*Astronomy & Astrophysics* manuscript no. hisa_filament
January 16, 2020


# Cloud formation in the atomic and molecular phase: H I self absorption (HISA) towards a giant molecular filament⋆

Y. Wang[1], S. Bihr[1], H. Beuther[1], M. R. Rugel[1,2], J. D. Soler[1], J. Ott[3], J. Kainulainen[4], N. Schneider[5], R. S. Klessen[6,7], S. C. O. Glover[6], N. M. McClure-Griffiths[8], P. F. Goldsmith[9], K. G. Johnston[10], K. M. Menten[2], S. Ragan[11], L. D. Anderson[12,13,14], J. S. Urquhart[15], H. Linz[1], N. Roy[16], R. J. Smith[17], F. Bigiel[18], T. Henning[1], and S. N. Longmore[19]

[1] Max Planck Institute for Astronomy, Königstuhl 17, 69117 Heidelberg, Germany
e-mail: wang@mpia.de
[2] Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, 53121 Bonn, Germany
[3] National Radio Astronomy Observatory, P.O. Box O, 1003 Lopezville Road, Socorro, NM 87801, USA
[4] Chalmers University of Technology, Department of Space, Earth and Environment, SE-412 93 Gothenburg, Sweden
[5] I. Physik. Institut, University of Cologne, Zülpicher Str.77, 50937 Cologne, Germany
[6] Universität Heidelberg, Zentrum für Astronomie, Institut für Theoretische Astrophysik, Albert-Ueberle-Str. 2, 69120 Heidelberg, Germany
[7] Universität Heidelberg, Interdisziplinäres Zentrum für Wissenschaftliches Rechnen, INF 205, 69120, Heidelberg, Germany
[8] Research School of Astronomy and Astrophysics, The Australian National University, Canberra, ACT, Australia
[9] Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA
[10] School of Physics and Astronomy, University of Leeds, Leeds LS2 9JT, UK
[11] School of Physics and Astronomy, Cardiff University, Queen's Buildings, The Parade, Cardiff, CF24 3AA, UK
[12] Department of Physics and Astronomy, West Virginia University, Morgantown, WV 26506, USA
[13] Adjunct Astronomer at the Green Bank Observatory, P.O. Box 2, Green Bank WV 24944, USA
[14] Center for Gravitational Waves and Cosmology, West Virginia University, Chestnut Ridge Research Building, Morgantown, WV 26505, USA
[15] School of Physical Sciences, University of Kent, Ingram Building, Canterbury, Kent CT2 7NH, UK
[16] Department of Physics, Indian Institute of Science, Bengaluru 560012, India
[17] Jodrell Bank Centre for Astrophysics, School of Physics and Astronomy, The University of Manchester, Oxford Road, Manchester, M13 9PL, UK
[18] Argelander Institut für Astronomie, Auf dem Hügel 71, 53121 Bonn, Germany
[19] Astrophysics Research Institute, Liverpool John Moores University, IC2, Liverpool Science Park, 146 Brownlow Hill, Liverpool L3 5RF, UK

Received dd, mm, yyyy; accepted dd, mm, yyyy

**ABSTRACT**

Molecular clouds form from the atomic phase of the interstellar medium. However, characterizing the transition between the atomic and the molecular interstellar medium (ISM) is a complex observational task. Here we address cloud formation processes by combining H I self absorption (HISA) with molecular line data. Column density probability density functions (N-PDFs) are a common tool for examining molecular clouds. One scenario proposed by numerical simulations is that the N-PDF evolves from a log-normal shape at early times to a power-law-like shape at later times. To date, investigations of N-PDFs have been mostly limited to the molecular component of the cloud. In this paper, we study the cold atomic component of the giant molecular filament GMF38.1-32.4a (GMF38a, distance=3.4 kpc, length~ 230 pc), calculate its N-PDFs, and study its kinematics. We identify an extended HISA feature, which is partly correlated with the $^{13}$CO emission. The peak velocities of the HISA and $^{13}$CO observations agree well on the eastern side of the filament, whereas a velocity offset of approximately 4 km s$^{-1}$ is found on the western side. The sonic Mach number we derive from the linewidth measurements shows that a large fraction of the HISA, which is ascribed to the cold neutral medium (CNM), is at subsonic and transonic velocities. The column density of the CNM part is on the order of $10^{20}$ to $10^{21}$ cm$^{-2}$. The column density of molecular hydrogen, traced by $^{13}$CO, is an order of magnitude higher. The N-PDFs from HISA (CNM), H I emission (the warm and cold neutral medium), and $^{13}$CO (molecular component) are well described by log-normal functions, which is in agreement with turbulent motions being the main driver of cloud dynamics. The N-PDF of the molecular component also shows a power law in the high column-density region, indicating self-gravity. We suggest that we are witnessing two different evolutionary stages within the filament. The eastern subregion seems to be forming a molecular cloud out of the atomic gas, whereas the western subregion already shows high column density peaks, active star formation, and evidence of related feedback processes.

**Key words.** ISM: clouds – ISM: atoms – ISM: molecules – Radio lines: ISM – Stars: formation

## 1. Introduction

Stars, one of the key components of our universe, form in molecular clouds which are composed mainly of molecular hydrogen

⋆ All the HISA data are available in fits format at the CDS via anonymous ftp to cdsarc.u-strasbg.fr (130.79.128.5) or via http://cdsweb.u-strasbg.fr/cgi-bin/qcat?J/A+A/.



Article published by EDP Sciences, to be cited as https://doi.org/10.1051/0004-6361/201935866

(e.g., Larson 2003; Stahler & Palla 2005; McKee & Ostriker 2007; Dobbs et al. 2014; Tan et al. 2014), yet the formation process of molecular clouds is still under debate. Various studies show molecular clouds form out of relatively diffuse atomic hydrogen gas (e.g., Larson 1981; Blitz et al. 2007; Clark et al. 2012; Hennebelle & Falgarone 2012; Dobbs et al. 2014; Sternberg et al. 2014; Klessen & Glover 2016). Different processes have been proposed (e.g., Hennebelle & Falgarone 2012; Dobbs et al. 2014; Klessen & Glover 2016). Basically, the atomic gas contracts and the increased column density shield the cloud from the interstellar UV radiation and cool down to form molecular gas. The cold neutral medium (CNM) with a typical temperature in the range of 40 to 100 K and volume density $> 10 \sim 100$ cm$^{-3}$ (McKee & Ostriker 1977; Wolfire et al. 1995; Wilson et al. 2010) is the key component connecting the diffuse atomic gas with the molecular gas. Thus, observational constraints on the physical properties of the CNM, such as density distribution and kinematics, are crucial to understanding the formation of the molecular cloud.

Although 21cm H I line emission offers a straightforward tool to study atomic hydrogen, it is difficult to determine the properties of the gas from which it arises. The main challenge is the coexistence of the warm neutral medium (WNM) and the CNM assumed to be in pressure equilibrium (McKee & Ostriker 1977; Wolfire et al. 1995, 2003). Studies of H I self absorption (HISA) overcome this problem, as they only trace the CNM. HISA was first detected by Heeschen (1954, 1955) towards the Galactic center. HISA features occur if cold, dense atomic hydrogen is in front of a warmer emission background (e.g., Knapp 1974). Since then, a number of observations that have been carried out with single dish telescopes and interferometers, and HISA features were found to be widespread in the Milky Way (e.g., Riegel & Crutcher 1972; Knapp 1974; Heiles & Gordon 1975; McCutcheon et al. 1978; Levinson & Brown 1980; Minn 1981; Shuter et al. 1987; van der Werf et al. 1988; van der Werf & Goss 1989; Montgomery et al. 1995; Gibson et al. 2000; Kavars et al. 2003; Gibson et al. 2005a; Kavars et al. 2005; Dénes et al. 2018). The spin temperature of the cold H I responsible for HISA ranges from ∼10-60 K, (e.g. Gibson et al. 2000; Kavars et al. 2005; McClure-Griffiths et al. 2006). A special case of the HISA features, so-called H I narrow self absorption (HINSA) features, were studied towards nearby molecular clouds, revealing small linewidths on the order of $\lesssim$1 km s$^{-1}$ (Li & Goldsmith 2003; Goldsmith & Li 2005; Krčo et al. 2008; Krčo & Goldsmith 2010; Zuo et al. 2018). However, studies characterizing the column density and the kinematic distribution of the CNM in large maps are still rare.

Recent observations have revealed a group of large ($\geq$ 100 pc) and massive ($\geq 10^5$ M$_\odot$) filaments, known as giant molecular filaments (GMFs), which may be linked to Galactic dynamics and trace the gravitational mid-plane in the Milky Way (MW) (Jackson et al. 2010; Goodman et al. 2014; Wang et al. 2015; Zucker et al. 2015; Abreu-Vicente et al. 2016; Li et al. 2016; Wang et al. 2016; Zucker et al. 2018; Zhang et al. 2019). These observations show that GMFs are the largest coherent gas structures in our Milky Way, and often contain different evolutionary stages of the star formation regions simultaneously in the same filament (Goodman et al. 2014; Zucker et al. 2015), which makes them ideal targets for studying the CNM properties in different environments that lead to molecular cloud formation.

A common tool to study molecular clouds are the probability density functions of the column density (N-PDFs) (see e.g., Ostriker et al. 2001; Lombardi et al. 2008; Kainulainen et al. 2009; Alves de Oliveira et al. 2014; Sadavoy et al. 2014; Abreu-Vicente et al. 2015; Stutz & Kainulainen 2015; Schneider et al. 2015; Lin et al. 2017; Chen et al. 2018). The shape of N-PDFs is predicted to depend on the physical processes acting within the cloud. In the early evolution of a molecular cloud, turbulent motions within the cloud dominate and the N-PDF reveals a log-normal shape. The width of the log-normal N-PDF is also determined by the turbulent motions (see e.g., Federrath et al. 2010; Ballesteros-Paredes et al. 2011; Kritsuk et al. 2011; Federrath & Klessen 2013; Burkhart et al. 2015a; Bialy et al. 2017b). In this scenario, more evolved clouds develop a high-density power-law tail, indicating that the cloud structure has evolved and gravity dominates. Observations indicate that star-forming clouds show such tails, lending support to this scenario (e.g., Kainulainen et al. 2009; Schneider et al. 2013). The slope of the power-law N-PDF can be related to evolutionary stages of the clouds with steeper slopes possibly indicating with earlier evolutionary stages (e.g., Kritsuk et al. 2011; Federrath & Klessen 2013; Ward et al. 2014).

High-mass star-forming regions reveal multiple power-laws, having a shallower slope for the highest density regions. This indicates a slower collapse for such regions (Schneider et al. 2015). Lombardi et al. (2015) and Alves et al. (2017) present a contrasting argument, reporting that all N-PDFs have a power-law shape and the log-normal shape could be an observational bias, a view point that has triggered considerable controversy (Ossenkopf-Okada et al. 2016; Chen et al. 2018; Körtgen et al. 2019a). Theoretical work and simulations of molecular clouds also reproduce N-PDFs in different forms (e.g., Vazquez-Semadeni 1994; Federrath et al. 2010; Federrath & Klessen 2012; Burkhart et al. 2015a). Burkhart et al. (2015b) and Imara & Burkhart (2016) studied nearby molecular clouds, and report H I N-PDFs with a log-normal shape, without any power-law tail. Rebolledo et al. (2017) studied the Carina and Gum 31 molecular complex, where the H I N-PDF also shows a log-normal shape.

To investigate the transition of atomic to molecular hydrogen in more detail, we examine the hydrogen content with HISA measurements in detail for GMF38.1-32.4a (GMF38a, Ragan et al. 2014). With a velocity range between 50 and 60 km s$^{-1}$ (Ragan et al. 2014), GMF38a is at a median distance of 3.4 kpc from the Sun (Galactocentric distance ∼5.9 kpc) estimated from the Bayesian Distance Estimator tool (Reid et al. 2016). The top panel of Fig. 1 shows the integrated $^{13}$CO emission from the GRS survey (Jackson et al. 2006). Our goal is to study the kinematics of this GMF in the molecular and atomic hydrogen traced by $^{13}$CO emission and HISA, respectively. Furthermore, we analyze N-PDFs for the atomic and molecular hydrogen, and compare their properties.

## 2. Observations and Methods

### 2.1. The H I 21 cm line and continuum

The H I 21 cm line observations of GMF38.1-32.4 are part of the H I, OH, Recombination line survey of the Milky Way (THOR; Beuther et al. 2016 and Wang et al. 2019). The survey observed a part of the first quadrant of the Galactic plane ($l = 14.0 - 67.4°$ and $|b| \leq 1.25°$) with the *Karl G. Jansky* Very Large Array (VLA) in C-configuration at L band from 1 to 2 GHz, covering the H I 21 cm line, 4 OH lines, 19 H$\alpha$ recombination lines, and eight continuum bands between 1 and 2 GHz (Beuther et al. 2016). Each pointing was observed for $4 \times 2$ minutes to ensure a uniform *uv*-coverage. The spectral window for the H I 21 cm line was set to have a bandwidth of 2 MHz ($\sim$ 400 km s$^{-1}$) and a spectral resolution of 3.91 kHz ($\sim$ 0.82 km s$^{-1}$). The data calibra-

tion was done with CASA[1] (McMullin et al. 2007). The flux and bandpass were calibrated with the quasar 3C 286. J1822-0938 was used for the phase and gain calibration (see also Beuther et al. 2016).

To recover the large scale structure, we combined the C-configuration data with the H I Very Large Array Galactic Plane Survey (VGPS, Stil et al. 2006), which consists of VLA D-configuration data combined with single-dish observations from the Green Bank Telescope (GBT). We subtracted the continuum in the visibility datasets (with the CASA command uvcontsub), and used the multiscale CLEAN in CASA[2] to image the three adjacent tiles of continuum-subtracted C-configuration data together with D-configuration data. A pixel size of 4″, a spectral resolution of 1.5 km s$^{-1}$, and a robust weighting value of 0.45 were used. The resulting images, which have a synthesized beam between 20″ to 40″ over the entire coverage of the THOR survey, were all smoothed to a common resolution of 40″. The images were further combined with the VGPS images (D+GBT) using the task "feather" in CASA to recover the large-scale structure. We compared the flux of the combined H I data with the single dish GBT data from VGPS. The flux agrees with each other within 5.7%. Considering that the typical absolute flux calibration uncertainty for the VLA at 1.4 GHz is ∼5% (Beuther et al. 2016), it is reasonable to conclude that our combined H I data fully recover the extended emission. The noise level in the line-free channel is about 4 K per 1.5 km s$^{-1}$.

Additionally, the THOR C-configuration only H I line with the continuum data (Beuther et al. 2016) are used to measure the H I optical depth towards bright continuum sources in the background. The THOR+VGPS 1.4 GHz continuum data (VLA C+D+GBT, Wang et al. 2018) are employed to estimate the diffuse continuum emission in the background. By comparing the flux density of the known SNRs (Green 2014), Anderson et al. (2017) showed that the flux retrieved from the combined continuum data is consistent with the literature. Thus, the continuum data also recover the extended emission.

### 2.2. H I *self absorption*

The integrated H I emission over the velocity range 50–60 km s$^{-1}$, shown in the top panel of Fig. 1, reveals diffuse emission covering a larger area than the $^{13}$CO emission. The strongest H I emission does not coincide with $^{13}$CO emission, but an anti-correlation between the H I and $^{13}$CO emission is suggested. Our analysis in the following section shows that this anti-correlation is due to the HISA: the cold atomic hydrogen absorbs the emission from an emitting atomic hydrogen cloud in the background, that is, H I self absorption. The terminology "H I self absorption" can be misleading. The emission and absorption processes can occur in the same cloud, but it is possible that the H I emission originates from a distant background cloud, which covers a similar or larger range of LSR velocities as the absorbing foreground cloud as illustrated in Fig. 2. A comprehensive discussion about the radiative transfer of HISA features can be found in Gibson et al. (2000), Kavars et al. (2003), Li & Goldsmith (2003), and Goldsmith & Li (2005). In general, we observe an emitting foreground and background H I cloud, which have spin temperatures, $T_{\rm fg}$ and $T_{\rm bg}$, respectively. The cold, absorbing H I cloud can be located between these two emitting clouds, having the spin temperature, $T_{\rm HISA}$. Furthermore, we observe 1.4 GHz continuum emission, which can be a diffuse Galactic component or arise from discrete strong sources. For simplicity, we assume that the continuum emission is situated in the background. In this, we will exclude the possibility of strong, discrete continuum sources and consider only the weak diffuse continuum background when estimating the HISA properties. In Sect. 4.3, we will utilize strong continuum sources to determine the optical depth of the atomic hydrogen, which can help us to constrain the spin temperature of HISA. Following the equation of radiative transfer in Rybicki & Lightman (1979), the measured on and off position brightness temperatures of the line above the continuum at a certain velocity are:

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

where $\tau_{\rm fg}$, $\tau_{\rm bg}$, $\tau_{\rm HISA}$ are the corresponding optical depths of each component shown in Fig. 2 and $T_{\rm cont}$ is the continuum brightness temperature. During the data reduction, we subtract the continuum emission from the H I visibility data (see also Sect. 2), which is indicated by the last term ($-T_{\rm cont}$, see Sect. 2). An example spectrum illustrating $T_{\rm on}$ and fitted $T_{\rm off}$ is shown in Fig. 3. Assuming on and off spectra share the same $T_{\rm cont}$ and calculating the difference, we get:

$$\begin{aligned} T_{\rm on-off} &= T_{\rm HISA}(1-{\rm e}^{-\tau_{\rm HISA}}){\rm e}^{-\tau_{\rm fg}} - T_{\rm bg}(1-{\rm e}^{-\tau_{\rm bg}}){\rm e}^{-\tau_{\rm fg}}(1-{\rm e}^{-\tau_{\rm HISA}})\\ &\quad - T_{\rm cont}(1-{\rm e}^{-\tau_{\rm HISA}}){\rm e}^{-(\tau_{\rm fg}+\tau_{\rm bg})}\\ &= \left(T_{\rm HISA} - T_{\rm bg}(1-{\rm e}^{-\tau_{\rm bg}}) - T_{\rm cont}{\rm e}^{-\tau_{\rm bg}}\right)\times(1-{\rm e}^{-\tau_{\rm HISA}}){\rm e}^{-\tau_{\rm fg}}. \end{aligned} \tag{2}$$

