Dust and gas in the Milky Way: a full-sky view on the multiphase interstellar medium

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

von
Daniel Lenz
aus
Bonn

Bonn, Juli 2016

Dieser Forschungsbericht wurde als Dissertation von der Mathematisch-Naturwissenschaftlichen Fakultät der Universität Bonn angenommen und ist auf dem Hochschulschriftenserver der ULB Bonn http://hss.ulb.uni-bonn.de/diss_online elektronisch publiziert.

1. Gutachter:	Privatdozent Dr. Jürgen Kerp
2. Gutachter:	Prof. Dr. Thomas Reiprich
Tag der Promotion:	12.09.2016
Erscheinungsjahr:	2016

Contents

Abstract

The Milky Way (MW) Galaxy is a unique laboratory to study the evolution of spiral galaxies, offering great sensitivities and angular resolutions with respect to observations of other galaxies. There, the large distance and consequently low linear resolution makes it impossible to study the accretion of matter and gas phase transitions through observations of individual clouds. The role of these high- and intermediate-velocity clouds (IVCs and HVCs), located in the Galactic halo, is still not fully understood. By using data of atomic neutral hydrogen (H I) and far-infrared (FIR) emission from dust grains, it is possible to probe a large range of densities, temperatures and gas phases. Here, the newly released data of the Effelsberg-Bonn HI Survey (EBHIS) and the FIR data from the *Planck* satellite allow to take this analysis to a new level in terms of sensitivity, spectral coverage, and angular resolution.

Classically, it is believed that HVCs originate from beyond the Galactic disk, either from the Galactic halo that surrounds the disk or from structures located beyond the MW. IVCs on the other hand are thought to be related to Galactic fountain processes in which supernovae eject gas into the halo where it eventually cools and returns to the disk.

We present evidence for the impact of a HVC onto a well studied IVC located close to the disk. The correlation of dust and gas allows us to infer the distribution and column density of molecular hydrogen. At the impact region, increased pressure reduces the formation time of molecular hydrogen. The morphology and velocity structure of the IVC/HVC system, combined with the dusty and molecular content of the IVC, points towards a physical connection between the two clouds.

To further investigate the properties of HVCs and their role in MW evolution, we conduct a detailed analysis of a very cold, bright, and clumpy HVC that is located at high Galactic latitude. These properties make this particular cloud one of the best candidates to detect FIR emission from interstellar dust. Owing to their large distance from the interstellar radiation field and their low metallicity, there have been only tentative detections of dust in very few HVCs. Here, the biggest challenge is to remove the foreground dust emission associated with the MW in order to detect the very faint signal from the HVC. We present a new approach to model the FIR foreground, demonstrate its advantages over previous methods, and verify its accuracy and robustness via simulations. Despite these advancements in cleaning the map from foreground emission, we find that no dust emission is present in that HVC. For this type of analysis, the fluctuations from the cosmic infrared background (CIB) are the limiting factor.

The work on dust in HVCs is the foundation for the final chapter of this thesis in which we present the first results on large-scale studies of the CIB. This background radiation is dominated by galaxies at redshifts between $z = 1..3$. Up to now, CIB maps were not corrected for foreground emission at all or only by a simple model that required manual inspection. Using the full spectral information of the H I data and modern methods of machine learning, we present a fully-automated approach to remove galactic foreground emission from CIB data for a large (20%) fraction of the sky. We compare our results on the CIB with previous studies to validate our approach. Finally, we compute the deconvolved, binned, and unbiased angular power spectrum that is used to compare the data to cosmological simulations. For the first time, fully cleaned images of the CIB on large scales and their power spectra are available to study phenomena such as the star-formation history and the connection between luminous and dark matter at

large scales. We thereby improve previous constraints on the CIB at the largest scales and further expand the angular range that is probed. This allows future cosmological models to be probed to a new level and with greater accuracy.

CHAPTER 1

Introduction

The multiphase interstellar medium (ISM)[1] in galaxies consists of different gaseous and dusty components, ranging from very thin, ionised H II regions to the densest molecular clouds with different molecular species and cold, large dust grains (Draine, 2011). Understanding the evolution of the ISM in a galaxy, its role in star formation and feedback mechanisms is the main driver for astronomers investigating the Milky Way and other spiral galaxies. Moreover, there are ongoing debates about how the Milky Way (MW) galaxy can have sustained its star formation over billions of years without any major mergers providing gaseous fuel for the formation of stars.

By studying the interplay of the different gaseous and dusty species in the MW, we can improve our understanding of the physics that relates these components to each other. In other galaxies, we lack the spatial resolution to resolve the complex structure of the ISM, making the MW a unique laboratory for these studies. At an angular resolution of 1′, we can resolve nearby MW clouds down to 0.14 pc. In the Large Magellanic Cloud (LMC) or in the nearby spiral galaxy M31, this drops to 15 and 227 pc, respectively.

The knowledge of the transitions and properties of the different ISM constituents opens a window to the extragalactic sky. An accurate model of the emission from the MW galaxy is an important prerequisite for studies of cosmological background radiation in any wavelength, from gamma-rays to the infrared. In extragalactic observations, foreground emission associated with the solar system and the MW galaxy obscures our view and is often a limiting factor for these studies. In order to investigate the distant Universe, it is thus necessary to understand the local Universe first.

We begin by introducing the different topics that are relevant for this thesis. Therefore, we cover the structure of neutral atomic hydrogen in the MW and its role in the evolution of the galaxy, the correlation between dust and gas, and the cosmic infrared background (CIB) that can eventually be studied if the foreground emission is successfully accounted for.

1.1 H I in the Milky Way

Neutral atomic hydrogen (H I) is an ubiquitous component in spiral galaxies and plays a key role in their evolution. It is traced by the 21-cm spectral line and is so abundant that it can be detected after a few seconds of integration time with a 100-m class telescope in any part of the sky. H I was first observed by Ewen and Purcell (1951) who thereby confirmed the prediction made by van de Hulst (1945).

[1] For an overview of the acronyms used throughout this work, see page 123.

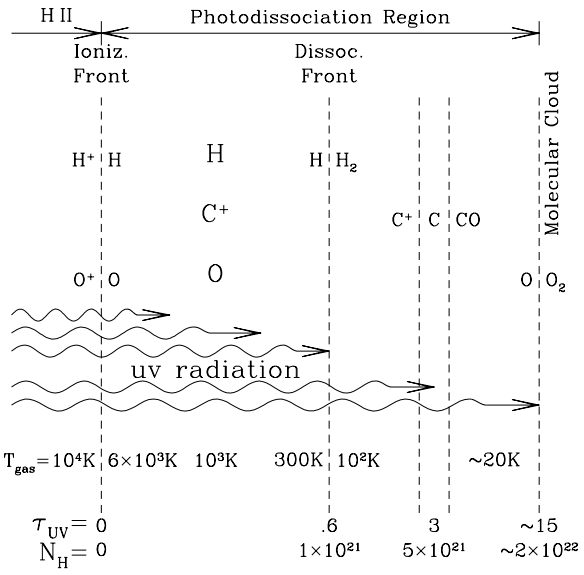


Figure 1.1: Structure of a PDR. It illustrates how the different regions of the PDR are mostly defined by the attenuation of the UV radiation field, determining the ionisation level and temperature of the region. The transitions between the individual layers are rather sharp, resulting in ionisation and dissociation fronts. Image taken from Draine and Bertoldi (1999).

1.1.1 ISM phases

The ISM has a very complex structure in terms of ionisation levels, temperatures and densities. In the following, we give a very brief overview on the different phases of the ISM with a emphasis on neutral atomic hydrogen which is the focus of this thesis. For a more detailed and extended discussion, we recommend the article by M. G. Wolfire et al. (2003) and the book by Draine (2011). For consistency, we use the numbers for different ISM properties such as temperatures and densities from Draine.

The structure of the ISM can be described by looking at the structure of a photodissociation region (PDR) (Fig. 1.1). The shape of a PDR depends mostly on the attenuation of UV radiation. This UV radiation is emitted by massive, young stars. With increasing hydrogen column density and hence increasing optical depth, the UV radiation is strongly attenuated and the molecules are shielded efficiently. Different molecules are dissociated at different photon energies, giving rise to the shell-like structure.

The hot ionised medium (HIM) has temperatures around 10^6 K and densities of approximately 10^{-3} cm^{-3}. It is heated by interstellar shocks and cooled via adiabatic expansion and the emission of X-rays which are the basis of HIM observations, combined with radio synchrotron and UV observations.

The warm ionised medium (WIM) is slightly cooler than the HIM with temperatures around 10^4 K and densities ranging from 0.3 to 10^4 cm^{-3}. Being ionised, it is heated through photoelectrons from H and He and cooled by optical line transitions, fine-structure emission and free-free emission, allowing to observe it in optical line emission and the thermal radio continuum.

Atomic neutral hydrogen exists in a cold and in a warm phase (McKee and Ostriker, 1977; M. G. Wolfire et al., 1995). These phases are in pressure equilibrium with $P/k_\mathrm{B} \approx 10^3$ K cm^{-3}. Here, P is the pressure and k_B is the Boltzmann constant. Fig. 1.2 (left) shows the pressure P as a function of the density n from M. G. Wolfire et al. (1995). In this diagram, regions of thermal stability occur where $dP/dn > 0$ (Field, 1965). With its characteristic pressure curve, the neutral atomic medium can stably exist in two phases that are in pressure equilibrium.

The warm neutral medium (WNM) is the neutral equivalent to the WIM with similar temperatures and densities. It is heated by photoelectrons from dust and is observed through H I 21-cm emission and absorption, and optical and UV lines. It has a diffuse structure and is not as clumpy as the colder atomic

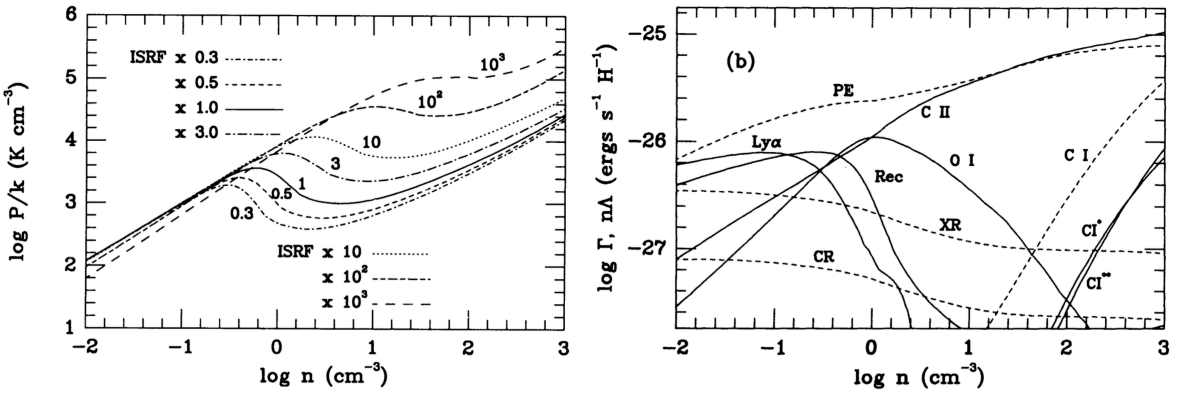


Figure 1.2: **Left:** $\log P$ versus $\log n$ diagram for different strengths of the interstellar radiation field (ISRF), illustrating the regions in which different gas phases can exists in thermal equilibrium. **Right:** Efficiency of the heating and cooling processes in the ISM as a function of the hydrogen density. Heating (dashed lines): Photoelectric heating from dust grains (PE), X-rays (XR), cosmic rays (CR), photoionisation of C (C I). Cooling (solid lines): C II fine-structure (C II), O I fine-structure, recombination onto dust grains (Rec), Lyman α transitions (Lyα), C I fine-structure at 609 μm (C I*), C I fine-structure at 370 μm (C I**). Images taken from M. G. Wolfire et al. (1995).

gas (Kalberla et al., 2016).

The cold neutral medium (CNM) is in pressure equilibrum with the WNM and has temperatures around 100 K and densities of approximately 10 cm^{-3}. It is mostly cooled through fine-structure line transitions and is, unlike its warm counterpart, rather clumpy and not diffuse.

The denser regions of the ISM are characterised by their molecular content and hence are often referred to as molecular medium (MM). It is heated by starlight and cosmic rays and cooled via fine-structure emission and dust far-infrared (FIR) emission. Common observational tracers are FIR dust emission, CO 2.6-mm emission and C I. The dust in molecular clouds acts as a catalyst for the formation of H_2 which cannot form efficiently in the gas-phase (Glover, 2003). Moreover, molecular hydrogen can hardly be observed directly because of its missing electric dipole. Instead, astronomers use tracer molecules such as CO or OH to infer the H_2 column density. The classical X_{CO} conversion factor is defined as $X_{CO} = N_{H_2}/W_{CO}$ and is hence used to convert from the CO line integral to H_2 column densities (e.g. Bolatto, M. Wolfire and Leroy, 2013). Its value is of the order of $X_{CO} = 2 \times 10^{20}\,\mathrm{cm}^{-2}/(\mathrm{K\,km/s})$ with a 30% uncertainty, reflecting its intrinsic variation across different clouds.

Heating and cooling mechanisms

The ISM is typically not in thermal equilibrium and different heating and cooling mechanisms constantly strive for balance. Depending on the density, temperature, and ionisation state, different mechanisms dominate this process (M. G. Wolfire et al., 1995; Draine, 2011). Simply spoken, heating and cooling add and remove translational kinetic energy of the particles in the gas. Fig. 1.2 (right) shows an overview of the different heating and cooling processes and how their efficiency depends on the hydrogen density.

Heating

Photoelectric heating by dust is the dominant heating process in the diffuse neutral ISM. Here, UV photons with energies as small as ≈ 4.5 eV can photo-ionise the larger carbonaceous grains (Moos et al.,

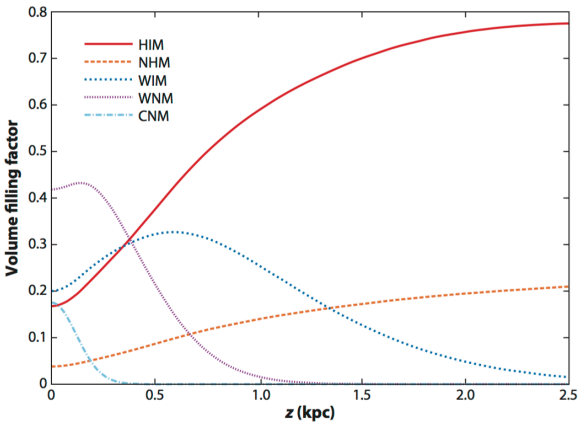


Figure 1.3: Volume filling factors for different ISM phases as a function of height above the disk. See Sect. 1.1.1 for details on the different phases. These results were derived using a hydrostatic model fit (Kalberla and Dedes, 2008) and the image is taken from Kalberla and Kerp (2009).

2001). However, it is generally most efficient for small ($\lesssim$ 150 nm) dust grains because of the large photoelectric efficiency (M. G. Wolfire et al., 1995).
X-rays ionise H and He atoms and are most efficient in the WNM. However, this effect is only important in clouds that are close to sources of X-ray emission such as supernova remnants.
Heating by cosmic rays was the first mechanism that was proposed to be responsible for heating the ISM (Field, Goldsmith and Habing, 1969). Particularly at low energies of a few MeV, cosmic rays are an efficient heating source in dense regions where photons fail to penetrate large optical depths.
Chemical heating occurs on the surface of dust grains which act as a catalyst for the formation of H_2. The merging of two H atoms yields an energy of 4.5 eV that is distributed over the rotational and vibrational modes of the dust grain.
Photoionisation of H, He, and metals such as C, Mg, and Si. by UV starlight is effective across many orders of magnitude and depends on the abundance of the elements. In H II regions, the photoionisation of H and He is an important heating source. For denser, more diffuse regions, this effect is negligible.

Cooling

Fine-structure line emission from C II 158 μm, O I 63 μm, and other lines are the most important coolants for the ISM at temperatures of $10 \lesssim T \lesssim 10^4$ K. Collisions excite these transitions and will eventually lead to the emission of a photon, removing kinetic energy from the gas.
Dust-gas collisions and recombinations of electrons onto dust grains is effective for $T \gtrsim 10^4$ K. The energy that is absorbed by the dust particle is re-emitted in the infrared (IR) regime.
Recombination radiation is relevant for the hot and warm ionised medium. Here, an electron is captured by the atom, commonly H, and the energy difference between the electron kinetic energy and the binding energy is radiated away.
Free-free emission and **synchrotron emission** occurs at high temperatures when the gas is ionised. Electrons that scatter off free ions or traverse magnetic fields emit Bremsstrahlung and dominate the cooling above 10^6 K.

1.1.2 Structure of the gaseous Milky Way

The MW galaxy, like other spiral galaxies, consists of a disk and a halo. The gaseous disk is made up of a thin and a thick gas component. According to the two-component model by McKee and Ostriker (1977), the CNM is thought to be in pressure equilibrium with the WNM. For these phases, the scale heights at

the location of the sun are 150 pc and 400 pc, respectively (Kalberla and Kerp, 2009). Fig. 1.3 shows the volume filling factors for different phases of hydrogen gas, based on hydrostatical simulations (Kalberla and Dedes, 2008). In addition to the already mentioned phases, the neutral halo medium (NHM) and the warm ionised medium (WIM) are shown. For a more detailed overview on the structure of the gaseous MW, we refer to Kalberla and Kerp (2009).

These disk components and their scale heights are not sharply defined and do not follow a strict pattern of differential rotation, but instead show strong flaring (Kalberla and Dedes, 2008). For both the CNM and WNM, the thickness of the gas layer increases with distance from the Galactic centre and the average density decreases.

1.1.3 Intermediate-velocity clouds

It is not straightforward to exactly define the transitional layer between the disk and the Galactic halo (the disk–halo interface, e.g. Putman, Peek and Joung, 2012). The H I sky is dominated by filaments, clumps and sheets and the transition at the disk–halo interface can be very smooth. A fair approach is to assume a model of differential Galactic rotation (Kalberla et al., 2007) and look for anomalies and deviations from that model, mostly via the deviation velocity $v_{\rm dev}$. Unfortunately, this model is not included in most studies and the radial velocity $v_{\rm LSR}$[2] of the H I is used instead to distinguish between local disk gas and halo gas components. For gas at high Galactic latitudes, which is the focus of the present work, this is a fair approximation. Because of the location right above and below the Sun's position in the Galactic disc, the relative motion tangential to the disk can be neglected in first order.

Typically, intermediate-velocity clouds (IVCs) are defined as clouds whose velocities are incompatible with a simple model of differential Galactic rotation. Their radial velocities $v_{\rm LSR}$ with respect to the LSR are loosely constrained to be about $40\,{\rm km\,s^{-1}} < |v_{\rm LSR}| < 90\,{\rm km\,s^{-1}}$, but this definition is debated because IVCs are commonly mixed with local H I emission in velocity space. Furthermore, this confusion explains why IVCs are difficult to isolate as individual objects.

IVCs are often located at the disk–halo interface, at altitudes of up to 2 kpc with respect to the Galactic disk (Wakker, 2001; Richter et al., 2001; Ben Bekhti et al., 2012). Their metallicities are typically solar or slightly lower (Wakker, 2001; Wakker et al., 2008; Fox et al., 2004; Richter et al., 2001), owing to their potential mixing with primordial gas in the Galactic halo (Fraternali et al., 2013, see also Sect. 1.2).

Using data on different metal absorption lines from the Far Ultraviolet Spectroscopic Explorer (FUSE) satellite, Richter et al. (2003) investigated the molecular hydrogen in IVCs with H I column densities of $N_{\rm HI} > 10^{19}\,{\rm cm^{-2}}$ via absorption towards extragalactic sources. They find detections in 14 out of 56 lines of sight with typical H_2 column densities between 10^{14} and $10^{17}\,{\rm cm^{-2}}$. These rather low molecular fractions indicate a physical association with gas that is dominated by the CNM.

Aside from the IVCs with low fractions of molecular hydrogen, there is a distinct population of IVCs that are CO-bright and hence contain significant amounts of H_2 (Hartmann, Magnani and Thaddeus, 1998; Magnani, Hartmann and Speck, 1996; Magnani and Smith, 2010). The work by Magnani and Smith has lead to a catalogue of a dozen of these intermediate-velocity molecular clouds (IVMCs), most of them associated with an IVC feature known as the IV Spur. Unlike molecular disk gas with a scale-height of approximately $z \approx 80$ pc (Cox, 2005), these IVMCs are located further away from the disk at the disk-halo interface at $z \approx 300 - 500$ pc (Benjamin et al., 1996; Gladders et al., 1998). Very recently, a full-sky catalogue of IVMCs was compiled by Röhser et al. (2016, submitted), using the correlation of atomic neutral hydrogen and dust grains.

[2] The local standard of rest (LSR) is a reference frame and describes the path of the sun around the centre of the MW. The solar system moves with a velocity of approximately $255\,{\rm km\,s^{-1}}$ around the centre which is at a distance of $d \approx 8.34$ kpc (Reid et al., 2014).

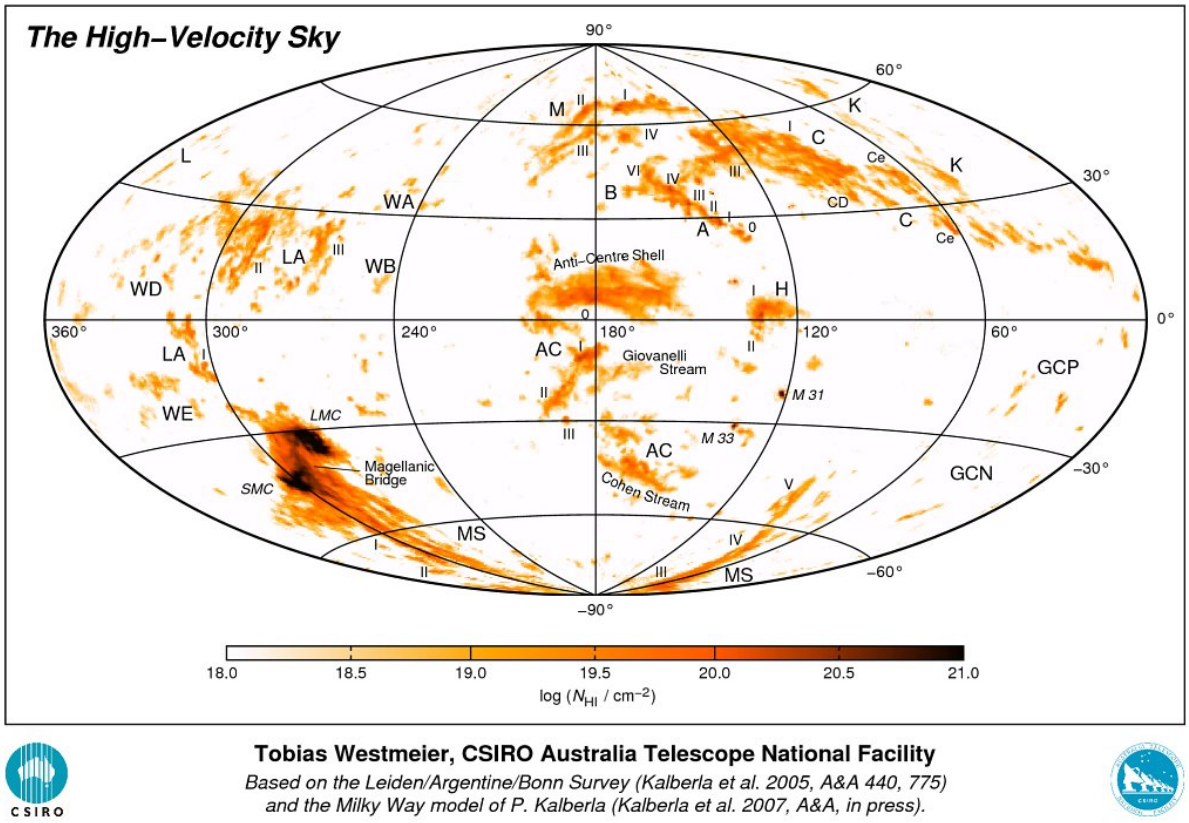


Figure 1.4: The HVC sky, divided into the classical HVC complexes that make up most of the HVC population. Image courtesy: Tobias Westmeier, based on LAB data and the MW model by Kalberla and Dedes (2008). Note that the Galactic anti-centre at $l = 180°$ is in the central meridian.

1.1.4 High-velocity clouds

Similar to IVCs, high-velocity clouds (HVCs) are defined as objects whose radial velocities are incompatible with a simple Galactic model of differential rotation. An overview of these objects, based on a MW model (Kalberla and Dedes, 2008) and the Leiden/Argentine/Bonn Survey (LAB, Kalberla et al. 2005) data is shown in Fig. 1.4. Aside from some compact, isolated HVCs, they are mostly found in larger HVC complexes. These halo objects were first detected by Muller, Oort and Raimond (1963) and are the densest and coldest components of the outer Galactic halo. They are usually defined via their radial velocities which exceed $|v_{\mathrm{LSR}}| \gtrsim 90\,\mathrm{km\,s^{-1}}$.

Even though our knowledge about HVCs is still very limited, there has been a lot of progress in the past two decades. An excellent overview is presented in the review by Wakker and van Woerden (1997) and the update in Wakker et al. (1999). For a more recent summary on HVCs, we once again recommend the review by Putman, Peek and Joung (2012).

The origin of HVCs is still debated and appears to be a mixture of Galactic fountains (Houck and Bregman, 1990; Fraternali et al., 2013, see also Sect. 1.1.4), accretion from the intergalactic medium (IGM) (Kaufmann et al., 2006; Peek, Putman and Sommer-Larsen, 2008), and stripping from the Magellanic stream (e.g. Gardiner, 1999; Nidever, Majewski and Butler Burton, 2008).

They have typical line widths of $20 - 30\,\mathrm{km\,s^{-1}}$, indicative of a WNM component, but many show a narrow line width, cold component as well (Kalberla and Haud, 2006). This two-component structure is

expected for objects that are in pressure equilibrium with a hot Galactic halo (M. G. Wolfire et al., 1995).

Using higher resolution data or more sophisticated methods, the large HVC complexes (sizes of up to 1000 deg^2) split up into a conglomeration of individual objects (e.g. Stanimirović et al., 2008; Winkel et al., 2011; Venzmer, Kerp and Kalberla, 2012), depending on the resolution of the H I data that can range from 30′ in the LAB data to 4′ in the Galactic Arecibo L-band Feed Array HI Survey (GALFA-HI) data (Saul, Peek and Putman, 2014). Even higher angular resolutions can be obtained by using radio-interferometric data.

The presence of dust could shed some light on the origins of HVCs. If dust were to be found in one of these objects, we could conclude that this material had already been processed by the star-formation cycle of the MW and was thereby enriched with metals. Moreover, the composition of the dust particles, their sizes, and temperatures would allow us to better understand their environments. Up to now, there have been mixed findings about the presence of dust in HVCs. Early studies disclose tentative detections of dust in the FIR (Miville-Deschênes et al., 2005b; Peek et al., 2009). However, the importance of careful foreground modelling was soon recognised (Peek et al., 2009; Planck Collaboration, 2011 XXIV; Williams et al., 2012; Planck Collaboration, 2014 XXX; Lenz, Flöer and Kerp, 2016) and is the major source of uncertainty. Studies that focus on this modelling and on the estimation of the uncertainty in the HVC's dust content do not find any significant detections (Planck Collaboration, 2011 XXIV; Saul, Peek and Putman, 2014; Lenz, Flöer and Kerp, 2016).

Deep H I observations have revealed that HVCs can be detected down to the limit at which hydrogen is dominantly present in the neutral atomic phase (Braun and Thilker, 2004). Furthermore, the clouds observed in H I emission are also connected to absorption line systems (Richter et al., 2001; Nidever et al., 2010; Ben Bekhti et al., 2012). Moreover, their metallicities are mostly significantly sub-solar (Wakker et al., 1999; Fox, Savage and Wakker, 2006) and they can also be associated with a hot, ionised phase (Fox et al., 2004).

There are different observational hints that imply an interaction of a HVC with the ambient medium. One of these phenomena are velocity bridges (VB) that continuously connect gas at different radial velocities (Pietz et al., 1996; Lenz et al., 2015). They cover velocity gaps of 20-100 km s^{-1} and are suggestive of an energetic interaction between HVCs and IVCs at the disk–halo interface. Other possible origins are the inelastic collision of HVCs with the Galactic wind or with the WIM, which extends up to 1 kpc above the Galactic plane (Reynolds-layer, Reynolds, 1991).

Another indicator of HVC interactions is the head-tail (HT) morphology, shown in Fig. 1.5. The cometary shape of the cloud suggests ram-pressure interaction with an ambient medium and is observed for a large number of HVCs (Brüns et al., 2000; Stanimirović et al., 2006; Peek et al., 2007).

Even though there have been numerous studies to determine the distance to HVCs and HVC complexes, large uncertainties remain. Most commonly, the distance to a HVC is measured by using a halo star of known distance and looking for absorption features in the star's spectrum at the radial velocity of the HVC. This distance bracketing can result in the detection of an absorption line (i.e. the HVC is in front of the star) or a significant non-detection which requires very sensitive observations (i.e. the HVC is behind the star). Typically, optical Ca II, H, K and Na II lines are used for this analysis. This approach has been applied by different researchers to disclose the distance to HVCs and IVCs across the entire sky (Wakker, 2001; Wakker et al., 2007; Wakker et al., 2008; Thom et al., 2008). Unfortunately, the number of background objects is often very limited and hence the distance estimate across a HVC complex is only very sparsely sampled. Moreover, the uncertainties on the distance estimate derived with this approach are of the order of kpc. For most of the HVC complexes, distances between 2 and 15 kpc are found. A noteworthy exception is the Magellanic Stream and the Leading Arm (see also Fig. 1.4, labels MS and LA) which are at a distance of approximately 55 kpc, similar to the distance of the Large Magellanic Cloud (Nidever, Majewski and Butler Burton, 2008; Nidever et al., 2010).

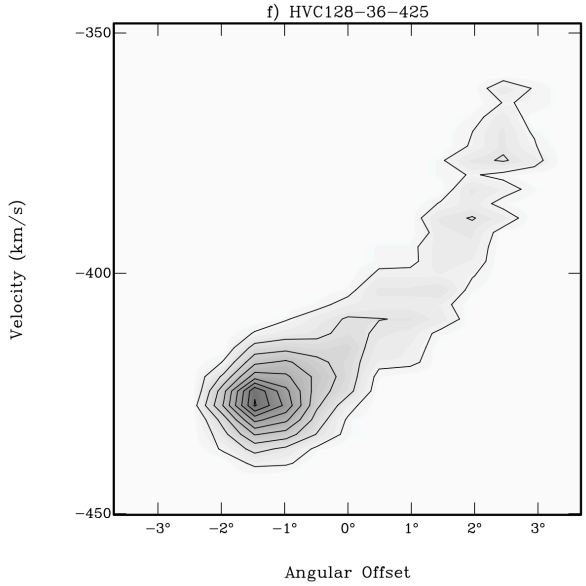


Figure 1.5: Example for a H I HVC with head-tail structure. The grey scale and the contours indicate the H I brightness temperature. Contour levels start at $T_B = 0.12\,K$ and increase in steps of $\Delta T_B = 0.12\,K$. The cometary shape is a strong indicator for ram-pressure interaction between the HVC and an ambient medium. This position-velocity diagram shows the radial velocity of the H I emission as a function of the positional offset. It is oriented along the velocity gradient axis. Image taken from Brüns et al. (2000).

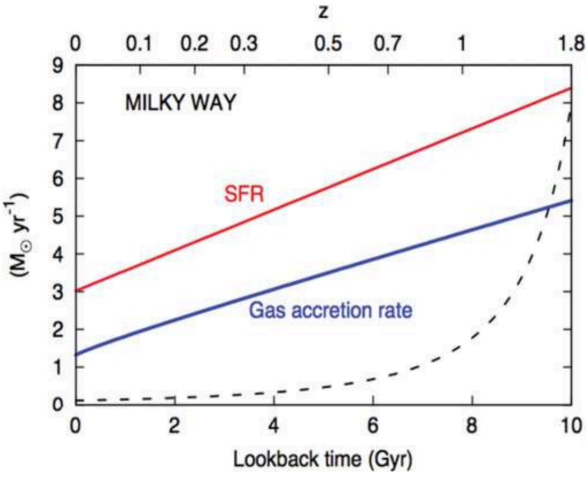


Figure 1.6: Star-formation history and accretion rate of the MW galaxy, illustrating that the ratio between gas accretion rate and star formation rate (SFR) are constant for most of the galaxy's lifetime. The dashed line corresponds to the SFR for a simulated closed-box scenario in which no accretion takes place. Image taken from Fraternali and Tomassetti (2012).

1.2 Gas flows and accretion in the Milky Way

Spiral galaxies in the local Universe have star-formation rates of $0.5 - 2\,M_\odot\,yr^{-1}$ with a value between 1.9 ± 0.4 and $4.0\,M_\odot\,yr^{-1}$ found for the MW galaxy (e.g. Diehl et al., 2006; Chomiuk and Povich, 2011). To balance this SFR, a constant inflow of gas is needed (Fig. 1.6, see also Fraternali and Tomassetti, 2012). Additionally, this material needs to have low metallicities to allow for the metallicities observed in the stellar population (G–dwarf problem, van den Bergh, 1962; Alibés, Labay and Canal, 2001; Schönrich and J. Binney, 2009). Some studies argue that recycled material from stellar mass loss can account for the sheer quantity of gas, but fail to solve the G–dwarf problem (Leitner and Kravtsov, 2011). In the following, we describe two processes that are thought to be responsible for re-fuelling the MW galaxy.

If this demand for new star-formation fuel is not met, stars will cease to form after a few Gyrs at most and the galaxy will slowly turn 'red and dead'. The summary by Fraternali (2014) contains further references and detailed explanations of this problem.

For other spiral galaxies, the cold gas accretion process has recently been studied by the Hydrogen

Accretion in LOcal GAlaxieS Survey (HALOGAS, Heald et al. 2011). The HALOGAS team used the Westerbork-Synthesis Radio Telescope (WSRT) to conduct very deep, high resolution H I observations of nearby spiral galaxies, probing the distribution and the kinematic of the cold, extraplanar gas (e.g. Kamphuis et al., 2013; Gentile et al., 2013). They find that the galaxies have a thick disk component, with scale heights of a few ($\sim$ 3) kpc. Moreover, the lag of the thick component for larger heights above the disk is estimated to be around $10\,\mathrm{km\,s^{-1}}$. These and other results allow to probe the different gas accretion processes in other galaxies, but the linear resolution is still very limited. Depending on the distance to the studied galaxy, it ranges between 0.5 kpc and 2 kpc. Hence, the MW is still a unique laboratory as it allows us to study these phenomena on much smaller scales ($\lesssim$ 1 pc).

1.2.1 Replenishment by high-velocity clouds

For decades after their discovery (Muller, Oort and Raimond, 1963), HVCs have been regarded as supplier of low-metallicity gas for the MW and other spiral galaxies. However, recent advancements in HVC research have led to distance estimates of around 10 kpc and mass estimates of less than $M_{\mathrm{H\,I}} = 10^6\,\mathrm{M_\odot}$ even for the largest HVC complexes (Wakker et al., 2008; Thom et al., 2008). Putman, Peek and Joung (2012) combined all these information on HVCs and come to an estimated accretion rate from HVCs of $\dot{M}_{\mathrm{accr}}^{\mathrm{HVCs}} = 0.08\,\mathrm{M_\odot\,yr^{-1}}$, with an upper limit of $0.4\,\mathrm{M_\odot\,yr^{-1}}$. This estimate already includes helium and the ionised gas phase. Hence, HVCs fall approximately one order of magnitude short of feeding the star formation in the Galactic disk.

