Properties of cold and warm H I gas phases derived from a Gaussian decomposition of HI4PI data

P. M. W. Kalberla and U. Haud

1 Argelander-Institut für Astronomie, Auf dem Hügel 71, 53121 Bonn, Germany
e-mail: pkalberla@astro.uni-bonn.de
2 Tartu Observatory, University of Tartu, 61602 Tõravere, Tartumaa, Estonia

Received 2 April 2018 / Accepted 2 June 2018

ABSTRACT

Context. A large fraction of the interstellar medium can be characterized as a multiphase medium. The neutral hydrogen gas is bistable with cold and warm neutral medium (CNM and WNM respectively) but there is evidence for an additional phase at intermediate temperatures, the lukewarm neutral medium (LNM) that is thermally unstable.

Aims. We use all sky data from the HI4PI survey to separate these neutral H I phases with the aim to determine their distribution and phase fractions \(f\) in the local interstellar medium.

Methods. HI4PI observations, gridded on an \(nside = 1024\) HEALPix grid, were decomposed into Gaussian components. From the frequency distribution of the velocity dispersions we infer three separate linewidth regimes. Accordingly we extract the H I line emission corresponding to the CNM, LNM, and WNM. We generated all-sky maps of these phases in the local H I gas with \(-8 < \theta_{LSR} < 8 \text{ km s}^{-1}\).

Results. Each of the H I phases shows distinct structures on all scales. The LNM never exists as a single phase but contributes on average 41% of the H I. The CNM is prominent only for 22% of the sky, contributes there on average 34% but locally up to 60% of the H I and is associated with dust at temperatures \(T_{\text{dust}} \sim 18.6\) K. Embedded cold filaments show a clear anti-correlation between CNM and LNM. Also the smoothly distributed WNM is anti-correlated with the CNM. It contributes for the rest of the sky 39% with dust associated at temperatures \(T_{\text{dust}} \sim 19.4\) K.

Conclusions. The CNM in filaments exists on small scales. Here the observed anti-correlation between LNM and CNM implies that both, filaments and the surrounding more extended LNM, must have a common origin.

Key words. ISM: general – ISM: structure – dust, extinction – ISM: clouds

1. Introduction

The interstellar medium, notably the H I, exists as a multiphase medium. Field et al. (1969) pointed out that most of the H I should exist in two stable phases in pressure equilibrium, the cold neutral medium (CNM) at excitation temperatures \(T < 300\) K, and the WNM at \(T \lesssim 10^4\) K. But soon Salpeter (1976), reviewing the formation and destruction of interstellar dust grains, objected and argued that there must be a third unstable phase at intermediate temperatures. This phase was dubbed by him the luke-warm neutral medium (LNM).

McKee & Ostriker (1977) extended the two-phase model of Field et al. (1969) and considered supernova (SN) explosions in an in-homogeneous environment. Their model of the interstellar medium (ISM) has three phases, including the ionized medium, and is dominated by individual SN explosions. The interior of SN remnants is filled by the hot ionized medium (HIM). Blast waves sweep up the gas inside the bubbles and pile it up into the shells. As soon as this gas is shocked, it starts to cool. It recombines rapidly and forms the CNM.

Subsequently the thermal equilibrium gas temperatures of the diffuse interstellar medium were calculated by Wolfire et al. (1995) an updated by Wolfire et al. (2003). These investigations have shown that there can be conditions where the H I gas is thermally unstable. However, the conclusion was that the local ISM is probably only marginally in the regime in which there must be more than two bistable H I phases.

From observations there were early indications for substantial amounts of unstable H I gas. Among others, significant contributions from unstable H I gas were observed by Dickey et al. (1977), Mebold et al. (1982), Kalberla et al. (1985), Spitzer & Fitzpatrick (1995), Fitzpatrick & Spitzer (1997). However, it was Heiles (2001) who initiated a vivid debate with the statement “Large amounts of thermally unstable gas are not allowed in theoretical models of the global interstellar medium”. From the Arecibo absorption line survey toward 79 sources there was evidence that 60% of the gas is WNM. At least 48% of this WNM turned out to be unstable (Heiles & Troland 2003a). In other words, 30% of the gas belongs to the LNM. More recently, Roy et al. (2013) obtained for 33 compact extragalactic radio sources deep high-velocity resolution absorption line data. They found that at least 28% of the gas must have temperatures in the thermally unstable range. A more recent study H I absorption in the direction of 57 background radio sources with the Very Large Array (Murray et al. 2015) has found an even smaller mass fraction of thermally unstable HI, about 20%.

Subsequent theoretical investigations have shown that considerable amounts of unstable gas can be induced by turbulence. Gazol et al. (2001) conclude that about 50% of the turbulent gas mass has temperatures that are characteristic for the LNM.
Audit & Hennebelle (2005) find also large fractions of thermally unstable gas. This fraction increases with the amplitude of the turbulent forcing. The thermally unstable gas tends to be organized in filamentary structures. Similar, de Avillez & Breitschwerdt (2005) find that up to 49% of the mass belongs to the thermally unstable regime. In general, these results show all that turbulence is playing a key role for phase transition in the H I gas (Saury et al. 2014).

The debate about the multiphase composition of the H I gas was summarized by Vázquez-Semadeni (2012) and amounts to the question whether classical discrete-phase models need to be replaced by a “phase continuum”. For details we refer to this excellent review.

In this paper we want to study phase dependencies in the local H I gas. We used high resolution observations of the all-sky H I brightness temperature distribution with large single dish radio telescopes. The data are decomposed into Gaussian components. We find three distinct linewidth-regimes that can be considered to represent the CNM, LNM, and WNM. Section 2 describes the observations, the data reduction, the Gaussian decomposition and selection criteria applied by us. Using Gaussian parameters we model separate H I distributions for the CNM, LNM, and WNM. These distributions are presented in Sect. 3 and we discuss the column density distributions in detail. The Sky appears to contain regions that are dominated by the CNM. In Sect. 4 we separate these regions and discuss phase dependent distributions for column densities and associated dust temperatures. Cold filamentary gas stands out and is discussed in more detail in Sect. 5. The interrelations between H I phases are shown in Sect. 6. Our results are discussed in Sect. 7, the summary is in Sect. 8.

2. Observations and data reduction

2.1. The HI4PI 21-cm line survey

For the northern hemisphere we use the Effelsberg–Bonn H I Survey (EBHIS) from observations with the 100-m Effelsberg radio telescope (Winkel et al. 2016a,b) and for the southern hemisphere the Galactic All Sky Survey (GASS), observed with the 64 m Parkes telescope (McClure-Griffiths et al. 2009; Kalberla et al. 2010; Kalberla & Haud 2015). Both surveys, EBHIS with the first data release (Winkel et al. 2016b) and GASS with its final data release (Kalberla & Haud 2015), were gridded in position to a common nside = 1024 HEALPix database (Górski et al. 2005). After calibration, this was the first step in merging both surveys and initially the original velocity vectors, different for both surveys, were kept. We used this intermediate data product, both surveys with their genuine spatial resolution and original velocity vectors, for our Gaussian decomposition as described in Sect. 2.2. When merging both surveys by HI4PI Collaboration (2016), the EBHIS beam shape was adapted to the GASS resolution. In addition the GASS spectra were smoothed and adapted to the EBHIS velocity grid. In our analysis, to decompose the spectra as accurate as possible into Gaussian components, we stayed as close as possible to the original database but used positions on the common HEALPix grid. To calculate maps for this publication we use Gaussian components to generate profiles on the HEALPix grid. Next the data are smoothed to a common effective resolution of 30′, afterward maps are generated.

Using an nside = 1024 HEALPix database implies an effective angular resolution of $\Theta_{\text{pix}} = 3\text{.}44^\circ$ for that grid (Górski et al. 2005). The EBHIS data were gridded to an effective beam sized with $FWHM = 10\text{.}8^\circ$, and the GASS to $FWHM = 16\text{.}2^\circ$.

This means that our database is over-sampled, neighboring positions are not independent from each other. Our choice of the nside = 1024 HEALPix database is motivated by the aim to enable an easy comparison between all-sky H I data and published Planck maps.

The Gaussian analysis enables us to separate different phases according to the linewidth-regimes derived in Sect. 2.3. We generate separate maps for the CNM, LNM, and WNM. The EBHIS and GASS overlap for declinations $-5^\circ < \delta < 1^\circ$. When selecting Gaussian components we use a border line between both surveys at $\delta = -2^\circ$. To avoid discontinuities in maps, caused by different beam sizes, we used a linear interpolation between both surveys for $-4^\circ < \delta < 0^\circ$.

2.2. Gaussian analysis

After generating the HEALPix database, we decomposed all brightness temperature spectra $T_b(v_i)$ into Gaussian components

$$T_b(v_i) = \sum_{j=1}^{N} T_{bc,j} \exp \left[ -\frac{(v_i - v_{c,j})^2}{2\sigma_j^2} \right] ,$$

(1)

where $T_{bc,j}$, $v_{c,j}$, and $\sigma_j$ are the adjustable parameters of the Gaussian component $j$, the sum is taken over $N$ components, describing the given profile and $v_i$ is the central velocity of the spectrometer channel $i$. For the decomposition we used mostly the same approach, which was described by Haud (2000) and applied earlier to the Leiden/Argentine/Bonn (LAB) data (Kalberla et al. 2005). In general, this is a rather classical Gaussian decomposition, but with two important additions, introduced for reducing the ambiguities inherent to the decomposition procedure. First of all, our decomposition algorithm does not treat each H I profile independently, but assumes that every observed profile shares some similarities with those in the neighboring sky positions, as expected for an over-sampled database. In addition, besides adding components into the decomposition, our algorithm also analyzes the results to find the possibilities for removing or merging some Gaussians without reducing too much the accuracy of the representation of the original profile with decomposition. For the acceptable decomposition we required that the RMS of the weighted deviations of the Gaussian model from the observed profile had to be no more than 1.006 times the weighted noise level of the emission-free Gaussian model from the observed profile had to be no more than 1.006 times the weighted noise level of the emission-free baseline regions of this profile. This condition corresponds to $|\ln(x^2/N_{\text{noise}})| < 0.015$, but we must keep in mind that here both, the estimates of the standard deviations of the data points and the fitted models are based on the same data set. The value of the multiplier has been chosen so that the averages of the noise levels and the RMS deviations of the models over all HEALPix profiles of the survey are equal (in practice, for the final decomposition the difference of these averages was less than 0.09% of their mean).

For the decomposition of the HI4PI, we also introduced some modifications to our old algorithm. First of all, the accuracy of the measured brightness temperatures is not the same for all profile channels. According to the radiometer equation the noise level depends on the signal strength and different corrections applied during the processing of the observed spectral dumps introduce additional uncertainties. During the decomposition, all this was considered through the weights assigned to all values of $T_b(v_i)$ of the profile in each HEALPix pixel. For the GASS part of the HI4PI, these weights were calculated from the RMS deviations of the spectral dumps from the corresponding average.
profile in each pixel, as described in Kalberla & Haud (2015). For EBHIS, the $T_{\text{sys}}(v_i)$ together with the RFI flags was used as a proxy for the RMS in the individual spectral dumps contributing to the HEALPix pixel.