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

$$p \equiv \frac{T_{\rm bg}(1-{\rm e}^{-\tau_{\rm bg}})}{T_{\rm off}}. \tag{3}$$

That means for $p = 1$, there is no foreground emission and for $p = 0.5$, the foreground and background emission are equal. Measuring $p$ is difficult and it usually has to be assumed. As a last simplification, we assume that the foreground and background clouds are optically thin and therefore $\tau_{\rm fg}$ and $\tau_{\rm bg}$ are small (Gibson et al. 2000). This results in:

$$T_{\rm on-off} = (T_{\rm HISA} - p\,T_{\rm off} - T_{\rm cont})\times(1-{\rm e}^{-\tau_{\rm HISA}}). \tag{4}$$

$T_{\rm on}$ and $T_{\rm off}$ can be derived from our H I emission line observations, $T_{\rm cont}$ is from our THOR+VGPS 1.4 GHz continuum data (see Sect. 2). With these observable quantities we can estimate the properties of the HISA using Eq. 4. Specifically, we derive the cloud spin temperature $T_{\rm HISA}$ and the optical depth $\tau_{\rm HISA}$. We cannot disentangle the spin temperature and the optical depth. Figure 4 shows an example of the relation between $\tau_{\rm HISA}$ and $T_{\rm HISA}$ for the region close to the strong continuum source G34.256+0.146 of $T_{\rm off}$ = 103 K, $T_{\rm on}$ = 50 K, and $T_{\rm cont}$ = 17 K. The different colors represent different values of $p$ from 0.4 to 1. The black vertical line indicates the temperature of the cosmic microwave background radiation (Fixsen et al. 1996; Fixsen 2009; Planck Collaboration et al. 2016) $T$ = 2.7 K. Since the L band continuum background emission in the Galactic plane is larger than 0 (bottom panel Fig. 1), the spin temperature must be larger than 2.7 K. The general interpretation of the curves is that

[1] http://casa.nrao.edu; version 4.1.0
[2] version 5.1.1

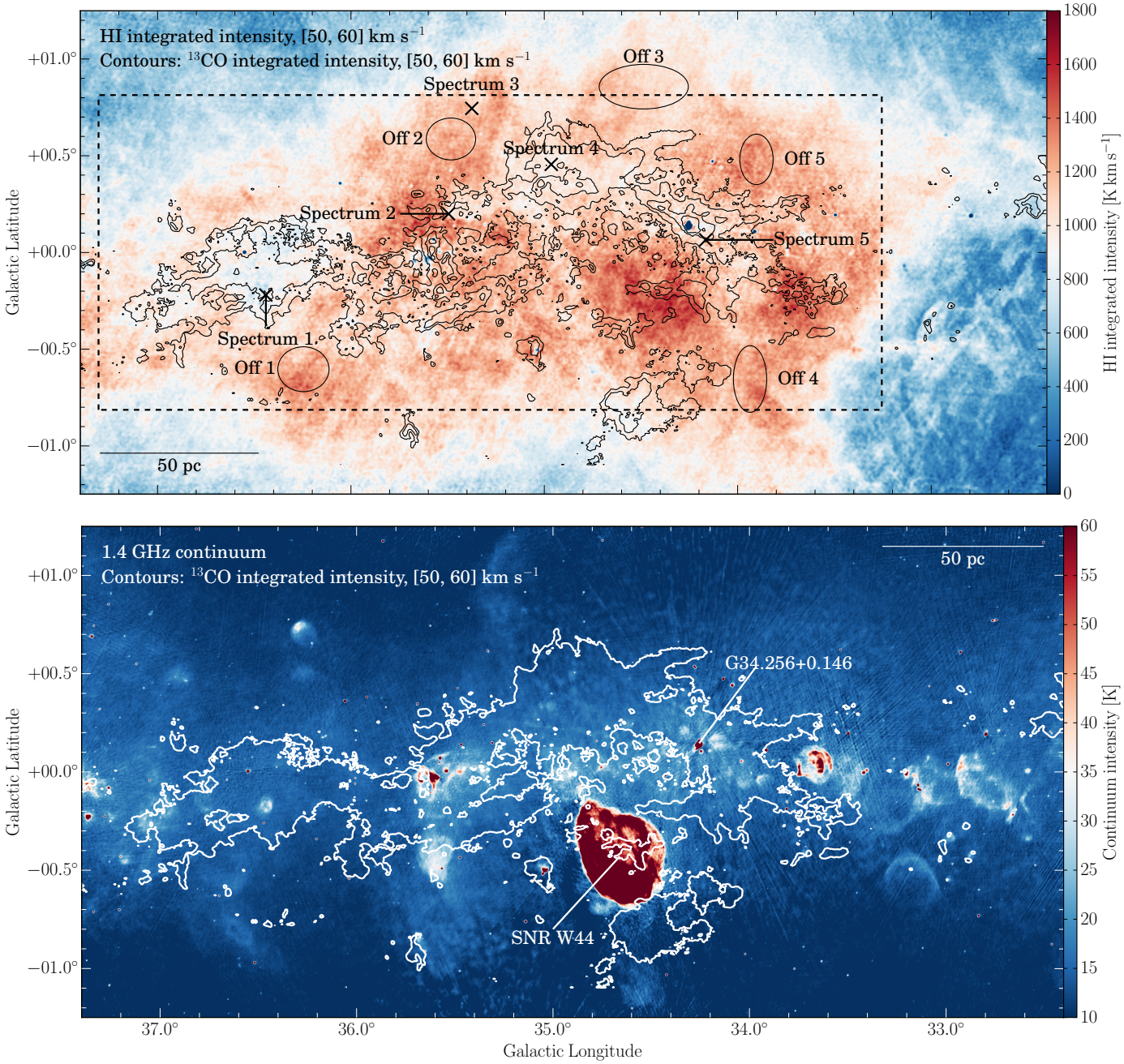


**Fig. 1.** Overview of the giant molecular filament GMF38a. The top panel shows the $^{13}$CO (GRS, Jackson et al. 2006) integrated intensity contours in the range $v_{\rm LSR}$ = 50 − 60 km s$^{-1}$ overlaid on integrated H I emission in the same velocity range. The black ellipses show the "off-positions" whose spectra are shown in Fig. 5 and the black "×" signs mark positions whose spectra are show in Fig. 6. The bottom panel shows the same $^{13}$CO integrated intensity contours overlaid on 1.4 GHz continuum emission from the THOR survey (Wang et al. 2018). The contours in the top panel indicate integrated $^{13}$CO emission levels of 5, 10, 20, and 30 K km s$^{-1}$. The contours in the bottom panel indicate integrated $^{13}$CO emission levels of 5 K km s$^{-1}$ for reference. The dashed box in the top panel outlines the region that is discussed in the following sections and shown in Fig.7, 9, 11, 14, and 19.

a higher optical depth is necessary to produce the assumed absorption feature for higher spin temperatures. This dependency becomes very steep at a certain point, depending on $p$. In Sect. 4 we shall discuss the relations among $\tau$, spin temperature, and $p$ in detail.

### 2.3. Background estimate to measure $T_{\rm off}$

To extract a reliable HISA feature, it is crucial to know the background H I emission. Different methods can be found in the literature to perform this task. The first one is to use absorption-free H I emission spectra, located close to the absorption feature (e.g., Gibson et al. 2000), referred to as "off-positions". This method assumes that the H I background emission stays spatially constant over the absorption feature, which might be true for spatially small HISA features. We tested this method by extracting H I spectra from five different regions, which are labeled as "Off 1" to "Off 5" in Fig. 1. These off-positions were chosen to be regions without significant 1.4 GHz continuum emission and without $^{13}$CO emission. Furthermore, these regions did not

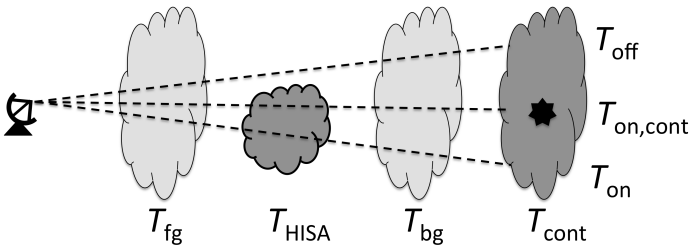


**Fig. 2.** Schematic view of the observed H I components. The cold absorbing cloud (HISA) with temperature $T_{\rm HISA}$ is surrounded by emitting clouds with temperature $T_{\rm fg}$ and $T_{\rm bg}$. Behind the H I clouds, several continuum sources can be situated, either diffuse or discrete (marked with a star).

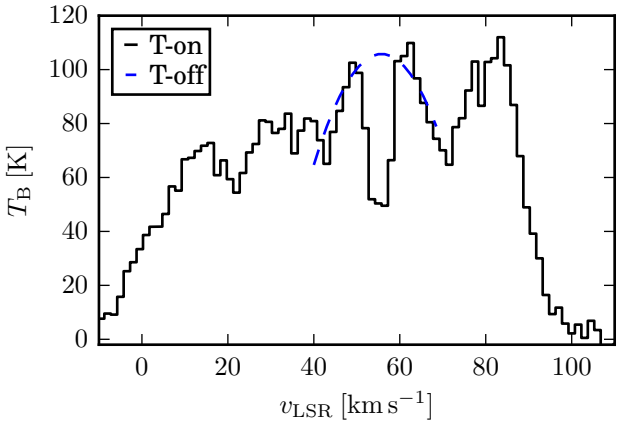


**Fig. 3.** Sample spectrum showing a prominent HISA feature around $v_{\rm LSR} \sim 55\,{\rm km\,s^{-1}}$. The actual H I spectra is shown in black ($T_{\rm on}$) and the estimated background emission using a second order polynomial fit (see Sect. 2.3) is shown in blue ($T_{\rm off}$).

show significant self absorption features at the velocity range of $v_{\rm LSR} = 50 - 60\,{\rm km\,s^{-1}}$. The corresponding spectra are presented in Fig. 5. These spectra reveal large variations, which makes it difficult to use them as a common off-position.

The second method utilizes a fit to the absorption free channels of the H I spectra to get $T_{\rm off}$. This method is applied frequently, using different functions to fit the H I emission, for example, linear fits (e.g., Minn 1981; Montgomery et al. 1995; McClure-Griffiths et al. 2006) or polynomials with different order (e.g. Myers et al. 1978; Bowers et al. 1980; Shuter et al. 1987; Kavars et al. 2003; Li & Goldsmith 2003). Figure 6 presents five different spectra from different positions indicated in Fig. 1. We used second and fourth order polynomials to fit the spectra for the velocity range around the HISA ($v_{\rm LSR} = 40 - 50$ and $60 - 70\,{\rm km\,s^{-1}}$). A polynomial function of the third order gave very similar results as the polynomial of the second order, thus, for clarity, we do not show it 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. 6 shows such a region and both functions represent the H I spectra well. Spectra 2 and 4 in Fig. 6 represents H I absorption features and the difference between the second and forth order polynomial is small. In contrast to this, Spectra 1 and 5 in Fig. 6 reveal a large difference between the fit functions. The fourth order polynomial fit is much higher (~50 K) than the second order polynomial for Spectrum 1, but much lower than the second order polynomial for Spectrum 5. It is not obvious which function describes the H I spectra more accurately. However, the fourth order polynomial might overestimate the actual spectra as steep slopes within the fitted velocity range would result in high values for the fitted spectra. 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 $T_{\rm off}$ and to extract HISA.

Another method was to utilize the second derivative representation of the spectrum as described in Krčo et al. (2008). Krčo et al. (2008) demonstrated that HINSA feature would become dominant in the second derivative representation. We also tested this method. For narrow HISA spectra (such as Spectrum 2 in Fig. 6) the second derivative technique can recover the HISA spectra relatively well. However, for broad spectra (e.g., Spec-

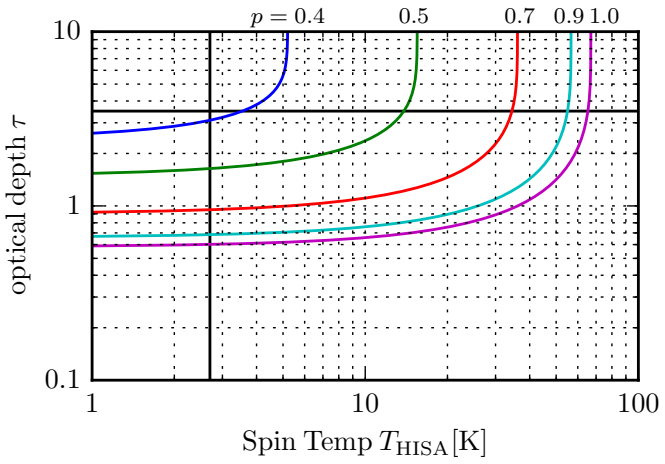


**Fig. 4.** Optical depth as a function of the spin temperature (as introduced in Eq. 4) for the region close to the strong continuum source G34.256+0.146 with $T_{\rm off} = 103\,{\rm K}$, $T_{\rm on} = 50\,{\rm K}$, and $T_{\rm cont} = 17\,{\rm K}$. The optical depth of $\tau = 3.5$ is indicated with a black horizontal line. The temperature of the CMB is indicated at $T = 2.7\,{\rm K}$ with a black vertical line.

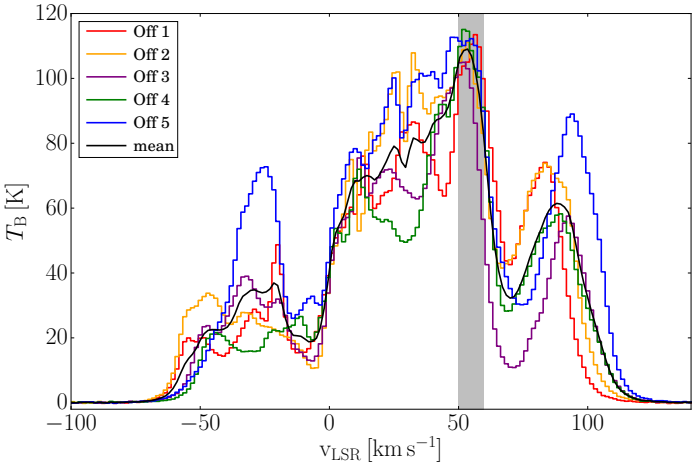


**Fig. 5.** Selected H I emission spectra around the GMF38a, which can be used as "off-positions". The regions we used for the extraction are shown in Fig. 1. The black line shows the mean spectrum of all five off-spectra and the gray shaded area indicates the velocity range of the HISA feature ($v_{\rm LSR}$ = 50 to 60 km s$^{-1}$).

trum 1 and Spectrum 5), the HISA spectra were filtered out by the method. Therefore, we do not use this method in our analysis.

McCutcheon et al. (1978), Winnberg et al. (1980), and Andersson et al. (1991) used one or several Gaussian profiles to fit the spectra and to derive the off spectra. However, as pointed out by McCutcheon et al. (1978), this method only works if the absorption feature is very narrow, the shape of the total spectrum is simple, and can be represented by a few Gaussians. On the other hand Dénes et al. (2018) used a machine learning method, the Autonomous Gaussian Decomposition algorithm (AGD; GAUSSPY) developed by Lindner et al. (2015), to decompose the emission spectra while masking out the HISA features to derive the off spectra. However, they only need to deal with 47 spectra and it is not clear how well this method would work for our region with ∼ 2 million spectra. It is definitely worthwhile to test the machine learning method in the future, but it is beyond the scope of this paper.

The mean spectrum of the five off-positions shown in Fig. 5 is shown in gray in Fig. 6 as well. While the mean off-position represents the H I spectra of $T_{\rm off}$ well in some cases (e.g., Spectrum 2 in Fig. 6), but in general it does not (e.g., Spectrum 1 or 5 in Fig. 6). There are apparent variations in the H I spectrum at velocities outside of the HISA feature, such the assumption of a uniform H I emission background does not seem to be adequate. The mean spectrum method was also discussed by Myers et al. (1978) and McCutcheon et al. (1978), and they concluded that it is not suitable for HISA studies for the same reason. Hence, we will use the polynomial fit method to extract the HISA feature rather than using a mean off-spectrum.

The noise of the extracted HISA spectra measured at the velocity ranges in $v_{\rm LSR}$ = 40 − 50 and 60 − 70 km s$^{-1}$ is ∼ 8 K per 1.5 km s$^{-1}$ for the second order polynomial fit, ∼ 5 K for the forth order fit. Since a forth order polynomial function can always fit the small bumps in the spectra better than the second order polynomial function, it is no surprise the HISA spectra extracted with the forth order fit have smaller noise.

### 2.4. HISA extraction

With the methods described in Sect. 2.3 we can estimate $T_{\rm off}$ in Eq. 4. Using this information, we can measure the depth of the absorption feature ($T_{\rm off} - T_{\rm on}$). 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. Fits that result in a peak intensity higher than 25 K (∼ 3 and 5 times the noise level of the HISA spectra extracted with second and forth oder fit, respectively), 1.5 km s$^{-1}$< full width half maximum (FWHM) linewidth (=channel width) <20 km s$^{-1}$ are considered as good fit. The peak values of the fitted Gaussian curves are shown in Fig. 7 for different ways of estimating $T_{\rm off}$. The absorption depth of the HISA shows values between ∼30 and ∼80 K.

As discussed in Sect. 2.3, different methods to estimate $T_{\rm off}$ resulted in differences for the peak value. Those based on the spatially averaged off-positions (bottom panel in Fig. 7) show unrealistically large values at the edges as well as in the presence of continuum sources. In the first case, as the Galactic H I emission drops at the edge, this method falsely identifies the edge regions as absorption features. In the second case, the method picks up the strong H I absorption towards the bright continuum emission in the background, as the structure of the continuum emission is clearly visible (bottom panel in Fig. 7).