Despite the insufficient accretion rate from HVCs, their influence in galaxy evolution should not be underestimated. Up to now, the individual contribution from the different processes that are responsible for the mixing of low-metallicity gas into spiral galaxies is not fully understood and hence HVCs might very well be a crucial component in this aspect of galaxy evolution.

The mixing of HVC halo gas with the disk is assumed to take place at the disk–halo interface, situated at an altitude of 0.8 to 1.5 kpc (Richter et al., 2001; Kalberla and Kerp, 2009). Because of the sub-solar metallicity of HVCs and the solar metallicities of most local clouds such as IVCs (Richter et al., 2001), the mixing must take place at this interface region.

Interaction of high-velocity clouds with the Galactic disk

During their passage through the Galactic halo, HVCs are disrupted by the interaction with the ambient medium (Heitsch and Putman, 2009). While much of the neural gas is stripped off of the infalling cloud or is ionised, some material remains neutral and impacts onto the disk. In the past decades, there have been numerous studies on these signatures of interaction of HVCs with the Galactic disk, some of which we describe in the following.

One region that is proposed to have undergone such an interaction in the past is the Gould Belt (Comeron and Torra, 1994, and references therein). Based on the morphology and the shape, size and formation timescales of the individual complexes of the Gould Belt, the authors conclude that an HVC impact is indeed the most likely origin of the Gould Belt system, a collection of different star-forming regions that is approximately 1 kpc across.

The Smith Cloud is a $10^6\,\mathrm{M_\odot}$ HVC, located in an advanced stage of accretion and situated only 2.9 kpc below the Galactic plane. It will reach the disk and interact with it in approximately 27 Myr (Lockman et al., 2008; Fox et al., 2016), contributing to its refuelling. The high metallicity of 0.5 solar measured in the Smith Cloud (Fox et al., 2016) indicates that its origin is not extragalactic, but is instead part of a Galactic fountain process (Sect. 1.2.2). While the metallicity alone could also be explained by stripping

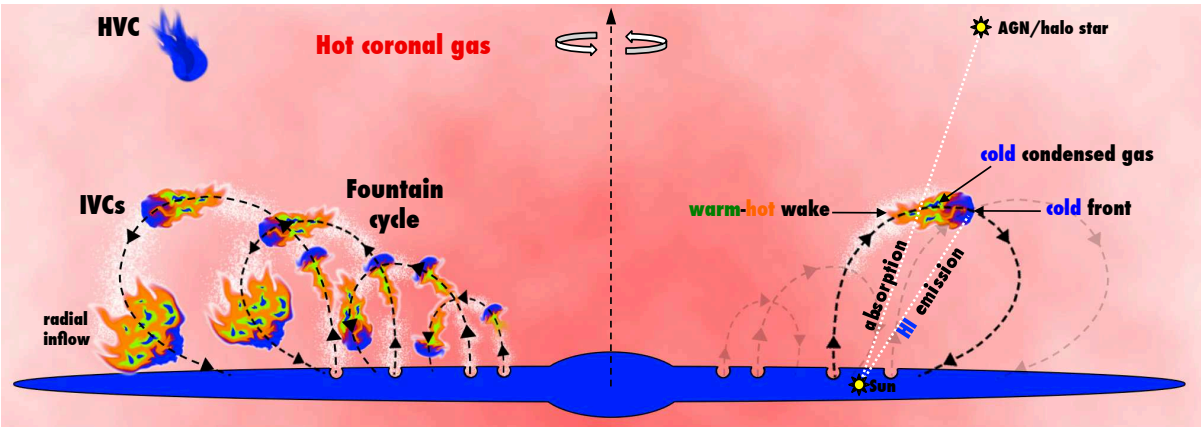


Figure 1.7: Schematic of a Galactic fountain process. **Left:** Gas is ejected towards the Galactic halo and hydrodynamical interactions between the clouds and the hot halo material initialise the cooling. Eventually, the ejected material will return to the disk, enriched by low-metallicity halo gas. **Right:** For an observer, multiple sight lines through the fountain object are available. The cold front of the cloud can be seen in H I emission and the hot, turbulent wake can be studied through absorption against a distant background source. Image taken from Marasco, Marinacci and Fraternali (2013)

gas from a low-metallicity galaxy such as the Magellanic clouds, the reconstructed trajectory of the Smith Cloud is indicative of an origin in the outer disk.

More examples of HVCs and HVC complexes that might have collided with the MW are given by Wakker and van Woerden (1997). Most of these clouds are located too far away from the plane of the galaxy to undergo an interaction at the disk–halo interface.

One observational hint on the interaction between HVCs and IVCs has been presented by Pietz et al. (1996): velocity bridges (VB) continuously connect these two cloud populations in velocity space and indicate that an inelastic collision has taken place. While present studies are focussed in the general direction of HVC complex C, VB have been found in other HVC complexes as well (Kerp et al., 1999; Lockman, 2003; Venzmer, Kerp and Kalberla, 2012; Henley, Shelton and Kwak, 2014).

1.2.2 Galactic fountains

The Galactic fountain process (Houck and Bregman, 1990) describes a feedback mechanism that is driven by the evolution of the high-mass stars in the MW galaxy. It is sketched in Fig. 1.7, taken from Marasco, Marinacci and Fraternali (2013). In this picture, gas is accelerated from the Galactic plane into the halo by stellar winds and supernova explosions. The gas is then ionised and lifted to heights of one to a few kpc above the disk (e.g. Melioli, Brighenti and D'Ercole, 2015, and references therein). Eventually, this gas will cool and fall back onto the Galactic plane as neutral gas clouds after approximately 10^8 yr.

This Galactic fountain process is thought to generate a major fraction of the IVC population that is observed. The mixing with the halo gas and the violent acceleration and deceleration of different gaseous phases leads to shattering of the dust grains in IVCs (Planck Collaboration, 2011 XXIV). Compared to dust in the Galactic plane, these grains are also significantly hotter ($T_{\mathrm{Dust}} \approx 20\,\mathrm{K}$) and have a lower dust opacity. Most of the HVCs cannot be the result of a fountain process because of their large distances above the plane or their high velocities (Marasco, Fraternali and J. J. Binney, 2012), even though there might be exceptions to this such as HVC complex C (Fraternali et al., 2015, see also Fig. 1.4).

Recently, the Galactic fountain process has been treated as a prime candidate to solve the issue of supplying the Galactic disk with sufficient amounts of cool, low-metallicity gas (Marasco, Fraternali

and J. J. Binney, 2012; Fraternali et al., 2013; Fraternali et al., 2015, see Fig. 1.7 for an illustration). At the interface between the Galactic disc and the halo, the ejected ionised, high-metallicity gas mixes with the hot, pristine halo material, triggering the cooling in the lower halo. The net metallicity of the gas is thereby reduced. This lower metallicity gas, enhanced in mass by the Galactic halo material, will eventually cool and fall back onto the Galactic disk. Within the individual fountain clouds (Fig. 1.7, right), we can observe the hot, ionised phase while the compressed head is already cooled (Marasco, Marinacci and Fraternali, 2013).

1.3 The relation of dust and gas

In the following, we outline the basics of the correlation between dust grains and gas in the ISM which is a cornerstone of this thesis. This relation allows us to expand our studies to denser, colder regions that are typically probed by FIR emission of dust.

The first correlation studies between dust and H I were conducted with data from the Infrared Astronomical Satellite (IRAS, Neugebauer et al. 1984). For the first time, the infrared cirrus clouds were detected and found to linearly correlate with the H I column density (Boulanger et al., 1996). These clouds are typically found at high Galactic latitudes and owe their name to their appearance, which is quite similar to clouds seen on earth. For a total hydrogen column density N_{H} and a FIR intensity I_ν at frequency ν, the correlation can be described as

$$I_\nu(\alpha, \delta) = \epsilon_\nu \, N_{\mathrm{H}}(\alpha, \delta) + N_\nu(\alpha, \delta) + R_\nu(\alpha, \delta) \tag{1.1}$$

Here, ϵ_ν is the dust emissivity per H-atom for a given frequency, N_ν is an instrumental noise term and R_ν describes the residual FIR emission that cannot be described by a simple linear model (see e.g. 1.4). The indices α and δ indicate sky coordinates.

At this point, it should be noted that the FIR intensity correlates with the *total* hydrogen column density N_{H}, but only the neutral atomic phase $N_{\mathrm{H\,I}}$ is traced through 21-cm spectroscopy. For a full analysis, we additionally need to include the ionised phase $N_{\mathrm{H\,II}}$ and the molecular phase $N_{\mathrm{H_2}}$:

$$N_{\mathrm{H}} = N_{\mathrm{H\,I}} + N_{\mathrm{H\,II}} + 2 \cdot N_{\mathrm{H_2}} \tag{1.2}$$

Molecular hydrogen consists of two hydrogen nuclei, hence the factor of 2.

The contribution from the ionised phase is considered to be small because the H II column densities are about an order of magnitude below that of the neutral atomic one. (Lagache et al., 1999; Lagache et al., 2000; Saul, Peek and Putman, 2014).

With large-scale H I surveys being readily available, astronomers usually focus on the correlation between H I and dust. Here, it has been found that the linear relation (Eq. 1.1) holds only up to H I column densities of about $N_{\mathrm{H\,I}} = 1 - 5 \times 10^{20}\,\mathrm{cm}^{-2}$ (Desert, Bazell and Boulanger, 1988; Reach, Wall and Odegard, 1998; Planck Collaboration, 2011 XXIV; Röhser et al., 2014; Lenz et al., 2015). Above this value of the column density, the molecular gas phase is thought to be formed. While the FIR intensity still scales with total hydrogen column density, the H I column density does not trace the total gas any more. This leads to FIR excess emission.

To quantify the correlation between dust and gas at higher column densities in the non-linear regime, the CO molecule is commonly used. By applying the X_{CO} conversion factor, the amount of molecular hydrogen can be inferred, and thus the total hydrogen column density can be reconstructed. Alternatively, the correlation between FIR intensity and H I column density can be extrapolated to higher column densities in order to estimate the contribution from $N_{\mathrm{H_2}}$ (Röhser et al., 2014) without having to rely on

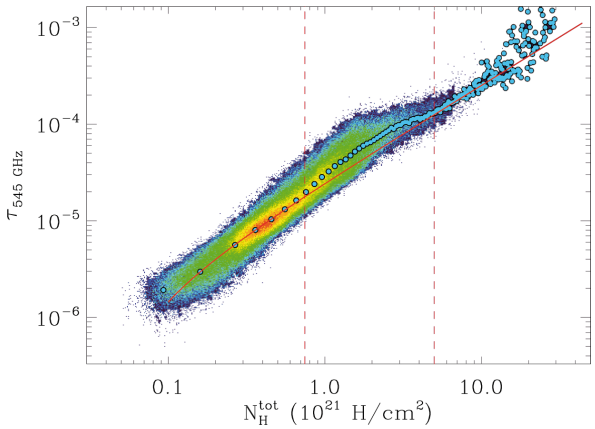


Figure 1.8: Correlation between the optical depth of dust at 545 GHz and total hydrogen column density for a large fraction of the sky. The colour scale indicates the number of data points on a log scale, red corresponds to the largest amount of data points. The blue circles show a binned representation of the correlation plot. In this double-log plot, most of these circles are at high column densities because the binning is performed on a linear scale. The red line shows the best fit for the low column densities, interpolated to the entire N_H range. For the highest column densities, a X_CO factor of $2.3 \times 10^{20}\,\mathrm{cm}^{-2}/(\mathrm{K\,km\,s}^{-1})$ is used to convert from CO line integral to H_2 column densities. At low and high total hydrogen column densities, the correlation is dominated by H I and CO, respectively. For intermediate column densities in between the vertical lines, the CO-dark molecular gas leads to FIR excess emission. This scatterplot covers data from 63% of the sky, based on data from *Planck*, GASS (Kalberla et al., 2010), and the CO-survey by Dame, Hartmann and Thaddeus (2001). Image taken from Planck Collaboration (2011 XIX).

CO data or the uncertain X_CO conversion factor (e.g. Lee et al., 2014).

By fixing the correlation of dust and gas at the low H I-dominated column densities and at the high CO-dominated column densities, FIR excess emission has been found at intermediate column densities between $N_\mathrm{H} = 5 \times 10^{20} - 5 \times 10^{21}\,\mathrm{cm}^{-2}$ (Planck Collaboration, 2011 XIX, see also Fig. 1.8). This excess emission is most commonly attributed to molecular gas that is not dense enough to form CO because of insufficient self-shielding. It is referred to as CO-dark molecular gas (Grenier, Casandjian and Terrier, 2005; M. G. Wolfire, Hollenbach and McKee, 2010; Planck Collaboration, 2011 XIX) .

A small fraction of this excess emission can also be attributed to optical thick H I (Fukui et al., 2014). Here, the H I column density is underestimated because of self-absorption. However, more recent studies have shown that this effect is not the dominant source of the dark gas (Lee et al., 2015). Nonetheless, a correction for H I self absorption should be applied when studying the correlation between dust and gas in regions of very high H I column densities ($N_\mathrm{H}^\mathrm{tot} \gtrsim 5 \times 10^{21}\,\mathrm{cm}^{-2}$).

To further explore the CO-dark molecular gas, other tracer molecules such as OH have been proposed. OH is formed in regions in which CO self-shielding is not efficient, but observations of OH require far more observing time to reach comparable sensitivities (Allen, Hogg and Engelke, 2015). Thus, it is useful for selected regions, but cannot be applied for spatially extended analyses of the dust and gas correlation.

As of today, the major source of uncertainty in the multiphase correlation between dust and the different hydrogen phases is the dust emissivity ϵ_ν, which is a function of illumination from the interstellar radiation field (ISRF), grain sizes, and grain temperatures, and other effects (Planck Collaboration, 2014 XVII). It varies not only intrinsically across a single cloud, but it also deviates for different clouds across the sky (Planck Collaboration, 2011 XIX; Planck Collaboration, 2014 XVII). While the assumption of a constant dust emissivity is a fair approximation for individual fields (Wakker and Boulanger, 1986; Miville-Deschênes et al., 2005b; Saul, Peek and Putman, 2014; Planck Collaboration, 2011 XXIV; Lenz et al., 2015), more advanced approaches need to be applied for high-sensitivity, large-scale models (e.g. Lenz, Flöer and Kerp, 2016).

1.4 The cosmic infrared background

The extragalactic background light (EBL) is the diffuse background radiation, emitted by the stellar population of galaxies at different redshifts (e.g. Hauser and Dwek, 2001). Most of this light stems from galaxies with redshifts ranging between 1 and 3 because the SFR peaks at a redshift of approximately $z = 1.5$ (Madau, Pozzetti and Dickinson, 1998). Commonly, the EBL is defined to cover wavelengths from 100 nm to 1000 μm, which corresponds to the UV, visible, and infrared parts of the electromagnetic spectrum. The cosmic infrared background (CIB) is the IR part of this emission and contains approximately half of its total energy (Dole et al., 2006).

Unlike in high-resolution studies conducted in the near-infrared, the individual sources cannot be resolved in the FIR. Here, we are mostly limited to space-based observatories such as *Planck* with resolutions of several arcminutes. Therefore, the CIB at long wavelengths is studied in a statistical manner via the anisotropies of the random field. These fluctuations trace the large-scale distribution of the star-forming galaxies across cosmic time and the distribution of the dark matter haloes in which these galaxies are located. Hence, the CIB can be used as a probe of gas accretion, star formation and the correlation of dark matter haloes and luminous galaxies.

The CIB was first described by Partridge and Peebles (1967) and is made up mostly from the dust in heavily star-forming galaxies, with a smaller contribution from the emission of Active Galactic Nuclei (AGN) (see Lagache, Puget and Dole, 2005, for a review). Recent pioneering work was made with *Herschel*-SPIRE data (Amblard et al., 2011) and with *Planck* data (Planck Collaboration, 2011 XVIII) and is further expanded by the studies which we present in the following.

With the help of these modern instruments, it has become feasible to probe the interplay between the luminous, star-forming galaxies and the dark matter halos. To achieve this, the CIB emission is compared to the gravitational lensing of the cosmic microwave background (CMB). Here, the CIB traces the luminous galaxies and the lensing traces the dark matter halos. The Planck Collaboration (2014 XVIII) found a correlation of 80% at the 19σ level, including systematic uncertainties, and are able to constrain the star formation rate density (SFRD) for individual redshift bins between $z = 1$ and $z = 7$.

In another work by the Planck Collaboration (2014 XXX), the focus was on the study of several smaller ($\sim 30\,\mathrm{deg}^2$) fields for which the CIB anisotropies were measured and the information on the clustering of dusty, star-forming galaxies was extracted. By comparing these data to different halo models, they constrained the star formation history (SFH) up to a redshift of $z = 2.5$. Moreover, it allows to study the temperature evolution of CIB galaxies and the mass at which dark matter (DM) haloes are most efficient at forming stars.

While these results are very impressive and demonstrate the capabilities of data sets such as those obtained with *Planck*, these previous studies are still limited by systematic uncertainties. One of the most challenging parts is the cleaning of the FIR intensity maps for the different foreground components to extract the very faint CIB anisotropies. These components include the CMB, the Sunyaev-Zeldovich effect (SZ, Sunyaev and Zeldovich 1970), extragalactic sources, and most importantly, Galactic thermal dust. For the component separation that has been applied to the CMB for decades, astronomers use the spectral behaviour of the different components to disentangle them and to generate maps of the individual components (e.g. Planck Collaboration, 2015 IX, for the latest application of such an approach).

For the CIB, this is unfortunately not possible because the spectral energy distribution (SED) of dust in the MW galaxy and of dust from the distant galaxies that make up the CIB are very similar. The SED can be described by a modified blackbody with a mean temperature of approximately 14.24 K (Planck Collaboration, 2011 XIX) for the Galactic dust and $24.4 \times (1 + z)^{0.36}$ K for the dust in the CIB galaxies (Planck Collaboration, 2014 XXX).

Instead, one can estimate the foreground dust emission by templates that are based on other tracers

of the ISM. For this approach, the most common technique is to use observations of H I and CO as an estimator of the foreground emission in the FIR intensity maps (e.g. Planck Collaboration, 2014 XXX), using the aforementioned correlation (Eq. 1.1).

Lastly, Planck Collaboration (2016 XLVIII) used the statistical spatial information of the foreground cirrus and the background CIB to distinguish the two components. They find that their results are in agreement with previous studies that use gas-phase templates to remove the CIB at the 5% level.

1.5 Motivation and outline

Over the past two decades, there has been a lot of progress in terms of understanding IVCs and HVCs and their role in galaxy evolution. Nonetheless, there have been only very limited efforts to actually study ongoing accretion and to understand which role the intermediate-velocity molecular clouds (IVMCs) play as the coldest, densest component of this process. Moreover, we aim to shed light on interactions of HVCs with IVCs and thus to infer the properties of the clouds such as their origin, distances, and trajectory. Furthermore, we can use such a system to study the presence of different gaseous phases and the collisionally-induced formation of molecular hydrogen.

We have outlined the work of some researchers who study the potential dust content in HVCs. These studies are often limited by very simple foreground models and an insufficient estimation of the uncertainties. By carefully modelling the foreground FIR intensity and verifying the methods through simulations, we can derive a more robust and more accurate estimation of the dust content of HVCs. This is particularly helpful to constrain their origin and their role in galaxy evolution.

The methods that are used to remove the foreground dust emission for the HVC are the foundation for the CIB studies that rely on a high accuracy and reliability. We aim at creating an accurate picture of the MW galaxy on large scales, eventually allowing us to visualise the CIB across large fractions of the sky. Thereby, we can enhance our knowledge of the Universe at high redshift ($z = 1 - 7$). Here, we use the latest generation of H I and FIR data sets (Planck/EBHIS/GASS) which allow us to open a new window on this extragalactic background component and to use it for the evaluation of cosmological models. Furthermore, it is essential to apply modern methods of applied statistics and machine learning to access the full potential of the high-quality data.

This thesis is structured as follows: In the present Chapter, we have described the basic principles and the historical framework of this thesis, covering the different ISM phases, the accretion processes in the MW, the correlation of dust and gas, and the CIB. Chapter 2 presents the data sets that are used throughout this work. Chapter 3 is based on Lenz et al. (2015) and describes the interaction of a HVC with an IVC at the disk–halo interface. In Chapter 4, we present the results by Lenz, Flöer and Kerp (2016) who investigate a cold, dense HVC for its potential dust content. The methods and analyses are refined in Chapter 5 and applied to the full sky, uncovering the CIB at large scales and comparing the results to previous studies and cosmological models. We conclude in Chapter 6.

CHAPTER 2

Data

In the following, we describe the data sets that are used throughout this work. We start off with the data of neutral atomic hydrogen in Sect. 2.1 and present the FIR data in Sect. 2.2. Table 2.1 gives an overview on the properties of these H I and FIR data. In Sect. 2.3 we give a brief introduction to the Hierarchical Equal Area isoLatitude Pixelation of the sphere (HEALPix) which is used frequently in this work.

2.1 Atomic neutral hydrogen

From the late 1990s until around 2010, the fundamental database for Galactic H I was the Leiden/Argentine/Bonn Survey (LAB, Kalberla et al. 2005). This survey is a joint project, merging the data from the southern hemisphere from the Instituto Argentino de Radioastronomía Survey (Arnal et al., 2000; Bajaja et al., 2005) and the Leiden/Dwingeloo Survey (Hartmann and Burton, 1997).

Despite the great legacy value of the LAB survey, the astronomical community asked for a modern H I survey, conducted with a 100-m class telescope. This was motivated by the incomplete beam-by-beam sampling of the LAB and the fact that many structures in the MW are smaller than the 30′ angular resolution of the LAB data (Lockman, 2002; Stanimirović et al., 2006; Peek et al., 2007). Here, a new full-sky survey can greatly expand the insights that have been gained by individual high-resolution H I observations and by the large-scale, but low-resolution LAB. In the following, we present the two projects that meet this demand.

2.1.1 Effelsberg-Bonn H I Survey

The Effelsberg-Bonn H I Survey (EBHIS, Winkel et al. 2010; Kerp et al. 2011; Winkel et al. 2016a) is a full-sky survey of the northern hemisphere with declination $\delta \geq -5°$ in H I, using the Effelsberg 100-m telescope. EBHIS is the first H I survey which simultaneously covers Galactic H I with high spectral resolution, and the extragalactic sky out to a redshift of $z \sim 0.07$. Figure 2.1 shows the Galactic H I column density across the full northern sky as seen with EBHIS.

The receiver bandwidth of 100 MHz is split into 16,384 spectral channels, using state-of-the-art field-programmable gate arrays (FPGAs, Stanko, Klein and Kerp, 2005; Klein et al., 2006). This results in a spacing of spectral channels of 1.29 km s^{-1} at a velocity resolution of 1.45 km s^{-1} which is well-suited to spectrally resolve the dynamics of the narrow line width features of the CNM.

EBHIS was measured on-the-fly, meaning that the telescope was continuously slewed. The observational setup was chosen such that the telescope scans along right ascension (R.A.) for a constant declination (Dec) and data are recorded every 500 ms. To increase the survey speed of the instrument,

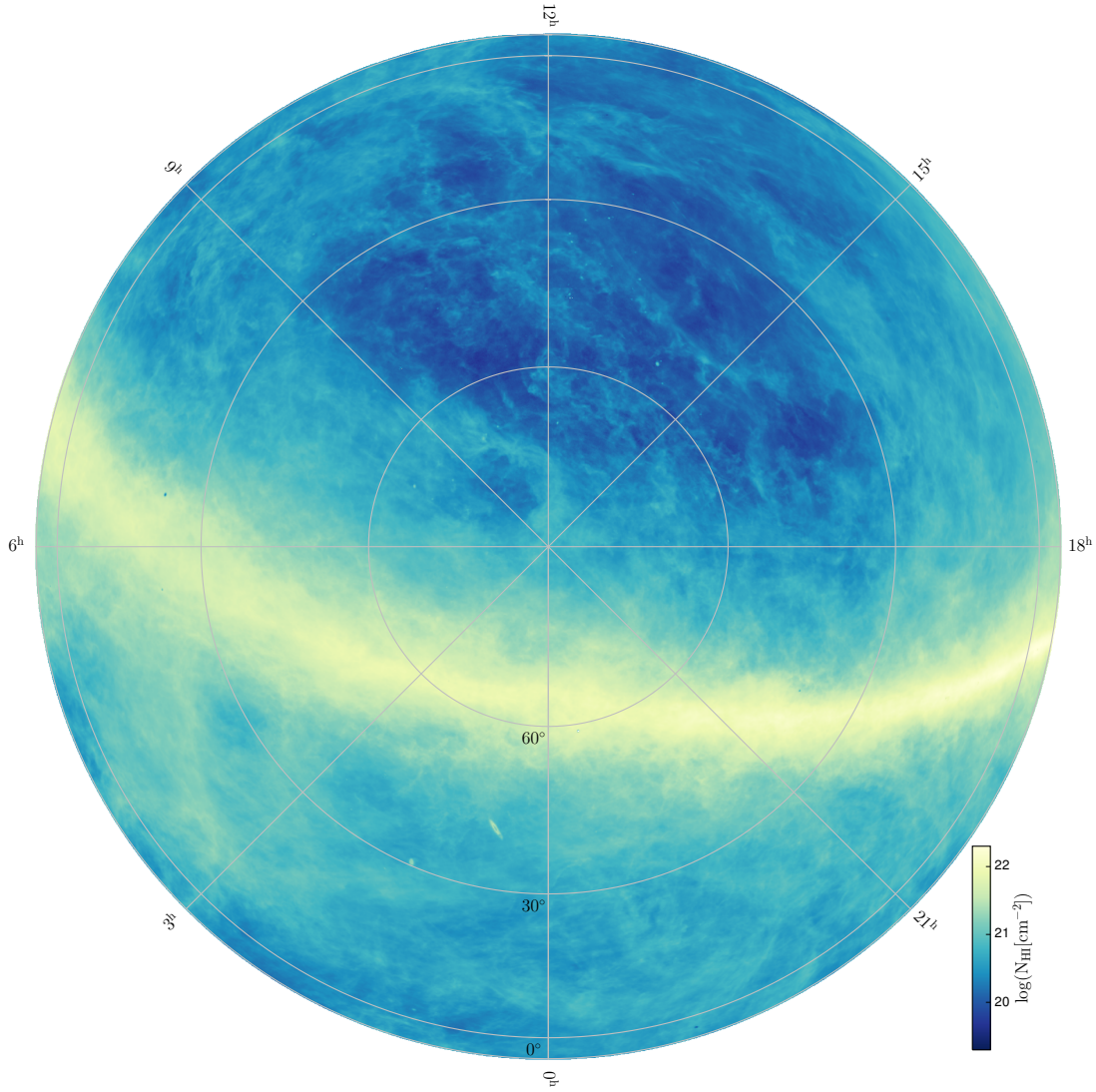


Figure 2.1: EBHIS H I column density map, integrated over the velocity range $-600\,\mathrm{km\,s^{-1}} \leq v_{\mathrm{LSR}} \leq +600\,\mathrm{km\,s^{-1}}$. This image shows the northern hemisphere, with Galactic longitude l increasing clockwise and Galactic latitude b increasing towards the centre of the map. Image taken from Winkel et al. (2016a).

Table 2.1: Properties of the H I and FIR data sets used in this study.

Survey	λ	ν [GHz]	θ [arcmin]	σ_{RMS}	σ_{cal}	Reference
EBHIS	21 cm	1.420	10.8	108 mK	$\lesssim 3\,\%$	(1)
GASS	21 cm	1.420	16.1	57 mK	$\lesssim 3\,\%$	(2)
IRIS	100 μm	2998	4.30	0.06 MJy sr^{-1}	13.5 %	(3)
Planck	350 μm	857	4.63	0.014 MJy sr^{-1}	$\lesssim 10\,\%$	(4)
Planck	550 μm	545	4.84	0.013 MJy sr^{-1}	$\lesssim 10\,\%$	(4)
Planck	850 μm	353	4.86	0.011 MJy sr^{-1}	$\lesssim 1.2\,\%$	(4)

References: (1) Winkel et al. (2016a); (2) Kalberla and Haud (2015); (3) Miville-Deschênes and Lagache (2005); (4) Planck Collaboration (2013 VI)
Notes: λ and ν are the reference wavelengths and frequencies. θ is the half-power beam width (HPBW). σ_{RMS} is the sensitivity per beam solid angle, and σ_{cal} is the calibration uncertainty. For the spectroscopic EBHIS and GASS, σ_{RMS} is calculated for a velocity resolution of 1.00 km s^{-1}.

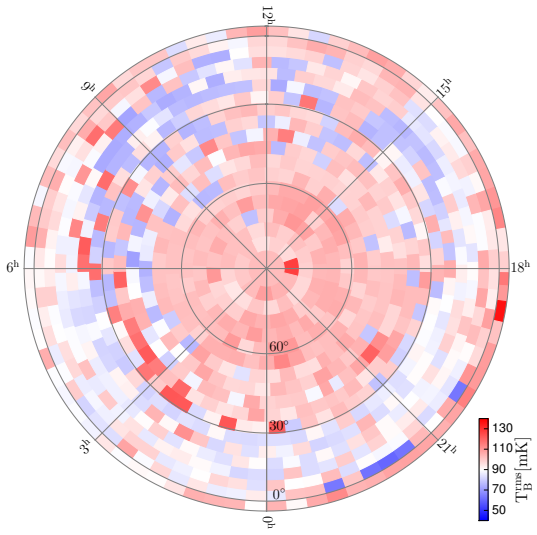


Figure 2.2: EBHIS noise map, showing the mean root mean square (RMS) for each of the 945 fields. The mean noise for each field depends mostly on the season and on the elevation at which the observation is carried out. Image taken from Winkel et al. (2016a).

a novel L-band 7-feed receiver array was used. For the first coverage that is published in Winkel et al. (2016a), a total of 945 fields with a size of $5° \times 5°$ were observed. Unlike the LAB, EBHIS is spatially fully Nyquist sampled (Shannon, 1949). This means that the spatial density of measurements is sufficiently high to reconstruct the full signal. For the LAB, this was not the case as it was only sampled beam-by-beam. The average noise for each of the fields is shown in Fig. 2.2. The differences of more than a factor of two have various reasons: The total observation time for each field was slightly modified through the course of project, resulting in a varying sensitivity. Moreover, the elevation of the observation and thus the angular separation from the ground reduces the influence of thermal ground radiation, increasing the sensitivity. Lastly, the temperature variations between summer and winter affect the noise level as well. For the entire EBHIS data set, the median noise is $\sigma_{\mathrm{RMS}} \approx 90\,\mathrm{mK}$.

For the observations, several constraints had to be applied to ensure the highest possible data quality. A mobile-communication tower had to be avoided at all times, restricting the available Azimuth-range of observations. Additionally, the angular distance to the Sun had to be kept as large as possible to avoid solar interference (see Barnes, Briggs and Calabretta, 2005, for the implications of this effect for the Parkes telescope).

In terms of data reduction, the software was developed completely from scratch. This includes the calibration (Winkel, Kraus and Bach, 2012), the correction for stray radiation (SR) (Kalberla, Mebold and Reich, 1980a; Kalberla, Mebold and Velden, 1980b), the baseline modelling, radio frequency interference (RFI) mitigation (Flöer, Winkel and Kerp, 2010), and the gridding (Winkel, Lenz and Flöer, 2016b). The individual steps are described in Winkel et al. (2010) and some updates are presented in Winkel et al. (2016a).

The observations started in the autumn of 2008 and the first coverage was completed in 2013. In 2016, the first data release was made available to the public (Winkel et al., 2016a). The full spectroscopic data with a total size of approximately 50 GB can be accessed via the Strasbourg astronomical Data Center (CDS)[1] in different formats and projections. Additionally, the H I column density can be queried on skyview[2].

A second sky coverage for EBHIS is well underway with approximately 50% of the sky north of $\delta = 30°$ already observed. Here, the scan direction is orthogonal to the first coverage, which makes it feasible to conduct basket-weaving (Winkel, Flöer and Kraus, 2012). This technique is used to remove the imprint of the scanning pattern on the data by minimising the signal difference at the crossing points. The second coverage will not only improve the sensitivity by a factor of $\sqrt{2}$, but will also help to address systematics such as residual SR, poor baseline fits, and RFI.

2.1.2 Parkes Galactic All Sky Survey

The Parkes Galactic All Sky Survey (GASS, McClure-Griffiths et al. 2009; Kalberla et al. 2010; Kalberla and Haud 2015) was conducted with the Parkes 64-m dish and covers Galactic H I over the entire southern hemisphere up to a declination of $\delta \leq 1°$. Using a 13-feed receiver array, the observations took place in 2005 and 2006. The entire sky was observed twice, with scan directions along Right Ascension and Declination, respectively.

The first data release (McClure-Griffiths et al., 2009) describes the observations and the initial data reduction. It is followed by a second data release (Kalberla et al., 2010) which improves the data through advanced RFI mitigation, the removal of SR, and the proper modelling of the instrumental baselines. The very recent third data release (Kalberla and Haud, 2015) examines remaining problems and explores the Gaussian decomposition (e.g. Haud, 2000) of the GASS data to identify them. Among the changes for

[1] http://cdsarc.u-strasbg.fr/viz-bin/qcat?J/A+A/585/A41
[2] http://skyview.gsfc.nasa.gov

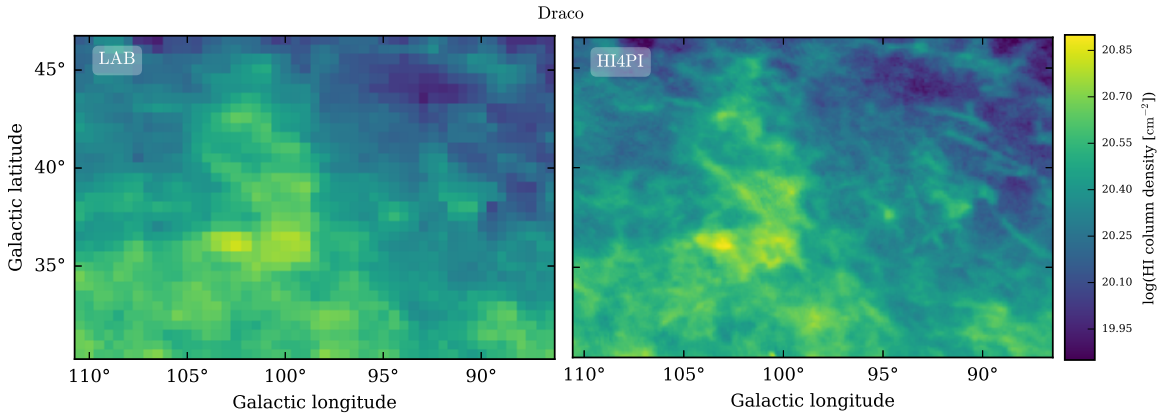


Figure 2.3: Comparison of LAB and EBHIS data in the framework of the HI4PI data release. The regions shows the total H I column density of the Draco nebula and illustrates the superior resolution of the EBHIS/HI4PI (right panel) compared to the LAB data (left panel). Image taken from HI4PI collaboration (2016, submitted)

this last release is an improved baseline solution, the removal of residual RFI, and the de-striping of the data.

Individual sight lines and small data cubes for the third GASS data release are available online[3]. Additionally, the first data release of the entire GASS column density and other data products can be accessed[4]. In the future, the full spectroscopic data will be made available online in various formats, similar to the way in which the EBHIS data are made public.

2.1.3 HI4PI Survey

The HI4PI Survey (HI4PI collaboration, 2016, submitted) merges the EBHIS and GASS data, creating a new data base for Galactic H I to replace the LAB. In addition to the two individually published surveys, its most important feature is a coherent map of the hydrogen column density $N_{\mathrm{H\,I}}$. This map is not only relevant for the radio astronomical community, but an estimate of the foreground level is of great importance for other areas in astronomy that wish to study extragalactic sources.

