Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **512,** 4765–4784 (2022) https://doi.org/10.1093/mnras/stac607
Advance Access publication 2022 March 7

On the accuracy of H I observations in molecular clouds – More cold H I than thought?

D. Seifried [1]★ H. Beuther,[2] S. Walch,[1] J. Syed [2] J. D. Soler,[2] P. Girichidis [3] and R. Wünsch [4]

[1] *I. Physikalisches Institut, Universität zu Köln, Zülpicher Str 77, D-50937 Köln, Germany*
[2] *Max Planck Institute for Astronomy, Königstuhl 17, D-69117 Heidelberg, Germany*
[3] *Zentrum für Astronomie, Institut für Theoretische Astrophysik, Universität Heidelberg, Albert-Ueberle-Str 2, D-69120 Heidelberg, Germany*
[4] *Astronomical Institute of the Czech Academy of Sciences, Bocní II 1401/1, CZ-141 00 Prague 4, Czech Republic*

Accepted 2022 March 2. Received 2022 March 2; in original form 2021 September 23

ABSTRACT

We present a study of the cold atomic hydrogen (H I) content of molecular clouds simulated within the SILCC-Zoom project for solar neighbourhood conditions. We produce synthetic observations of H I at 21 cm, including H I self-absorption (HISA) and observational effects. We find that H I column densities, $N_{\rm H\,I}$, of $\gtrsim 10^{22}$ cm^{-2} are frequently reached in molecular clouds with H I temperatures as low as ∼10 K. Hence, HISA observations assuming a fixed H I temperature tend to underestimate the amount of cold H I in molecular clouds by a factor of 3–10 and produce an artificial upper limit of $N_{\rm H\,I}$ around 10^{21} cm^{-2}. We thus argue that the cold H I mass in molecular clouds could be a factor of a few higher than previously estimated. Also, $N_{\rm H\,I}$ PDFs obtained from HISA observations might be subject to observational biases and should be considered with caution. The underestimation of cold H I in HISA observations is due to both the large H I temperature variations and the effect of noise in regions of high optical depth. We find optical depths of cold H I around 1–10, making optical depth corrections essential. We show that the high H I column densities ($\gtrsim 10^{22}$ cm^{-2}) can in parts be attributed to the occurrence of up to 10 individual H I–H_2 transitions along the line of sight. This is also reflected in the spectra, necessitating Gaussian decomposition algorithms for their in-depth analysis. However, also for a single H I–H_2 transition, $N_{\rm H\,I}$ frequently exceeds 10^{21} cm^{-2}, challenging one-dimensional, semi-analytical models. This is due to non-equilibrium chemistry effects and the fact that H I–H_2 transition regions usually do not possess a one-dimensional geometry. Finally, we show that the H I gas is moderately supersonic with Mach numbers of a few. The corresponding non-thermal velocity dispersion can be determined via HISA observations within a factor of ∼2.

Key words: MHD – radiative transfer – methods: numerical – ISM: atoms – ISM: clouds – radio lines: ISM.

1 INTRODUCTION

The chemical composition of molecular clouds (MCs) is still a field of active research. In particular, the transition from atomic hydrogen (H I) to molecular hydrogen (H_2) is of interest. During the initial formation of MCs, H I is continuously transformed into H_2 as the cloud collapses and its density increases (e.g. Glover & Mac Low 2007b; Glover et al. 2010; Clark et al. 2012b; Mac Low & Glover 2012; Seifried et al. 2017, but see also the review by Dobbs et al. 2014). Knowledge about the exact content of H I and H_2 in MCs would also be of great value to assess the amount of CO-dark H_2 gas, i.e. molecular gas that is not traced by CO emission and thus affects the $X_{\rm CO}$ factor (see e.g. the review by Bolatto, Wolfire & Leroy 2013).

Plenty of semi-analytical works studied the transition of H I- to H_2-dominated gas in MCs under various conditions with either plane-parallel or spherically symmetric models (e.g. van Dishoeck & Black 1986; Sternberg 1988; Röllig et al. 2007; Krumholz, McKee & Tumlinson 2008, 2009; Wolfire, Hollenbach & McKee 2010; Sternberg et al. 2014; Bialy & Sternberg 2016). For solar neighbourhood conditions, these models predict that the H I–H_2 transition occurs around column densities of $\sim 10^{20}$–10^{21} cm^{-2}. Furthermore, besides the column density where the *transition* occurs, Sternberg et al. (2014) and Bialy & Sternberg (2016) also find that the *total* H I column density, $N_{\rm H\,I}$, should be limited to a maximum of about 10^{21} cm^{-2}. These models are, however, typically highly idealized as they assume chemical equilibrium and are limited to a one-dimensional (1D) geometry, although attempts are made to extend them to turbulent environments (Bialy, Burkhart & Sternberg 2017b).

Modelling of the H I–H_2 transition in three-dimensional (3D), magnetohydrodynamical (MHD) simulations remains a highly challenging task. This is due to that fact that MCs are not necessarily in chemical equilibrium concerning their hydrogen content due to turbulent mixing of H_2 into the lower density environment (Glover et al. 2010; Valdivia et al. 2016; Seifried et al. 2017). Therefore, any simulation trying to study the physics and chemistry of the H I–H_2 transition in a self-consistent manner requires the inclusion of an on-the-fly chemical network. Moreover, in order to achieve a converged H_2 (and thus H I) content in the simulations, a high spatial resolution of ∼0.1 pc is required (Seifried et al. 2017), which was confirmed subsequently by means of semi-analytical considerations (Joshi et al. 2019). Despite these difficulties, there have been a great number of studies attempting to model the chemical composition of the (dense) interstellar medium (ISM; e.g. Gnedin, Tassis & Kravtsov 2009;

★ E-mail: seifried@ph1.uni-koeln.de

Glover et al. 2010; Mac Low & Glover 2012; Valdivia et al. 2016; Bialy et al. 2017b; Clark et al. 2019; Joshi et al. 2019; Nickerson, Teyssier & Rosdahl 2019; Bellomi et al. 2020; Smith et al. 2020, and many more). However, not all of these studies fulfil the requirements of non-equilibrium chemistry and a sufficient spatial resolution. In agreement with semi-analytical results, the H I–H_2 transition is found to occur around $\sim 10^{20}$–10^{21} cm^{-2} (Gnedin et al. 2009; Valdivia et al. 2016; Seifried et al. 2017; Bellomi et al. 2020).

From the observational perspective, the H I content in the ISM is often determined via observations of the H I 21 cm emission line, ultraviolet (UV) absorption measurements, and far-infrared studies. A large number of observations on both Galactic (e.g. Savage et al. 1977; Kalberla et al. 2005; Gillmon et al. 2006; Rachford et al. 2009; Barriault et al. 2010; Lee et al. 2012, 2015; Stanimirović et al. 2014; Burkhart et al. 2015; Imara & Burkhart 2016) and extragalactic scales (e.g. Wong & Blitz 2002; Browning, Tumlinson & Shull 2003; Blitz & Rosolowsky 2004, 2006; Bigiel et al. 2008; Wong et al. 2009; Schruba et al. 2011) find that the H I–H_2 transition occurs around $\sim 10^{20}$–10^{21} cm^{-2}. Furthermore, besides the value for the *transition*, some of these observations (e.g. Wong & Blitz 2002; Bigiel et al. 2008; Barriault et al. 2010; Schruba et al. 2011; Lee et al. 2012; Stanimirović et al. 2014; Burkhart et al. 2015) also suggest an upper *threshold* of $N_{\rm H\,I}$ around 10^{21} cm^{-2}, similar to the aforementioned semi-analytical models.

On smaller scales of individual MCs, the measurement of their cold H I content via the H I 21 cm line is challenging due to the simultaneous emission of H I in the warm neutral medium. This problem can be overcome by the study of H I self-absorption (HISA) first reported by Heeschen (1954, 1955). These HISA features arise when cold H I in the foreground absorbs the emission of warmer H I in the background (e.g. Knapp 1974). Over the last few decades, there have been numerous HISA observations studying the properties of H I gas in MCs (e.g. Riegel & Crutcher 1972; Knapp 1974; van der Werf, Goss & Vanden Bout 1988; Goodman & Heiles 1994; Montgomery, Bates & Davies 1995; Gibson et al. 2000, 2005; Kavars et al. 2003, 2005; Li & Goldsmith 2003; Goldsmith & Li 2005; Klaassen et al. 2005; Krčo et al. 2008; Krčo & Goldsmith 2010; Dénes et al. 2018; Beuther et al. 2020; Syed et al. 2020; Wang et al. 2020b). Similar to the observations on galactic scales and semi-analytical models, a large number of these HISA observations towards MCs seem to confirm an upper column density threshold of a few 10^{21} cm^{-2} for cold H I. However, there have been observations that report partly significantly higher H I column densities up to $\sim 10^{22}$ cm^{-2} (Motte et al. 2014; Bihr et al. 2015; Dénes et al. 2018). Also, some indirect measurements suggest higher H I column densities (Fukui et al. 2014, 2015; Okamoto et al. 2017). These observations thus challenge the picture of a saturation of $N_{\rm H\,I}$ around 10^{21} cm^{-2} obtained from other observational works and semi-analytical methods.

Some of the first synthetic H I observations are presented by Douglas et al. (2010) and Acreman et al. (2010, 2012), modelling the H I emission on galactic scales with resolutions around a few 1 pc, i.e. stemming from the more diffuse, atomic ISM. On similar scales, Kim, Ostriker & Kim (2014), and in associated follow-up studies Murray et al. (2015, 2017), conclude that H I absorption observations can trace parameters like the $N_{\rm H\,I}$ and the spin temperature with an accuracy of a few 10 per cent. However, the authors do not include a chemical network for H I–H_2 and only apply a simplified radiative transfer method without observational effects such as noise or beam smearing. On smaller scales of individual MCs, both Hennebelle, Audit & Miville-Deschênes (2007) and Heiner, Vázquez-Semadeni & Ballesteros-Paredes (2015) present H I spectra from simulations without any chemical network and only a simplified radiative transfer method. However, on these scales H I and H_2 coexist and are out of equilibrium due to turbulent mixing (Glover & Mac Low 2007b; Valdivia et al. 2016; Seifried et al. 2017) and thus the predictive power of these studies is limited. Fukui et al. (2018) present synthetic H I spectra of colliding flow simulations including a chemical network for H I and H_2. The authors find that for a significant portion of the pixels, $N_{\rm H\,I}$ can be underestimated in a non-negligible manner by several 10 per cent and more, in particular in pixels with high optical depths. They do, however, not report any HISA features, nor include observational effects. Furthermore, as the authors consider an MC at an extremely early evolutionary stage, their findings rather apply to the cold and warm neutral medium. Finally, Soler (2019) applies the method of oriented gradients to synthetic HISA observations of MCs modelled by Clark et al. (2019).

Following these works, here we present fully self-consistent synthetic HISA observations of MCs, i.e. including 3D, MHD simulations with a chemical network, high spatial resolution ($\sim$0.1 pc), self-consistent radiative calculations, and observational effects. We investigate the accuracy of HISA observations towards MCs (Section 3) and discuss implications of our findings for the postulated saturation of $N_{\rm H\,I}$ around 10^{21} cm^{-2} (Section 4). In particular, we show that this saturation of $N_{\rm H\,I}$ could be a purely observational effect and/or a consequence of oversimplifying assumptions in semi-analytical models (Section 5). We conclude our work in Section 6.

2 NUMERICS

2.1 Simulations

We here briefly describe the numerics behind the simulations; for more details, we refer to Seifried et al. (2017, 2019). All simulations are performed with the FLASH code (Fryxell et al. 2000; Dubey et al. 2008). We include a chemical network for H_2, H, H^+, C^+, C, O, CH, OH, CO, HCO^+, He, He^+, M, and M^+ (Nelson & Langer 1999; Glover & Mac Low 2007a, b; Glover & Clark 2012) with updates in the network described in Mackey et al. (2019). Here, M and M^+ represent the contribution of metals, where we specifically consider Si and its first ionized state. We also include the most relevant heating and cooling mechanisms. In addition, we calculate the attenuation of the interstellar radiation field ($G_0 = 1.7$ in Habing units; Habing 1968; Draine 1978) using the TREERAY/OPTICALDEPTH module (Clark, Glover & Klessen 2012a; Walch et al. 2015; Wünsch et al. 2018). The cosmic ray ionization rate of atomic hydrogen is set to 3×10^{-17} s^{-1}. The Poisson equation for self-gravity is solved using a tree-based method (Wünsch et al. 2018).

The simulations are part of the SILCC-Zoom project (Seifried et al. 2017, 2020a), where we model the formation and evolution of MCs located in a part of a stratified galactic disc, which in turn is part of the SILCC project (Walch et al. 2015; Girichidis et al. 2016). The disc has an initial Gaussian density profile given by

$$\rho(z) = \rho_0 \times \exp\left[-\frac{1}{2}\left(\frac{z}{h_z}\right)^2\right], \quad (1)$$

with $h_z = 30$ pc and $\rho_0 = 9 \times 10^{-24}$ g cm^{-3}, resulting in a total gas surface density ($\Sigma_{\rm gas}$) of 10 $M_\odot$ pc^{-2}. We run two simulations, one without a magnetic field and one with. For the magnetized run, we initialize the magnetic field as

$$B_x = B_{x,0}\sqrt{\rho(z)/\rho_0}\ ,\ B_y = 0\ ,\ B_z = 0, \quad (2)$$

with $B_{x,0} = 3$ μG in accordance with observations (Beck & Wielebinski 2013). We emphasize that the magnetic field is dynamically

important for the (chemical) evolution of the MCs (Seifried et al. 2020a, b). In addition to the gas self-gravity, we include a background potential from the old stellar component modelled as an isothermal sheet with a scale height of 100 pc and $\Sigma_{\rm star} = 30\ {\rm M}_\odot\ {\rm pc}^{-2}$.

In the initial simulation phase, up to a time t_0, the spatial resolution is 4 pc and we drive turbulence by injecting supernovae (SNe) with a rate of 15 SNe Myr^{-1} (see Walch et al. 2015; Girichidis et al. 2016; Gatto et al. 2017, for details). At t_0, we stop the SN injection to allow for the formation of MCs unaffected by nearby SN remnants, which could influence their evolution (Seifried et al. 2018). For both the unmagnetized and magnetized runs, we pick two regions, each with a typical size of $\sim$(100 pc)^3, in which MCs are about to form. We thus have in total four MCs, henceforth denoted as MC1-HD, MC2-HD, MC1-MHD, and MC2-MHD, where the first two are non-magnetized runs and the latter two include a dynamically relevant magnetic field.[1] The typical H_2 masses of these clouds are around (20–50) $\times\ 10^3\ {\rm M}_\odot$, and for the magnetized runs the volume-weighted magnetic field is around 4 μG. Starting at t_0, we then progressively increase the spatial resolution in these zoom-in regions over 1.65 Myr, reaching a maximum resolution of 0.06 pc. Afterwards, we evolve the clouds on this resolution for a few more Myr. We note that throughout the paper all times refer to the time elapsed since t_0, i.e. the start of the zoom-in procedure. We have $t_0 = 11.9$ and 16.0 Myr for the runs without and with magnetic fields, respectively.

2.2 Radiative transfer

The radiative transfer simulations are performed in a post-processing step with RADMC-3D (Dullemond et al. 2012) for the H I 21 cm emission line of atomic hydrogen. In order to calculate the (two-)level population, we apply the method to calculate the spin temperature of atomic hydrogen, $T_{\rm s}$, described in Kim et al. (2014, their equations 4–7). In particular, we include the Wouthuysen–Field (WF) effect (Wouthuysen 1952; Field 1958) assuming $T_\alpha = T_{\rm gas}$ for the effective temperature of the Ly α field (Field 1959) and $n_\alpha = 10^6\ {\rm cm}^{-3}$ for the Ly α photon density (Liszt 2001). We emphasize that including or excluding the WF effect has only a marginal impact as for the temperatures of $\lesssim 10^2$ K typical for MCs, the WF effect does barely affect $T_{\rm s}$, which remains close to the actual gas temperature in both cases (fig. 2 of Kim et al. 2014). Furthermore, the Einstein coefficient of the H I line is $A_{\rm ul} = 2.8843 \times 10^{-15}\ {\rm s}^{-1}$ (Gould 1994) and the spectral resolution is set to 200 m s^{-1} over a range of ±20 km s^{-1}, resulting in 201 velocity channels.

We consider the emission of the four MCs in isolation, i.e. the emission stemming from the gas in the aforementioned zoom-in regions only. This allows us to focus on the HISA signal originating from the MCs themselves (and to a smaller extent from warm H I in the zoom-in region), avoiding any foreground contamination. We use a resolution of 0.06 pc, i.e. identical to the maximum resolution of the underlying simulations. We investigate the emission for two points in time, i.e. $t_{\rm evol} = 2$ and 3 Myr. As the results for both times are qualitatively similar, for most of the plots we focus on $t_{\rm evol} = 2$ Myr. In order to model the emission (and its absorption) of a diffuse H I background, we include a (spatially and spectrally) fixed background radiation field with a brightness temperature of 100 K in the radiative transfer calculation. This background temperature is motivated by results from recent HISA observations within the galactic plane (Syed et al. 2020; Wang et al. 2020b) and is also used in the numerical study presented in Soler et al. (2019). This makes our HISA observations sensitive to the absorption of H I gas with $T_{\rm s}$ below 100 K; warmer gas will be seen in emission. We note that we have also tested the usage of a background temperature of 200 K. The observed changes are, however, only very moderate, which is why we focus on the case of 100 K here.

2.2.1 *Adding observational effects*

In a final step, we incorporate observational effects into our obtained (ideal) H I emission maps. For this purpose, we (i) convolve our emission maps with a Gaussian beam of 80 arcsec (at the chosen distance; see below); (ii) reduce the spectral resolution from 200 m s^{-1} to 1 km s^{-1} by summing up the contribution of 5 neighbouring channels; i.e. the data cubes analysed in the following have only 40 velocity channels (one-fifth of the original cubes); and (iii) finally add random Gaussian noise with a standard deviation of 3 K to the obtained emission maps.[2] All the above stated values are average values from recent HISA observations towards MCs (e.g. Dénes et al. 2018; Beuther et al. 2020; Syed et al. 2020; Wang et al. 2020b). We choose two different distances for the observed clouds of 150 pc and 3 kpc, corresponding to a physical beam size of 0.06 and 1.2 pc, respectively. However, as the results for both distances are relatively similar, in the following we focus on the distance of 150 pc.