As the velocity resolution of the EBHIS data is the lowest among the decomposed surveys, it revealed a problem with the original decomposition algorithm: near the steepest gradients in the observed profiles the decomposition results often contained groups of very high ($|T_{\text{bc}}| > |T_{\text{bc}}(v_i) + T_{\text{bc}}(v_{i+1})/2|$), but narrow ($\sigma < \Delta v$, where $\Delta v \equiv v_{i+1} - v_i$ is the channel separation of the survey) Gaussians, which centers were located between the consecutive spectrometer channels $v_i$ and $v_{i+1}$. The obtained decompositions fitted well the observed profiles at $v_i$ and $v_{i+1}$, but with strong fluctuations between the measured data points. To suppress such a behavior, we used the penalty function approach. During the decomposition, we calculated at $v_{i,j}$ of all Gaussians the deviations of the model from the cubic spline representation of the observed profile and added the weighted squares of these deviations to the RMS of the fit. The weights of these penalty points were calculated as $w_j = \frac{\Delta v_{\text{min}}}{v_{i,j} - v_{i,j}} W$, where $\Delta v_{\text{min}}$ is the positive velocity difference between the $v_{i,j}$ and the nearest $v_i$. $W$ was obtained from the cubic spline fit of the weights of the observed profile at velocities $v_{i,j}$.

After finding this problem in the decomposition of the EBHIS, we performed a special search in the results of the GASS decomposition and found similar narrow Gaussians also there. The only difference was that in the GASS such components appeared in general one at a time and therefore did not cause so obvious oscillations of the resulting model profiles as in EBHIS. Nevertheless, in the final decomposition, the modification for suppressing such Gaussians was applied to both the EBHIS and GASS data.

Considerable change in the decomposition algorithm was made available by the increased power of the computers. In the LAB survey, we used the decomposition results of one of the neighboring profiles of any given profile as an initial estimate of the Gaussian parameters, and we only started the decomposition of the first profile with one roughly estimated component at the highest maximum of the profile. In the HI4PI study, the decomposition was divided into two stages. In the first run, we decomposed all profiles in the HEALPix database independently of their neighbors, starting with one Gaussian at the brightest tip of the profile. In the second stage, we compared the results for neighboring profiles. To accomplish this, we used the decomposition obtained so far for each profile as an initial approximation for all eight nearest neighbors of this profile, and checked whether this led to better decomposition of these neighboring profiles. If the decomposition of a neighbor was improved, the new result was used as an initial solution for the eight neighbors of this profile, and so on. The process was repeated until no more improvements were found (on the order of 500 runs through the full database).

During this process, we estimated the goodness of the obtained fits using two criteria. First of all, as described above, we did not accept the decompositions for which the RMS of the weighted deviations of the Gaussian model from the observed profile exceeded more than 1.006 times the weighted noise level of the emission-free baseline regions of this profile. If for some profiles the RMS criterion was satisfied for more than one trial decomposition with different numbers of Gaussians, we accepted the solution with the smallest number of the components. In the case of acceptable decompositions with equal numbers of the Gaussians, we chose the decomposition with the smallest RMS as the best. The decomposition process is described in more detail in Sects. 3.1 and 3.2 of Haud (2000).

Finally, in the case of the LAB, we only used positive Gaussians ($T_{\text{bc}} > 0$ K) for decomposition and the parts of the profiles with $T_{\text{bc}}(v_i) < 0$ K were not considered at all. With the HI4PI, we also fitted negative Gaussians to the regions of the profiles where the brightness temperature was on average below zero. For fitting the regions of the profiles, where $T_{\text{bc}}(v_i) > 0$ K, we still used only positive Gaussians and not a combination of positive and negative components. Therefore, only strong absorption or the baseline problems, where the $T_{\text{bc}}(v_i)$ of the profile drops below the continuum level, induced negative Gaussians. In general, absorption was modeled as a gap between positive Gaussians.

2.3. Defining different HI phases

The hypothesis that the H I component line shapes are Gaussian rests on the assumption that motions within any H I cloud have a random velocity distribution that can be described by a Gaussian function. This assumption is supported by the finding that the H I absorption features are often well represented by Gaussians in optical depth $\tau(v)$ (Mohan et al. 2004). The decomposition of the emission profiles is physically meaningful only when $\tau(v) \ll 1$. Nevertheless, both the opacity and brightness profiles of most sources are easily decomposed into Gaussians, so whether or not this model is physically correct, it works empirically and is convenient.

As a result, the Gaussian decomposition offers an opportunity to differentiate between components with different line widths (Haud & Kalberla 2007). In combination with absorption-line studies (Dickey et al. 2003; Heiles & Troland 2003a), components with broad or narrow line widths have been found to arise from the WNM and CNM, respectively (Dickey & Lockman 1990; Wolfire et al. 2003; Hennebelle & Falgarone 2012) and there could be substantial amounts of thermally unstable LNM as well (Heiles & Troland 2003a; Haud & Kalberla 2007; Hennebelle & Falgarone 2012; Roy et al. 2013; Saury et al. 2014).

Our decomposition of the full HEALPix database of the HI4PI profiles gave on average 8.472 Gaussians per profile, but the complexity of the profiles greatly varies between different sky regions and often the Gaussians in the decomposition are heavily blended with each other. This considerably complicates the physical interpretation of the results, as for such profiles the decomposition may be not unique. This means that several quite different solutions may approximate the observed profile almost equally well, and the decomposition provides no satisfactory means for choosing between these solutions, while others, equally good or even better ones, may not be found at all. As described in Sect. 2.2, we have tried to reduce this problem by keeping the number of the used Gaussians as low as possible and by considering during the decomposition of each profile the information about the neighboring profiles.

In this way we hope that for similar profiles the algorithm prefers similar decompositions, which may better represent the general properties of the H I gas than the set of more independent solutions. Nevertheless, these precautions cannot guarantee that our decomposition results mostly describe the properties of the underlying H I gas and are not dominated by the random factors inherent to the decomposition process. Therefore, we decided to compare first the results, obtained from independent observations (the LAB vs. HI4PI) with different decomposition
algorithms (as described in Sect. 2.2). We expect that the similarity of the results, obtainable from such studies indicate that the main properties of the results are determined by the nature of the galactic gas but in cases where uncertainties in the Gaussian decomposition are dominant the results may yield nearly incomparable outcomes.

It is clear that the blending and non-uniqueness problems are weaker for the simplest profiles and for the strongest Gaussians in the decompositions of the more complicated profiles. Therefore, as with the LAB (Haud & Kalberla 2007), we started the analysis of the HI4PI data with the simplest profiles in the database, which are located at relatively high galactic latitudes and followed the changes in the results caused by gradually adding more complicated profiles at lower latitudes (Fig. 1). With the HI4PI data we used for these tests only the Gaussians which belong to the central peaks of the observed profiles. We defined the central peak as the maximum of $T_b(v)$ closest to $v_{LSR} = 0$ km s$^{-1}$. We considered this peak to extend to the velocities where the $T_b(v)$, calculated from the decomposition results, drops below the noise level of the corresponding observed profile.

This approach had an additional advantage that it eliminated from the analysis nearly all noise Gaussians without applying any formal selection criteria (as Eqs. (3) and (4) in Haud & Kalberla 2007). Only a rather small contamination by RFI remained. In some cases such approach included also intermediate- (IVC) or even high-velocity clouds (HVC) into our sample, but as in the following we will apply also different velocity limits on the studied components, this is not a considerable problem. Moreover, we will use the number of the Gaussians in the central peak of the profiles, $N_G$, as an indicator of the profile complexity and study only the profiles with moderate complexity. As the inclusion of IVCs and HVCs as local gas increases $N_G$ for the corresponding pixels, such cases have higher probability to be excluded as too complex profiles. Figure 1 gives the sky distribution of the described complexity estimates for $N_G < 8$.

Using the LAB data, we have demonstrated (Haud & Kalberla 2007) that by considering only the simplest H$\text{I}$ profiles and relatively small LSR velocities ($-9 \leq v_L \leq 4$ km s$^{-1}$) it is possible to distinguish three groups of preferred line-widths, which are defined by fitting the sum of log-normal functions to the frequency distribution of the Gaussian widths. It is found that on the basis of the LAB profiles decomposed into one, two, or three low velocity Gaussians, the estimates for the mean line-widths of the CNM, LNM, and WNM gas are $4.9 \pm 0.1$, $12.0 \pm 0.3$, and $24.4 \pm 0.2$ km s$^{-1}$ (Fig. 4 in Haud & Kalberla 2007). In the same way, we obtained from the decomposition of the HI4PI survey the values $4.7 \pm 0.4$, $11.2 \pm 0.6$, and $24.2 \pm 0.4$ km s$^{-1}$ (Fig. 2). For CNM and WNM the results from LAB and HI4PI are rather close to each other. The agreement is slightly worse for the LNM, but this is understandable, as the frequency distribution of the widths of the LNM Gaussians is blended by both, CNM and WNM. This makes the separation of this gas phase more uncertain.

Besides the curves for CNM, LNM, and WNM Fig. 2 (also Fig. 4 in Haud & Kalberla 2007) contains a fourth curve, labeled here as XNM. At first sight, the nature of this component is not obvious and therefore we made a special study of the profiles, responsible for this component (see Appendix A). It turned out that the corresponding Gaussians are weak and in many cases they describe the wide non-Gaussian wings of the profiles, but sometimes (mostly in the EBHIS part of the HI4PI but in

![Fig. 1](image1.png)

**Fig. 1.** Sky positions of the 7 855 871 profiles with less than eight Gaussians in the central emission peak. The positions of the simplest profiles are marked by the reddest colors. Among the profiles with an equal number of Gaussians those with lower column density of the local H$\text{I}$ are plotted with redder color.

![Fig. 2](image2.png)

**Fig. 2.** Distribution of the widths of Gaussians for the profiles decomposed to one (upper panel), two (middle panel), or three (lower panel) components.
less than 3% of all positions) are caused by instrumental problems. We found from different models that these Gaussians represent 3.2%–5.3% of the total HI and in the following they are considered as belonging to WNM.

In the two lower panels of the Fig. 2 a similar, but even weaker enhancement is visible also on the left edge of the distribution. This is caused by rare and very narrow noise Gaussians, still remaining in the analyzed data sample. As their contribution is very small, we did not use an additional log-normal curve for describing this enhancement and in the following the corresponding Gaussians are considered as part of CNM.

As can be seen from Fig. 2, with increasing complexity of the profiles the relative number of CNM Gaussians increases and up to a certain complexity limit this permits better determination of the parameters of the CNM gas phase from the width distribution of the corresponding Gaussians. However, at considerably higher complexity levels \( N_G > 10 \) of the profiles the definition of the LNM phase becomes more and more questionable as the decompositions of the complex profiles usually contain many weak Gaussians with rather ill-defined parameters. Such components add noise to the distribution of the widths of the Gaussians, and the CNM, LNM, and WNM components of the width distribution start to merge into a relatively structureless curve, where the maximum corresponding to LNM is completely blended by CNM and WNM. Therefore, the separation of CNM, LNM, and WNM phases is best determined for the profiles with some moderate complexity.