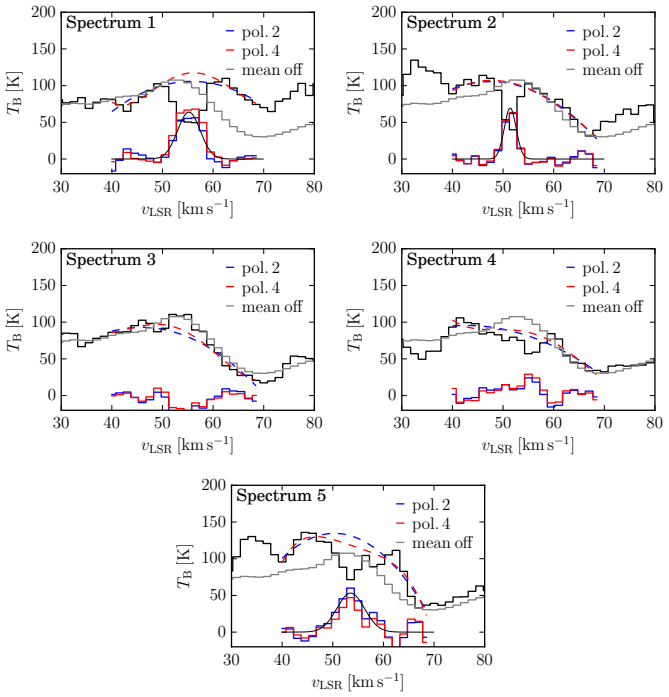


**Fig. 6.** Selected H I emission spectra corresponding the positions indicated in Fig. 1. The black lines represent the H I spectra, the colors correspond to the HISA. The gray spectra indicate the mean of the five off-positions presented in Fig. 5. The blue and red dashed lines represent the fits of the background to the HISA spectrum for a polynomial of second and fourth order, respectively, using the velocity range of $v_{\rm LSR}$ = 40−50 and 60 − 70 km s$^{-1}$ for the the baseline of the fit. The blue and red solid lines show the difference between fitted H I spectra and the measured H I spectra ($T_{\rm on-off}$ in Eq. 4) for a polynomial of second and fourth order, respectively. We fitted Spectra 1, 2, and 5 using a Gaussian function, shown by the black solid curve on top of the HISA spectra.

Fitting $T_{\rm off}$ of each H I spectra with a polynomial function circumvents these problems. Hence, the polynomial fitting method is more appropriate for our analysis and in the following we focus on the determination of $T_{\rm off}$ from HISA-free channels in the $T_{\rm on}$ spectrum.

We found large differences in $T_{\rm off}$ estimates made with polynomials of second order forth order, in particular towards regions around $l$ = 35° and $l$ = 33.8°(Fig. 7). Using a forth order polynomial for the background estimate, we found significantly more absorption due to a possible overestimate of the background emission (Sect. 2.3). Other regions are not affected significantly by the choice of the fit function, for instance, the regions around spectrum 1 or 2. The uncertainties of the HISA properties introduced by different polynomial function fittings will be discussed in Sect. 4.7.

## 3. Results

### 3.1. Kinematics

In this section, we discuss the kinematic properties of the HISA features. In contrast to the absorption depth, the peak velocity is not significantly affected by the choice of the fit function. Therefore we present here only the velocity structure using the second order polynomial for the determination of $T_{\rm off}$. The velocity structure revealed by the fourth order polynomial is similar.

To compare the kinematics of the HISA feature with those of the $^{13}$CO data, we resampled both the $^{13}$CO and H I data to the

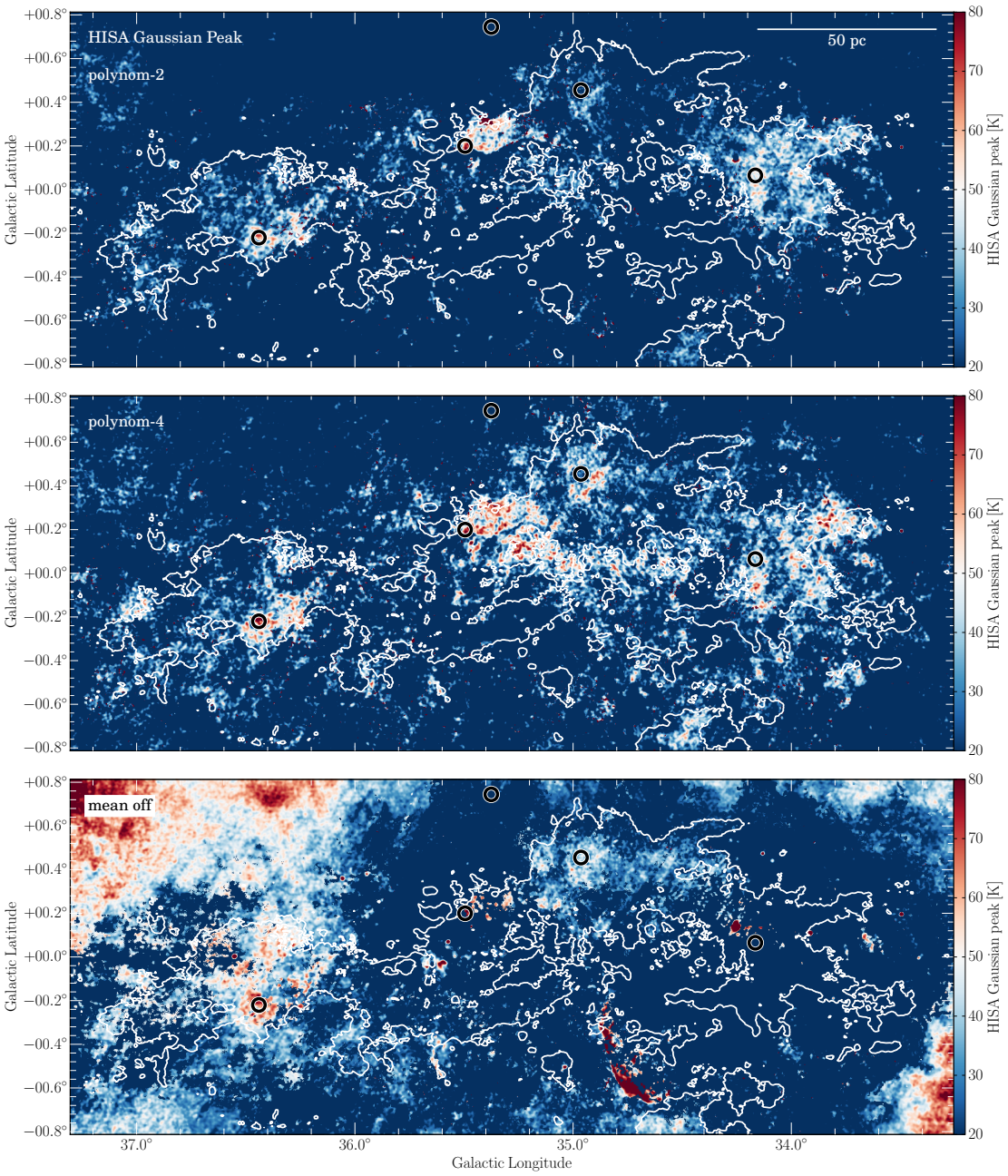


**Fig. 7.** Peak value of a Gaussian fit for the absorption depth $T_{\rm off} - T_{\rm on}$ for different methods of estimating $T_{\rm off}$. The top and middle panel show the method using a polynomial fits with second and forth order, respectively. The bottom panel represents the method using a spatially averaged off spectrum shown in Fig. 5. The white contours represent the integrated $^{13}$CO emission at levels of 5 K km s$^{-1}$ for reference. The circles from left to right marked the same positions as in Fig. 1 where Spectra 1 to 5 shown in Fig. 6 are extracted, respectively.

same spatial and velocity resolution (pixel size of 22″ and spectral channel width of 1.5 km s$^{-1}$) and applied Gaussian fitting to the data sets pixel by pixel to determine the peak velocity and FWHM linewidth. Due to limited sensitivity and spectral resolution of our H I data, we cannot disentangle multiple velocity components. Therefore, we chose to fit both the $^{13}$CO and HISA data with a single Gaussian component for simplicity and consistency. The Gaussian fitting to the HISA spectra is described in Sect. 2.4. For the $^{13}$CO data, fits that result in a peak intensity higher than 2 K (∼10 times noise level of the $^{13}$CO datacube at 1.5 km s$^{-1}$ channel width), 1.5 km s$^{-1}$ < FWHM linewidth <20 km s$^{-1}$ are considered good fits for the $^{13}$CO data.

The peak velocity maps are presented in Fig. 9. The $^{13}$CO peak velocity shows that the majority of the filament is at ∼54–58 km s$^{-1}$. The western part of the $^{13}$CO filament around $l$ = 34° is slightly red-shifted compared to rest of the filament, and has a velocity of ∼57-58 km s$^{-1}$. For this region, the peak velocities revealed by the HISA feature are at ∼54-55 km s$^{-1}$, which is about 3-4 km s$^{-1}$ lower than the $^{13}$CO velocity. This can also be seen in the right panel of Fig. 10, where we present a histogram of the HISA and $^{13}$CO peak velocities for the eastern and western side of the GMF, respectively. In contrast to this, the eastern side of the filament around $l$ = 36.5° shows a close correlation of the peak velocities as shown in the left panel of Fig. 10. We will dis-

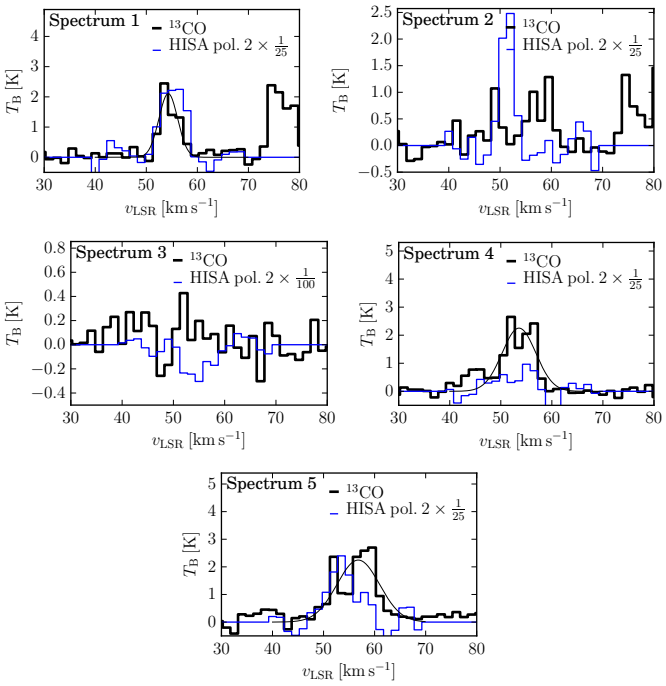


**Fig. 8.** Examples of $^{13}$CO spectra extracted from the same positions as in Fig.6 and the Gaussian fitting results if available. The $^{13}$CO spectra are shown in thick black lines. The HISA spectra from the same locations are scaled to 1/25 of its original intensity and shown in thin blue lines. The positions corresponding to each spectrum are shown in Fig. 1. For Spectra 1, 4, and 5, we can fit the spectrum with a Gaussian component to the $^{13}$CO spectrum successfully, we show the Gaussian fitting result with a black solid curve in each panel. For Spectra 2 and 3, the Gaussian fitting results did not meet the criteria we described in Sect. 3.1 and were ignored.

cuss this effect in Sect. 4.1. The western and eastern region we refer to here are both under the Galactic coordinate framework.

The linewidths of both $^{13}$CO and HISA are shown in Fig. 11. The linewidth for the $^{13}$CO emission shows extremely high values of more than 10 km s$^{-1}$ for the central region of the filament around $l = 35°$. However, these values have to be treated cautiously as the $^{13}$CO emission exhibits multiple lines in this region and we only use a single Gaussian function to fit them. On the eastern side of the filament around $l = 36.5°$ we find mostly single components for the $^{13}$CO emission and the linewidth is $\Delta v \sim 2$ to 4 km s$^{-1}$. The linewidth of the HISA feature shows values $\Delta v \sim 3 - 6$ km s$^{-1}$ for the whole filament. This can also be seen in the left panel of Fig. 12, where a histogram of the linewidths is shown. The linewidth distribution of the $^{13}$CO emission is systematically higher in the western region of the filament, whereas the linewidth of the HISA feature is similar to the eastern region. This result has to be treated cautiously as we see multiple components for the $^{13}$CO line within the western region, which increases the linewidth.

To estimate the contribution of the non-thermal component in the HISA features and the molecular emission, we assume the relation $\sigma_{\rm nth} = \sqrt{\sigma^2_{\rm obs} - \sigma^2_{\rm th} - \sigma^2_{\rm res}}$, where $\sigma_{\rm obs}$ is the measured velocity dispersion, $\sigma_{\rm th}$ is the radial component of the thermal velocity dispersion, and $\sigma_{\rm res}$ is the velocity dispersion introduced by the channel width of our data (1.5 km s$^{-1}$). Assuming a Gaussian line profile with the FWHM linewidth $\Delta v$ obtained from the aforementioned fittings we get $\sigma_{\rm obs} = \Delta v / \sqrt{8\ln 2}$, and $\sigma_{\rm res} = 1.5/\sqrt{8\ln 2}$ km s$^{-1}$. Assuming a Maxwell-Boltzmann velocity distribution, $\sigma_{\rm th} = \sqrt{k_{\rm B} T_{\rm k}/(\mu m_{\rm H})}$, where $k_{\rm B}$ is the Boltzmann constant, $\mu$ is the molecular weight, $m_{\rm H}$ is the mass of the hydrogen atom, and $T_{\rm k}$ is the kinetic temperature. For HISA, we assume $T_{\rm k} = T_{\rm HISA} = 40$ K. The peak excitation temperature of $^{13}$CO we derived for the filament is $\sim$ 25 K (see Sect. 3.2.1) which agrees to what Roman-Duval et al. (2010) found for the GRS molecular clouds. If we assume that the brightest $^{13}$CO emission is coming from regions where the line is optically thick and thermalized, then this excitation temperature will be comparable to the actual gas kinetic temperature, which must therefore have a value close to 20 K. Furthermore, simulations of molecular clouds in a variety of different radiation fields show that the CO mass-weighted temperature of the gas is typically in the range 10–30 K, with very little dependence on the local environment (Peñaloza et al. 2018). Therefore, we assume a uniform $T_{\rm k}$ of 20 K for $^{13}$CO. Assuming spatial isotropy, the Mach number is estimated to be $\sqrt{3}\sigma_{\rm nth}/c_{\rm s}$, where $c_{\rm s}$ is the sound speed estimated using a mean molecular weight $\mu = 2.34$ for the molecular cloud and $\mu = 1.27$ for the H I cloud (Allen 1973; Cox 2000). The distribution of the Mach number for HISA and $^{13}$CO across the whole filament (Fig. 13) shows that the $^{13}$CO emission is dominated by supersonic motions, whereas HISA features have a much smaller Mach number in general with a significant fraction of the HISA features, that is, a significant fraction of the CNM being at subsonic and transonic velocities. If the HISA and $^{13}$CO lack spatial isotropy, we could be overestimating the Mach number by a maximum factor of $\sqrt{3}$.