2.3 HISA calculations

In order to investigate the properties of the absorbing atomic hydrogen gas, i.e. the HISA features, we follow the approach described in Wang et al. (2020b; see their section 2.2) to convert the absorption spectrum into an effective emission spectrum of the cold H I. In short, assuming a temperature of the absorbing, cold H I gas, $T_{\rm HISA}$, and that both foreground and background emission are optically thin, one can relate the observed emission to the optical depth, $\tau_{\rm HISA}$, of the absorption layer via

$$T_{\rm off-on} = T_{\rm off} - T_{\rm on} = (pT_{\rm off} + T_{\rm cont} - T_{\rm HISA}) \times (1 - {\rm e}^{-\tau_{\rm HISA}}). \quad (3)$$

For the sake of readability, we have omitted the dependence on the velocity channel in the above equation. Here, $T_{\rm on}$ is the actually observed brightness temperature and $T_{\rm off}$ is the brightness temperature at an off-position, i.e. the brightness temperature that would be measured if no absorbing cold H I gas were present along the line of sight (LOS) towards the observer. Furthermore, $T_{\rm cont}$ is the brightness temperature of the diffuse continuum background, and the dimensionless quantity p parametrizes the ratio of foreground to background emission (Feldt 1993; Gibson et al. 2000) and is usually estimated to be close to 1 (McClure-Griffiths et al. 2006; Rebolledo et al. 2017; Dénes et al. 2018; Wang et al. 2020b).

In Wang et al. (2020b), $T_{\rm off}$ – which is not accessible via observations – is inferred by fitting a polynomial to the absorption-free channels, i.e. those channels where no HISA feature is present. We use a fixed background brightness temperature of 100 K in our synthetic observations, which varies neither spatially nor spectrally. Therefore, we cannot test additional inaccuracies arising from the

[1] See also Seifried et al. (2020a), where we discuss their chemical properties concerning H_2 and CO. In that publication, the MHD clouds were named MC3-MHD and MC4-MHD. Also note that, due to the nature of the SILCC simulations, the runs including magnetic fields are completely independent from those without magnetic fields, and so are the resulting clouds.

[2] Note that it is important to apply step (iii) as the final step.

uncertainty to determine $T_{\rm off}$ in actual observations – a caveat to keep in mind throughout the paper. However, due to this simplifying assumption, in our case the spectrum in the absorption-free channels is practically flat as the emission of H I gas warmer than 100 K in the considered zoom-in region is almost negligible, similar to the findings of Soler et al. (2019). For this reason, we can set $T_{\rm off} = 100$ K for our data, which yields

$$T_{\rm off-on}(v) = 100\,{\rm K} - T_{\rm on}(v). \quad (4)$$

The quantity $T_{\rm off-on}(v)$ is positive and gives the depth of the absorption feature seen in $T_{\rm on}(v)$. In the following, we refer to these kinds of spectra as HISA spectra.

Furthermore, using a constant background temperature and assuming that the emission of H I warmer than 100 K is negligible are congruent with setting $p = 1$ and $pT_{\rm off} + T_{\rm cont} = 100$ K.[3] Hence, we can simplify equation (3) to yield

$$T_{\rm off-on}(v) = (100\,{\rm K} - T_{\rm HISA}) \times (1 - {\rm e}^{-\tau_{\rm HISA}(v)}). \quad (5)$$

We apply the spectral analysis tool BTS[4] (Clarke et al. 2018) to $T_{\rm off-on}(v)$ to find the location and properties of the HISA feature. BTS identifies and fits Gaussian peaks in a spectrum. Assuming that the HISA feature has approximately the shape of a single Gaussian, we restrict the fitting function used in BTS to a single Gaussian. We set the noise level to 3 K and the required signal-to-noise ratio to 3. Of the obtained fitting values, here only the line width, $\sigma_{\rm BTS}$, will be used later in the paper. Next, for each pixel for which BTS identifies a (Gaussian) HISA feature, we calculate the column density of the cold H I responsible for the HISA feature.

For this purpose, for most of our paper we adopt a fixed value for $T_{\rm HISA}$ for the entire map, identical to the approach applied in recent observations (e.g. Syed et al. 2020; Wang et al. 2020b, but see below and Section 3.5 for a different approach). We then solve equation (5) for the optical depth of the HISA feature, $\tau_{\rm HISA}$, for each channel independently, which yields

$$\tau_{\rm HISA}(v) = -\ln\left(1 - \frac{T_{\rm off-on}}{100\,{\rm K} - T_{\rm HISA}}\right). \quad (6)$$

With this, we can calculate the column density of the HISA feature via (Wilson, Rohlfs & Hüttemeister 2013)

$$N_{\rm H\,I,obs} = 1.8224 \times 10^{18}{\rm cm}^{-2}\,\frac{T_{\rm s}}{1\,{\rm K}}\int \tau(v)\frac{{\rm d}v}{1\,{\rm km\,s}^{-1}}, \quad (7)$$

where we assume $T_{\rm s} = T_{\rm HISA}$ for the H I spin temperature and $\tau(v) = \tau_{\rm HISA}(v)$ for the optical depth.

Though being straightforward to use, this method has the disadvantage that a fixed value of $T_{\rm HISA}$ for every pixel of the map has to be assumed, which might not be the case. Furthermore, when choosing a fixed $T_{\rm HISA}$, equation (3) might not yield a result for $\tau_{\rm HISA}$ for every channel. The maximum usable value for each channel, $T_{\rm HISA,max}(v)$, is obtained by assuming $\tau_{\rm HISA} \rightarrow \infty$ (Wang et al. 2020b), which results for our set-up in

$$T_{\rm HISA,max}(v) = -T_{\rm off-on}(v) + pT_{\rm off}(v) + T_{\rm cont} = T_{\rm on}(v). \quad (8)$$

Hence, if the assumed $T_{\rm HISA}$ exceeds the observed brightness temperature $T_{\rm on}(v)$ of a given channel, this channel has to be dropped and cannot be taken into account for the calculation of the column density (see Section 3.2).

[3]Effectively, we now do not differentiate any more between $pT_{\rm off}$ and $T_{\rm cont}$.
[4]Acronym for 'Behind The Spectrum', https://github.com/SeamusClarke/BTS.

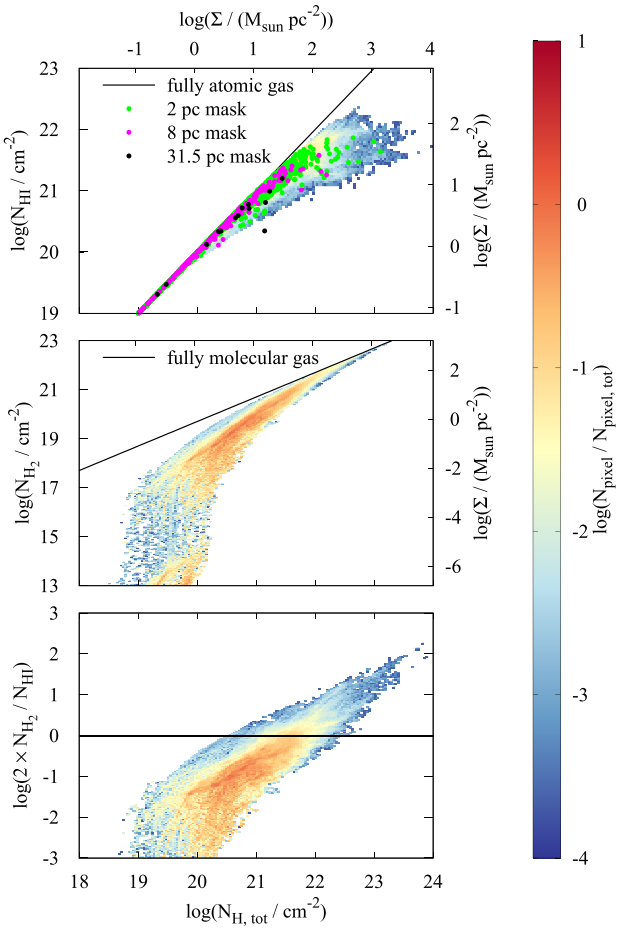

Figure 1. 2D PDF of $N_{\rm H\,I}$ (top) and $N_{\rm H_2}$ (middle) and the ratio of both column densities versus $N_{\rm H,tot}$ (bottom) for MC1-HD at 3 Myr for one LOS. The H I–H_2 transition occurs around 10^{21} cm^{-2}, in rough agreement with observational results. However, the H I column density levels off only around a few 10^{22} cm^{-2}, thus higher than that typically obtained in observations. Coloured dots denote the average column density when using pixels with a side length of 2, 8, and 31.5 pc (see Section 5). Note the different y-axis scaling for the top and middle panels.

An alternative way to determine the optical depth, which simultaneously leaves $T_{\rm HISA}$ as a free parameter, is given by Knapp (1974) expressing $\tau_{\rm HISA}(v)$ via

$$\tau_{\rm HISA}(v) = \tau_0 e^{-\frac{1}{2}\left(\frac{v-v_0}{\sigma_0}\right)^2}. \quad (9)$$

Here, τ_0, v_0, and σ_0 are free parameters (together with $T_{\rm HISA}$) that are determined by fitting the observed spectrum $T_{\rm off-on}(v)$ with equation (9) inserted into equation (5). As before, the fit is only performed for those pixels where we identify a Gaussian HISA feature with BTS. Besides the approach using a fixed $T_{\rm HISA}$, we will also test this approach in the following.

3 RESULTS

In order to get a first impression about the H I and H_2 content of the simulated MCs, in Fig. 1 we show the 2D PDFs of $N_{\rm H_2}$, $N_{\rm H\,I}$, and their ratio versus $N_{\rm H,tot}$ for a selected simulation snapshot. Here, $N_{\rm H,tot} = N_{\rm H\,I} + 2\,N_{\rm H_2} + N_{\rm H^+}$ denotes the total hydrogen column density. We find that H_2 starts to form above $N_{\rm H,tot} \simeq 10^{20}$ cm^{-2}. The transition from atomic- to molecular-hydrogen-dominated gas, however, occurs

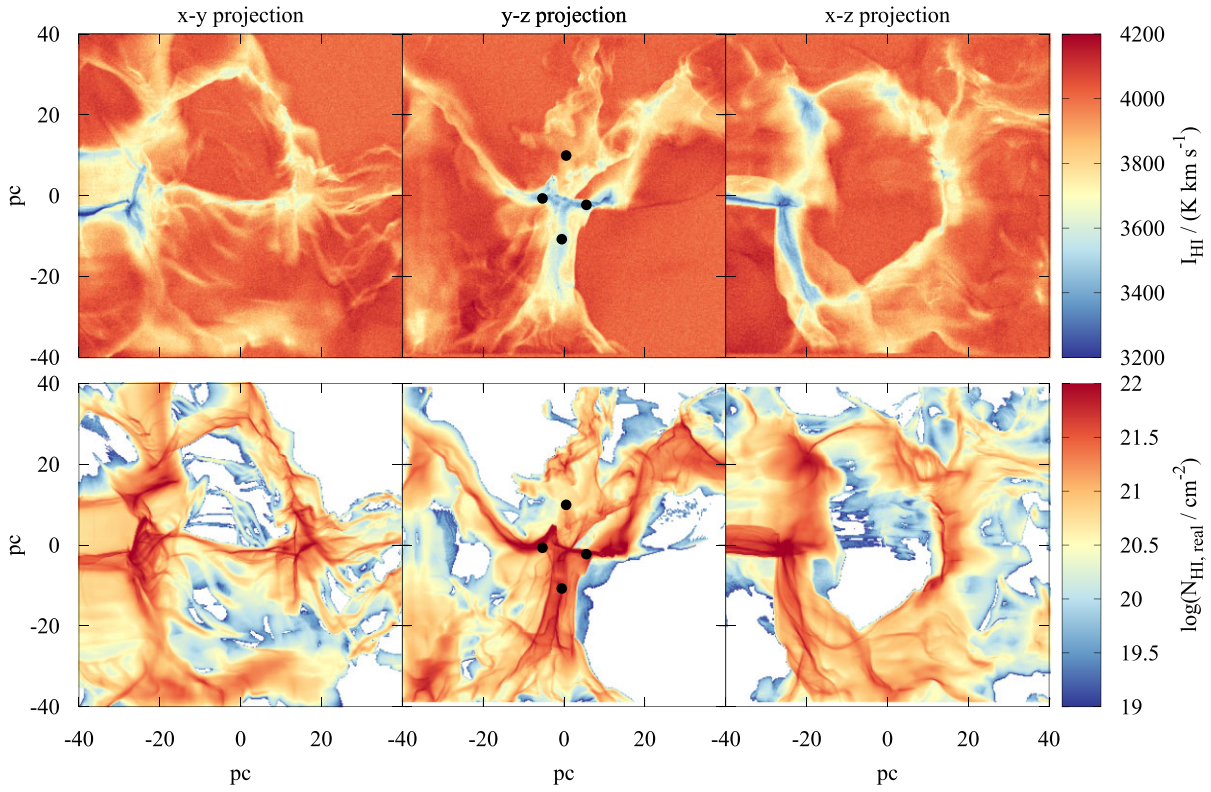

Figure 2. Top row: Integrated H I intensity for three different directions of MC1-HD at $t_{evol} = 2$ Myr, including observational effects for an assumed distance of 150 pc (beam size corresponds to 0.06 pc). Bottom row: Column density of H I gas with temperatures below 100 K calculated directly from the simulation data. The HISA feature traces well the high-column density regions. Note that the high integrated intensities around 4000 K km s^{-1} are due to the assumed constant background brightness temperature of 100 K integrated over a velocity range of ±20 km s^{-1}. The black dots in the middle column show the positions of the two LOSs for which the spectra are plotted in Fig. 3.

around $N_{H,tot} \simeq 10^{21}$ cm^{-2} ($\simeq 8$ M$_\odot$ pc^{-2}), as discussed in detail in Seifried et al. (2020a). This is in agreement with other numerical and semi-analytical works (Krumholz et al. 2008, 2009; Gnedin et al. 2009; Sternberg et al. 2014; Bialy & Sternberg 2016; Valdivia et al. 2016; Bellomi et al. 2020).

However, despite H_2 forming rapidly above $N_{H,tot} \simeq 10^{20}$ cm^{-2}, also N_{HI} continues to rise. Most of the H I has column densities around 10^{21} cm^{-2} ($\simeq 8$ M$_\odot$ pc^{-2}) similar to theoretical predictions (Krumholz et al. 2008, 2009; Sternberg et al. 2014; Bialy & Sternberg 2016). However, even significantly higher H I column densities up to a few 10^{22} cm^{-2} are reached; i.e. H_2 and H I coexist on the projected maps. Hence, having the H I–H_2 transition around a certain value (e.g. 10^{21} cm^{-2}) does not exclude the occurrence of significantly higher H I column densities.

3.1 Deriving N_{HI} using a fixed T_{HISA}

In the following, we will assess how accurately the *cold* H I content in MCs can be determined via HISA observations.[5] We define cold H I as all H I with temperatures below 100 K. This definition is motivated by the chosen background brightness temperature (Section 2.3), which makes our synthetic HISA observations sensitive to H I with temperatures below 100 K. Depending on the simulation, we find that around and in our MCs 22–43 per cent of the H I is warmer than 100 K; i.e. in all cases, the HISA observations are sensitive to more than 50 per cent of the entire H I mass. In the top row of Fig. 2, we show the velocity-integrated intensity, I_{HI}, of our synthetic H I observations of MC1-HD from three different directions at $t_{evol} = 2$ Myr, including observational effects at an assumed distance of 150 pc (see Section 2.2.1). For $t_{evol} = 3$ Myr, the results are qualitatively similar. In the bottom row, we show the cold H I column density of MC1-HD inferred directly from the simulation, $N_{HI,real}$. We emphasize that throughout this paper $N_{HI,real}$ refers to the cold H I gas ($T \leq 100$ K) as described above.

Both the synthetic H I emission and the $N_{HI,real}$ maps show complex filamentary structures with extended envelopes. There is a clear anticorrelation between I_{HI} and $N_{HI,real}$. In order to demonstrate that the drop in I_{HI} is due to the self-absorption of radiation, we show in Fig. 3 two example spectra from pixels in the high-$N_{HI,real}$/low-I_{HI} areas in the middle column of Fig. 2 (black dots). For both pixels, there are clear HISA features recognizable caused by the cold H I. For the other MCs and an assumed distance of 3 kpc, the obtained results are qualitatively and quantitatively very similar. We note that the rather high integrated intensities around 4000 K km s^{-1} are due to the assumed constant background brightness temperature of 100 K integrated over a velocity range of ±20 km s^{-1}, which is, however, automatically taken into account via equation (5).

Next, we investigate the H I column densities obtained from the HISA observations, denoted as $N_{HI,obs}$ (equation 7), assuming a fixed

[5]In Seifried et al. (2020a), we have already discussed a new approach to determine the H_2 content by means of combined CO(1–0) and dust emission observations, which allows for an accurate determination of H_2 within a factor of 1.8.

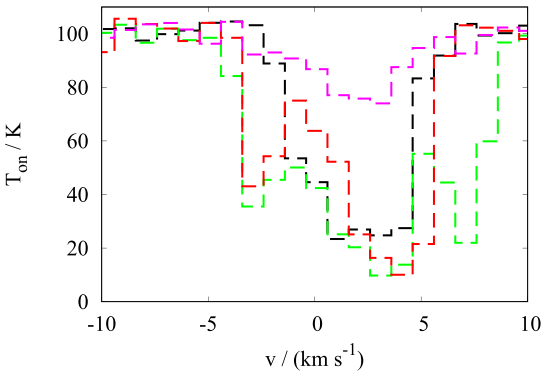

Figure 3. Synthetic H I spectra including observational effects for the positions indicated in the map in the middle column of Fig. 2 (black dots) showing the variety of spectral shapes obtained in our synthetic observations. The black line shows a spectrum potentially suffering from opacity broadening at the line centre. In the spectra shown by the red and green lines, multiple absorption features seem to be present.