To estimate the reasonable complexity limit, we analyzed the width distributions of the Gaussians from the profiles with \( 1 \leq N_G \leq 15 \) and found that the uncertainties of the parameters of the LNM phase started to grow more rapidly for \( N_G \geq 8 \). As for \( N_G \leq 3 \) we obtained for \( N_G \leq 7 \) the following mean line-widths of the CNM, LNM, and WNM gas: 3.6 ± 0.1, 9.6 ± 1.3, and 23.3 ± 0.6 km s\(^{-1}\). The distribution of the Gaussian widths for \( N_G \leq 7 \) together with the corresponding model curves are given in Fig. 3. The sky distribution of all profiles used for these estimates is in Fig. 1. Despite the fact that here we used a considerably simplified approach to the determination of these averages (we dropped the analysis of the reliability of the decompositions of the profiles with \( N_G \geq 4 \), as it was done by Haud & Kalberla (2007)), these estimates are comparable with the final results of Haud & Kalberla (2007; FWHM = 3.9 ± 0.6, 11.8 ± 0.5, and 24.1 ± 0.6 km s\(^{-1}\)).

As a result, we may recognize that Gaussian fits to observed HI emission or absorption profiles have been applied by a number of authors (e.g., Takakubo & van Woerden 1966; Mebold 1972; Mebold et al. 1982; Heiles & Troland 2003b; Nidever et al. 2008; Murray et al. 2015). We have used it with the LAB, GASS, and EBHIS data (Haud & Kalberla 2007; Kalberla & Haud 2015; Kalberla et al. 2016, 2017). We think that it is out of question that such fits are correct in a mathematical sense. Whether the derived physical parameters are meaningful (here the FWHM is a different question and by selecting example profiles one can always find good or bad examples. Even for perfect observations it may happen that the derived parameters do not represent a correct model for the gas distribution. We cannot exclude such cases but we believe that at least on average the Gaussian fits deliver a reasonable model for the HI distribution. Simplicities in decompositions of the LAB, GASS, and EBHIS profiles support this and regardless of possible ambiguities we believe that the results give some information on the properties of the CNM, LNM, and WNM gas phases. This agrees with the conclusion, pointed out by Murray et al. (2017) that Gaussian fits to synthetic HI lines are able to recover gas structures with excellent completeness at high Galactic latitude, and this completeness declines with decreasing latitude due to strong velocity-blending of spectral lines. Moreover, as mentioned by Roy et al. (2013), while there are certainly possible drawbacks to using Gaussian components to model HI 21-cm spectra, there is no obvious alternative route to determining physical conditions in the ISM.

2.4. Separating HI phases

So far we have modeled the average frequency distribution of the widths of the Gaussians from profiles with \( N_G \leq 7 \) in the velocity range, preselected in Haud & Kalberla (2007), but when plotting the frequency distribution of these Gaussians in the component central velocity – line-width plane (Fig. 4), we can see that at different velocities the distribution of the widths is different. In our trial to separate three HI phases, we decided to take into account also the velocity dependence of the Gaussian widths in different gas phases. For this we assumed that in each narrow velocity range the distribution of the widths of the Gaussians can be modeled with the sum of four log-normal functions, as in Figs. 2 and 3, but the parameters of the log-normal model distributions may change with the velocity.

To estimate the reliability of the obtained models, we decided to construct many different models and to analyze their average results together with corresponding deviations. Altogether six families of the models were constructed, each with many sub-models. All these models were based on the same Gaussian decomposition of the HI4PI HI profiles but differed from each other by the approaches, used for the modeling of the distribution of the Gaussian parameters in their central velocity – line-width plane. For the following we define the model family as a group of models, based on some more or less general assumptions or fitting strategies, related to all models of this group. The sub-models in the family differ then from each other by the usage of different velocity resolutions, profile complexities and initial approximations for the non-linear model fitting. For example, in all families separate models were made for the complexity limits \( N_G \leq 3, 4, 5, 6, \) or 7 local Gaussians per observed HI profile. A brief overview of our model families is given in Table 1. The differences between the model families are defined in the second and the third columns of this table and the last column describes the velocity ranges, used in the sub-models of the corresponding family.

In the first two model families, we assumed that the changes in the parameters of the four log-normal distributions with the velocity can be described with the polynomial functions, which
Table 1. Families of the models for the width distributions of the Gaussians at different velocities.

| Family | Minimized | Model restrictions | $|v_c| \leq$ |
|--------|-----------|--------------------|-----------|
| 1      | $\sum(N_O - N_M)^2$ | $g_{ij} = \sum_{k=0}^3 p_k \arctan(p_5 + p_6 v_c)$ | 12.5, 15, 25, 30, 50 |
| 2      | $\sum(N_O - N_M)^2$ | $g_{ij} = \sum_{k=0}^3 p_k \arctan(p_5 + p_6 v_c)$ | 12.5, 15, 25, 30, 50, 99 |
| 3      | $\sum(N_O - N_M)^2$ | – | 50 |
| 4      | $\sum(N_O - N_M)^2$ | FWHM$_{CNM}$ < FWHM$_{LNM}$ < FWHM$_{WNM}$ < FWHM$_{XNM}$ | 50 |
| 5      | $\sum(\text{lg}(N_O + 1) - \text{lg}(N_M + 1))^2$ | – | 50 |
| 6      | $\sum(\text{lg}(N_O + 1) - \text{lg}(N_M + 1))^2$ | FWHM$_{CNM}$ < FWHM$_{LNM}$ < FWHM$_{WNM}$ < FWHM$_{XNM}$ | 50 |

Notes. $g_{ij}$ is the $i$th parameter ($1 \leq i \leq 3$) of the log-normal model distribution of the gas phase $j$ ($1 \leq j \leq 4$ for CNM to XNM). Parameters $p_k$ are adjusted for obtaining the two-dimensional model of the frequency distribution of the Gaussians in the component central velocity – line-width plane. $N_O$ and $N_M$ are the actual and model numbers of the Gaussians in different counting bins. Each family has sub-models for $N_O \leq 3, 4, 5, 6, 7$.

Fig. 4. Frequency distribution of the 42,488,690 Gaussians, used for the modeling of the line-width distributions of the gas phases, in the component central velocity – line-width plane. Blue points mark the local maxima in the frequency distribution of the Gaussian line widths at different velocities and the magenta lines represent initial estimates of the centers of four log-normal model distributions of the Gaussian widths in different gas phases. As abscissa we have used the quantities of the velocity distribution of the used Gaussians. The numbers mark the values of the deciles.

contain adjustable parameters. In the first family we used the sixth order polynomial of velocity and in the second family the fourth order polynomial of the function $\arctan(p_5 + p_6 v_c)$, where $p_5$ and $p_6$ were also adjustable model parameters. For sub-models we analyzed profiles with different complexity $N_G$ and their Gaussians from different velocity ranges (as specified in Table 1). We divided all Gaussians in the analyzed velocity ranges into 200 groups, each containing the same number of components with similar velocities. In this way the usage of different velocity ranges gave us also different velocity resolutions of the corresponding models. By adjusting the parameters of the polynomials, describing the velocity dependencies of three parameters in each of four log-normal distribution, we represented the complete distribution of the Gaussians (as in Fig. 4) with some model distribution (one possible example in Fig. 5) by minimizing the RMS deviations between observed and model distributions in all velocity and line-width bins.

To define the initial approximation for all models in these two families, we searched in each velocity bin of Fig. 4 for the line widths, corresponding to the local frequency maxima and recorded their velocities, Gaussian widths and the breadths, $\Delta$, of the lg(FWHM) regions, in which the corresponding point was a global maximum. The results are given as blue points in Fig. 4. The sizes of the points are proportional to $\Delta$. The found maxima were classified by hand into four groups, corresponding to four log-normal functions, fitting the frequency distribution of the Gaussians. The initial dependencies of the centers of the log-normal functions on the velocity were then defined as fits of the polynomials through the corresponding points of the local maxima in the observed frequency distribution. For this fitting, the values of $\Delta$ were used as weights. The results are given in Fig. 4 by magenta lines. Initial approximations of the heights of the log-normal distributions were determined by fitting the family specific polynomials through the Gaussian frequencies at the points calculated from the functions found for the centers of the log-normal distributions.

From the preceding experience on the modeling of the Gaussian frequency distributions in the fixed velocity range (Figs. 2 and 3), it was clear that the most uncertain property of the log-normal distributions is their width. Therefore, the initial approximations for the widths of the log-normal distributions were determined as constants from the velocity range $-3.39 < v_c < -1.26$ km s$^{-1}$, where the LN is most clearly separated from CNM and WNM (Fig. 4). When starting the final model fitting, these constants were used for all velocities (the coefficients of the velocity dependent terms of the corresponding functions were all taken to be equal to zero), but during the fitting of the models the coefficients of all terms were allowed to vary freely.
For the fitting of the models, we used the multidimensional downhill simplex method and for the first two families of the models we generated up to five subfamilies with different algorithms for the generation of the initial simplex. When the model converged to some solution, we generated a new large simplex around the obtained result and repeated the fitting until the repetitions did not yield any considerable improvements in the results. Due to a large number of the free fitting parameters, such repetitions generally found deeper minima than those obtained in the first run of the fitting.

All together we obtained in both model families up to 150 different models. In each obtained model we searched the velocity dependent values of the line widths at which the modeled frequency of the Gaussians in two neighboring (on the order of increasing line widths) gas phases were equal. The obtained values of $FWHM_{CL}(v_c)$, $FWHM_{WX}(v_c)$, and $FWHM_{WX}(v_c)$ for the borders between CNM – LNM, LNM – WNM, and WNM – XNM components were converted into corresponding Doppler temperatures $T_{D,CL}(v_c)$, $T_{D,LW}(v_c)$, and $T_{D,WX}(v_c)$, where

$$T_D(v_c) = 21.8618 \cdot FWHM^2(v_c),$$

and averaged over all models of one family. The standard deviations of the results from different models were considered as error estimates of the results.

An example of the model from the second family for $|v_c| < 100$ km s$^{-1}$ and $N_G \leq 7$ is given in Fig. 5. Here the isolines represent the observed distribution from Fig. 4 and the background colors give the model approximation of these lines. Blue dashed curves give the locations of the peaks of the log-normal model curves (the magenta lines from Fig. 4 after the model fitting) and the solid blue lines are the line widths, which separate different gas phases.

The upper left corner of Fig. 5 demonstrates that at higher velocities the behavior of the models was sometimes so wild that it was impossible to define the separating line widths. At the same time, at low velocities the polynomial models were even more stable than we expected in advance. Therefore, we decided to take a step further and to construct additional model families, in which we dropped the requirement of the polynomial dependence of the parameters of the log-normal distributions on the velocity, but analyzed the width distributions only for $|v_c| < 50$ km s$^{-1}$. In the last four families of the models we turned back to the fitting of the width distributions in the narrow velocity ranges, as it was described in Sect. 2.3, but for each fit we defined an independent velocity region for the fitting. The velocity ranges were constructed randomly so that they contained at least 19 000 and at most 210 000 Gaussians and all Gaussians had equal probabilities to be included in some velocity range. For each model the upper limit of the profile complexity was also randomly chosen so that $N_G \leq 3$, 4, 5, 6, or 7.