### 3.2. Column density

#### 3.2.1. $H_2$ column density

We use the $^{13}$CO(1–0) data from the Galactic Ring Survey (GRS, Jackson et al. 2006) to derive the column density and kinematic properties of the molecular gas of the filament. The data have an angular resolution $\Theta = 46''$ and a velocity resolution of $\Delta v = 0.21$ km s$^{-1}$. We can estimate the column density of the $^{13}$CO molecule including opacity correction with the equation (Wilson et al. 2010):

$$N(^{13}\mathrm{CO}) = 3.0 \times 10^{14} \frac{\tau_{13}}{1 - \mathrm{e}^{-\tau_{13}}} \frac{\int T_{\rm MB}(v)\, dv}{1 - \mathrm{e}^{-5.3/T_{\rm ex}}}, \quad (5)$$

where $N(^{13}$CO) is the column density of the $^{13}$CO in units of cm$^{-2}$, $dv$ is the velocity in km s$^{-1}$, $\tau_{13}$ is the $^{13}$CO opacity, $T_{\rm MB}$ is the main beam brightness temperature, and $T_{\rm ex}$ is the excitation temperature. We do not have a direct measurement for the excitation temperature, and we assume that the excitation temperatures of the $^{12}$CO and $^{13}$CO are the same. Assuming that the $^{12}$CO line is optically thick, we used the $^{12}$CO(1–0) data from the FOREST unbiased Galactic plane imaging survey with the Nobeyama 45 m telescope (FUGIN; Umemoto et al. 2017) to estimate $T_{\rm ex}$ following the formula (Wilson et al. 2010):

$$T_{\rm ex} = \frac{5.5}{\ln\left(1 + \frac{5.5}{T_{\rm mb}(^{12}\mathrm{CO}) + 0.82}\right)}, \quad (6)$$

where $T_{\rm mb}(^{12}$CO) is the peak main-beam brightness temperature of $^{12}$CO(1–0) line. We calculated $T_{\rm ex}$ for regions where $T_{\rm mb}(^{12}$CO) is above the $5\sigma$ level (2 K), which results in a $T_{\rm ex}$ between ~5 to 25 K. For regions where $T_{\rm mb}(^{12}$CO) is below the $5\sigma$ level (2 K), an upper limit of 5 K for $T_{\rm ex}$ is applied. Following Eq. B.6 in Schneider et al. (2016), we can derive $\tau_{13}$ from $T_{\rm ex}$ and $T_{\rm mb}(^{13}$CO). For regions where $T_{\rm mb}(^{13}$CO) is above $5\sigma$

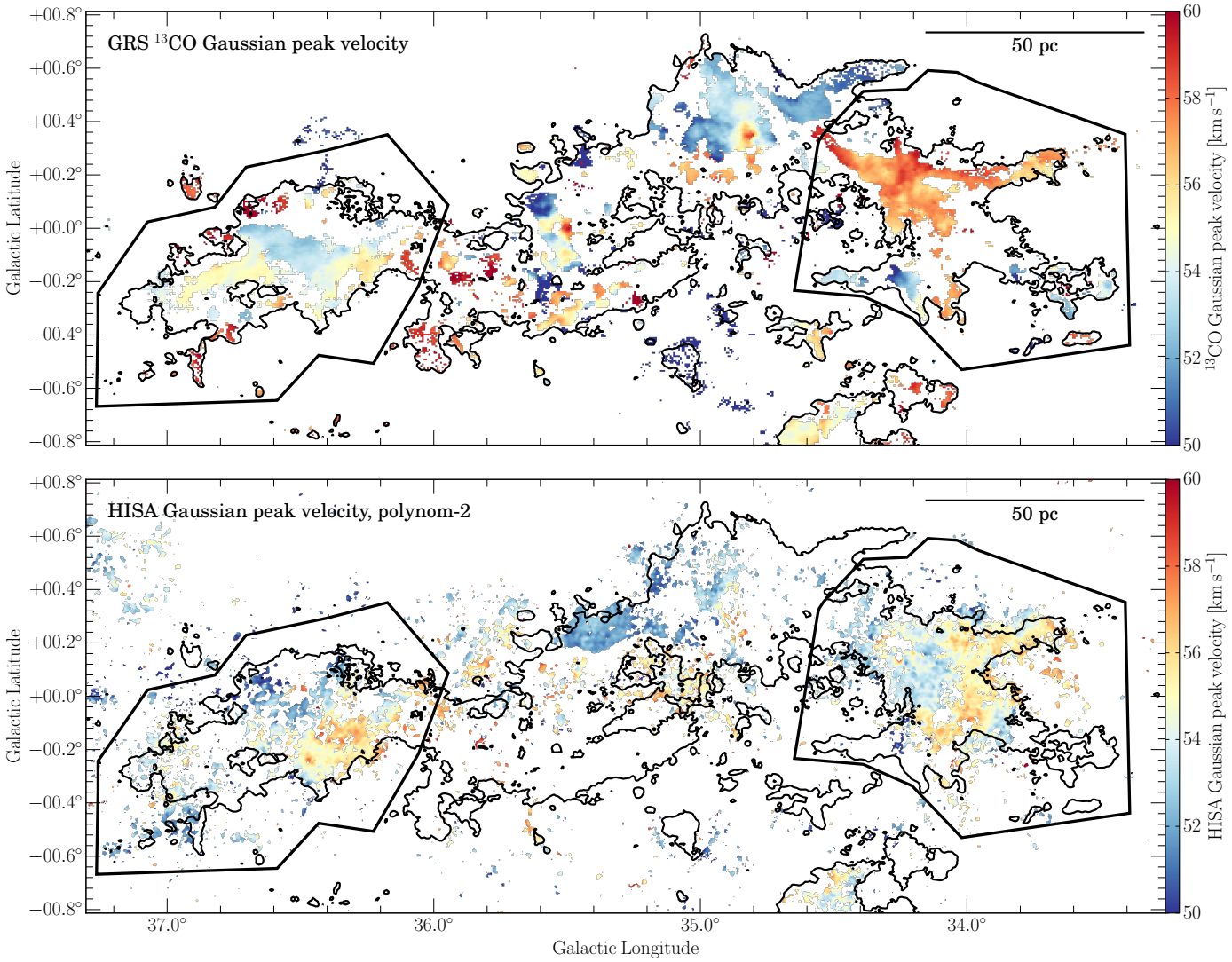


**Fig. 9.** Peak velocity of the Gaussian fit of the $^{13}$CO data (top panel) and the HISA data (bottom panel), respectively. The HISA are extracted by using a second order polynomial to estimate the background emission. The black contours indicate the integrated $^{13}$CO emission at levels of 5 K km s$^{-1}$. The polygons mark the eastern (around $l$ = 36.5°) and western (around $l$ = 34°) subregions, for which we present the histograms of the peak velocities in Fig. 10.

level (1.05 K), $\tau_{13}$ is estimated to be between ~0.1 and 3. For regions where $T_{\rm mb}(^{13}{\rm CO})$ is below 5$\sigma$ level (1.05 K), an upper limit of 0.1 for $\tau_{13}$ is applied. The majority of the region along the filament has a $\tau_{13} \lesssim 1.0$, only the region around ($l$ = 34.30, $b$ = 0.18) and a few pixels around ($l$ = 35.55, $b$ = 0.0) have a $\tau_{13} > 2.0$.

For the Galactocentric distance of 5.9 kpc of G38a, the fractional abundance of $^{13}$CO relative to $H_2$ is estimated to be 2.9×10$^{-6}$ following the relations reported by Giannetti et al. (2014). With this abundance, we converted the $N(^{13}{\rm CO})$ to $N(H_2)$. Zhang et al. (2019) employed a similar method (uniform CO abundance, $T_{\rm mb}(^{12}{\rm CO}) \rightarrow T_{\rm ex}$) to estimate the column density and mass of a sample of GMFs (they estimated the mass of GMF38a to be $\sim 3.8 - -11.0 \times 10^5\ M_\odot$), and discussed in detail the uncertainties brought in by $T_{\rm ex}$ and the $^{13}$CO abundance. According to their results, the 1$\sigma$ uncertainty of the column density estimated from $^{13}$CO is ~ 50%. Simulations show that the abundance of $^{13}$CO could vary and we could underestimate the column density for the low column density part by ~ 40% ($N(^{13}{\rm CO}) < 10^{16}$, or $N(H_2) < 3.4 \times 10^{21}$ in this paper, Szűcs et al. 2014, 2016).

### 3.2.2. CNM column density from HISA measurements

Besides the kinematics, the column density of H I is also a critical cloud parameter. To estimate the column density of the HISA feature we use the equation given by Wilson et al. (2010):

$$N_{\rm H} = 1.8224 \times 10^{18}\ T_{\rm S} \int \tau(v)\, dv, \qquad (7)$$

where $T_{\rm S}$ and $\tau$ are the spin temperature and optical depth, respectively. For the CNM traced by HISA, $T_{\rm S} = T_{\rm HISA}$ and $\tau = \tau_{\rm HISA}$ (Eq. 4). However, as mentioned in Sect. 2.2 we measure the spin temperature and the optical depth together and disentangling them is difficult. Hence, we assume a constant spin temperature over the cloud and calculate the optical depth using Eq. 4. As we have no measurement for $p$, this value is difficult to estimate. Considering GMF38a is at the near side of the Milky way, we assume a value of $p$ = 0.9 in the following and discuss the corresponding uncertainties in Sect. 4.2 and 4.7. We

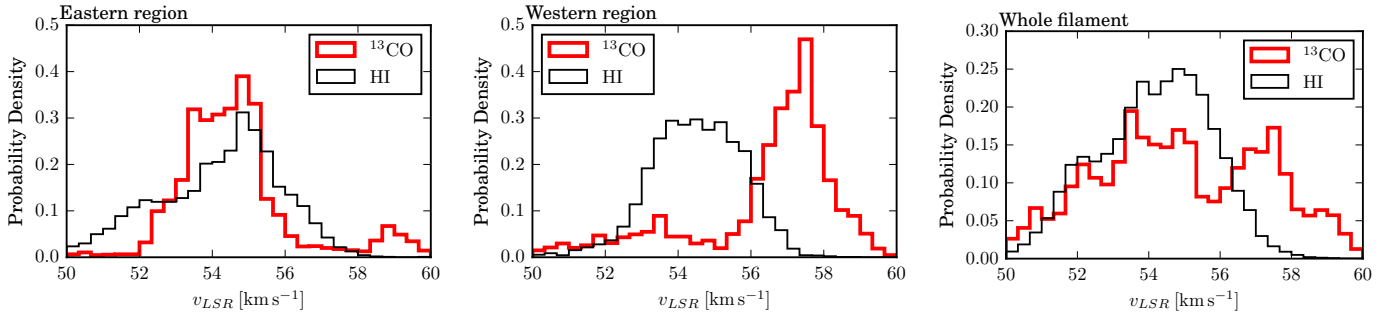


**Fig. 10.** Histograms of the HISA and $^{13}$CO peak velocity, shown in black (thin) and red (thick) respectively. The left and middle panels show all extracted pixels within the marked eastern (around $l = 36.5°$) and western (around $l = 34°$) polygon in Fig. 9. The right panel shows all extracted pixels for the whole filament (red polygon in Fig. 9).

will discuss the uncertainty of column density brought in by the method we choose to estimate the background temperature $T_{\rm off}$ in Sect. 4.7

We integrate between 50 and 60 km$^{-1}$ and derive the column density map shown in Fig. 14, assuming for the HISA feature a spin temperature $T_{\rm S} = 40$ K, $p = 0.9$ and using a second order polynomial to estimate the background temperature $T_{\rm off}$. Larger spin temperatures do not change the structure of the column density map significantly, but will increase its value everywhere.

The column density peaks in the $^{13}$CO map do not coincide well with the column density peaks of the atomic hydrogen. As shown in Fig. 1, the highest peak in the $^{13}$CO (around $l = 34°$) coincides with a strong continuum source and hence makes the determination of the HISA feature at this position impossible. However, we use this continuum source to constrain the optical depth, which we present in Sect. 4.3. The highest column density peak for the atomic hydrogen can be found in the eastern area of the filament ($l = 36.5°$). In this region, the $H_2$ is diffuse and its column density is low. Another CNM column density peak can be found in the center of the filament around $l = 35.4°$, $b = +0.3°$. This CNM feature has almost a round shape and only a very weak counterpart in the $^{13}$CO emission.

Assuming a typical CNM thermal pressure of $P_{\rm CNM}/k$ ~4000 K cm$^{-3}$ (Heiles 1997; Jenkins & Tripp 2011; Goldsmith 2013), and $T_{\rm K} = T_{\rm HISA} = 40$ K, we can estimate the size of the CNM along the line of sight by dividing the column density by $P_{\rm CNM}/k/T_{\rm K}$. Depending on the column density, the line of sight size is estimated to be ~0.5–3 pc, which is much smaller than the width and length of the filament (~ 25 pc and ~230 pc, respectively). Arguing the other way round, if the line of sight size of the CNM is similar to the width of the filament (~ 25 pc), the thermal pressure would be significantly lower, and the CNM could be under-pressured or the pressure may be dominated by some other (magnetic or turbulent) component.

### 3.2.3. Atomic gas column density from H I emission

Since HISA traces only the CNM, it is likely that this component is surrounded or even mixed with a warm component. We also derive the column density traced by the H I emission (between 50 and 60 km s$^{-1}$, $32.65° < l < 37.27°$ and $|b| \leq 1.25°$), which traces both the CNM and WNM. Following the method described in Bihr et al. (2015), we derived the mean H I optical depth map from the strong continuum sources in the background (G33.498+0.194, G33.810-0.189, G33.915+0.110, G34.133+0.471, G35.053–0.518, G35.574+0.068, G35.947+0.379, G36.056+0.357, and G36.551+0.002; Wang et al. 2018). Following the method described in Bihr et al. (2015), the optical depth measured towards the continuum source is:

$$\tau = -\ln\left(\frac{T_{\rm on,\ cont} - T_{\rm off,\ cont}}{T_{\rm cont}}\right), \quad (8)$$

where $T_{\rm on,\ cont}$ is the on-continuum-source brightness temperature, and $T_{\rm off,\ cont}$ is the off-continuum-source temperature. Since we use the THOR C array data to calculate $\tau$, the smooth, large-scale structure is mostly filtered out (Beuther et al. 2016). We can neglect the off emission $T_{\rm off,\ cont}$ and simplify Eq. 8 to:

$$\tau_{\rm simplified} = -\ln\left(\frac{T_{\rm on,cont}}{T_{\rm cont}}\right). \quad (9)$$

For channels with a $T_{\rm on,\ cont}$ value smaller than 3 times the rms, we use the $3\sigma$ value to get a lower limit of $\tau$. The mean optical depth varies between 1.1 to 1.9 from 50 to 60 km s$^{-1}$. The optical depth corrected spin temperature is $T_{\rm S} = T_{\rm B}/(1 - {\rm e}^{-\tau})$, where $T_{\rm B}$ is the brightness temperature of the H I emission. The optical depth corrected atomic hydrogen column density is calculated with Eq. 7. Since the absorption features towards these continuum sources between 50 to 60 km s$^{-1}$ often saturates, the optical depth we derived is a lower limit as shown in Fig. 18 and we are underestimating the H I column density.

The column density we derived from H I emission is a combined result from the far (10.3 kpc) and near side (3.4 kpc) due to the kinematic distance ambiguity. Assuming the atomic gas in the Galactic plane is approximately axisymmetric with respect to the Galactic center (Kalberla & Dedes 2008), the atomic gas at near side and far side that are at the same Galactocentric distance share the same density distribution in the vertical direction. We used the average vertical density profile described by Lockman (1984) (see Eq. 5 and Table. 1 in their paper) to estimate how much gas is at near distance for each line of sight and derived the column density map of H I emission at 3.4 kpc shown in Fig. 14.

By comparing the column density maps of CNM and $H_2$, we produce the CNM-to-$H_2$ ratio map and show it for the Eastern and Western regions in Fig. 15. We masked out regions outside the column density threshold contours shown in Fig. 14. In both subregions, the CNM-to-$H_2$ ratio varies between ~ 0.5 – 25% with a median value of ~ 9%. Figure 15 shows that the outer layers of the filament have high CNM-to-$H_2$ ratio, while the inner regions show lower CNM-to-$H_2$ ratio. This change from the outside to the inside of the cloud can be interpreted as signature of

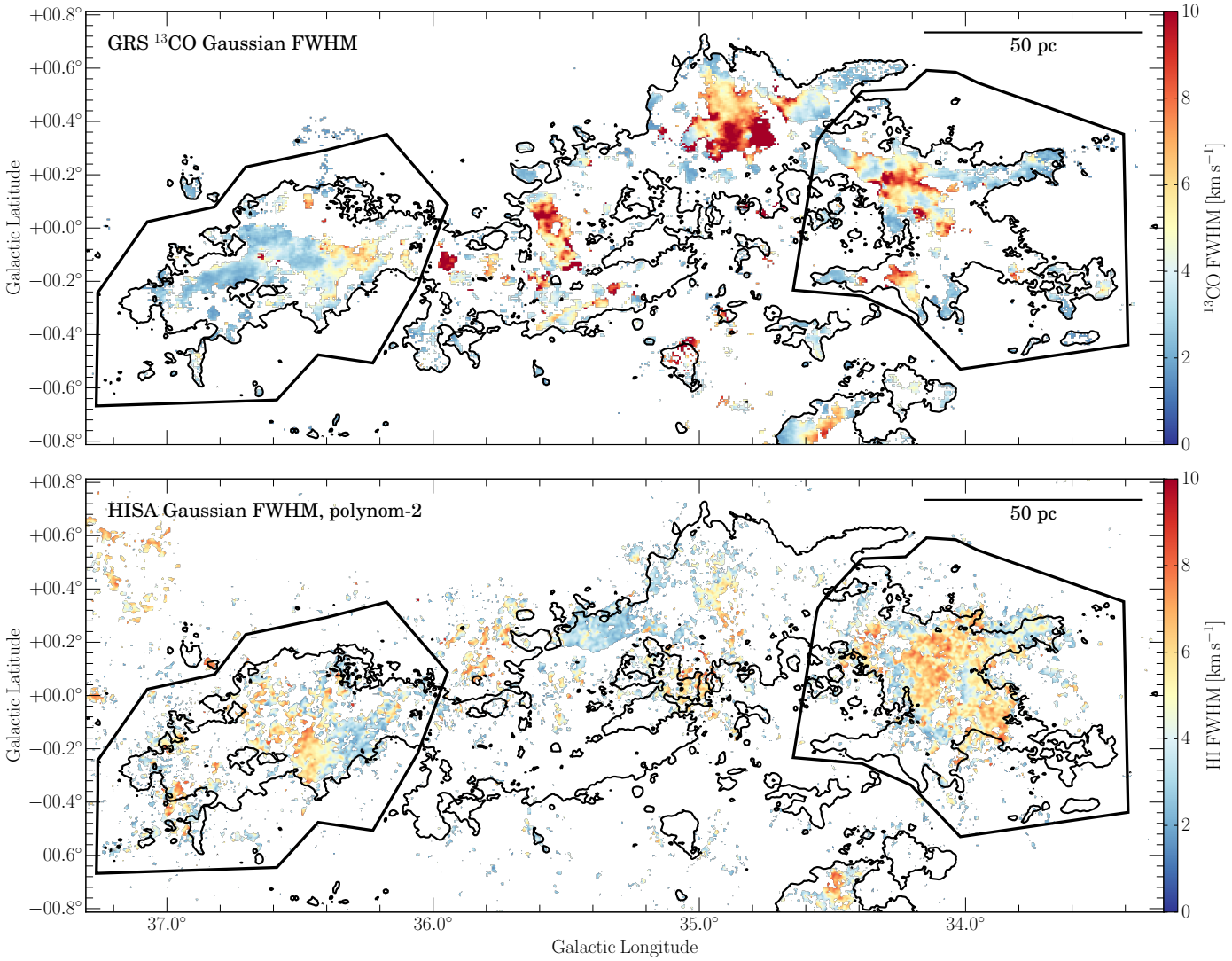


**Fig. 11.** Full width at half maximum (FWHM) of the $^{13}$CO emission (top panel) and the HISA feature (bottom panel), respectively. The linewidth is determined by using a single Gaussian fit. The contours indicate the integrated $^{13}$CO emission at levels of 5 K km s$^{-1}$ for reference. The polygons mark the eastern (around $l$ = 36.5°) and western (around $l$ = 34°) subregions, for which we present the histograms of the FWHM linewidths in Fig. 12.

the conversion of atomic to molecular gas with increasing density. Zuo et al. (2018) studied HINSA towards nearby clouds and found a much lower ratio between 0.2 and 2%.

