully contained within the observed region. These are called 'closed contours' and are shown in Fig. 4.13. The corresponding value is also presented as a vertical line in the column density PDF (Fig. 4.19). Values below this point have to be treated cautiously as we miss data points for the PDF in these low column density regions.

As expected from the column density maps (Fig. 4.13) and mass estimates, the H I column density PDFs peak at lower values than the H_2 column density PDF. To describe the shape of the column density PDF, we can either use a log-normal or power-law function. As can be seen in Fig. 4.19, the log-normal function represents the H I and the H_2 column density PDF well. For comparison, we fit the high column density part with a power-law function. However, we do not observe a high column density tail, which would require an extra power-law function. Our data is well represented by a single log-normal shape function. This implies that turbulent motions might be the main driver and gravitational collapse is not seen in our data on the scale of the full filament.

Beside the overall structure of the long filament, we can study smaller sub-regions within the filament. The column density PDFs for the left and right regions, indicated with polygons in Fig. 4.13, are shown in Fig. 4.20. As already mentioned in Sect. 4.4.5, the left region reveals a rather diffuse and low H_2 column density and does not exhibit any prominent peaks. Hence, the corresponding PDF shows a steep drop for the high column density regime. This can be either described by a steep power-law or by a narrow log-normal function. The H I column density reveals a broader column density distribution.

In contrast to this, the H_2 column density PDF of the right region is poorly represented by a single log-normal function, but shows a high column density power-law tail with a slope of $\alpha = -2.1$. Hence we witness gravitational collapse in this region of the molecular hydrogen, which creates high column density peaks. The H I column density PDF does not show a high column density power-law tail and can be represented well by a single log-normal function.

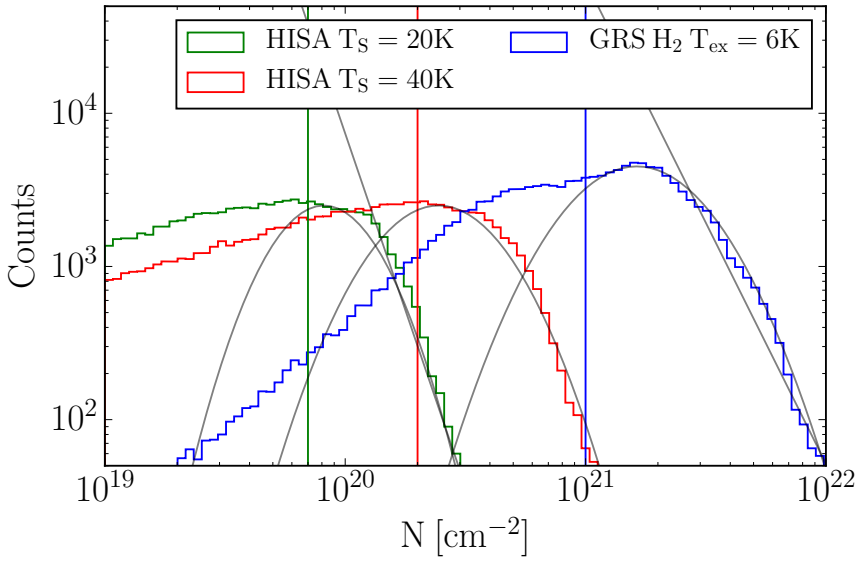

Figure 4.19: The column density PDF of the atomic and molecular hydrogen for the entire filament indicated with the white box in Fig. 4.13. The atomic hydrogen is estimated for the HISA feature assuming a spin temperature of T_S = 20 and 40 K and using a polynomial fit of second order to estimate the background. The molecular hydrogen is estimated using ^{13}CO emission and assuming optically thin emission with an excitation temperature of T_{ex} = 6 K. The gray curves indicate the corresponding log-normal shapes or power-law functions.

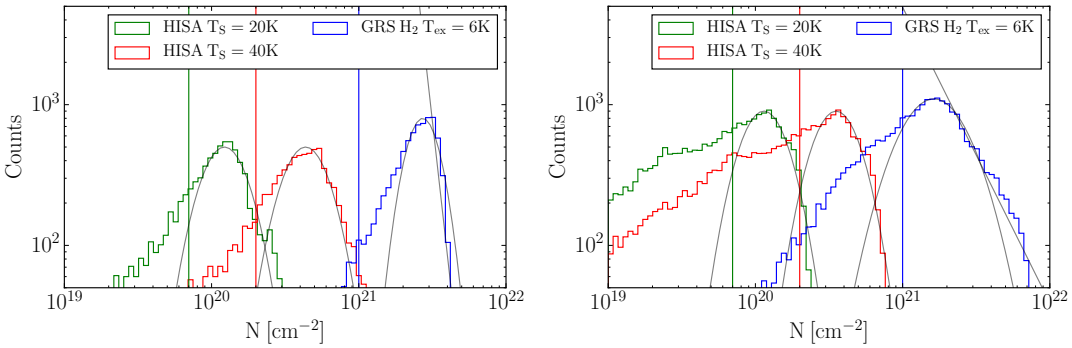

Figure 4.20: Same as Fig. 4.19, but for sub-regions within the filament. The left and right panel show the PDF of the left and right region, respectively, indicated with polygons in Fig. 4.13.

4.5.5 Uncertainties for the determined HISA properties

Several factors introduce uncertainties to the determined properties of the HISA features. In the following we will discuss three factors: the ratio of foreground to background emission, different methods to determine the background emission and the assumption of the spin temperature.

The ratio of the foreground and background emission, parameterized with p, is difficult to estimate. Different assumptions can be found in the literature. For example, McClure-Griffiths et al. (2006) assumes that $p = 1$ for the observed Riegel-Crutcher cloud, as the corresponding distance is small (~125 pc). Since the background and foreground H I emission occurs from warm and diffuse H I clouds, we do not expect fluctuations of this emission on small scales. Hence, the assumption of a constant p for the entire filament is reasonable. Furthermore, we can constrain a lower limit for p. As shown in Fig. 4.17, low values of $p \lesssim 0.3$ are not feasible, as the spin temperature would become smaller than the temperature of the CMB. Larger values of p result in smaller values for the optical depth, assuming a constant spin temperature. Hence, the column density becomes smaller as well. Depending on the used p value, the column density can change by at most a factor of 5. This is shown in Fig. 4.21 for the column density PDF of the entire filament assuming four differnet values for p. However, the column density structure stays constant, but the actual values are shifted for different p values.

As discussed in Sect. 4.4.2, the chosen method for the background estimate can influence the absorption depth of the HISA feature and thus the column density. We showed that the best method is a fit to the H I spectra and interpolate for the HISA feature. The difference for a second or fourth order polynomial is negligible for most regions and we use a second order polynomial.

The most important factor is the assumption of a constant spin temperature for the cloud. This is obviously a poor assumption, but due to additional measurements we can constrain the range of the spin temperature. The most important constrain is the upper limit for the spin temperature introduced in Sect. 4.5.3. Using this information, we can constrain the spin temperature to values of $T_S < 40$ K for the majority of the HISA features. For the H I mass estimate given in Sect. 4.4.6 we assumed a spin temperature of $T_S = 20$ and 40 K. As seen in Fig. 4.19, the column density distribution does not change significantly, but larger spin temperatures introduce larger column densities and masses. In Sect. 4.4.6, we showed that the mass is about a factor of three larger for $T_S = 40$ K instead of 20 K.

Summarizing, it is difficult to quantify exactly the uncertainty of the H I column density and mass. However, considering all assumptions, the H I mass will have an uncertainty of a factor of 2-4, which is similar to the H_2 mass uncertainty derived using ^{13}CO. However, we could

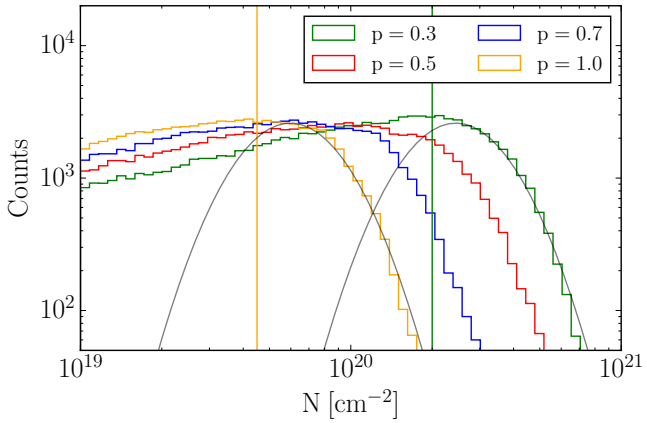

Figure 4.21: The column density PDF for the atomic hydrogen for different ratios of foreground to background emission, paramterized with p. The PDF is measured over the entire filament, indicated with a white polygon in Fig. 4.13, assuming a constant spin temperature of $T_S = 20$ K. The yellow and green vertical line presents the closed contours shown in Fig. 4.13.

show that the shape of the column density PDF is robust.