In each of these four additional families we used a different approach to the fitting of the models and generated 60 000 independent sub-models for different velocity ranges and profile complexities. For all these models the initial approximations were taken from the final models of the second polynomial family. For families three and four we minimized the parameter $\sum(N_0 - N_M)^2$, but for families five and six we used $\sum(\log(N_0 + 1) - \log(N_M + 1))^2$, where $N_0$ and $N_M$ are the numbers of the Gaussians in different width bins from our Gaussian decomposition and from the corresponding model, respectively. In this way the families three and four were mostly sensitive to the ridges of the frequency distributions, but five and six gave more weight to the wings of the distributions. In families three and five we did not apply any additional restrictions to the model fitting, but in families four and six we also demanded that during the fitting process the FWHM, corresponding to the peaks of the model distributions, had to follow the relation $FWHM_{CNM} < FWHM_{LNM} < FWHM_{WNM} < FWHM_{XNM}$. It was accomplished through the penalty function so that each violation of this rule multiplied the sum of the squares of the residuals by 2.0. Some examples of the models from the last four families are given in Fig. 6.

After constructing the models, the interpretation of the results was similar to that of the polynomial families one and two. We calculated the line widths at which the modeled frequencies of the Gaussians in the neighboring gas phases were equal.

Fig. 6. Three examples of the models from the last four families. In the upper panel is the model from the velocity range, where the LNM is most clearly visible in Fig. 4. The middle panel corresponds to the velocities, close to the limit $|v_c| \leq 8$ km s$^{-1}$, used for the discussions in the present paper and the lower panel illustrates the situation at extreme velocities $|v_c| = 25$ km s$^{-1}$, used for Fig. 7.
The obtained values of the FWHM were converted into the Doppler temperatures, averaged in regular velocity bins over all models of each family and the corresponding standard deviations were calculated as estimates of the result uncertainties.

The comparison of the results indicated that up to about \( |v_c| = 15 \text{ km s}^{-1} \) the different model families were in good mutual agreement. At higher velocities the deviations started to increase (especially at negative velocities for \( T_{D,CL} \) – the Doppler temperature, which separates CNM from the LNM), but remained acceptable up to about \( |v_c| = 25 \text{ km s}^{-1} \). At even higher velocities the deviations grow rapidly, which makes the results extremely uncertain. This is rather understandable. We planned to study the local interstellar medium and to test how far our approach may be viable. In Fig. 4 the bands of the CNM and WNM are well visible up to \( |v_c| \approx 15 \text{ km s}^{-1} \). The blue dots defining the LNM are considerably more scattered, but still the presence of this phase is felt. Beyond \( |v_c| \approx 15 \text{ km s}^{-1} \) the bands of CNM and WNM turn upward. CNM at negative and WNM at positive velocities becomes also less obvious and the definition of the LNM becomes rather doubtful. At \( |v_c| > 25 \text{ km s}^{-1} \) the identification of the LNM with the intermediate- and high-velocity clouds (the strong frequency enhancements at about \( \text{lg}(FWHM) \approx 3.35 \)) near the left and right edges of the figure, see Haud (2008) is already more than questionable. These results are also understandable from the comparison of the example models, given in Fig. 6. At low velocities the CNM, LNM, and WNM phases are well defined in the distribution of the Gaussian widths, but at higher velocities their identification becomes more and more doubtful. To avoid possible misidentifications we eventually decided to restrict our analysis to \( |v_c| < 8 \text{ km s}^{-1} \), see Sect. 3.2.

The results for \( |v_c| \leq 25 \text{ km s}^{-1} \) from all model families are summarized in Fig. 7. The values for \( T_{D,CL} \) are given in blue, those for \( T_{D,LW} \) are in green and red corresponds to \( T_{D,WX} \). The error-bars illustrate the standard deviations of the results in the different model families. The dark thick curves are the polynomial fits (using corresponding standard deviations for computing the weights) to all the presented results and the lighter thick curves illustrate the median uncertainties (for \( |v_c| \leq 25 \text{ km s}^{-1} \)) of the dark curves.

The error bars correspond to the standard deviations of the sub-models in each family. The polynomial fits to the results from all families are plotted with thick dark lines and the thick lighter lines give the corresponding median uncertainties in the velocity range \( |v_c| \leq 25 \text{ km s}^{-1} \). Two vertical black lines indicate the velocity range \( |v_c| \leq 8 \text{ km s}^{-1} \), used for the discussions.

3. Derived properties of discrete \( \text{H} I \) phases

So far, when deriving criteria to distinguish Gaussians belonging to different phases, we used only profiles with simple structures, corresponding to decompositions with \( N_G \leq 7 \) components. For a meaningful analysis of the \( \text{H} I \) distribution on the sky we need to generalize our results and assume that the derived criteria do not depend significantly on the number of clouds along the line of sight or the complexity of the profiles, represented by \( N_G \).

We sort Gaussian components according to their center velocities and Doppler temperatures, using the velocity dependent thresholds from Fig. 7, and group the components for separate phases. These different groups of Gaussians are then used to generate individual synthetic profiles \( T_{SP} \) for the phases \( P = \text{CNM}, \text{LNM}, \) and \( \text{WNM} \) with a velocity grid of \( \delta v_{LSR} = 1 \text{ km s}^{-1} \) on an inside \( = 1024 \) HEALPix grid in position. Thus we replace the observed \( \text{H} I \) distribution by three separate distributions to model the CNM, LNM, and WNM. As mentioned earlier, the WNM includes the contributions from the XNM. The sum of these phase dependent distributions fits to the observed brightness distribution \( T_b \); we only omit Gaussians representing obvious instrumental problems (Kalberla & Haud 2015). When generating maps we smooth these brightness temperatures to an effective beam-size of 30' \( \text{FWHM} \).

To measure, how much of the \( \text{H} I \) gas belongs to a particular phase we define phase fractions \( f_P \)

\[
f_P = \frac{\int_{v_L}^{v_H} T_{bP}(v_{LSR}) \delta v_{LSR} dv_{LSR}}{\int_{v_L}^{v_H} T_{b}(v_{LSR}) \delta v_{LSR} dv_{LSR}}
\]

for \( P = \text{CNM}, \text{LNM}, \) and \( \text{WNM} \).

Phase fractions are in general velocity dependent but for simplicity we drop this notation in the following. We will see later that phase fractions depend significantly on environmental effects.

3.1. Velocity dependencies

Velocity dependencies of derived single channel phase fractions are shown in Fig. 8. We display a comparison of the average all-sky phase fractions for \( -8 < v_{LSR} < 8 \text{ km s}^{-1} \) (thick lines) for the individual phases. Velocity dependencies are obvious, in particular \( f_{\text{CNM}} \) and \( f_{\text{WNM}} \) show systematic fluctuations.
Figure 8 includes also estimates for the uncertainties of the derived average phase fractions (thin lines). For all of the phases we used median standard deviations for the thresholds between the linewidth regimes (see Fig. 7) to derive upper and lower limits for the phase fractions. The uncertainties are significant, up to 50%, in particular for the LNM since this phase is affected by uncertainties in the thresholds for both, the CNM and the WNM phase.

3.2. All-sky maps for CNM, LNM, and WNM

In this section we show all-sky distributions of separate phases to give a first impression of morphological differences between the different phases. We use a plate carrée projection\(^1\).

In Fig. 9 we use a single velocity channel at \(v_{\text{LSR}} = 0 \text{ km s}^{-1}\) to show the separate components. At the top we display the brightness temperature contributions for CNM (left), LNM (middle), and WNM (right). At the bottom we show with the same sequence the corresponding phase fractions \(f_{\text{CNM}}, f_{\text{LNM}}, \) and \(f_{\text{WNM}}\) according to Eq. (3).

These all-sky maps display a surprising wealth of large and small scale structures. For the CNM and the LNM there are dominant filamentary features but even the \(f_{\text{WNM}}\) map shows remarkable structures, either with high or low \(f_{\text{WNM}}\) fractions. The WNM maps are barely compatible with a diffuse medium, the usual description we find in textbooks. In fact, in case of ongoing thermal instabilities we expect some anti-correlations between CNM on one hand and LNM or WNM on the other side. Such anti-correlations will be discussed in Sects. 4 and 6.

Figure 10 gives an alternative representation of the all-sky column density distribution for the velocity range \(-8 < v_{\text{LSR}} < 8 \text{ km s}^{-1}\). With the same sequence of the phases as in Fig. 9 we display at the top \(\log(N_{\text{H}}/[\text{cm}^{-2}])\). At the bottom we show the corresponding phase fractions. Interestingly, phase fractions in Figs. 9 and 10 (bottom) show in general very similar structures. This result can be explained by the velocity distribution of dominant local \(\text{H}_1\) structures (Kalberla et al. 2016; Sect. 5.13) that can be fit well with a Gaussian of \(\text{FWHM}\) of 16.8 \(\text{km s}^{-1}\), centered at \(v_{\text{LSR}} = 0 \text{ km s}^{-1}\). The local \(\text{H}_1\) gas is dominated by low velocity structures in this range.

The top panels of Figs. 9 and 10 for the WNM (right hand side) show enhancements of the \(\text{H}_1\) emission in the Galactic plane with a \(\sin(2\ell)\) modulation. A similar but less obvious effect is visible for the LNM (middle panels). These enhancements are caused by differential Galactic rotation (Mebold 1972) and imply for the Galactic plane that some of the observed gas must originate from large distances. At these positions our basic assumption, that we consider only local \(\text{H}_1\) gas, is violated. Surprisingly, when calculating phase fractions (lower panels of Figs. 9 and 10) these contamination are less obvious. Apparently the contaminations cancel mostly when calculating phase fractions. Anyhow, our caveat is that phase fractions at low Galactic latitudes are in general less reliable, this includes also larger uncertainties from the Gaussian analysis at low latitudes.

The separation of individual \(\text{H}_1\) phases in Sect. 2.4 leads to significant velocity dependencies of the \(T_{\text{B}}\) thresholds, as shown in Fig. 7. These velocity dependencies are relatively weak for the velocity range \(-8 < v_{\text{LSR}} < 8 \text{ km s}^{-1}\) but increase significantly at higher velocities. Since this velocity range covers most of the local \(\text{H}_1\) gas features, as presented in Figs. 9 and 10, we decided to restrict our analysis to this velocity range.

3.3. Average column density distributions

The column density distributions for different phases can be derived in two ways. Either by selecting Gaussian components according to the velocity dependent thresholds derived in Sect. 2.4 and generating then maps for a \(\text{FWHM}\) of 30’. This results in the spatial distributions displayed in Figs. 9 and 10. Alternatively we may derive the column density distributions directly from the Gaussian database, using appropriate phases and center velocities. Figure 11 shows the results for the velocity range \(-8 < v_{\text{LSR}} < 8 \text{ km s}^{-1}\). We display data for the CNM, LNM, and WNM (left to right) restricted to latitudes \(|b| > 20^\circ\) but also for all-sky.