Figure 16 shows the surface density comparison between the atomic hydrogen (CNM+WNM) and the total gas (CNM+WNM+$H_2$). The surface density of the atomic hydrogen rises up to $\sim 14-23\ M_\odot$ pc$^{-2}$ ($\sim 1.8-2.9\times10^{21}$ cm$^{-2}$) and then saturates to an almost flat distribution. This turnover is at lower values than that found by Bihr et al. (2015) towards W43 (50–80 M$_\odot$ pc$^{-2}$) but it is still higher than the 10 M$_\odot$ pc$^{-2}$ observed towards nearby clouds (Lee et al. 2015) and predicted by models (e.g., Krumholz et al. 2008, 2009; Sternberg et al. 2014). Such higher than predicted H I column densities can be explained by the clumpy nature of the ISM with several H I-to-H2 transitions along the line of sight (Bialy et al. 2017a).

### 3.3. Mass estimate

As we know the column density and the distance to the cloud (~3.4 kpc), we can directly estimate the mass of H I and $H_2$ gas. We do so for three different regions, the "full filament" (red polygon in Fig. 14), the "eastern region" (eastern green dashed polygons in Fig. 14) and the "western region" (western green dashed polygons in Fig. 14). Table 1 summarizes the mass measurements. The molecular hydrogen mass for the entire filament is ~$3.6\times10^5\ M_\odot$ and the CNM mass traced by the HISA is significantly less, showing values of $3.4\times10^3$ to $9.5\times10^3\ M_\odot$, depending on the assumed spin temperature.

Furthermore, we studied the cold atomic to the molecular hydrogen mass ratio $M$(CNM)/$M$($H_2$). For the entire filament, this value is between 1% and 3%, again depending on the assumed spin temperature. However, this ratio has to be treated cautiously. 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 higher values (3% to 4%). The $H_2$ column density shows significantly higher values for the western region in comparison to the eastern region. In contrast to this, the H I column density reveals a prominent peak on the eastern side and hence the $M$(CNM)/$M$($H_2$) ratio is lower for the western than

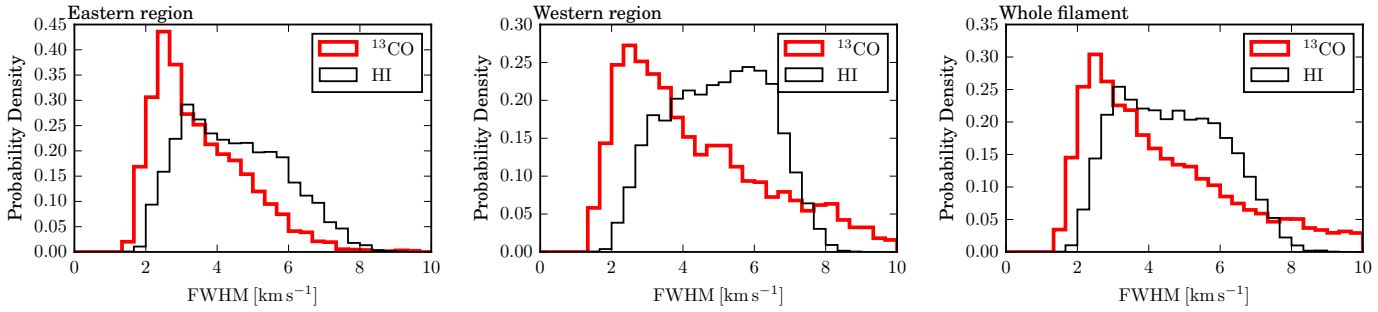


**Fig. 12.** Histograms of the FWHM linewidth of the HISA and $^{13}$CO emission in black (thin) and red (thick), respectively. The left and middle panels show all extracted pixels within the marked eastern (around $l$ = 36.5°) and western (around $l$ = 34°) polygon in Fig. 9. The right panel shows all extracted pixels for the whole filament (red polygon in Fig. 9).

**Table 1.** Mass estimates of the GMF.

| Region | $M(H_2)$ [$M_\odot$] | $M$(CNM) (20 K) [$M_\odot$] | $M$(CNM) (40 K) [$M_\odot$] | $M$(CNM)/M($H_2$) (40 K) | $M$(H I emission) [$M_\odot$] |
|---|---|---|---|---|---|
| Full filament | $3.6\times10^5$ | $3.4\times10^3$ | $9.5\times10^3$ | 3% | $2.3\times10^5$ ($6.1\times10^{5*}$) |
| Eastern region | $7.9\times10^4$ | $1.1\times10^3$ | $3.4\times10^3$ | 4% | $5.0\times10^4$ |
| Western region | $1.1\times10^5$ | $1.1\times10^3$ | $2.9\times10^3$ | 3% | $7.1\times10^4$ |

∗This mass is calculated including all H I emission between 50 and 60 km s$^{-1}$, 32.65° < $l$ < 37.27°, $|b| \leq 1.1°$ (bottom panel in Fig. 14).

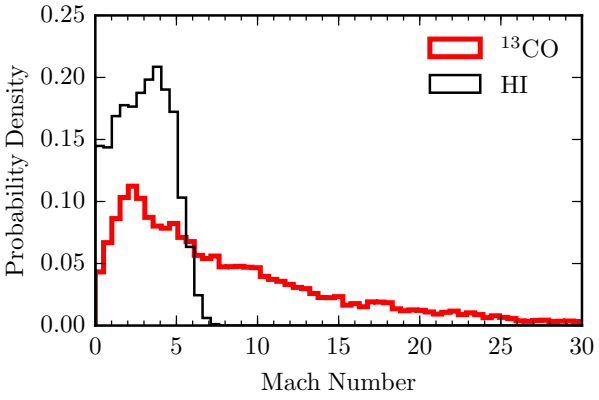


**Fig. 13.** Histograms of the Mach number of the HISA and $^{13}$CO emission, shown in black and red, respectively.

that for the eastern region. Considering that $^{13}$CO does not trace all the molecular hydrogen gas (Pineda et al. 2008; Goodman et al. 2009; Gong et al. 2018), the $M$(CNM)/$M(H_2)$ ratio we derive could be just an upper limit.

We further estimate the mass of the atomic component traced by H I emission. Within the same area (polygons in Fig. 14, the atomic gas traced by H I emission has lower mass as the molecular gas (Table 1). However, since there is no clear boundary of the GMF shown in the H I column density map (see bottom panel in Fig. 14), we can assume all H I emission between 50 to 60 km s$^{-1}$ within 32.65° < $l$ < 37.27° and $|b| \leq 1.1°$ (same latitude range as the GRS coverage) is associated with the filament to obtain the mass. The mass is estimated to be $6.1\times10^5$ $M_\odot$, which is about ~60 times larger than the mass estimated for the CNM traced by HISA (40 K). The mass of the atomic hydrogen is about 60% larger than the molecular hydrogen mass ($3.6\times10^5$ $M_\odot$), which makes sense since the molecular cloud is surrounded by a large reservoir of atomic gas.

## 4. Discussion

### 4.1. Kinematics

For nearby galaxies, the ratios of the CO to H I linewidth is around $\sigma_{HI}/\sigma_{CO} = 1 - 1.4$ (Caldú-Primo et al. 2013; Mogotsi et al. 2016). The linewidth values found in these studies are approximately $\sigma \sim 6 - 12$ km s$^{-1}$, which corresponds to $\Delta v_{FWHM} \sim 14 - 28$ km s$^{-1}$ for both the H I and CO lines. These measurements 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 in the supersonically turbulent ISM (e.g., Klessen & Glover 2016). Our Galactic HISA measurements of GMF38a show significantly smaller values for the linewidth of $\Delta v_{FWHM} \sim 2 - 8$ km s$^{-1}$ for CNM. The reason is that we observe a much smaller region and we are able to separate multiple components. Furthermore, they observe H I emission and the linewidth is dominated by WNM, whereas we observe cold H I absorption features produced by CNM.

To study the linewidth ratio in detail, we determine this ratio for the eastern region, which is indicated in Fig. 11. We focus on this region as it is not significantly affected by multiple component line spectra. A histogram of the H I/$^{13}$CO ratio is shown in Fig. 17. The mean values for the linewidths are $\Delta v_{FWHM}(^{13}CO) = 3.6$ km s$^{-1}$ and $\Delta v_{FWHM}$(H I) = 4.5 km s$^{-1}$.

### 4.2. The percentage of background emission – $p$

The percentage of background emission, parameterized with $p$, is difficult to estimate. Different assumptions can be found in the literature. For example, McClure-Griffiths et al. (2006) and Dénes et al. (2018) assume that $p$ = 1 and 0.9 for the observed Riegel-Crutcher cloud, respectively, as the corresponding dis-

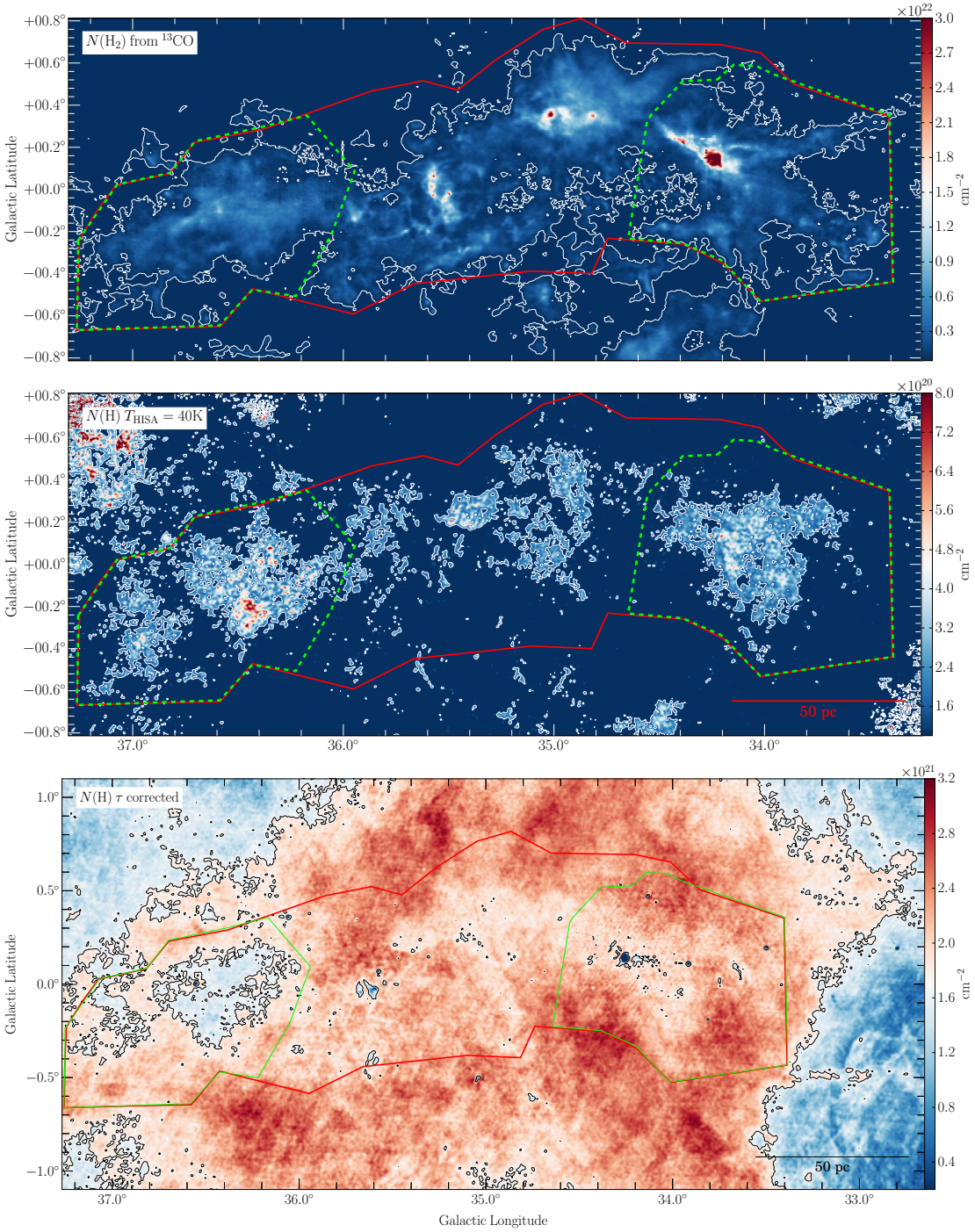


**Fig. 14.** Top panel: Column density map of the molecular cloud derived from $^{13}$CO emission (see Sect. 3.2.1). Middle panel: The column density map of the CNM derived from HISA, assuming $T_S = 40$ K, $p = 0.9$ and using a second order polynomial to estimate the background temperature $T_{off}$. Bottom panel: The optical depth corrected column density map of atomic hydrogen derived from H I emission. The contours indicate column density threshold used for the N-PDFs (see Sect. 4.5). For the $H_2$, HISA, and H I emission they have levels of $6\times10^{20}$, $1.5\times10^{20}$, and $1.7\times10^{21}$ cm$^{-2}$, respectively. The red and green dashed polygons mark the region for the mass estimates and the column density PDF measurements shown in Figs. 20 and 22, respectively.

tance is small (~125 pc). Rebolledo et al. (2017) studied the HISA features in the Gum 31 molecular complex with a simple two component assumption ($p = 1$). Li & Goldsmith (2003) studied the H I narrow self-absorption towards dark clouds in the Taurus and Perseus region. Since these clouds are located at high Galactic latitude away from the Galactic mid-plane, they can assume a simple Gaussian Galactic H I disk model (Lockman 1984) and estimate the factor $p$ with the complementary error function.

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 obtain a lower limit for $p$. As shown in Fig. 4, low values of $p \lesssim 0.4$ are not feasible as the spin temperature would become

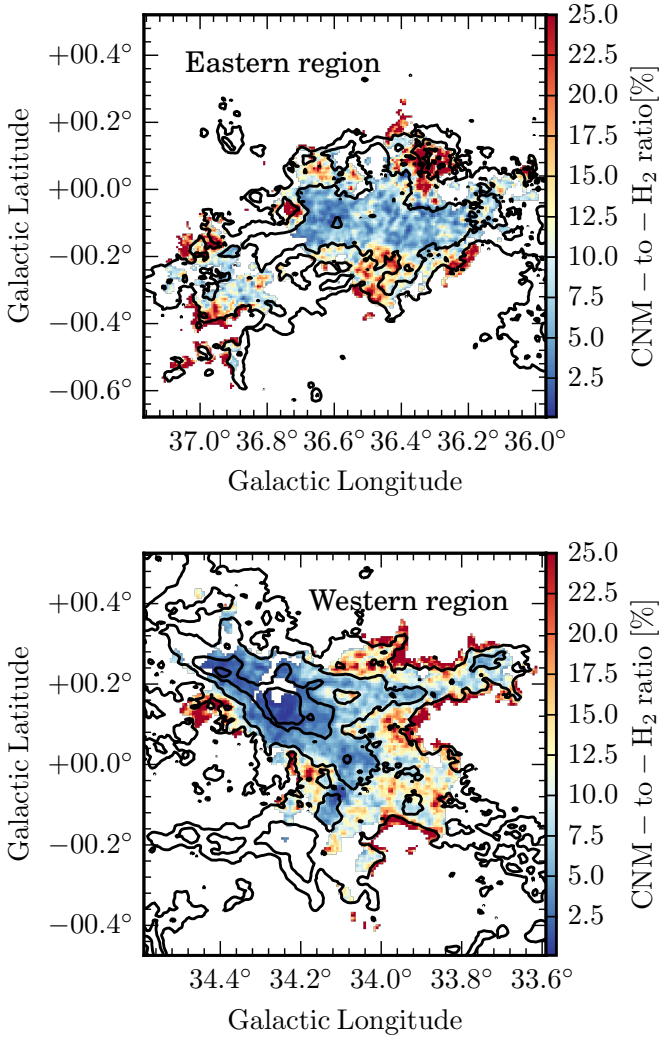


**Fig. 15.** Ratio of the CNM and $H_2$ column densities in percentage of the Eastern region (top panel) and Western region (bottom panel). The contours in both panels indicate the integrated $^{13}$CO emission at levels of 5, 10, 20, and 30 K km s$^{-1}$.

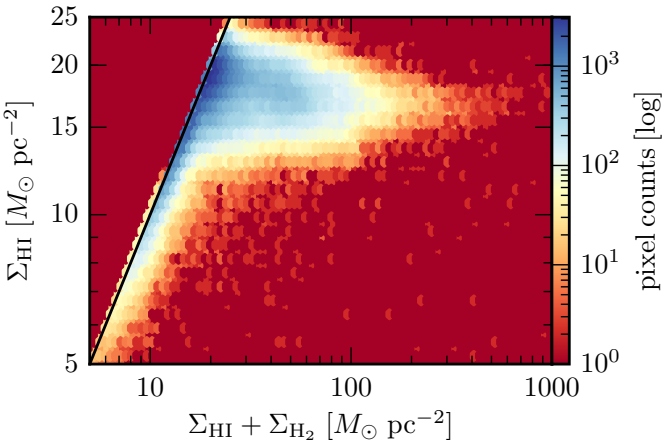


**Fig. 16.** Surface density of the atomic hydrogen ($v_{lsr}$ = 50 – 60 km s$^{-1}$) as a function of the total gas surface density ($\Sigma_{HI} + \Sigma_{H_2}$). The solid line represents a one-to-one relation.