HISA temperature, $T_{\rm HISA}$. In the top row of Fig. 4, we show $N_{\rm H\,I,obs}$ for $T_{\rm HISA}$ = 20, 40, and 60 K for MC1-HD at $t_{\rm evol}$ = 2 Myr for one LOS at an assumed distance of 150 pc. In the bottom row, we show the ratio of $N_{\rm H\,I,obs}$ and $N_{\rm H\,I,real}$. Most prominently, we find that for all three values of $T_{\rm HISA}$, the observed column densities $N_{\rm H\,I,obs}$ in the central and most dense regions are a factor of $\gtrsim$10 too low. This trend becomes more pronounced with increasing $T_{\rm HISA}$. In the outer regions, $N_{\rm H\,I,obs}$ could not be calculated for all pixels as here the absorption features are partly too weak and therefore the spectral analysis tool BTS does not identify any HISA feature. For the remaining pixels in the outer parts, $N_{\rm H\,I,real}$ is also typically underestimated ranging from a few 10 per cent up to a factor of a few. However, increasing $T_{\rm HISA}$ pushes $N_{\rm H\,I,obs}$ in the outer regions closer to $N_{\rm H\,I,real}$. Overall, however, the match between the actual and observed H I column density is rather poor with a clear tendency to underestimate $N_{\rm H\,I,real}$ by factors up to ∼10 (see also Section 3.4 for the effect on total estimated H I mass). This does not change when considering different LOSs, times, and MCs or assuming a distance of 3 kpc. Reasons for this underestimation will be discussed in detail in Sections 3.2 and 3.3.

In Fig. 5, we show the mean value of $N_{\rm H\,I,obs}$ as a function of $N_{\rm H\,I,real}$ for the four different MCs placed at 150 pc at $t_{\rm evol}$ = 2 Myr focusing on the case of $T_{\rm HISA}$ = 20, 40, and 60 K (coloured lines). Overall, the underestimation shown in Fig. 4 is observed in all cases; the degree of underestimation becomes more pronounced with increasing $N_{\rm H\,I,real}$ and decreasing $T_{\rm HISA}$. Lowering or increasing $T_{\rm HISA}$ beyond the values shown only amplifies the trends, which is why we do not explicitly discuss them here. In the following, we focus on the range of $N_{\rm H\,I,real} \geq 10^{19.5}$ cm^{-2}.

First, we note that $T_{\rm HISA}$ = 20 K (blue lines) is apparently a rather bad choice resulting in an underestimation by about one order of magnitude (and more). Considering $T_{\rm HISA}$ = 60 K (red lines), we find that in the range of $10^{19.5}$ cm^{-2} $\lesssim N_{\rm H\,I,real} \lesssim 10^{21}$ cm^{-2}, the actual values are underestimated by a factor of ∼2–5. In the same range, for $T_{\rm HISA}$ = 40 K (green lines), the actual values are underestimated even more severely by a factor of ∼3–10. Moreover, for $N_{\rm H\,I,real} > 10^{21}$ cm^{-2}, $N_{\rm H\,I,obs}$ seems to level off around $\sim 10^{20.5\text{–}21}$ cm^{-2}, for both $T_{\rm HISA}$ = 40 and 60 K. This results in an increasing underestimation of the actual column density when going to denser and denser regions. We note that this artificial levelling-off around $N_{\rm H\,I,obs} \simeq 10^{21}$ cm^{-2} matches well the maximum values reported in recent HISA observations (Kavars et al. 2003, 2005; Li & Goldsmith 2003; Goldsmith & Li 2005; Klaassen et al. 2005; Krčo et al. 2008; Barriault et al. 2010; Krčo & Goldsmith 2010; Syed et al. 2020; Wang et al. 2020b, but see also Section 4.1 for a further discussion). Finally, even for a single MC there is a large scatter of the measured $N_{\rm H\,I,obs}$ for a given $N_{\rm H\,I,real}$ (shown by the full distribution in greyscale in the background for MC1-HD for $T_{\rm HISA}$ = 40 K). This further lowers the accuracy with which HISA observations seem to be able to constrain the actual H I column density. An additional uncertainty in observations arises from the unknown value of $T_{\rm off}$ (and thus $T_{\rm off\text{–}on}$), which in our case is chosen to be constant ($T_{\rm off}$ = 100 K). Equation (6) for the optical depth implies that this uncertainty increases even further the scatter found for individual MCs at a given $N_{\rm H\,I,real}$.

3.2 The H I temperature

Investigating equation (3) shows that there is a certain upper threshold for $T_{\rm HISA}$, above which the equation is not solvable for $\tau_{\rm HISA}$ for at least some of the velocity channels, which then would have to be omitted for the calculation of $N_{\rm H\,I}$. We denote this upper threshold as $T_{\rm dip}$, which is set by the minimum of $T_{\rm HISA,max}(v)$ (equation 8) over all velocity channels for a given pixel, i.e.

$$T_{\rm dip} = \min(T_{\rm HISA,max}(v)) = \min(T_{\rm on}(v)). \quad (10)$$

The denomination as $T_{\rm dip}$ is motivated by the fact that it corresponds to the temperature at the dip of the observed absorption spectrum $T_{\rm on}(v)$. For the sake of clarity, in Table 1 we give a short summary of the most relevant temperature definitions used in this paper.

In the left-hand panel of Fig. 6, we show the map of $T_{\rm dip}$ for one snapshot. In addition, the right-hand panel shows the actual mass-weighted, LOS-averaged H I temperature, $T_{\rm H\,I,mw}$.[6] Both $T_{\rm dip}$ and $T_{\rm H\,I,mw}$ show a strong drop towards the central, high-column density areas (see also top panel of Fig. 8). Interestingly, we find that both temperature measures show an agreement within about 20 K (Fig. 7). Based on this, in Section 3.5 we investigate whether the usage of $T_{\rm dip}$ as an approximation for the temperature of the cold H I along the LOS (and thus for $T_{\rm HISA}$) is suitable.

The strong variations of $T_{\rm dip}$ and $T_{\rm H\,I,mw}$ down to values as low as ∼10 K cause the significant underestimation of $N_{\rm H\,I,real}$ seen in Figs 4 and 5 by means of two effects explained in the following and sketched in Fig. 8:

(i) $T_{\rm HISA}$ chosen too high ($>T_{\rm dip}$): These regions typically correspond to high $N_{\rm H\,I,real}$ ($\gtrsim 10^{21}$ cm^{-2}) (top panel of Fig. 8). Here, equation (3) does not yield any results for at least some of the velocity channels. This happens for ≲10 per cent, 5–20 per cent, and 30–40 per cent of the pixels for $T_{\rm HISA}$ = 20, 40, and 60 K, respectively. Hence, for these pixels *some* of the channels have to be neglected, which reduces $N_{\rm H\,I,obs}$ significantly (equation 7). As with decreasing $T_{\rm dip}$, i.e. increasing $N_{\rm H\,I,real}$, more and more velocity channels have to be omitted (at a fixed $T_{\rm HISA}$); this leads to the observed artificial levelling-off of $N_{\rm H\,I,obs}$ at $\sim 10^{21}$ cm^{-2}. As a consequence, for central, high-column density regions of MCs, $N_{\rm H\,I,real}$ is underestimated by a factor of about 10 and more (Fig. 5).

(ii) $T_{\rm HISA}$ chosen too low ($<T_{\rm s}$): This leads to an underestimation of the optical depth $\tau_{\rm HISA}$ (equation 6) as well as $N_{\rm H\,I,obs}$ (equation 7,

[6] As the synthetic H I observations are only sensitive to H I with temperatures below 100 K, also for this average (as for $N_{\rm H\,I,real}$) only H I gas with $T \leq 100$ K is considered.

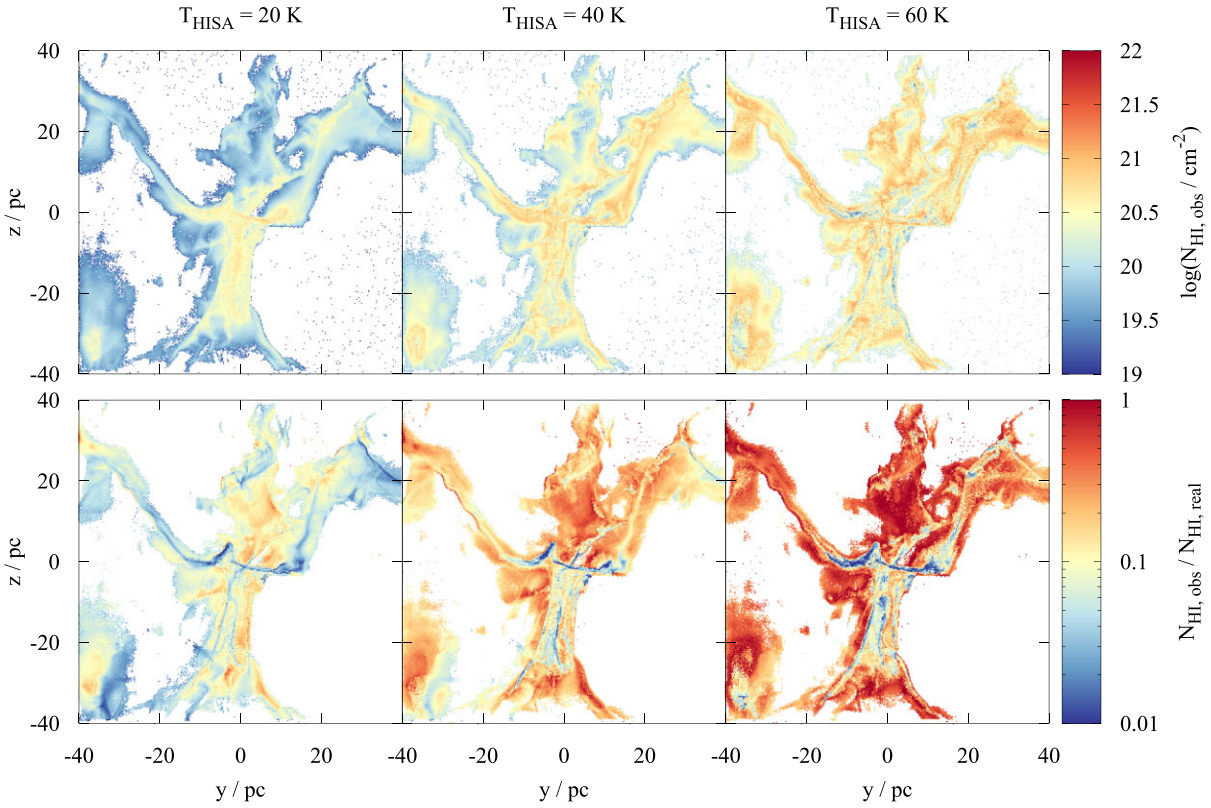

Figure 4. Top row: Observed H I column density assuming a fixed $T_{\rm HISA}$ of 20, 40, and 60 K (from left to right) for MC1-HD at $t_{\rm evol} = 2$ Myr at an assumed distance of 150 pc (beam size corresponds to 0.06 pc). Bottom row: Ratio of the observed H I column density shown in the top row to the actual H I column density (for $T < 100$ K). Overall, assuming a fixed $T_{\rm HISA}$ underestimates $N_{\rm H\,I,real}$ by almost about one order of magnitude in the central, high-column density regions. Increasing $T_{\rm HISA}$ improves the match only in the outer regions.

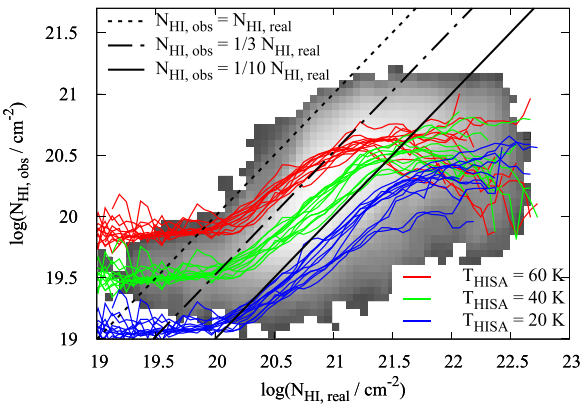

Figure 5. Mean value of $N_{\rm H\,I,obs}$ against $N_{\rm H\,I,real}$ for the three different directions of all four MCs placed at a distance of 150 pc at $t_{\rm evol} = 2$ Myr and using three different $T_{\rm HISA}$ (coloured lines). In the background, the full distribution for one snapshot (MC1-HD, $T_{\rm HISA} = 40$ K) is shown in greyscale. The black lines show lines of constant ratio $N_{\rm H\,I,obs}/N_{\rm H\,I,real}$ to guide the readers' eye. In general, the actual column density is underestimated significantly, and the maximum $N_{\rm H\,I,obs}$ levels off around 10^{21} cm^{-2}.

both via $\tau_{\rm HISA}$ and the assumption $T_{\rm s} = T_{\rm HISA}$). This effect is dominant mainly in the low- to intermediate-column density regions in the outer parts of the MCs ($N_{\rm H\,I,real} \lesssim 10^{21}$ cm^{-2}). Here, $N_{\rm H\,I,obs}$ underestimates $N_{\rm H\,I,real}$ on average by a factor of 3–10. Due to the linear dependence of $N_{\rm H\,I,obs}$ on $T_{\rm HISA}$, the actual value of $N_{\rm H\,I,obs}/N_{\rm H\,I,real}$ increases with increasing $T_{\rm HISA}$ at high $T_{\rm dip}$ (coloured lines in the bottom panel of Fig. 8).

Overall, our results demonstrate that finding an accurate value for $T_{\rm HISA}$ is crucial but at the same time *not possible* when using a single value for the entire map. In addition, both choosing a too high and too low value for $T_{\rm HISA}$ lead to an underestimation of the H I column density.

3.2.1 *H I temperature variations*

Interestingly, even for regions where $T_{\rm HISA} \simeq T_{\rm dip}$ (indicated by the peak of the coloured lines in the bottom panel of Fig. 8), a ratio of $N_{\rm H\,I,obs}/N_{\rm H\,I,real}$ close to 1 is barely reached. An additional source of uncertainty causing this is the significant variation of $T_{\rm H\,I,mw}$ across the map (right-hand panel of Fig. 6). Similar variations will also occur for each individual pixel *along* the LOS. Hence, even for an individual pixel the assumption of a constant $T_{\rm HISA}$ presents an oversimplification, which in turn results in the observed underprediction of the H I column density particularly for the dense regions.

We emphasize that $T_{\rm H\,I,mw}$ can be a few 10 K higher than the temperature of the coldest H I gas along each LOS, denoted as $T_{\rm H\,I,min}$ (not shown), due to H I gas warmer than $T_{\rm H\,I,min}$ along the LOS. This also explains why there are regions where $T_{\rm dip} < T_{\rm H\,I,mw}$ (Fig. 7), whereas $T_{\rm H\,I,min}$ is, as expected, always smaller than $T_{\rm dip}$. This also indicates that $T_{\rm dip}$ only gives an upper limit to the actual spin temperature $T_{\rm s}$ of the absorbing H I layer. The non-isothermality of

Table 1. List of the most relevant temperature definitions used in this paper, including a short explanation and the nature of each temperature.

T_s	*Actual* spin temperature of the H I gas used for the radiative transfer (Section 2.2), close to the actual gas temperature of the simulation	Spin temperature, varies along the LOS
T_{HISA}	*Assumed* spin temperature of the HISA features needed for the calculation of the H I optical depth (equation 6) and subsequently the column density (equation 7)	Spin temperature
$T_{H\,I,mw}$	Mass-weighted, LOS-averaged H I temperature *calculated* from the simulation data for a given pixel; see right-hand panel of Fig. 6	Kinetic gas temperature
$T_{H\,I,min}$	Minimum H I temperature *calculated* from the simulation data along the LOS for a given pixel	Kinetic gas temperature
$T_{on}(v)$	Spectrum of the *measured* H I brightness temperature for a given pixel; see Fig. 3	Brightness temperature
T_{dip}	Lowest temperature of the $T_{on}(v)$ spectrum for a given pixel; see equation (10) and left-hand panel of Fig. 6	Brightness temperature

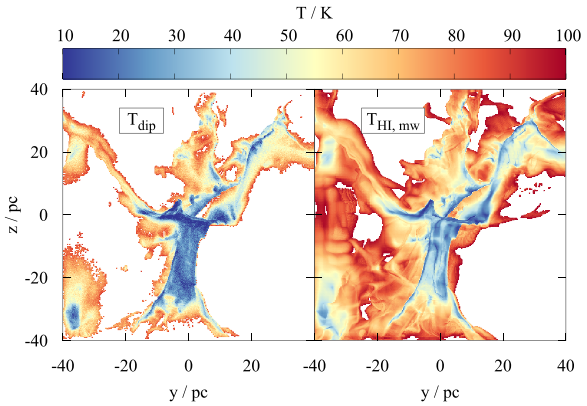

Figure 6. Map of T_{dip} (left, equation 10), which can be used such that equation (3) yields a result for every velocity channel, and the mass-weighted average of the temperature of the H I gas (right) for the same snapshot as shown in Fig. 4. Both quantities show a reasonable agreement within about 20 K. However, T_{dip} is quite low in the central areas. This explains the poor match of the observed with the actual column density (Fig. 4), as here often $T_{HISA} > T_{dip}$ (depending on the actual choice of T_{HISA}).