---

\(^1\) On request HEALPix side = 1024 FITS files that can be used to generate other projections are available from the first author.
Fig. 10. Top panel: All-sky display of the observed H\textsc{i} column density distributions, integrating over a velocity range $-8 < v_{\text{LSR}} < 8$ km s\textsuperscript{-1} and distinguishing contributions from the CNM (left), LNM (middle), and WNM (right). The color scale is logarithmic, $\lg(N_{\text{H}}/\text{cm}^{-2})$. Bottom panel: H\textsc{i} phase fractions $f_{\text{CNM}}$ (left), $f_{\text{LNM}}$ (middle), and $f_{\text{WNM}}$ (right) for the same data. Galactic coordinates are used, the Galactic Center is in the middle.

Fig. 11. Distribution functions for column densities $\lg(N_{\text{H}}/\text{cm}^{-2})$ of the CNM (left), LNM (middle), and WNM (right), derived from Gaussian components in the velocity range $-8 < v_{\text{LSR}} < 8$ km s\textsuperscript{-1}. The light and dark blue colors distinguish distributions for all Sky and $|b| > 20^\circ$. The green lines show in each case least square Gaussian fits, resulting in log-normal distributions. For the CNM (left panel) we include the distributions of major filamentary USM structures as observed by Kalberla et al. (2016; their Fig. 12). On the right hand side the XNM components are also included for comparison.

Figure 11 shows in all cases column density distributions that are close to log-normal, the best fit Gaussians are included for comparison. For the CNM we plot also the distributions for filamentary structures derived from unsharp masked (USM) maps (Kalberla et al. 2016; their Fig. 12). Such features emphasize the H\textsc{i} distribution at high spatial frequencies. This gas is cold and aligned with polarized dust emission (Clark et al. 2014; Kalberla et al. 2016). The USM structures have low column densities and represent only a subset of the CNM emission. These features have barely direct counterparts in the list of Gaussian components. In the right hand panel of Fig. 11 we show also the H\textsc{i} distribution for the XNM, the fourth phase from Figs. 2 to 7 that we consider as part of the WNM.

3.4. Column densities depending on phase fractions

Here we derive dependencies between column densities and phase fractions according to Eq. (3). We use the velocity range $-8 < v_{\text{LSR}} < 8$ km s\textsuperscript{-1}. Figure 12 displays separately for each of the gas phases 2D density distributions for column densities and phase fractions. On the top the data are restricted to high latitudes $|b| > 20^\circ$, below we plot all-sky data.

High latitude data show the cleanest relations between column densities and phase fractions. The all-sky distributions are more complex due to some additional structures originating from regions close to the Galactic plane. However, in both cases the LNM stands out with a distribution that differs significantly from the CNM and the WNM. The LNM exists mostly with phase fractions $0.3 \leq f_{\text{LNM}} \leq 0.6$.

Figure 12 shows a clear dependency between column density distributions of individual H\textsc{i} phases and the corresponding phase fractions. This may imply that phases with distinct different temperatures are separated from each other. Figures 9 and 10 show in fact some evidence for regions that are dominated by one or the other phase.

3.5. Average phase fractions

For the velocity range $-8 < v_{\text{LSR}} < 8$ km s\textsuperscript{-1} we determine all-sky average phase fractions $f_{\text{CNM}} = 0.24 \pm 0.04$, $f_{\text{LNM}} = 0.44 \pm 0.02$, and $f_{\text{WNM}} = 0.32 \pm 0.05$. Here the listed uncertainties come only from the velocity dependent fluctuations of phase fractions for $-8 < v_{\text{LSR}} < 8$ km s\textsuperscript{-1}. Systematic errors may be much larger, see Sect. 2.4 and Fig. 8. Restricting our sample to high Galactic latitudes $|b| > 20^\circ$ we obtain $f_{\text{CNM}} = 0.25 \pm 0.02$, $f_{\text{LNM}} = 0.41 \pm 0.02$, and $f_{\text{WNM}} = 0.34 \pm 0.03$. As detailed in Sect. 3.2 we consider the derived parameters close to the Galactic plane as uncertain. Still the all-sky results agree within the errors with average phase fractions from high latitudes.
Fig. 12. 2D density distribution functions in the velocity range \(-8 < v_{LSR} < 8 \) km s\(^{-1}\) for CNM (left), LNM (middle), and WNM (right), showing the relations between phase fractions and column densities \(\log(N_H/\text{cm}^{-2})\) for each phase. Top panel: positions with \(|b| > 20^\circ\); bottom panel: all sky data.

4. CNM and WNM dominated regions

A comparison between Figs. 9 and 10 shows that the all-sky distribution of the CNM phase is most prominent at velocities close to zero. We use therefore the range \(-1 < v_{LSR} < 1 \) km s\(^{-1}\) to separate CNM dominated regions from the rest of the \(\text{H}\ I\) distribution by applying a mask. To define CNM dominated regions we first select only positions with \(f_{\text{CNM}} > 0.5\). This map of delta functions, representing cold spots, is smoothed by a Gaussian function with \(FWHM = 5^\circ\); the resulting map is truncated at 10% of the peak. Values above this threshold are set to one and values below to zero respectively. Approximately 22% of the total sky is this way masked as CNM dominated.

The application of this mask allows a separation of CNM dominated regions including a homogeneous area around the most prominent CNM structures. Figure 13 shows in the upper panel CNM dominated areas separated from the rest of the sky. The CNM structures are surrounded by LNM but there is little WNM associated with the CNM. Some borders of the CNM dominated regions appear to have enhanced WNM as a transition to the regions displayed on the lower panels of Fig. 13 that have been masked out from the CNM dominated areas. Our masking procedure is subjective, we have tried to work out an environment around the CNM with a smooth boundary at the transition between cold (CNM dominated) and warm (WNM dominated) \(\text{H}\ I\) regions.

4.1. Average phase fractions in CNM and WNM dominated regions

Calculating average phase fractions for \(|b| > 20^\circ\) and \(-8 < v_{LSR} < 8 \) km s\(^{-1}\) we come for CNM dominated regions to the striking result that these regions are actually not dominated by the CNM but by LNM: \(f_{\text{LNM}} = 0.40 \pm 0.03, f_{\text{CNM}} = 0.34 \pm 0.07,\) and \(f_{\text{WNM}} = 0.26 \pm 0.08\). The situation is similar for the rest of the sky, the WNM dominated part: \(f_{\text{LNM}} = 0.41 \pm 0.03, f_{\text{WNM}} = 0.39 \pm 0.01,\) and \(f_{\text{CNM}} = 0.20 \pm 0.03\). Thus, we obtain the unexpected result that in both cases the LNM is dominant. The average LNM phase fraction \(f_{\text{LNM}} = 0.41 \pm 0.02,\) determined in Sect. 3.5, appears to be characteristic for most of the sky if we consider the velocity range \(-8 < v_{LSR} < 8 \) km s\(^{-1}\).

4.2. Phase fractions around cold spots

To define the CNM dominated regions shown in Fig. 13 on top we selected a smoothing kernel \(K\) with a \(FWHM\) of \(5^\circ\). For a more general analysis, proposed by the referee, we extend here our calculations for the range \(0.5^\circ \leq K \leq 10^\circ\). In addition we use several velocity ranges for the determination of average phase fractions.

The top left panel of Fig. 14 shows a generalization of the results from the previous subsection. \(f_{\text{CNM}}\) decreases for increasing \(K\). At the same time we find an increase of \(f_{\text{LNM}}\). This implies that the CNM is on average localized; as we increase the area around the cold spots we observe on average less CNM but more LNM. \(f_{\text{CNM}}\) and \(f_{\text{LNM}}\) are anti-correlated; the LNM surrounds the cold spots. In addition we find that also \(f_{\text{WNM}}\) is gradually increasing with \(K\). The response of the WNM is much smoother than that of the LNM, implying than the WNM is distributed on average over large scales. This confirms essentially the visual impression we get from the right hand side of Fig. 13.

The middle and right hand side panels on top of Fig. 14 show even clearer correlations between CNM, LNM, and WNM for restricted velocity windows \(-4 \leq v_{LSR} \leq 4 \) km s\(^{-1}\) and \(-1 \leq v_{LSR} \leq 1 \) km s\(^{-1}\). \(f_{\text{CNM}}\) increases with decreasing velocity width but \(f_{\text{LNM}}\) remains almost unaffected. The implication is that the CNM in cold spots is not only spatially localized but also restricted in velocity space. The LNM is spatially extended, surrounding the CNM, and covers in addition a larger velocity spread. The WNM has a line-width of typically 24 km s\(^{-1}\). \(f_{\text{WNM}}\) contributes correspondingly less with decreasing velocity width. This effect apparently tends to balance the LNM phase fraction, we observe \(f_{\text{LNM}} \sim 0.4\) for different choices of the averaging line-width. In the lower panels of Fig. 14 we present our results for an all-sky analysis, confirming essentially the conclusions from high latitudes.

4.3. Dust temperatures associated with \(\text{H}\ I\) phases

The all-sky maps displayed in Figs. 9, 10, and 13 show for the CNM some common structures with filamentary \(\text{H}\ I\) gas and the \textit{Planck} HFI Sky Map at 353 GHz (see Kalberla et al. 2016; their Fig. 2); the CNM is associated with polarized dust emission. To
find out whether our masking for cold and warm H\textsc{I} gas results also in regions with physically distinct different dust temperatures we intend to correlate now the H\textsc{I} distribution with the dust temperature distribution.

Currently the most reliable dust temperatures are available from Planck Collaboration Int. XLVIII (2016). These authors used a tailored component-separation method, the so-called generalized needlet internal linear combination (GNILC) method, which uses spatial information (angular power spectra) to disentangle the Galactic dust emission and anisotropies in the cosmic infrared background.

To determine phase dependent relations between H\textsc{I} column densities and dust temperatures we use Gaussian components with center velocities in the range $-8 < v_{\text{LSR}} < 8$ km s$^{-1}$ at latitudes $|b| > 20^\circ$. Corresponding to the cases selected in Fig. 13 (CMN top and WNM bottom) we plot in Fig. 15 phase dependent 2D density distribution functions for dust temperatures and column densities for different H\textsc{I} phases.

The average dust temperature determined by Planck Collaboration Int. XLVIII (2016) is $T_{\text{dust}} = 19.41 \pm 1.54$ K. All H\textsc{I} phases in the CNM dominated part of the sky (Fig. 15 top) are characterized by low dust temperatures. H\textsc{I} column densities are high, in particular for the CNM and LNM. For the rest of the sky (bottom) we find warmer dust with temperatures around 19.4 K for the CNM and WNM, but for the LNM even some of the H\textsc{I} gas reaches dust temperatures close to 20 K.
4.4. Phase dominated positions for $f > 2/3$

Next we consider the question how far dust temperatures depend on H$_1$ phase fractions. As before, we consider CNM and WNM dominated regions on the sky (Fig. 13). We select there positions where each of the H$_1$ phases is dominant. To decide whether a phase is predominant we use the condition $f/(1 - f) > C$. Thus we demand that the amount of H$_1$ in an individual phase has to be a factor of $C > 1$ larger than the sum of the remaining phase fractions for the other phases. Figure 16 shows the distribution of dust temperatures derived for the case $C = 2$ or $f > 2/3$. This factor is somewhat arbitrary (though widely used to define a qualified majority) but our results do not depend significantly on $C$. For all $C > 1$ we get well defined dust temperature distributions that can be approximated by Gaussians. To the left of Fig. 16 we plotted the selected cold CNM dominated part, to the right the warm WNM dominated sky and in the middle the unconstrained sample. Each of the distributions is fitted by a Gaussian to obtain mean dust temperatures with dispersions that are listed in Table 2.