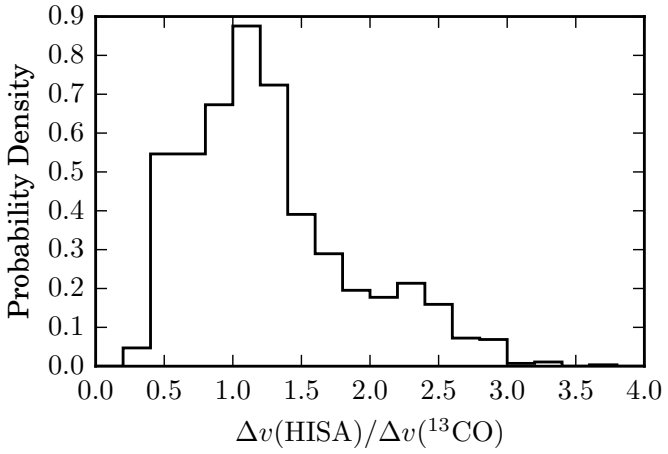


**Fig. 17.** Ratio of HISA and $^{13}$CO linewidths in the eastern region indicated in Fig. 11.

smaller than the temperature of the CMB. As we further discussed in Sect. 4.4, low values of $p \lesssim 0.7$ would also result an unrealistic low spin temperature.

For Galactocentric radius $7 \lesssim R \lesssim 35$ kpc, Kalberla & Dedes (2008) reported the average mid-plane volume density distribution of the atomic gas follows $n(R) \sim n_0\, e^{-(R-R_\odot)/R_n}$ with $n_0 = 0.9$ cm$^{-3}$, and $R_n$=3.15 kpc, $R_\odot$ =8.5 kpc (IAU recommendations). Assuming $n(R < 7\ \mathrm{kpc}) = n(7\ \mathrm{kpc})$, we integrate the density along the line of sight of GMF38a ($l = 35.5°$, $b = 0°$) and obtain the amount of gas in the foreground and background of GMF38a. Assuming optically thin emission, $p$ is estimated to be 0.91, which agrees with our adopted value of 0.9.

### 4.3. Optical depth measurement toward a strong continuum source

We can use strong continuum sources to estimate the optical depth of the CNM. The UCH II region G34.256+0.146 (see Fig. 1) is an ideal candidate to perform this task as it is very bright ($T_{\mathrm{cont}}(max) \sim 1300$ K) and slightly extended ($d \sim 70''$). We use the THOR data to extract the absorption spectrum towards the continuum peak and determine the optical depth and the lower-limit of the optical depth using Eq. 9. The optical depth spectrum is shown in Fig. 18. In the velocity range of the HISA feature, the absorption spectrum saturates and the determined optical depth $\tau = 3.5$ is a lower limit. Furthermore, since the UCH II region is at the same distance as the filament (Anderson et al. 2014) and associated with the filament, there could be CNM that are behind the UCH II region and are not traced by the absorption spectrum. Therefore, the determined optical depth represents a lower limit.

As explained in Sect. 2.2, the general HISA extraction method measures the optical depth together with the spin temperature and we are not able to disentangle them. 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 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 $\tau = 3.5$ using a black horizontal line. Assuming $p = 0.9$ reveals a spin temperature of $T_{\mathrm{HISA}} \sim 55$ K. This is a bit higher than the assumed spin temperature for the column density determination presented in Fig.14 (40 K). Since

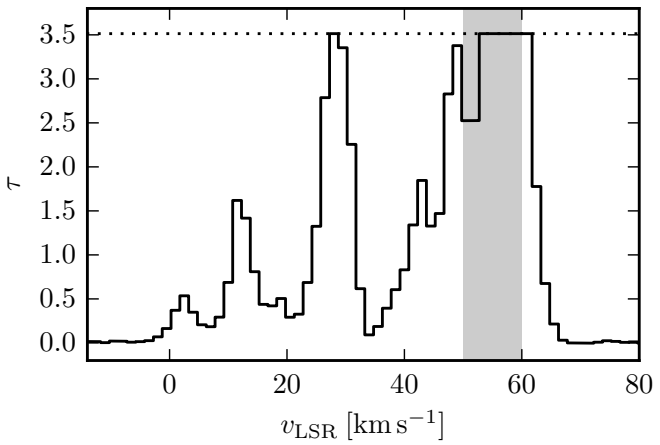


**Fig. 18.** Spectrum of the H I optical depth towards the UCH II region G34.256+0.146 using the THOR data. For some channels, the absorption spectra saturates and the measured optical depth is a lower limit of $\tau = 3.5$, which is indicated by the dotted line. The gray shaded area indicates the velocity range of the HISA feature ($v_{\rm LSR} = 50-60$ km s$^{-1}$).

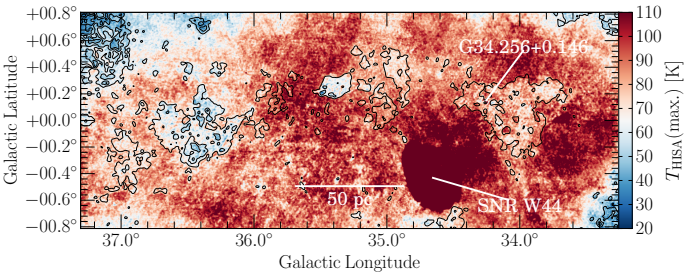


**Fig. 19.** Maximum spin temperature (Eq. 11) of the absorption features assuming $p = 0.9$. The black contours show the HISA column density map assuming $T_{\rm HISA}$ =40 K, $p = 0.9$, and the contour levels are 1.5, 3.5, and $5.5\times10^{20}$ cm$^{-2}$ with a smoothing of 3 pixels. Regions with strong continuum emission are not reliable, such as SNR W44 and G34.256+0.146 (see Fig. 1).

we measure the spin temperature close to an UCH II region, we expect rather high values.

A problem for HISA studies is that low level of background emission can be interpreted as an absorption feature. Studying very narrow absorption features, HINSA (e.g., Li & Goldsmith 2003; Goldsmith & Li 2005; Goldsmith et al. 2007; Krčo & Goldsmith 2010), can avoid this problem, since the steep absorption profiles of these HINSA features cannot be induced by two broad emission profiles on each side of the absorption feature. The broad HISA features we identified could be caused by two emission components. Fortunately, the optical depth information from the spectrum toward the strong continuum source helps to solve this problem. Since the optical depth is high ($\tau > 3.5$) for the velocity range of the HISA feature (Fig. 18), we are confident that we actually indeed observe a HISA feature rather than missing H I emission. Furthermore, the correlation of the HISA feature with the $^{13}$CO emission is another indicator of cold, dense H I.

### 4.4. Maximum spin temperature

As explained in Sect. 4.3, we can use strong continuum sources to measure the optical depth and therefore disentangle the spin temperature and the optical depth. However, this is only possible in a selected number of locations, that is, in the vicinity of a strong continuum source. In general we can only give the spin temperature as a function of the optical depth. Figure 4 shows that the function becomes very steep for certain spin temperatures. Hence, this 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 for $T_{\rm HISA}$:

$$T_{\rm HISA} = \frac{T_{\rm on-off}}{1-{\rm e}^{-\tau}} + p\,T_{\rm off} + T_{\rm cont}. \quad (10)$$

Since $T_{\rm on-off}$ is always negative, $T_{\rm HISA}$ reaches an upper limit for a given $T_{\rm off}$ and $T_{\rm cont}$ when $\frac{T_{\rm on-off}}{1-{\rm e}^{-\tau}}$ has its minimum value, which occurs for $\tau \to \infty$. This means the maximal $T_{\rm HISA}$ is:

$$T_{\rm HISA}(max.) = T_{\rm on-off} + p\,T_{\rm off} + T_{\rm cont}. \quad (11)$$

This equation depends on the assumption of the ratio of foreground and background emission, which is described by the factor $p$. For $p = 1$, the upper limit of the spin temperature reaches 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 $p = 0.9$. The result is given in Fig.19. We focus the discussion on regions offset from strong continuum sources, since for strong continuum sources the attenuation of the continuum exceeds by far the contribution of self-absorption to the H I absorption spectrum. As expected, we see a clear anti-correlation of the upper limit for the spin temperature and the column density of the HISA feature. A weaker absorption feature ($T_{\rm on-off}$ is always negative) will result in a higher $T_{\rm HISA}$, and a lower column density. The lowest values are found for the compact HISA feature in the center around $l = 35.5°$ with values around $T_{\rm HISA}(max.) \sim 25$ K. Similar values can be found for the eastern region of the filament, whereas the western side of the filament shows in general higher values around $T_{\rm HISA}(max.) \sim 75$ K. As this is only an upper limit for the spin temperature, we cannot directly infer the actual temperature. However, it is plausible that the H I spin temperature is higher on the western side of the cloud due to star formation activity and feedback processes, such as the prominent UCH II region. We will discuss this aspect further in Sect. 4.6.

For the H I column density determination in Sect. 3.2, we assumed a spin temperature of $T_{\rm HISA} = 40$ K. As seen in Fig. 19, this is higher than the upper limit of the spin temperature for certain cold regions. Hence for a few channels we cannot determine a column density for these regions, we exclude these pixels in these channels from our column density calculation. Since we observe only small regions with $T_{\rm HISA}(max.) < 40$ 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 would make the determined column density unreliable.

Furthermore, as indicated in Equation 11, the factor $p$ also affects the value of $T_{\rm HISA}(max.)$. If we take $p = 0.7$, the $T_{\rm HISA}(max.)$ for the whole filament would drop $\sim$ 20 K. Regions around $l = 35.5°$ and in the eastern part of the filament would have a $T_{\rm HISA}(max.) < 10$ K, which is highly unlikely, since simulations find very little H I has a temperature $\lesssim$20 K (Glover & Smith 2016). Furthermore, previous HISA study towards nearby molecular clouds reveals a $T_{\rm S}$ of 20-80 K (Gibson et al. 2000, 2005b; Dénes et al. 2018). Therefore, our assumption of $T_{\rm HISA} = 40$ K and $p = 0.9$ in the previous sections is reasonable.

### 4.5. Column density probability density functions (N-PDFs)

The column density maps derived in Sect. 3.2 can be utilized to determine the probability density functions of column densities (N-PDFs). We resampled all the column density maps into the same spatial resolution and constructed the N-PDFs. Figure 20 presents the N-PDFs in units of hydrogen atoms per square cm for the entire filament traced by HISA, H I emission and $^{13}$CO. For the HISA feature, we assumed a spin temperature of $T_{\rm HISA}$ = 40 K, $p$ = 0.9, and used a second order polynomial to estimate the background emission, which are the same assumptions used to produce Fig. 14. We derived the $H_2$ column density from $^{13}$CO (see Sect. 3.2.1), and converted it into the unit of hydrogen atoms per square centimeter for easy comparison with HISA and H I emission. We calculated the column density traced by H I emission with optical depth correction (see Sect. 3.3).

In the following, we first consider the quantification of the shapes of the observed N-PDFs and then the interpretation of the observed shapes. It has been argued that it may be necessary to consider "completeness" of the N-PDFs when defining the column density range that can be studied (Kainulainen & Tan 2013, see also Brunt 2015; Lombardi et al. 2015; Alves et al. 2017), although, it is unclear if such a requirement is meaningful when studying possibly turbulence-dominated gas (Körtgen et al. 2019b). In our analysis, we define the completeness for the HISA and $^{13}$CO data, but do not apply such requirement for the H I emission data that we expect to be clearly turbulence-dominated. We define the completeness with the help of the lowest "closed contours" in the column density maps. These lowest closed contours are $1.5 \times 10^{20}$ cm$^{-2}$ for the HISA data and $1.2 \times 10^{21}$ cm$^{-2}$ for the $^{13}$CO data and lead to regions that are marked with red polygons in Fig. 14. For H I emission, we include all the data within $32.65^\circ < l < 37.27^\circ$, $b < 1.1^\circ$ and use the column density level of $1.7 \times 10^{21}$ cm$^{-2}$ to define the range over which the range is analyzed. We normalize all N-PDFs by the mean column density of the tracer (shown in Table 2).

We first note that the N-PDFs of all components appear curved in log-log representation, even when only taking into account the data above the completeness levels. Therefore, we do not make an effort to quantify the N-PDFs with single power-law functions. Given the appearance of the N-PDFs, we decided to quantify their shapes through fits of log-normal functions. The N-PDF from H I emission shows a clear log-normal shape. We cannot identify a clear peak in the N-PDFs of HISA and $^{13}$CO, but proceed with the log-normal fit nevertheless. The best fit for HISA is not well constrained and shows some excess over the fit at the high column density side. The best fit for $^{13}$CO agrees well with a log-normal function over a wide range at low column densities, but shows an excess over it at higher columns; a power-law could also describe the high column densities relatively well. We fitted the power-law tail of the $^{13}$CO N-PDF with a power-law ($p(x) \propto x^{-\alpha}$) from the optimal column density threshold $N_{\rm min}$, which results in the minimum Kolmogorov-Smirnov distance between the fit and the N-PDF. The fit was performed with the python package *Powerlaw* (Alstott et al. 2014). The fitted parameters of the log-normal functions and power-laws, when applicable, are listed in Table 2.

The N-PDF from the H I emission peaks at $2.2{\times}10^{21}$ cm$^{-2}$, corresponding to $A_V$ = 1 (Güver & Özel 2009). Studies towards nearby clouds have found that the N-PDFs from H I emission there peak around $\sim 1-2\times10^{21}$ cm$^{-2}$ (Imara & Burkhart 2016), slight lower than we find here for our target. The widths of the fitted log-normal models vary among the tracers; the H I emission has the smallest width, followed by HISA and the $^{13}$CO that has the widest width (see Table 2). The narrowness of the H I emission N-PDF indicates that H I emission is relatively smoothly distributed without large variations and substructure (Fig. 14). The width of H I emission N-PDF is similar to those seen towards nearby clouds (Burkhart et al. 2015b; Imara & Burkhart 2016). We note that the width of the HISA N-PDF is much larger than the error introduced by our HISA extraction method (see Appendix B); we consider it is robustly between the widths of H I emission and $^{13}$CO.

We also examine the N-PDF of "all gas", derived by adding together the column density maps from all three tracers (Fig. 21). We note that "all gas" does not, in fact, trace all the gas, since there is some amount of $H_2$ that $^{13}$CO does not trace (Pineda et al. 2008; Goodman et al. 2009), aka., CO-dark gas. This fraction could be as high as 26% to 79% of the total $H_2$ gas (Gong et al. 2018). The N-PDF of "all gas" (Fig. 21) can not be fitted with single log-normal or power-law function. Similar to what we applied to the $^{13}$CO, we fitted the N-PDF with a log-normal function and the high-column density side with a power-law, but obviously, a single power law cannot account for the peak in the observed distribution.

Next, we discuss the N-PDFs of the two subregions of the filament. The N-PDFs for the eastern and western regions, indicated with dashed polygons in Fig. 14, are shown in Fig. 22. The N-PDFs of the cold dense H I traced by HISA for both subregions still can be described by a single log-normal function with a similar width as that found over the whole filament. The N-PDFs of the molecular gas show very different shapes for two subregions. While the N-PDF for the eastern subregion can be described by a log-normal function, the N-PDF for the western subregion shows a clear power-law shape.

Recall that we pointed out in Sect. 1 that theoretical studies predicted that the shapes of the N-PDFs are depend on the physical processes acting within the cloud, a log-normal N-PDF indicates that turbulent dominates and a power-law indicates that gravity dominates (e.g., Federrath et al. 2010; Ballesteros-Paredes et al. 2011; Kritsuk et al. 2011; Federrath & Klessen 2013; Burkhart et al. 2015a). A possible explanation for the different N-PDF for different subregions is that the western subregion shows ongoing high-mass star formation activities, that is, UCH II region, indicating that the molecular cloud is dominated by gravitational collapse.

One possible reason we do not see a power-law tail in the N-PDF of the molecular gas in the eastern subregion is that the $^{13}$CO could be frozen onto the dust grains in the densest and coldest part of the molecular cloud (e.g., Giannetti et al. 2014). In this case we could not recover the high density part, whereas in the western subregion the feedback effects from the UCH II have released the $^{13}$CO molecules from the dust grains. As the line is not excited at very low density, altogether $^{13}$CO only traces the intermediate density gas between the dense star forming cores, and so these do not show up in the N-PDF (see e.g., Ossenkopf 2002 for a similar effect in the $\Delta$-variance.)

We estimated the $^{13}$CO depletion factor by comparing the molecular cloud column density we derived from $^{13}$CO with the $H_2$ column density derived from the ATLASGAL dust continuum. Four ATLASGAL dense clumps (AGAL036.406+00.021, AGAL036.666-00.114, AGAL036.826-00.039, and AGAL036.839-00.022) that have velocities within the respective $^{13}$CO velocity range are located in the eastern subregion (Urquhart et al. 2018). We smoothed the ATLASGAL image into the same pixel size and angular resolution calculated the column density of molecular hydrogen following Eq. 15 in Giannetti et al. (2014). The same dust

**Table 2.** Results of the fits to the N-PDFs.

| Component | $N$ Threshold $\log_{10}(N(\mathrm{H}))$ | $\langle N(\mathrm{H})\rangle$ $\log_{10}(N(\mathrm{H}))$ | Log-normal width $\sigma$ | power-law index $\alpha$ | $N_{\mathrm{min}}$ $\log_{10}(N(\mathrm{H}))$ |
|---|---|---|---|---|---|
| CNM | 20.18 | 20.39 | 0.31 | – | – |
| H I emission | 21.23 | 21.34 | 0.12 | – | – |
| $H_2$ | 21.08 | 21.70 | 0.71 | 3.40 | 22.25 |
| All gas | 21.23 | 21.59 | 0.51 | 3.59 | 22.33 |
| CNM West | 20.18 | 20.42 | 0.27 | – | – |
| CNM East | 20.18 | 20.38 | 0.35 | – | – |
| $H_2$ West | 21.64 | 22.03 | 0.71 | 2.61 | 21.64 |
| $H_2$ East | 21.64 | 21.86 | 0.71 | – | – |

opacity $\kappa = 1.8$ cm$^2$ g$^{-1}$ as that used by Giannetti et al. (2014) is employed, and we assume the dust temperature from Herschel Hi-GAL results (Marsh et al. 2017). The $^{13}$CO depletion factors we derived are between 2 and 4, with a mean value of 3, which agrees with the numbers Giannetti et al. (2014) derived for infrared dark sources. Therefore, the high column density part of the N-PDF in the eastern subregion could be underestimated due to $^{13}$CO depletion.