HI4PI covers the full sky (12.56 sr or 4π) with an overlap area of the two surveys of approximately 5 degrees in declination along the equatorial plane. For this area, the two surveys are linearly interpolated. The HI4PI column density map will be available in their original angular resolution (16.1′ for the southern hemisphere, 10.8′ for the northern one) and with a consistent GASS resolution of 16.1′.

Fig. 2.3 shows the Draco nebula as seen in the LAB and in the HI4PI data, illustrating the superior resolution of the new data set. Due to the different projections and formats in which HI4PI will be published, it can easily be used as a drop-in replacement to the LAB data.

2.2 Far-infrared data

To complement the data of the gaseous phases that are traced via the 21-cm line, we explore the state-of-the-art dust data for the full sky as obtained from the *Planck* mission (Sect. 2.2.1) and the Improved

[3] https://www.astro.uni-bonn.de/hisurvey/AllSky_profiles/index.php
[4] http://www.atnf.csiro.au/research/GASS/index.html

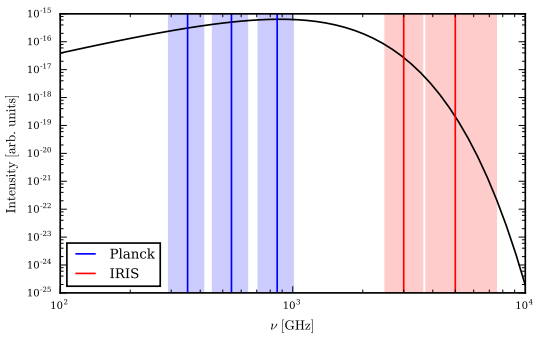


Figure 2.4: Planck and IRIS frequency coverage of dust emission, compared to dust emission which follows a blackbody with $T_{\mathrm{Dust}} = 15\,\mathrm{K}$ (black line). The shaded areas indicate the bandwidth of the individual detectors.

Reprocessing of the IRAS Survey (IRIS, Miville-Deschênes and Lagache 2005) (Sect. 2.2.2).

Fig. 2.4 shows the frequency coverage of the high-frequency *Planck* receivers and the two IRIS frequencies. The blackbody corresponds to dust emission with $T_{\mathrm{Dust}} = 15\,\mathrm{K}$ and illustrates how the two individual infrared missions complement each other and are ideally suited to probe the FIR emission of dust.

2.2.1 *Planck* mission

The *Planck*[5] satellite (Tauber et al., 2010; Planck Collaboration, 2011 I) is a mission of the European space agency, dedicated to study the early Universe and its evolution. It was launched on May 14th 2009 and, after arriving at the Lagrangian point $\mathcal{L}2$, has surveyed the sky continuously from August 13th 2009 until October 23rd 2013, completing almost five full coverages of the entire sky.

The *Planck* satellite carried two instruments: the High Frequency Instrument (HFI) and the Low Frequency Instrument (LFI). The former covers three frequencies between 30 GHz and 70 GHz and the latter covers six frequencies between 100 GHz and 857 GHz, yielding an unprecedented view on the full sky FIR emission. For this work, the channels at 353 GHz, 545 GHz, and 857 GHz are of main interest because Galactic thermal dust and the CIB dominate the spectral energy distribution at these frequencies.

The *Planck* mission is not only designed to study the total intensity and polarisation of the CMB, but also galaxy clusters through the Sunyaev-Zel'dovich effect (Sunyaev and Zeldovich, 1970), AGN, the gravitational lensing of the CMB, and the foreground emission from the Milky Way. At the time of writing, this amounts to a total of more than 150 publications by the *Planck* collaboration and additional publications from researchers from all around the globe who have benefited from the open data access.

Among the advantages of the *Planck* mission are the higher angular resolution, but more importantly the large frequency coverage with nine channels. The predecessors, Cosmic Background Explorer (COBE, Boggess et al. 1992) (Boggess et al., 1992) and Wilkinson Microwave Anisotropy Probe (WMAP, Bennett et al. 2003) (Bennett et al., 2013), were designed with only three and five frequencies, respectively. This improved spectral coverage of the CMB allowed the *Planck* team to perform a separation of the diffuse Galactic foregrounds at a new level, using a Bayesian framework (Planck Collaboration, 2015 X) to disentangle the individual contributions. Each of these components is available as a full-resolution image, including synchrotron emission, free-free emission, the CMB, CO, and Galactic thermal dust emission.

The different publications, from the earliest bluebook to the latest results, are available through the

[5] Planck (http://www.esa.int/Planck) is a project of the European Space Agency (ESA) with instruments provided by two scientific consortia funded by ESA member states and led by Principal Investigators from France and Italy, telescope reflectors provided through a collaboration between ESA and a scientific consortium led and funded by Denmark, and additional contributions from NASA (USA).

ESA website[6]. The different data products can be accessed through the Planck Legacy Archive[7].

Far-infrared intensity

For this work, we only use the FIR intensity maps at 353 GHz, 545 GHz, and 857 GHz. The publicly available maps are corrected for Zodiacal light (Kelsall et al., 1998), but not for the emission from the CMB. An overview of the properties of these maps is given in Planck Collaboration (2015 VIII, Table 6)

Throughout this work, we use MJy sr^{-1} as the standard unit for FIR intensities. Another common unit in this field is the differential CMB temperature K_{CMB} which is particularly useful when working with CMB data. An overview of the conversion factors is given in Planck Collaboration (2014 XXX, Table 1).

Carbon monoxide

Among the components extracted from the diffuse component separation (Planck Collaboration, 2015 X) are different line transitions in the FIR regime. In the 100, 217, and 353 GHz *Planck* HFI channels, emission from the CO(1–0), (2–1), and (3–2) transition is well detected. Some other line transitions from HCN, CN, HCO+, and other molecules contribute to this component of line transitions, but can be isolated and treated individually, independent of the CO emission.

Due to the design of the *Planck* mission, the CO map published by the *Planck* collaboration is a measure of the velocity-integrated CO emission and is not resolved spectroscopically. Other surveys such as the one conducted by Dame, Hartmann and Thaddeus (2001) are dedicated mappings of the CO line with spectrometers over large fractions of the sky ($\lesssim 20\,\%$), allowing to infer the radial velocity of the CO molecules. However, the prime advantage of the *Planck* data is the coverage of the full sky which is not available for other data sets of CO.

Cosmic microwave background

The ultimate goal of the *Planck* mission is to provide maps of the CMB with the highest possible quality, allowing to constrain the values of the cosmological parameters with high precision. Moreover, it is possible for the first time to measure the signal from the cosmological polarisation, allowing unprecedented limits for example on the amplitude of primordial non-Gaussianity through the E-mode polarisation. Four different map-making algorithms are employed to achieve this, each with different advantages and caveats which allows the users to select the map that suits their purpose in the optimal way.

A very detailed description of this procedure is given by Planck Collaboration (2015 IX), with many topics that expand earlier investigations of the CMB component separation (e.g. Planck Collaboration, 2014 XV). However, this work is still ongoing, especially in order to properly measure the polarisation of the E- and B-modes of primordial polarisation (Planck Collaboration, 2015 XI).

For this thesis, the CMB is mostly negligible because the investigated frequencies are dominated by thermal dust and the CIB signal. While the CMB contribution to the total intensity drops rapidly above 100 GHz, thermal dust and the CIB peak around 1000 GHz. Nonetheless, we subtract the CMB signal for the 353 and 545 GHz frequency maps because some residual emission might lead to increased uncertainties or systematics in the subsequent analysis (Planck Collaboration, 2014 XVII, their Fig. C.1).

[6] http://www.cosmos.esa.int/web/planck/publications
[7] http://pla.esac.esa.int/pla/

2.2.2 Improved Reprocessing of the IRAS Survey

The Infrared Astronomical Satellite (IRAS, Neugebauer et al. 1984) (Neugebauer et al., 1984) was the first mission to conduct an all-sky survey in the FIR regime with frequencies of 24 000, 12 000, 5 000, and 3 000 GHz (corresponding to wavelengths of 12, 25, 60, and 100 μm). The satellite was launched on January 25th 1983 and scanned the sky for ten months. It had an enormous scientific impact on many different fields in astrophysics, particularly on the understanding of the infrared cirrus clouds that are the topic of the present work. Focussing on the 2998 and 5000 GHz bands, it is a very suitable complement for the *Planck* HFI data which covers frequencies between 100 and 857 GHz (see Fig. 2.4). Moreover, the angular resolution of both data sets of approximately 5′ match nicely (Table 2.1), providing the astronomical community with a full-sky data set in the FIR.

In order to fully utilise the IRAS data from 1984 in combination with the state-of-the-art *Planck* data, several corrections had to be applied to the 30-year old IRAS maps. This was addressed by Miville-Deschênes and Lagache (2005) in the Improved Reprocessing of the IRAS Survey (IRIS, Miville-Deschênes and Lagache 2005). Here, the calibration, the zero level of the data, the de-striping of the maps, and the subtraction of zodiacal light of the original IRAS data was improved. For the calibration and zero level, data from the Diffuse Infrared Background Experiment (DIRBE) aboard the COBE satellite were used. The full-sky data set which is used in this work is publicly available online[8] in different formats.

2.3 HEALPix

The Hierarchical Equal Area isoLatitude Pixelation of a sphere (HEALPix)[9] is a hierarchical pixelation scheme for full-sky data sets in which each pixel covers the same surface area or solid angle (Górski et al., 2005). It was developed for high resolution, multi-frequency CMB missions and was heavily used for the WMAP data and now for the *Planck* data. For these data sets, tasks involving the Fourier transform (harmonic decomposition, power spectrum estimation), object detection, and nearest-neighbour searches need to be implemented fast and efficient, giving the increasing size of data samples over the past decades. Over the course of 20 years, this size has grown from ~6000 pixel sky maps at three wavelengths for COBE data to ~50 million pixel sky maps at nine wavelengths for the *Planck* mission.

Developed by the CMB community, HEALPix has also advanced to other areas of astronomy and is widely used in the processing of the EBHIS data (Winkel, Lenz and Flöer, 2016b). Moreover, this format has been used dominantly in the present work.

Fig. 2.5 illustrates the hierarchical nature of HEALPix: Each of the 12 base pixels is divided into another 4 pixels for each increment in resolution. This is an essential requirement for very large data sets because points that are nearby on the sphere are also nearby in the tree structure of the data set. This is a prerequisite for fast look-ups of nearest neighbours which is essential for many algorithms on the HEALPix grid (e.g. Winkel, Lenz and Flöer, 2016b). Moreover, the fact that each pixel covers the same area ensures that sky signals are sampled without any directional dependence. Lastly, the pixels in HEALPix are distributed along lines of constant latitude. This iso-latitude distribution is necessary for all analyses involving spherical harmonics.

For numbering the pixels within the HEALPix framework, two different schemes are available: The ring ordering simply sorts the pixels from north to south along each iso-latitude ring. This is internally used for any computations that require the spherical harmonics such as the estimation of the angular

[8] http://www.cita.utoronto.ca/~mamd/IRIS/IrisOverview.html
[9] http://healpix.sourceforge.net

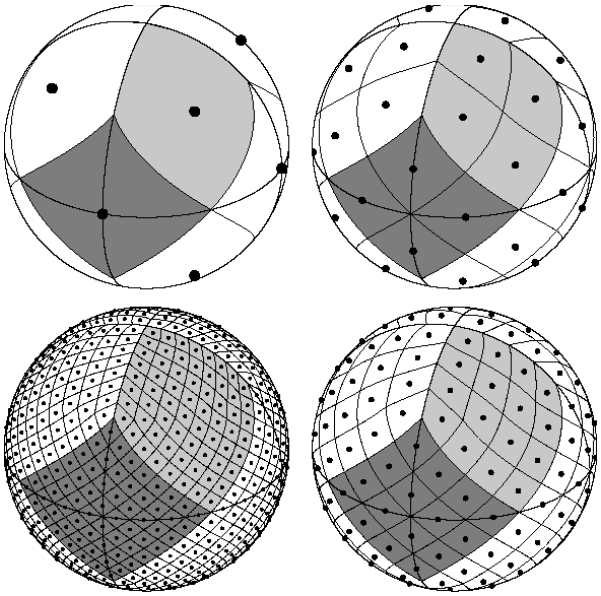


Figure 2.5: Orthographic projection of the same base-resolution spherical tessellations. For each increment in resolution, the pixels are divided into another four pixels, generating a hierarchical structure. Figure taken from Górski et al. (2005).

power spectrum. The nested ordering reflects the hierarchical tree structure of the framework and can be used for nearest neighbour searches.

Software packages for HEALPix exists in different programming languages such as IDL, C++, Python, and Fortran. For this work, most of the software was developed in Python, hence the reference implementation healpy[10] was used.

[10] http://healpy.readthedocs.org/en/latest/

CHAPTER 3

Far-infrared excess emission as a tracer of disk-halo interaction

This Chapter is based on the content published in Lenz et al. (2015), Far-infrared excess emission as a tracer of disk-halo interaction, A&A 573, A83, 2015, reproduced with permission © ESO.

Abstract

Given the current and past star-formation in the Milky Way in combination with the limited gas supply, the re-fuelling of the reservoir of cool gas is an important aspect of Galactic astrophysics. The infall of H I halo clouds can, among other mechanisms, contribute to solving this problem.

We study the intermediate-velocity cloud IVC135+54 and its spatially associated high-velocity counterpart to look for signs of a past or ongoing interaction. Using the Effelsberg-Bonn H I Survey data, we investigated the interplay of gas at different velocities. In combination with far-infrared Planck and IRIS data, we extended this study to interstellar dust and used the correlation of the data sets to infer information on the dark gas.

The velocity structure indicates a strong compression and deceleration of the infalling high-velocity cloud (HVC), associated with far-infrared excess emission in the intermediate-velocity cloud. This excess emission traces molecular hydrogen, confirming that IVC135+54 is one of the very few molecular halo clouds. The high dust emissivity of IVC135+54 with respect to the local gas implies that it consists of disk material and does not, unlike the HVC, have an extragalactic origin.

Based on the velocity structure of the HVC and the dust content of the IVC, a physical connection between them appears to be the logical conclusion. Since this is not compatible with the distance difference between the two objects, we conclude that this particular HVC might be much closer to us than complex C. Alternatively, the indicators for an interaction are misleading and have another origin.

3.1 Introduction

To constantly form stars, disk galaxies such as the Milky Way need to acquire gas across cosmic times to re-fuel their gaseous reservoir (e.g. Putman, Peek and Joung, 2012). The required mass flow needs to be in excess of or equal to the star formation rate (SFR) of $1.9 \pm 0.4\,\mathrm{M}_\odot\,\mathrm{yr}^{-1}$ (Chomiuk and Povich, 2011). Leitner and Kravtsov (2011) argued that recycled gas from stellar mass loss can account for much

of the required matter, but an infall of pristine or sub–solar gas is required to generate the observed metallicity of stars (G–dwarf problem, Alibés, Labay and Canal, 2001; Schönrich and J. Binney, 2009). This indicates that a significant gas fraction needs to come from regions beyond the Milky Way galaxy disk.

Another potential source of star formation material is recycled disk gas, mixed with halo gas via Galactic fountains (Shapiro and Field, 1976; Houck and Bregman, 1990). Recent simulations showed that this formation channel yields a sufficient mass inflow to successfully model the evolution of the Milky Way (Marasco, Marinacci and Fraternali, 2013; Fraternali et al., 2013). The models assume a mixture of rising Galactic fountain gas with halo material of low angular momentum. The results agree well with the observed kinematics of nearby spiral galaxies (Fraternali and J. J. Binney, 2008).

Since their discovery by Muller, Oort and Raimond (1963), high-velocity clouds (HVCs) have been considered to be candidates for inflowing gas. However, recent studies showed that the HVC gas accretion mass flow is only of the order of $0.08\,M_\odot\,yr^{-1}$ (Putman, Peek and Joung, 2012). In combination with the merging of gas-rich satellite galaxies, accretion rates of $0.1-0.2\,M_\odot\,yr^{-1}$ are expected, corresponding to only 10% of the SFR (Sancisi et al., 2008).

By comparing simulations and observations, Peek, Putman and Sommer-Larsen (2008) confirmed this HVC accretion rate of $0.2\,M_\odot\,yr^{-1}$ and furthermore concluded that neutral atomic hydrogen is just a minor (30 %) constituent of HVCs, which are thought to be predominantly ionised. These conclusions are consistent with those of Barger et al. (2012) who investigated the Hα content of HVC complex A and those of Fox, Savage and Wakker (2006) who study highly ionised species such as O VI and C III for various lines of sight and deduced that HVCs have a multiphase structure.

Here, we aim to investigate observationally the role of infalling sub–solar metallicity HVC gas by studying the gas–to–dust relation of high Galactic altitude clouds.

3.1.1 Interaction of HVCs with the disk–halo interface

In the past decades the interaction of HVCs with their ambient medium has been studied extensively. The Gould Belt (Comeron and Torra, 1994, and references therein) is a prime example for such a region because it has been proposed that it was shaped by the collision of an HVC with the Galactic disk. The Smith Cloud and its possible interaction with the coronal halo medium have been studied by Lockman et al. (2008). In their review on HVCs, Wakker and van Woerden (1997) considered the evidence for and against the interaction of a number of HVCs with the disk or intermediate-velocity clouds (IVCs) at the disk–halo interface. With the exception of HVC complex M (Danly, Albert and Kuntz, 1993), the confirmed distance of HVCs, determined by absorption line studies, reaches up to around 10 kiloparsec. Thus, the neutral gas is too far away to physically interact with the gas located at the disk–halo interface.

It has been proposed that a physical connection between HVC and IVC gas is detected by H I 21 cm line emission (Pietz et al., 1996). So–called velocity bridges (VB) in H I appear to continuously connect the HVC and IVC in H I spectra in the general direction of HVC complex C. Spatially correlated with these H I VBs, Kerp et al. (1999) found large–scale soft X-ray enhancements in *ROSAT* (Voges et al., 1999) all–sky survey data. The emission measure and plasma temperature suggest a compression of coronal halo gas at the leading edge of the HVC neutral rims, leading to a volume density enhancement by an order of magnitude at a presumed altitude of 4 kpc. This implies that HVC gas is decelerated, the ambient halo gas being compressed and finally becoming neutral in the disk–halo interface region, which is assumed to be located at an altitude between 0.8 and 1.5 kpc (Richter et al., 2001; Ben Bekhti et al., 2012).

Collisions of HVCs with Milky Way disk gas force the formation of density enhancements, yielding a mixture of disk and sub–solar metallicity gas. Accordingly, targets of our search for potential interactions

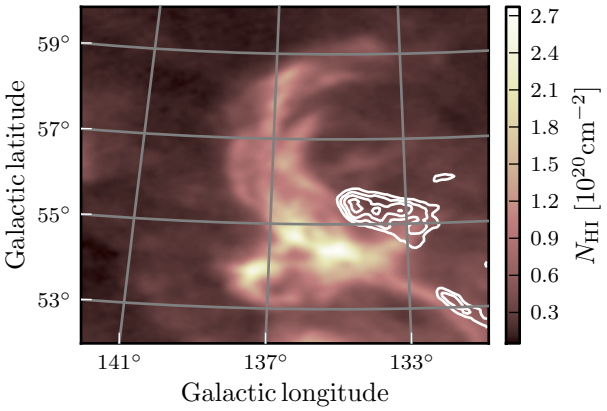


Figure 3.1: EBHIS (Kerp et al., 2011) column density map of IVC135+54 with the HVC contours superposed. For the IVC, velocities ($-52\,\mathrm{km\,s^{-1}} \leq v_{\rm LSR} \leq -31\,\mathrm{km\,s^{-1}}$) are taken into account. For the HVC, we use ($-190\,\mathrm{km\,s^{-1}} \leq v_{\rm LSR} \leq -90\,\mathrm{km\,s^{-1}}$). Contours start at 6×10^{19} cm^{-2} and increase in steps of 2×10^{19} cm^{-2}.

of HVC with the Galactic disk are IVCs with sub–solar metallicities and a close positional association with HVCs.

3.1.2 IVC135+54

An example of an IVC with sub–solar metallicity and close positional correlation with an HVC is located at Galactic coordinates $(l, b, v_{\rm LSR}) = (135\,^\circ, +54\,^\circ, -45\,\mathrm{km\,s^{-1}})$ (Figure 3.1, hereafter IVC135+54, also catalogued as IV21 in Wakker 2001). IVC135+54 has been extensively studied over the past three decades across a wide range of wavelengths. The cloud and its associated HVC are located at the north-eastern tip of the HVC complex C (Wakker and van Woerden, 1997). Thom et al. (2008) located large parts of complex C at a distance of approximately 10 kpc. This distance, combined with the metallicity of $\sim 0.15\,Z_\odot$ (Fox et al., 2004), led to the hypothesis that complex C is either accreted from the intergalactic medium (IGM), or stripped off from a satellite galaxy (Richter, 2012).

Using Na I absorption lines, Benjamin et al. (1996) bracketed the distance to IVC135+54 and found a best estimate of $d = (355 \pm 95)$ pc. The high Galactic latitude of the IVC allows its height above the Galactic plane to be relatively precisely calculated at $z = (285 \pm 75)$ pc. This is approximately the thickness of the H I layer (Kalberla and Kerp, 2009). It is one of only seven known intermediate-velocity molecular clouds (IVMCs, Magnani and Smith, 2010) that show high column densities ($\geq 10^{19}$ cm^{-2}) of molecular hydrogen.

Using the full sky data of the new Effelsberg-Bonn H I survey (EBHIS, Winkel et al., 2010; Kerp et al., 2011) together with Planck (Planck Collaboration, 2013 VI) and IRIS (Miville-Deschênes and Lagache, 2005) data, we investigate the interaction scenario and the implications for the transition of gas phases in the IVC. The main aim of the present work is to test whether IVC and HVC are in physical contact and to investigate the properties of dust and gas in the clouds.

3.2 Analysis

To ensure a proper comparison of the different data sets and to minimise systematic uncertainties resulting from interpolations, we smooth those data with higher spatial resolution to the appropriate lower resolution of the EBHIS data by convolving them with a Gaussian kernel. The fact that the Planck beam is not a perfectly symmetric Gaussian does not have a measurable impact, as the EBHIS resolution is much lower.

We use data in the HEALPix format (Górski et al., 2005) for pixel-based comparisons which require

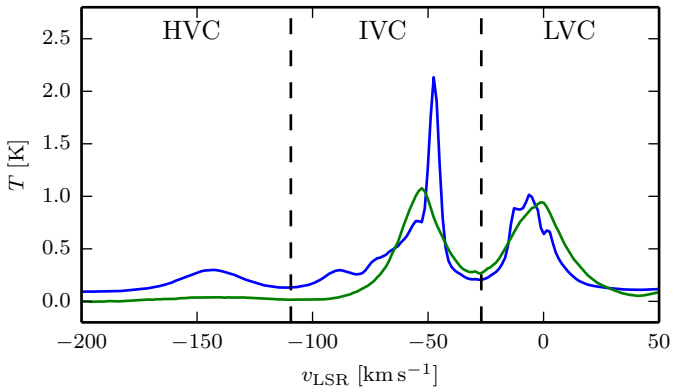


Figure 3.2: EBHIS median spectrum (blue) and standard deviation spectrum (green) of IVC135+54. Minima in the noise spectrum are used to separate the H I components at low, intermediate, and high velocities.

an equal-area projection. The presented maps are projected from HEALPix onto a Sanson-Flamsteed grid (Calabretta and Greisen, 2002) for displaying purposes.

3.2.1 H I data

Using the EBHIS measurements of neutral atomic hydrogen, we studied the three-dimensional structure of the interstellar medium and analysed the interplay of the different velocity regimes with high spectral resolution. The Doppler radial velocity information of the H I data allows us to infer the acceleration and deceleration of gas.

The EBHIS column density map of IVC135+54 is shown in Figure 3.1. The IVC in H I consists of three distinct clumps and an extended arc-like shape that was also pointed out by Hernandez et al. (2013). In addition to the emission from IVC135+54 itself, we find H I emission in the general direction of the IVC in the form of fine filamentary structures and diffuse background gas. We estimate the H I mass of IVC135+54 to be $650^{+400}_{-300}\,\mathrm{M}_\odot$ at a distance of 355 pc (Benjamin et al., 1996). The uncertainty in this mass estimate is dominated on the one hand by the distance uncertainty and on the other from diffuse and filamentary emission in the direction of IVC135+54. Thus, our calculations serve only as an order-of-magnitude estimate. Adopting this mass and spherical symmetry, we find a mean H I volume density of $n = 300\,\mathrm{cm}^{-3}$.

Superposed onto the IVC in Figure 3.1 is the HVC column density distribution in white contours. The H I mass of the HVC is approximately $M_{\mathrm{HVC}} = 8 \times 10^4 \cdot \left(\frac{D}{10\,\mathrm{kpc}}\right)^2\,\mathrm{M}_\odot$.

An average H I spectrum of the region indicated in Fig. 3.1 is shown in Fig. 3.2. For each channel of the three-dimensional datacube, we computed the median and the standard deviation. Based on the standard deviation spectrum, we decomposed the H I data into three components at low, intermediate, and high velocities. Following the approach of the Planck Collaboration (2011 XXIV), minima in the noise spectrum were used to separate the H I components. The column density maps for each component are shown in Fig. 3.3.

Velocity structure

To investigate the velocity structure and the proposed connection of high-velocity and intermediate-velocity gas, we generated a longitude-velocity diagram of IVC135+54 (Figure 3.4). The IVC is the brightest component, located around $v_{\mathrm{LSR}} \approx -50\,\mathrm{km\,s}^{-1}$. The right-hand panel is a magnified version that has been spectrally smoothed to a resolution of $5\,\mathrm{km\,s}^{-1}$. Contours mark the 5σ-level and 8σ-level. The local gas at radial velocities around $0\,\mathrm{km\,s}^{-1}$ is rather faint with $T_{\mathrm{B}} \approx 1$ K. Towards high negative

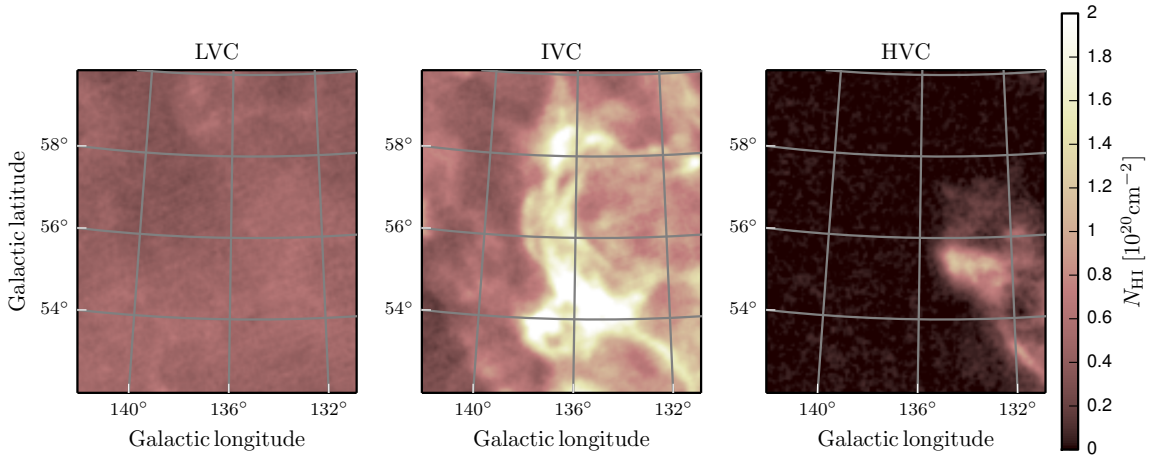


Figure 3.3: EBHIS column density maps of LVC, IVC, and HVC gas based on the velocity thresholds shown in Fig. 3.2. The LVC component is rather faint and diffuse, while the IVC is very bright with peak brightness temperatures of 4 K. For the HVC, there is hardly any background emission and the HVC has a steep brightness gradient in the eastern direction, towards the IVC.

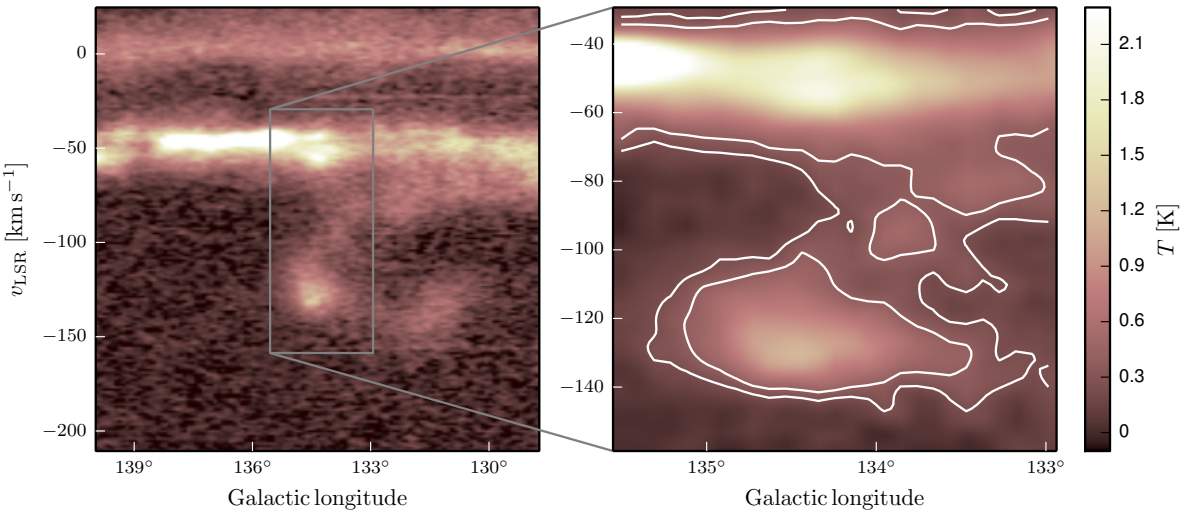


Figure 3.4: **Left**: EBHIS position-velocity diagram of IVC135+54 with Galactic latitude fixed at 55.8 °. **Right**: Zoom, smoothed to a velocity resolution of 5 km s^{-1}. The VB connecting HVC and IVC is highlighted by white contours at the 5σ-level and 8σ-level, corresponding to brightness temperatures of 250 mK and 400 mK.

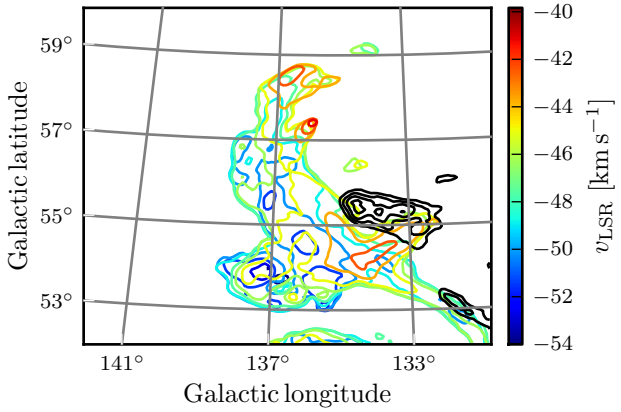


Figure 3.5: EBHIS Renzogram of the IVC. For each colour-coded velocity channel, a contour line at the fixed brightness temperature of 5 K is drawn. Superposed is the HVC with contours starting at 6×10^{19} cm^{-2} and increasin g in steps of 2×10^{19} cm^{-2} (black).

radial velocities around $-130\,\mathrm{km\,s^{-1}}$, the HVC is visible. The velocity difference between the HVC components and its appearance in the column density map (Fig. 3.3) indicates a head-tail (HT) structure (Brüns et al., 2000).

We confirm the velocity bridge that connects IVC and HVC gas, as detected in previous studies by Pietz et al. (1996). The VB has a brightness temperature of approximately 0.4 K and implies that the HVC gradually slows down and merges with the IVC in velocity space. The VB column density is in the range of $2-3.5\times10^{19}\,\mathrm{cm^{-2}}$ and thus three times lower than the HVC, which has peak column densities around $1.2\times10^{20}\,\mathrm{cm^{-2}}$. To convert the column density into a volume density and a mass for the VB, we estimated its radial extent. Adopting the hypothesis of a physical contact between the neutral gas phase of the IVC and the HVC locates the VB at a distance of 355 pc (Benjamin et al., 1996). Assuming spherical symmetry, the angular extent of approximately $2^\circ\times2^\circ$ yields an H I number density of $n_{\mathrm{H\,I}}\lesssim1\,\mathrm{cm^{-3}}$ and an H I mass of about 10-20 $\mathrm{M_\odot}$.

To study the IVC and its velocity structure, we generated a Renzogram (e.g. Schiminovich et al., 1997), which allowed us to investigate the spatial axes simultaneaously with the velocity structure (Figure 3.5). For each spectral channel of our observations ($\Delta v = 1.3\,\mathrm{km\,s^{-1}}$), we drew a contour line at the fixed brightness temperature of 5 K and colour-coded each contour with the corresponding radial velocity. This type of plot is preferred over velocity-weighted intensity maps because it provides different information on the velocity substructure of the IVC. We also covered a broader velocity range because no averaging is involved in the Renzogram. We find a velocity shift of approximately $15\,\mathrm{km\,s^{-1}}$ across the IVC. At the hypothetical impact position of the HVC, that is, between its head and the arc in the IV regime, the radial velocity shift is reduced to approximately $10\,\mathrm{km\,s^{-1}}$. This is slightly different from the $15\,\mathrm{km\,s^{-1}}$ shift towards the high-b or low-l end of the IVC and is consistent with the interpretation of momentum transfer from the HVC to the IVC.

To deduce a coherent view of IVC135+54 it is essential to consider not only the IVC and the VB, but also the HVC, which has been proposed to interact with IVC135+54. We investigate the HVC for signatures of shocks and their potential influence on the dust grains and on the hydrogen phase transitions (e.g. Guillard et al., 2009). The sound velocity of the HVC head is determined by fitting two Gaussian components to the HVC spectrum. The location is chosen via the HVC head column density maximum (Fig. 3.3). This returns line widths of $15.2\,\mathrm{km\,s^{-1}}$ (cold component) and $36.0\,\mathrm{km\,s^{-1}}$ (warm component). We converted the line width to upper limits of the kinetic temperature via the relation

$$T_{\mathrm{kin}} \leq \frac{\Delta v^2\cdot m_{\mathrm{H}}}{8k_B\ln 2} = 21.85\cdot\left(\Delta v\,[\mathrm{km\,s^{-1}}]\right)^2. \tag{3.1}$$

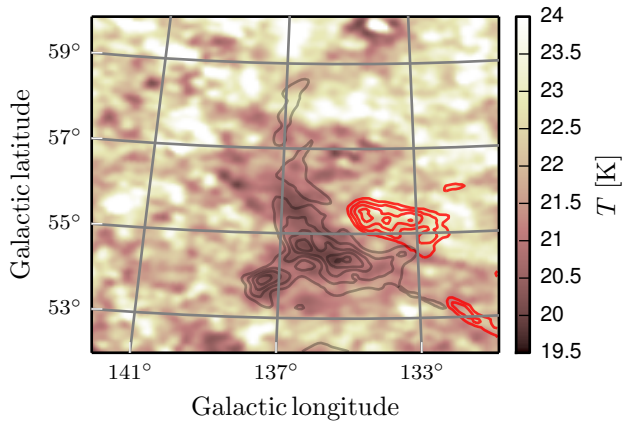


Figure 3.6: Dust temperature map of IVC135+54, extracted from a modified blackbody fit (Planck Collaboration, 2013 XI). Superposed are the EBHIS HVC contours, starting at 6×10^{19} cm^{-2} and increasing in steps of 2×10^{19} cm^{-2} (red) and the IVC contours starting at 1.1×10^{20} cm^{-2} and increasing in steps of 4×10^{19} cm^{-2} (grey).