4.5.6 Evolutionary stages?

As mentioned in Sect. 4.4.5, we observe a significant difference of the H_2 column density for the left and right sub-region of the filament. The left sub-region shows a more diffuse column density, whereas the right sub-region reveals several high column density peaks. Hence, the right sub-region reveals a high column density power-law tail in the H_2 column density PDFs shown in Fig. 4.20 and the left sub-region shows a log-normal shaped PDF. Furthermore, we see a strong UCH II region within the right sub-region, whereas the left sub-region does not harbor significant continuum emission. Ragan et al. (2014) studied also high density tracers, using the ATLASGAL survey (Schuller et al. 2009). Similar to the H_2 column density, they find several high density peaks in the right sub-region, whereas such peaks are absent in the left sub-region. All these different tracers disclose an actively star-forming region on the right side of the filament, while the left side shows no star formation activity. An intriguing question is if we witness different evolutionary stages for these two regions in the filament and do we see such signatures also for the atomic hydrogen?

The kinematics of the HISA do not exhibit significant differences between the left and right sub-region and the linewidth discloses similar values. However, comparing the HISA and CO kinematics shows an interesting difference. The left sub-region shows similar peak velocities for the H I and CO, whereas the right sub-region reveals a difference of $\sim$4 km s^{-1}. Studying the H I column density, we found higher column density peaks on the left side in comparison

to a more diffuse column density structure on the right side. The maximal spin temperature also discloses smaller values on the left side. This might be an indication of a young, colder and more dense H I cloud on the left side in comparison to a more evolved cloud on the right side. It is possible that the dense H I cloud on the left side of the filament is about to become a dense molecular cloud, forming high density peaks and subsequently form stars. However, further observations or simulations are needed to support this statement.
Simulations of cloud formation and H I to H_2 conversion so far do not give clear predictions of the signatures one would find in the atomic or molecular tracers and whether they vary for example in models of colliding or converging gas flows compared to less dynamic cloud contraction models. Here, we present observables, in particular the steepness of the column density PDF and the variation in peak velocities between the molecular and atomic gas. While we have to extend these studies to larger samples, theorists need to provide comparable parameters to test the validity of their models in comparison to the data.

4.6 Conclusions

We study the atomic and molecular hydrogen within the molecular filament GMF38.1-32.4. The molecular hydrogen is traced via observations of ^{13}CO, whereas the cold H I is observed via H I self absorption features. The main results can be summarized as:

1. We extracted H I self absorption features (HISA) by estimating the background emission with different methods. For the observed large filament, a polynomial fit of second order is the most reliable method to estimate the background emission.
2. The HISA features and the CO emission reveal a spatial correlation. However, the peak velocity shows a step of $\sim$4 km s^{-1} for the right part of the filament, whereas the left part shows also a close correlation in velocity space.
3. Assuming a spin temperature of $T_S = 20 - 40$ K for the H I and an excitation temperature of $T_{ex} = 6$ K for the CO, we are able to determine the column densities. The corresponding peaks do not coincide and the H I column density shows in general a diffuse structure. The H_2 column density reveals prominent column density peaks on the right side of the filament whereas the left side appears more diffuse.
4. Studying the column density probability distribution functions (PDFs), we are able to determine the physical processes within the cloud. Both, the H I and H_2 column density PDFs are represented well by a log-normal function, which indicates turbulent motions as the main driver. Only the H_2 column density of the right sub-region within the filament reveals a high column density power-law tail, which indicates star formation activity.

5. We speculate that the left and right side of the filament represent different evolutionary stages. The left side might represent an earlier stage, which forms currently a dense molecular cloud out of the atomic reservoir. We do not observe high molecular column density peaks, nor signs of active star formation for this sub-region, but the H I shows low spin temperatures and high column densities. In contrast to this, the right side of the filament shows high H_2 column density peaks, sings of active star formation, such as UCH II regions and in general a warmer and less dense atomic counterpart. This provides interesting observables for theoretical models and simulations.

5

Summary, conclusions and outlook

5.1 Summary and Conclusions

This thesis is dedicated to the early formation mechanism of molecular cloud formation starting in the atomic phase. In the context of the H I, OH, Recombination Line Survey of the Milky Way (THOR), we do not just answer scientific questions but also tackle technical difficulties. The THOR survey covers a large portion of the Galactic disk ($l = 15 - 67°$, $|b| \leq 1°$) in the 21cm H I line, 4 OH lines, 19 Hα recombination lines and the continuum from 1-2 GHz using the VLA in the C-array configuration. This results in about 4 TB raw data and the reduction of such a massive data set is not a trivial undertaking.

Chapter 3 provides a catalog of 4422 continuum sources within the first half of the survey, providing an insight into the data reduction and source extraction methods. Beside these technical details, we could extract spectral indices α with the form $I(\nu) \propto \nu^{\alpha}$ for 1840 sources. The distribution of α reveals a double peaked profile, peaking at $\alpha = -1$ and $\alpha = 0$, which indicates non-thermal and thermal radiation, respectively. Using the spectral index, we could also classify and study super nova remnants (SNR), confirming four SNR candidates showing a spectral index of $\alpha = -0.5$. Even though such a continuum source catalog itself does not answer fundamental questions about molecular cloud formation, it is a crucial database for subsequent absorption studies of the HI line as well as the OH transitions.

The THOR survey started with the pilot study around the giant molecular cloud (GMC) associated with W43 ($l = 29.2 - 31.5°$, $|b| \leq 1°$) which is presented in Chapter 2. We could show that the assumption of optically thin emission is not sufficient to describe the mass or column density of such a cloud. Using the strong continuum source W43-Main in the center, we estimated the optical depth with a lower limit of $\tau \sim 2.7$. This information allowed us to correct the column density and mass and we measured the HI mass of W43 with a lower limit of $M = 6.6_{-1.8} \times 10^6\,\mathrm{M}_{\odot}$, which is a factor of ~2.4 larger than the mass assuming optically thin emission. We used the corrected column density to study the spatial structure of the HI

and assumed an elliptical structure for W43 and we found a linearly decreasing H I volume density towards the center with values from $n_{\mathrm{H\,I}} = 20$ to almost $0\,\mathrm{cm}^{-3}$. In contrast to this, the molecular hydrogen, traced via Herschel dust observations, increases exponentially from $n_{\mathrm{H_2}} = 15$ to $200\,\mathrm{cm}^{-3}$. For W43, we do not observe a clear transition zone from H I to H_2, neither do we observe column density thresholds for the atomic phase, at which the formation of molecular hydrogen should set in. Such a threshold of $N_{\mathrm{H\,I}} \sim 10\,\mathrm{M}_\odot\,\mathrm{pc}^{-2}$ was proposed by analytic cloud formation models (Krumholz et al. 2008, 2009). However, we find values up to $N_{\mathrm{H\,I}} \sim 150\,\mathrm{M}_\odot\,\mathrm{pc}^{-2}$. Since W43 harbors a central cluster, its internal radiation field is strong enough to dissociate large amount of the molecular hydrogen. We might also observe several transition zones along the line of sight. This finding highlights the importance of the environment on the formation of molecular hydrogen out of the atomic phase.

To disentangle the warm neutral medium (WNM) and the cold neutral medium (CNM) of the atomic hydrogen, we extracted HI self absorption features (HISA) for the long molecular filament GMF38.1-32.4. Since the HISA traces only the cold HI component we could study its interplay with the molecular hydrogen, which we traced via ^{13}CO observations. The observed linewidths are not thermal and broaden due to turbulent motion of superposition of different velocity components. Furthermore, we studied column density probability distribution functions (PDFs) for the atomic and molecular phase. The PDFs from the atomic hydrogen reveal a log-normal shape, which indicates turbulent motions as the main driver. Most of the PDFs from the molecular counterpart show the same structure. However, we find one sub-region, which reveals a power law tail for high column densities, indicating gravitational collapse. We speculate that two sub-regions within the filament represent different evolutionary stages, whereas the left sub-region is in an early stage and currently forming a molecular cloud out of the atomic phase. In contrast to this, the right sub-region is more evolved, containing high column density peaks and showing signs of active star formation. These signatures provide parameters, which can be used to validate or disprove current cloud formation models.

5.2 Outlook

In the following I will introduce two projects that I would like to do in the near future. Both are a continuation of the presented work, either with existing data or upcoming datasets.

5.2.1 A statistical study of HISA features

In Chapter 4 we presented the HI self absorption along one long filament and determined its properties, such as the spin temperature, column density or linewidth. Furthermore, we compared our findings with molecular tracers and we chose ^{13}CO to perform this. However,

further molecular tracers are available at similar resolution and sensitivity. The obvious next step will be the comparison to Herschel dust data, similar to the data presented in Chapter 2. Even though we cannot determine kinematical properties from the Herschel data, it will provide us with a measurement of the entire gas content, including HI and H_2. This will allow us to compare the entire gas content to the corresponding CO and HI content and calibrate the spin and excitation temperature. Furthermore, we can study the so called CO-dark-H_2 gas (Smith et al. 2014).