### 4.6. Evolutionary stages

As mentioned in Sect. 3.2, we observe a significant difference in the distribution of the $H_2$ column density for the eastern and western subregion of the filament. The eastern subregion shows a more diffuse column density distribution, whereas the western subregion reveals several high column density peaks. Hence, the western subregion reveals a power-law in the N-PDFs shown in Fig. 22 and the eastern subregion shows a log-normal shaped PDF. Furthermore, we see a luminous UCH II region within the western subregion, whereas the eastern subregion does not harbor significant continuum emission. The ATLASGAL survey (Schuller et al. 2009) shows strong extended submillimeter emission with a group of star forming clumps in the western subregion, whereas only a few unresolved continuum clumps are found with velocities within the respective $^{13}$CO velocity range in the eastern subregion (Ragan et al. 2014; Urquhart et al. 2018). All these different tracers are indicative of active high mass star formation with strong feedback effects on the western side of the filament, while the eastern side shows no effects of feedback.

The kinematics of the HISA do not exhibit significant differences between the eastern and western subregion. However, comparing the HISA and $^{13}$CO kinematics shows an interesting difference. The eastern subregion shows similar peak velocities for the H I and CO, whereas the western subregion reveals a difference of ~4 km s$^{-1}$. Using the newly developed histogram of oriented gradients (HOG) tool, Soler et al. (2018) confirmed morphological correlation between the H I and $^{13}$CO emission in velocity channels separated by 3 to 4 km s$^{-1}$ towards the western subregion. The H I at ~54 km s$^{-1}$ is spatially correlated with $^{13}$CO emission at ~57.5 km s$^{-1}$ (Fig. 15 in Soler et al. 2018). They further demonstrated by applying the HOG analysis on synthetic observations from MHD simulations that this velocity offset between $^{13}$CO and H I could arise from more general molecular cloud formation conditions. Another explanation for this velocity offset could be the feedback from the expanding UCH II region G34.256+0.146. A significant amount of the $^{13}$CO gas at ~ 54 km s$^{-1}$ has been driven away by the radiation from the forming high-mass star.

Studying the cold H I column density, we found higher column density peaks on the eastern side in comparison to a more diffuse H I column density structure on the western side. The maximum spin temperature also shows smaller values on the eastern side. This might be an indication of a younger, colder, and more dense H I cloud on the eastern side in comparison to a more evolved cloud on the western side. It is possible that the dense H I cloud on the eastern 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 hypothesis.

### 4.7. 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 contributions: the ratio of foreground to background emission – factor $p$, different methods to determine the background emission and the assumption of the spin temperature $T_{\mathrm{HISA}}$.

Figure 4 also shows that for a fixed spin temperature $T_{\mathrm{HISA}}$, the larger the value of $p$ is, the lower the optical depth $\tau$ is, hence the lower the column density is. Depending on the $p$ value we choose (between 0.7 and 0.9), the column density can change by at most a factor of ~ 2. This is shown in Fig. 23 for N-PDFs of the entire filament assuming three different values for $p$. The column density structure stays almost constant, but the actual values are shifted for different $p$ values.

As discussed in Sect. 2.3, the chosen method for the background estimate can influence the absorption depth of the HISA feature and thus the column density. We showed in Sect. 2.3 that the best method is a polynomial 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. As we can see in the bottom panel of Fig. 23, the column density structures of the second and fourth order polynomial are almost identical, the mean column density of the fourth order polynomial fit is ~ 3% higher than the one of second order. Thus we chose to use a second order polynomial fit for $T_{\mathrm{off}}$.

Another important factor is the assumption of a constant spin temperature for the cloud. This is obviously a poor assumption, but using additional measurements we can constrain the range of the spin temperature. The most important one is the upper limit for the spin temperature introduced in Sect. 4.4. Using this information, we can constrain the spin temperature to values of $T_{\mathrm{HISA}} < 70$ K for the majority of the HISA features. For the CNM mass estimation given in Sect. 3.3, we assumed a spin temperature of $T_{\mathrm{HISA}} = 20$ and 40 K. As seen in Fig. 23, the N-PDF does not change significantly, but higher spin temperatures result in larger column densities and masses. In Sect. 3.3, we showed

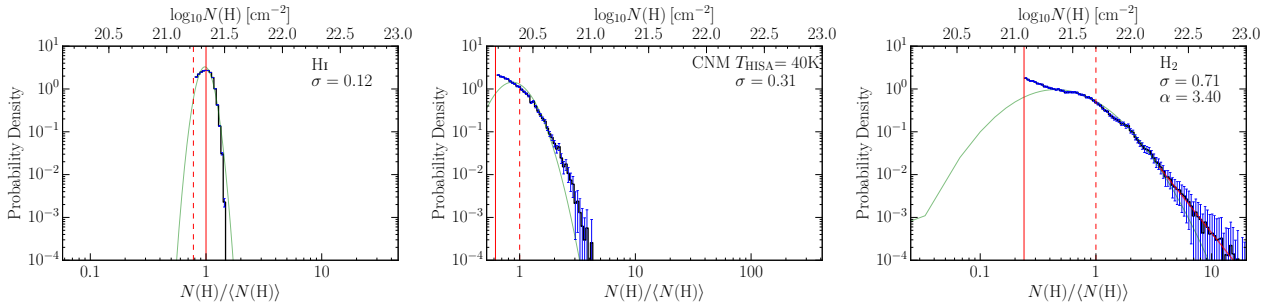


**Fig. 20.** N-PDFs of the gas traced by H I (left panel) emission, HISA (middle panel), and $^{13}$CO (right panel) in units of hydrogen atoms per square cm. The light green lines show the log-normal fit in each panel. $\sigma$ is the width of the log-normal distribution. The dashed vertical lines in each panel mark the column density threshold, and the solid vertical lines mark the mean column densities. The red line in the $^{13}$CO panel shows the power-law fit with an index $-\alpha$ to the high column density tail.

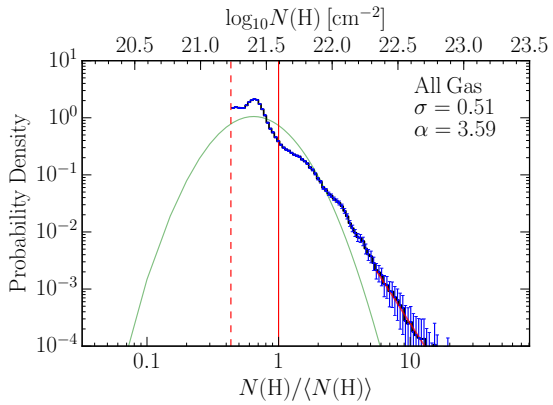


**Fig. 21.** N-PDF of the combination of atomic and molecular gas over the entire region in Fig. 14. The green line shows the log-normal fit. The dashed vertical line mark the column density threshold (the contour levels in Fig. 14), and the vertical solid line mark the mean column density. The red line shows the power-law fit with an index $-\alpha$ to the high column density tail.

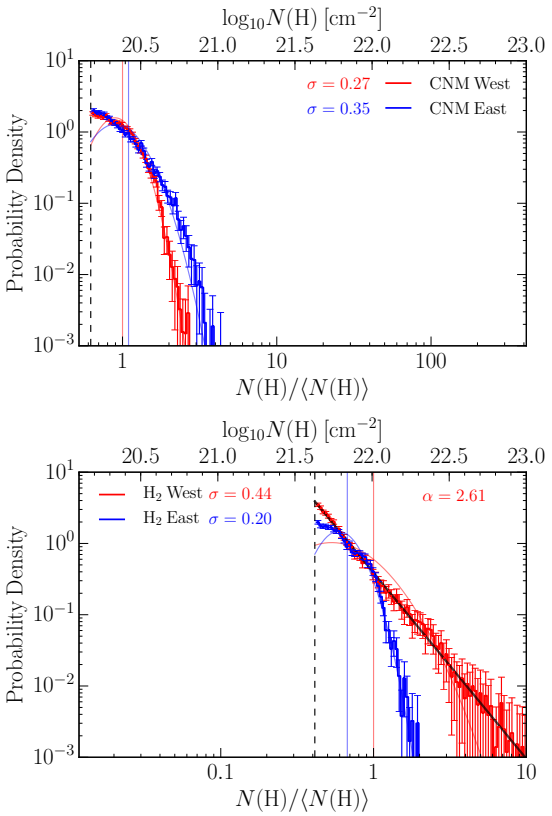


**Fig. 22.** N-PDFs for the two subregions within the filament marked with dashed polygons in Fig. 14. The top and bottom panel show the N-PDF of atomic and molecular hydrogen derived with HISA and $^{13}$CO measurements, respectively. The lighter lines show the log-normal (top and bottom panel) and power-law (bottom panel) fits. The black dashed lines in each panel mark the column density threshold, and the solid vertical lines indicate the mean column densities. The bottom x-axes in both panels are the relative column density for the N-PDF of the western subregion, i.e., the red histogram in both panels. Other N-PDFs are shifted to the correct absolute column density for comparison. The thick black line in the molecular component panel shows the power-law fit with an index $-\alpha$ to the column density distribution of the western region.

that the mass is about a factor of three larger for $T_{\rm HISA}$ = 40 K with respect to that assuming $T_{\rm HISA}$ = 20 K.

Furthermore, we performed our HISA extraction method on model images (see Appendix B) to estimate the uncertainty brought in by the $T_{\rm off}$ fitting method to the width of the N-PDF. The test results show that the "instrument broadening" introduced by the fitting method to the N-PDF is not very large and the log-normal width of the CNM we derive in the paper is robust.

In summary, it is difficult to quantify exactly the uncertainty of the CNM column density and mass. Considering all assumptions, the CNM mass has an uncertainty of a factor of 2-3 (dominated by the uncertainty of $T_{\rm S}$), which is similar to the $H_2$ mass uncertainty based on $^{13}$CO. However, we showed that the shape of the column density PDF is robust.

## 5. Conclusions

We studied atomic and molecular gas components of the giant molecular filament GMF38a. The molecular component is traced via observations of $^{13}$CO, whereas the cold atomic gas is observed via HISA features and H I 21 cm emission. The main results can be summarized as:

1. We extracted HISA by estimating the background emission with different methods. For the observed giant filament, a polynomial fit of second order to the neighboring channels of each HISA is the most reliable method from the different options we tested to estimate the background emission.
2. The HISA features and the $^{13}$CO emission are spatially correlated. While in the eastern subregion they correlate well,

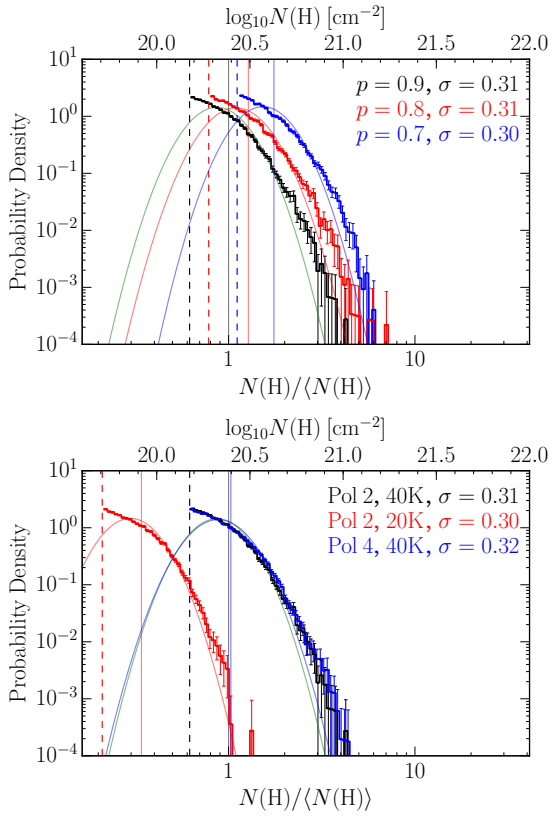


**Fig. 23.** Top panel: N-PDFs for the atomic hydrogen for different $p$. The PDF is measured over the entire filament indicated with a red polygon in Fig. 14, assuming a constant spin temperature of $T_{\rm HISA}$ = 40 K. Bottom panel: N-PDFs for the atomic hydrogen for different spin temperature and different polynomial functions fit to get $T_{\rm off}$. The dashed vertical lines in each panel indicate the column density threshold, and the solid vertical lines indicate the mean column densities. The bottom x-axes in both panels are the relative column densities for the N-PDF of "Pol 2, $p$=0.9, 40K", i.e., the black histogram in both panels. Other N-PDFs are shifted to the correct absolute column density for comparison.

**Table 3.** Results of the fits to the N-PDFs of CNM for different parameters (Fig. 23).

| Pol. | p | $T_{\rm HISA}$ [K] | $N$ Threshold $\log_{10}(N(\rm H))$ | $\langle N(\rm H)\rangle$ $\log_{10}(N(\rm H))$ | width $\sigma$ |
|---|---|---|---|---|---|
| Pol 2 | 0.9 | 40 | 20.18 | 20.39 | 0.31 |
| Pol 2 | 0.8 | 40 | 20.28 | 20.49 | 0.31 |
| Pol 2 | 0.7 | 40 | 20.43 | 20.63 | 0.30 |
| Pol 2 | 0.9 | 20 | 19.71 | 19.92 | 0.30 |
| Pol 4 | 0.9 | 40 | 20.18 | 20.40 | 0.32 |

the peak velocities of the two tracers show an offset of ~4 km s$^{-1}$ for the western subregion of the filament.

3. Although the linewidth ratio between HISA and $^{13}$CO is around unity, the Mach number estimation shows that the $^{13}$CO emission is dominated by supersonic turbulent motions, whereas a large fraction of the CNM is at subsonic or transonic velocities.
4. Assuming a spin temperature of $T_{\rm HISA}$ = 40 K for the H I, we determined the column densities of the cold dense H I and compared them to the $H_2$ column density distribution derived from the $^{13}$CO emission. The column density peaks do not coincide and the H I column density shows in general a diffuse structure. The $H_2$ column density reveals prominent peaks in the western subregion of the filament whereas the eastern subregion appears more diffuse. The CNM-to-$H_2$ column density ratio varies between 0.5 to 25% with a median value of ~ 9%. The outer layer of the filament exhibits a higher ratio than that toward the cloud centers. The surface density of atomic hydrogen peaks at ~ 14 – 23 $M_\odot$ pc$^{-2}$ (corresponding to ~ $1.8-2.9\times10^{21}$ cm$^{-2}$). Furthermore, the mass traced by HISA is only ~ 3% of the molecular mass, and ~ 1.6% of the mass traced by atomic H I emission.
5. Studying the N-PDFs, we are able to provide constraints on the physical processes within the cloud. The location of the HISA N-PDF is strongly dependent on the assumed parameters, but the width is not. The N-PDFs of CNM, H I emission, and $H_2$ can be described by a log-normal function, which indicates turbulent motions as the main physical driver. Only the $H_2$ column density of the western subregion within the filament is characterized by a high column density power-law structure, consistent with the observed star formation activity. Adding the column density maps of all three tracers (CNM, H I emission, and $H_2$) up, we generated a column density map of "all gas."
6. We hypothesize that the eastern and western sides of the filament represent different evolutionary stages. The eastern side represents an earlier stage, which is currently forming a dense molecular cloud out of the atomic reservoir. As we do not observe high molecular column density peaks, the H I shows low spin temperatures and high column densities. In contrast, the western side of the filament shows high $H_2$ column density peaks, signs of active star formation, such as UCH II regions, and, in general, a warmer and less dense atomic counterpart. These differences provide interesting constraints for theoretical models and simulations of the formation of molecular clouds.

*Acknowledgements.* The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. Y.W., H.B., S.B., and J.D.S. acknowledge support from the European Research Council under the Horizon 2020 Framework Program via the ERC Consolidator Grant CSF-648505, and R.S.K. via the ERC Advanced-Grant 339177 (STARLIGHT). H.B., S.C.O.G., and M.R. acknowledge support from the Deutsche Forschungsgemeinschaft in the Collaborative Research Center (SFB 881) "The Milky Way System" (subproject B1, B2, B8). R.S.K. and S.C.O.G also acknowledge support from the DFG via Germany's Excellence Strategy EXC-2181/1 - 390900948 (the Heidelberg STRUCTURES Cluster of Excellence). This work was carried out in part at the Jet Propulsion Laboratory which is operated for NASA by the California Institute of Technology. N.S. acknowledges support by the french ANR and the german DFG through the project "GENESIS" (ANR-16-CE92-0035-01/DFG1591/2-1). F.B. acknowledges funding from the European Union's Horizon 2020 research and innovation program (grant agreement No 726384). This research made use of Astropy and affiliated packages, a community-developed core Python package for Astronomy (The Astropy Collaboration et al. 2018), Python package *SciPy*[3], APLpy, an open-source plotting package for Python (Robitaille & Bressert 2012), and software TOPCAT (Taylor 2005). The authors thank the anonymous referee for the constructive comments that improve the paper.