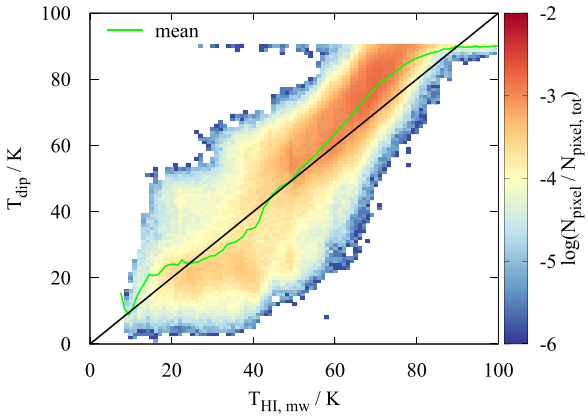

Figure 7. 2D PDF of T_{dip} versus $T_{H\,I,mw}$ and its mean value (green line) for the same snapshot as shown in Fig. 6. The black line corresponds to $T_{dip} = T_{H\,I,mw}$. Overall, there is a rough correspondence between both quantities with a typical scatter of ∼10–20 K.

the cold H I gas is further emphasized by its temperature distribution in Fig. 9 showing the cumulative PDF of H I gas above a certain threshold temperature for all four MCs at $t_{evol} = 2$ Myr. The amount of H I gas is rising steadily with decreasing temperature, independent

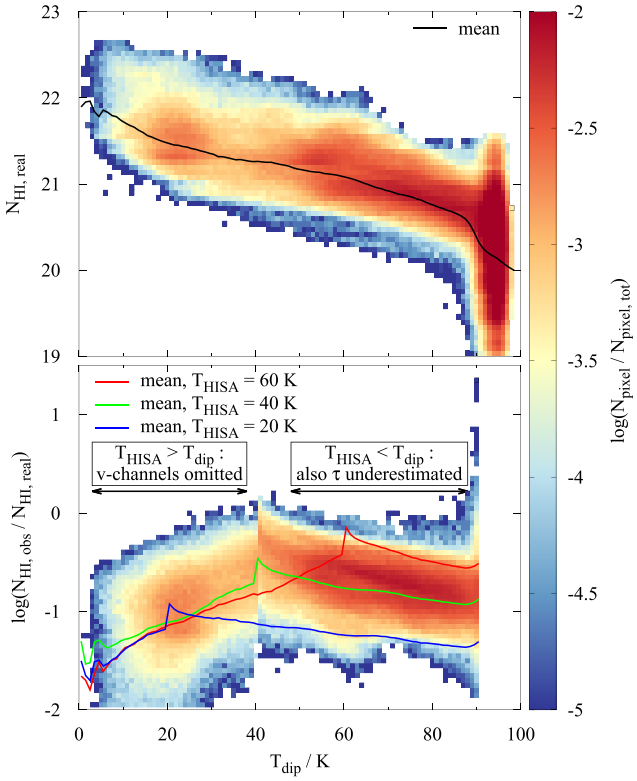

Figure 8. Top: Phase diagram of $N_{H\,I,real}$ versus T_{dip} for the same snapshot as in Fig. 4. The highest H I column densities are associated with low T_{dip}. Bottom: Phase diagram of the ratio of $N_{H\,I,obs}$ and $N_{H\,I,real}$ versus T_{dip} for the same snapshot as in the top panel using $T_{HISA} = 40$ K. The coloured lines show the mean value of the distribution for three different T_{HISA}. Overall, the actual H I column density is underestimated by up to a factor of ∼10. Changing T_{HISA} increases the accuracy only locally, i.e. where that assumed T_{HISA} roughly corresponds to T_{dip}, which represents the real temperature of the HISA feature. We indicated the reasons of the underestimation in the two temperature ranges above and below $T_{dip} = T_{HISA}$.

of the considered MC. This shows that no *single* temperature can be used to describe the H I content of MCs. As an example, for $T_{HISA} = 40$ K, the HISA observations would (at least) miss out 20–40 per cent of the cold H I mass.

The above results explain why also in the absence of observational effects like noise and limited spectral or spatial resolution the poor match between the observed and actual H I column density remains (see Fig. A1 in Appendix A). Also, for different assumed distances

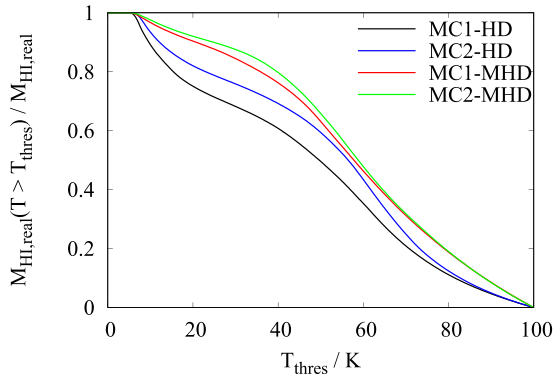

Figure 9. Cumulative temperature PDF showing the amount of cold H I above a certain threshold temperature for all four MCs at $t_{\rm evol} = 2$ Myr. The steady rise with decreasing temperature indicates that no single temperature choice for $T_{\rm HISA}$ is suitable to accurately determine the amount of cold H I in the clouds. Note that gas above 100 K is not considered here.

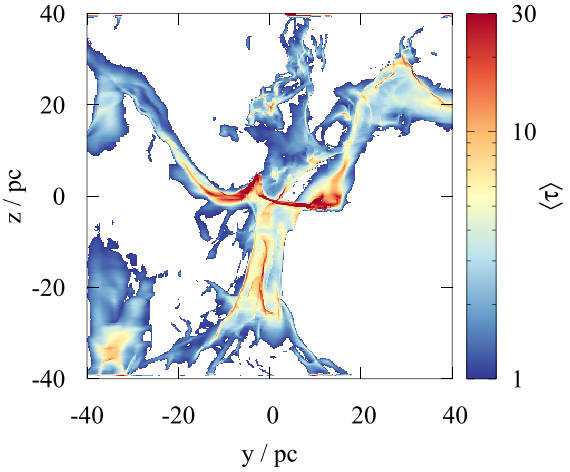

Figure 10. Map of the H I optical depth proxy $\langle\tau\rangle$ (see Appendix B) for MC1-HD at 2 Myr. The dense cloud region (compare with bottom middle panel of Fig. 2) has an average optical depth $\gtrsim$1; thus, optical depth corrections cannot be neglected when calculating H I column densities.

of 150 pc and 3 kpc we find little differences. This further supports our claim that the underestimation of $N_{\rm H\,I,real}$ can be attributed – at least in parts (see Section 3.3) – to the non-uniform H I temperatures present in the clouds.

Finally, we note that the rather low H I temperatures found in our simulations ($\lesssim$40 K, Fig. 6) are in good agreement with a number of observations of Galactic MCs, which find typical H I temperatures between 10 and 40 K (Gibson et al. 2000; Kavars et al. 2003, 2005; Klaassen et al. 2005; Fukui et al. 2014, 2015; Stanimirović et al. 2014; Dénes et al. 2018; Nguyen et al. 2019). They are, however, lower than typical temperatures found by Wang et al. (2020b) in the giant molecular filament GMF38a. A possible reason for this might be stellar feedback heating the gas in GMF38a. As a consequence, the H I column densities determined in Wang et al. (2020b) might be more accurate than in our case.

3.3 The H I optical depth

In order to investigate the typical optical depths in our clouds, in Fig. 10 we plot a proxy for the H I optical depth, $\langle\tau\rangle$, for MC1-HD at 2 Myr. The definition of $\langle\tau\rangle$ is given in Appendix B. It represents a channel-averaged approximation to the real optical depth that is accurate within a few 10 per cent above $\langle\tau\rangle = 1$, i.e. in the optically thick regions we are interested in here. For optically thin regions, the approximation is not applicable, which is why we do not show these regions here. The values of $\langle\tau\rangle$ span a wide range, from the moderately optically thick regime up to highly optically thick regions with $\langle\tau\rangle \sim 10$. In particular, the entire area of central cloud (compare with bottom middle panel of Fig. 2) has an optical depth $\gtrsim$1, which is in excellent agreement with recent observations (e.g. Fukui et al. 2014, 2015; Bihr et al. 2015; Dénes et al. 2018; Murray et al. 2018; Nguyen et al. 2019; Syed et al. 2020; Wang et al. 2020b). Our results thus demonstrate that optical depth effects cannot be neglected in H I observations of MCs, also when calculating $N_{\rm H\,I}$ from H I emission observations.

We note that at first view the maximum values of $\langle\tau\rangle > 10$ appear high in comparison with those found in the aforementioned observational works. However, as these measurements are limited by observational noise, ΔT, the observationally reported values have to be taken as lower limits (see e.g. fig. 10 of Bihr et al. 2015).

In regions of high optical depth, the observed brightness temperature $T_{\rm on}$ will be close to the spin temperature $T_{\rm s}$ of the absorbing, cold H I layer, i.e. $T_{\rm off-on} \simeq 100\,{\rm K} - T_{\rm s}$. Hence, choosing $T_{\rm HISA} \lesssim T_{\rm s}$ will result in an underestimation of $N_{\rm H\,I,real}$ for optically thick regions as well, as discussed in Section 3.2. Moreover, for $T_{\rm HISA} \simeq T_{\rm s}$ an additional source of error in such optically thick regions is caused by the observational noise ΔT, as now $T_{\rm off-on} \simeq 100\,{\rm K} - T_{\rm HISA} - \Delta T$ (from equation 5). Inserting this into equation (6) yields

$$\begin{aligned}\tau_{\rm HISA,noise} &= -\ln\left(1 - \frac{100\,{\rm K} - T_{\rm HISA} - \Delta T}{100\,{\rm K} - T_{\rm HISA}}\right)\\ &= -\ln\left(\frac{\Delta T}{100\,{\rm K} - T_{\rm HISA}}\right).\end{aligned} \tag{11}$$

Analysing equation (11) shows that the observational uncertainty ΔT in highly optically thick regions (if $T_{\rm HISA} \simeq T_{\rm s}$) results in an underestimation of $N_{\rm H\,I,obs}$ *regardless* of its sign:

(i) $\Delta T < 0$: If noise artificially lowers $T_{\rm on}$, this *increases* $T_{\rm off-on}$ beyond a value of 100 K $- T_{\rm HISA}$ in a highly optically thick region. Hence, equation (11) would contain a negative expression in the logarithm and the contribution from the corresponding velocity channel has to be omitted.

(ii) $\Delta T > 0$: If noise, but also the potential emission of warm and diffuse H I in the foreground, increases $T_{\rm on}$ (and thus decreases $T_{\rm on-off}$), this results in an underestimation of the true value of $\tau_{\rm HISA}$ (which can be larger than $\tau_{\rm HISA,noise}$) and thus also $N_{\rm H\,I,obs}$ (equation 7). The effect of foreground emission is thus also related to the problem of identifying $T_{\rm HISA}$ correctly.

Hence, even if one were to choose the correct value of $T_{\rm HISA}$, $N_{\rm H\,I,obs}$ is in general underestimated in optically thick regions (see regions of high $N_{\rm H\,I,real}$ in Fig. 5). The observational noise contributes to the fact that even at the peaks of the mean-value lines in the bottom panel of Fig. 8, where $T_{\rm HISA} \simeq T_{\rm dip}$, the real H I column density is on average underestimated. We emphasize that this underestimation due to ΔT adds on top of the problem to determine a reasonable value of $T_{\rm HISA}$. This effect also contributes to the more pronounced underestimation at low values of $T_{\rm dip}$ (higher values of $N_{\rm H\,I,real}$, bottom panel of Fig. 8): the lower $T_{\rm dip}$, the higher is the H I column density and thus the optical depth, which amplifies the issue arising from this effect.

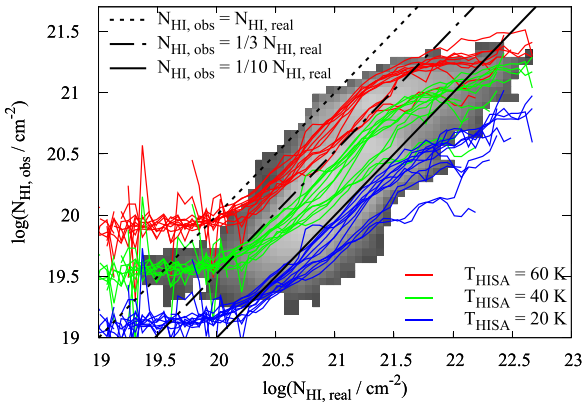

Figure 11. Same as in Fig. 5 but now including the correction in optically thick regions (see the text). Overall, the match is somewhat improved. However, depending on the choice of $T_{\rm HISA}$, $N_{\rm H\,I,real}$ is still underestimated by a factor of a few up to ∼10 in particular in the densest regions.

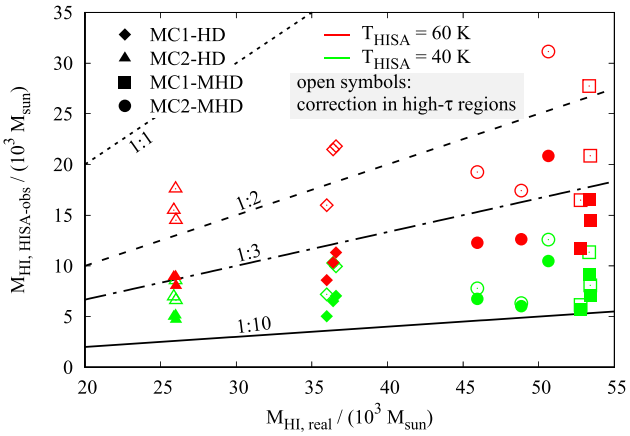

Figure 12. Accuracy of the H I mass inferred from HISA observations for the four different MCs at 2 Myr (symbols) for $T_{\rm HISA} = 40$ K (green) and 60 K (red). The black lines show the different mass ratios to guide the readers' eye. Overall, we find that the H I mass is underestimated by a factor of a few up to ∼10. Correction for optically thick channels (open symbols) improves the accuracy only moderately. Note that depending on the chosen projection direction, the mass in the observable area is different for the same MC.

3.3.1 Opacity correction in MCs

Motivated by the large extent of optically thick regions in MCs (Fig. 10), we suggest a method to improve the accuracy of $N_{\rm H\,I,obs}$: A significant underestimation occurs in the high-$N_{\rm H\,I}$/high-optical depth regions of the MCs, where equation (6) yields no result for $\tau_{\rm HISA}$ any more and velocity channels have to be omitted (see Fig. 8). Hence, for these velocity channels we use an optical depth set by the typical rms noise (ΔT) of the observation, which is given by $\tau_{\rm HISA,noise}$ (equation 11), e.g. for $T_{\rm HISA} = 40$ K and the adopted noise of 3 K (Section 2.2.1), we obtain $\tau_{\rm HISA,noise} = 3.0$. We emphasize that this estimate is still a conservative estimate as the actual optical depth is likely to be higher. A similar approach is also followed by Bihr et al. (2015) for H I emission maps. We note that this approach has to be considered under the premise that, as shown before, a constant value of $T_{\rm HISA}$ over the entire map is an oversimplification, which in addition does also not account for the temperature variations along the LOS (see Section 3.2).

The obtained column density maps are shown in Fig. A2 in Appendix A. As can be seen, $N_{\rm H\,I,obs}$ in the denser parts of the MCs is represented better than before (compare with Fig. 4). In the very densest parts, however, $N_{\rm H\,I,real}$ is still significantly underestimated. This is also visible in Fig. 11, where we show the mean values of $N_{\rm H\,I,obs}$ for all MCs and directions at 2 Myr using this correction. There is an improvement in all areas compared to the case without any correction (see Fig. 5). However, $N_{\rm H\,I,real}$ can still be underestimated by a factor of a few to ∼10. Hence, the suggested method has only a moderate impact on increasing the accuracy, both due to the non-isothermality of the H I gas and the fact that $\tau_{\rm HISA,noise}$ is most likely lower than the real optical depth.

3.4 The cold H I budget of molecular clouds

Summarizing the findings of the previous sections, we find that the uncertainty in determining the H I column density is due to (i) the assumption of a fixed temperature $T_{\rm HISA}$ for the calculation of the HISA column densities and (ii) noise in the temperature brightness measurement. As a consequence, either the optical depth and the true $T_{\rm HISA}$ are underestimated (mainly in the outer parts of MCs) or velocity channels have to be omitted for the calculation of the column density (mainly in the densest parts of MCs). Overall, this results in a significant underestimation of the actual H I column densities by a factor of 3–10 (and even more in the densest regions of clouds).

This is also reflected in the total mass of cold H I in MCs inferred from HISA observations, $M_{\rm H\,I,HISA-obs}$, shown in Fig. 12. Here, we add up the observed H I mass of all pixels for which an HISA feature is identified (i.e. pixels outside the coloured regions in Fig. 4 are ignored) using $T_{\rm HISA} = 40$ and 60 K. For the actual mass, $M_{\rm H\,I,real}$, we only take into account H I gas with $T < 100$ K, to which our HISA observations are sensitive to. As for the column densities, also the total, cold H I mass is typically underestimated by a factor of a 3–10, when no correction in the optically thick regions is applied (filled symbols). The correction (Section 3.3.1), however, improves the accuracy only moderately by a factor of ∼1.5–2 (open symbols). We emphasize that increasing $T_{\rm HISA}$ to obtain apparently more accurate mass estimates should be considered with caution. This merely leads to an overestimation of $N_{\rm H\,I}$ at low column densities compensating the underestimation at high column densities (see Fig. 5 and also Section 3.5).

We emphasize that our results do not change significantly among the different MCs considered, i.e. whether or not dynamically important magnetic fields are present. This indicates that HISA observations in general tend to significantly underestimate the cold H I budget in MCs. This is markedly different to a complementary study for the more diffuse ISM (on scales $\gtrsim$1 pc) of Kim et al. (2014) and Murray et al. (2015, 2017), who find that in this regime H I absorption observations can trace the H I mass with an accuracy of a few 10 per cent. We tentatively attribute this to the fact that the authors probe several and more diffuse H I clouds along significantly longer LOS of several 100 pc length. These clouds might have lower optical depths and are thus less prone to the measurement uncertainties mentioned in Section 3.3. Furthermore, H I masses obtained from emission observations (e.g. Bihr et al. 2015), which correct for the optical depth and which also implicitly take into account warm H I ($T > 100$ K, in our case 22–47 per cent), might achieve more accurate H I masses, a topic not investigated in this study.