As a next step we will study further HISA features and apply the developed observing methods on a larger sample including other long filaments from the sample of Ragan et al. (2014) and from other surveys (e.g., Wang et al. 2015). We will study other HISA clouds and create a consistent HISA catalog, similar to the work on the Canadian Galactic Plane Survey (CGPS) presented by Gibson et al. (2005a,b). This will allow us to study HI clouds with different properties, such as mass and temperature in different environments and compare our findings to theoretical models or simulations of molecular cloud formation, such as the SILCC (SImulating the LifeCycle of molecular Clouds) project (Walch et al. 2015).

5.2.2 Re-weighting the Milky Way

To date, most H I studies consider the 21cm emission to be optically thin and calculate properties such as the column density using this assumption. In Chapter 2, we could show that this assumption underestimates the mass significantly by at least a factor of 2-3 for the GMC associated with the W43 star-formation complex (Bihr et al. 2015). However, other studies of low mass clouds such as Perseus showed that the assumption of optically thin H I emission is reasonable (Lee et al. 2015). Milky Way studies also use the assumption of optically thin emission to determine the H I mass of the Milky Way, which is $\sim 8 \times 10^9$ $M_\odot$ (Kalberla & Kerp 2009). However, we would like to answer the question: Is the assumption of optically thin H I emission on a Galactic scale reasonable or do we miss a significant fraction of the H I mass? The answer for this question will have various implications on different topics, such as rotation curve studies of the Milky Way, including the dark matter content as well as theoretical studies of the ISM in the spiral galaxies in general. The capability of the THOR survey will make it possible to tackle this challenging question. We could use a large sample of H I absorption and emission spectra toward extragalactic continuum sources to disentangle the different H I phases and determine the optical depth as well as the spin temperature. This would allow us to create an optical depth profile of the Milky Way and determine the correct mass. Initial studies examining this topic showed an increasing optical depth toward the center of the Milky Way around a Galactic radius of 4 to 8 kpc (Kolpak et al. 2002). However, this study is very limited by the number of continuum sources (~50) and therefore suffers of low sampling. Furthermore,

the sensitivity ($\sigma_{\rm fi}$ = 0.2) of their observations is low and therefore a lot of absorption spectra saturate. Hence, it cannot be used to determine an accurate H I mass on a Galactic scale.
Within the first half of the THOR survey field, we find over 100 sources, which are brighter than 200 K and hence suitable for absorption studies. This number will approximately double for the entire survey field. About half of these sources reveal a negative spectral index, which indicates that these sources are extragalactic. These extragalactic sources are ideal to do absorption studies, because they are located behind the Milky Way and the absorption spectra reveal the full spiral structure. Beside the extragalactic continuum sources, Galactic continuum sources, such as H II regions can also be used for such a study. Due to their position within the Galactic plane, the interpretation of the absorption spectra can be challenging. However, they offer the possibility to solve the near/far distance ambiguity. Furthermore, largely extended continuum sources can be used to investigate H I absorption spatially. Using all this information will allow us to elucidate the H I content of our own Galaxy and to re-weight the Milky Way.
However, there are also a few drawbacks. As we do not have a uniform coverage of background sources, we have to develop a suitable method to interpolate between the different sources. This becomes complicated if we include the Galactic continuum sources, as they trace the H I only until their position in the Milky Way. Another major drawback is the saturation of the absorption spectra, which depends on the noise of the data and the strength of the continuum sources (see Eq. 2.7). This means a bright background source can be used to give estimates of higher optical depths than weaker sources. We also have to consider this effect when interpolating between sources, which is not a trivial problem. Higher sensitivity will help to overcome this problem, which we could achieve by spectral smoothing. For extended sources, we could also use a spatially averaged spectra to achieve a higher signal-to-noise ratio.
To minimize the mentioned drawbacks further, additional observations would be needed. The capability of the Australian Square Kilometre Array Pathfinder (ASKAP) telescope and the associated 'Galactic ASKAP Survey' (GASKAP) (Dickey et al. 2013) will revolutionize our view of the Milky Way. Having similar resolutions than the THOR survey, but about an order of magnitude higher sensitivity and a larger survey area, will allow us to utilize ~15000 continuum sources in the low latitude area of the Milky Way ($|b| < 2.5°$) with an optical depth sensitivity of $\sigma_{\rm fi} \leqslant 0.05$. This will improve the sampling dramatically and simplify the problem of the interpolation between the sources. The first observations of this survey might start end of 2016.
A logical extension to this project would be to examine extragalactic systems. Strong continuum sources within the center of a galaxy provide us with a unique possibility to measure the optical depth of the H I line and therefore to determine the correct column density. Galaxies with different inclination angles would allow us to probe different regions within the disk and

hence also get a better picture of our own Milky Way.

A

Appendix for THOR continuum catalog (Chap. 3)

A.1 Source Examples

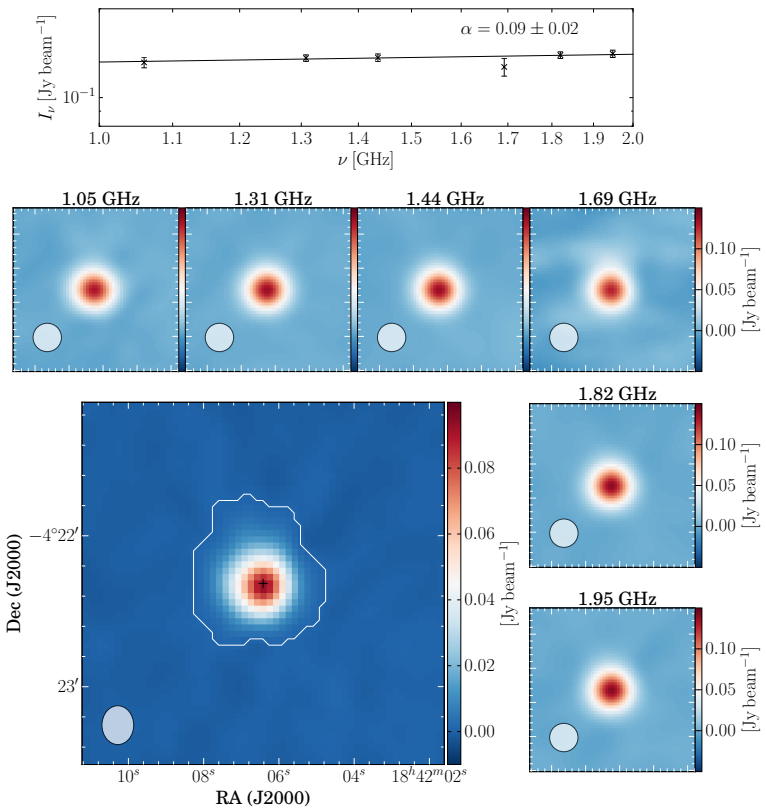

Figure A.1: Example image of the THOR source G27.978+0.078, which corresponds to the WISE H II region G027.980+00.080 (Anderson et al. 2014). The large image represents an averaged image of the two spectral windows around 1.4 and 1.8 GHz, which we used for the source extraction (see Sects. 3.4.1 and 3.4.3). The white contours shows the extent of the source determined by the BLOBCAT algorithm. The black cross marks the peak position, which we used to determine the spectral index. The small images show each spectral window separately and the top panel presents the peak intensity for each spectral window and the corresponding spectral index fit. In each image the synthesized beam is given in the lower left corner.