These histograms show for all phases that the CNM dominated part of the local H$_1$ distribution is cold with typical dust temperatures $T_{dust} \sim 18.7$ K. There is little WNM with $f_{WNM} \sim 2/3$. Opposite, the WNM dominated regions have $T_{dust} \sim 19.4$ K and there is only little CNM. The unconstrained H$_1$ distribution is dominated by the LNM with intermediate temperatures, $T_{dust} \sim 19$ K. The upper plots in Fig. 16 are for high Galactic latitudes, $|b| > 20^\circ$, the lower plots display the all-sky distribution. The reliability of the all-sky histograms is somewhat questionable but we include these to demonstrate that they share the same trend as the histograms for high latitudes.

Our results are consistent with the average $T_{dust} = 19.41 \pm 1.54$ K as determined by Planck Collaboration Int. XLVIII (2016) at high latitudes. However, Fig. 16 indicates a general dependency that dust temperatures decrease in regions that are dominated by the CNM. Each of the dust components can be fitted by Gaussians (see Table 2). For the CNM dominated sample the fits are well defined with low dispersions around one K. The distributions for the WNM dominated sky are broader and may contain several components with some scatter. The lower panel in Fig. 16 indicates that also some more scatter may exist at latitudes $|b| \lesssim 20^\circ$.

Figure 16 confirms the trends from Fig. 15 but we need to point out that in particular positions with $f_{CNM} \gtrsim 2/3$ that are associated with cold dust are almost exclusively found in the CNM dominated part of the sky. Opposite, positions with $f_{WNM} \gtrsim 2/3$ and warm dust are less frequent in this areas. For the rest of the sky we observe a dominance of the WNM, but surprisingly also there the LNM is significant. In all cases, for $f \gtrsim 2/3$, we find low-dust temperatures for CNM dominated regions and
opposite high temperatures in case of the WNM dominated sky, see Table 2. Thus, the dust temperatures are more smoothly distributed compared to the H I gas that may have significant local fluctuations in phase fractions.

5. Cold and filamentary H I

A significant fraction of the cold H I gas exists in filamentary structures that can be worked out for high resolution H I observations by unsharp masking (Kalberla et al. 2016). Many of these filaments are associated with dust and the most prominent filaments are aligned with the magnetic field. In this section we want to explore temperature dependencies of the different phases in presence of such filamentary structures.

Using the USM data from Kalberla et al. (2016), we search at each position for the strongest filament in the velocity range $-8 < v_{LSR} < 8$ km s$^{-1}$ and calculate its Doppler temperature $T_{D}$. We demand an USM peak temperature of at least one K. If such a feature is found we determine the velocity channel corresponding to its center velocity $v_{0}$. We use one additional channel on both sides and determine for this velocity window of $3$ km s$^{-1}$ the phase fractions $f_{CNM}$, $f_{LNM}$, and $f_{WNM}$. Our analysis is restricted to filamentary H I structures at high latitudes $|b| > 20^\circ$. Figure 17 displays on top the derived 2D density distributions for Doppler temperatures and phase fractions $f$.

Next we consider the question how the observed dust temperatures might depend on H I phase fractions for filamentary H I. For a comparison between dust temperatures and H I Doppler temperatures we determine at each position with a known USM Doppler temperature $T_{D}$ the corresponding GNILC dust temperature $T_{dust}$. In Fig. 17 we present on bottom 2D density distributions for dust temperatures associated with filamentary H I structures. Comparing Doppler temperatures at top of Fig. 17 with dust temperatures at bottom we find similar trends for different phase fractions: $f_{CNM}$, $f_{LNM}$, and $f_{WNM}$ (left to right).

These plots show that Doppler and dust temperatures in filaments decrease on average with increasing CNM fraction $f_{CNM}$. The vertical stripe at $f_{CNM} \sim 0$ in the left panels of Fig. 17 indicates that some of the USM filaments are not identified with proper Gaussian components. For an approximate threshold $f_{CNM} \leq 0.05$ we find that less than 4% of the USM filaments are affected. For the other data we fit $\lg(T_{D}/[K]) = 2.5222 \pm 0.0008 - (0.451 \pm 0.002) \cdot f_{CNM}$ for the H I and $T_{dust} = 19.455 \pm 0.002 - (0.643 \pm 0.004) \cdot f_{CNM}$ for the dust. The median Doppler temperature is $T_{D} = 203$ K, corresponding to a median dust temperature of $T_{dust} = 19.08$ K.

The median temperature for dust associated with filamentary H I is colder than the average temperature of $T_{dust} = 19.41 \pm 1.54$ K (Planck Collaboration Int. XLVIII 2016). The H I filaments selected in the current analysis are also slightly colder than the filaments studied in Kalberla et al. (2016) with a median Doppler temperature of $T_{D} = 223$ K. Prominent cold filaments reach $f_{CNM} \sim 0.65$ (see Fig. 18, also Appendix B). Typical Doppler temperatures are $170$ K, with dust temperatures of $19$ K. Such Doppler temperatures correspond to a FWHM line width of $3.2$ km s$^{-1}$ and our sliding velocity window with $\delta v_{LSR} = 3$ km s$^{-1}$ matches well with this line width. We conclude that the ISM in local filamentary structures is comparatively cold.

The middle and lower plots in Fig. 17 show that the Doppler and dust temperatures in filaments are also correlated with the component fraction $f_{LNM}$ and $f_{WNM}$ of the associated warmer gas. For the LNM we fit for the H I $\lg(T_{D}/[K]) = 2.1789 \pm 0.0008 + (0.3600 \pm 0.002) \cdot f_{LNM}$; for the dust $T_{dust} = 19.165 \pm 0.002 - (0.12 \pm 0.004) \cdot f_{LNM}$. For the WNM we obtain $\lg(T_{D}/[K]) = 2.24195 \pm 0.0005 + (0.50156 \pm 0.003) \cdot f_{WNM}$ and $T_{dust} = 18.994 \pm 0.002 + (1.070 \pm 0.006) \cdot f_{WNM}$.

This implies that the Doppler temperatures of USM filaments tend to increase moderately up to typical temperatures around $T_{D} \sim 228$ K and $T_{dust} = 19.17$ K for an LNM environment with a component fraction of $f_{LNM} = 0.5$. Such a H I Doppler temperature differs only slightly from the median of $T_{D} \sim 223$ K determined by Kalberla et al. (2016). For filaments embedded in WNM with $f_{WNM} = 0.5$ we obtain a significantly higher Doppler...
temperature $T_D = 311$ K and correspondingly a higher dust temperature $T_{\text{dust}} = 19.53$ K. The temperature differences would increase further for larger WNM component fractions, however we observe little filaments associated with WNM in this state (Fig. 17 bottom). This deficiency may imply that we observe here a transition to ionized gas.

Calculating average phase fractions for filaments we find that such regions are dominated by the CNM: $f_{\text{CNM}} \sim 0.46$, $f_{\text{LNM}} \sim 0.37$, and $f_{\text{WNM}} \sim 0.17$. Here $f_{\text{CNM}}$ is a lower limit only, our data are not corrected for optical depth effects, for discussion see Kalberla et al. (2016: Sect. 5.9). Optical depth effects however are mostly negligible for $\log(N_H/\text{cm}^{-2}) \lesssim 20.6$ (Lee et al. 2015). Figures 12 and 15 show that most of the CNM at high latitudes is below this limit.

6. H I phases along the line of sight

The results from the previous sections imply that the different H I phases are associated with each other; the observed gas composition changes according to the environment. In Sect. 4.2, considering the extension of CNM dominated regions in the plane of the sky, we demonstrated that there is an anti-correlation between $f_{\text{CNM}}$ and $f_{\text{LNM}}$.

To study further the relationship between CNM, LNM, and WNM we determine the 2D frequency distributions of the observed phase fractions at the same position, hence along the line of sight. We use latitudes $|b| > 20^\circ$ and the velocity range $-8 < v_{\text{LSR}} < 8$ km s$^{-1}$. Figure 18 shows how frequent the CNM with a phase fraction $f_{\text{CNM}}$ is associated with LNM or WNM and phase fractions $f_{\text{LNM}}$ or $f_{\text{WNM}}$ respectively. To the left we display unconstrained data. In this case the CNM contributes typically little H I gas with $f_{\text{CNM}} \lesssim 0.15$. The distribution is bimodal. The associated warmer gas belongs either to the LNM with typically $f_{\text{LNM}} \sim 0.55$ or to the WNM with $f_{\text{WNM}} \sim 0.3$. Thus the LNM is dominant, in agreement with Fig. 16.

The filamentary H I gas has a very different composition. This is demonstrated on the right hand side of Fig. 18. In this case only filamentary H I structures from the USM database have been selected. Clearly, filaments are outstanding with high phase fractions $f_{\text{CNM}}$. Further a sharp division between the associated LNM and WNM phases is visible. The LNM typically occupies regions with larger phase fractions $f_{\text{LNM}}$ compared to the WNM. The sharp diagonal stripe in the upper panel on the right hand side of Fig. 18 confirms that also in filamentary H I structures CNM and LNM are closely anti-correlated. The anti-correlation between CNM and WNM, shown in the panel below, has a larger scatter.

7. Discussion

Our most important result is the large average phase fraction $f_{\text{LNM}} = 0.41$ and the close anti-correlation between LNM and CNM. Most obvious are the phase relations from the distribution of the different phases around cold spots (Sect. 4.2). The detailed anti-correlation between $f_{\text{CNM}}$ and $f_{\text{LNM}}$ on small and intermediate scales proves that the clumpy CNM is embedded in LNM. The WNM surrounds both phases with an anti-correlation on larger scales. The same kind of layered structure is observed around cold filaments. These filaments are associated with cold dust. Doppler temperatures as well as dust temperatures decrease with increasing $f_{\text{CNM}}$. 
When distinguishing H\textsc{i} associated with cold filamentary structures as described in Sect. 5 from the unconstrained H\textsc{i} distribution it needs to be considered that filamentary structures most probably are caused by sheets, seen edge-on (Heiles & Troland 2005; Kalberla et al. 2016, 2017). Similar structures with different orientations, such that we do not have a favorable tangential view (amplifying projected column densities), may exist but these parts of the H\textsc{i} distribution will not be observable with an outstanding morphology. On the other hand, also H\textsc{i} clouds that are warmer than the filaments may contribute to the diffuse H\textsc{i} phase.

From H\textsc{i} observations alone it is difficult to decide whether the WNM shown in Fig. 18 on the right hand side is physically related to the CNM but we see in Fig. 17 that a significant part of WNM in filaments is associated with particular cold dust that appears to be more smoothly distributed than the H\textsc{i}. So we are apparently left with a miracle: warm H\textsc{i} gas associated with cold dust filaments. Thus the observed steepening of local power spectra, implies therefore that the LNM must be related to each other.