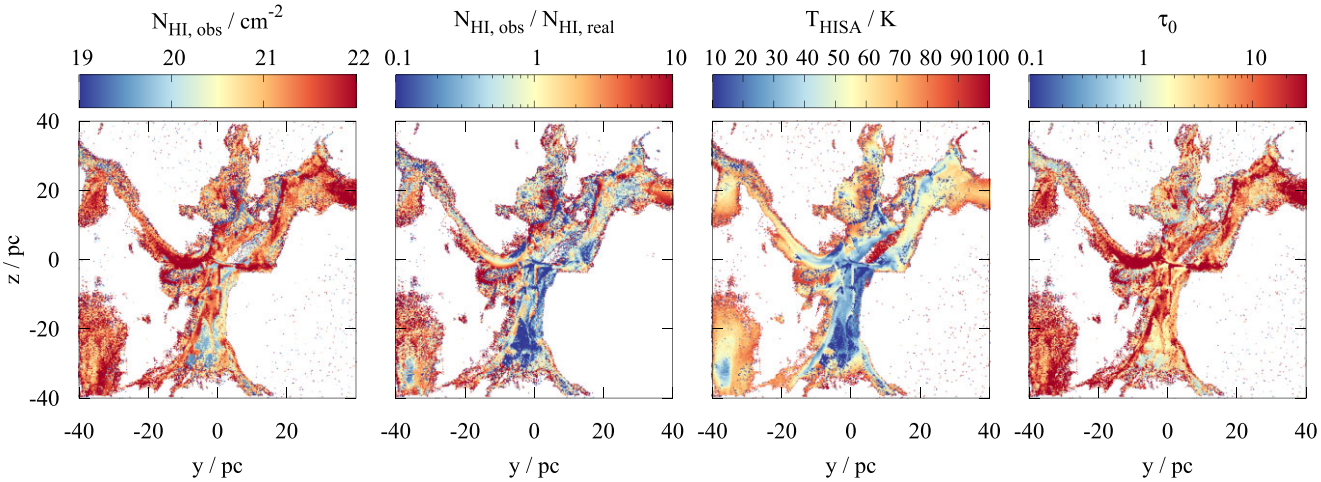

Figure 13. Maps of $N_{\rm H\,I,obs}$ and its ratio to the actual H I column density (two left-hand panels) determined by leaving $T_{\rm HISA}$ and τ_0 (two right-hand panels) as free parameters. The maps are shown for the same snapshot as in Fig. 4. Overall, allowing $T_{\rm HISA}$ to be a free parameter does not increase the accuracy compared to assuming a fixed $T_{\rm HISA}$ (compare Fig. 4). Although the determined $T_{\rm HISA}$ is similar to the mass-weighted H I temperature (right-hand panel of Fig. 6), the optical depth shows a partly unphysical behaviour with high values in the outer, low-column density areas.

3.5 Deriving $N_{\rm H\,I}$ with a variable $T_{\rm HISA}$

3.5.1 T_{HISA} *as a free fit parameter*

Following the results of Section 3.2, we next leave $T_{\rm HISA}$ as a free parameter. We determine its value and the optical depth by assuming a Gaussian optical depth profile, i.e. inserting equation (9) into equation (5) and fitting the observed H I spectrum. In Fig. 13, we show the various quantities obtained by the approach for MC1-HD at 2 Myr and an assumed distance of 150 pc. For other directions, times, clouds, and the 3 kpc-distance case, we find qualitatively and quantitatively similar results. Overall, we find a quite poor match between the observed and actual $N_{\rm H\,I}$ (second panel from the left in Fig. 13 and top panel of Fig. 14): Although the mean values of the distribution (orange lines in the top panel of Fig. 14) show a reasonable match for $N_{\rm H\,I,obs}$ and $N_{\rm H\,I,real}$ below $\sim 10^{21.5}$ cm^{-2}, there is a significant scatter of more than 1 dex.

We attribute this rather poor match mainly to (i) the occurrence of multiple Gaussian absorption components in the spectra (see Fig. 3), which are not accounted for in our simplistic model, and – to a lesser extent – to (ii) the lack of spectral resolution (1 km s^{-1}) and (iii) observational noise. As a consequence, it is not possible to reliably determine the optical depth with our fitting approach, which would required accurate spectral information, also about the wings of the spectrum. This is visible in the obtained values of τ_0 (rightmost panel in Fig. 13), which show no clear correlation with the underlying column density distribution $N_{\rm H\,I,real}$ as opposed to the optical depth proxy $\langle\tau\rangle$ shown in Fig. 10.

We emphasize that when repeating the method for the noiseless, high-resolution spectra (200 m s^{-1}), we obtain a similar poor match between $N_{\rm H\,I,obs}$ and $N_{\rm H\,I,real}$. This further supports our assumption that the poor match is in parts due to the occurrence of multiple Gaussian absorption components not accounted for here and not due to a generic problem of this approach. We therefore suggest that multiple Gaussian components with individual temperatures have to be taken into account (e.g. Heiles & Troland 2003; Stanimirović et al. 2014; Murray et al. 2015; Dénes et al. 2018) to get a better match with $N_{\rm H\,I,real}$. This might, to some extent, also remedy the temperature problem discussed in Section 3.2 as each component can be assigned an individual temperature. We will, however, postpone this investigation to future work.

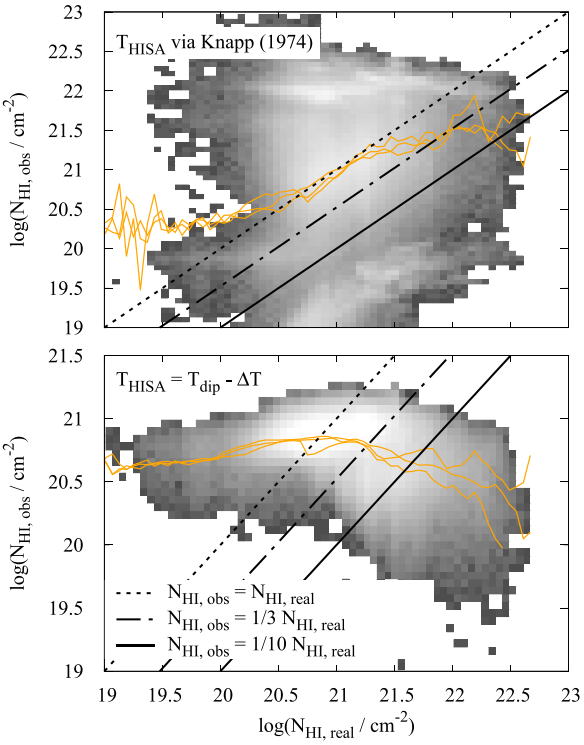

Figure 14. Mean value of $N_{\rm H\,I,obs}$ against $N_{\rm H\,I,real}$ (orange lines) for three different directions of MC1-HD at $t_{\rm evol} = 2$ Myr using the method of Knapp (1974) (top) and equation (12) with $\Delta T = 3$ K (bottom). In the background, the full distribution for one LOS is shown in greyscale. The black lines show lines of constant ratio $N_{\rm H\,I,obs}/N_{\rm H\,I,real}$ to guide the readers' eye. In general, the qualitative match between $N_{\rm H\,I,obs}$ and $N_{\rm H\,I,real}$ is rather poor with a significant scatter for the Knapp (1974) method. Note the different ranges on the y-axis.

Finally, we note that the fitted values of $T_{\rm HISA}$ (second panel from the right in Fig. 13) appear roughly comparable to the mass-weighted mean temperatures (right-hand panel of Fig. 6). However, also here strong temperature variations along the LOS (Section 3.2) can affect the fit value of $T_{\rm HISA}$. As a consequence, as τ_0 and $T_{\rm HISA}$ are

degenerate, overestimating (underestimating) $T_{\rm HISA}$ requires a higher (lower) τ_0 to match the observed $T_{\rm off-on}$ at the dip of the absorption spectrum (equation 5). Following equation (7), this directly leads to a too high (low) value of $N_{\rm H\,I,obs}$.

3.5.2 T_{HISA} given by T_{dip}

Given the similarity of $T_{\rm dip}$ and the mass-weighted H I temperature $T_{\rm H\,I,mw}$ (see Fig. 7), as well as the fact that the highest accuracy for $N_{\rm H\,I,obs}$ was found where $T_{\rm HISA} \simeq T_{\rm dip}$ (bottom panel of Fig. 8), we also try an alternative approach by setting $T_{\rm HISA}$ close to, but slightly below $T_{\rm dip}$. In detail, we set

$$T_{\rm HISA} = T_{\rm dip} - \Delta T \tag{12}$$

with $\Delta T = 3$ K being the noise level of the synthetic observations and test the approach for MC1 at $t_{\rm evol} = 2$ Myr.

As for the method from Knapp (1974), we find a qualitatively poor match between $N_{\rm H\,I,obs}$ and $N_{\rm H\,I,real}$ (bottom panel of Fig. 14). The method gives a rather flat distribution of $N_{\rm H\,I,obs}$ with too high values at low $N_{\rm H\,I,real}$ and drops towards higher $N_{\rm H\,I,real}$. Overall, the reasons for this poor match are again the temperature variations along the LOS (Section 3.2), the partly high optical depths (for the high-$N_{\rm H\,I}$ regions, Section 3.3) as well as the degeneracy of $T_{\rm HISA}$ and $\tau_{\rm HISA}$ (for the low-$N_{\rm H\,I}$ regions, Section 3.5.1). We emphasize that also the usage of $T_{\rm H\,I,mw}$ – which is not accessible to an observer – for $T_{\rm HISA}$ does not improve the situation but gives qualitatively similar results as using equation (12). We attribute this to the fact that $T_{\rm H\,I,mw}$ and $T_{\rm dip}$ are similar within a scatter of 10–20 K (see Fig. 7).

To summarize, even when choosing $T_{\rm HISA}$ by a physically motivated approach, the quality of the obtained H I column density maps does not increase, but partly even decreases. Contrary to the approach of a fixed $T_{\rm HISA}$, for these approaches not only the quantitative agreement but also the qualitative agreement between $N_{\rm H\,I,obs}$ and $N_{\rm H\,I,real}$ is lost.

3.6 The H I velocity dispersion

Finally, we consider the accuracy of the H I velocity dispersion inferred from HISA observations. For this purpose, we compare the non-thermal velocity dispersion $\sigma_{\rm obs}$ identified via the BTS tool (Section 2.3) with the actual H I velocity dispersion along each LOS directly inferred from the simulation data, $\sigma_{\rm real}$. For $\sigma_{\rm real}$, we only take into account the velocity component along the LOS and H I with a temperature below 100 K and then calculate for each pixel the H I mass-weighted LOS average. For $\sigma_{\rm obs}$, we correct the value obtained by BTS, $\sigma_{\rm BTS}$, for the contribution from the limited channel width of 1 km s^{-1} and thermal motions, i.e.

$$\sigma_{\rm obs} = \left(\sigma_{\rm BTS}^2 - \left(\frac{1\ {\rm km\ s^{-1}}}{\sqrt{8\ln 2}} \right)^2 - c_{\rm s}^2 \right)^{1/2}, \tag{13}$$

with $c_{\rm s}$ being the sound speed for H I gas, where, for simplicity, we assume an average temperature of 40 K (right-hand panel of Fig. 6). The factor $\sqrt{8\ln 2}$ accounts for the conversion of channel width into the standard deviation. In the following, we only consider pixels where $\sigma_{\rm obs} > 0$.

In the left-hand and middle panels of Fig. 15, we plot the distribution of $\sigma_{\rm real}$ and $\sigma_{\rm obs}$ and its mean value (black line) as a function of $N_{\rm H\,I,real}$ for MC1-HD at $t_{\rm evol} = 2$ Myr for one direction assuming a distance of 150 pc for the beam size. We note that the following results also hold for the other clouds and times. First, we find that the scatter for $\sigma_{\rm real}$ appears to be somewhat larger than for $\sigma_{\rm obs}$. Secondly, for $\sigma_{\rm real}$ there is only a moderate increase with $N_{\rm H\,I,real}$ in particular for $N_{\rm H\,I,real} > 10^{21}$ cm^{-2}, whereas for $\sigma_{\rm obs}$ the increase is more pronounced. The latter could be attributed to opacity broadening occurring for high $N_{\rm H\,I,real}$ (see black line in Fig. 3), which we expect to happen frequently, given the high optical depths found in our MCs (Fig. 10). The presence of multiple Gaussian components in the spectrum (red line in Fig. 3), however, cannot explain the somewhat larger values of $\sigma_{\rm obs}$ compared to $\sigma_{\rm real}$: Although the result of the single-component fit for $\sigma_{\rm BTS}$ will be broader than the velocity dispersion of the individual H I components, multiple components will also increase $\sigma_{\rm real}$. We also note that the non-thermal velocity dispersions of H I of a few 1 km s^{-1} reported here are in general in agreement with the velocity dispersion of dense gas ($n > 100$ cm^{-3}) in these clouds (Seifried et al. 2017).

The black lines in the right-hand panel of Fig. 15 show the mean of $\log(\sigma_{\rm obs}/\sigma_{\rm real})$ for all clouds and projection directions at 2 Myr and an assumed distance of 150 pc. Despite the differences seen in the left-hand and middle panels, $\sigma_{\rm obs}$ appears to trace the actual velocity dispersion with a reasonable accuracy. For $N_{\rm H\,I,real} \lesssim 10^{22}$ cm^{-2}, the mean of $\log(\sigma_{\rm obs}/\sigma_{\rm real})$ is one average within ± 0.3 dex around a value of 0, which would indicate a perfect agreement. Also, the standard deviations of the various distributions (grey lines) are about 0.2–0.3 dex. Hence, we argue that for typical H I column densities between $\sim 10^{20}$ and $\sim 10^{22}$ cm^{-2}, HISA observations are able to probe the non-thermal velocity dispersion of H I with an accuracy of a factor of ~ 2, even in the case of a limited spectral resolution of ~ 1 km s^{-1}. Only for very high H I column densities ($N_{\rm H\,I,real} > 10^{22}$ cm^{-2}), the non-thermal velocity dispersion might be somewhat overestimated, which we tentatively attribute to the aforementioned opacity broadening of the absorption feature. We note that these results also hold for $t_{\rm evol} = 3$ Myr and an assumed distance of 3 kpc.

Finally, in Fig. 16 we show the distribution of the Mach numbers $\sqrt{3}\sigma/c_{\rm s}$ (assuming $T = 40$ K) for MC1-HD at 2 Myr. The values are inferred directly from the simulation using $\sigma_{\rm real}$ (black lines) and from the HISA observation using $\sigma_{\rm obs}$ (red lines). For other clouds, we find similar results. Overall, the H I gas is moderately supersonic with the distribution peaking around Mach numbers of a few, although also values up to ~ 10 are reached. The observationally determined Mach numbers somewhat underrepresent the lowest values. This could be related to the aforementioned opacity broadening or a broadening due to the limited spectral resolution. The overall similarity between both distributions, however, confirms the accuracy of a factor of ~ 2 between $\sigma_{\rm real}$ and $\sigma_{\rm obs}$ reported before. Furthermore, the Mach numbers found are also in good agreement with recent HISA observations (Burkhart et al. 2015; Nguyen et al. 2019; Syed et al. 2020; Wang et al. 2020b).

4 DISCUSSION

4.1 The $N_{\rm H\,I}$ PDF: comparison with observations

As discussed in Section 3.1, we find that HISA measurements tend to underestimate the actual column density of cold H I by a factor of 3–10 in the outer parts of MCs and potentially by an even higher factor in the central, high-density parts. We attribute this to (i) a large temperature variation of the H I gas and the assumption of a fixed $T_{\rm HISA}$ and (ii) the effect of noise in the brightness temperature measurements in particular for regions of high optical depth.

This underestimation is again demonstrated in Fig. 17 where we show the area-weighted PDFs of $N_{\rm H\,I,real}$ and $N_{\rm H\,I,obs}$ for MC1-HD at

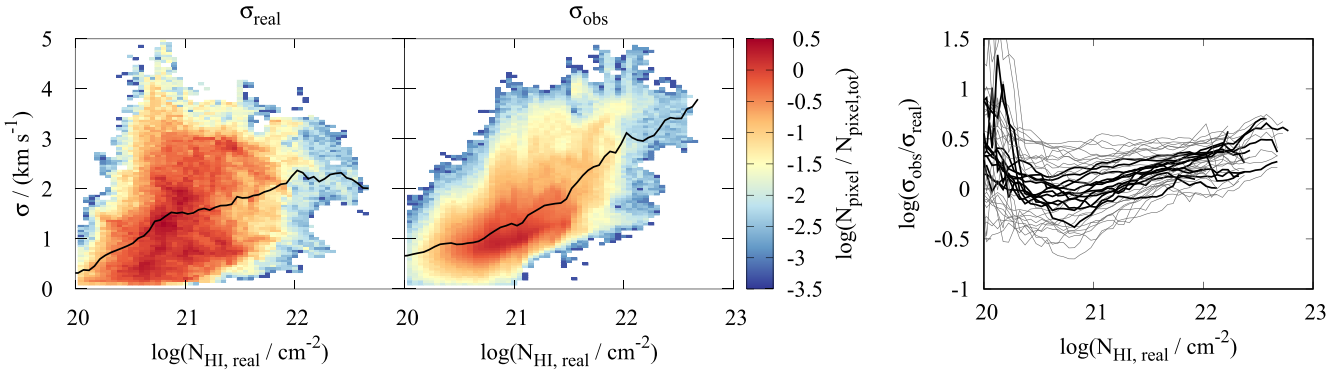

Figure 15. Left-hand and middle panels: Distribution of the real non-thermal H I velocity dispersion, $\sigma_{\rm real}$, and the observed one, $\sigma_{\rm obs}$, as a function of $N_{\rm H\,I,real}$ for MC1-HD at $t_{\rm evol} = 2$ Myr for one direction assuming a distance of 150 pc. Overall, there is an increase of σ with the column density, although the effect is less pronounced for $\sigma_{\rm real}$. The stronger increase of $\sigma_{\rm obs}$ could be due to opacity broadening. Right-hand panel: Ratio of $\sigma_{\rm obs}$ and $\sigma_{\rm real}$ as a function of $N_{\rm H\,I,real}$. The black line shows the mean for all clouds and three directions at $t_{\rm evol} = 2$ Myr assuming a distance of 150 pc; the grey lines represent the interval of one standard deviation. Overall, HISA observations trace the velocity dispersion with an accuracy of a factor of ∼2.

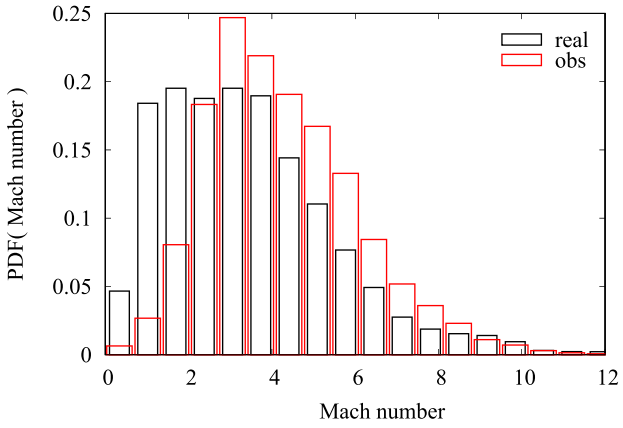

Figure 16. Mach number distribution of the H I gas for the same snapshot as in Fig. 4 assuming a temperature of 40 K. The black lines show the distribution obtained from the simulation data directly ($\sigma_{\rm real}$), and the red that from the observed velocity dispersion ($\sigma_{\rm obs}$). The H I gas is moderately supersonic with Mach numbers around 1–10.