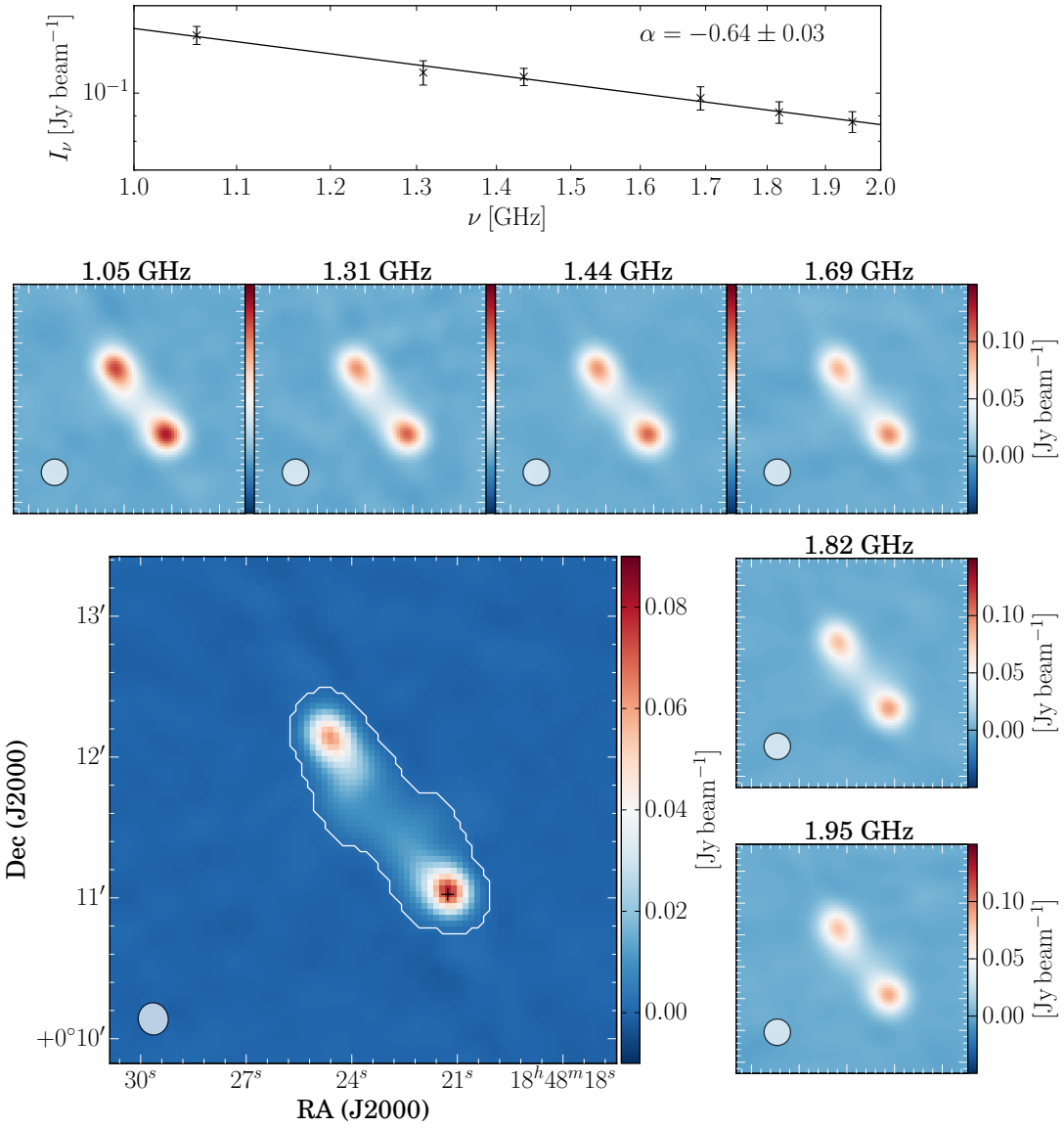

Figure A.2: Example image of the THOR source G32.744+0.770, which is most likely an extragalactic jet. The large image represents an averaged image of the two spectral windows around 1.4 and 1.8 GHz, which we used for the source extraction (see Sects. 3.4.1 and 3.4.3). The white contours shows the extent of the source determined by the BLOBCAT algorithm. The black cross marks the peak position, which we used to determine the spectral index. The small images show each spectral window separately, and the top panel presents the peak intensity for each spectral window and the corresponding spectral index fit. In each image the synthesized beam is given in the lower left corner.

A.2 Completeness maps

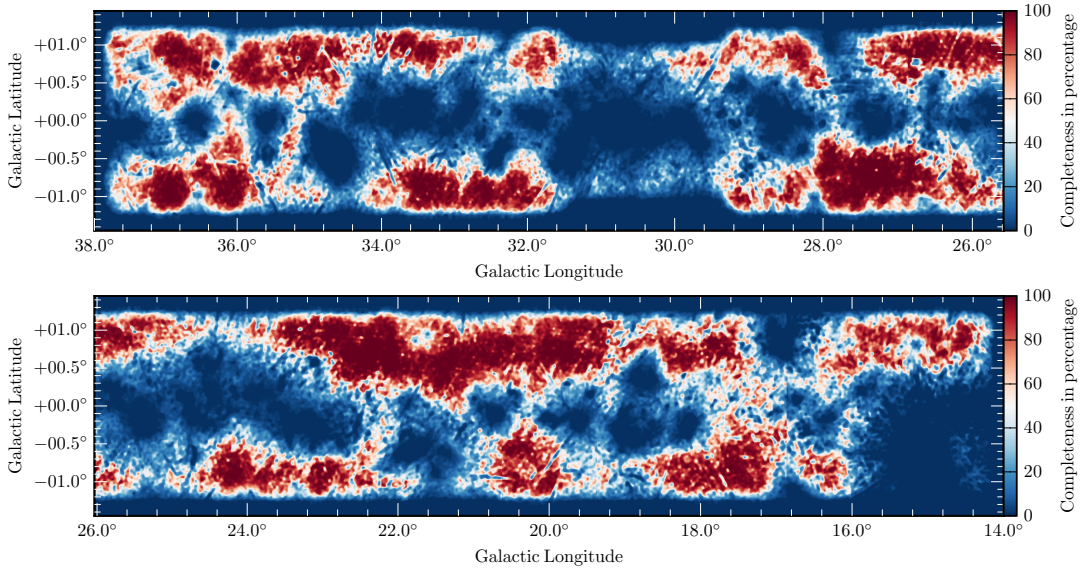

Figure A.3: Completeness map in percentage for sources with a peak intensity of 2 mJy beam^{-1}.

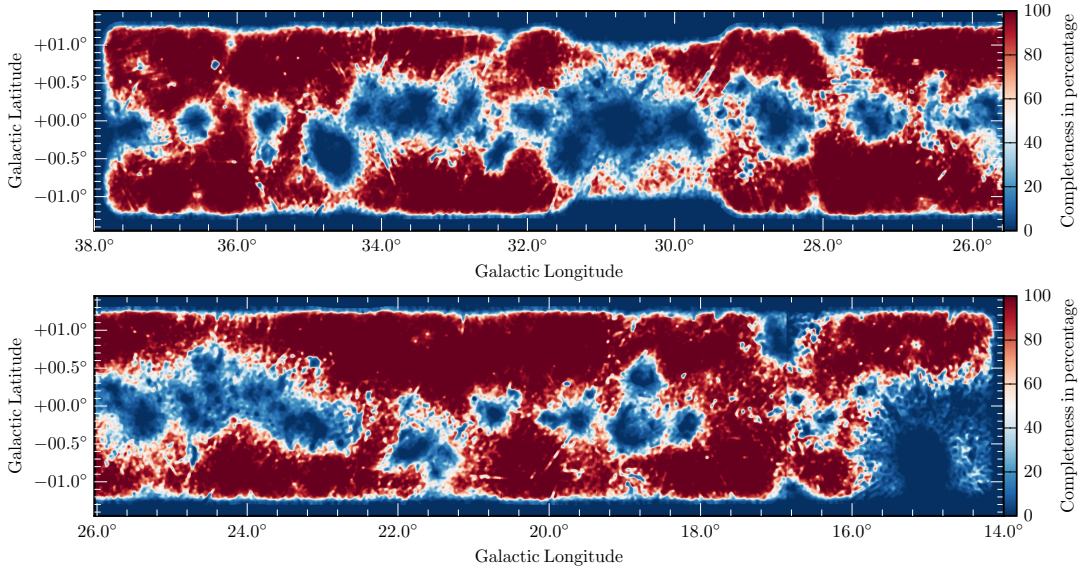

Figure A.4: Completeness map in percentage for sources with a peak intensity of 3 mJy beam^{-1}.

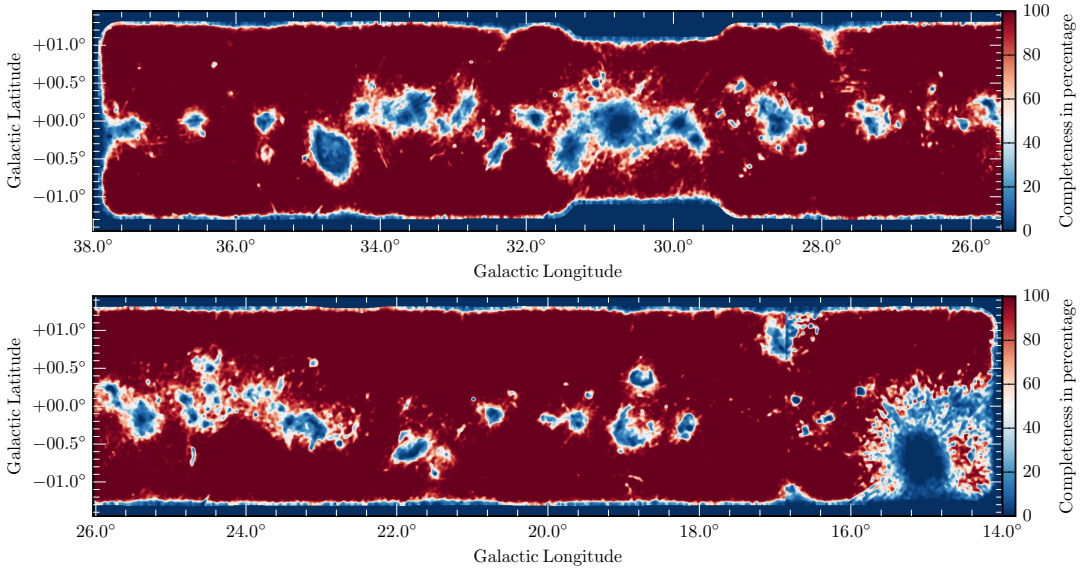

Figure A.5: Completeness map in percentage for sources with a peak intensity of 5 mJy beam^{-1}.