The adiabatic index γ of atomic hydrogen is 1.67. The resulting upper limits to the kinetic temperature are 8400 K (cold component) and 2.9×10^4 K (warm component), which significantly exceeds typical temperatures (see e.g. Draine (2011), who gives $\sim$ 100 K and $\sim$ 5000 K). Hence, we suggest that the line width is dominated by turbulence and not by the kinetic temperature.

Conditions for shocks can be determined by converting the temperatures into sound velocities c_S

$$c_S = \left(\frac{\gamma k_B T_{\rm kin}}{m_H} \right)^{1/2} . \tag{3.2}$$

Based on the temperature estimates, the calculated upper limits to the sound velocities range from 6.5 km s^{-1} to 15.6 km s^{-1}. Because of the large velocity difference ($\gtrsim$ 40 km s^{-1}) between HVC and IVC, the proposed HVC impact causes supersonic shock waves.

3.2.2 Dust data

In addition to the H I data, information on interstellar dust grains as extracted from the FIR IRIS and Planck data allow us also to probe the molecular and ionised gas phases. This analysis can only be applied to the IVC however, because the HVC is not detected in the FIR data. For a distant HVC, the relatively weak interstellar radiation field (ISRF) at high Galactic altitudes decreases the illumination of dust grains by UV-starlight. Additionally, the suggested extragalactic origin of HVCs and especially of HVC complex C would be expected to be associated with low metallicities ($\sim 0.15\,Z_\odot$, Fox et al., 2004) and thus in small dust abundances.

The dust temperature map of IVC135+54 is displayed in Fig. 3.6 (Planck Collaboration, 2013 XI). It shows that the dust in the IVC is cooler ($\lesssim$ 20 K) than in the environment ($\gtrsim$ 22 K), which is associated with the higher H I column densities and the presence of different molecular species in the cloud (e.g. Weiß et al., 1999). However, recent studies (Planck Collaboration, 2013 XI) find that the coupling between dust and gas temperature is weak. Thus, dust temperature fluctuations might instead be interpreted as signature of dust evolution.

3.2.3 H I-dust correlation

The correlation between neutral hydrogen and FIR dust emission has been shown to be linear up to H I column densities of 1.5 to 5×10^{20} cm^{-2} (Desert, Bazell and Boulanger, 1988; Boulanger et al., 1996; Reach, Wall and Odegard, 1998). For higher column densities, the FIR emission is in excess

with respect to the linear model (Reach, Wall and Odegard, 1998; Planck Collaboration, 2011 XXIV; Planck Collaboration, 2013 XI).

This excess emission can partly be attributed to H I self–absorption. If the assumption of optically thinness is violated, the observed H I column density is systematically underestimated. However, the contribution from this process is only about a few percent in the present case because of the low H I column densities in IVC135+54 (Planck Collaboration, 2011 XXIV; Braun, 2012).

Since FIR dust emission traces not only hydrogen in the atomic phase, but actually the total hydrogen column density

$$N_{\mathrm{H}} = N_{\mathrm{H\,I}} + 2 \cdot N_{\mathrm{H_2}} + N_{\mathrm{H\,II}}, \tag{3.3}$$

the excess emission can be used to trace ionised and molecular hydrogen.

To disclose possible deviations from the correlation for instance as a result of variable dust emissivity (Planck Collaboration, 2013 XI), we used not only the FIR intensity to trace interstellar dust, but also the dust opacity

$$\tau_\nu(\lambda) = \frac{I_\nu(\lambda)}{B_\nu(T_{\mathrm{D}})} \tag{3.4}$$

at 353 GHz provided by Planck Collaboration (2013 XI), which is a measure of the total hydrogen column density. We dropped the index ν because we only used this frequency. In the following, τ is used as a measure for the dust column density for all equations and expressions related to the H I-dust correlation.

To account for the different dust emissivities

$$\epsilon = \tau / N_{\mathrm{H\,I}} \tag{3.5}$$

of local gas as well as intermediate– and high–velocity gas, we adopted the approach of the Planck Collaboration (2011 XXIV). Hence, we interpret the total dust opacity as

$$\tau(x, y) = \sum_i \epsilon^i \cdot N^i_{\mathrm{H\,I}}(x, y) + Z + R(x, y). \tag{3.6}$$

Here, i cycles over the LVC, IVC, and HVC component of the H I emission presented in Fig. 3.3. Z is a global offset to the full map and R is the residual emission that disagrees with this linear model. To properly account for the FIR excess emission of the IVC, we modified Eq. (3.6) to differentiate between two emissivities in the IVC regime: For low H I column densities, we expect both the atomic and the ionised phase but used a single value because the FIR emissivity scales only according to the number of nucleons. At high H I column densities, the molecular hydrogen contributes to the dust opacity, yielding a higher emissivity. Hence, these emissivities reflect the atomic and the molecular hydrogen of the IVC, respectively. The transition between the two emissivities $\epsilon^{\mathrm{IVC,\,low}}$ and $\epsilon^{\mathrm{IVC,\,high}}$ is realised using a step function $\theta(N^0_{\mathrm{H\,I}})$. The IVC contribution to the dust opacity, τ^{IVC}, can then be written as

$$\tau^{\mathrm{IVC}} = \epsilon^{\mathrm{IVC,\,low}} N^{\mathrm{IVC}}_{\mathrm{H\,I}} + \epsilon^{\mathrm{IVC,\,high}} (N^{\mathrm{IVC}}_{\mathrm{H\,I}} - N^0_{\mathrm{H\,I}}) \cdot \theta(N^0_{\mathrm{H\,I}}). \tag{3.7}$$

We solved Eq. (3.6), using a Bayesian analysis realised with a Markov chain Monte Carlo (MCMC) approach that is implemented in PyMC (Patil, Huard and Fonnesbeck, 2010). Fig. 3.7 shows the result of this model. A pixel-based evaluation of the results is shown in Fig. 3.8. To compute the dust opacity for the different velocity intervals, the contribution from sources other than the one shown were subtracted from the total opacity, that is,

$$\tau^{\mathrm{i}} = \tau - Z - \sum_{j \neq i} \epsilon^{\mathrm{j}} N^{\mathrm{j}}_{\mathrm{H\,I}}. \tag{3.8}$$

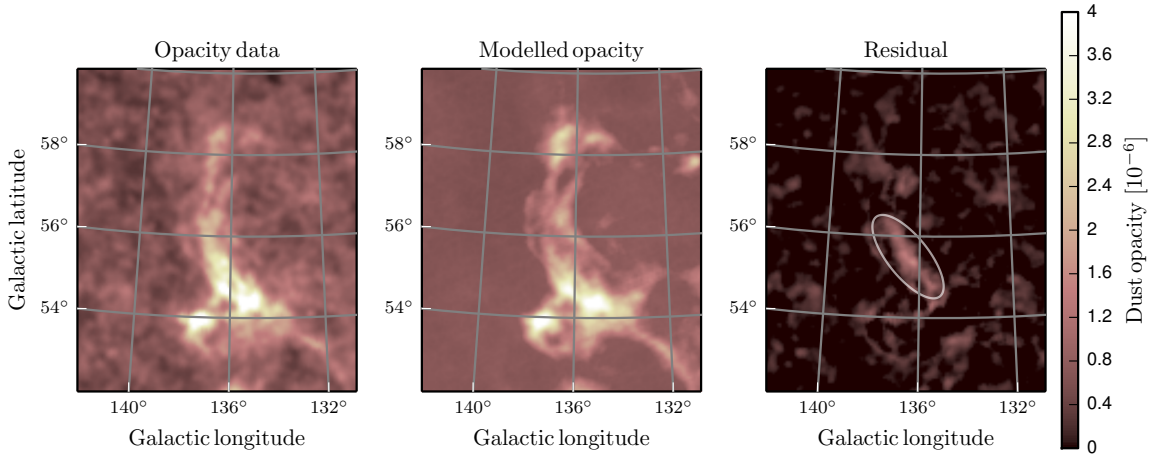


Figure 3.7: **Left:** IVC135+54 dust opacity at 353 GHz as measured by Planck (Planck Collaboration, 2011 XXIV). **Centre:** Reconstruction of the opacity via velocity–dependent emissivities according to Eqs. (3.6) and (3.7). **Right:** Residual emission ($\tau^{\mathrm{Data}} - \tau^{\mathrm{Model}}$). The white ellipse marks the pronounced front that is connected with the HVC head.

The right-hand side of Fig. 3.8 clearly illustrates that a simple, linear approach cannot describe the data properly. Unlike the Planck Collaboration (2011 XXIV), for instance, we did not mask the data points that contain FIR excess emission, but included them in our model, which allowed us to better quantify this excess emission. The switchpoint $N^0_{\mathrm{H\,I}}$ of the transition is another additional parameter that is sampled by the Bayesian model. With this we also determined the column density threshold at which molecular hydrogen starts to dominate the atomic hydrogen. Table 3.1 summarises the fit parameters for the H I column density correlation with the FIR intensity and the dust opacity.

Data	I_{857}	τ_{353}
ϵ^{LVC}	$0.44^{+0.01}_{-0.02}$	$0.78^{+0.03}_{-0.04}$
$\epsilon^{\mathrm{IVC,\,low}}$	$0.06^{+0.01}_{-0.01}$	$0.09^{+0.02}_{-0.01}$
$\epsilon^{\mathrm{IVC,\,high}}$	$1.00^{+0.00}_{-0.02}$	$2.00^{+0.03}_{-0.02}$
ϵ^{HVC}	$-0.04^{+0.02}_{-0.01}$	$-0.09^{+0.03}_{-0.03}$
Z	$0.35^{+0.00}_{-0.01}$	$0.38^{+0.01}_{-0.01}$
$N^0_{\mathrm{H\,I}}$	$0.75^{+0.01}_{-0.01}$	$0.77^{+0.01}_{-0.01}$

Table 3.1: Emissivities ϵ^i for the different velocity components, according to Eq. (3.6). The emissivities are derived via the H I column density correlation with the FIR intensity (middle column) and the dust opacity (right column). For the intensity data I_{857}, units are 10^{-20} MJy sr^{-1} cm^2 for the emissivities and MJy sr^{-1} for the offset. For the opacity data τ_{353}, units are 10^{-26} cm^2 for the emissivities and 10^{-6} for the offset. For $N^0_{\mathrm{H\,I}}$, units are 10^{20} cm^{-2}. The uncertainties denote the 95% uncertainty interval and only account for the statistical error of the fit. A brief investigation of the systematic uncertainties is presented in Sect. 3.2.4.

This approximation yields an emissivity ϵ^{HVC} lower than zero, which cannot be motivated physically. The very low dust content of the HVC, if any, in combination with the statistical and systematical uncertainties biases the performed fit to this result. Although the MCMC error for ϵ^{HVC} is small, we show below that additional systematic uncertainties have to be taken into account (Sect. 3.2.4). One of these

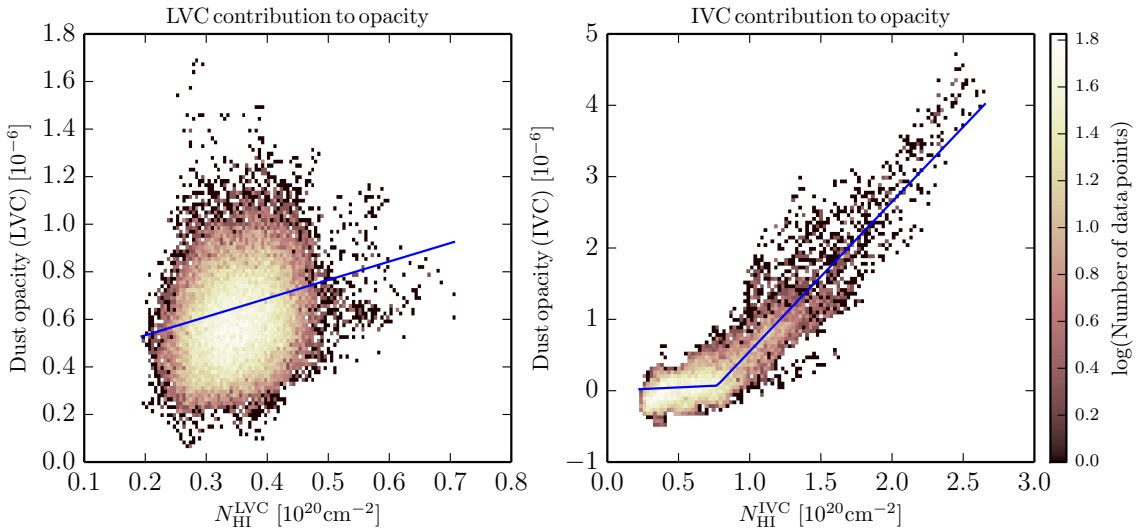


Figure 3.8: Dust opacity at 353 GHz as a function of the EBHIS column density for the LVC and IVC gas. The H I column densities correspond to the maps shown in Fig. 3.3. For the opacities, the contribution from sources other than the one shown are subtracted from the total opacity according to Eq. (3.8). The linear correlation, corresponding to the emissivity ϵ, is shown in blue. For the IVC, two different emissivities are used to model the correlation (Eq. 3.7).

systematic effects is the quantitative subtraction of the FIR emission of the cosmic infrared background (CIB) that needs to be accounted for (Puget et al., 1996; Planck Collaboration, 2013 XI).

As expected, the IVC emissivity at low H I column densities, $\epsilon^{\rm IVC,\,low}$, is extremely low. This results from foreground opacity that is modelled via $\epsilon^{\rm LVC}$ and Z. Hence, $\epsilon^{\rm IVC,\,low}$ can be interpreted as systematic uncertainty to our foreground correction. For the high H I column density part of the IVC, $\epsilon^{\rm IVC,\,high}$, the spatial correlation between H I and dust is very tight and properly reflected by the high emissivity. This FIR excess part of the diagram allows us to infer information on the molecular hydrogen.

For the H I column density $N_{\rm H\,I}^{0}$ at which the data deviate from a linear correlation, we find a value of 0.75×10^{20} cm^{-2}. This is lower than the commonly accepted value of 1.5 to 5×10^{20} cm^{-2} (Desert, Bazell and Boulanger, 1988; Boulanger et al., 1996), but it must be noted that this transition was computed after accounting for the LVC foreground gas with median column densities of about 0.4×10^{20} cm^{-2}. Therefore, we arrive at a classical threshold of approximately 1.15×10^{20} cm^{-2}, which is consistent with previous studies. Furthermore, the pronounced front in the residual map (right-hand side of Fig. 3.7) is connected with the dense HVC head, which appears to be strongly decelerated with respect to the HVC tail.

As argued earlier, the dust opacity is more resistent to variations in the ISRF than the FIR intensity. Hence, we used the emissivities derived with the opacity for subsequent steps of the analysis. The FIR intensity-based emissivities are used for a comparison with other studies (Sect. 3.3.1).

3.2.4 Displacement–map method

The derivation of dust emissivities ϵ^{i} assumes independent and identically distributed Gaussian uncertainties in the residuals (R in Eq. 3.6). The right-hand side plot in Fig. 3.7 clearly shows that this assumption is not valid for our data, given the structure and large–scale variations in the residual.

Therefore, we investigated the uncertainty $\sigma_{\epsilon}^{\rm pos}$, introduced by varying the position of the cloud across

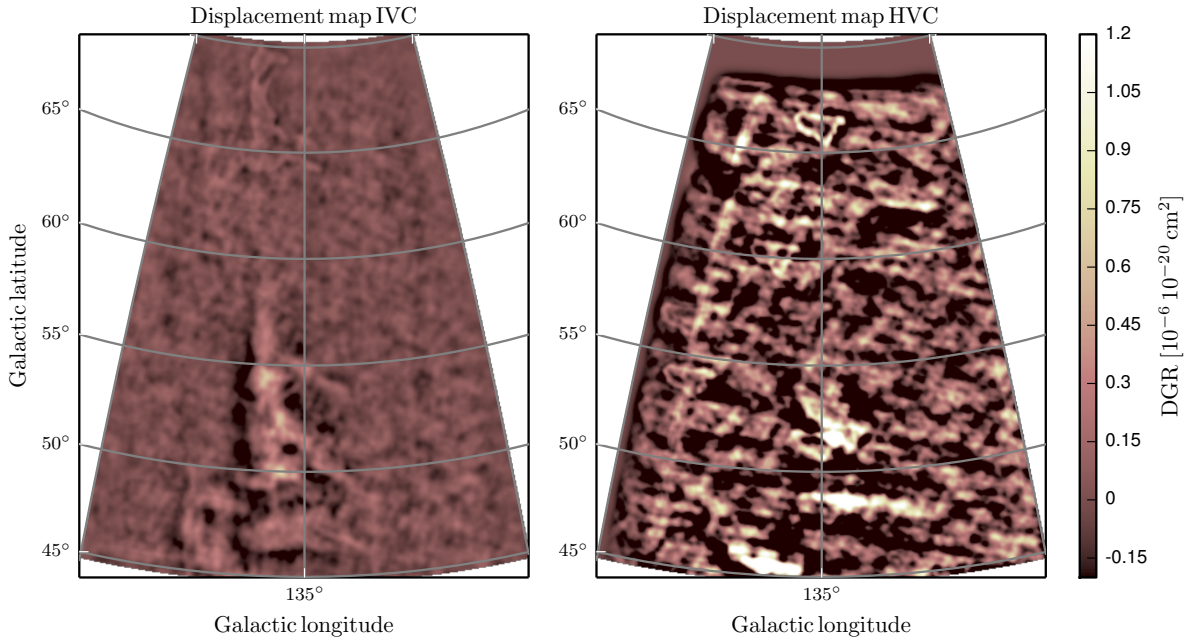


Figure 3.9: Application of the displacement-map technique (Peek et al., 2009). **Left:** Normalised cross-correlation of IVC opacity and H I column density. The variations in this map allow us to measure the uncertainty of ϵ^{IVC}. **Right:** Same image as left, but with the HVC displaced across the map.

the sky and evaluating the model for each position shift. The displacement–map method was applied by Peek et al. (2009) to investigate the dust content of different HVCs and IVCs. Peek et al. (2009) aimed to obtain a reliable estimate of the uncertainties on the emissivities ϵ^i. For a full description we refer to Peek et al. (2009), particularly to their Fig. 6.

In brief, we performed template matching of the dust opacity and H I column density map of IVC135+54. That is, we cross–correlated the measured H I map and the calculated opacity of the IVC. This cross–correlation is highest at the location of the cloud. The amplitude of the chance correlation across the map can be used as measure of the uncertainty. The displacement–map for the IVC and HVC is shown in Fig. 3.9.

To evaluate the uncertainty $\sigma_\epsilon^{\mathrm{pos}}$ we computed the standard deviation in signal-free regions in the displacement–map for the IVC and HVC. We find these uncertainties to be $\sigma_\epsilon^{\mathrm{pos}} = 0.33 \times 10^{-26}\,\mathrm{cm}^2$ for the HVC and $\sigma_\epsilon^{\mathrm{pos}} = 0.07 \times 10^{-26}\,\mathrm{cm}^2$ for the IVC. This means that the negative emissivity of $0.09 \times 10^{-26}\,\mathrm{cm}^2$ for the HVC is not significant, but results from the chance correlation. For the much more luminous IVC, the displacement–map method yields an uncertainty consistent with the MCMC error estimate (see Table 3.1).

3.2.5 FIR excess emission as a tracer of molecular hydrogen

Given the high H I column densities of IVC135+54, the molecular phase is expected to dominate the ionised phase. To test this hypothesis, we used the Wisconsin Hα mapper survey (WHAMS) data (Haffner et al., 2003) to estimate the column density of ionised hydrogen. Since the angular resolution of WHAMS

is about 1°, it does not allow a spatially precise comparison with the EBHIS and Planck data. We only evaluated the average intensity here, which yields a value of 0.46 ± 0.03 Rayleigh that corresponds to a column density (Lagache et al., 2000, Eq. 2) of $N_{\mathrm{H}\alpha} = 0.4 \times 10^{19}\,\mathrm{cm}^{-2}$. Because this is an order of magnitude below that of neutral atomic hydrogen, neglecting the emission of ionised hydrogen is a valid first order approach. It must however be noted that with the current generation of Galactic H I and dust data, we approach sensitivities at which H II can contribute to the correlation of H I and dust. Accordingly, a careful treatment of the ionised phase will be mandatory for future models, given the quality of EBHIS/GASS and Planck/IRIS, to name only two.

Because the H I column densities are not sufficiently high (Braun, 2012; Planck Collaboration, 2011 XXIV), we do not expect signifiant H I self–absorption. Therefore, the spatial information of the FIR excess emission and the emissivity $\epsilon^{\mathrm{IVC\,high}}$ describe the molecular content of the IVC. We re-arranged Eqs. (3.3) and (3.6) and solved for the molecular column density in the intermediate-velocity regime:

$$\tau^{\mathrm{IVC}} = \epsilon^{\mathrm{IVC,\,high}} \cdot (N_{\mathrm{H\,I}}^{\mathrm{IVC,\,high}} + 2\,N_{\mathrm{H_2}}^{\mathrm{IVC}}) \tag{3.9}$$

$$\Rightarrow N_{\mathrm{H_2}}^{\mathrm{IVC}} = \frac{\tau^{\mathrm{IVC}}}{2\epsilon^{\mathrm{IVC,\,high}}} - \frac{N_{\mathrm{H\,I}}^{\mathrm{IVC,\,high}}}{2} \tag{3.10}$$

The factor of two accounts for the fact that H_2 consists of two hydrogen atoms. This equation can be applied to generate maps of upper limits to the molecular hydrogen column density (Fig. 3.10). This result must be treated carefully because it is sensitive to $\epsilon^{\mathrm{IVC,\,high}}$, which is difficult to measure properly since it requires an accurate model of the foreground LVC gas.

The information on the molecular column density can be compared with the values inferred by Weiß et al. (1999) from measuring ^{12}CO and ^{13}CO. They mapped the small region in IVC135+54, indicated by the rectangle in Fig. 3.10. Using the line integral of ^{12}CO, they found $W_{\mathrm{CO}} = 1 - 10\,\mathrm{K\,km\,s^{-1}}$. This can be converted into molecular hydrogen column densities by adopting an $\mathrm{X_{CO}}$ conversion factor of approximately $\mathrm{X_{CO}} = 4 \times 10^{20}\,\frac{\mathrm{cm^{-2}}}{\mathrm{K\,km\,s^{-1}}}$ (Draine et al., 2007; Narayanan et al., 2012). Using this, Weiß et al. (1999) deduced molecular column densities of $N_{\mathrm{H_2}} = 0.4 - 4 \times 10^{20}\,\mathrm{cm}^{-2}$. For the same region, we find molecular column densities in the range of $1.3 \times 10^{20}\,\mathrm{cm}^{-2}$ to $2.1 \times 10^{20}\,\mathrm{cm}^{-2}$, traced by the FIR excess emission (Figure 3.10).

Furthermore, we note that potential differences in molecular hydrogen column density derived from (a) FIR excess emission and (b) CO line–integral can arise from systematic uncertainties. In their analysis of the infrared cirrus clouds Spider and Ursa Major, Barriault et al. (2010) found that the peaks of FIR excess emission and ^{12}CO emission peaks do not coincide spatially. A follow-up analysis (Barriault, Joncas and Plume, 2011) showed that low densities at the location of the FIR excess peak cause a low CO excitation temperature, which was argued to be the reason for the faint CO emission that causes the misalignment. This highlights the importance of investigating the properties of different tracers and applying them only under the proper conditions.

3.3 Discussion

3.3.1 Dust-to-gas ratios

Unlike other studies of the dust-to-gas ratio (DGR) in IVCs, we did not mask the FIR excess emission, but included it in the model. Thus, a comparison with other studies is not straightforward. However, the foreground emissivities ϵ^{LVC} and $\epsilon^{\mathrm{IVC,\,low}}$ can be used to compare our results with those found by the Planck Collaboration (2011 XXIV), who investigated the velocity-dependent emissivity of different

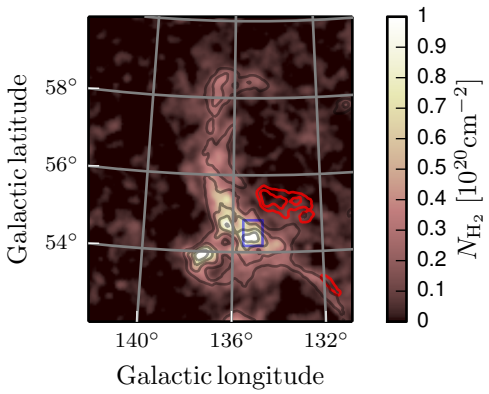


Figure 3.10: Distribution of the cold, dominantly molecular, gaseous phase of IVC135+54. Using Eq. (3.10), we convert the FIR excess emission (Fig. 3.7 and 3.8) into upper limits to the molecular column density. The blue rectangle indicates the region in which Weiß et al. (1999) have detected ^{12}CO and ^{13}CO. Superposed are the HVC contours, starting at 6×10^{19} cm^{-2} and increasing in steps of 2×10^{19} cm^{-2} red) and the IVC contours starting at 1.1×10^{20} cm^{-2} and increasing in steps of 4×10^{19} cm^{-2} (black).

fields at high Galactic latitudes. For the LVC and IVC component, typical emissivities are around $\epsilon_{857\,\mathrm{GHz}} \approx 0.3 - 0.7 \times 10^{-20}$ MJy sr^{-1} cm^2, which agrees with our results for the correlation with $I_{857\,\mathrm{GHz}}$ (Table 3.1). Hence, IVC135+54 is an average halo cloud in terms of its DGR for low H I column densities. For the DGR at higher column densities, a comparison is not possible since this part of the data is masked in the study by the Planck Collaboration (2011 XXIV).

For IVC135+54 in particular, the DGR can be compared with the study performed by Weiß et al. (1999), who fitted a linear correlation to the IRAS and Effelsberg data of the cloud. They did not distinguish between a linear part of the correlation and FIR excess emission. They found the mean emissivities across the entire cloud to be around $\epsilon_{3000\,\mathrm{GHz}} \approx 1 \times 10^{-20}$ MJy sr^{-1} cm^2. An inspection of the ratios of IRIS intensity at 3000 GHz and EBHIS total column density yields similar values between 0.7 and 1.2×10^{-20} MJy sr^{-1} cm^2.

The latest study on IVC135+54 by Hernandez et al. (2013) claims that the cloud has an unusual low DGR of $\epsilon_{3000\,\mathrm{GHz}} \approx 0.32 \times 10^{-20}$ MJy sr^{-1} cm^2, using IRAS and Green Bank Telescope data. Their values are not consistent with those reported here. The main difference probably arises from fact that Hernandez et al. (2013) used IRAS instead of the reprocessed IRIS data. A direct comparison of the two data sets exhibits main differences in de-striping, calibration, and, most importantly here, on the absolute FIR luminosity level.

3.3.2 Origin of the HVC–IVC system

The literature on IVC135+54 suggests different origin scenarios on the observed IVC–HVC system (Weiß et al., 1999; Hernandez et al., 2013). We focus on two different questions: (1) are the observed structures remnants of infalling material from the IGM, or do they result from Galactic fountains, and (2) do IVC and HVC have a common origin, or are they shaped by an HVC impact on the disk–halo interface?

Galactic or extragalactic origin

In their analysis on the H I, dust, and metallicity of IVC135+54, Hernandez et al. (2013) focused on question (1) and argued that the distinction between Galactic fountain and extragalactic HVC infall is only possible through evaluating the metallicity (similar to Heitsch and Putman, 2009). Hernandez et al. concluded that the observed sub–solar metallicity of $\log(Z/Z_\odot) = -0.43 \pm 0.12$ dex is a signature of extra- or circumgalactic infall via HVCs. This proposal is consistent with the simulations of Marasco, Marinacci and Fraternali (2013), showing that the scale height of the hot material of Galactic fountains is about ~ 1.2 kpc at $R = R_\odot$. Moreover, the HVC associated with IVC135+54 is part of complex C, which

is proposed to be accreted from the IGM or stripped off a satellite galaxy (Richter, 2012). This makes an extragalactic origin of the HVC most likely.

For the IVC, we find that its dust emissivity at 353 GHz is similar to or higher than the emissivity of disk gas at low velocities (Table 3.1). Since the emissivity is a tracer of metallicity in first order, we conclude that the IVC has approximately solar metallicity. This agrees with the ratio of the dust intensity at 100 μm and the H I column density, which we find to be of the order of 1×10^{-20} MJy sr^{-1} cm^2 across the IVC. Similar ratios are found for other molecular IVCs in the Galactic halo (e.g. Weiß et al., 1999; Magnani and Smith, 2010). Thus, we concluded that the IVC is of Galactic origin, whereas the HVC is associated with complex C and therefore probably of extragalactic origin.

Impact onto the IVC or remnant of HVC infall

To answer the question of whether IVC and HVC interact physically or whether both components are remnant of a larger, infalling HVC, different facts need to be taken into account. For the IVC, we found a head-tail structure (Figure 3.5), but the velocity shift of approximately $\Delta v_{\mathrm{LSR}} = 14\,\mathrm{km\,s^{-1}}$ can also result from a possible confusion with line-of-sight effects. Brüns et al. (2000) noted that these effects can be $10\,\mathrm{km\,s^{-1}}/[^\circ]$. This means that we do not find compelling evidence of ram-pressure interaction of the IVC with the ambient medium, as would be expected if the IVC were indeed descending through the disk–halo interface.

The assumption that the IVC-HVC system has been shaped by the interaction of a static IVC with an infalling HVC agrees with the negative spatial correlation of both clouds and also with the detected VB that connects the two components. The inferred deceleration of the HVC down to the IVC regime is a strong indicator for inelastic interaction. Furthermore, we detected an arc of FIR excess emission that indirectly traces molecular hydrogen in the IVC and is coincident with the proposed impact position of the HVC.

The FIR excess emission can originate from enhanced dust production and thus from associated formation of molecular hydrogen. We calculated the sound velocities in this regime to be $15\,\mathrm{km\,s^{-1}}$ at most, therefore the velocity difference of approximately $40\,\mathrm{km\,s^{-1}}$ between HVC and IVC implies supersonic shocks that enhance the pressure locally, which decreases the formation time of H_2 (Guillard et al., 2009; Röhser et al., 2014). The molecular hydrogen is "invisible" to H I observations, but is consistent with the observed FIR excess emission.

Distance difference between HVC and IVC

The distance to IVC135+54 was determined by Benjamin et al. (1996) to be approximately 355 pc. This is 30 times closer than the 10 ± 2.5 kpc at which HVC complex C is located (Wakker et al., 2007; Thom et al., 2008). The discrepancy between the distances to these objects is inconsistent with the proposed interaction scenario. Both distances were computed via bracketing with stellar absorption lines. A detailed and helpful description of this procedure is given by Schwarz, Wakker and van Woerden (1995).

For complex C, the non-detections and thus the lower limits to the distance are only based on three sight lines that Thom et al. labelled S135, S139 and S441 and that are located towards the low-l, low-b part of complex C. This part has a true angular distance to IVC135+54 of more than 30°. Indeed, Thom et al. argued that recent data suggest that complex C extends much closer to the plane. Moreover, they noted that the low-l, low-b part in which the sight-lines for the lower limits are located might be farther removed, thus resulting in a distance gradient across the complex.

Furthermore, the large projected angular extent of complex C of about 40 ° in combination with the assumed 10 kpc distance leads to a linear extent of 8.4 kpc. If we drop the hypothesis that the complex is

a sheet with equal distance to all points, the HVC associated with IVC135+54 can be much closer than the commonly cited 10 kpc, but a distance of only 350 pc appears very unlikely.

To conclude the discussion on the distance discrepancy, the three-dimensional structure of complex C remains elusive, and more work is required to generate a distance map. Based on the work of Thom et al. (2008), additional sight-lines for bracketing via stellar absorption lines could improve the distance map. For the HVC related to IVC135+54, the indicators for an interaction disagree with the distance of complex C. This means that either the distance to that particular HVC is heavily overestimated, or that another, unknown process that is not a physical connection between HVC and IVC causes the velocity bridge and the associated formation of dust and molecular hydrogen in the IVC.

3.4 Conclusion

Using the latest large–scale data sets for neutral hydrogen and interstellar dust, we have investigated the well-studied cirrus cloud IVC135+54, which is located at the disk–halo interface. EBHIS, Planck, and IRIS data were used to study the gas at both intermediate and high velocities, allowing us to probe the three-dimensional structure as well as different gas phases. Given the different conclusions on the origin of the IVC–HVC system in previous studies on IVC135+54, we focussed our analysis on the origin of the clouds.

(1) We find that the high-velocity cloud has a head-tail structure in the EBHIS data, combined with a velocity bridge that continuously connects gas from the high– to the intermediate–velocity regime. This indicates a deceleration of the high-velocity cloud.

(2) The far-infrared excess emission in IVC135+54 does not follow the distribution of either the neutral hydrogen or the dust opacity, but is instead enhanced in an arc that is connected with the high-velocity cloud head and the velocity bridge (Fig. 3.7 and 3.10).

(3) This combination of deceleration and far-infrared excess emission gives rise to the hypothesis that the high-velocity cloud is falling onto the disk–halo interface and interacts with IVC135+54.

(4) From the FIR excess emission we created maps of molecular hydrogen that confirm the presence of H_2 and make IVC135+54 one of the very few intermediate-velocity molecular clouds.

(5) The overall dust-to-gas ratio of IVC135+54 is compatible with the ratios of other infrared cirrus clouds and with previous studies of IVC135+54. By comparing the low H I column density part of the cloud with IRIS data at 3000 GHz and Planck data at 857 GHz, we found $\epsilon_{3000\,\mathrm{GHz}} = 0.7 - 1.2 \times 10^{-20}\,\mathrm{MJy\,sr^{-1}\,cm^2}$ and $\epsilon_{857\,\mathrm{GHz}} \approx 0.5 \times 10^{-20}\,\mathrm{MJy\,sr^{-1}\,cm^2}$, respectively.

Despite the advancement in understanding IVC135+54 and its association with high-velocity cloud complex C, there are several questions that remain to be answered to complete the picture:

(a) What is the role of the ionised hydrogen? Can high-spatial resolution Hα observations track ionisation generated by the HVC impact?

(b) The indicators for an interaction between IVC135+54 and the HVC are inconsistent with the large distances determined for other parts of HVC complex C. Could the HVC be closer to us than current distance estimations suggest, or are the indicators for an interaction misleading and have another origin?

These questions need to be addressed in future studies related to IVC135+54 to ultimately solve the question of its origin. This requires spatially extended observations of the other gas phases, that is, H II and molecular gas tracers such as CO, but also a more stringent distance determination to complex C and especially the high-velocity cloud associated with IVC135+54.

Resume

The following section is not part of the original publication by Lenz et al. (2015) on which this Chapter is based. Instead, we relate the content of this Chapter to the thesis as a whole and highlight the scientific and methodological progress.

The research that is published in this article is among the pioneering work that was done in the H I research group at the Argelander Institute for Astronomy (AIfA). The research on high- and intermediate-velocity cloud has a long history in Bonn, and the correlation with the *Planck* data of interstellar dust offered a new window into the Galactic halo. The collaboration between the Bonn H I group and the *Planck* collaboration turned out to be very fruitful, resulting in two Phd theses (the present work and Röhser, 2016) and multiple publications.