## References

Abreu-Vicente, J., Kainulainen, J., Stutz, A., Henning, T., & Beuther, H. 2015, A&A, 581, A74
Abreu-Vicente, J., Ragan, S., Kainulainen, J., et al. 2016, A&A, 590, A131
Allen, C. W. 1973, Astrophysical quantities
Alstott, J., Bullmore, E., & Plenz, D. 2014, PLoS ONE, 9, e85777
Alves, J., Lombardi, M., & Lada, C. J. 2017, A&A, 606, L2
Alves de Oliveira, C., Schneider, N., Merín, B., et al. 2014, A&A, 568, A98
Anderson, L. D., Bania, T. M., Balser, D. S., et al. 2014, ApJS, 212, 1
Anderson, L. D., Wang, Y., Bihr, S., et al. 2017, A&A, 605, A58

[3] https://www.scipy.org

Andersson, B.-G., Wannier, P. G., & Morris, M. 1991, ApJ, 366, 464
Ballesteros-Paredes, J., Vázquez-Semadeni, E., Gazol, A., et al. 2011, MNRAS, 416, 1436
Beuther, H., Bihr, S., Rugel, M., et al. 2016, A&A, 595, A32
Bialy, S., Bihr, S., Beuther, H., Henning, T., & Sternberg, A. 2017a, ApJ, 835, 126
Bialy, S., Burkhart, B., & Sternberg, A. 2017b, ApJ, 843, 92
Bihr, S., Beuther, H., Ott, J., et al. 2015, A&A, 580, A112
Blitz, L., Fukui, Y., Kawamura, A., et al. 2007, Protostars and Planets V, 81
Bowers, P. F., Kerr, F. J., & Hawarden, T. G. 1980, ApJ, 241, 183
Brunt, C. M. 2015, MNRAS, 449, 4465
Burkhart, B., Collins, D. C., & Lazarian, A. 2015a, ApJ, 808, 48
Burkhart, B., Lee, M.-Y., Murray, C. E., & Stanimirović, S. 2015b, ApJ, 811, L28
Caldú-Primo, A., Schruba, A., Walter, F., et al. 2013, AJ, 146, 150
Chen, H. H.-H., Burkhart, B., Goodman, A., & Collins, D. C. 2018, ApJ, 859, 162
Clark, P. C., Glover, S. C. O., Klessen, R. S., & Bonnell, I. A. 2012, MNRAS, 424, 2599
Cox, A. N. 2000, Allen's astrophysical quantities
Dénes, H., McClure-Griffiths, N. M., Dickey, J. M., Dawson, J. R., & Murray, C. E. 2018, MNRAS, 479, 1465
Dobbs, C. L., Krumholz, M. R., Ballesteros-Paredes, J., et al. 2014, Protostars and Planets VI, 3
Federrath, C. & Klessen, R. S. 2012, ApJ, 761, 156
Federrath, C. & Klessen, R. S. 2013, ApJ, 763, 51
Federrath, C., Roman-Duval, J., Klessen, R. S., Schmidt, W., & Mac Low, M.-M. 2010, A&A, 512, A81
Feldt, C. 1993, A&A, 276, 531
Fixsen, D. J. 2009, ApJ, 707, 916
Fixsen, D. J., Cheng, E. S., Gales, J. M., et al. 1996, ApJ, 473, 576
Giannetti, A., Wyrowski, F., Brand, J., et al. 2014, A&A, 570, A65
Gibson, S. J., Taylor, A. R., Higgs, L. A., Brunt, C. M., & Dewdney, P. E. 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
Glover, S. C. O. & Smith, R. J. 2016, MNRAS, 462, 3011
Goldsmith, P. F. 2013, ApJ, 774, 134
Goldsmith, P. F. & Li, D. 2005, ApJ, 622, 938
Goldsmith, P. F., Li, D., & Krčo, M. 2007, ApJ, 654, 273
Gong, M., Ostriker, E. C., & Kim, C.-G. 2018, ApJ, 858, 16
Goodman, A. A., Alves, J., Beaumont, C. N., et al. 2014, ApJ, 797, 53
Goodman, A. A., Pineda, J. E., & Schnee, S. L. 2009, ApJ, 692, 91
Green, D. A. 2014, Bulletin of the Astronomical Society of India, 42, 47
Güver, T. & Özel, F. 2009, MNRAS, 400, 2050
Heeschen, D. S. 1954, AJ, 59, 324
Heeschen, D. S. 1955, ApJ, 121, 569
Heiles, C. 1997, ApJ, 481, 193
Heiles, C. & Gordon, M. A. 1975, ApJ, 199, 361
Hennebelle, P. & Falgarone, E. 2012, Astronomy and Astrophysics Review, 20, 55
Imara, N. & Burkhart, B. 2016, ApJ, 829, 102
Jackson, J. M., Finn, S. C., Chambers, E. T., Rathborne, J. M., & Simon, R. 2010, ApJ, 719, L185
Jackson, J. M., Rathborne, J. M., Shah, R. Y., et al. 2006, ApJS, 163, 145
Jenkins, E. B. & Tripp, T. M. 2011, ApJ, 734, 65
Kainulainen, J., Beuther, H., Henning, T., & Plume, R. 2009, A&A, 508, L35
Kainulainen, J. & Tan, J. C. 2013, A&A, 549, A53
Kalberla, P. M. W. & Dedes, L. 2008, A&A, 487, 951
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
Klessen, R. S. & Glover, S. C. O. 2016, Star Formation in Galaxy Evolution: Connecting Numerical Models to Reality, Saas-Fee Advanced Course, Volume 43. ISBN 978-3-662-47889-9. Springer-Verlag Berlin Heidelberg, 2016, p. 85, 43, 85
Knapp, G. R. 1974, AJ, 79, 527
Körtgen, B., Federrath, C., & Banerjee, R. 2019a, MNRAS, 482, 5233
Körtgen, B., Federrath, C., & Banerjee, R. 2019b, MNRAS, 482, 5233
Kritsuk, A. G., Norman, M. L., & Wagner, R. 2011, ApJ, 727, L20
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
Krčo, M., Goldsmith, P. F., Brown, R. L., & Li, D. 2008, ApJ, 689, 276
Larson, R. B. 1981, MNRAS, 194, 809
Larson, R. B. 2003, Reports on Progress in Physics, 66, 1651
Lee, M.-Y., Stanimirović, S., Murray, C. E., Heiles, C., & Miller, J. 2015, ApJ, 809, 56
Levinson, F. H. & Brown, R. L. 1980, ApJ, 242, 416
Li, D. & Goldsmith, P. F. 2003, ApJ, 585, 823
Li, G.-X., Urquhart, J. S., Leurini, S., et al. 2016, A&A, 591, A5
Lin, Y., Liu, H. B., Dale, J. E., et al. 2017, ApJ, 840, 22
Lindner, R. R., Vera-Ciro, C., Murray, C. E., et al. 2015, AJ, 149, 138
Lockman, F. J. 1984, ApJ, 283, 90
Lombardi, M., Alves, J., & Lada, C. J. 2015, A&A, 576, L1
Lombardi, M., Lada, C. J., & Alves, J. 2008, A&A, 489, 143
Marsh, K. A., Whitworth, A. P., Lomax, O., et al. 2017, MNRAS, 471, 2730
McClure-Griffiths, N. M., Dickey, J. M., Gaensler, B. M., Green, A. J., & Haverkorn, M. 2006, ApJ, 652, 1339
McCutcheon, W. H., Shuter, W. L. H., & Booth, R. S. 1978, MNRAS, 185, 755
McKee, C. F. & Ostriker, E. C. 2007, ARA&A, 45, 565
McKee, C. F. & Ostriker, J. P. 1977, ApJ, 218, 148
McMullin, J. P., Waters, B., Schiebel, D., Young, W., & Golap, K. 2007, in Astronomical Society of the Pacific Conference Series, Vol. 376, Astronomical Data Analysis Software and Systems XVI, ed. R. A. Shaw, F. Hill, & D. J. Bell, 127
Minn, Y. K. 1981, A&A, 103, 269
Mogotsi, K. M., de Blok, W. J. G., Caldú-Primo, A., et al. 2016, AJ, 151, 15
Montgomery, A. S., Bates, B., & Davies, R. D. 1995, MNRAS, 273, 449
Myers, P. C., Ho, P. T. P., Schneps, M. H., et al. 1978, ApJ, 220, 864
Ossenkopf, V. 2002, A&A, 391, 295
Ossenkopf-Okada, V., Csengeri, T., Schneider, N., Federrath, C., & Klessen, R. S. 2016, A&A, 590, A104
Ostriker, E. C., Stone, J. M., & Gammie, C. F. 2001, ApJ, 546, 980
Peñaloza, C. H., Clark, P. C., Glover, S. C. O., & Klessen, R. S. 2018, MNRAS, 475, 1508
Pineda, J. E., Caselli, P., & Goodman, A. A. 2008, ApJ, 679, 481
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, A&A, 594, A13
Ragan, S. E., Henning, T., Tackenberg, J., et al. 2014, A&A, 568, A73
Rebolledo, D., Green, A. J., Burton, M., et al. 2017, MNRAS, 472, 1685
Reid, M. J., Dame, T. M., Menten, K. M., & Brunthaler, A. 2016, ApJ, 823, 77
Riegel, K. W. & Crutcher, R. M. 1972, A&A, 18, 55
Robitaille, T. & Bressert, E. 2012, APLpy: Astronomical Plotting Library in Python, Astrophysics Source Code Library
Roman-Duval, J., Jackson, J. M., Heyer, M., Rathborne, J., & Simon, R. 2010, ApJ, 723, 492
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
Schneider, N., André, P., Könyves, V., et al. 2013, ApJ, 766, L17
Schneider, N., Bontemps, S., Girichidis, P., et al. 2015, MNRAS, 453, 41
Schneider, N., Bontemps, S., Motte, F., et al. 2016, A&A, 587, A74
Schuller, F., Menten, K. M., Contreras, Y., et al. 2009, A&A, 504, 415
Shuter, W. L. H., Dickman, R. L., & Klatt, C. 1987, ApJ, 322, L103
Soler, J. D., Beuther, H., Rugel, M., et al. 2018, arXiv e-prints, arXiv:1809.08338
Stahler, S. W. & Palla, F. 2005, The Formation of Stars, 865
Sternberg, A., Le Petit, F., Roueff, E., & Le Bourlot, J. 2014, ApJ, 790, 10
Stil, J. M., Taylor, A. R., Dickey, J. M., et al. 2006, AJ, 132, 1158
Stutz, A. M. & Kainulainen, J. 2015, A&A, 577, L6
Szűcs, L., Glover, S. C. O., & Klessen, R. S. 2014, MNRAS, 445, 4055
Szűcs, L., Glover, S. C. O., & Klessen, R. S. 2016, MNRAS, 460, 82
Tan, J. C., Beltrán, M. T., Caselli, P., et al. 2014, Protostars and Planets VI, 149
Taylor, M. B. 2005, in Astronomical Society of the Pacific Conference Series, Vol. 347, Astronomical Data Analysis Software and Systems XIV, ed. P. Shopbell, M. Britton, & R. Ebert, 29
The Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, ArXiv e-prints [arXiv:1801.02634]
Umemoto, T., Minamidani, T., Kuno, N., et al. 2017, Publications of the Astronomical Society of Japan, 69, 78
Urquhart, J. S., König, C., Giannetti, A., et al. 2018, MNRAS, 473, 1059
van der Werf, P. P. & Goss, W. M. 1989, A&A, 224, 209
van der Werf, P. P., Goss, W. M., & Vanden Bout, P. A. 1988, A&A, 201, 311
Vazquez-Semadeni, E. 1994, ApJ, 423, 681
Wang, K., Testi, L., Burkert, A., et al. 2016, ApJS, 226, 9
Wang, K., Testi, L., Ginsburg, A., et al. 2015, MNRAS, 450, 4043
Wang, Y., Beuther, H., Rugel, M. R., et al. 2019, arXiv e-prints, arXiv:1912.08223
Wang, Y., Bihr, S., Rugel, M., et al. 2018, A&A, 619, A124
Ward, R. L., Wadsley, J., & Sills, A. 2014, MNRAS, 445, 1575
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)
Winnberg, A., Grasshoff, M., Goss, W. M., & Sancisi, R. 1980, A&A, 90, 176
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
Zhang, M., Kainulainen, J., Mattern, M., Fang, M., & Henning, T. 2019, A&A, 622, A52
Zucker, C., Battersby, C., & Goodman, A. 2015, ApJ, 815, 23
Zucker, C., Battersby, C., & Goodman, A. 2018, ApJ, 864, 153
Zuo, P., Li, D., Peek, J. E. G., et al. 2018, ApJ, 867, 13

## Appendix A: All gas

## Appendix B: CNM N-PDF width test

To test how much our HISA identification and fitting method broadens the N-PDF of the CNM, we tested our HISA extraction method on artificial H I emission maps with absorption features.

We made a model H I map with a relative uniform peak $T_{\rm B}$ ~ 85 K with a noise of ~ 4 K, which is similar to the real data we have. The H I spectra in the model map have a random linewidth varying between ~30 to 45 km s$^{-1}$, and a random peak velocity between 47.5 to 52.5 km s$^{-1}$. We also made a model continuum map with a relative uniform $T_{\rm B}$ ~ 13 K with a noise of 0.7 K, and both values are similar to the diffuse continuum emission flux and the noise level of the real data. Both model images have the same pixel size (10″) and beam size (40″) as the real data.

For the first test, we generate artificial absorption features from a single column density value with $T_{\rm HISA}$ =40 K and $p$ = 0.9, and added them into the model H I image. This column density value equals to the mean CNM column density of the GMF38a in Table 2 ($\log_{10}(N({\rm H}))$ = 20.39 cm$^{-2}$). The artificial absorption features peak at 50 km$^{-1}$ and have a linewidth of 5 km s$^{-1}$. Following the method described in Sect. 2 and 3, we extracted the HISA spectra, estimated the column density and construct the N-PDF shown in the top panel of Fig. B.1. The absorption features we put in the model image all have the same column density, so the modeled N-PDF is a delta function. Due to noise and the uncertainty we brought in through our HISA extraction method, the N-PDF we derived has a width of 0.19, and about 10% lower mean column density than the input one.

For the second test, instead of a single column density value, we generate artificial absorption features from a log-normal distribution column density with $T_{\rm HISA}$ =40 K and $p$ = 0.9, and added them into the model H I image. The input log-normal distribution has a width of 0.15, and a peak at the mean CNM column density of the GMF38a in Table 2 ($\log_{10}(N({\rm H}))$ = 20.39 cm$^{-2}$). Similarly, we ran our procedure on the model image and the N-PDF is shown in the bottom panel of Fig. B.1. The input width of 0.15 is broadened to 0.22, and the mean column density is also about 10% lower than the input one.

Both tests demonstrate that the "instrument broadening" contribution to the N-PDF of our method is not very large and the log-normal width of the CNM we derive in the paper is robust.

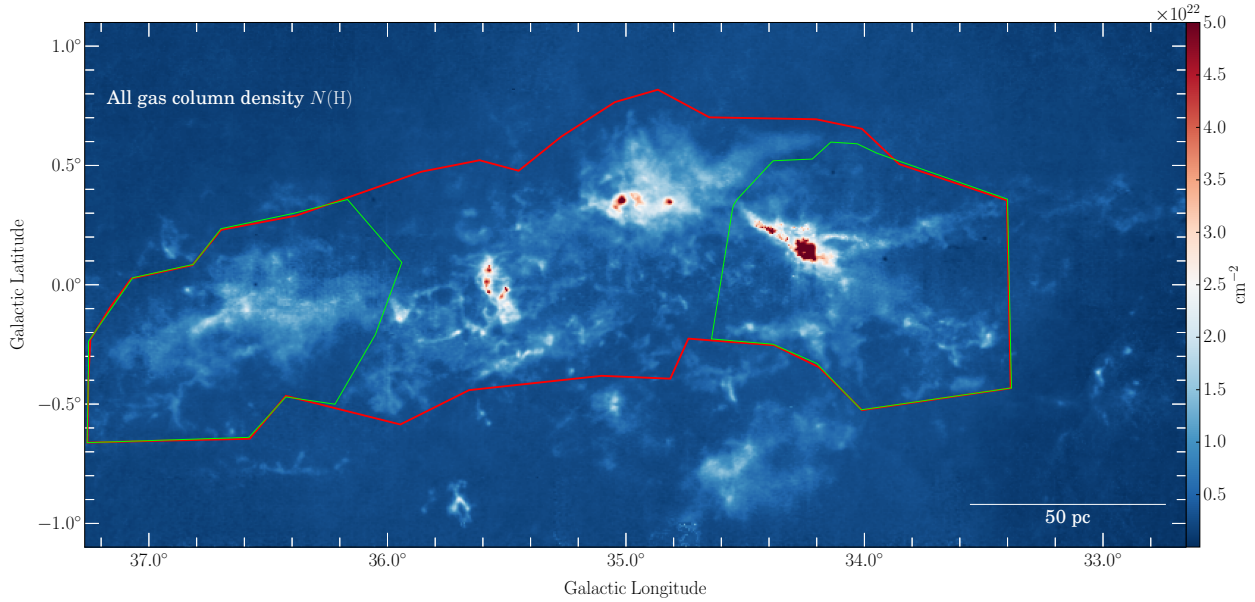


**Fig. A.1.** The column density map of the gas combined molecular component (from $^{13}$CO), CNM (from HISA), and WNM+CNM (from H I emission) (see Sect. 3.2.1). The red and green dashed polygons mark the region for the mass estimates and the column density PDF measurements (only molecular component and CNM) shown in Figs. 20 and 22, respectively.

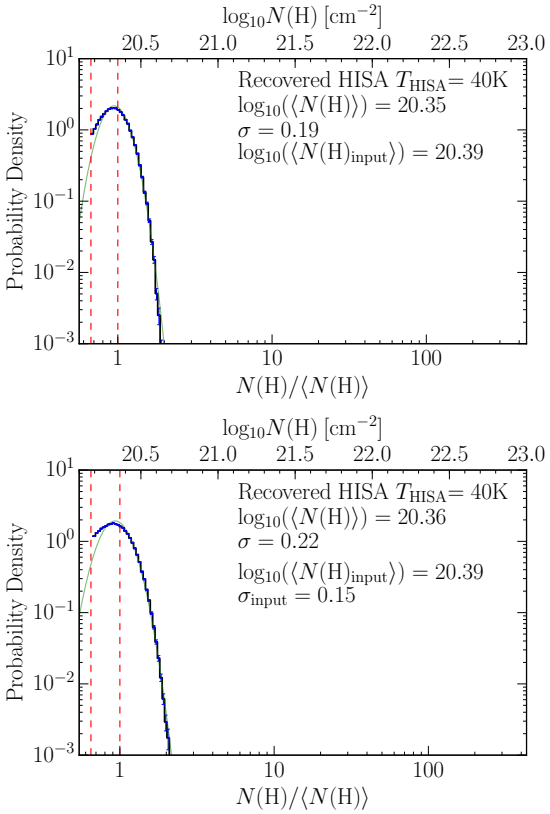


**Fig. B.1.** The N-PDFs of the CNM recovered from the tests.