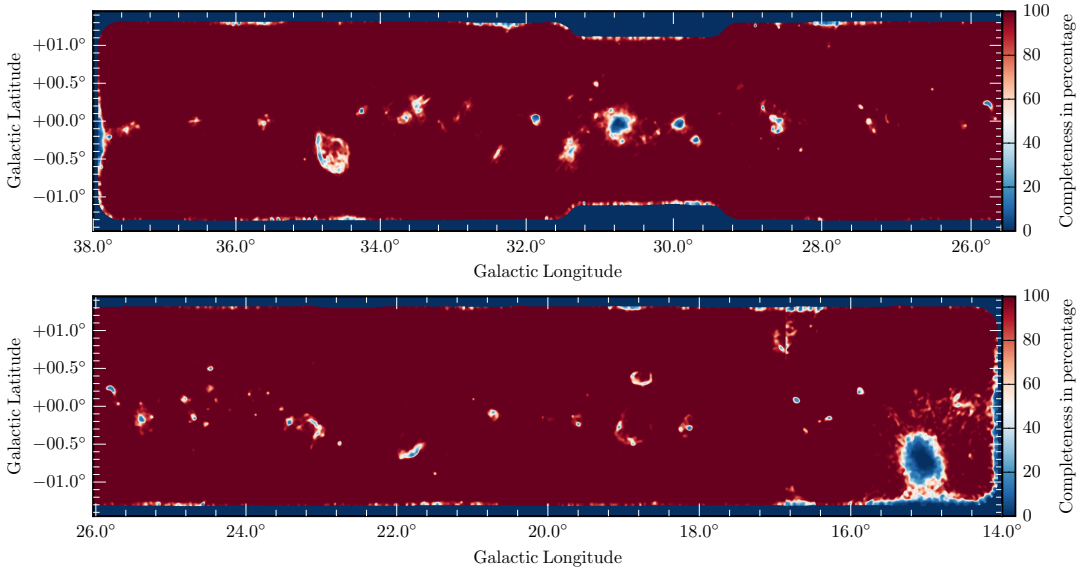

Figure A.6: Completeness map in percentage for sources with a peak intensity of 10 mJy beam^{-1}.

A.3 Example Table

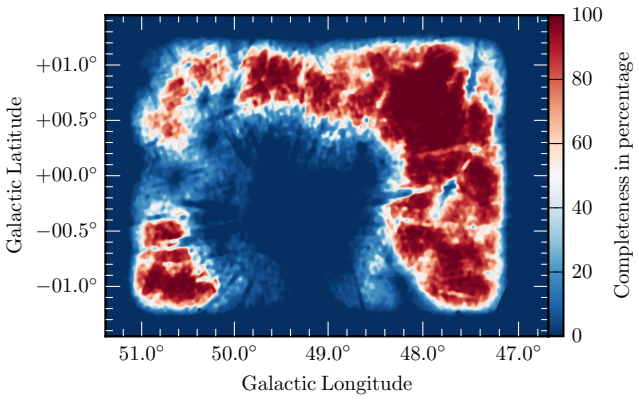

Figure A.7: Completeness map in percentage for sources with a peak intensity of 2 mJy beam^{-1}.

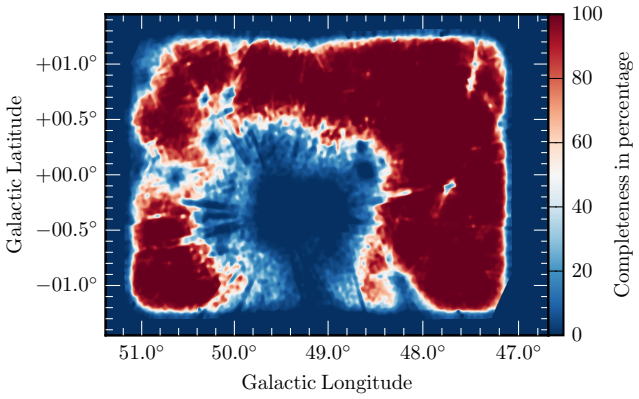

Figure A.8: Completeness map in percentage for sources with a peak intensity of 3 mJy beam^{-1}.

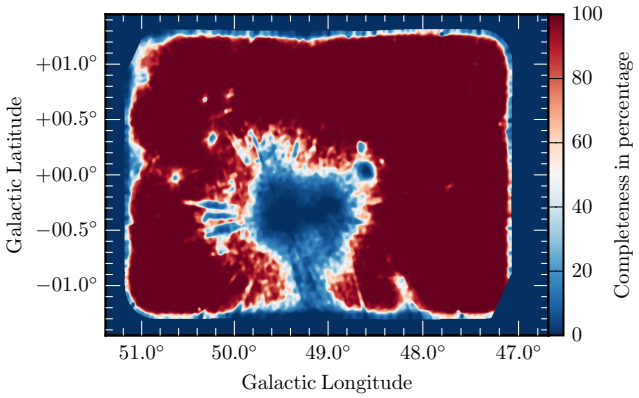

Figure A.9: Completeness map in percentage for sources with a peak intensity of 5 mJy beam^{-1}.

Table A.1: Example sources of the published catalog.