Despite the insights on the HVC/IVC system that have been gained in this Chapter, the methods that have been employed are not much different than similar approaches that were already applied in the late 1990s. After the work on this publication was finished, we therefore aimed to improve the investigation of the correlation between dust and gas in the following way: First, the dust content of the HVC in the IVC/HVC system would have been another strong indicator that the system is indeed in physical contact. Unfortunately, the model that was chosen and the systematics that were investigated did not suffice to make a clear statement on the dust in the HVC. Second, the dust model was based simply on the H I column density and made only very limited use of the H I spectral information. Third, the far-infrared background was only treated as a constant offset and the spatial structure of this cosmic infrared background was neglected.

In the upcoming Chapter, we address these issues and develop a novel method to investigate the correlation of dust and gas. Moreover, we will demonstrate how this correlation can be used to probe the cosmic infrared background radiation.

CHAPTER 4

Dust in a compact, cold, high-velocity cloud: A new approach to removing foreground emission

This Chapter is based on the content published in Lenz, Flöer and Kerp (2016), Dust in a compact, cold, high-velocity cloud: A new approach to removing foreground emission, A&A 586, A121, 2016, reproduced with permission © ESO.

Abstract

Because isolated high-velocity clouds (HVCs) are found at great distances from the Galactic radiation field and because they have subsolar metallicities, there have been no detections of dust in these structures. A key problem in this search is the removal of foreground dust emission.

Using the Effelsberg-Bonn H I Survey and the Planck far-infrared data, we investigate a bright, cold, and clumpy HVC. This cloud apparently undergoes an interaction with the ambient medium and thus has great potential to form dust.

To remove the local foreground dust emission we used a regularised, generalised linear model and we show the advantages of this approach with respect to other methods. To estimate the dust emissivity of the HVC, we set up a simple Bayesian model with mildly informative priors to perform the line fit instead of an ordinary linear least-squares approach.

We find that the foreground can be modelled accurately and robustly with our approach and is limited mostly by the cosmic infrared background. Despite this improvement, we did not detect any significant dust emission from this promising HVC. The 3σ-equivalent upper limit to the dust emissivity is an order of magnitude below the typical values for the Galactic interstellar medium.

4.1 Introduction

Since their discovery by Muller, Oort and Raimond (1963), high-velocity clouds (HVCs) have been the target of a wide range of studies (see Wakker and van Woerden (1997) for a review). It is thought that they are located at distances of several kpc (Wakker, 2001) and contribute to the fuelling of low-metallicity gas into the Galaxy (Putman, Peek and Joung, 2012).

Early attempts to detect dust emission from HVCs were unsuccessful (Wakker and Boulanger, 1986; Desert, Bazell and Boulanger, 1988) and were generally considered to be difficult because of the clouds'

low metallicities and hence low dust-to-gas ratios (Fox et al., 2004). Moreover, HVCs are located far from the interstellar radiation field (ISRF), which implies a faint illumination by UV light that is absorbed and re-emitted by the dust grains. The cosmic infrared background radiation (CIB) has anisotropies on angular scales that are comparable to the typical sizes of HVCs and is therefore another source of confusion (Planck Collaboration, 2011 XXIV; Planck Collaboration, 2014 XXX). Very recently, the Planck Collaboration (2014 XVII) has reported that the variation of dust emissivities across the field of interest is the limiting source of uncertainty when modelling the dust data.

Recent attempts to disclose the far-infrared (FIR) emissivity of HVCs are in line with these findings. Neither the investigation of different high-latitude clouds (Planck Collaboration, 2011 XXIV) nor the stacking of GALFA-H I compact clouds (Saul, Peek and Putman, 2014) has detected significant FIR emission.

Despite these odds, Miville-Deschênes et al. (2005b) generate a simple model of the dust emission from HVC complex C based on its H I column density and find a faint, but significant dust emissivity for the HVC regime. A similar approach, combined with an investigation of the chance correlation of H I and dust emission, yields a $> 3\sigma$ detection of dust towards complex M (Peek et al., 2009).

Here we investigate the HVC located at $(l, b, v_{\mathrm{LSR}}) = (125\,^{\circ}, 41\,^{\circ}, -207\,\mathrm{km\,s^{-1}})$, hereafter HVC125. The cloud has been previously studied with the Effelsberg 100 m telescope (Brüns, Kerp and Pagels, 2001) and the Westerbork Synthesis Radio Telescope (Braun and Burton, 2000). In combination, these data sets disclose a two-phase structure and a head-tail morphology of the HVC of interest. This might indicate ram-pressure interaction which in turn results in a reduced formation time of H_2 and dust (Guillard et al., 2009; Röhser et al., 2014). The single-dish data suggest that the warm phase is stripped off the core of the HVC. The cold component has an extraordinarily narrow line width around $\Delta v = 2\,\mathrm{km\,s^{-1}}$ FWHM, equivalent to a kinetic temperature of $T_{\mathrm{kin}} \lesssim 85\,\mathrm{K}$. The compact spatial structure of a few arcminutes, high brightness temperature of $T_B \gtrsim 10\,\mathrm{K}$ in the single-dish data, and low kinetic temperature make HVC125 one of the most promising HVCs in terms of detection probability of FIR dust emission.

4.2 Data

We used data from the recently finished Effelsberg-Bonn H I Survey (EBHIS, Kerp et al., 2011; Winkel et al., 2010; Winkel et al., 2016a) to study the neutral atomic hydrogen in HVC125.

For the FIR data, we used the latest release of the Planck data at 857 GHz (Planck Collaboration, 2015 VIII). We chose the highest frequency of the Planck data because the relative contributions of the cosmic microwave background and the cosmic infrared background to the uncertainty in modelling the foreground both decrease in proportion to the frequency (Planck Collaboration, 2014 XVII, their Fig. C.1).

To account for the differences in angular resolution, the dust data were smoothed to the angular resolution of the EBHIS data by Gaussian convolution. The final angular resolution is 10.83′. The EBHIS data have a spectral resolution of $1.49\,\mathrm{km\,s^{-1}}$ at a channel spacing of $1.28\,\mathrm{km\,s^{-1}}$.

4.3 Analysis

In the following, we use N to refer to the H I column density $N_{\mathrm{H\,I}}$ and I to refer to the FIR intensity at 857 GHz. When analysing the dust content of HVCs by comparing their H I column density to their dust content, the most challenging step is the estimation of the foreground dust emission. Because of the complexity and uncertainty of the correlation of dust and gas, an accurate and robust determination of the foreground component is of the utmost importance (e.g. Peek et al., 2009; Saul, Peek and Putman, 2014).

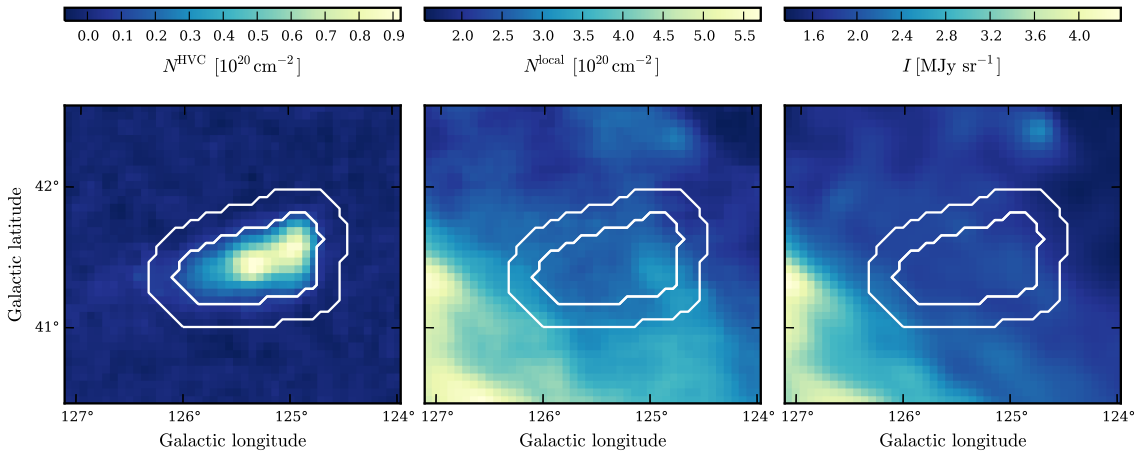


Figure 4.1: **Left:** EBHIS column density map N^{HVC} of the HVC ($-230\,\mathrm{km\,s^{-1}} < v_{\mathrm{LSR}} < -190\,\mathrm{km\,s^{-1}}$) **Center:** EBHIS column density map N^{local} of the local foreground emission ($-190\,\mathrm{km\,s^{-1}} < v_{\mathrm{LSR}} < +30\,\mathrm{km\,s^{-1}}$). **Right:** FIR intensity $I_{857\,\mathrm{GHz}}$ from Planck Collaboration (2015 VIII). The inner white contour line corresponds to the 5σ noise level in the HVC column density map. The outer contour marks an annulus that contains the same number of pixels as the tight HVC mask.

The H I data allow us to distinguish between local foreground emission and HVC emission via the radial velocity. For HVC125, we selected the velocity range ($-230\,\mathrm{km\,s^{-1}}, -190\,\mathrm{km\,s^{-1}}$) for the HVC and the remaining range ($-190\,\mathrm{km\,s^{-1}}, +30\,\mathrm{km\,s^{-1}}$) for the foreground emission. The corresponding column density maps are shown in Fig. 4.1. Moreover, we present an image of the FIR intensity at 857 GHz in the direction of HVC125. The inner contour outlines the 5σ level of the HVC H I column density. The outer contour marks an annulus around the HVC that contains as many pixels as the inner, narrow mask. In the following, we use these masks to determine the foreground dust contribution and the dust emissivity of the HVC.

For this analysis, we decided not to include the statistical uncertainties from the data noise of $\sigma_{\mathrm{RMS}} = 90\,\mathrm{mK}$ for the EBHIS data and $\sigma_{\mathrm{RMS}} = 0.014\,\mathrm{MJy\,sr^{-1}}$ for the Planck data at 857 GHz. We refer again to Fig. C.1 in Planck Collaboration (2014 XVII), which shows that at this frequency the analysis is dominated by uncertainties from the foreground estimation.

4.3.1 Standard approach

The standard approach (e.g. Miville-Deschênes et al., 2005b; Planck Collaboration, 2011 XXIV) used to evaluate the dust content of HVCs is the superposition of different H I column density maps N^i to model the FIR intensity:

$$I(x,y) = \epsilon^{\mathrm{local}} N^{\mathrm{local}}(x,y) + \epsilon^{\mathrm{HVC}} N^{\mathrm{HVC}}(x,y) + Z \tag{4.1}$$

Here, I is the FIR intensity, ϵ denotes the dust emissivity per hydrogen nucleon, and Z is a constant offset; I and N^i are two-dimensional images with (x,y) denoting the spatial position of the pixel.

We fitted Eq. (4.1) with a least-squares approach to quantify the parameters $\epsilon^{\mathrm{local}}$, ϵ^{HVC}, and Z. To investigate the influence of the spatial area that is fitted on the results, we performed the fit for three different spatial masks: a tight mask around the HVC, a more extended mask, and the full image. The results are compiled in Table 4.1. The large variations for different spatial masks and the uncertainties

Mask	$\epsilon^{\mathrm{local}}$	ϵ^{HVC}	Z
Small	0.26 ± 0.48	0.05 ± 0.38	1.14 ± 1.31
Wide	0.41 ± 0.09	0.029 ± 0.11	0.76 ± 0.26
None	0.61 ± 0.005	-0.05 ± 0.03	0.26 ± 0.02

Table 4.1: Emissivities ϵ^i for the different velocity components and offset Z, according to Eq. (4.1). Units are MJy sr^{-1}/10^{20} cm^{-2} for the emissivities and MJy sr^{-1} for the offset. The fit uncertainties are asymptotic standard errors, taken from the covariance matrix.

on the fit parameters emphasise that this approach is very sensitive to changes in the area of interest. Furthermore, this approach only relies on two parameters (Eq. 4.1) to describe the foreground: the local dust emissivity $\epsilon^{\mathrm{local}}$ and the offset Z and so it cannot account for multiple features at different radial velocities, possibly exposed to different physical environments. For the full field, we show the modelled FIR intensity I^{Standard} and the residual $I - I^{\mathrm{Standard}}$ in Fig. 4.3, left column.

4.3.2 Generalised linear model for foreground estimation

To overcome the limitations of the standard approach, we applied a generalised linear model (GLM; for a review see Madsen and Thyregod, 2010; de Souza et al., 2014) to the data. For this, we assume that each channel T_B^i of the H I cube can contribute individually to the FIR intensity. In consequence, we do not rely on a vague definition of the velocity range for local and HVC gas.

Within the GLM, the FIR intensity can be written as

$$I^{\mathrm{local}}(x, y) = \sum_i T_B^i(x, y)\beta^i + Z. \tag{4.2}$$

The β^i are the GLM coefficients and can be understood as emissivity per spectral channel. The parameter Z is a global offset to the model. Because of the co-linearity between neighbouring spectral channels, the assumption of independent data for ordinary least-squares fitting is violated. To break this degeneracy, we controlled the GLM with lasso regularisation (Tibshirani, 1996) and minimised the term

$$||I^{\mathrm{local}} - T_B^i\beta^i - Z||^2 + \alpha \cdot ||\beta^i||_1. \tag{4.3}$$

The first part of this term corresponds to the regular least-squares approach and gives the residual sum of squares. The second part is the penalty term and ensures that the coefficients β^i are chosen as sparsely as possible. The strength of this second term is scaled by α. We used simulations and cross validation to optimise this regularisation strength (see Sect. 4.4). The regularised GLM and the cross validation was implemented via `scikit-learn` (Pedregosa et al., 2011), a machine-learning package for `python`.

For our application to the H I and dust data, only data points outside of the wide mask and with $v_{\mathrm{LSR}} > -190\,\mathrm{km\,s^{-1}}$ were considered because they are unrelated to the HVC. This ensures that the HVC signal is not accidentally removed by our foreground subtraction. While the distinction between HVC and local gas is very straightforward in this particular case, there are other cases where gas at high or intermediate velocities is difficult to disentangle (e.g. HVC Complex M, Wakker, 2001). In these cases, we cannot simply remove the foreground by applying a threshold in radial velocity, but have to rely only on spatial masking of the HVC.

The resulting GLM coefficients β^i and the mean H I spectrum are shown in Fig. 4.2. The offset is $Z = 0.73\,\mathrm{MJy\,sr^{-1}}$. We find that the dust emission towards HVC125 can be modelled well with approximately seven different emissivities. The narrowness of the GLM coefficients is the product of

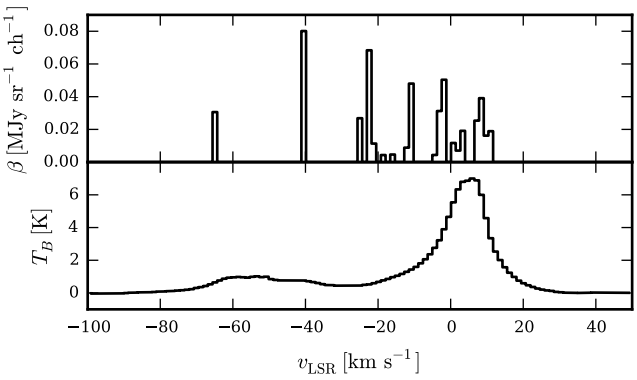


Figure 4.2: **Top:** GLM coefficients β^i for each channel based on the cross-validated lasso regression (Eqs. 4.2 and 4.3). **Bottom:** Mean H I spectrum of the data cube.

the applied regularisation. Visually inspecting the HI data cube discloses that each individual GLM coefficient is indeed associated physically with an individual HI structure. However, the exact matching of H I structures to FIR emissivity peaks varies with the regularisation strength and demands further justification (Sect. 4.4). The sparse filling of the whole HI spectrum with GLM coefficients argues strongly for a localisation of the dust within the cold neutral medium (CNM) filaments and is inconsistent with a homogenous mixture of dust and gas on all linear scales. However, we note that the discrete nature of the FIR emissivities for each spectral channel is primarily a schematic description.

Similar to the standard approach, we show the modelled FIR intensity and the residual (Fig. 4.3, right column). A comparison of the two different approaches shows that the GLM can cover a wider range of FIR intensities and manages to account for a larger number of different features. The residual is less structured and the remaining structures stem from the CIB anisotropies (compare Sect. 4.4).

Furthermore, we show the histograms from the residual FIR intensity in Fig. 4.4. For both models, we consider only data points outside of the wide mask because we are only interested in the capability to model the foreground emission. For the standard approach, we find that the residual is rather broad and irregular. In contrast, the GLM residual is narrow and symmetric and its shape can be approximated as Gaussian.

4.3.3 Measurement of the HVC dust content

To quantify the hypothetical dust emission from HVC125, we corrected the FIR intensity map for the local foreground emission by using the GLM, $I^{\rm HVC} = I - I^{\rm GLM}$ (Fig. 4.3, bottom right). We investigated the correlation between the HVC H I column density and the HVC FIR intensity within the narrow mask with a linear correlation in the Bayesian framework (e.g. D'Agostini, 2003). Thus, we sampled the posterior given by

$$p(\epsilon, \sigma|\mathcal{D}) \propto \mathcal{L}(\mathcal{D}|\epsilon, \sigma) p(\epsilon, \sigma) \tag{4.4}$$

where $\mathcal{D}$ is the data vector $\mathcal{D} = (N^{\rm HVC}, I^{\rm HVC})$. The likelihood is given by

$$\mathcal{D}|\epsilon, \sigma \sim \mathcal{N}(I^{\rm HVC} - \epsilon \cdot N^{\rm HVC}, \sigma^2). \tag{4.5}$$

The parameter ϵ is the dust emissivity per hydrogen nucleus and σ is the standard deviation of the Gaussian likelihood.

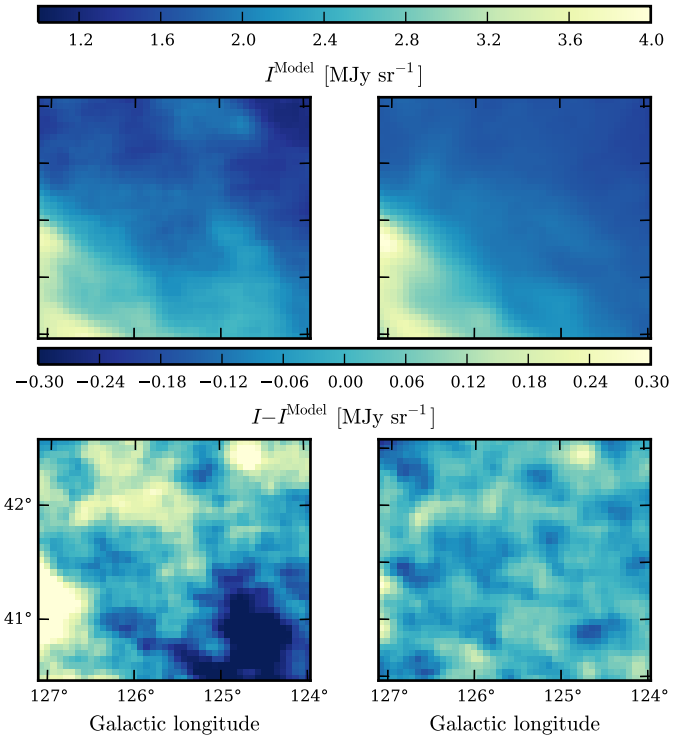


Figure 4.3: Model of the FIR intensity (top) and residual emission (bottom). We show the results of the standard approach (Sect. 4.3.1) in the left column and the GLM (Sect. 4.3.2) in the right column.

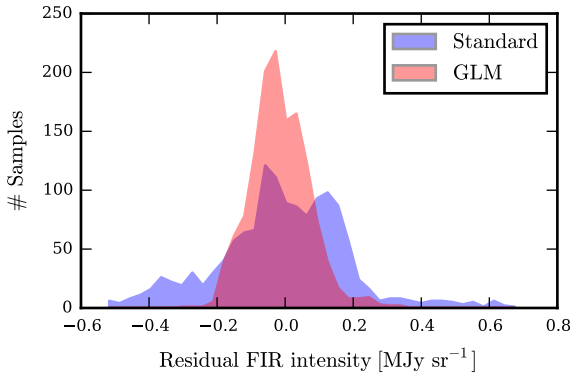


Figure 4.4: Residual histograms after correcting for the local foreground FIR intensity outside of the wide mask for the two different methods.

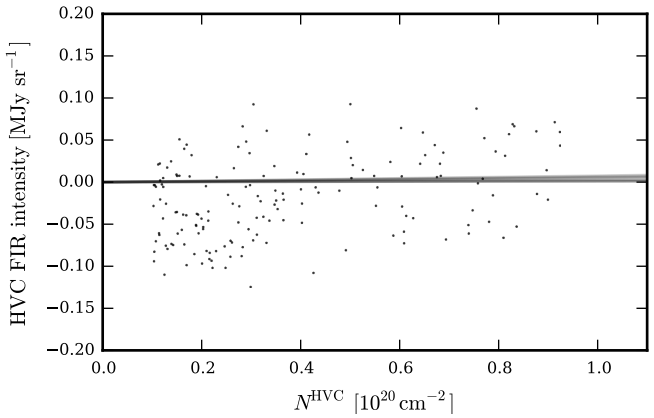


Figure 4.5: Linear correlation between HVC H I column density and foreground-subtracted FIR intensity in the narrow mask. The lines correspond to thirty randomly chosen MC samples (Eqs. 4.4 to 4.6) from the Bayesian line fit.

We selected minimally informative priors for our parameters:

$$\begin{aligned} \epsilon &= \tan\phi \\ p(\phi) &= \mathrm{Uniform}(\phi, 0, \pi/2) \\ p(\log\sigma) &= \mathrm{Uniform}(\log\sigma, -8, 4) \end{aligned} \tag{4.6}$$

For the emissivity ϵ, we sampled uniformly in $\arctan(\epsilon)$ to avoid a bias towards larger values. A simple scale-invariant prior was chosen for the scatter σ (Jeffreys, 1946) in a reasonable range. Furthermore, we reasonably chose to restrict ϵ to positive values. An offset FIR intensity is already part of the GLM, hence we choose not to include a further offset here. Lastly, we note that we did not account for the spatial covariance in the image owing to the non-flat CIB angular power spectrum. For a complete and proper treatment of this effect, we would also require an accurate determination of the spatial covariance due to the beam shapes of the different data sets and the sampling on the pixel grid.

The model was sampled with `emcee` (Foreman-Mackey et al., 2013), a `python` implementation of the affine-invariant ensemble sampler for Markov chain Monte Carlo (MCMC, Goodman and Weare, 2010). Figure 4.5 shows randomly drawn samples of this model, applied to the narrow mask H I and dust data of HVC125. The posterior distribution of the individual parameters is presented in Fig. 4.6.

As in the standard approach (Table 4.1), we not only evaluated the narrow mask, but also the wide mask and the full field. The resulting fit parameters are summarised in Table 4.2.

We find that the emissivity ϵ and scatter σ parameters are well-sampled. The HVC emissivity is not normally distributed and illustrates that the model strongly prefers zero emissivity. The 99.87 % upper limit, corresponding to 3σ, is 0.021 MJy sr^{-1}/10^{20} cm^{-2} and thus an order of magnitude below typical Galactic ISM values (Planck Collaboration, 2011 XXIV). Moreover, the parameters vary only slightly for different masks. As expected, the scatter increases with the size of the mask, or equivalently with the number of data points.

4.4 Verification of the GLM

In the following, we present our thorough investigation of the performance of the GLM; we conducted a series of simulations and tests to explore its advantages and limitations. Moreover, we used these simulations to properly select the regularisation strength α (Eq. 4.3).

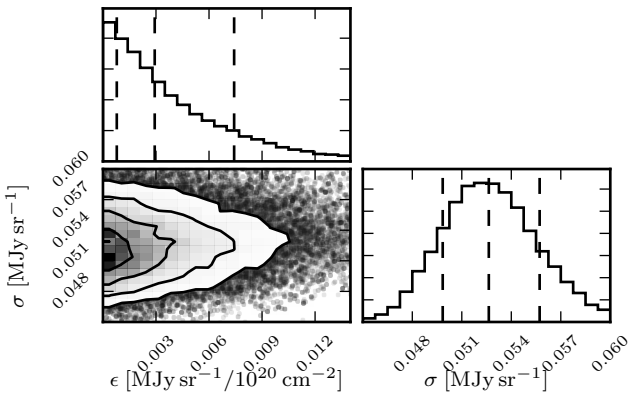


Figure 4.6: Posterior distribution for ϵ and σ, obtained from sampling Eq. (4.4). The lines in the histogram indicate the 16th, 50th, and 84th percentile, equivalent to mean and $\pm 1\sigma$ for a Gaussian posterior.

	ϵ [MJy sr^{-1}/10^{20} cm^{-2}]			σ [MJy sr^{-1}]		
Mask	16%	Median	84%	16%	Median	84%
Small	0.001	0.003	0.007	0.049	0.053	0.056
Wide	0.001	0.002	0.006	0.045	0.047	0.049
None	0.001	0.003	0.008	0.064	0.065	0.066

Table 4.2: Emissivity ϵ and scatter σ for the different masks. The numbers indicate the median of the posterior distribution and the 16th and 84th percentile. For a Gaussian posterior, this is equivalent to median $\pm 1\sigma$.

4.4.1 Construction of the simulations

The simulations were generated in the following way:

1. We generated an artificial, noisy spectrum of GLM coefficients;
2. we convolved this coefficient spectrum with the measured H I data cube to generate a map of foreground FIR intensity;
3. We added a random realisation of the CIB to the foreground.

Because we did not know the real nature of dust emissivities at different radial velocities, we tested different approaches. For all of them, we assumed that the dust emissivity occurs only in spectral channels in which significant H I emission is found. We distinguished between the following types of GLM coefficients (see also Fig. 4.7):

Spiky: One spectral channel wide, between 2 and 7 components. The spikes vary in amplitude by up to 40 %.

Smooth: Similar to the spiky input, but the spectrum of GLM coefficients is smoothed with random Gaussian kernels with FWHM between 2 km s^{-1} and 8 km s^{-1}.

Flat: This represents the standard approach to investigating the H I-FIR correlation. We choose a constant dust emissivity with 10 % random fluctuation for local gas ($v_{\mathrm{LVC}} > -30$ km s^{-1}) and a 20 % lower dust emissivity for the gas at intermediate velocities. (-95 km s^{-1} < v_{IVC} < -30 km s^{-1})

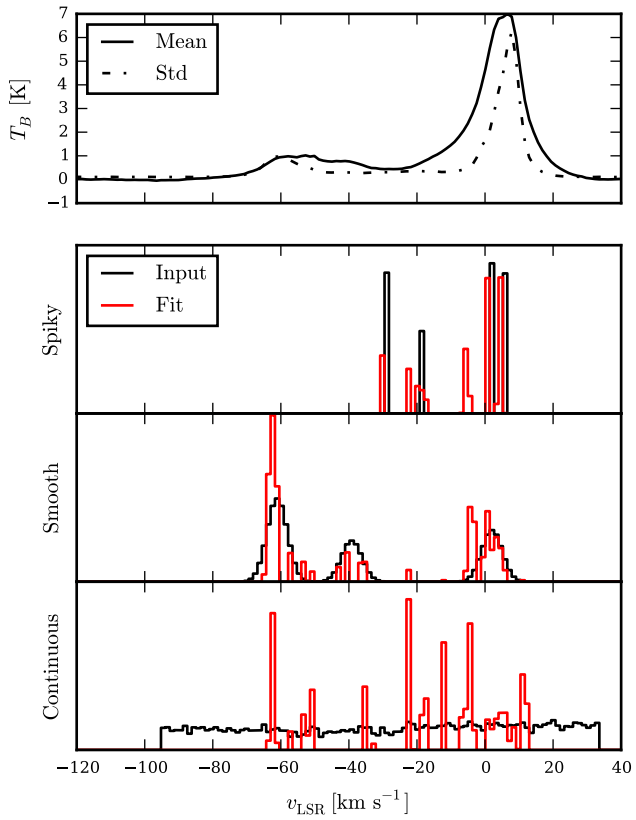


Figure 4.7: Top panel: Spectrum of mean and standard deviation of the full H I data cube. Bottom panels: Input spectra (black) of GLM coefficients to simulate FIR intensity maps. The reconstruction by the GLM is shown in red and has been shifted to the left by one channel for illustration purposes. See the text for a detailed description.

The amplitudes of all these GLM coefficient spectra were normalised such that the resulting mean intensity of the foreground FIR map was equal to the value found in the measured FIR intensity map (Fig. 4.1, right panel). This ensured that we modelled the proper ratio of foreground to background.

To combine the foreground FIR intensity with the CIB, we generated random realisations of the Gaussian random field based on the CIB angular power spectrum taken from Planck Collaboration (2014 XXX). We extrapolated the angular power spectrum from their Table D.2 with a power-law. This extrapolation does not hold for large angular separations i.e. small multipoles, but this effect is negligible for the present field size of only $2° \times 3°$. Here, a power law is a valid approximation (G. Lagache, priv. comm.). To obtain the proper ratio of foreground and background components, we also scaled the simulated CIB to match the mean and fluctuation amplitude given in Planck Collaboration (2011 XVIII, their Table 5). Despite the non-flat angular power spectrum of the CIB, this is possible because the angular size of the fields investigated in Planck Collaboration (2011 XVIII) is similar to the field size in the present study. Finally, the simulated CIB was smoothed to the angular resolution of 10.8′, which was used for all data throughout this study.

We generated 1000 of these simulations for each type of spectrum (spiky, smooth and flat) and applied the GLM to reconstruct the total FIR intensity. We compare the input and outcome of the GLM coefficient spectra in Fig. 4.7.

4.4.2 GLM performance on simulations and choice of regularisation strength

The visual inspection of Fig. 4.7 shows that, despite the CIB confusion, we were able to properly reconstruct the shape and position of the GLM coefficients for the spiky and smooth case. Because of the

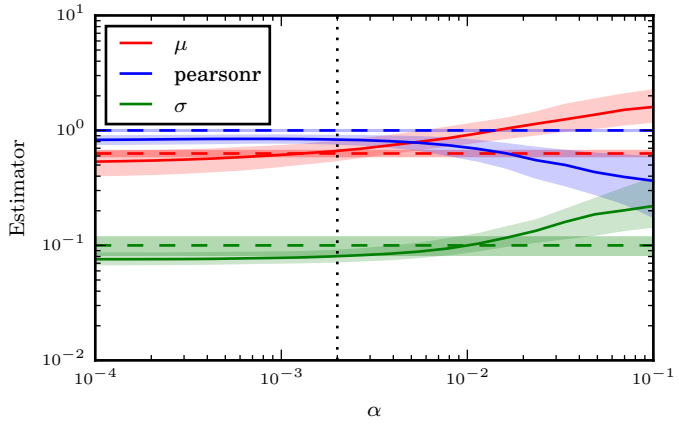


Figure 4.8: Evaluation of the reconstructed CIB mean μ, CIB fluctuation amplitude σ and Pearson's r of input CIB image and reconstruction. This is based on the smooth input GLM coefficients (third panel from the top in Fig. 4.7). Dashed lines indicate the input, solid lines the reconstructed quantities. Contours correspond to 1σ uncertainties. The vertical line indicates our choice of α.

GLM design, we did not recover the exact shape of the flat input spectrum, but used the most relevant channels to create an approximation that produces an accurate model. We note again that the discrete description of the dust emissivities for the spectral channels is a result of our approach and does not necessarily reflect the physical conditions.

The results are furthermore evaluated via three quantities: the mean strength of the CIB, the strength of its fluctuations and the Pearson's r correlation coefficient of the reconstructed and the input CIB image. This mainly ensures that the GLM neither over- nor underfits the data. For the CIB mean μ and fluctuation amplitude σ, the values are taken from Planck Collaboration (2011 XVIII).

We investigated how these quantities vary as a function of the regularisation strength α (Fig. 4.8). Our example was generated for the smooth input GLM coefficients. For the other cases, the results are presented in Sect. C. Dashed lines indicate the input values, solid lines are the results of our application of the GLM to the simulated data. The contours correspond to 1σ uncertainties.

We find that for $\alpha \lesssim 4 \times 10^{-3}$, there is an agreement between input and reconstruction for all three estimators within their respective uncertainties. We chose to set $\alpha = 2 \times 10^{-3}$ for our analysis of HVC125 (vertical line in Fig. 4.8).

The **Pearson's *r*** correlation coefficient between the input and reconstructed CIB is remarkably constant and close to 1, even for a very weak regularisation that allows a great number of GLM coefficients. This illustrates that chance correlation by individual H I spectral channels is not very efficient in mimicking the CIB signal. For a very strong regularisation, the model cannot account for all the dust-emitting H I components and the residual map is dominated by foreground emission, yielding a poor correlation to the input CIB.

The **CIB mean μ** is the quantity that varies strongest for different values of the regularisation strength α. We find that for $\alpha = 10^{-3}$ to $\alpha = 4 \times 10^{-3}$, the CIB mean is properly estimated.

The **CIB fluctuation amplitude σ** is systematically underestimated if the regularisation is too weak, meaning that the model overfits the data. Other than in the Pearson's r correlation, this hints towards a mimicking of background CIB by chance correlation with some H I channels. This effect, however, is of the order of 10 % and is within the uncertainties for our choice of $\alpha = 2 \times 10^{-3}$.

Furthermore, we find that the choice of α and the evaluation of the different quality estimators does not vary strongly for different types of input GLM coefficients (Fig. C.1, C.2). Because we cover a variety of shapes (Fig. 4.7) and demonstrate that the GLM can properly remove the foreground and uncover the faint CIB emission in each individual case, we conclude that it is well suited for the search of dust in HVCs.

To conclude, our approach is not constructed to precisely measure the dust emissivity of the individual

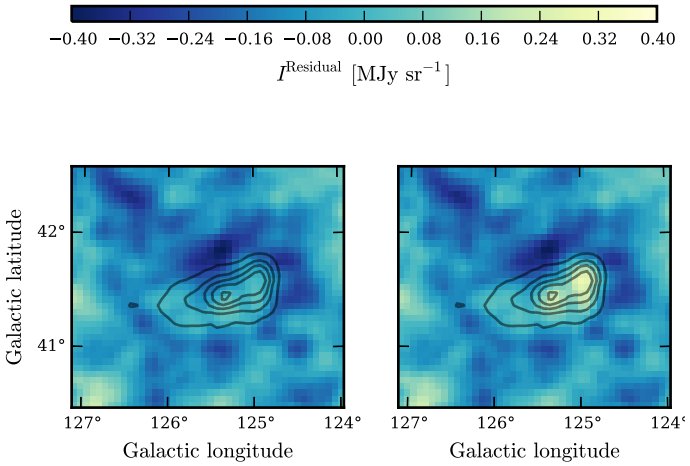


Figure 4.9: GLM reconstruction of a simulated dusty HVC. The image is dominated by the simulated CIB anisotropies. The HVC dust emissivity corresponds to 10% (left) and 30% (right) of typical Galactic values. The black contours correspond to the HVC column density starting at 10^{19} cm^{-2} and increasing in steps of 2×10^{19} cm^{-2}.