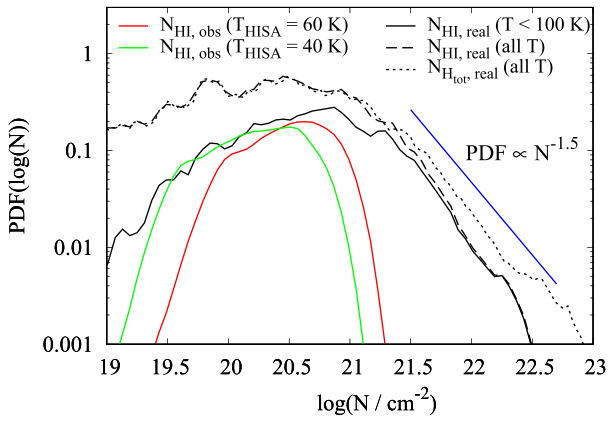

Figure 17. Column density PDFs of the cold H I (black solid line), the total H I (black dashed line), and all hydrogen nuclei ($H_{\rm tot}$, black dotted line) inferred directly from the simulation of MC1-HD at 2 Myr for one direction. In addition, the corresponding PDFs determined from the synthetic HISA observations assuming $T_{\rm HISA} = 40$ and 60 K (green and red lines, respectively) are shown. The $N_{\rm H\,I,obs}$ PDFs peak at significantly lower values than those of $N_{\rm H\,I,real}$ and show a roughly lognormal distribution. In contrast, the $N_{\rm H\,I,real}(N_{\rm H,tot})$ PDFs exhibit a power-law tail roughly proportional to $N^{-1.5}$ (blue line), which indicates that also H I is gravitationally unstable.

$t_{\rm evol} = 2$ Myr for one direction at an assumed distance of 150 pc.[7] The latter is derived for $T_{\rm HISA} = 40$ and 60 K (coloured lines). As expected, the peak of the $N_{\rm H\,I,obs}$ PDF is shifted by a factor of 3–10 towards lower column densities with respect to that of the $N_{\rm H\,I,real}$ PDF (black solid line). Moreover, also the shapes of the two PDFs are very different with important implications. First, the $N_{\rm H\,I,real}$ PDF is significantly broader than the $N_{\rm H\,I,obs}$ PDF. This indicates that the width of the $N_{\rm H\,I,obs}$ PDF obtained from HISA observations might not be a good quantity to assess turbulence statistics (see e.g. Burkhart & Lazarian 2012, for an application to the $N_{\rm H,tot}$ PDF). Secondly, the $N_{\rm H\,I,real}$ PDF shows signs of a power-law tail at column densities above a few 10^{21} cm^{-2}, which is roughly proportional to $N^{-1.5}$, similar to that of the $N_{\rm H,tot}$ PDF (black dotted line; see also Kainulainen et al. 2009; Kritsuk, Norman & Wagner 2011; Girichidis et al. 2014; Schneider et al. 2015; Auddy, Basu & Kudoh 2018; Veltchev et al. 2019). This power-law tail indicates that also the dense H I gas is undergoing gravitational collapse. We emphasize, however, that – as the dense gas ($N \gtrsim 10^{21}$ cm^{-2}) is predominantly molecular (bottom panel of Fig. 1) – the gravitational force in this range is dominated by gas in form of H_2 with which the H I is mixed.

Our $N_{\rm H\,I,real}$ PDFs are markedly different from PDFs found in recent HISA observations (Burkhart et al. 2015; Imara & Burkhart 2016; Syed et al. 2020; Wang et al. 2020b) that find a lognormal shape, indicating that the cold H I is not gravitationally unstable. This apparent contradiction could have its origin in possible observational biases, indicated by some striking similarities between our synthetic H I observations and that of the aforementioned authors: Our synthetic and the actually observed $N_{\rm H\,I,obs}$ PDFs are of roughly lognormal shape, are in a similar range ($N_{\rm H\,I} = 10^{20-21}$ cm^{-2}), and are at significantly lower column densities than that of either H_2 and

[7]We note that the integrated area under the curves is not necessarily unity as many pixels have no observed H I column density (Fig. 4). Hence, the relative area under the curves gives the reader a direct indication of how many pixels are omitted (i.e. have $N_{\rm H\,I,obs} = 0$) compared to the PDF of all H I.

the total H I observed in emission (Syed et al. 2020; Wang et al. 2020b) or that of $N_{\rm H\,I,tot}$ measured via dust emission (Burkhart et al. 2015; Imara & Burkhart 2016). Taking these similarities into account, we suggest that there indeed is an observation bias in the shape of observed $N_{\rm H\,I}$ PDFs. This could be particularly pronounced at the high end of the PDF, which is often characterized by a power law. In addition, the $N_{\rm H\,I}$ values obtained should rather be considered as lower thresholds. We note, however, that an MC at a very early evolutionary stage might not have undergone gravitational collapse, i.e. might not yet have developed high column densities ($\gtrsim 10^{21}$ cm^{-2}) and the associated power-law tail in the *N*-PDF. Therefore, for such an MC, the assessment of the H I column densities and masses via HISA observations might still be somewhat better compared to the findings presented here.

4.2 Multiple H I–H_2 transitions: comparison with analytical results

As stated before, recent semi-analytical works predict that, for ISM conditions comparable to that of the solar neighbourhood, the transition from H I to H_2 occurs at column densities of $\lesssim 10^{21}$ cm^{-2} (Krumholz et al. 2008, 2009; Sternberg et al. 2014; Bialy & Sternberg 2016). These models also suggest an upper limit of $N_{\rm H\,I}$ around this value. Contrary to that, for the clouds simulated in this work, we find H I column densities partly well above this value (see Figs 1 and 2). Also, recent observations of W43 (Motte et al. 2014; Bihr et al. 2015) and indirect estimates towards Perseus (Okamoto et al. 2017) and clouds outside the Galactic plane (Fukui et al. 2014, 2015) have revealed H I column densities of up to a few 10^{22} cm^{-2}, thus well comparable to our findings, but in apparent contradiction to the theoretical predictions. Also, observations of the Magellanic Clouds by Welty, Xue & Wong (2012) seem to challenge the prediction for the value of $N_{\rm H,tot}$, where the transition to H_2-dominated gas is supposed to occur (Krumholz et al. 2008, 2009; McKee & Krumholz 2010).

However, strictly speaking, the suggested upper limit around 10^{21} cm^{-2} only applies to a single H I–H_2 transition. As pointed out by Motte et al. (2014), a possible solution for this contradiction could thus be the presence of several transitions along the LOS (see also Bialy et al. 2017a). Indeed, the occurrence of multiple absorption components (see Fig. 3) and the highly complex structure of our simulated MCs (see Fig. 2) and of real MCs indicate that the assumption of a single H I–H_2 transition might be an oversimplification and that rather several transitions are present.

In the following, we test this hypothesis in physical space (i.e. considering spatial distances) as opposed to parts of the analysis done before in velocity space (Section 3). For this purpose, we identify the number of H I-density peaks along rays which intersect the entire length of the zoom-in regions and which are distributed uniformly over the entire area of the emission maps with a spacing of 0.24 pc. For this purpose, we first determine the profile of the logarithm of the H I density, $\log(n_{\rm H\,I})(l)$, along each ray, i.e. along the LOS of each pixel. A few selected profiles are shown in Fig. A4, demonstrating their variability for different pixels, which necessitates a more systematic study. Secondly, we identify the positions of the peaks of $\log(n_{\rm H\,I})$ along the LOS, $l_{\rm peak,k}$, as well as the position of their left and right bases, $l_{\rm base,left,k}$ and $l_{\rm base,right,k}$, respectively (see Fig. 18 for a schematic view). The base of a peak (green squares) is the minimum of the density profile between this and the neighbouring peak. In order to account for the influence of small-scale density fluctuations, we discard peaks that (i) have a prominence (blue arrow) of less than 0.5 (in log-space), or (ii) have a peak height $\log(n_{\rm H\,I,peak})$ of less than 0.5 ($n_{\rm H\,I,peak} \simeq 3$ cm^{-3}), or (iii) are separated from the next (and higher) peak by less than two grid cells along the LOS.[8] Furthermore, we note that as $n_{\rm H\,I}$ comes from the chemical network implemented in the simulations, a low $n_{\rm H\,I}$ could indicate either a low total gas density or a high gas density, where hydrogen is already predominantly in form of H_2.

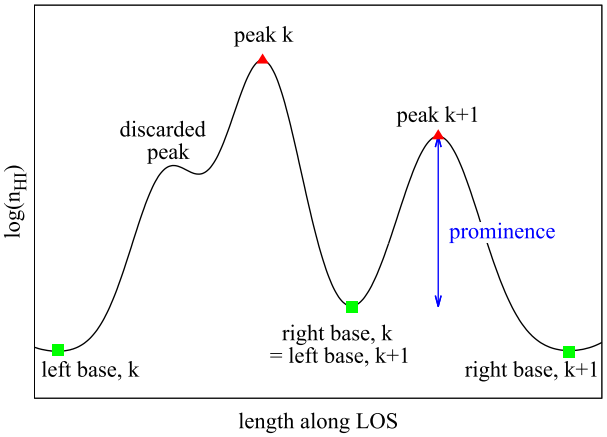

Figure 18. Sketch for the identification of H I peaks (red triangles) and the position of their bases (green squares), also for a case where a local peak is discarded. The blue arrow indicates the prominence of peak $k+1$.

Next, we calculate the column density of each H I-density peak as

$$N_{\rm H\,I,peak,k} = \int_{l_{\rm base,left,k}}^{l_{\rm base,right,k}} n_{\rm H\,I} {\rm d}l, \tag{14}$$

and determine the average column density, $\langle N_{\rm H\,I,peak}\rangle$, of all peaks along a given LOS. In addition, we identify the most massive peak, i.e. the peak that accumulates most of the H I along the given LOS, and determine its column density, $N_{\rm H\,I,peak,max}$. In Fig. 19, we plot the number of peaks along each LOS, $\langle N_{\rm H\,I,peak}\rangle$, $N_{\rm H\,I,peak,max}$, and the ratio $N_{\rm H\,I,peak,max}/N_{\rm H\,I,real}$ for MC1-HD at 2 Myr along one direction. The last value shows how much the most massive peak contributes to the overall H I column density.

The left-hand panel of Fig. 19 shows that there are indeed up to ∼10 H I-density peaks along the LOS as suggested by Bialy et al. (2017a) for the case of W43. There is a moderate tendency of a higher number of peaks towards the centre of the MC, i.e. with increasing $N_{\rm H\,I,real}$, probably caused by the filamentary substructure of the MCs. This increase causes $\langle N_{\rm H\,I,peak}\rangle$ to remain below 10^{21} cm^{-2} for the vast majority of rays (second panel of the left); only for about 5–25 per cent of all rays (depending on the cloud, direction, and time) it exceeds this value. Hence, on first view this appears to be in rough agreement with analytical predictions (Krumholz et al. 2008, 2009; Sternberg et al. 2014; Bialy & Sternberg 2016).

However, the column density of the dominant peak, $N_{\rm H\,I,peak,max}$, exceeds the value of 10^{21} cm^{-2} for a large number of rays (second panel from the right). We find that $N_{\rm H\,I,peak,max}$ exceeds 10^{21} cm^{-2} for 30–50 per cent of the rays, i.e. more than twice as often as

[8]When we discard a peak, the minimum between this discarded and the neighbouring peak *k* is *not* taken as the base of the peak *k*. Rather, the base of the peak *k* is shifted beyond the discarded peak such that the discarded peak now lies between $l_{\rm base,k}$ and $l_{\rm peak,k}$ (see leftmost green square in Fig. 18). We also note that we tested the approach by discarding the lower of two peaks when they are separated by less than four cells. This, however, affected the findings only marginally, which is why we do not follow this further here.

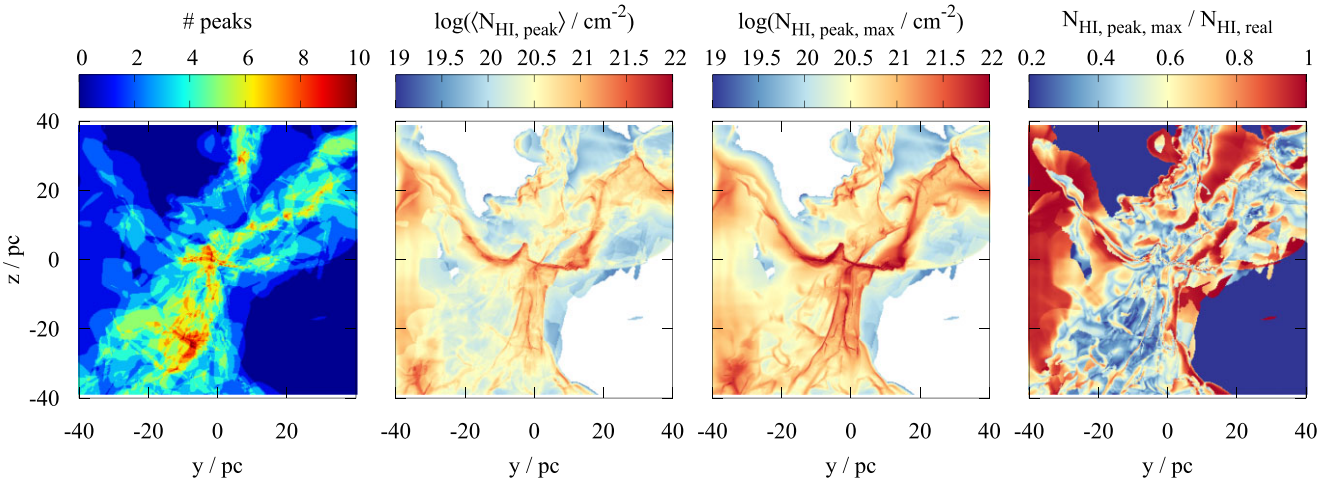

Figure 19. Maps of the number of H I-density peaks per LOS, the mean H I column density per H I-density peak, the H I column density of the most prominent H I-density peak, and the ratio of this value to the total H I column density (from left to right). The leftmost panel demonstrates the occurrence of up to 10 peaks, which corresponds to the same number of H I–H_2 transitions along the LOS. As a consequence, the average H I column density per peak (centre-left panel) mostly matches the theoretical predictions of $\lesssim 10^{21}$ cm^{-2}. However, the most prominent H I-density peak (centre-right panel) often has a column density that is considerably larger than this value and contributes significantly to $N_{\rm H\,I,real}$ ($\gtrsim$40 per cent, right-hand panel).

the corresponding fraction for $\langle N_{\rm H\,I,peak}\rangle$. Again, the exact fraction depends on the considered MC, direction, and time. However, we do not see any dependence on the presence or absence of magnetic fields in the simulations, despite the fact that the field is dynamically important for the overall (chemical) evolution of the MCs (see Seifried et al. 2020a, b). Furthermore, the dominant peak accounts on average for about 40–60 per cent and even more of $N_{\rm H\,I,real}$ (right-hand panel) and is thus indeed dominating the overall H I budget of the clouds. Hence, these findings are in contrast to the theoretical models, and we will discuss their implications in detail in the following section.

5 IS THERE MORE COLD H I THAN THOUGHT?

5.1 The theoretical perspective

As discussed so far, our results indicate that semi-analytical models tend to underestimate the maximum column density of cold H I in MCs. This can be attributed to several reasons. First, as suggested by Motte et al. (2014) and Bialy et al. (2017a) and explicitly shown here for the first time, in realistic models of MCs there appear up to $\sim$10 H I–H_2 transitions along the LOS. Secondly, non-equilibrium effects can increase the H I content of MCs due to the limited time available for H_2 to form (Glover & Mac Low 2007b; Glover et al. 2010; Mac Low & Glover 2012; Motte et al. 2014). We investigate this effect by artificially evolving the chemistry for a selected snapshot to chemical equilibrium (see Appendix A for details). Doing so, we find that this reduces the H I content in our MCs by a factor of 2–2.5 (see Fig. A3). Hence, the assumption of chemical equilibrium in semi-analytical models indeed results in too low H I abundances compared to the actual non-equilibrium H I present in dynamically evolving MCs.

In addition, we here suggest a third reason why even for a single H I–H_2 transition $N_{\rm H\,I}$ could be higher (centre-right panel of Fig. 19) than predicted by semi-analytical models with a 1D geometry (Krumholz et al. 2008, 2009; Sternberg et al. 2014; Bialy & Sternberg 2016). For this purpose, we consider the effect of assuming an idealized plane-parallel slab in more detail. In such a configuration, the gas in the cloud is irradiated only from one side; i.e. a parcel of gas having a high column density in the slab direction will receive very little radiation. In other words, for a plane-parallel slab – at a given total hydrogen nuclei column density $N_{\rm H,tot}$ towards the direction of the incident radiation – the extinction is maximal as radiation coming from other possible directions is neglected. As a consequence, also the amount of H I is minimal due to the lack of H_2 dissociation. The same chain of arguments can also be made for a spherically symmetric configuration.

In contrast, under realistic conditions in turbulent MCs, there may exist large low-density voids through which the radiation can propagate into the cloud (almost) unhindered. This is sketched in the left-hand panel of Fig. 20: A dense region shielded completely against UV radiation from one direction (horizontal) can still be irradiated from another direction (vertical). Hence, if observed from the horizontal direction, this region would *appear* completely optically thick (i.e. a high value of $A_{\rm V,1D\text{-}slab}$), corresponding to the assumption of a plane-parallel configuration. In reality, however, the region might still receive up to $\sim$50 per cent of the radiation coming from the vertical direction; i.e. the real $A_{\rm V}$ might be significantly lower.[9] As a consequence, H_2 might still be dissociated by UV radiation even in the central regions of MCs, thus increasing the amount of H I. Considering a 3D graphical representation of MC1-HD at 2 Myr (right-hand panel of Fig. 20) indeed shows that there exist these large low-density voids (bluish) through which radiation can travel (almost) unattenuated, thereby dissociating H_2 in the densest regions (reddish). Depending on the considered snapshot, 55–83 per cent of the volume of the zoom-in regions is covered by gas with densities below 1 cm^{-3}. We note that also for internal, stellar radiative feedback the actual cloud's substructure and shielding parameters have a similar importance (Haid et al. 2019).