Galactic_ID	RA	Dec	S_p [mJyb^{-1}]	SNR	S_int [mJy]	BMAJ [″]	BMIN [″]	BPA [°]	n_pix	ext.	art.	S_p 1.06 GHz [mJy b^{-1}]	S_p 1.31 GHz [mJy b^{-1}]	S_p 1.44 GHz [mJy b^{-1}]	S_p 1.69 GHz [mJy b^{-1}]	S_p 1.82 GHz [mJy b^{-1}]	S_p 1.95 GHz [mJy b^{-1}]	alpha alpha	delta _alpha	reliable _alpha
G16.008+1.022	274.0295	-14.5304	4.00	13.48	3.81	18.1	11.1	-176.7	62	0	0	8.61	6.07	5.44	5.27	2.73	3.08	-1.88	0.37	1
G16.026-0.804	275.7037	-15.3780	12.56	27.21	10.12	18.1	11.1	-176.7	64	0	0	19.46	12.56	12.70	10.06	8.88	6.22	-1.58	0.17	1
G16.029-0.663	275.5768	-15.3096	57.97	101.00	56.76	18.1	11.1	-176.7	121	0	0	78.56	63.65	63.81	56.13	53.21	50.58	-0.71	0.05	1
G16.055+0.828	274.2288	-14.5818	9.92	24.01	7.64	18.1	11.1	-176.7	59	0	0	6.43	9.15	8.75	7.56	8.69	6.69	-0.05	0.26	1
G16.059+0.875	274.1880	-14.5560	5.97	15.21	6.97	18.1	11.1	-176.7	98	0	0	9.28	5.04	6.29	4.47	4.57	4.02	-1.30	0.37	1
G16.070+0.346	274.6755	-14.7975	2.36	7.05	2.23	18.1	11.1	-176.7	39	0	1	3.16	2.21	3.68	-0.14	1.30	1.32	0.00	0.00	0
G16.079+0.812	274.2554	-14.5680	3.42	9.09	2.61	18.1	11.1	-176.7	37	0	0	3.46	0.82	3.24	1.75	2.88	1.71	-0.50	1.61	0
G16.084+0.042	274.9591	-14.9290	11.56	17.24	16.22	18.1	11.1	-176.7	113	1	0	23.84	15.58	16.39	12.55	10.79	10.41	-1.37	0.24	1
G16.086-0.546	275.4974	-15.2042	2.77	8.36	4.34	18.1	11.1	-176.7	74	1	1	9.11	-0.64	2.37	3.30	2.97	3.46	-1.78	0.68	0
G16.091+0.333	274.6970	-14.7844	3.58	9.75	3.89	18.1	11.1	-176.7	57	0	0	4.77	5.34	4.78	3.38	3.42	2.33	-1.46	1.70	0
G16.101+0.472	274.5751	-14.7105	2.27	8.53	1.94	18.1	11.1	-176.7	41	0	0	6.50	2.66	2.43	0.50	1.84	2.82	-1.29	0.63	0
G16.106-0.984	275.9083	-15.3919	13.10	30.19	14.38	18.1	11.1	-176.7	95	0	0	21.56	14.68	14.93	12.35	14.24	13.22	-0.56	0.23	1
G16.112+0.702	274.3722	-14.5913	7.47	20.13	5.72	18.1	11.1	-176.7	55	0	0	10.07	6.94	7.45	5.96	5.31	3.50	-1.31	0.27	1
G16.125+1.227	273.9014	-14.3305	5.73	9.45	5.66	18.1	11.1	-176.7	50	0	0	11.28	8.18	6.72	5.17	5.63	6.13	-1.44	0.65	0
G16.134+0.399	274.6576	-14.7155	4.16	14.27	2.94	18.1	11.1	-176.7	43	0	0	6.60	5.71	4.63	4.00	1.85	2.79	-1.28	0.46	0
G16.134+0.491	274.5744	-14.6717	4.91	17.16	3.98	18.1	11.1	-176.7	57	0	0	7.03	5.70	5.44	3.01	3.19	2.10	-1.79	0.34	1
G16.140-0.126	275.1396	-14.9589	2.20	5.49	1.62	18.1	11.1	-176.7	23	0	1	4.08	1.87	2.67	2.21	0.60	-0.04	0.00	0.00	0
G16.143+0.041	274.9893	-14.8769	7.18	14.44	5.12	18.1	11.1	-176.7	45	0	0	12.19	6.49	6.01	4.40	4.57	3.83	-2.09	0.59	0
G16.145+0.009	275.0188	-14.8901	9.50	17.45	8.14	18.1	11.1	-176.7	62	0	0	6.98	4.07	7.31	9.86	11.06	7.90	0.48	0.66	0
G16.155-0.432	275.4266	-15.0891	1.87	6.24	1.67	18.1	11.1	-176.7	32	0	1	2.73	-0.66	2.75	1.08	1.22	2.66	-0.11	1.34	0
G16.170+0.953	274.1720	-14.4219	3.97	11.65	3.05	18.1	11.1	-176.7	43	0	0	6.62	7.24	4.11	4.32	2.89	2.74	-1.49	0.44	1
G16.174+0.411	274.6663	-14.6746	1.61	5.84	1.43	18.1	11.1	-176.7	31	0	1	2.56	1.88	1.81	1.33	1.19	-0.89	0.00	0.00	0
G16.178-0.711	275.6924	-15.2003	2.39	8.01	1.74	18.1	11.1	-176.7	33	0	0	-1.97	6.91	2.87	1.34	1.13	0.75	0.00	0.00	0
G16.182+0.950	274.1799	-14.4122	3.16	8.97	2.45	18.1	11.1	-176.7	39	0	0	5.25	5.15	3.36	4.76	2.27	2.96	-1.09	0.59	0
G16.183+0.768	274.3460	-14.4975	36.51	72.15	35.82	18.1	11.1	-176.7	113	0	0	57.19	44.52	41.68	32.96	32.53	28.64	-1.10	0.06	1
G33.118-1.244	284.0517	-0.4017	4.61	5.86	7.68	13.5	13.3	3.3	51	1	1	9.64	7.78	7.09	6.05	4.63	4.70	-1.01	0.64	0
G33.120-0.894	283.7413	-0.2406	23.31	93.90	23.35	13.5	13.3	3.3	134	0	0	23.04	23.44	23.27	23.87	22.79	22.14	-0.05	0.06	1
G33.128+0.364	282.6253	0.3406	8.34	20.43	7.06	13.5	13.3	3.3	60	0	0	8.67	9.58	6.72	7.72	7.60	5.90	-0.39	0.32	1
G33.133-0.093	283.0336	0.1365	187.07	176.50	315.49	13.5	13.3	3.3	392	1	0	130.33	165.62	179.39	206.92	219.04	228.78	0.86	0.03	1
G33.141-0.673	283.5538	-0.1212	2.45	12.09	2.56	13.5	13.3	3.3	63	0	0	4.45	4.41	2.60	2.04	2.08	2.64	-1.66	0.60	0
G33.143-0.066	283.0149	0.1580	122.27	111.80	230.62	13.5	13.3	3.3	252	1	0	239.64	190.95	180.47	157.71	144.76	128.60	-0.96	0.03	1
G33.144-0.041	282.9927	0.1698	5.22	5.02	29.34	13.5	13.3	3.3	135	1	1	11.20	9.81	12.98	15.20	12.06	7.48	-0.54	0.62	0
G33.147+1.044	282.0287	0.6677	26.93	104.80	26.50	13.5	13.3	3.3	114	0	0	30.77	28.74	27.58	26.61	25.98	24.19	-0.36	0.06	1
G33.149-0.557	283.4545	-0.0607	1.61	5.32	1.13	13.5	13.3	3.3	19	0	1	1.82	0.98	1.24	0.08	0.61	0.29	0.00	0.00	0
G33.166-1.165	284.0031	-0.3233	6.74	23.71	6.30	13.5	13.3	3.3	69	0	0	8.94	9.12	6.65	6.63	5.77	6.75	-0.73	0.23	1
G33.169-0.016	282.9816	0.2038	7.30	7.41	339.77	13.5	13.3	3.3	2011	1	0	22.89	22.11	23.00	20.30	20.61	19.51	-0.30	0.28	1
G33.176+0.424	282.5933	0.4101	2.34	5.56	4.74	13.5	13.3	3.3	56	1	1	5.19	4.40	4.72	2.68	2.62	1.80	-0.31	1.13	0
G33.176-0.196	283.1454	0.1275	6.55	9.84	7.09	13.5	13.3	3.3	50	0	0	11.93	8.06	8.66	7.00	5.04	5.57	-1.18	0.47	0
G33.181-0.187	283.1392	0.1365	4.22	6.45	5.61	13.5	13.3	3.3	44	1	1	6.54	3.79	5.27	3.46	4.14	2.85	0.00	0.00	0

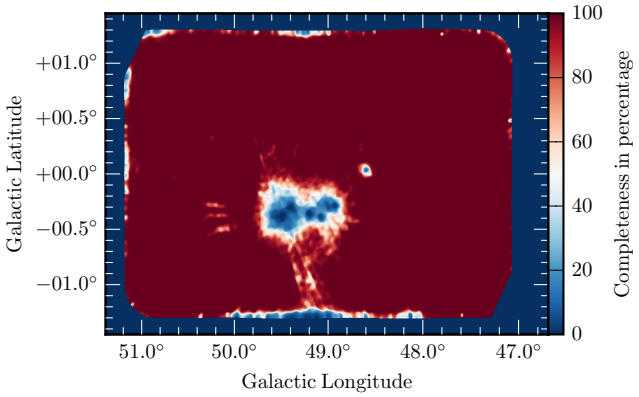

Figure A.10: Completeness map in percentage for sources with a peak intensity of 10 mJy beam^{-1}.

A.4 SNR Green and THOR comparison

Table A.2: Matching of SNR between the THOR catalog and the SNR catalog presented by Green (2014).