One of the main drivers for the launch of Planck was the quest to observe the cosmic microwave background B-mode signal induced by primordial gravitational waves during cosmic inflation. However, there are Galactic foregrounds and it is now established that these are a main limiting factor for accurate cosmological parameters. Polarized Galactic foreground observations with Planck and H\textsc{i} emission data indicate that most of the filamentary dust structures are embedded in filamentary CNM structures (Clark et al. 2014; Kalberla et al. 2016). Based on this finding Ghosh et al. (2017) constructed a phenomenological dust model, combining H\textsc{i} data with an astrophysically motivated description of the large-scale and turbulent Galactic magnetic field. They modeled the large-scale polarized dust emission over the southern Galactic cap by using GASS H\textsc{i} data with selected velocity dispersions. They were able to reproduce the Planck dust observations at 353 GHz for a region of 3500 deg$^2$ at the southern Galactic cap. They tested their model for H\textsc{i} velocity dispersions $\sigma < 3$ km s$^{-1}$ and $3 < \sigma < 7.5$ km s$^{-1}$ and found no significant differences. These two ranges correspond approximately to our CNM and LNM linewidth regimes. Thus the results by Ghosh et al. (2017) are consistent with our conclusion that cold CNM filaments and the surrounding LNM must be related to each other.
8. Summary and conclusion

We used a Gaussian decomposition of the HI4PI survey to separate and model the CNM, LNM, and WNM distributions. Our main results are as follows:

Phase fractions depend on the selected velocity range. For the velocity range \( |b| > 20\degree \), an extrapolation to lower latitudes suffers from large uncertainties but closer to the Galactic plane we did not get significantly different results. It is difficult to define clear criteria for the separation of CNM, LNM, and WNM. The derived H I phase fractions may be uncertain by 50\%.

Previous studies have shown that cold filamentary H I structures in the diffuse interstellar medium are associated with dust ridges, aligned with the magnetic field measured on the structures by Planck at 353 GHz (Clark et al. 2014; Kalberla et al. 2016). These are also the structures where we find high values for \( f_{\text{CNM}} \), and low dust temperatures. It is becoming increasingly evident that the state of the ISM is significantly affected by feedback processes.

Acknowledgements. We thank the referee, Carl Heiles, for constructive criticism that helped to improve the paper, also Jürgen Kerp for valuable discussions and continuous support. U.H. acknowledges the support by the Estonian Research Council grant IUT26-2, and by the European Regional Development Fund (TK133). This research has made use of NASA’s Astrophysics Research Council grant IUT26-2, and by the European Regional Development Fund (TK133). This research has made use of NASA’s Astrophysics Research Council grant IUT26-2, and by the European Regional Development Fund (TK133). This research has made use of NASA’s Astrophysics Research Council grant IUT26-2, and by the European Regional Development Fund (TK133). This research has made use of NASA’s Astrophysics Research Council grant IUT26-2, and by the European Regional Development Fund (TK133).

Note added in proof. After acceptance of the manuscript the 21-SPONGE H I absorption line survey against 57 background continuum sources became available (Murray et al. 2018). These authors distinguish H I phases for derived spin temperatures with \( T_s < 250 \) K for the CNM, \( 250 < T_s < 1000 \) K for the unstable LNM, and \( T_s > 1000 \) K for the WNM. Accordingly they conclude that 28\% of the H I mass is associated with the CNM, 20\% belongs to the LNM, and 52\% to the WNM. We determine for the region covered by our survey mostly dust temperatures between 19 and 19.5 K, independent of preferences in phase fractions, and \( f_{\text{CNM}} = 0.24, f_{\text{LNM}} = 0.45, \) and \( f_{\text{WNM}} = 0.31. Murray et al. (2018) \) do not detect a significant mass fraction for H I at \( 1000 \lesssim T_s \lesssim 4000 \) K. The LNM is a transient phase. We explain discrepancies in the derived LNM and WNM phase fractions as an indication that only about half of the LNM observed by us is currently evolving to cold structures with recognizable optical depths.

References

Audit, E., & Hennebelle, P. 2005, A&A, 433, 1
Clark, S. E., Peek, J. E. G., & Putman, M. E. 2014, ApJ, 789, 82
de Avillez, M. A., & Breitschwerdt, D. 2005, A&A, 436, 585
Dickey, J. M., & Lockman, F. J. 1990, ARA&A, 28, 215
Dickey, J. M., Salpeter, E. E., & Terzian, Y. 1977, ApJ, 211, L77
Dickey, J. M., McClure-Griffiths, N. M., Gaensler, B. M., & Green, A. J. 2003, ApJ, 585, 801
Egger, R. J., & Aschenbach, B. 1995, A&A, 294, L25
Field, G. B., Goldsmith, D. W., & Habing, H. J. 1969, ApJ, 155, L149
Fitzpatrick, E. L., & Spitzer, L., Jr. 1997, ApJ, 475, 623
Frisch, P., & Dwarkadas, V. V. 2018 [arXiv:1801.06223]
Gazol, A., Vázquez-Semadeni, E., Sánchez-Salcedo, F. J., & Scalo, J. 2001, ApJ, 557, L121
Ghosh, T., Boulanger, F., Martin, P. G., et al. 2017, A&A, 601, A71
Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759
Haud, U. 2000, A&A, 364, 83
Haud, U. 2002, A&A, 438, 461
Haud, U., & Kalberla, P. M. W. 2007, A&A, 466, 555
Heiles, C. 2001, ApJ, 551, L105
Heiles, C., & Troland, T. H. 2003a, ApJ, 586, 1067
Heiles, C., & Troland, T. H. 2003b, ApJS, 145, 329
Heiles, C., & Troland, T. H. 2005, ApJ, 624, 773
Hennebelle, P., & Falgarone, E. 2012, A&ARv, 20, 55
HI4PI Collaboration, (Ben Bekhti, N., et al.) 2016, A&A, 594, A116
Jones, A. P., Tielens, A. G. G. M., & Hollenbach, D. J. 1996, ApJ, 469, 740
Kalberla, P. M. W., & Haud, U. 2015, A&A, 578, A78
Kalberla, P. M. W., Schwarz, U. J., & Goss, W. M. 1985, A&A, 144, 27
Kalberla, P. M. W., Burton, W. B., Hartmann, D., et al. 2005, A&A, 440, 775
Kalberla, P. M. W., McClure-Griffiths, N. M., Pisano, D. J., et al. 2010, A&A, 521, A37
Kalberla, P. M. W., Kerp, J., Haud, U., et al. 2016, ApJ, 821, 117
Lee, M.-Y., Stanimirović, S., Murray, C. E., Heiles, C., & Miller, J. 2015, ApJ, 809, 56
McClure-Griffiths, N. M., Pisano, D. J., Calabretta, M. R., et al. 2009, ApJS, 181, 398
McKee, C. F., & Ostriker, J. P. 1977, ApJ, 218, 148
McBuld, U. 2012, A&A, 19, 13
McBuld, U., Winnberg, A., Kalberla, P. M. W., & Goss, W. M. 1982, A&A, 223
Micelotta, E. R., Jones, A. P., & Tielens, A. G. G. M. 2010, A&A, 510, A36
Mitchell, A. G. G., & Zmiksky, M. W. 1971, Resonance radiation and excited atoms (Cambridge: Cambridge University Press)
Mohan, R., Dwarakanath, K. S., & Srinivasan, G. 2004, JAA&a, 25, 143
Murray, C. E., Stanimirović, S., Goss, W. M., et al. 2015, ApJ, 804, 89
Murray, C. E., Stanimirović, S., Kim, C.-G., et al. 2017, ApJ, 837, 55
Murray, C. E., Stanimirović, S., Goss, W. M., et al. 2018, ApJS, 238, 14
Nidever, D. L., Majewski, S. R., & Butler Burton, W. 2008, ApJ, 679, 432
Planck Collaboration Int. XLVIII. 2016, A&A, 596, A109
Roy, N., Kanekar, N., & Chengalur, J. N. 2013, MNras, 436, 2366
Salpeter, E. E. 1976, ApJ, 206, 673
Saury, E., Miville-Deschênes, M.-A., Hennebelle, P., Audist, E., & Schmidt, W. 2014, A&A, 567, A16
Spitzer, L., Jr. & Fitzpatrick, E. L. 1995, ApJ, 445, 196
Takakubo, K., & van Woerden, H. 1966, Bull. Astron. Inst. Netherlands, 18, 488
Tielens, A. G. G. M. 2008, ARA&A, 46, 289
Vázquez-Semadeni, E., & Scalo, J. 2001, ApJ, 557, L121
Winkle, B., Kerp, J., Flöer, L., et al. 2016a, A&A, 585, A41
Winkle, B., Lenz, D., & Flöer, L. 2016b, A&A, 591, A12
Wolfire, M. G., Hollenbach, D., McKee, C. F., Tielens, A. G. G. M., & Bakes, E. L. O. 1995, ApJ, 443, 152
Wolfire, M. G., McKeel, C. F., Hollenbach, D., & Tielens, A. G. G. M. 2003, ApJ, 587, 278
Ysard, N., Köhler, M., Jones, A., et al. 2015, A&A, 577, A110

A58, page 17 of 20
Appendix A: On the XNM

In this paper we discuss CNM, LNM, and WNM phases of H\textsc{i}, but Figs. 2, 3, and 6 contain also the fourth curve – XNM. We mentioned that in many cases, but not always, the XNM Gaus-
sians seem to be caused by the wide but weak wings of the lines from other phases. In a Gaussian decompo-
sition such lines cannot be described by a single component as the Gaussians fall off too rapidly compared to the data. However, such lines are often successfully modeled with the Voigt profile (Mitchell & Zemansky 1971) – a convolution of a Gaussian function with a Lorentzian, or, due to the computational expense of the con-
volution operation, the Voigt profile is approximated with a pseudo-Voigt one. Here we try to give a brief test, if such a more complicated fitting procedure may help us to avoid the XNM components in our decomposition.

From the middle panel of Fig. 2, we can see that the XNM is well visible already for profiles, decomposed with only two low-velocity Gaussians. With such profiles it is easy to estimate the presumable improvements in the results, obtainable by using of Voigt profiles. However, the Voigt profiles are only able to describe symmetric non-Gaussian wings. Therefore, it is important to estimate, how symmetric is the XNM component in each profile located relative to another Gaussian of the same H\textsc{i} profile.

To study the symmetry of the simple profiles, we selected from our Gaussian database 51 726 profiles with $N_G = 2$, in which at least one Gaussian had the width, typical for XNM ($lg(FWHM) > 1.64$ from Fig. 2). For each profile we computed the positive velocity difference $\Delta v_c$ between the centers of the two Gaussians. Figure A.1 gives the distribution of the contributions (the percentage from the total column density of the XNM, represented by the selected profiles), of the XNM components with different shifts relative the main components of the profiles. We can see that the maximum of the distribution corresponds to a considerable shift between the centers of the Gaussians. However, due to the observational noise we cannot expect that the centers of the Gaussians, which describe the main H\textsc{i} peak and its wide wings coincide exactly and the acceptable mismatch increases with the increasing width of the XNM component.