[9]In order to calculate the average visual extinction in a point, one must not take the average $\langle A_{\rm V,i}\rangle$ over different directions i, but must take the logarithm of $\langle \exp(-\gamma A_{\rm V,i}\rangle$, as the latter describes the amount of incident radiation. The latter averaging puts more emphasis on the low-$A_{\rm V}$ directions.

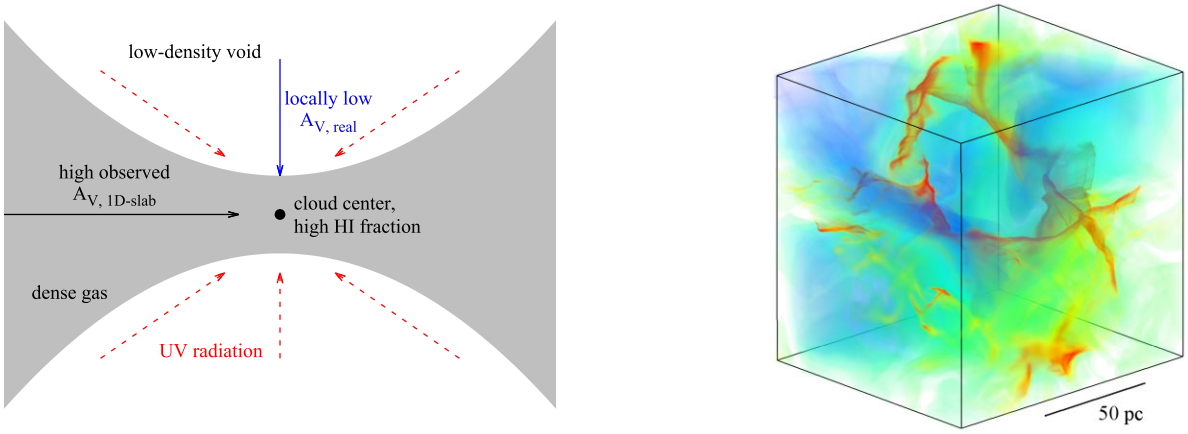

Figure 20. Left: Sketch of a situation in an MC (dense gas depicted with grey). Here, the observer's assumption of a 1D, plane-parallel configuration as applied in semi-analytical models overestimates the shielding (black arrow, $A_{V,1D\text{-}slab}$) and thus underestimates the amount of H I in the cloud's centre. Under realistic conditions, radiation (red arrows) dissociating H_2 might be able to propagate through low-density voids (white areas) towards the centre of the cloud. This results in a rather low actual visual extinction, $A_{V,real}$. Right: Volume rendering of MC1-HD at 2 Myr showing the highly complex and filamentary structure of the dense gas (reddish). Large low-density voids (bluish) are recognizable through which radiation can propagate into the cloud thus increasing the H I fraction. This resembles the simplified picture shown in the left-hand panel.

One could interpret this effect as either increasing the radiation strength or decreasing the *effective* shielding of UV radiation in the cloud. In the semi-analytical models of Sternberg et al. (2014), Bialy & Sternberg (2016), and Bialy et al. (2017b), this is parametrized by their parameter αG. There, α indicates the radiation strength and G the shielding factor including H_2 self-shielding and dust attenuation; the lower G, the better the radiation is shielded. We speculate that large density voids can easily reduce the (self-)shielding of the surrounding dust and H_2 by an order of magnitude. The corresponding *increase* of G (and thus αG) in the models of the aforementioned authors would thus increase their $N_{H\,I}$ (see e.g. equation 40 and fig. 9 in Sternberg et al. 2014) by a factor of a few. This behaviour is thus in general accordance with the interpretation presented here and would bring their upper limits for $N_{H\,I}$ closer to the values reported here.

5.2 The observational perspective

As observational works on Galactic (e.g. Savage et al. 1977; Kavars et al. 2003, 2005; Klaassen et al. 2005; Gillmon et al. 2006; Krčo et al. 2008; Barriault et al. 2010; Krčo & Goldsmith 2010; Lee et al. 2012, 2015; Stanimirović et al. 2014; Burkhart et al. 2015; Imara & Burkhart 2016) and extragalactic scales (e.g. Wong & Blitz 2002; Browning et al. 2003; Blitz & Rosolowsky 2004, 2006; Bigiel et al. 2008; Wong et al. 2009) tend to find upper limits of $N_{H\,I}$ around 10^{21} cm^{-2} (equivalent to 8 M$_\odot$ pc^{-2}), it could be argued that the high H I column densities in MCs and the associated underestimation of H I by a factor of 3–10 suggested in this work are rather exceptional. Given the various reasons discussed in this work, however, we argue that the underestimation is indeed rather common. In addition, carefully investigating the observational literature we find further evidence that an upper $N_{H\,I}$ threshold of 10^{21} cm^{-2} could be artificial:

(i) Some of the aforementioned H I measurements are emission observations, and for some of them the contribution of the cold H I might be omitted; e.g. Lee et al. (2012) report two absorption features in their H I spectra of the Perseus MCs, which they do not consider in their $N_{H\,I}$ calculations. Furthermore, other highly resolved H I emission observations indeed find H I column densities well above 10^{21} cm^{-2} (Motte et al. 2014; Bihr et al. 2015; Syed et al. 2020; Wang et al. 2020b).

(ii) Observations in emission often assume optically thin emission. However, as large parts of MCs have optical depths well above 1 (Fig. 10), optical depth corrections are crucial to infer the correct H I column densities. This is in line with observations of W43 by Bihr et al. (2015), who find an increase in H I by a factor of $\sim$2 compared to the optically thin assumption (Motte et al. 2014). Similar correction factors were found for indirect H I measurements of off-Galactic plane gas (Fukui et al. 2015) and the Perseus MC (Okamoto et al. 2017), although for the latter Lee et al. (2012, 2015) argue for a correction of only $\sim$20 per cent. For the THOR survey, Wang et al. (2020a) determined the correction factor of $\sim$31 per cent. All these correction factors are lower limit as optical depth estimates have an upper limit set by the observational noise (Bihr et al. 2015, and our equation 11).

(iii) Finally, extragalactic observations typically have spatial resolutions of a few 100 pc. Hence, they average over clouds and the surrounding diffuse ISM, which can lower the maximum value of $N_{H\,I}$ significantly. We show this by calculating the total H I column density (now including again H I with $T > 100$ K) for our simulations using pixels with a side length of 31.5, 8, and 2 pc (black, magenta, and green dots, respectively, in the top panel of Fig. 1). Overall, this reduces the maximum $N_{H\,I}$ values to $\sim 10^{22}$ cm^{-2} for 2 pc pixels and even further to $\sim 10^{21}$ cm^{-2} for 31.5 pc pixels, i.e. in parts by more than one order of magnitude compared to the maximum around a few 10^{22} cm^{-2} for our highest resolution.

To summarize, we suggest that (i) H I column densities well beyond 10^{21} cm^{-2} ($\sim$8 M$_\odot$ pc^{-2}) are significantly more common in MCs than thought (Fig. 5) and (ii) also the entire mass of cold H I gas in clouds could be a factor of a few higher ($\gtrsim$3) than thought (Fig. 12). Vice versa, we argue that (iii) 1D PDR models might underestimate the amount of cold H I in a typical H I–H_2 transition layer as those are in general not plane-parallel or spherically symmetric objects and (iv) H I observations might underestimate the H I content in MCs by a factor of a few due to the various systematic observational biases discussed in this work.

6 CONCLUSIONS

In this work, we present the first fully self-consistent synthetic H I 21 cm observations including self-absorption (HISA) of MCs simulated within the SILCC-Zoom project. The synthetic observations are based on 3D MHD simulations, including a non-equilibrium H I–H_2 chemistry, detailed radiative transfer calculations, and realistic observational effects like noise and a limited spectral and spatial resolution adapted to actual observations. In addition, we analyse in detail the actual content of cold H I in the simulated clouds and compare it with the results obtained from the synthetic HISA observations. We summarize our main results in the following:

(i) We show that HISA observations, which assume a fixed H I temperature, typically tend to underestimate column densities of cold H I, $N_{\rm H\,I}$, and the total cold H I mass in MCs by a factor of 3–10. This effect is particular pronounced towards the central regions, which frequently reach column densities up to $\gtrsim 10^{22}$ cm^{-2}. It occurs for MCs under various conditions, e.g. with and without dynamically important magnetic fields.

(ii) We show that the underestimation of $N_{\rm H\,I}$ in HISA observations can be attributed to the following two effects. (i) The large temperature variations of cold H I (∼10 K up to 100 K) make a reliable determination of $T_{\rm HISA}$ not possible. This leads to the fact that the real $T_{\rm HISA}$ and thus $N_{\rm H\,I}$ are underestimated and that velocity channels have to be omitted for the calculation of $N_{\rm H\,I}$. (ii) Observational noise and the emission of warm H I in the foreground either reduce the inferred optical depth or – as before – cause individual velocity channels to be omitted for the calculation of $N_{\rm H\,I}$. This effect is particularly pronounced in regions of high optical depth. In combination, both effects (i + ii) can lead to an artificial upper limit in observation of $N_{\rm H\,I,obs}$ around 10^{21} cm^{-2}.

(iii) We suggest a method to correct for the aforementioned omission of high-optical depth channels. This correction reduces the underestimation of the H I mass budget by a factor of 1.5–2.

(iv) We find that clouds typically have H I optical depths around 1–10. This implies that the optically thin H I assumption is usually not suitable and that optical depth corrections are essential when calculating $N_{\rm H\,I}$ from H I observations.

(v) We show that the high H I column densities ($\gtrsim 10^{22}$ cm^{-2}) can (in parts) be attributed to the occurrence of up to 10 individual H I–H_2 transitions along the LOS. This emphasizes the necessity of Gaussian decomposition algorithms to fully analyse the individual components that constitute the HISA spectra.

(vi) Also for individual H I–H_2 transitions, $N_{\rm H\,I}$ frequently exceeds a value of 10^{21} cm^{-2}, thus challenging 1D, semi-analytical models. This can be attributed to non-equilibrium chemistry effects, which are included in our models, and to the fact that H I–H_2 transitions usually do not have a 1D geometry, i.e. to the fractal structure of MCs.

(vii) We demonstrate that $N_{\rm H\,I}$ PDFs obtained from HISA observations with a fixed temperature assumption should be considered with great caution concerning both the position of the peak and the width. Due to the underestimation of H I, the observed PDFs appear to lack the high-$N_{\rm H\,I}$ end, which in reality seems to be characterized by a power law.

(viii) Finally, we show that the cold H I gas in MCs is moderately supersonic with Mach numbers of up to a few. The corresponding non-thermal velocity dispersion can be determined via HISA observations with an accuracy of a factor of ∼2.

To summarize, our results indicate that measuring the H I content in MCs via HISA observations is a challenging task and that the amount of cold H I in MCs could be a factor of 3–10 higher than previously thought.

ACKNOWLEDGEMENTS

We thank the referee for a very thorough and helpful report, which helped to improved the clarity of this work. DS likes to thank H. Dénes for helpful discussions. DS and SW acknowledge support of the Bonn-Cologne Graduate School, which is funded through the German Excellence Initiative as well as funding by the Deutsche Forschungsgemeinschaft (DFG) via the Collaborative Research Center SFB 956 'Conditions and Impact of Star Formation' (subprojects C5 and C6). SW acknowledges support via the ERC starting grant number 679852 'RADFEEDBACK'. HB and JDS acknowledge support from the European Research Council under the Horizon 2020 Framework Program via the ERC Consolidator Grant CSF-648505. HB and JS acknowledge support from the DFG in the Collaborative Research Center SFB 881 – Project-ID 138713538 – 'The Milky Way System' (subproject B1). PG acknowledges funding from the European Research Council under ERC-CoG grant CRAGSMAN-646955 and the ERC Synergy Grant ECOGAL (grant 855130). RW acknowledges support by project 19-15008S of the Czech Science Foundation and by the institutional project RVO:67985815. The FLASH code used in this work was partly developed by the Flash Center for Computational Science at the University of Chicago. The authors acknowledge the Leibniz-Rechenzentrum Garching for providing computing time on SuperMUC via the project 'pr94du' as well as the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu).

DATA AVAILABILITY

The data underlying this article can be shared for selected scientific purposes after request to the corresponding author.

REFERENCES

Acreman D. M., Douglas K. A., Dobbs C. L., Brunt C. M., 2010, MNRAS, 406, 1460
Acreman D. M., Dobbs C. L., Brunt C. M., Douglas K. A., 2012, MNRAS, 422, 241
Auddy S., Basu S., Kudoh T., 2018, MNRAS, 474, 400
Barriault L. et al., 2010, MNRAS, 406, 2713
Beck R., Wielebinski R., 2013, in Oswalt T. D., Gilmore G., eds, Planets, Stars and Stellar Systems, Vol. 5. Springer Sci.+Bus. Media, Dordrecht, p. 641
Bellomi E., Godard B., Hennebelle P., Valdivia V., Pineau des Forêts G., Lesaffre P., Pérault M., 2020, A&A, 643, A36
Beuther H. et al., 2020, A&A, 638, A44
Bialy S., Sternberg A., 2016, ApJ, 822, 83
Bialy S., Bihr S., Beuther H., Henning T., Sternberg A., 2017a, ApJ, 835, 126
Bialy S., Burkhart B., Sternberg A., 2017b, ApJ, 843, 92
Bigiel F., Leroy A., Walter F., Brinks E., de Blok W. J. G., Madore B., Thornley M. D., 2008, AJ, 136, 2846
Bihr S. et al., 2015, A&A, 580, A112
Blitz L., Rosolowsky E., 2004, ApJ, 612, L29
Blitz L., Rosolowsky E., 2006, ApJ, 650, 933
Bolatto A. D., Wolfire M., Leroy A. K., 2013, ARA&A, 51, 207
Browning M. K., Tumlinson J., Shull J. M., 2003, ApJ, 582, 810
Burkhart B., Lazarian A., 2012, ApJ, 755, L19
Burkhart B., Lee M.-Y., Murray C. E., Stanimirović S., 2015, ApJ, 811, L28
Clark P. C., Glover S. C. O., Klessen R. S., 2012a, MNRAS, 420, 745
Clark P. C., Glover S. C. O., Klessen R. S., Bonnell I. A., 2012b, MNRAS, 424, 2599