Galactic_ID	art.	npix	res.	α	$\Delta\alpha$	SNR name[1]	SNR α[2]
G15.913+0.183	0	2753	1	-0.78	0.07	G015.9+00.2	-0.63
G15.907+0.233	0	1303	1	-1.12	0.20	G015.9+00.2	-0.63
G16.742+0.088	0	9207	1	-0.17	0.10	G016.7+00.1	-0.60
G17.030-0.069	1	620	1	-0.19	0.64	G017.0-00.0	-0.50
G17.448-0.063	0	815	1	0.19	0.29	G017.4-00.1	-0.70
G18.107-0.134	0	760	1	-0.72	0.25	G018.1-00.1	-0.50
G18.193-0.174	0	6624	1	-0.37	0.10	G018.1-00.1	-0.50
G18.171-0.213	0	952	1	-0.68	0.25	G018.1-00.1	-0.50
G18.128-0.218	1	35	0	-0.36	1.37	G018.1-00.1	-0.50
G18.610-0.316	0	2784	1	0.17	0.22	G018.6-00.2	-0.40
G18.761+0.287	0	31818	1	-1.07	0.06	G018.8+00.3	-0.46
G18.908-0.922	1	1385	1	-0.85	0.27	G018.9-01.1	-0.39
G19.954-0.250	0	162	1	-1.12	0.34	G020.0-00.2	-0.10
G19.952-0.169	0	11499	1	-0.32	0.19	G020.0-00.2	-0.10
G20.075-0.181	0	913	1	-1.30	0.44	G020.0-00.2	-0.10
G20.502+0.155	0	1411	1	-0.54	0.33	G020.4+00.1	-0.10
G21.503-0.884	0	1713	1	-0.02	0.00	G021.6-00.8	-0.50
G21.765-0.631	0	49153	1	-0.77	0.04	G021.8-00.6	-0.56
G21.948-0.416	0	11333	1	-0.31	0.12	G021.8-00.6	-0.56
G23.124-0.199	0	6108	1	-1.13	0.11	G023.3-00.3	-0.50
G23.015-0.288	0	3619	1	-1.27	0.10	G023.3-00.3	-0.50
G23.105-0.411	0	10590	1	0.04	0.18	G023.3-00.3	-0.50
G23.062-0.376	0	935	1	-1.71	0.18	G023.3-00.3	-0.50
G23.539+0.268	0	9419	1	-0.26	0.39	G023.6+00.3	-0.30
G24.664+0.620	0	2798	1	-0.66	0.22	G024.7+00.6	-0.20
G24.689-0.589	0	7553	1	-0.73	0.17	G024.7-00.6	-0.50
G27.365+0.014	0	8048	1	-0.49	0.04	G027.4+00.0	-0.68
G28.610-0.142	0	2919	1	-0.79	0.04	G028.6-00.1	–
G28.672-0.108	0	2015	1	-0.64	0.04	G028.6-00.1	–
G29.567+0.094	0	468	1	-0.36	0.54	G029.6+00.1	-0.50
G29.689-0.242	0	5059	1	-0.64	0.01	G029.7-00.3	-0.63
G31.869+0.064	0	18727	1	-0.32	0.01	G031.9+00.0	–
G32.423+0.079	0	60	0	-0.74	0.44	G032.4+00.1	–
G32.415+0.076	1	37	1	0.00	0.00	G032.4+00.1	–
G32.929+0.021	0	5006	1	-0.74	0.24	G032.8-00.1	-0.20
G33.748+0.025	0	3496	1	-1.04	0.13	G033.6+00.1	-0.51
G33.651+0.051	0	6623	1	-0.82	0.10	G033.6+00.1	-0.51
G33.667+0.100	0	3002	1	-1.31	0.15	G033.6+00.1	-0.51
G33.607+0.089	0	2519	1	-0.55	0.19	G033.6+00.1	-0.51
G34.588-0.238	0	8050	1	-0.34	0.06	G034.7-00.4	-0.37
G34.568-0.630	0	11697	1	-1.01	0.07	G034.7-00.4	-0.37
G34.834-0.439	0	45702	1	-0.80	0.06	G034.7-00.4	-0.37
G34.681-0.635	0	3854	1	-1.06	0.09	G034.7-00.4	-0.37
G35.583-0.448	0	37	0	-3.09	1.18	G035.6-00.4	-0.50
G35.602-0.548	1	226	1	0.55	0.94	G035.6-00.4	-0.50
G49.016-0.731	0	4875	1	-1.05	0.12	G049.2-00.7	-0.30
G49.059-0.777	1	473	1	0.42	0.38	G049.2-00.7	-0.30
G49.190-0.801	0	5343	1	-1.00	0.21	G049.2-00.7	-0.30

Visually matched sources between the THOR catalog and the SNR catalog by Green (2014). The first six columns are taken from the THOR continuum catalog, whereas the last two columns are presented in Green (2014). As the SNR are very clumpy, we find several THOR continuum sources, which overlap with the same SNR.

[1] Following the naming in Green (2014).

[2] Taken from Green (2014). The spectral index in Green (2014) is negatively defined, and we adapt the values according to our definition of the spectral index.

References

Aguirre, J. E., Ginsburg, A. G., Dunham, M. K., et al. 2011, ApJS, 192, 4

Anderson, L. D., Bania, T. M., Balser, D. S., et al. 2014, ApJS, 212, 1

Anderson, L. D., Bania, T. M., Balser, D. S., & Rood, R. T. 2011, ApJS, 194, 32

André, P., Di Francesco, J., Ward-Thompson, D., et al. 2014, Protostars and Planets VI, 27

Ballesteros-Paredes, J., Vázquez-Semadeni, E., Gazol, A., et al. 2011, MNRAS, 416, 1436

Bally, J., Anderson, L. D., Battersby, C., et al. 2010, A&A, 518, L90

Benjamin, R. A., Churchwell, E., Babler, B. L., et al. 2003, PASP, 115, 953

Berry, D. S. 2015, Astronomy and Computing, 10, 22

Beuther, H., Ragan, S. E., Ossenkopf, V., et al. 2014, A&A, 571, A53

Beuther, H., Tackenberg, J., Linz, H., et al. 2012, A&A, 538, A11

Bhatnagar, S., Rau, U., & Golap, K. 2013, ApJ, 770, 91

Bhatnagar, S., Rau, U., Green, D. A., & Rupen, M. P. 2011, ApJ, 739, L20

Bihr, S., Beuther, H., Ott, J., et al. 2015, A&A, 580, A112

Bihr, S., Johnston, K. G., Beuther, H., et al. 2016, ArXiv e-prints

Blum, R. D., Damineli, A., & Conti, P. S. 1999, AJ, 117, 1392

Brunthaler, A., Falcke, H., Bower, G. C., et al. 2005, A&A, 435, 497

Brunthaler, A., Reid, M. J., Menten, K. M., et al. 2011, Astronomische Nachrichten, 332, 461

Burkhart, B., Lee, M.-Y., Murray, C. E., & Stanimirović, S. 2015, ApJ, 811, L28

Caldú-Primo, A., Schruba, A., Walter, F., et al. 2013, AJ, 146, 150

Carey, S. J., Noriega-Crespo, A., Mizuno, D. R., et al. 2009, PASP, 121, 76

Carlhoff, P., Nguyen Luong, Q., Schilke, P., et al. 2013, A&A, 560, A24

Chengalur, J. N., Kanekar, N., & Roy, N. 2013, MNRAS, 432, 3074

Churchwell, E., Babler, B. L., Meade, M. R., et al. 2009, PASP, 121, 213

Clark, B. G. 1965, ApJ, 142, 1398

Condon, J. J., Cotton, W. D., Greisen, E. W., et al. 1998, AJ, 115, 1693

Cornwell, T. J. 2008, IEEE Journal of Selected Topics in Signal Processing, 2, 793

Cornwell, T. J., Golap, K., & Bhatnagar, S. 2008, IEEE Journal of Selected Topics in Signal Processing, 2, 647

Csengeri, T., Urquhart, J. S., Schuller, F., et al. 2014, A&A, 565, A75

Dickey, J. M., McClure-Griffiths, N., Gibson, S. J., et al. 2013, PASA, 30, 3

Dobbs, C. L., Krumholz, M. R., Ballesteros-Paredes, J., et al. 2014, Protostars and Planets VI, 3

Dobbs, C. L., Pringle, J. E., & Burkert, A. 2012, MNRAS, 425, 2157

Draine, B. T. 2011, Physics of the Interstellar and Intergalactic Medium (Princeton University Press)

Dubner, G. & Giacani, E. 2015, A&A Rev., 23, 3

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

Fukui, Y., Torii, K., Onishi, T., et al. 2015, ApJ, 798, 6

Gibson, S. J., Taylor, A. R., Higgs, L. A., Brunt, C. M., & Dewdney, P. E. 2005a, ApJ, 626, 195

Gibson, S. J., Taylor, A. R., Higgs, L. A., Brunt, C. M., & Dewdney, P. E. 2005b, ApJ, 626, 214

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. & Clark, P. C. 2012, MNRAS, 421, 9

Goldsmith, P. F. & Li, D. 2005, ApJ, 622, 938

Goodman, A. A., Alves, J., Beaumont, C. N., et al. 2014, ApJ, 797, 53

Gould, R. J. 1994, ApJ, 423, 522

Green, D. A. 2014, Bulletin of the Astronomical Society of India, 42, 47

Grobler, T. L., Nunhokee, C. D., Smirnov, O. M., van Zyl, A. J., & de Bruyn, A. G. 2014, MNRAS, 439, 4030

Hales, C. A., Murphy, T., Curran, J. R., et al. 2012, MNRAS, 425, 979

Hancock, P. J., Murphy, T., Gaensler, B. M., Hopkins, A., & Curran, J. R. 2012, MNRAS, 422, 1812

Heiles, C. & Troland, T. H. 2003a, ApJS, 145, 329

Heiles, C. & Troland, T. H. 2003b, ApJ, 586, 1067

Heitsch, F., Hartmann, L. W., Slyz, A. D., Devriendt, J. E. G., & Burkert, A. 2008, ApJ, 674, 316

Helfand, D. J., Becker, R. H., White, R. L., Fallon, A., & Tuttle, S. 2006, AJ, 131, 2525

Hill, T., Motte, F., Didelon, P., et al. 2011, A&A, 533, A94

Hoare, M. G., Purcell, C. R., Churchwell, E. B., et al. 2012, PASP, 124, 939

Hollenbach, D. J. & Tielens, A. G. G. M. 1997, ARA&A, 35, 179

Immer, K., Brunthaler, A., Reid, M. J., et al. 2011, ApJS, 194, 25

Irwin, J. A. 2007, Astrophysics: Decoding the Cosmos (Wiley-VCH Verlag)

Jackson, J. M., Finn, S. C., Chambers, E. T., Rathborne, J. M., & Simon, R. 2010, ApJ, 719, L185