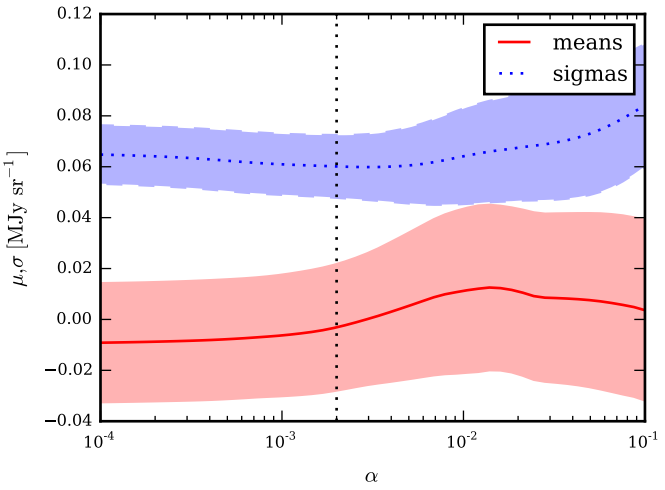


Figure 4.10: Results of the cross validation to investigate our choice of the regularisation strength. The solid lines correspond to the mean and standard deviation of the residual emission in the test sample. Contours correspond to 1σ uncertainties. The vertical line illustrates our choice of α.

clouds and filaments; rather, it is designed to remove the local foreground for studies of faint FIR signals such as CIB emission or dust in HVCs. To further verify this, we simulated a FIR intensity map in which the HVC has a dust emissivity of 10% and 30% of typical Galactic values and removed the foreground emission (Fig. 4.9). We find that for an emissivity of only 10% Galactic, the HVC can barely be distinguished from the residual CIB fluctuations. For the higher emissivity, the HVC signal is significantly stronger than these fluctuations and would imply a detection in our Bayesian line fit.

4.4.3 Estimation of alpha and the uncertainties via cross validation

To further validate the regularisation strength α in Eq. (4.3), we used structured cross validation (e.g. Picard and Cook, 1984) with the HVC mask as kernel. For this purpose, we shifted the wide HVC mask to random positions in our field. The GLM was then computed on the data outside of this mask (training sample). Subsequently, we compared the data and the model inside the mask (test sample) by inspecting the residual mean and its standard deviation. This was done for 1000 different, random HVC mask positions. The resulting means and 1σ uncertainties of residual mean and residual standard deviation in

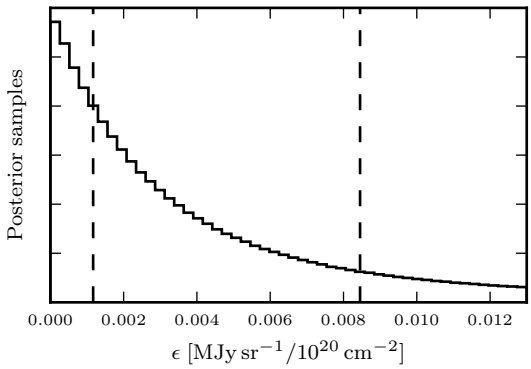


Figure 4.11: Stacked posterior distribution for the HVC dust emissivity ϵ from simulated data. The procedure is similar to the one described in Sect. 4.3.3. The lines in the histogram indicate the 16th and 50th percentile.

the test sample are shown in Fig. 4.10 for a range of regularisation strengths.

If the regularisation is too strong, the model will eventually fail and will not properly account for the complexity of the data. Based on the simulations, we chose $\alpha = 2 \times 10^{-3}$. This is supported by the findings of the present cross validation which shows that this value is the strongest possible regularisation before the model begins to lose accuracy.

We used the same technique to estimate the uncertainties for a fixed value of the regularisation strength $\alpha = 2 \times 10^{-3}$ and find that the mean residual FIR intensity is -0.003 ± 0.025 MJy sr^{-1}. The standard deviation, equivalent to the amplitude of the CIB anisotropies for the present field-size is 0.060 ± 0.012 MJy sr^{-1}.

The same cross validation analysis for the standard approach yields a mean of -0.040 ± 0.129 MJy sr^{-1} and a standard deviation of 0.129 ± 0.054 MJy sr^{-1}. The larger uncertainties with respect to the GLM illustrate another advantage of the novel method. We note, however, that the standard deviation of the residuals differs significantly between the two methods. We discuss the implications further in Sect. 4.5.1.

4.4.4 HVC dust emissivities in simulations

To cross-check our result on the HVC dust emissivity (Sect. 4.3.3), we quantified its posterior distribution using simulated data in which no HVC dust emission is present. To generate the simulated FIR intensity maps, we used a smooth shape for the GLM coefficients (Fig. 4.7). After removing the foreground emission, we evaluated the HVC dust emissivity using a similar procedure to the one described in Sect. 4.3.3. We present the stacked posterior distribution for all 1000 simulations in Fig. 4.11. The 16th, 50th, 84th and 99.87th percentiles correspond to a dust emissivity of 0.001, 0.008, 0.111, and 0.555 MJy sr^{-1}/10^{20} cm^{-2}, respectively.

Despite the absence of a HVC signal in these simulations, the emissivities are lower than the values derived from real data (Fig. 4.6). The large values for the 84th and 99.87th percentile are the consequence of the large sample of simulations and yield a heavy tail in the posterior.

4.5 Discussion

4.5.1 Quality of the foreground model

To address the potential dust content of HVC125, a reliable estimation of the foreground FIR intensity is the most challenging step and of the utmost importance.

After early studies of the H I-dust correlation did not account for this foreground emission (Wakker and Boulanger, 1986), the often-applied standard approach assumes a simple linear relation between FIR intensity and H I column density for different H I column density maps (Eq. 4.1, e.g. Miville-Deschênes et al., 2005b; Planck Collaboration, 2011 XXIV). This approach is limited by the uncertain separation of H I in different foreground and cloud components. Furthermore, it only has a low number of degrees of freedom and hence cannot properly treat the multiple, complex features at different radial velocities. Thus, clouds with different emissivities or illumination by the ISRF cannot be accounted for if they are in the same column density map. In the case of HVC125, the separation between HVC emission and foreground emission is straightforward and the results are not affected by uncertainties introduced by different definitions of N^{HVC} and N^{local}. Commonly, a clean separation of the different components for the more complex intermediate-velocity and local gas is hardly feasible. The residual histogram (Fig. 4.4) shows that the standard approach yields a broad and asymmetric residual signal. Moreover, Table 4.1 shows that the fit parameters vary strongly and are poorly determined for different spatial masks.

We apply a GLM to use each individual channel of the H I observations, restricted by the regularisation to overcome the degeneracy between the individual, correlated channels and to avoid overfitting. We justified the choice of the regularisation strength via simulations and cross validation, yielding consistent results. The success is well demonstrated by the narrow, symmetric distribution in the FIR intensity histogram. The results of our cross validation yields no bias with a mean residual of $0.003 \pm 0.025\,\mathrm{MJy\,sr^{-1}}$. The mean standard deviation of the residual maps of $0.060 \pm 0.012\,\mathrm{MJy\,sr^{-1}}$ result from the underlying CIB fluctuations (Planck Collaboration, 2011 XVIII; Planck Collaboration, 2014 XXX). Moreover, the parameters are barely affected by different mask sizes, unlike the standard approach (Table 4.2). The spectrum of GLM coefficients illustrates that the local H I does not contribute uniformly to the foreground FIR intensity, but different features need to be accounted for. Aside from the complex, bright emission from the gas around $0\,\mathrm{km\,s^{-1}}$, we find different filaments at intermediate velocities. Within the GLM framework, these features are recognised and the FIR intensity is successfully modelled.

The investigation of the standard deviation via cross validation generates significantly different results between the standard approach and the GLM. Here, the former is in agreement with studies of high-latitude fields of similar sizes using the very same approach (Planck Collaboration, 2011 XVIII; Planck Collaboration, 2014 XXX). However, our simulations show that we can properly reconstruct the different simulated CIB properties with the GLM approach. Accordingly, the tension in the strength of the CIB fluctuations requires further investigation in a future study. We note at this point that we constrain ourselves to this particular field on the sky which can only provide very limited insights into the global CIB properties.

4.5.2 Dust content of the HVC

We find that the foreground-corrected FIR intensity map does not contain any significant contribution from the HVC. This is clearly seen in the posterior distribution of the emissivity ϵ (Fig. 4.6, top), which strongly favours zero dust emissivity. Given our model, there is a 99.87% probability that the emissivity of the HVC is below $0.02\,\mathrm{MJy\,sr^{-1}}/10^{20}\,\mathrm{cm^{-2}}$. This is an order or magnitude lower than typical Galactic values found by Planck Collaboration (2011 XXIV).

Similar non-detections have been obtained in other studies of dust in HVCs (Planck Collaboration, 2011 XXIV; Saul, Peek and Putman, 2014). A noteworthy exception is HVC complex M (Peek et al., 2009). However, complex M is on the transition of HVC/IVC classification and could possibly be related to the IV arch (Wakker, 2001).

Because of the very low kinetic temperature, its compact structure, and high brightness temperature, HVC125 is one of the most promising candidates for the detection of dust in HVCs. Moreover, the

head-tail structure can indicate the formation of dust and molecules via the increased pressure (Gillmon et al., 2006; Guillard et al., 2009; Röhser et al., 2014). Our non-detection shows that even for this candidate, the upper limit is significantly below typical ISM dust emissivities.

4.6 Conclusion

To explore the properties and the origin of HVCs, their potential dust content can be a powerful tool. We pointed out the importance and the difficulty of estimating the foreground contribution to the FIR intensity in the classical framework. Without an accurate and robust determination of this foreground emission, it is not possible to evaluate the expected, very faint dust emission from HVCs.

We have presented a new approach to address this issue by applying a GLM to evaluate the correlation of atomic neutral hydrogen and FIR dust emission. The GLM offers the opportunity to attribute an individual dust emissivity to each spectral channel of the H I data. To regularise the fit, we introduced linear penalty terms for the GLM coefficients. The investigation of the residual FIR intensity shows that the GLM yields significantly lower residual emission than the standard approach. Furthermore, the distribution function of the residual is more symmetric, has fewer outliers, and the derived model is more robust to variations of the spatial area that is approximated.

After correcting for the foreground dust emission via this GLM, we analysed the potential dust content of the HVC. Using a line fit in the Bayesian framework, we derived that the dust emissivity at 857 GHz is 0.021 MJy $\mathrm{sr}^{-1}/10^{20}\,\mathrm{cm}^{-2}$ at the 99.87% confidence level. This is more than an order of magnitude lower than typical ISM emissivities. This shows that even for this promising candidate with low kinetic temperatures, high brightness temperature, and head-tail structure, the detection of dust is not feasible via the correlation of dust and neutral gas.

Resume

The following section is not part of the original publication by Lenz, Flöer and Kerp (2016) on which this Chapter is based. Instead, we relate the content of this Chapter to the thesis as a whole and highlight the scientific and methodological progress.

For this Chapter, we have developed a new approach to removing Galactic foreground emission from FIR intensity maps. It was possible to illustrate accurately and significantly that even for one of the best candidate HVCs, there is no hint of dust emission. Furthermore, we have shown that the CIB fluctuations are the dominant source of uncertainty. The algorithm presented in this Chapter has the potential to remove the Galactic foreground automatically and therefore access the CIB. In the following Chapter, we will continue to improve this approach and present the first large-scale measurements of the infrared background light as seen by *Planck*, robustly corrected for the foreground component.

With respect to Chapter 3, we have improved the methods to investigate the correlation by employing Bayesian statistics, modern machine learning tools, and new data formats such as the HEALPix, based on the work of Górski et al. (2005).

CHAPTER 5

Large-scale studies of the cosmic infrared background

Preface

The following section relates the content of this Chapter to the previous ones and highlights the scientific and methodological progress throughout this thesis.

In the previous Chapter, the search for dust in a cold, compact high-velocity cloud (HVC) has eventually lead to the conclusion that this search is limited by the anisotropies of the cosmic infrared background (CIB). With the insights gained during the study of HVC125, we are at an excellent starting point to study the CIB on larger scales: The basic principles have been developed, we can benefit from the vast expertise of our colleagues from the *Planck* collaboration, and we have the optimal data sets.

Furthermore, the present Chapter requires that we familiarise ourselves with the basic principles and methods of cosmology, including primarily the work with angular power spectra and cosmological models. While much of this thesis is based on the science of the interstellar medium (ISM), it marks the transition to cosmology. The interdisciplinary nature of this work therefore merges the expertise from the two different fields, allowing each of them to excel previous results.

Abstract

The cosmic infrared background (CIB) contains a vast amount of information on the properties and evolution of distant far-infrared (FIR) galaxies. Many studies on the CIB are limited either by the insufficient removal of foreground Galactic dust or by a very limited sky coverage. We aim to uncover this background light on large scales, covering 20% of the sky in total and thereby doubling the currently available sky fraction. Thereby, we create the foundation for precise evaluations of cosmological models of the FIR galaxies and their dark matter halos. For this work, we greatly benefit from the latest generation of *Planck* FIR and HI4PI H I data. Using the full spectral information of the Galactic H I, we employ an automated approach to remove the foreground dust emission accurately and robustly. For this approach, the weighted sum of all H I spectral channels is combined via a generalised linear model and serves as template for the dust. Furthermore, we correct the FIR intensity maps for the cosmic microwave background (CMB) anisotropies and mask extragalactic point sources. We are finally able to generate maps of the amplitude of the CIB fluctuations at 353, 545, and 857 GHz for 20% of the full sky, which

are verified with respect to previous studies. Additionally, we provide the deconvolved, unbiased, and binned angular power spectra. We find that our power spectra are also in line with previous results from the *Planck* collaboration. These power spectra, particularly at the larger scales which have not been probed before, will help to put tighter constrains on the parameters of cosmological simulations, allowing the access to the reionisation history of the Universe and the star formation history (SFH) in dark matter halos.

5.1 Introduction

The extragalactic background light (EBL) contains the relic emission from galaxies across cosmic time (see Fig. 5.1 for the spectral energy distribution (SED) of the background light). Between wavelengths of 1 to 1000 μm, excluding the CMB, it is referred to as CIB. This part of the EBL consists of the infrared (IR) emission from dust within galaxies and was first discovered in the Cosmic Background Explorer (COBE, Boggess et al. 1992) data. The dust grains are heated by absorbing the UV radiation from young, luminous stars and re-emit the energy in the IR. By studying the ensemble properties and the redshift distribution of the CIB galaxies, we can improve our understanding of the formation and evolution of these galaxies. Throughout this Chapter, we will constrain ourselves to unresolved FIR emission of the CIB. Other studies, for example with the Spitzer Space telescope, focus on source counts at higher frequencies (e.g. Dole et al., 2004). For an extensive summary on this and on other properties of the CIB, we recommend the reviews by Hauser and Dwek (2001) and Lagache, Puget and Dole (2005), with the former containing a very detailed historical view on CIB predictions and measurements.

Individual dusty, star-forming galaxies are very difficult to detect at high redshifts because of the high density of unresolved sources (e.g. Lagache, Dole and Puget, 2003; Dole et al., 2004b). Using the statistical properties of the CIB, we can study the distribution of these sources in a statistical sense (Knox et al., 2001). The anisotropies directly probe the distribution and the properties of the FIR galaxies. The angular power spectrum can be used to study the SFH across cosmic times and the temperature evolution of dust in CIB galaxies (Planck Collaboration, 2011 XVIII; Planck Collaboration, 2014 XVIII). Furthermore, it can be used to constrain the galaxy bias, which is the relation of the distribution of galaxies and the underlying dark matter distribution (Planck Collaboration, 2014 XXX; Serra et al., 2014).

Up to now, only small, selected regions have been accurately removed for FIR foreground emission, leading to low sensitivities and missing information in the power spectrum on to largest scales. Here, we present an automatic and robust approach that allows to access a major fraction of the sky, resulting in maps of CIB anisotropies that can be used to test cosmological models on a new level.

To put constraints on the properties and evolution of the ensemble of CIB galaxies, it is necessary to compare the measured power spectra to cosmological models of the halo occupation of CIB galaxies. Throughout this Chapter, we constrain ourselves to deriving CIB maps at the highest possible quality. A comparison to simulations will be performed at a future point.

5.1.1 Disentangling Galactic thermal dust and the CIB

Incorrect modelling of thermal dust might be responsible for a significant bias on the estimation of the cosmological parameters (e.g. Remazeilles et al., 2016), extracted from measurements of the CMB fluctuations. Therefore, the correct disentanglement of CIB and Galactic thermal dust is a crucial part of the analysis. At lower frequencies, radiation processes such as synchrotron, free-free, and molecular line transitions need to be accounted for as well (Planck Collaboration, 2015 IX). The Planck Collaboration

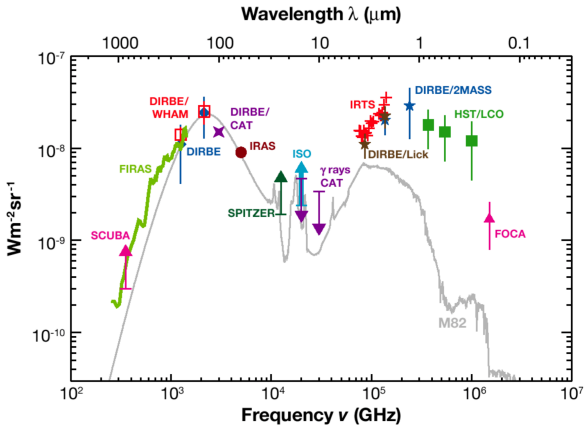


Figure 5.1: SED of the EBL over a wide range of frequencies, from the near UV to the millimetre regime. For comparison, the grey line shows the SED of the nearby starburst galaxy M82 (Chanial, 2003), normalised to the EBL intensity at 140 μm. Image taken from Lagache, Puget and Dole (2005). See their Table 1 for a full list of the references.

(2015 X) extracted multiple components from the multifrequency *Planck* and ancillary data, but no distinction is made between Galactic thermal dust and CIB emission. Thus, CIB emission leaks into the dust foreground map and introduces a systematic bias into this map (Planck Collaboration, 2016 XLVIII). For an estimate of the frequency-dependent ratio of the different components and to illustrate which process dominates which bands, we refer to (Planck Collaboration, 2014 XVII, their Fig. C.1).

To improve this part of the CMB component separation and to achieve an accurate distinction between Galactic dust and the CIB, two methods can be used and will briefly be described in the following: First, using the different statistical spatial properties (power spectrum) of the two components and second, using templates for the Galactic dust such as H I and CO.

By analysing its power spectrum, it has been found that the CIB scales approximately as l^{-1} (Planck Collaboration, 2011 XVIII, and references therein) where l is the multipole of the angular power spectrum. For the Galactic FIR cirrus, the power spectrum is steeper and scales with $l^{-2.6}$ (Bracco et al., 2011). There have been several approaches to disentangle the different components in the FIR sky. The Planck Collaboration (2014 XII) uses four different approaches to separate the CMB from other components, one of them being a needlet version of the internal linear combination (ILC) algorithm (NILC). For an introduction to the algorithm, we refer to Delabrouille et al. (2009), who applied this to remove foreground emission for Wilkinson Microwave Anisotropy Probe (WMAP, Bennett et al. 2003) data. Oppermann and Enßlin (2014) extend this work by applying information field theory (IFT) methods, assuming Gaussian priors for the CMB and log-normal priors for the foreground components. Compared to flat priors which are commonly used, they find that their results for simulated data are superior and even valid if the assumption of log-normal fields is not precisely fulfilled. Very recently, the Planck Collaboration (2016 XLVIII) apply a variation of the NILC algorithm to the High Frequency Instrument (HFI) *Planck* data in order to separate CIB and Galactic cirrus for the first time, using their characteristic spatial properties. They find an excellent agreement with previous results and their foreground and CIB maps will be publicly available in the near future. Unfortunately, all of these approaches yield a variable resolution of the final maps and thus contain specific information on different angular scales. Fig. 5.2 shows the resulting resolution of the study conducted by Planck Collaboration (2016 XLVIII) for the CIB at large scales. In order to deconvolve the angular power spectrum to correct for the beam smearing, it has to be assumed that the full width at half-maximum (FWHM) of the beam is constant. Hence, the variable resolution makes it very difficult to estimate the unbiased, deconvolved power spectrum (Hivon et al., 2002; Ponthieu, Grain and Lagache, 2011, see also Sect. 5.3.2).

The more traditional approach to clean CIB images from Galactic cirrus is to use templates such as H I and CO (Planck Collaboration, 2011 XVIII; Planck Collaboration, 2014 XXX, and references therein).

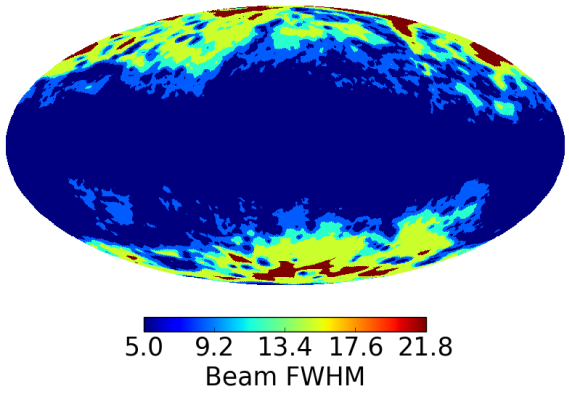


Figure 5.2: Effective beam FWHM of the dust maps derived in Planck Collaboration (2016 XLVIII). The spatially varying beam FWHM is the same for all frequencies. Their approach preserves the 5′-scale power of the thermal dust in the high signal-to-noise regions of the sky, but leads to a high variation in the final resolution of the image at higher Galactic latitude. Image taken from Planck Collaboration (2016 XLVIII).

They have the benefit of producing a constant angular resolution across the entire image. Up to now, these studies had only access to low resolution H I data such as the Leiden/Argentine/Bonn Survey (LAB, Kalberla et al. 2005) for larger scales, or only a very limited sky coverage for high resolution H I data (e.g. Martin et al., 2015). Furthermore, the applied methods require manual fine-tuning of the parameters and are thus very time-consuming when applied to larger areas.

We aim to expand the latter approach, using new full-sky H I data at resolutions four times better than previous large-scale H I data. Moreover, we use advanced methods (Sect. 5.3) that have previously been verified to improve the analysis of selected fields (Chapter 4) and can be applied to large areas on the sky without manual work. With these large scales, we increase the statistical significance of the power spectrum at smaller multipole l. Thus, it will be possible to put tighter constraints the parameters of cosmological models that are fitted to the power spectra. Thereby, we will be able to improve our understanding of CIB galaxies and their evolution.

The structure of this Chapter is as follows. In Sect. 5.2, we describe the data that has been used for this particular study. We introduce the standard approach of H I as a template for Galactic cirrus and our improved method in Sect. 5.3. In this section, we also introduce a formalism to compute accurate and unbiased binned angular power spectra which are required to compare the CIB data to simulations. In Sect. 5.4, we compare the two methods on a small sample of previously studied fields at high Galactic latitude. We then apply our improved method to 20% of the full sky in Sect. 5.5 and present the first angular power spectra at different frequencies for the CIB as seen in the *Planck* data. In Sect. 5.6, we suggest several points for future studies to further increase the quality of the foreground removal technique. We conclude in Sect. 5.7.

5.2 Data

In the following, we briefly describe the particular data sets that have been used for this Chapter. A more general description of the individual data sets and the data format was already given in Chapter 2.

5.2.1 H I data

For the H I data, we use data from the Effelsberg-Bonn H I Survey (EBHIS, Winkel et al. 2010; Kerp et al. 2011; Winkel et al. 2016a) and the Galactic All Sky Survey (GASS, McClure-Griffiths et al. 2009; Kalberla et al. 2010; Kalberla and Haud 2015) for the northern and southern hemisphere, respectively. The two surveys have recently been merged to create the HI4PI Survey, a new database of Galactic H I (HI4PI collaboration, 2016, submitted). The angular resolution of the joint dataset is 16.1′ and the

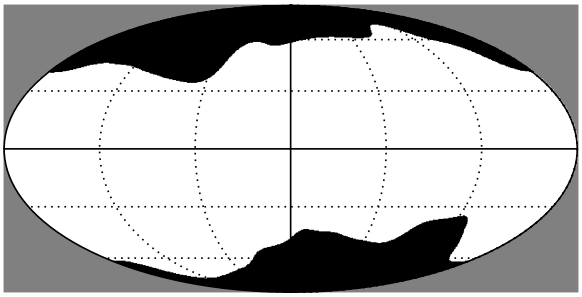


Figure 5.3: Galactic plane mask for 20% of the sky, taken from (Planck Collaboration, 2015 I). Black corresponds to valid values and white to values that are masked. For all plots in Mollweide projection, the Galactic centre is located in the centre of the image and Galactic longitude l increase towards the left.

velocity resolution is $1.49\,\mathrm{km\,s^{-1}}$. Due to the spectral smoothing for the GASS data and the spatial smoothing for the EBHIS data, the sensitivity of approximately 43 mK of the HI4PI data is very similar for both hemispheres.

5.2.2 FIR data

The H I data is complemented by the dust data from the *Planck* Public Release 2 (PR2, Planck Collaboration, 2015 I). In the present work, we use data from the 353, 545, and 857 GHz HFI bands (Planck Collaboration, 2015 VIII). At these frequencies, the contribution from thermal dust over other components such as the CMB, free-free, or synchrotron emission is maximised. This makes it easier to focus on the separation of Galactic thermal dust and the CIB signal.

To minimise the CMB contamination in the final CIB map, we correct the FIR intensity maps by subtracting the SMICA (Cardoso et al., 2008) map of CMB anisotropies (Planck Collaboration, 2015 IX). The Spectral Matching Independent Component Analysis (SMICA) is one of four different map-making algorithms that is applied to extract the CMB from the *Planck* data. While the relative strength of the CMB with respect to other components is maximal in the 143 and 217 GHz bands, its impact onto the CIB map, particularly at 353 GHz, should be accounted for by subtracting the CMB fluctuations (Planck Collaboration, 2014 XXX).

The extragalactic point sources need to be accounted for in the derivation of the CIB images. Unlike the statistical CIB fluctuations which are caused by distant, dusty galaxies, the detected point sources are often nearby galaxies or Sunyaev-Zeldovich effect (SZ, Sunyaev and Zeldovich 1970) sources. We use the *Planck* mask of point sources for each individual frequency (Planck Collaboration, 2015 VIII), based on the Planck catalogue of compact sources (Planck Collaboration, 2015 XXVI). For the foreground removal, these sources are masked and do not affect the fitting procedure.

The complexity and the column density of the foreground emission increases greatly from high Galactic latitudes to the Galactic plane and from the anti-centre at $l = 180°$ to the centre of the Galaxy at $l = 0°$. Thus, we constrain our analysis in Sect. 5.5 to 20% of the full sky at high Galactic latitudes. Instead of using a fixed threshold for the latitude, we use the Galactic plane mask that is provided by Planck Collaboration (2015 I). An illustration of this footprint is shown in Fig. 5.3. It is based on the FIR intensity in the 353 GHz data and is also available for 40, 60, 70, 80, 90, 97, and 99% sky coverage.

The *Planck* FIR data are complemented by the Improved Reprocessing of the IRAS Survey (IRIS, Miville-Deschênes and Lagache 2005) data at 100 μm, corresponding to 3000 GHz. The advantage of this additional data set is the coverage of the dust emission at frequencies higher than the blackbody peak which is located at 1200 GHz for a dust temperature of 20 K.

5.2.3 Data format

Throughout this Chapter, all data sets have been used at an angular resolution of 16.1′, which corresponds to the lowest common resolution of all the data, set by the HI4PI survey. For higher resolution data sets, the resolution was degraded by convolution with a Gaussian kernel.

The data is stored and processed in the Hierarchical Equal Area isoLatitude Pixelation of a sphere (HEALPix) format at at `NSIDE` parameter of 1024, which results in a pixel size of approximately 3.4′. For Sect. 5.4, we re-grid the data onto smaller fields with the convolution-based gridding code cygrid[1] (Winkel, Lenz and Flöer, 2016b).

The fully reduced data of the HI4PI survey, which makes up the largest part of the data used in this study, amounts to 100 GB in size. To optimise the access times and avoid memory issues, we use the Hierarchical Data Format 5 (HDF5, The HDF Group 1997-NNNN)[2]. It is a completely portable data format with no restriction on the file sizes, allows easy reading and writing for multiple threads, and its very flexible application programming interface (API) is available for many programming languages, including Python.

5.3 Methods

To compare the CIB to cosmological models and thus study the ensemble properties and the evolution of the CIB galaxies, two steps are required: The FIR intensity maps must be corrected for all foreground components and the unbiased angular power of the cleaned image must be computed. In the following, we introduce the necessary steps to conduct this analysis.

5.3.1 H I as a template for Galactic cirrus

Similar to many previous studies on the subject (e.g. Planck Collaboration, 2014 XXX, for the latest application of this approach on FIR data), we use the H I data as a template for Galactic dust. In the following, we briefly recall the usage of the H I column density and the use of the full spectral data, which has already been applied on a single field in Chapter 4. A quick illustration of the different resulting CIB images is shown in Fig. 5.4 for the AG field, which was investigated in Planck Collaboration (2014 XXX). It is a region that contains gas at different velocities, covering low, intermediate and high velocities. It is also part of the Green Bank Telescope (GBT) survey of selected fields (Martin et al., 2015). We use this field for illustration purposes here and in an upcoming section, but the choice is purely random and we could have selected any field from the Planck Collaboration (2014 XXX) sample. The central panel of Fig. 5.4 is the reference image from the Planck Collaboration. It can be seen that the generalised linear model (GLM, J. A. Nelder 1972; Madsen and Thyregod 2010; de Souza et al. 2014) approach yields very consistent results with respect to the *Planck* image, while the standard, $N_{\mathrm{H\,I}}$-based approach cannot reproduce the results. A more detailed comparison of the two methods is presented in Sect. 5.4.

The H I column density-based approach

The standard approach to remove the foreground FIR intensity is to model its value based on the H I column density (e.g. Boulanger et al., 1996).

$$I_\nu(\alpha, \delta) = \epsilon_\nu \cdot N_{\mathrm{H\,I}}(\alpha, \delta) + Z_\nu \tag{5.1}$$

[1] https://github.com/bwinkel/cygrid
[2] https://www.hdfgroup.org

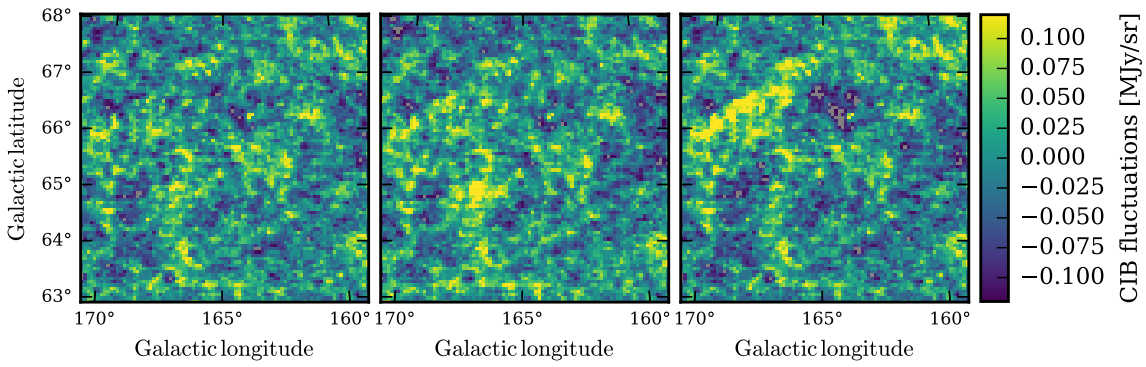


Figure 5.4: Amplitude of the CIB fluctuations at 545 GHz for the AG field. **Left:** Foreground removal via the GLM. **Centre:** Reference image, derived by Planck Collaboration (2014 XXX). **Right:** Foreground removal via the total H I column density. We find that the results of the GLM and the *Planck* results are consistent, while the standard, H I column density-based approach fails to accurately recover the CIB emission.

Here, I_ν is the FIR intensity, ϵ_ν is the dust emissivity per hydrogen nucleon, $N_{\mathrm{H\,I}}$ is the H I column density, and Z_ν is a constant reflecting the mean CIB intensity. The subscript ν indicates that this equation is solved at each frequency individually and the arguments (α, δ) indicate the position on the sky.

Eq. (5.1) can easily be solved with a least-squares approach. In practice, other terms such as the instrumental noise and dust residuals, not associated with H I, need to be accounted for.

If more than one component can be clearly identified in the H I spectrum, Eq. (5.1) is often expanded by one or more terms in the form $\epsilon_{\nu,b} \cdot N_{\mathrm{H\,I},b}(\alpha, \delta)$ (e.g. Peek et al., 2009; Planck Collaboration, 2011 XXIV). The additional component often corresponds to features in the H I data at non-local velocities such as an intermediate-velocity cloud (IVC) or a high-velocity cloud (HVC). In that case, the column density for each term is computed by integrating the H I data over the appropriate radial velocities. This approach is motivated by the fact that the dust emissivity ϵ_ν is different for multiple components because of a different composition or illumination by the interstellar radiation field (ISRF). By accounting for this difference, the FIR intensity can be modelled more accurately. In practice, the separation of the H I data cube into multiple column density maps is done manually and can be very subjective.

Using the full spectral H I data

We expand the idea of assuming different emissivities for different radial velocities by allowing each spectral channel to have an individual dust emissivity (see also Lenz, Flöer and Kerp, 2016, Chapter 4). This technique is referred to as generalised linear model (GLM, J. A. Nelder 1972; Madsen and Thyregod 2010; de Souza et al. 2014). Building up on Eq. (5.1), the GLM for the FIR intensity $I_\nu(\alpha, \delta)$ can be written as weighted sum for all spectral channels:

$$I_\nu(\alpha, \delta) = \sum_{\mathrm{chan}} \beta_\nu^{\mathrm{chan}} \cdot T_B^{\mathrm{chan}}(\alpha, \delta) + Z_\nu. \tag{5.2}$$

Here, $T_B^{\mathrm{chan}}(\alpha, \delta)$ denotes the H I brightness temperature in each spectral channel and $\beta_\nu^{\mathrm{chan}}$ is the emissivity for each individual channel. Similar to Eq. (5.1), Z_ν is a constant offset that accounts for the mean CIB intensity.

Many of the H I spectral channels contain no significant emission and are masked when solving Eq. (5.2). We use the $3\sigma_{\mathrm{RMS}}$ of the H I data as clip level. This avoids that CIB fluctuations are falsely fitted by using hundreds of noise channels, thereby creating an artificial, false image of the background component.

To further constrain our model, we use lasso regularisation for the individual GLM coefficients $\beta_\nu^{\mathrm{chan}}$ (Tibshirani, 1996). Thereby, we break the degeneracy between neighbouring, correlated channels and avoid overfitting. In practice, this means that we do not simply minimise the residual sum of squares (RSS) of the data, but penalise the term by adding the total sum of the coefficients:

$$\begin{aligned} \mathrm{RSS} \quad &+ \quad \alpha \cdot \|\beta_\nu^{\mathrm{chan}}\|_1 \\ \|I_\nu - \sum_{\mathrm{chan}} T_B^{\mathrm{chan}} \beta^{\mathrm{chan}} - Z_\nu\|^2 \quad &+ \quad \alpha \cdot \|\beta_\nu^{\mathrm{chan}}\|_1 \end{aligned} \tag{5.3}$$

Here, $\|\cdot\|_1$ denotes the L^1-norm and α is the strength of the regularisation. The second term yields a sparse set of coefficients.

To scale the regularisation strength α, we use cross validation (e.g. Picard and Cook, 1984). This technique is commonly used to optimise hyperparameters such as α, which are not directly evaluated on the data, but need to be determined before the actual fitting procedure. For this purpose, the image is split into n parts (so-called folds). For the n-fold cross validation, the data is fitted on $n-1$ folds (training sample) and the quality of the fit is evaluated on the n th fold (test sample). This procedure is repeated n times, so each fold serves as test sample exactly once. For our purpose, we work with values of $n = 3..5$ and find that the exact choice of n does not affect the results significantly.

The regularisation strength α is then varied across a large range of values and the best value for this hyperparameter is determined via cross validation. The selection is based on the residual mean and standard deviation of the test sample (see also Fig. 4.10). Therefore, the GLM is not evaluated just once for an image, but instead hundreds of times to find an appropriate solution.