Clark P. C., Glover S. C. O., Ragan S. E., Duarte-Cabral A., 2019, MNRAS, 486, 4622
Clarke S. D., Whitworth A. P., Spowage R. L., Duarte-Cabral A., Suri S. T., Jaffa S. E., Walch S., Clark P. C., 2018, MNRAS, 479, 1722
Dénes H., McClure-Griffiths N. M., Dickey J. M., Dawson J. R., Murray C. E., 2018, MNRAS, 479, 1465
Dobbs C. L. et al., 2014, in Beuther H., Klessen R. S., Dullemond C. P., Henning T., eds, Protostars and Planets VI. Univ. Arizona Press, Tucson, AZ, p. 3
Douglas K. A., Acreman D. M., Dobbs C. L., Brunt C. M., 2010, MNRAS, 407, 405
Draine B. T., 1978, ApJS, 36, 595
Dubey A. et al., 2008, in Pogorelov N. V., Audit E., Zank G. P., eds, ASP Conf. Ser. Vol. 385, Numerical Modeling of Space Plasma Flows. Astron. Soc. Pac., San Francisco, p. 145
Dullemond C. P., Juhasz A., Pohl A., Sereshti F., Shetty R., Peters T., Commercon B., Flock M., 2012, Astrophysics Source Code Library, record ascl:1202.015
Feldt C., 1993, A&A, 276, 531
Field G. B., 1958, Proc. IRE, 46, 240
Field G. B., 1959, ApJ, 129, 551
Fryxell B. et al., 2000, ApJS, 131, 273
Fukui Y. et al., 2014, ApJ, 796, 59
Fukui Y., Torii K., Onishi T., Yamamoto H., Okamoto R., Hayakawa T., Tachihara K., Sano H., 2015, ApJ, 798, 6
Fukui Y., Hayakawa T., Inoue T., Torii K., Okamoto R., Tachihara K., Onishi T., Hayashi K., 2018, ApJ, 860, 33
Gatto A. et al., 2017, MNRAS, 466, 1903
Gibson S. J., Taylor A. R., Higgs L. A., Dewdney P. E., 2000, ApJ, 540, 851
Gibson S. J., Taylor A. R., Higgs L. A., Brunt C. M., Dewdney P. E., 2005, ApJ, 626, 195
Gillmon K., Shull J. M., Tumlinson J., Danforth C., 2006, ApJ, 636, 891
Girichidis P., Konstandin L., Whitworth A. P., Klessen R. S., 2014, ApJ, 781, 91
Girichidis P. et al., 2016, MNRAS, 456, 3432
Glover S. C. O., Clark P. C., 2012, MNRAS, 421, 116
Glover S. C. O., Mac Low M.-M., 2007a, ApJS, 169, 239
Glover S. C. O., Mac Low M.-M., 2007b, ApJ, 659, 1317
Glover S. C. O., Federrath C., Mac Low M.-M., Klessen R. S., 2010, MNRAS, 404, 2
Gnedin N. Y., Tassis K., Kravtsov A. V., 2009, ApJ, 697, 55
Goldsmith P. F., Li D., 2005, ApJ, 622, 938
Goodman A. A., Heiles C., 1994, ApJ, 424, 208
Gould R. J., 1994, ApJ, 423, 522
Habing H. J., 1968, Bull. Astron. Inst. Neth., 19, 421
Haid S., Walch S., Seifried D., Wünsch R., Dinnbier F., Naab T., 2019, MNRAS, 482, 4062
Heeschen D. S., 1954, AJ, 59, 324
Heeschen D. S., 1955, ApJ, 121, 569
Heiles C., Troland T. H., 2003, ApJS, 145, 329
Heiner J. S., Vázquez-Semadeni E., Ballesteros-Paredes J., 2015, MNRAS, 452, 1353
Hennebelle P., Audit E., Miville-Deschênes M. A., 2007, A&A, 465, 445
Imara N., Burkhart B., 2016, ApJ, 829, 102
Joshi P. R., Walch S., Seifried D., Glover S. C. O., Clarke S. D., Weis M., 2019, MNRAS, 484, 1735
Kainulainen J., Beuther H., Henning T., Plume R., 2009, A&A, 508, L35
Kalberla P. M. W., Burton W. B., Hartmann D., Arnal E. M., Bajaja E., Morras R., Pöppel W. G. L., 2005, A&A, 440, 775
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
Kim C.-G., Ostriker E. C., Kim W.-T., 2014, ApJ, 786, 64
Klaassen P. D., Plume R., Gibson S. J., Taylor A. R., Brunt C. M., 2005, ApJ, 631, 1001
Knapp G. R., 1974, AJ, 79, 527
Krčo M., Goldsmith P. F., 2010, ApJ, 724, 1402
Krčo M., Goldsmith P. F., Brown R. L., Li D., 2008, ApJ, 689, 276
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
Lee M.-Y. et al., 2012, ApJ, 748, 75
Lee M.-Y., Stanimirović S., Murray C. E., Heiles C., Miller J., 2015, ApJ, 809, 56
Li D., Goldsmith P. F., 2003, ApJ, 585, 823
Liszt H., 2001, A&A, 371, 698
McClure-Griffiths N. M., Dickey J. M., Gaensler B. M., Green A. J., Haverkorn M., 2006, ApJ, 652, 1339
McKee C. F., Krumholz M. R., 2010, ApJ, 709, 308
Mackey J., Walch S., Seifried D., Glover S. C. O., Wünsch R., Aharonian F., 2019, MNRAS, 486, 1094
Mac Low M.-M., Glover S. C. O., 2012, ApJ, 746, 135
Montgomery A. S., Bates B., Davies R. D., 1995, MNRAS, 273, 449
Motte F. et al., 2014, A&A, 571, A32
Murray C. E. et al., 2015, ApJ, 804, 89
Murray C. E., Stanimirović S., Kim C.-G., Ostriker E. C., Lindner R. R., Heiles C., Dickey J. M., Babler B., 2017, ApJ, 837, 55
Murray C. E., Peek J. E. G., Lee M.-Y., Stanimirović S., 2018, ApJ, 862, 131
Nelson R. P., Langer W. D., 1999, ApJ, 524, 923
Nguyen H., Dawson J. R., Lee M.-Y., Murray C. E., Stanimirović S., Heiles C., Miville-Deschênes M. A., Petzler A., 2019, ApJ, 880, 141
Nickerson S., Teyssier R., Rosdahl J., 2019, MNRAS, 484, 1238
Okamoto R., Yamamoto H., Tachihara K., Hayakawa T., Hayashi K., Fukui Y., 2017, ApJ, 838, 132
Rachford B. L. et al., 2009, ApJS, 180, 125
Rebolledo D. et al., 2017, MNRAS, 472, 1685
Riegel K. W., Crutcher R. M., 1972, A&A, 18, 55
Röllig M. et al., 2007, A&A, 467, 187
Savage B. D., Bohlin R. C., Drake J. F., Budich W., 1977, ApJ, 216, 291
Schneider N. et al., 2015, A&A, 575, A79
Schruba A. et al., 2011, AJ, 142, 37
Seifried D. et al., 2017, MNRAS, 472, 4797
Seifried D., Walch S., Haid S., Girichidis P., Naab T., 2018, ApJ, 855, 81
Seifried D., Walch S., Reissl S., Ibáñez-Mejía J. C., 2019, MNRAS, 482, 2697
Seifried D., Haid S., Walch S., Borchert E. M. A., Bisbas T. G., 2020a, MNRAS, 492, 1465
Seifried D., Walch S., Weis M., Reissl S., Soler J. D., Klessen R. S., Joshi P. R., 2020b, MNRAS, 497, 4196
Smith R. J. et al., 2020, MNRAS, 492, 1594
Soler J. D., 2019, A&A, 629, A96
Soler J. D. et al., 2019, A&A, 622, A166
Stanimirović S., Murray C. E., Lee M.-Y., Heiles C., Miller J., 2014, ApJ, 793, 132
Sternberg A., 1988, ApJ, 332, 400
Sternberg A., Le Petit F., Roueff E., Le Bourlot J., 2014, ApJ, 790, 10
Syed J. et al., 2020, A&A, 642, A68
Valdivia V., Hennebelle P., Gérin M., Lesaffre P., 2016, A&A, 587, A76
van der Werf P. P., Goss W. M., Vanden Bout P. A., 1988, A&A, 201, 311
van Dishoeck E. F., Black J. H., 1986, ApJS, 62, 109
Veltchev T. V., Girichidis P., Donkov S., Schneider N., Stanchev O., Marinkova L., Seifried D., Klessen R. S., 2019, MNRAS, 489, 788
Walch S. et al., 2015, MNRAS, 454, 238
Wang Y. et al., 2020a, A&A, 634, A83
Wang Y. et al., 2020b, A&A, 634, A139
Welty D. E., Xue R., Wong T., 2012, ApJ, 745, 173
Wilson T. L., Rohlfs K., Hüttemeister S., 2013, Tools of Radio Astronomy. Springer-Verlag, Berlin
Wolfire M. G., Hollenbach D., McKee C. F., 2010, ApJ, 716, 1191
Wong T., Blitz L., 2002, ApJ, 569, 157
Wong T. et al., 2009, ApJ, 696, 370
Wouthuysen S. A., 1952, AJ, 57, 31
Wünsch R., Walch S., Dinnbier F., Whitworth A., 2018, MNRAS, 475, 3393

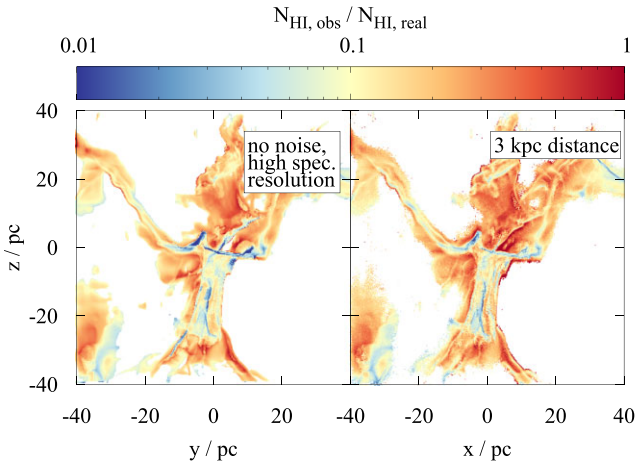

Figure A1. Map of the ratio of the observed and actual H I column density for the same snapshot as in Fig. 4, now for the case of no noise and a high spectral resolution (left) and an assumed distance of 3 kpc including observational effects like noise (right) inferred from the HISA observation assuming $T_{\rm HISA}$ = 40 K. The poor match is also found for the ideal observations (left), supporting the conclusion that strong temperature variations are (in parts) the cause for it.

APPENDIX A: SUPPLEMENTARY FIGURES

In the left-hand panel of Fig. A1, we show the ratio of $N_{\rm H\,I,obs}$ to $N_{\rm H\,I,real}$ for MC1-HD at 2 Myr, where $N_{\rm H\,I,obs}$ is calculated from the noiseless, high-spectral resolution (200 m s^{-1}) maps obtained directly from RADMC-3D using $T_{\rm HISA}$ = 40 K. We find a comparable poor match as for the case when observational effects are included. In the right-hand panel, we show the results obtained assuming a distance of 3 kpc (again including observational effects). Little differences are found compared to a distance of 150 pc (compare with the bottom middle panel of Fig. 4). This result also holds for the other snapshots considered in this work.

In Fig. A2, we show the inferred H I column density for MC1-HD at 2 Myr now including the correction in optically thick regions. The results are discussed in Section 3.3.1.

In Fig. A3, we show the effect of post-processing the chemical state of one of our simulations, i.e. pushing it towards a chemical equilibrium state. This is done exemplarily for MC1-HD taking two snapshots at $t_{\rm evol}$ = 2 and 3 Myr considered in this work. We stop the MHD evolution at these times, i.e. freeze the total density, velocity, etc., and only evolve the chemistry for additional 100 Myr. The chemical post-processing time (measured from $t_{\rm evol}$ onwards) is denoted as $t_{\rm chem}$. The H I content quickly drops to 40–50 per cent of the actual (non-equilibrium) H I content (at $t_{\rm chem}$ = 0), which indicates that equilibrium models generally underestimate the amount of H I in MCs.

In Fig. A4, we show the $\log(n_{\rm H\,I})$ profile for five selected pixels for MC1 at 2 Myr along the x-direction used for the analysis in Section 4.2. The profiles show a large variability concerning the number of peaks, their widths and positions. Some of the peaks are discarded as they do not have either the required minimum prominence or the minimum peak height.

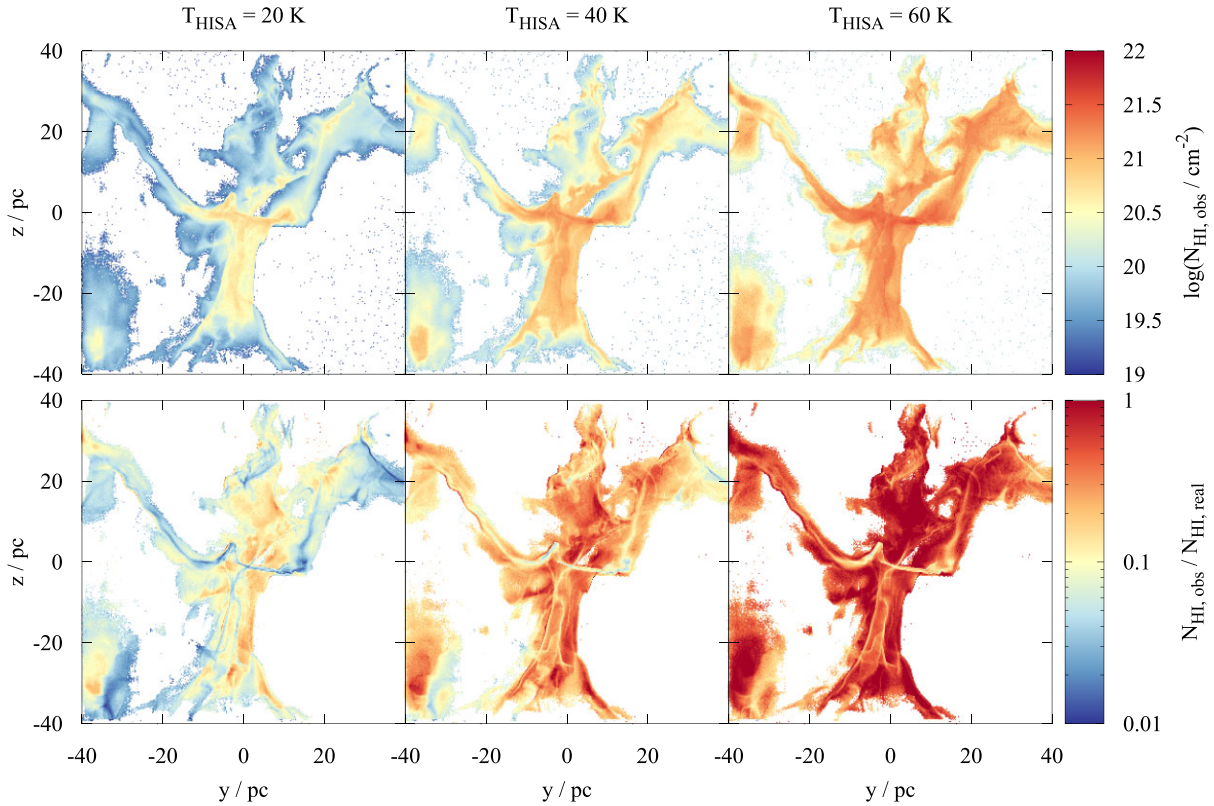

Figure A2. Same as in Fig. 4, but now with the correction in optically thick regions where for channels, where equation (6) does not yield any result, we assume an optical depth given by $\tau_{\rm H\,I,noise}$ (equation 11). Overall, the match in the moderately dense gas is improved, whereas in the most densest parts $N_{\rm H\,I,real}$ is still significantly underestimated.

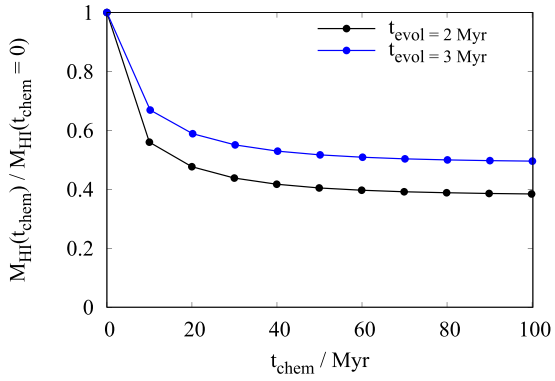

Figure A3. Evolution of the H I content of MC1-HD when post-processing the chemical abundances for a time of $t_{\rm chem}$ relative to the actual (non-equilibrium) H I content (at $t_{\rm chem} = 0$). Assuming chemical equilibrium would reduce the H I content by a factor of 2–2.5.

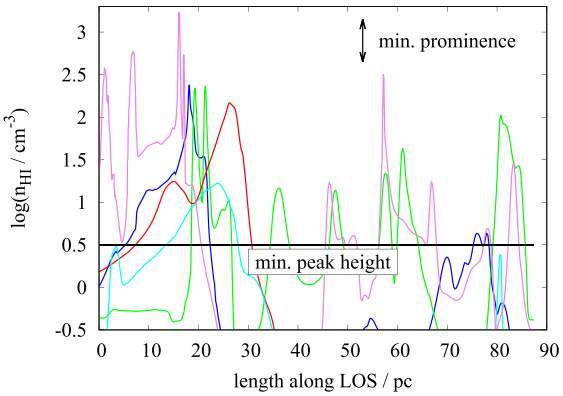

Figure A4. Profiles of $\log(n_{\rm H\,I})$ for five selected pixels for MC1 at 2 Myr along the x-direction. The profiles show a large variability. In addition, we show the minimum prominence and minimum peak value (both in black), which a peak must have to be considered and not discarded.

APPENDIX B: AN APPROXIMATION FOR THE OPTICAL DEPTH

In order to estimate the optical depth of the H I gas in our simulations, we repeat one radiative transfer calculation for MC1-HD, however, now setting the background temperature to 0 K. Hence, the observed emission purely stems from the H I gas in the zoom-in region.

As noted in equation (7), but now written down for a single velocity channel, the H I column density is given by

$$\mathrm{d}N_{\rm H\,I} = 1.8224 \times 10^{18}\mathrm{cm}^{-2}\,\frac{T_{\rm s}}{1\,\mathrm{K}}\tau(v)\frac{\mathrm{d}v}{1\,\mathrm{km\,s}^{-1}}. \tag{B1}$$

Next, we consider the expression

$$T_{\rm rad} = \frac{h\nu_{\rm H\,I}}{k_{\rm B}}\left(f(T_{\rm s}) - f(T_{\rm bg})\right)\left(1 - e^{-\tau}\right), \tag{B2}$$

for the observed brightness temperature $T_{\rm rad}$. Here, $T_{\rm bg}$ is the background temperature, h is the Planck constant, $k_{\rm B}$ is the Boltzmann constant, $\nu_{\rm H\,I} = 1420$ MHz is the frequency of the H I 21 cm line, and

$$f(T) = \frac{1}{\exp\left(\frac{h\nu_{\rm H\,I}}{k_{\rm B}T}\right) - 1}. \tag{B3}$$

Considering that $h\nu_{\rm H\,I}/k_{\rm B} = 0.068\ \mathrm{K} \ll T_{\rm s}$ and $f(T_{\rm bg}) \ll f(T_{\rm s})$, and inserting equation (B2) into equation (B1), yields

$$\mathrm{d}N_{\rm H\,I} \simeq 1.8224 \times 10^{18}\mathrm{cm}^{-2}\,\frac{\tau}{1 - e^{-\tau}}\frac{T_{\rm rad}}{1\,\mathrm{K}}\frac{\mathrm{d}v}{1\,\mathrm{km\,s}^{-1}}. \tag{B4}$$

We now integrate over all velocity channels using a definition of a $T_{\rm rad}$-weighted, channel-averaged approximation of the optical depth

$$N_{\rm H\,I} = 1.8224 \times 10^{18}\mathrm{cm}^{-2}\,\langle\tau\rangle\int\frac{T_{\rm rad}}{1\,\mathrm{K}}\frac{\mathrm{d}v}{1\,\mathrm{km\,s}^{-1}}. \tag{B5}$$

Here, we have defined

$$\langle\tau\rangle = \frac{\int\frac{\tau}{1-e^{-\tau}}T_{\rm rad}\mathrm{d}v}{\int T_{\rm rad}\mathrm{d}v}. \tag{B6}$$

The interpretation of $\langle\tau\rangle$ as an approximation for a $T_{\rm rad}$-weighted, channel-averaged optical depth can be understood, when considering the fact that $\frac{\tau}{1-e^{-\tau}} \to \tau$ with $\tau \to \infty$. For $\tau = 1$, the expression $\frac{\tau}{1-e^{-\tau}}$ is only $\sim$50 per cent larger than τ, and for $\tau = 2$ only $\sim$15 per cent. For optically thin regions ($\tau < 1$), the approximation is not applicable. However, as in Section 3.3 we are mainly interested in high-optical depth regions, we consider our definition of $\langle\tau\rangle$ as a reasonable approximation for the typical optical depth of H I in our simulations.

Next, using the integrated intensity from the radiative transfer calculations without any background radiation field and the real H I column density from the simulation data, $N_{\rm H\,I,real}$, we can now calculate $\langle\tau\rangle$ via

$$\langle\tau\rangle = \frac{N_{\rm H\,I,real}}{1.8224 \times 10^{18}\mathrm{cm}^{-2}\int\frac{T_{\rm rad}}{1\,\mathrm{K}}\frac{\mathrm{d}v}{1\,\mathrm{km\,s}^{-1}}}, \tag{B7}$$

where the denominator is describing the H I column density obtained from H I emission under the assumption of optically thin emission.

This paper has been typeset from a TEX/LATEX file prepared by the author.