We have studied the influence of the observational noise to the uncertainties of the Gaussian parameters in Haud (2000) and on the basis of this study we expect that for our selected XNM Gaussians with average $FWHM = 56.5$ km s$^{-1}$, the shifts up to about $\Delta v_c = 6$ km s$^{-1}$ (broad green line in Fig. A.1) between the centers of two components in each profile may yet be acceptable for fitting both of these Gaussians with the symmetric Voigt profile. From Fig. A.1 we can see that most (about 79\%) of the XNM column density in studied simple profiles comes from the Gaussians with larger displacements relative to the main component. Corresponding profiles must be clearly asymmetric and for these we cannot expect good fits with symmetric Voigt profiles. Moreover, the Gaussians, which represent base-line problems or other asymmetric features, may be centered anywhere on the velocity axis, but the wings, describable by the Voigt profiles must be centered near our present main component of the corresponding profile. Therefore, even at $\Delta v_c < 6$ km s$^{-1}$ we may expect some contribution from features for which modeling with the Voigt profiles may be questionable and therefore $100 - 79 = 21\%$ is an upper limit for the fraction of simple profiles that can be modeled successfully with Voigt profiles.

Of course, the estimate for the permissible displacement of the Gaussians for successful fitting with the Voigt profile,

\begin{align*}
l = 265.541^\circ \\
b = -44.053^\circ \\
FWHM_G = 25.8 \text{ km s}^{-1} \\
FWHM_V = 24.3 \text{ km s}^{-1}
\end{align*}

Fig. A.1. Fractions of the total column density of the XNM from profiles with $N_G = 2$ as a function of the velocity difference $\Delta v_c$ between the centers of their Gaussian components. The shifts up to the thick green line are likely explainable with the uncertainties in the Gaussian decomposition and the thin green line marks the median of the distribution.

\begin{align*}
l = 276.390^\circ \\
b = -48.825^\circ \\
FWHM_G = 23.6 \text{ km s}^{-1} \\
FWHM_V = 24.3 \text{ km s}^{-1}
\end{align*}

\begin{align*}
l = 265.341^\circ \\
b = -44.053^\circ \\
FWHM_G = 25.8 \text{ km s}^{-1} \\
FWHM_V = 28.1 \text{ km s}^{-1}
\end{align*}

Fig. A.2. Two example fits of the observations with the Gaussian com-
ponents and with the pseudo-Voigt profile. The upper panel corre-
sponds to the nearly ideal case for the fitting with the pseudo-Voigt profile and the lower panel represents the typical situation. Thick lines correspond to the observations and the models. Thin lines illustrate the Gaussian and the Lorentzian contributions to the final models. Crosses mark the points, which define FWHM$_{\text{XNM}}$ in two different models.
is relatively uncertain and it may be possible that contrary to our expectations even at higher displacements satisfactory fits with Voigt profiles exist. To test this, we give two examples in Fig. A.2. The upper panel corresponds to nearly ideal case for fitting with the Voigt profile with \( \Delta v_c = 0.06 \) km s\(^{-1}\). The reduced \( \chi^2 \) values for fits with Gaussian components and with the pseudo-Voigt profile are \( \chi^2/N_{\text{dof}}^G = 0.981 \) and \( \chi^2/N_{\text{dof}}^V = 0.975 \), respectively. Both these values satisfy our requirements for a good fit (see Sect. 2.2) and as the fit with the pseudo-Voigt profile contains less free parameters we would prefer this.

For the lower panel of Fig. A.2 we have chosen a more typical example. \( \Delta v_c = 10.14 \) km s\(^{-1}\) is close to the maximum of the distribution in Fig. A.1, but still less than the median of that distribution at \( \Delta v_c = 13.2 \) km s\(^{-1}\) (thinner green line in Fig. A.1). The contribution of the XNM component to the column density of this profile \( (N_{\text{XNM}}/N_{\text{Tot}} = 0.39) \) is also close to the median \( (N_{\text{XNM}}/N_{\text{Tot}} = 0.40) \) of the corresponding distribution. As we can see, this profile is fitted well with Gaussians \( (\chi^2/N_{\text{dof}}^G = 1.022) \), but the fit with the pseudo-Voigt profile is not any more a satisfactory replacement for the XNM component \( (\chi^2/N_{\text{dof}}^V = 1.165) \). The change of the line width is also considerably larger (from \( FWHM_G = 25.8 \) km s\(^{-1}\) to \( FWHM_V = 28.1 \) km s\(^{-1}\)) but we have to note that this is for a final fit we need at least one more pseudo-Voigt profile, the interpretation of the changes becomes already complicated.

Indirect indication that the modeling of the HI profiles with more general line shapes, other than Gaussians, cannot solve the problems with the XNM, can be seen also from Fig. A.3, which gives the distribution of the XNM fraction, \( f_{\text{XNM}} \), in the sky. Except the galactic plane, the \( f_{\text{XNM}} \) is often the highest in the sky areas, where the total column density is low (see Fig. 2. in HI4PI Collaboration 2016). This is in particular true for the prominent region \( 130 < l < 190^{o} \) and \( 50 < b < 70^{o} \). There are extended regions with brightness temperatures \( T_B < 0.5 \) K at \( v_{\text{LSR}} \sim 0 \) km s\(^{-1}\) and for such a low emission remaining instrumental problems (RFI, baseline problems and residual stray radiation errors) can easily trigger Gaussian components that are interpreted as XNM. Problems of this kind are recognizable as boxy structures, discontinuities at the borders of individual observing sessions in RA/DEC. About 0.5% of all positions are affected. Demanding for a better reliability of XNM components a more stringent limit of \( T_B \leq 1 \) K, we find that 3% of the XNM Gaussians may possibly be affected by instrumental problems. The XNM fraction is high also in the first and fourth quadrants near the galactic plane, where the column density of the HI is the highest, but there the different features in the line profiles are so heavily blended with each other that any decomposition of these profiles becomes unreliable, yielding many components with very different properties. About 6% of all positions are affected. In our study we have avoided making conclusions on the basis of the observations near the galactic plane and have used corresponding Gaussians only for tests of the limits, to which the results, obtained from higher latitudes, are applicable.

Therefore, there are certainly cases, where the usage of the Voigt profiles may give better results for the profile decomposition, but we cannot expect that they will solve most of the problems with XNM. Using Voigt profiles for the real symmetric wings of the HI lines improves (increases) somewhat our line-width estimates, but by fitting asymmetric lines or some combinations of the real lines with more questionable features may introduce additional uncertainties to the line-width estimates and as we cannot in all cases automatically distinguish the line wings and the more problematic features, the real gain from the more complicated mathematics may remain rather questionable. As a result, we consider the usage of the Voigt profiles in the present work impractical.

We conclude that a Gaussian decomposition appears to provide a reasonable parametrization of the XNM. About 3% of the XNM may be explainable as instrumental, additional 6% in the Galactic plane as affected by confusion. The rest appears to be associated with WNM structures, mostly asymmetric. This is of course not a proof that such structures must be real. The problem is more general, the question is how to treat extended profile wings when processing instrumental baselines of HI survey observations. Extended asymmetric profile wings may easily be fitted away as baseline errors but the approaches for the LAB (Kalberla et al. 2005), GASS (Kalberla & Haud 2015), and EBHIS (Winkel et al. 2016a) were conservative, taking care that such wings are not eliminated.

Appendix B: Total column density distributions with uncertainties

With Fig. 12 we presented 2D distributions for column densities associated with individual phases. A more concise presentation is possible if we summarize the column densities for all gas at a particular phase fraction. For an optical thin medium, assuming further that the distances of the three different observed gas phases do on average not vary significantly, the total column density is proportional to the total mass of the gas associated with a particular phase fraction.

Figure B.1 shows for each of the phases how phase fraction and contributing column densities are related to each other. We distinguish between HI associated with cold USM filaments (top) and HI observed outside such filamentary structures (bottom). As detailed in Sect. 2.4, the separation of the CNM, LNM, and WNM is burdened with rather large uncertainties (see also Fig. 8). To get an impression how far such uncertainties affect the derived distributions, we compare several cases; the best fit result (middle) together with distributions that were derived after applying thresholds offset by the median standard deviation when separating the individual gas phases. We distinguish
Comparing the upper panels of Fig. B.1 with the lower ones, shows how far the variation of the thresholds with phase content that we annotate as min or max affect the derived CNM, LNM, and WNM distributions. In general the column density distributions outside filaments at constant velocities \( |v_{LSR}| < 1 \text{ km s}^{-1} \) are not associated with filaments (bottom) is characterized by a broad range of observed phase fractions for all phases, in particular for the CNM. The LNM tends to have predominantly \( f_{\text{LNM}} \approx 0.5 \) while \( f_{\text{WNM}} \approx 0.5 \).

The middle panels of Fig. B.1 show that CNM is dominant in USM filaments with large phase fractions. A large amount of the gas is concentrated in cold condensations, \( f_{\text{CNM}} \approx 0.65 \). Most of the CNM is associated with LNM, correspondingly with \( f_{\text{LNM}} \sim 0.4 \). The WNM is unimportant with \( f_{\text{WNM}} \sim 0.15 \). H\(_{\text{i}}\) gas that is not associated with filaments (bottom) is characterized by a broad range of observed phase fractions for all phases, in particular for the CNM. The LNM tends to have predominantly \( f_{\text{LNM}} \approx 0.5 \) while \( f_{\text{WNM}} \approx 0.5 \).

The upper panels of Fig. B.1 show that CNM is dominant in USM filaments with large phase fractions. A large amount of the gas is concentrated in cold condensations, \( f_{\text{CNM}} \approx 0.65 \). Most of the CNM is associated with LNM, correspondingly with \( f_{\text{LNM}} \sim 0.4 \). The WNM is unimportant with \( f_{\text{WNM}} \sim 0.15 \). H\(_{\text{i}}\) gas that is not associated with filaments (bottom) is characterized by a broad range of observed phase fractions for all phases, in particular for the CNM. The LNM tends to have predominantly \( f_{\text{LNM}} \geq 0.5 \) while \( f_{\text{WNM}} \leq 0.5 \).

Comparing the upper panels of Fig. B.1 with the lower ones, confirms our previous finding that filamentary H\(_{\text{i}}\) is special in the sense that the filamentary CNM is dominant for these structures. Correspondingly the total CNM fraction is in this case twice as high as for the non-filamentary sample. Filamentary H\(_{\text{i}}\) contains little WNM, only half of the fraction derived for the non-filamentary sample, see Fig. 18 for a complementary view.

The middle panels of Fig. B.1 show how far the variation of the thresholds with phase content that we annotate as min or max affect the derived CNM, LNM, and WNM distributions. In general the column density distributions are shifted to higher phase fractions for the max cases and opposite to lower for the min cases. However the observed variations do not invalidate the conclusions drawn above that filamentary H\(_{\text{i}}\) structures are enhancements in CNM, associated predominantly with LNM envelopes that have significant phase fractions \( f_{\text{LNM}} \). The WNM envelopes are in total less important.