An illustration of the GLM approach is given in Fig. 5.5 for the AG field, which we have already introduced in Fig. 5.4. We show the mean H I spectrum for this field and the GLM coefficients, both as a function of radial velocity v_{LSR}. Instead of using a fixed dust emissivity for the column density of atomic hydrogen, our approach allows to assign different emissivities to different gas components. Thereby, it is possible to treat H I features at different radial velocities individually. This allows, for example, a local cloud to have a higher dust emissivity than an IVC, which very well reflects the actual physical properties of these objects (e.g. Planck Collaboration, 2011 XXIV). Instead of doing this manually by creating different maps of H I column density for different radial velocity intervals, the GLM approach allows to perform this task automatically and in a very flexible manner.

We note that the discrete nature of the FIR emissivities that is assumed in this approach is primarily a schematic description and does not necessarily yield an accurate value of the physical dust emissivity for each spectral channel of the H I data. Our main goal here is to create a robust foreground model, free of residuals from Galactic dust.

Lastly, we note that the standard H I column density-based approach can be used to construct a proper model of foreground FIR intensity, especially if the H I datacube is accurately divided into multiple column density maps. However, this approach requires very subjective, manual work and is thus hardly reproducible and cannot be automatically applied to large areas. Here, the GLM is a viable alternative and of key importance to remove the foreground for large areas.

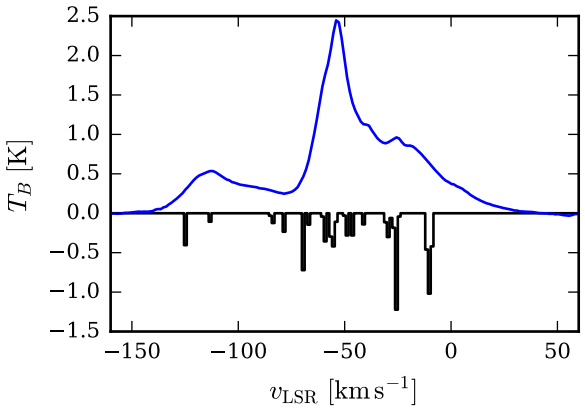


Figure 5.5: GLM coefficients (black) and mean H I spectrum (blue) for the AG field as investigated in (Planck Collaboration, 2014 XXX). To match the scale, the coefficients were normalised to the peak of the H I spectrum and furthermore multiplied by -1 for visualisation purposes. This approach allows to account for varying dust emissivities for components at different radial velocities and thus different physical properties.

5.3.2 Reconstructing the angular power spectrum

Whether it is the emission from the CMB or the CIB, much of the statistical information from the map is contained in the angular power spectrum C_l, which is a measure for the amplitude of the fluctuations on a given angular scale. The angular scale is commonly measured in terms of the multipole l. By comparing these C_l to simulations, one can infer information about the distribution and properties of FIR galaxies (CIB) or on the cosmological parameters (CMB). In practice, the process of computing the angular power spectrum involves performing a Fourier transform of the map and then to bin the modulus square into multipole bins. For most measurements, the maps have boundaries or holes due to point sources, which biases the resulting angular power spectrum if not accounted for. Furthermore, the power spectrum needs to be corrected for the convolution with the beam, the pixel grid, and it needs to be binned to increase the signal-to-noise ratio (SNR).

Until the beginning of this century, the proper derivation of the angular power spectrum scaled as $O(n^3)$ or worse (see the review by Borrill, 1999). At that time, it was recognised that this would lead to computational constraints when analysing the soon-available WMAP data (Bennett et al., 2003), among other data sets.

In the following, we introduce the basic principles of the Monte Carlo Apodized Spherical Transform Estimator (MASTER, Hivon et al. 2002) code and our implementation of this algorithm in Python, which is publicly available online[3]. MASTER was developed to deal with the aforementioned limitations. It is based on the decomposition of the signal into spherical harmonics (MacRobert and Sneddon, 1967, their Chapter 4). For a full, very detailed description of the method, we refer to Hivon et al., 2002. An application of this algorithm for flat fields is given in the description of the P of k EstimatoR code (POKER, Ponthieu, Grain and Lagache 2011) and a demonstration of this algorithm applied to CIB data can be found for example in Serra et al. (2014) and Planck Collaboration (2014 XXX).

Mathematical introduction

We introduce the concept that is required to derive an unbiased estimate of the angular power spectrum and provide some details about the involved quantities.

A scalar map of temperature fluctuations $\Delta T(\vec{n})$, such as the CIB or CMB anisotropies, can be decomposed into the coefficients of the spherical harmonics:

$$a_{lm} = \int d\vec{n} \Delta T(\vec{n}) Y^*_{lm}(\vec{n}), \tag{5.4}$$

[3] https://github.com/DanielLenz/powspecpy

Here, * indicates complex conjugation and $\vec{n}$ denotes the position vector, which can be substituted by spherical coordinates (θ, ϕ). $\Delta T(\vec{n})$ is then given by

$$\Delta T(\vec{n}) = \sum_{l>0}^{l_{\max}} \sum_{m=-l}^{l} a_{lm} Y_{lm}(\vec{n}). \tag{5.5}$$

If the temperature fluctuations are Gaussian distributed, we can assume that each a_{lm} is an independent Gaussian random variable and

$$\langle a_{lm} \rangle = 0 \tag{5.6}$$

$$\langle a_{l_1 m_1} a^*_{l_2 m_2} \rangle = \delta_{l_1 l_2} \delta_{m_1 m_2} \langle C_{l_1} \rangle. \tag{5.7}$$

Here, δ denotes the Kronecker delta function, $\langle \cdot \rangle$ is the ensemble average, and C_l is the unbiased angular power spectrum.

C_l and the a_{lm} are related via

$$C_l = \frac{1}{2l+1} \sum_{m=-l}^{l} |a_{lm}|^2 . \tag{5.8}$$

In practice, the fraction of the sky that can be investigated is limited because of missing or contaminated data. In our case, we need to mask point sources, the Galactic plane, and residual emission that cannot be modelled properly. Additionally, the mask is apodized by convolution with a Gaussian kernel of 16.1′, corresponding to the beam size. This removes the sharp edges in the image domain which would lead to strong oscillations in Fourier space. Hence, a position depended weighting function $W(\vec{n})$ is introduced

$$w_{lm} = \int d\vec{n} W(\vec{n}) Y^*_{lm}(\vec{n}) \tag{5.9}$$

$$\mathcal{W}_l = \frac{1}{2l+1} \sum_{m} |w_{lm}|^2 , \tag{5.10}$$

where $\mathcal{W}_l$ is the power spectrum of the weighting function.

We can only directly infer the pseudo power spectrum $\langle \tilde{C}_l \rangle$, which is measured by the spherical harmonic transformation and biased by the limited sky coverage that is described by $W(\vec{n})$. To estimate the unbiased angular power spectrum $\langle C_l \rangle$, we need to compute the mode-mode coupling matrix $M_{ll'}$. This matrix is based on the position-dependent weighting power spectrum $\mathcal{W}_l$ and accounts for the partial sky coverage. Its computation is the most challenging aspect of this analysis, so we provide some details on the practical aspects on the following section.

Furthermore, we need to account for the beam shape and the pixel size. Hence, we introduce the effective beam function

$$B_l = P_l D_l, \tag{5.11}$$

with the beam window function D_l and the pixel window function P_l. Both functions are obtained by computing the power spectrum of the beam shape and the pixel shape. They can easily be obtained within the HEALPix framework. Moreover, the *Planck* beam window function is publicly available.

Finally, $\langle C_l \rangle$ can be computed by simply solving the matrix equation (5.12) below:

$$\langle \tilde{C}_{l_1} \rangle = \sum_{l_2} M_{l_1 l_2} B^2_{l_2} \langle C_{l_2} \rangle + \langle \tilde{N}_{l_1} \rangle. \tag{5.12}$$

$\langle \tilde{N}_l \rangle$ is the average power spectrum of the noise. For the *Planck* mission, we can use the difference of the individual ring-halves to characterise this term (see Sect. 5.5).

Computing the mode-mode coupling matrix

For a detailed derivation of the mode-mode coupling matrix $M_{ll'}$, we refer to Wandelt and Hivon (1999) and Hivon et al. (2002). In particular, we recommend the appendix A.2 of Hivon et al. We briefly outline our approach to calculate this expression in the following with the focus on reproducing these steps in practice.

$$M_{l_1 l_2} = \frac{2l_2 + 1}{4\pi} \sum_{l_3} (2l_3 + 1) \mathcal{W}_{l_3} \begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix}^2 \tag{5.13}$$

Here, $\mathcal{W}_{l_3}$ is the aforementioned power spectrum of the weighting map and $\begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix}$ is the Wigner 3-j symbol. This object denotes a scalar and represents the coupling of three angular momentum vectors. It is only non-zero if the triangle relation

$$|l_1 - l_2| \leq l_3 \leq |l_1 + l_2| \tag{5.14}$$

is satisfied. Furthermore, it is zero if $L = l_1 + l_2 + l_3$ is odd. We use the following expression, which is valid for $m_1 = m_2 = m_3 = 0$, to compute the Wigner 3-j coefficients:

$$\begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix} = (-1)^{L/2} \left[\frac{(L - 2l_1)!(L - 2l_2)!(L - 2l_3)!}{(L + 1)!} \right]^{1/2} \frac{(L/2)!}{(L/2 - l_1)!(L/2 - l_2)!(L/2 - l_3)!} \tag{5.15}$$

To ensure numerical stability and improve the performance, we compute the factorials in log-space.

Illustration and validation of the angular power spectrum estimation

We provide some examples and details to the individual steps to complement the mathematical framework introduced above. The challenge is the following: We have a map of fluctuations for which we want to estimate the unbiased, deconvolved angular power spectrum. This means that we need to account for the partial sky coverage and that we need to deconvolve the image, thereby accounting for the beam smearing.

For the example here, we use data in the HEALPix format at an `NSIDE` of 256 (corresponding to a resolution of 14′) and we reconstruct the power spectrum up to an `lmax` of 512. For the artificial beam, we use a FWHM of 35′. To improve the SNR, we bin the unbiased power spectrum C_b by a factor of 32, which means that the bin size Δl is 16. Another binning scheme such as even spacing in log space could be realised easily as well.

We start by creating an artificial angular power spectrum that we wish to recover. Thus, we generate a spectrum of the form

$$C_l = \frac{a}{(l + 1)^b} \tag{5.16}$$

with random numbers a and b. Then, we generate a random realisation of the map, corresponding to that power spectrum. For this step, tools such as `healpy.synfast` can be used. Lastly, we convolve the map with an artificial beam function to simulate the response function of a real instrument. In Fig. 5.6, the input power spectrum and the map is shown.

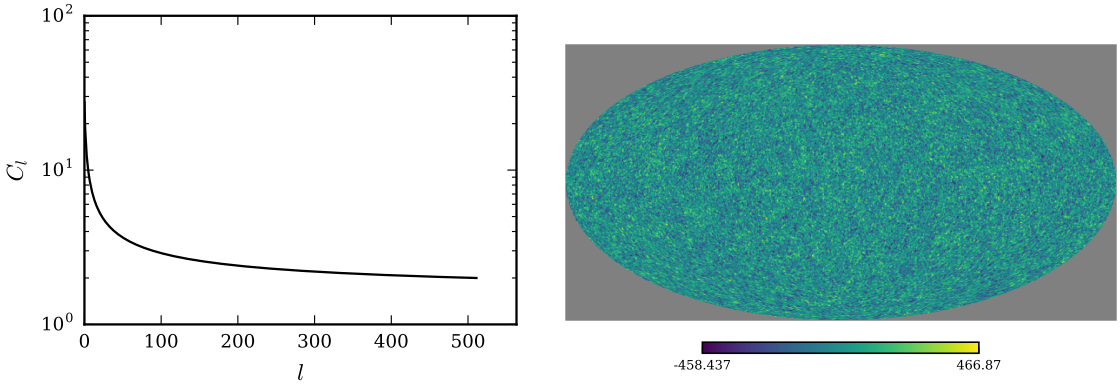


Figure 5.6: Simulated angular power spectrum that we wish to recover (left). Based on this power spectrum, we generate a realisation of the image with arbitrary units (right). The map is convolved with a Gaussian kernel to simulate the response of a real instrument.

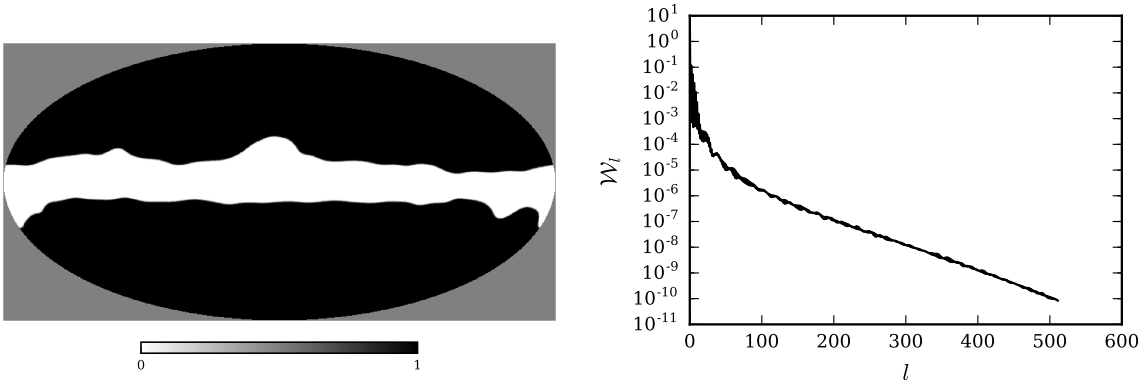


Figure 5.7: Map of the mask and the corresponding power spectrum. Here, black corresponds to valid values and white to values that are masked. We use the 80% Galactic plane mask from (Planck Collaboration, 2015 VIII) to simulate a real measurement with a limited sky coverage. Furthermore, we apodize the mask via convolution with the beam. The power spectrum of the mask is required to reconstruct the unbiased power spectrum, based on the measured map (Fig. 5.6, right).

To simulate the partial sky coverage that is very common for a real measurement, we mask the Galactic plane. Thus, we use the *Planck* Galactic plane mask (Planck Collaboration, 2015 VIII) with a sky coverage of 80% (i.e. the innermost 20% are masked). To avoid sharp edges in the map which will lead to oscillations in Fourier space, we apodize the mask by convolving it with the beam function. Finally, we compute the pseudo power spectrum of the mask $\mathcal{W}_l$ using `healpy.anafast`. The mask and its power spectrum are shown in Fig. 5.7.

We compute the mode-mode coupling matrix $M_{l_1 l_2}$ according to Eq. (5.13). This step is the most time-consuming one and the most challenging, but $M_{l_1 l_2}$ depends only on the pseudo power spectrum of the mask $\mathcal{W}_l$ and the maximum l to which the power spectrum is reconstructed. Thus, it needs to be computed only once for a given problem, even if the input map should change. This is particularly

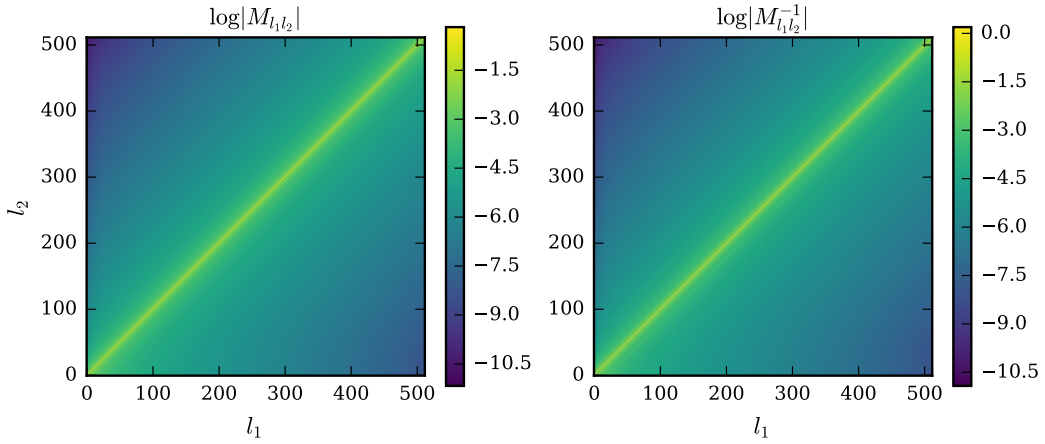


Figure 5.8: Mode-mode coupling matrix $M_{l_1 l_2}$ and its inverse. We show the logarithm of the absolute value of the matrices. This matrix accounts for the partial sky coverage which biases the derived power spectrum. An accurate and numerically stable derivation of $M_{l_1 l_2}$ is the key to overcoming this bias and thereby computing the unbiased power spectrum.

helpful when the uncertainties are computed via Markov chain Monte Carlo (MCMC) methods. We show an illustration of the matrix and its inverse in Fig. 5.8, yielding the rather constant diagonal and the suppressed off-diagonal elements. An inspection of this matrix is particularly helpful if the estimation of the unbiased power spectrum does not return the correct results. As expected, the matrix is dominated by the diagonal elements, indicating that the value of the unbiased power spectrum $\langle C_{l_1} \rangle$ for the multipole l_1 depends foremost on a similar l range of the naive estimate of the power spectrum $\langle \tilde{C}_{l_2} \rangle$.

Finally, we can reconstruct the unbiased, binned power spectrum by solving Eq. (5.12). For this example, we ignore the noise term $\langle \tilde{N} \rangle$. Fig. 5.9 (left) demonstrates the result of the algorithm, comparing the input power spectrum, the naive, biased estimate of the power spectrum, and our reconstruction. We find that our unbiased power spectrum is an excellent reconstruction of the input power spectrum.

To estimate the uncertainties, there are two viable approaches. First, they can be computed analytically (Serra et al., 2014):

$$\sigma_{C_b} = \left(\frac{2C_b^2}{(2l+1)\Delta l} \right)^{1/2} \tag{5.17}$$

Here, Δl is the centre value of each l-bin. Second, the uncertainties can be estimated using simulations of the map for which the power spectrum is reconstructed. In our case, we can easily simulate different realisations of the image and use the different samples to estimate the standard deviation of the resulting power spectra. The covariance matrix of the binned power spectrum is

$$C_{bb'} = \left\langle (C_b - \langle C_b \rangle_{\mathrm{MC}}) (C_{b'} - \langle C_{b'} \rangle_{\mathrm{MC}}) \right\rangle_{\mathrm{MC}} \tag{5.18}$$

with $\langle \cdot \rangle_{\mathrm{MC}}$ being Monte-Carlo averaging. The numerically computed uncertainties can be derived from the diagonal of the covariance matrix:

$$\sigma^{\mathrm{MC}}_{C_b} = (C_{bb})^{1/2} \tag{5.19}$$

For real data, the estimation of the uncertainties via Monte Carlo samples requires an accurate simulation of the maps, including instrument noise terms and the shot noise (see Sect. 5.5 for a discussion of these terms).

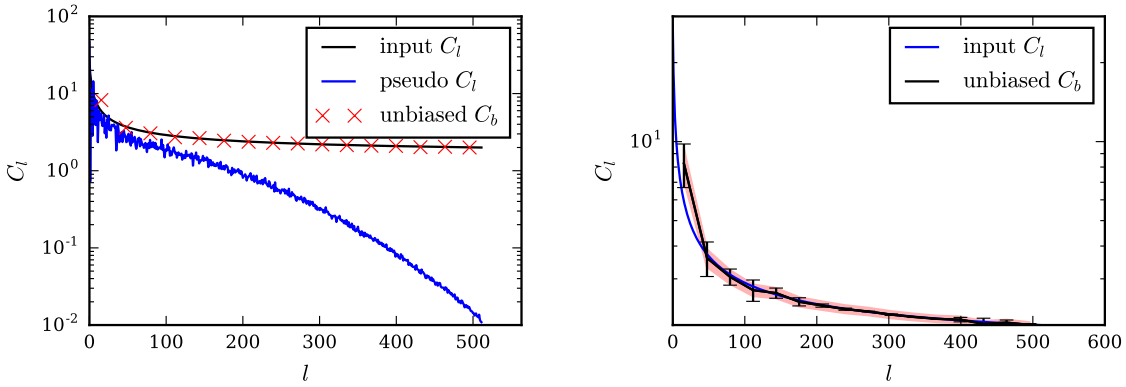


Figure 5.9: **Left:** Input C_l (black) and the binned, deconvolved, and unbiased reconstruction C_b (red crosses). Furthermore, we show the naive estimate $\langle \tilde{C}_l \rangle$ of the power spectrum (blue). The two factors that had to be corrected are the convolution by the effective beam and the limited sky coverage of the input map. **Right:** Input power spectrum (blue) and the unbiased, binned power spectrum (black) including the uncertainties. Because of the small relative size of the errors, we show the 3σ error bars from MC simulations by black error bars and the analytical 3σ uncertainties as red contours.

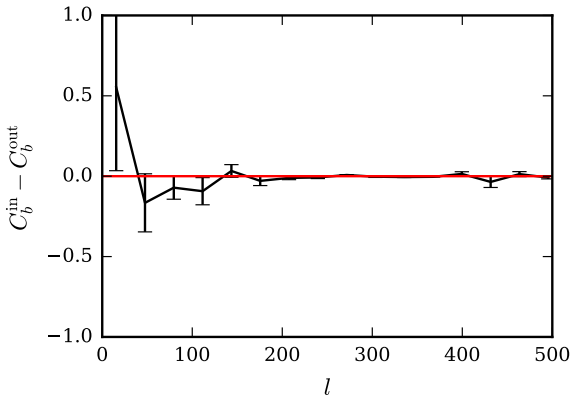


Figure 5.10: Residual power spectrum $C_b^{\rm in} - C_b^{\rm out}$ between the input power spectrum and our reconstruction. The error bars are 1σ uncertainties, derived from the simulations. The red line indicates the zero line for which the reconstruction would be perfect.

We show the 3σ uncertainties in Fig. 5.9 (right) with the error bars indicating the uncertainties from 500 Monte Carlo samples (Eq. 5.19) and the shaded areas showing the analytical estimate (Eq. 5.17). We selected the 3σ level for visualisation purposes because the 1σ uncertainties are too small. The uncertainties decrease with increasing l because of the increasing number of samples that can probe these scales.

Lastly, we show the difference between our binned input $C_b^{\rm in}$ and the unbiased, binned power spectrum $C_b^{\rm out}$ in Fig. 5.10. This demonstrates that we can accurately reconstruct the power spectrum and are only limited by sample variance at the lowest l (i.e. at the largest scales).

5.4 Comparison to small-scale *Planck* results

5.4.1 Introduction and data

We have validated the GLM approach to remove foreground dust emission in great detail in Chapter 4. However, this analysis only included a single field and the focus of that work was to recover the dust

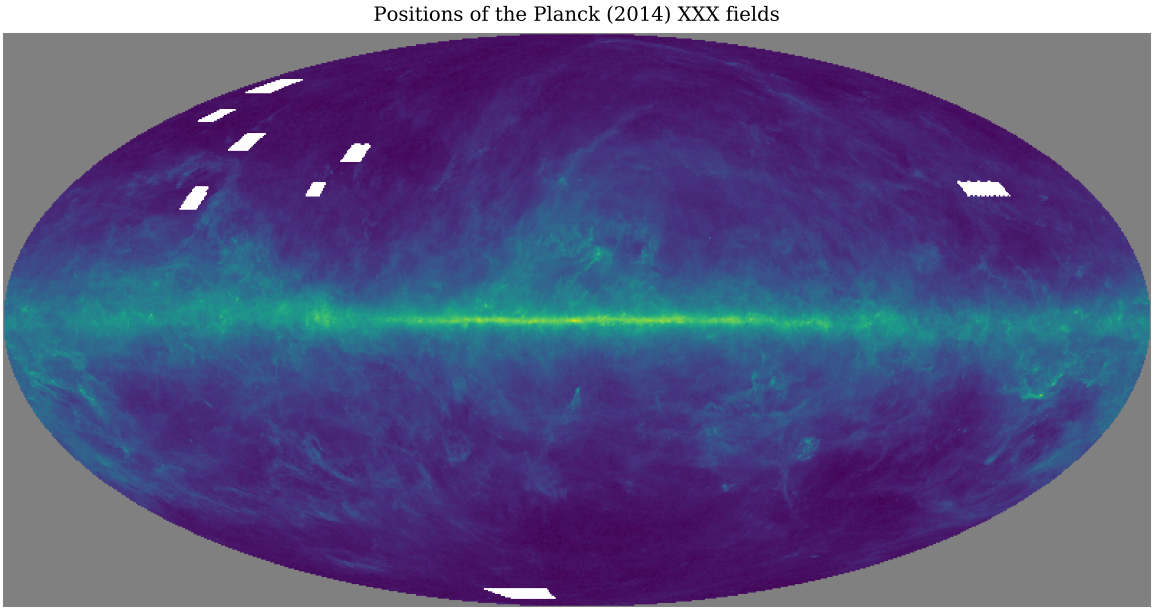


Figure 5.11: Positions and sizes of the individual fields investigated in Planck Collaboration (2014 XXX), superposed onto the *Planck* FIR intensity at 545 GHz on a log-scale.

emission from an HVC. Here, we validate our method on the Planck Collaboration (2014 XXX) sample of high-latitude fields and compare the derived CIB images. Furthermore, we show the higher quality of the CIB reconstruction with respect to the simple H I column density-based approach.

Fig. 5.11 shows the position of the fields investigated in Planck Collaboration (2014 XXX). For these fields, H I data from the GBT have been used as a template for the Galactic cirrus. For the present analysis, we use the EBHIS data at a resolution of 10.8′ instead of the lower resolution HI4PI data because it is closer to the 10′ resolution of the GBT data.

On top of the smaller GBT fields, the analysis of the Planck Collaboration involved larger fields from EBHIS and GASS, but they were treated differently from the GBT fields and will thus be disregarded in our comparison. For an overview of the GBT observational setup and the comparison to other surveys, including EBHIS, we refer to Martin et al. (2015).

For the FIR data, we use the exact same data as in Planck Collaboration (2014 XXX) that were provided by the corresponding author G. Lagache (priv. comm.). As we have only access to the 545 GHz data, we constrain our comparison to this single frequency. Our main goal here is to illustrate the difference between the standard, $N_{\mathrm{H\,I}}$-based approach and the GLM, for which a single frequency is sufficient. A larger coverage of the spectral bands is used for further steps of this analysis (Sect. 5.4.2).

5.4.2 Results

Individual fields

We start by removing the FIR foreground, based on the simple approach that uses the H I column density as introduced in Sect. 5.3.1. The results of this approach are shown in Fig. 5.12. For clarity, we only show the fields N1, NEP4, and AG from Planck Collaboration (2014 XXX) here and the remaining ones in appendix A. A description of the individual panels is given in the figure caption. We use the Pearson's r correlation coefficient (Pearson, 1895) to quantify the correlation of two images. This value ranges from -1 to +1, where +1 is total positive correlation, 0 is no correlation, and -1 is total negative correlation.

We find that the derived map of CIB anisotropies (second row) is more structured and has a larger amplitude when derived with the simple $N_{\mathrm{H\,I}}$-based method than the one seen in the *Planck* results. Using the Pearson's r correlation coefficient, we find that neither the *Planck* CIB fields nor the ones derived with the $N_{\mathrm{H\,I}}$-based approach are correlated with the H I column density.

The scatter between our CIB results and the *Planck* results (third row) is fairly large with a mean value of approximately 0.1 MJy sr^{-1}. The overall ratio of the two CIB images is significantly lower than unity while the offset is zero.

The difference between the two CIB images (fourth row) reveals a lot of structure, which indicates that the results from this simple method and the *Planck* approach do not yield consistent results. For the Pearson's r between the two CIB images, we find values between 0.72 and 0.87 for the full sample.

In rows 5 to 7, we check whether the difference in the CIB maps stems from regions of high $N_{\mathrm{H\,I}}$ (row 5), the noise in the H I data (row 6), or an incorrect subtraction of the CMB fluctuations (row 7). Analysing the entire sample, we find that only the H I column density yields a significant correlation with the CIB difference map. Hence, we can conclude that the differences in the derived CIB maps between this work and the Planck Collaboration (2014 XXX) results do not stem from CMB or noise in the H I data, but instead on differences in the Galactic foreground modelling. Due to the excellent agreement between the GBT and the EBHIS data (Martin et al., 2015), we assume that the quality of H I data is not responsible for this discrepancy.

We repeat the analysis for the Planck Collaboration (2014 XXX) fields with the GLM approach, generating the same set of images to compare the outcome. We find that the derived CIB image contains much less structure and is very similar to the *Planck* results. Moreover, the scatterplot indicates that the scatter is approximately 50% smaller with respect to the previous approach. Nonetheless, the slope is significantly below unity, similar to the findings of Planck Collaboration (2016 XLVIII), who separate the Galactic FIR emission and the CIB based on the spatial statistical information.

The difference of the CIB images is much smaller and less structured than in the previous example. Lastly, the correlation of the CIB difference map and the H I column density is distributed around zero.

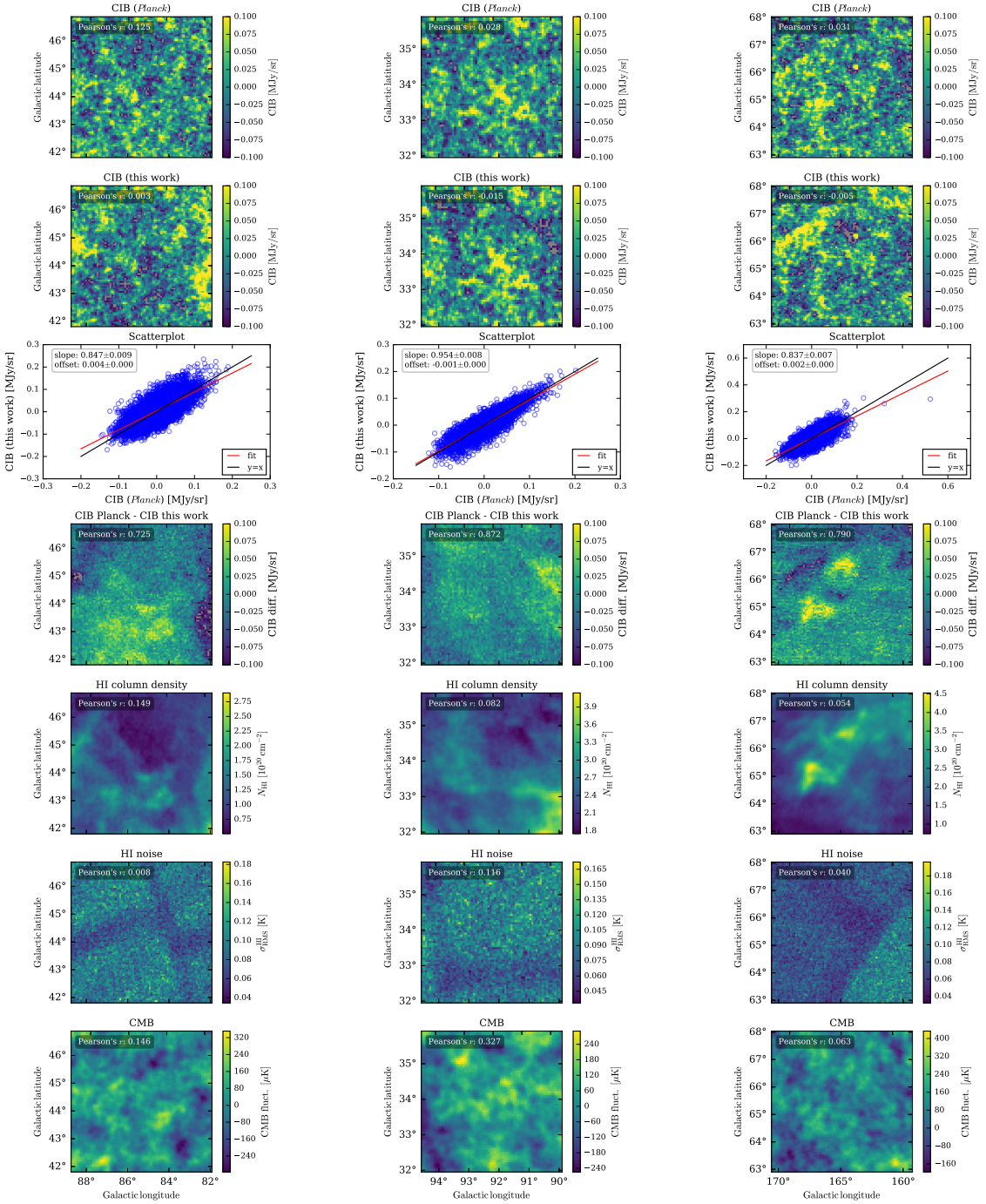


Figure 5.12: Comparison of CIB reconstruction at 545 GHz between Planck Collaboration (2014 XXX) and this work, using a simple linear relation between FIR intensity and H I column density. Each column corresponds to a different field: N1 (**left column**), NEP4 (**central column**), and AG (**right column**). Row description from top to bottom: **(1)** CIB image derived by Planck Collaboration (2014 XXX), Pearson's r w.r.t to $N_{\mathrm{H\,I}}$. **(2)** CIB image derived in this work, using a simple linear relation between FIR intensity and H I column density. Pearson's r w.r.t to $N_{\mathrm{H\,I}}$. **(3)** Scatterplot between individual CIB values found by Planck Collaboration and this work. **(4)** Difference map between the the *Planck* results and this work. The Pearson's r indicates the similarity of the two CIB images (rows 1 and 2). **(5)** Total EBHIS H I column density. **(6)** EBHIS noise map. **(7)** CMB map, taken from Planck Collaboration (2015 IX). For the rows (5) to (7), the Pearons's r indicates the correlation between the individual map and the CIB difference map (row 4). Thereby, we check if the difference between the derived image of CIB anisotropies can be explained by one of the quantities in the bottom rows.

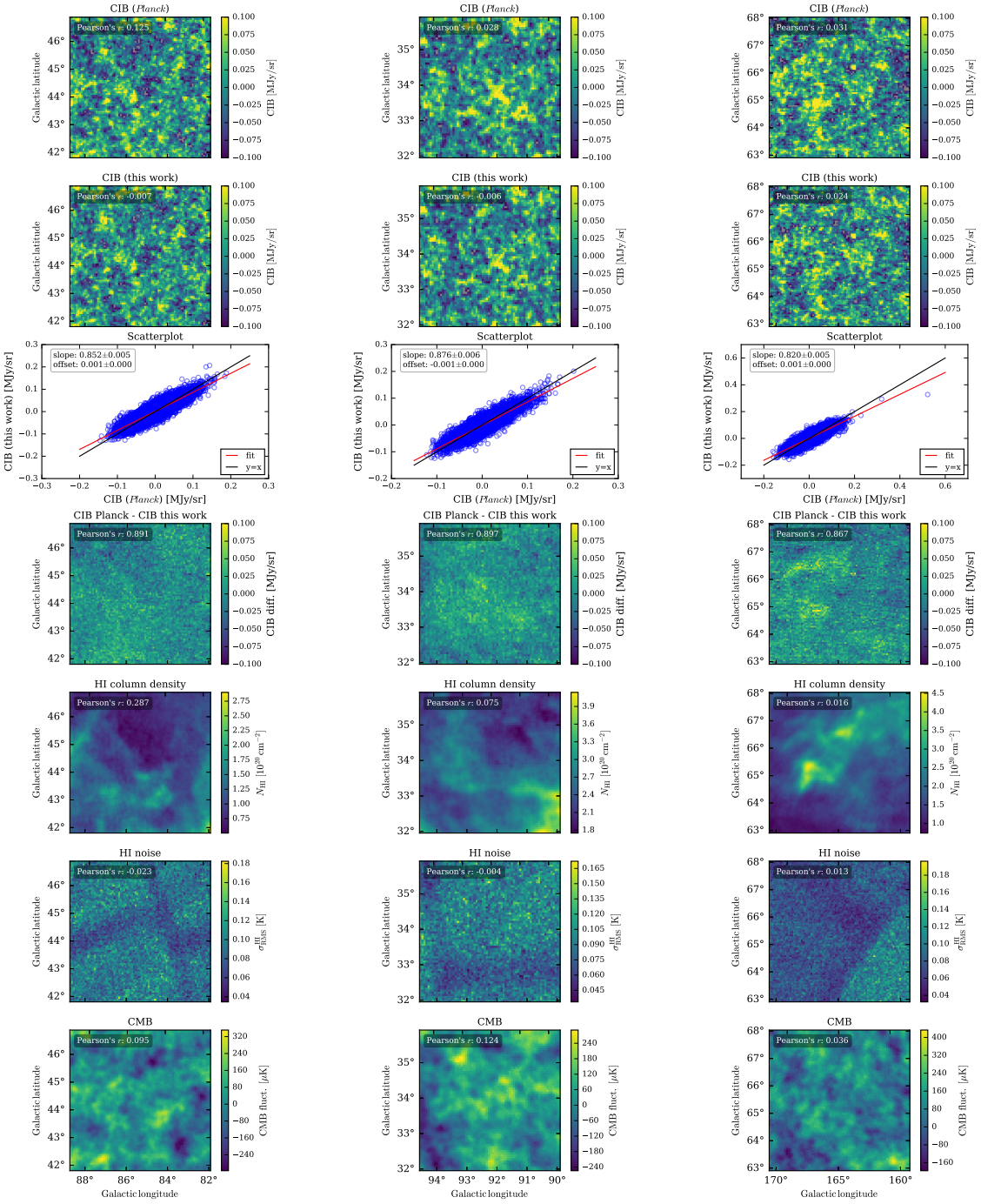


Figure 5.13: Identical to Fig. 5.12, but now using the GLM to remove the foreground.