Jackson, J. M., Rathborne, J. M., Foster, J. B., et al. 2013, PASA, 30, 57

Jackson, J. M., Rathborne, J. M., Shah, R. Y., et al. 2006, ApJS, 163, 145

Johanson, A. K. & Kerton, C. R. 2009, AJ, 138, 1615

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

Kainulainen, J., Federrath, C., & Henning, T. 2014, Science, 344, 183

Kainulainen, J., Ragan, S. E., Henning, T., & Stutz, A. 2013, A&A, 557, A120

Kainulainen, J. & Tan, J. C. 2013, A&A, 549, A53

Kalberla, P. M. W. & Kerp, J. 2009, ARA&A, 47, 27

Kavars, D. W., Dickey, J. M., McClure-Griffiths, N. M., Gaensler, B. M., & Green, A. J. 2003, ApJ, 598, 1048

Kavars, D. W., Dickey, J. M., McClure-Griffiths, N. M., Gaensler, B. M., & Green, A. J. 2005, ApJ, 626, 887

Keto, E. 2003, ApJ, 599, 1196

Klessen, R. S. 2011, in EAS Publications Series, Vol. 51, EAS Publications Series, ed. C. Charbonnel & T. Montmerle, 133–167

Klessen, R. S. & Glover, S. C. O. 2014, ArXiv e-prints

Kolpak, M. A., Jackson, J. M., Bania, T. M., & Dickey, J. M. 2002, ApJ, 578, 868

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

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

Krčo, M. & Goldsmith, P. F. 2010, ApJ, 724, 1402

Lada, C. J., Alves, J. F., & Lombardi, M. 2007, Protostars and Planets V, 3

Lee, M.-Y., Stanimirović, S., Douglas, K. A., et al. 2012, ApJ, 748, 75

Lee, M.-Y., Stanimirović, S., Murray, C. E., Heiles, C., & Miller, J. 2015, ApJ, 809, 56

Lester, D. F., Dinerstein, H. L., Werner, M. W., et al. 1985, ApJ, 296, 565

Li, D. & Goldsmith, P. F. 2003, ApJ, 585, 823

Liszt, H. S., Braun, R., & Greisen, E. W. 1993, AJ, 106, 2349

Lombardi, M., Alves, J., & Lada, C. J. 2015, A&A, 576, L1

Louvet, F., Motte, F., Hennebelle, P., et al. 2014, A&A, 570, A15

Lucas, P. W., Hoare, M. G., Longmore, A., et al. 2008, MNRAS, 391, 136

Mac Low, M.-M. & Glover, S. C. O. 2012, ApJ, 746, 135

Mac Low, M.-M. & Klessen, R. S. 2004, Reviews of Modern Physics, 76, 125
McClure-Griffiths, N. M., Dickey, J. M., Gaensler, B. M., et al. 2012, ApJS, 199, 12
McClure-Griffiths, N. M., Dickey, J. M., Gaensler, B. M., Green, A. J., & Haverkorn, M. 2006, ApJ, 652, 1339
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
Meisenheimer, K. 1999, in Lecture Notes in Physics, Berlin Springer Verlag, Vol. 530, The Radio Galaxy Messier 87, ed. H.-J. Röser & K. Meisenheimer, 188
Mezger, P. G. & Henderson, A. P. 1967, ApJ, 147, 471
Milam, S. N., Savage, C., Brewster, M. A., Ziurys, L. M., & Wyckoff, S. 2005, ApJ, 634, 1126
Mogotsi, K. M., de Blok, W. J. G., Caldú-Primo, A., et al. 2016, AJ, 151, 15
Molinari, S., Swinyard, B., Bally, J., et al. 2010, A&A, 518, L100
Motte, F., Nguyên Luong, Q., Schneider, N., et al. 2014, A&A, 571, A32
Motte, F., Schilke, P., & Lis, D. C. 2003, ApJ, 582, 277
Murray, C. E., Lindner, R. R., Stanimirović, S., et al. 2014, ApJ, 781, L41
Nguyen-Lu'o'ng, Q., Motte, F., Carlhoff, P., et al. 2013, ApJ, 775, 88
Nguyen Luong, Q., Motte, F., Schuller, F., et al. 2011, A&A, 529, A41
Nielbock, M., Launhardt, R., Steinacker, J., et al. 2012, A&A, 547, A11
Offner, S. S. R., Clark, P. C., Hennebelle, P., et al. 2014, Protostars and Planets VI, 53
Oka, T., Hasegawa, T., Sato, F., et al. 2001, ApJ, 562, 348
Peek, J. E. G., Heiles, C., Douglas, K. A., et al. 2011, ApJS, 194, 20
Pineda, J. L., Goldsmith, P. F., Chapman, N., et al. 2010, ApJ, 721, 686
Pineda, J. L., Langer, W. D., Velusamy, T., & Goldsmith, P. F. 2013, A&A, 554, A103
Purcell, C. R., Hoare, M. G., Cotton, W. D., et al. 2013, ApJS, 205, 1
Purcell, C. R., Longmore, S. N., Walsh, A. J., et al. 2012, MNRAS, 426, 1972
Radhakrishnan, V., Murray, J. D., Lockhart, P., & Whittle, R. P. J. 1972, ApJS, 24, 15
Ragan, S. E., Henning, T., Tackenberg, J., et al. 2014, A&A, 568, A73
Rau, U., Bhatnagar, S., & Owen, F. N. 2014, ArXiv e-prints
Reid, M. J., Menten, K. M., Brunthaler, A., et al. 2014, ApJ, 783, 130
Reynoso, E. M. & Walsh, A. J. 2015, MNRAS, 451, 3044
Rich, J. W., de Blok, W. J. G., Cornwell, T. J., et al. 2008, AJ, 136, 2897
Roman-Duval, J., Jackson, J. M., Heyer, M., Rathborne, J., & Simon, R. 2010, ApJ, 723, 492
Rosolowsky, E., Dunham, M. K., Ginsburg, A., et al. 2010, ApJS, 188, 123
Roy, N., Kanekar, N., Braun, R., & Chengalur, J. N. 2013a, MNRAS, 436, 2352
Roy, N., Kanekar, N., & Chengalur, J. N. 2013b, MNRAS, 436, 2366
Rybicki, G. B. & Lightman, A. P. 1979, Radiative processes in astrophysics

Sadavoy, S. I., Di Francesco, J., André, P., et al. 2014, ApJ, 787, L18

Schinnerer, E., Meidt, S. E., Pety, J., et al. 2013, ApJ, 779, 42

Schneider, N., Bontemps, S., Girichidis, P., et al. 2015a, MNRAS, 453, 41

Schneider, N., Ossenkopf, V., Csengeri, T., et al. 2015b, A&A, 575, A79

Schuller, F., Menten, K. M., Contreras, Y., et al. 2009, A&A, 504, 415

Smith, R. J., Glover, S. C. O., Clark, P. C., Klessen, R. S., & Springel, V. 2014, MNRAS, 441, 1628

Stahler, S. W., Palla, F., & Palla, F. 2005, The Formation of Stars (Physics Textbook), 1st edn. (Wiley-VCH)

Stil, J. M., Taylor, A. R., Dickey, J. M., et al. 2006, AJ, 132, 1158

Strasser, S. & Taylor, A. R. 2004, ApJ, 603, 560

Strasser, S. T., Dickey, J. M., Taylor, A. R., et al. 2007, AJ, 134, 2252

Stutz, A. M. & Kainulainen, J. 2015, A&A, 577, L6

Taylor, A. R., Gibson, S. J., Peracaula, M., et al. 2003, AJ, 125, 3145

Taylor, G. B., Carilli, C. L., & Perley, R. A., eds. 1999, Astronomical Society of the Pacific Conference Series, Vol. 180, Synthesis Imaging in Radio Astronomy II

Vallée, J. P. 2008, AJ, 135, 1301

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

Vázquez-Semadeni, E., Ryu, D., Passot, T., González, R. F., & Gazol, A. 2006, ApJ, 643, 245

Vollmer, B., Krichbaum, T. P., Angelakis, E., & Kovalev, Y. Y. 2008, A&A, 489, 49

Walch, S., Girichidis, P., Naab, T., et al. 2015, MNRAS, 454, 238

Walsh, A. J., Beuther, H., Bihr, S., et al. 2016, MNRAS, 455, 3494

Walsh, A. J., Breen, S. L., Britton, T., et al. 2011, MNRAS, 416, 1764

Wang, K., Testi, L., Ginsburg, A., et al. 2015, MNRAS, 450, 4043

Watson, C., Povich, M. S., Churchwell, E. B., et al. 2008, ApJ, 681, 1341

Williams, J. P., de Geus, E. J., & Blitz, L. 1994, ApJ, 428, 693

Wilson, T. L., Rohlfs, K., & Hüttemeister, S. 2010, Tools of Radio Astronomy (Astronomy and Astrophysics Library), softcover reprint of hardcover 5th ed. 2009 edn. (Springer)

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, B., Moscadelli, L., Sato, M., et al. 2014, ApJ, 781, 89