ν [GHz]	Standard approach	GLM	Planck Collaboration (2014 XXX)
353	0.55 ± 0.20	0.53 ± 0.09	0.53 ± 0.02
545	2.52 ± 0.91	2.18 ± 0.25	2.30 ± 0.10
857	8.08 ± 4.19	6.33 ± 0.97	7.70 ± 0.20
3000	39.2 ± 27.6	29.0 ± 10.4	13.1 ± 1.0

Table 5.1: CIB mean level for the fields investigated in Planck Collaboration (2014 XXX), compared to the results derived with the standard approach and the GLM. Units are nW m^{-2} sr^{-1}. The uncertainties for our results are based on the sample variance, whereas the Planck Collaboration (2014 XXX) uncertainties are based on a complex halo model.

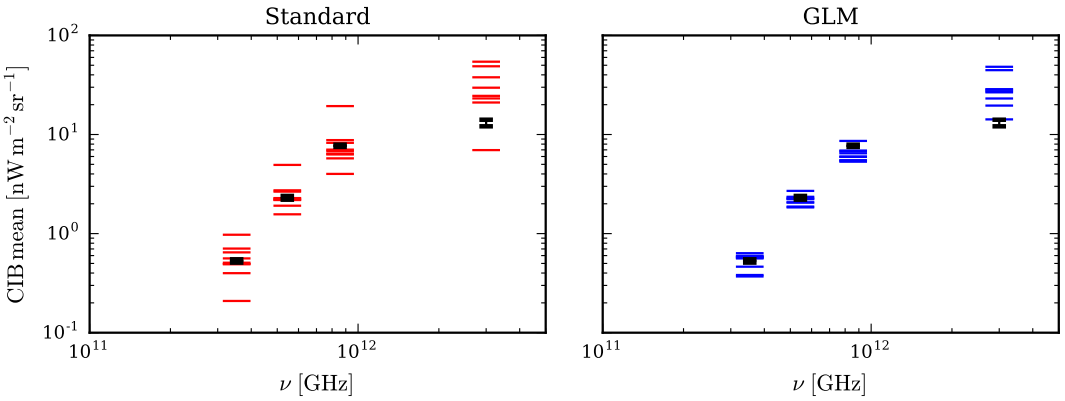


Figure 5.14: CIB mean levels for the fields investigated in Planck Collaboration (2014 XXX), derived with the standard $N_{\mathrm{H\,I}}$-based approach (left) and the GLM (right). Each horizontal line corresponds to a single field. The black error bar indicates the *Planck* result, for which the uncertainties are derived after evaluating a halo model (see the text for a description).

CIB mean levels

Typically, the CIB mean level is expressed in νI_ν [nW m^{-2} sr^{-1}]. To convert from MJy sr^{-1} to this unit, we use

$$\nu I_\nu \text{ [nW m}^{-2}\text{ sr}^{-1}\text{]} = \nu \text{ [GHz]} \times I_\nu \text{ [MJy sr}^{-1}\text{]} \times 10^{-11}. \tag{5.20}$$

In addition to the shape and structure of the CIB fluctuations, we compare the mean level of the derived CIB for each field. This is done for both the $N_{\mathrm{H\,I}}$-based approach and for the GLM (Fig. 5.14 and Table 5.1). To access all frequencies from 353 GHz to 3000 GHz, we used the public *Planck* and IRIS data and re-gridded them onto the corresponding fields. For the uncertainties, we simply use the ensemble mean and standard deviation of the individual fields. In contrast to that, the values from Planck Collaboration (2014 XXX) are derived after evaluating a halo model that is fit to the power spectra at different frequencies (see Cooray and Sheth, 2002, for a review on halo models). Because of the different approach to estimate the uncertainties in Table 5.1, the error bars can only be compared with caution.

Nonetheless, the comparison shows that the mean values scatter strongly for the standard approach and are more consistent for the GLM. Compared to the results from (Planck Collaboration, 2014 XXX), our results agree well except for the IRIS data at 3000 GHz. Here, a different version of the IRIS data might be responsible for tension, but a more detailed analysis will help to resolve the tension between the different results.

Finally, we characterise the mean CIB signal, derived with the GLM approach, from all the fields with

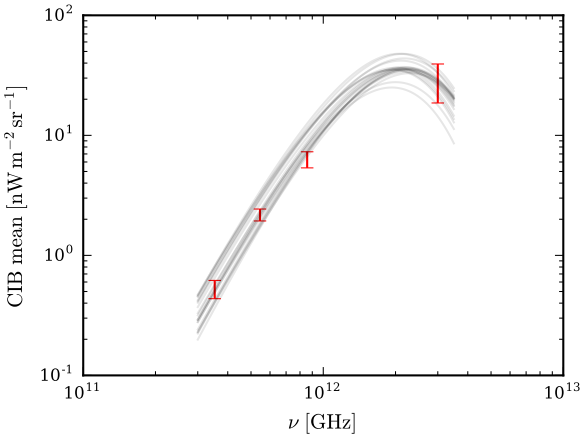


Figure 5.15: Modified blackbody fit to the CIB means, derived with the GLM. Grey lines correspond to MCMC samples, the red error bars indicate the mean and standard deviation of our results on the CIB across the different fields.

a modified blackbody (Draine and Li, 2007).

$$B_\nu^{\mathrm{mod}}(T,\beta) = A\left(\frac{\nu}{\nu_0}\right)^\beta \frac{2hc^3}{c^2}\frac{1}{\exp\left(h\nu/k_BT\right)-1} \tag{5.21}$$

A denotes the amplitude and the reference frequency ν_0 is set to 545 GHz. Fig. 5.15 shows the result of this fit. The estimated values are $T = 20.7 \pm 1.4\,\mathrm{K}$, $\beta = 1.9 \pm 0.1$, and $A = (2.6 \pm 0.8) \times 10^{15}$. These values are the posterior means and the corresponding 1σ uncertainties, equivalent to the 16th and 84th percentile. For a more detailed analysis and a comparison to literature values, it will be required to fit the power spectra with cosmological models.

5.5 Large-scale foreground modelling

5.5.1 Introduction

By expanding the GLM foreground removal technique introduced in the previous sections, we can access large fractions of the CIB sky for the first time. Currently, the largest CIB image at the *Planck* frequencies covers approximately 10% of the full sky and was derived by using the GASS H I column density as a template for the foreground. Therefore, current estimate of the power spectrum at the largest scales have large uncertainties or are only upper limits (Planck Collaboration, 2014 XXX). For the first time, we will remove the foreground emission on these scales in an automatic, accurate manner and thereby open the window to studying the reionisation history and the SFH in dark matter halos with unprecedented sensitivity.

Due to the unresolved tension for the 100 GHz IRIS data with respect to the Planck Collaboration (2014 XXX) results, we constrain our analysis to the 353, 545, and 857 GHz FIR data. Furthermore, the point source masks that are available for the *Planck* data are not easily derived for the IRIS data, which might introduce another systematic uncertainty.

5.5.2 Method

For previous sections, we have assumed that the relation of dust and gas does not vary across the field, which is a fair assumption for smaller fields (e.g. Peek et al., 2009; Planck Collaboration, 2014 XXX). For larger areas, this assumption will inevitably be violated, so a new approach is required. The Planck Collaboration (2014 XVIII) has used large patches of tens of square degrees that they fit independently to determine the emission properties of dust from the FIR to microwave frequencies. Unfortunately, such an approach would lead to edges in our CIB map that make a further use impossible (see Fig. 2.2 for an example of such a pattern).

Instead, we fit the correlation of dust and gas by evaluating the GLM for all data points in an aperture around a single pixel. Based on this fit, we compute the modelled FIR foreground intensity and the model parameters (Eq. 5.2) for this particular pixel. This procedure is repeated for each position that is covered by the 20% Galactic plane mask (Planck Collaboration, 2015 VIII). For the size of the aperture, we find that a radius of 1.5° offers the best balance between sufficient number of data points for the fit and only a minor variation of the dust emissivities across the spatial area.

5.5.3 Results

We show the results on the CIB maps in the following section for the 545 GHz data. For the 353 GHz and 857 GHz, the images and tables can be found in appendix B.

The CIB map at 545 GHz for 20% of the sky is shown in Fig. 5.16, using a zenith equal-area projection. A zoom in on a high Galactic latitude field with a size of $10° \times 10°$ is shown in Fig. 5.17, highlighting the typical structure of CIB anisotropies and lack of artefacts. Each of the data points in this image is analysed by only evaluating its close neighbours, and the size of the aperture is much smaller than the extent of the image. We note that no imprint of this aperture is seen and that the resulting CIB fluctuations are very coherent.

To validate that the derived CIB image is independent from other contributions such as CMB or H I column density, we generate scatter plots for these quantities in Fig. 5.18 by comparing the CIB image to these theoretically unrelated components. While the slope and offset for the comparison of CIB and CMB is compatible with zero, there is a marginal correlation of total H I column density and the amplitude of CIB fluctuations.

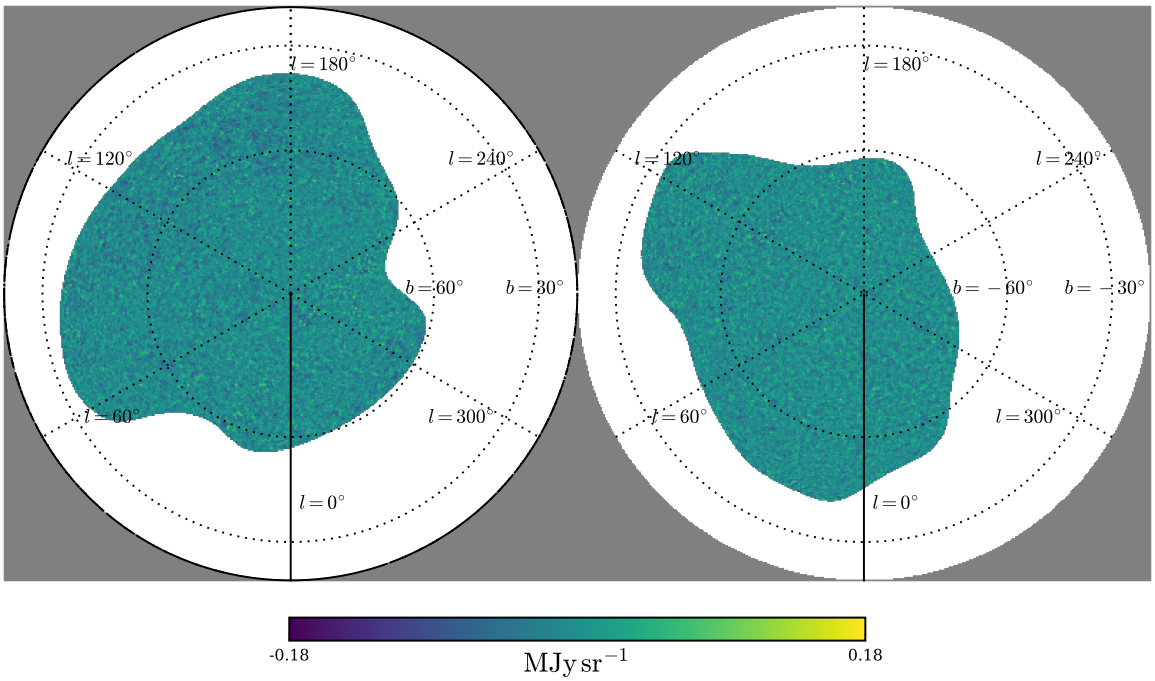


Figure 5.16: Large-scale CIB in zenith equal-area projection. The thick lines indicates $l = 0$, with l increasing clockwise. The Galactic poles are in the centre of the image (northern Galactic pole left, southern Galactic pole right). Concentric circles correspond to $|b| = 30°$ and $60°$.

Furthermore, we inspect the offset level Z_ν that is subtracted to derive the map of CIB fluctuations (Fig. 5.19(a)). We find that this offset scatters between 0 and approximately 3.2 nW m^{-2} sr^{-1} across the sky, but depends only very weakly on Galactic latitude. Its mean value is 2.3 nW m^{-2} sr^{-1} and thus in agreement with the values found in Planck Collaboration (2014 XXX) and reported here in Sect. 5.4.2. Regarding the large scatter, it is expected that Galactic dust is partially responsible for the offset on top of the pure CIB mean level.

Fig. 5.19(b) shows for how many spectral channels in the H I data cube a non-vanishing dust emissivity is found. As expected, the increasing complexity of the interstellar medium (ISM) structure towards the Galactic plane leads to an increasing number of GLM coefficients. Towards these complex regions, the automatic and robust performance of the GLM is particularly superior to the H I column density approach that would rely on manually constructing many different $N_{\mathrm{H\,I}}$ maps for each sight line.

For the *Planck* satellite, the so-called half-ring maps are generated using only the first or the second half of each pointing period. These periods have durations of typically 20 minutes. Thus, there are two half-ring maps for each frequency. This was designed to validate the mission internally and to characterise the noise by investigating the difference of these half-ring maps. Here, we use these two maps to characterise the angular power spectrum of the instrumental noise (see also Eq. 5.12). Fig. 5.19(c) shows the difference CIB map after removing the foreground for each ring-half. We find that the structure corresponds to the characteristic *Planck* noise pattern, given mostly by its scanning pattern. The power spectrum of the noise $\langle \tilde{N}_l \rangle$ is approximately two orders of magnitude weaker than the pseudo power spectrum $\langle \tilde{C}_l \rangle$, but nevertheless accounted for in our estimation of the unbiased power spectrum. Lastly, we note the excellent deterministic behaviour of the GLM without which we would not be able to

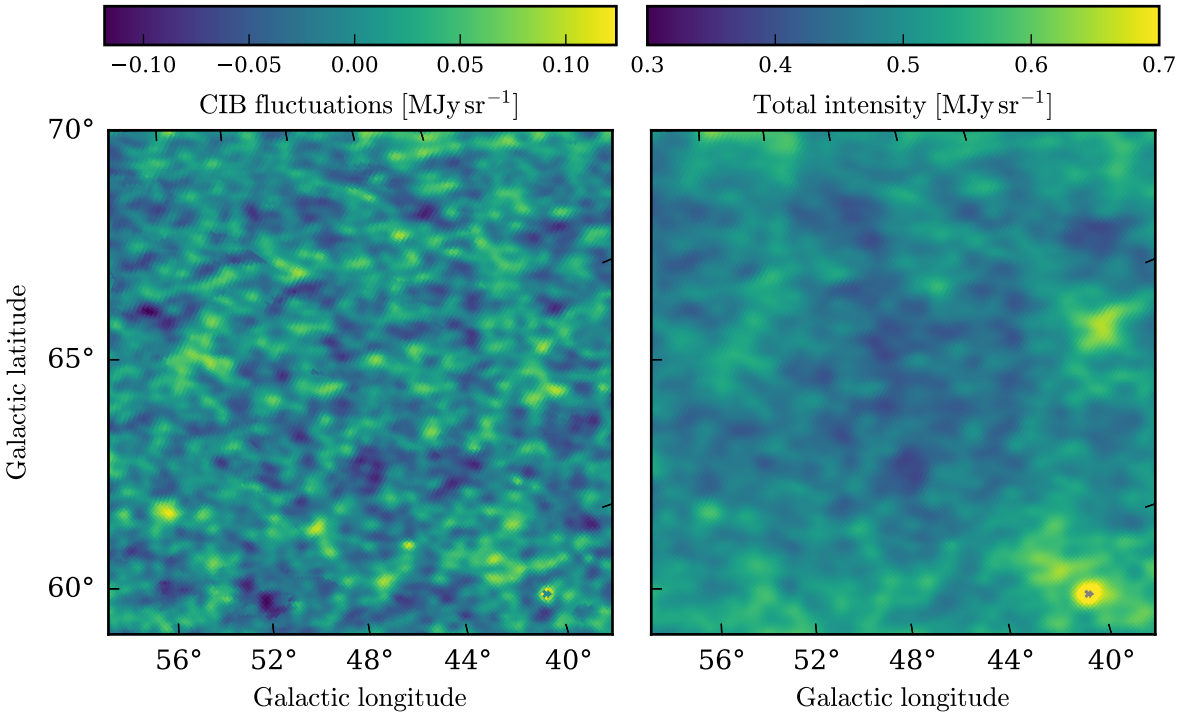


Figure 5.17: CIB (**left**) and total FIR intensity (**right**) for a smaller field, illustrating the quality of the foreground removal technique. The point source in the bottom right corner is covered by the *Planck* point source mask. Note the different scale for the colorbars. For the CIB anisotropies, the foreground was removed for each line of sight independently, only using data in an aperture of 1.5°.

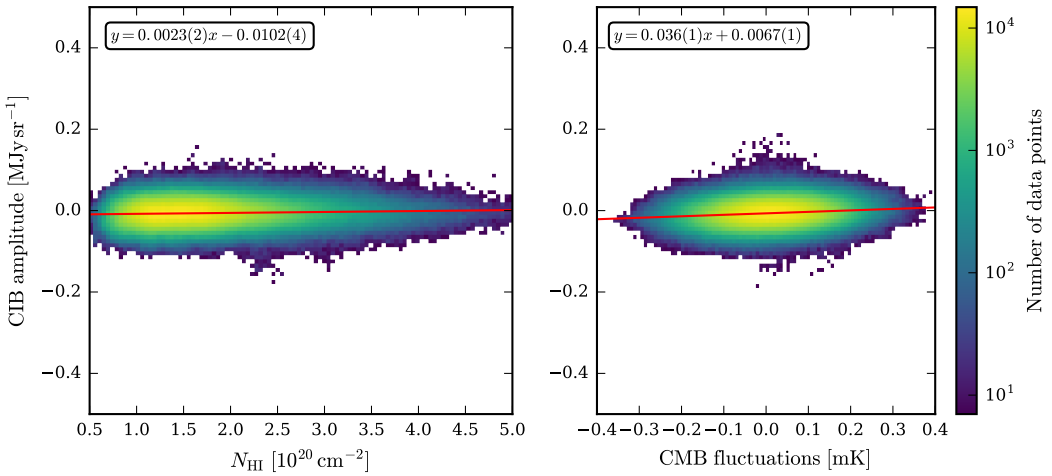


Figure 5.18: **Left:** Scatterplot of total H I column density against the derived CIB at 545 GHz. **Right:** Scatterplot of CMB fluctuations against the derived CIB at 545 GHz. For both panels, the red line shows a linear fit to the data points. Fit parameters are shown in each panel. The entire sky coverage of 20% (see also Fig. 5.16) is used for this comparison. Units are MJy sr^{-1} for the offset and MJy sr^{-1}/10^{20} cm^{-2} or MJy sr^{-1}/mK for the slope, respectively.

derive the consisting results for the two individual ring-halfs. The distribution of the GLM coefficients across the H I data cube is not sensitive to small variations in the dust map and therefore another hint for its robustness.

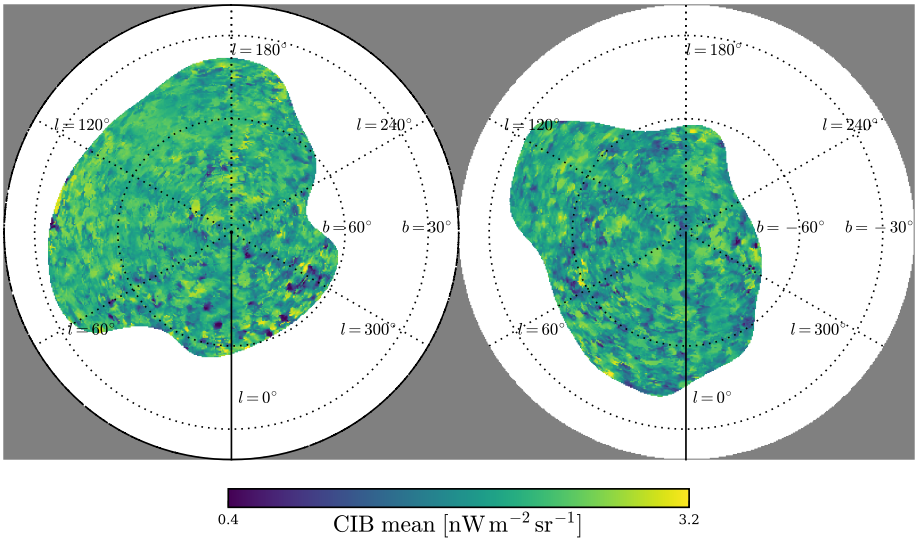
Lastly, we use the aforementioned approach to compute the unbiased, binned angular power spectrum. The resulting spectrum is shown in Fig. 5.20. The black error bars show our estimate for the binned power spectrum C_b and 1σ uncertainties, estimated with 200 Monte Carlo samples. The red line shows the results from Planck Collaboration (2014 XXX). The *Planck* uncertainties are derived in a very different way, using simulations to estimate the quality of the correction for residual Galactic dust and the CMB. Due to the lack of comparability, we decide to focus on our estimate of the power spectrum and only show the Planck Collaboration (2014 XXX) as reference. The black dotted line in Fig. 5.20 indicates the approximate multipole that corresponds to the resolution limit of $16.1'$ in this Chapter. For the *Planck* data, the resolution of approximately $4.5'$ corresponds to a multipole of $l \approx 2400$, which explains our larger uncertainties towards smaller scales.

We note that for the foreground removal, the Planck Collaboration (2014 XXX) created the dust model at the angular resolution of the H I data of approximately $10'$. However, the removal was done at the resolution of the FIR data of $4.3'$ to $4.7'$. They point out that this does not affect the cleaning process because the Galactic FIR cirrus has a power-law of approximately $l^{-2.6..2.8}$ (Miville-Deschênes et al., 2007; Bracco et al., 2011). Therefore, it has negligible power between the resolutions of the H I and the FIR data sets, compared to the power spectrum of the CIB which scales as l^{-1} (Planck Collaboration, 2011 XVIII). We carefully note that the subtraction of a low resolution map from its high-resolution counterpart describes an image sharpening technique, which is referred to as unsharp masking. This unsharp masking removes large-scale features, focusing the smaller scales. It has, for example, been applied to study the small-scale structure of the ISM (Kalberla et al., 2016), where it is used to remove the warm neutral medium (WNM) component and highlight the cold neutral medium (CNM). Additionally, the resolution of the map that results from this process is possible not well-determined and might affect the deconvolution of the angular power spectra. Thus, we are cautious about the results on the smallest scales in the *Planck* studies and decided to work with all data sets at the same resolution.

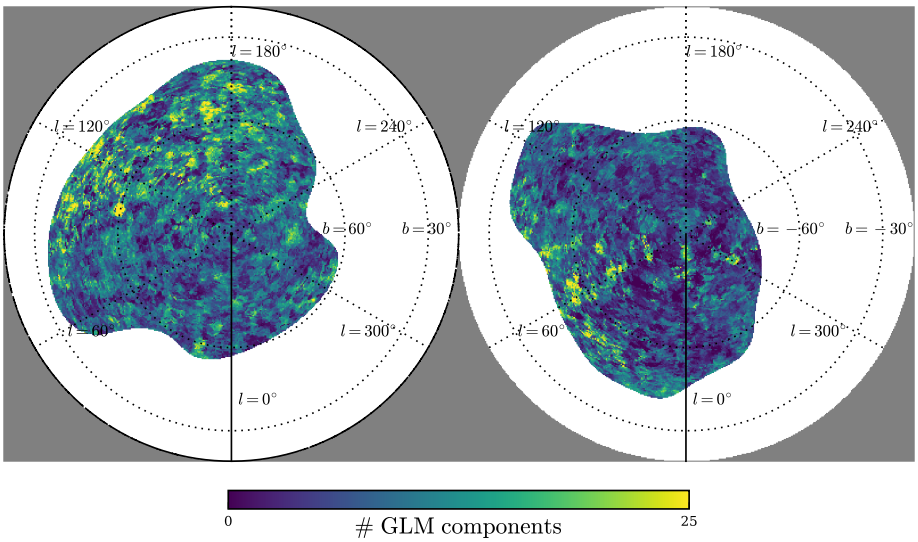
For the larger scales, we can provide first estimates of the foreground corrected power spectrum that have not been probed before. Other studies of the CIB at larger scales have not applied any correction to account for Galactic foreground emission (e.g. Serra et al., 2014) because of the complexity and difficulty in modelling the ISM at large scales. Here, we can provide the community with accurately and robustly cleaned CIB images at different frequencies that will greatly reduce the systematics e.g. in the correlation between CIB anisotropies and weak lensing, tracing the interplay of dark matter halos and dusty FIR galaxies.

5.6 Future improvements

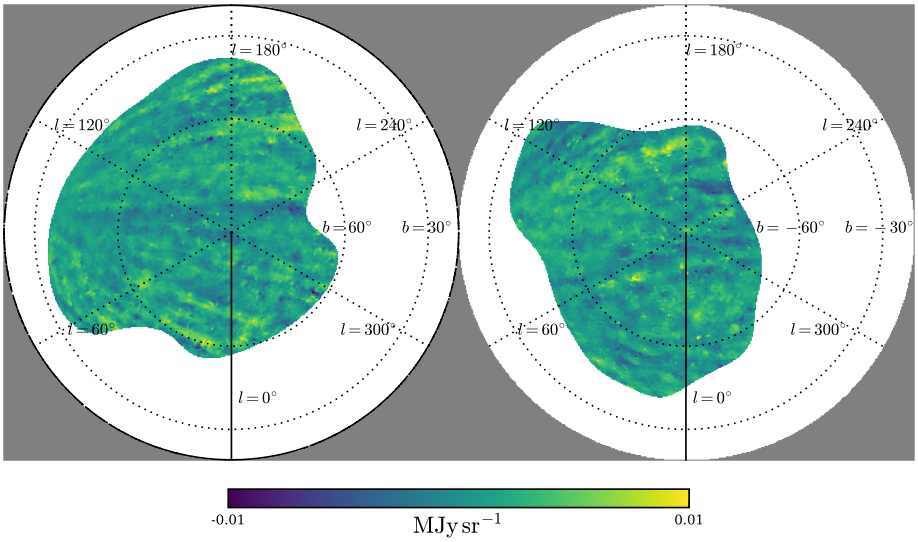
To access lower Galactic latitudes and thus regions of higher column density, we need to accurately **model the molecular content** of the objects at these locations. CO-dark and CO-bright molecular gas will contribute to the FIR foreground intensity, but cannot be completely probed by H I. Unfortunately, most surveys of CO lines are only available for the Galactic plane (Dame, Hartmann and Thaddeus, 2001; Mizuno and Fukui, 2004). The full sky CO images, obtained from the *Planck* component separation (Planck Collaboration, 2013 XIII; Planck Collaboration, 2015 X), do not offer the sensitivity required to detect some of the known CO bright intermediate-velocity molecular clouds (IVMCs) such as IVC135 (Weiß et al., 1999; Lenz et al., 2015) or other high latitude molecular gas (Magnani, Hartmann and Speck, 1996). Nonetheless, a model that makes proper use of large-scale CO data could help to extend



(a)



(b)



(c)

Figure 5.19: **(a)** FIR offset, derived with the model. While this offset is dominated by the CIB mean level, other Galactic components additionally increase or decrease this quantity. **(b)** Number of GLM coefficients for each line of sight. As expected, the number of components required to accurately describe the correlation between dust and gas increases towards the Galactic plane. **(c)** CIB difference map for the two *Planck* ring halfs, showing the typical *Planck* scanning pattern.

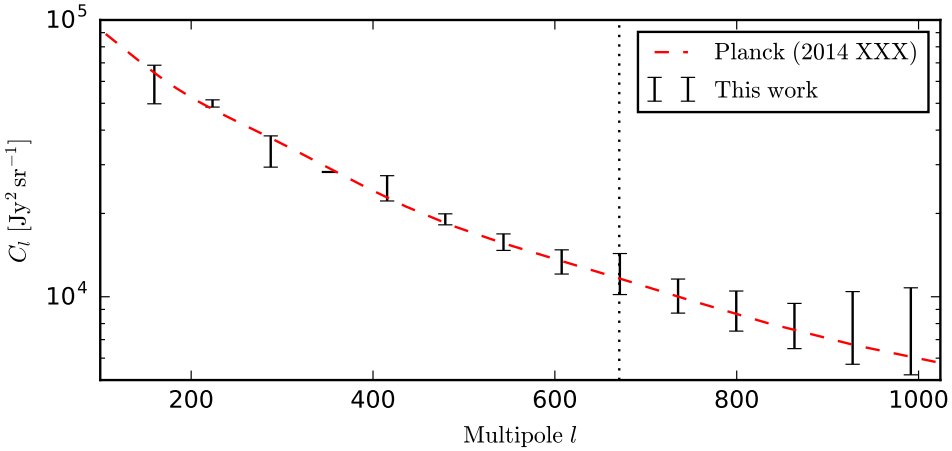


Figure 5.20: Unbiased, binned CIB power spectrum at 545 GHz for 20% of the sky. The black error bars indicate the 1σ uncertainties, derived with 200 Monte Carlo estimates. The red line corresponds to the results from Planck Collaboration (2014 XXX) and the dashed red line indicates the approximate resolution of the data sets used in the present study. Due to the very complex and different approach to quantify the uncertainties in the Planck Collaboration (2014 XXX) results, we abstain from showing them as the two estimates for the errors would not be comparable.

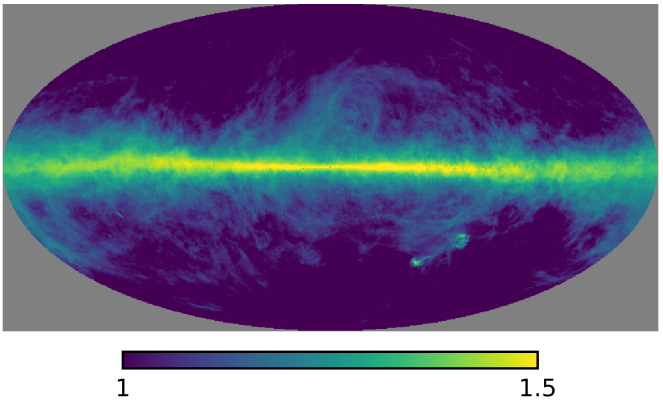


Figure 5.21: Correction factor f for optical thick H I, derived from the model in Lee et al. (2015) and based on the HI4PI H I column density (HI4PI collaboration, 2016, submitted). By multiplying the observed $N_{\mathrm{H\,I}}$ values with the correction factor, the actual H I column density can be derived.

the foreground removal to lower Galactic latitudes and to the molecular cirrus clouds.

One of the biggest challenges in modelling the foreground intensity is the **CO-dark molecular gas**. Here, we rely on other tracers of the molecular phase such as OH (Allen, Hogg and Engelke, 2015) or C^+ (Pineda et al., 2013; Langer et al., 2014; Tang et al., 2016) to understand the structure of the ISM at these intermediate column densities. For this gas phase, we cannot rely solely on H I and CO data. However, only small areas mostly in the Galactic plane have been studied up to now and an application to larger regions will need to be developed in the future.

A part of the dark gas can be explained by **optical thick H I**. The commonly applied equation to derive the H I column density is given by the velocity integral over the H I brightness temperature distribution T_B:

$$N_{\mathrm{H\,I}}\,[\mathrm{cm}^{-2}] = 1.822 \times 10^{18} \int \mathrm{d}v\; T_B\,[\mathrm{K\,km\,s}^{-1}] \tag{5.22}$$

If the H I becomes optically thick (i.e. $\tau \gg 1$), Eq. (5.22) is not valid anymore and an estimate of the H I spin temperature T_S is required to accurately calculate the column density:

$$T_{\mathrm{spin}} = \frac{T_B}{1 - e^{\tau}} \tag{5.23}$$

$$N_{\mathrm{H\,I}}\,[\mathrm{cm}^{-2}] = 1.823 \times 10^{18} \int \mathrm{d}v\, \tau\, T_{\mathrm{spin}}\,[\mathrm{K\,km\,s}^{-1}] \tag{5.24}$$

According to Fukui et al. (2014), this is the exclusive explanation for the dark gas, but several studies find that CO-dark molecular gas is instead responsible for the major fraction of the dark gas (Planck Collaboration, 2011 XIX; Lee et al., 2015; Reach, Heiles and Bernard, 2015). Based on H I absorption line measurements in the direction of Perseus, Lee et al. (2015) estimate the correction factor for the observed H I column density as $f = \log_{10}\left(N_{\mathrm{H\,I}}^{\mathrm{obs}}/10^{20}\right) \times (0.32 \pm 0.06) + (0.81 \pm 0.05)$. Even though this is an empirical correction factor for a small portion of the full sky and should be taken with caution, we extrapolate this correction factor f to the full sky, based on the HI4PI data (Fig. 5.21). We find that for the present analysis of the CIB at high Galactic latitudes, the amount is dark gas due to H I self absorption is negligible. Towards the Galactic centre and the Galactic plane, the modelling of the foreground is much more difficult and a correction for self absorption should be included as well, given the large values of $f \gtrsim 1.5$ for the inner Galaxy.

Independent of the model, the foreground removal algorithm would benefit from being realised in a **Bayesian framework** instead of just maximising the posterior. This would yield a better estimate of the uncertainties and could potentially provide a more accurate solution. Moreover, it could help to understand the correct assumptions and the limitations of our model. However, this is computationally very demanding and cannot be realised in its current form. On a modern 32 core machine, it takes approximately 6 days to clean a single FIR map at one frequency from the foreground dust. The expansion to a fully probabilistic treatment with stable MC sampling would increase this to the timescale of multiple weeks. For a future analysis, we will develop the tools to perform this type of analysis, but it is beyond the scope of the present work.

Instead of masking the point sources, there have been very recent and very promising applications of the **point source inpainting technique** (Remazeilles et al., 2016; Planck Collaboration, 2016 XLVIII). This technique uses a spline interpolation to locally fill the holes produced by point sources, which helps to improve the robustness of the power spectrum estimation. This process is hindered by the presence of many small masked regions, leading to oscillations and uncertainties in the power spectra and therefore in the derived results.

To validate the regularisation strength α and other hyperparameters on a larger scale, it would be beneficial to generate a full foreground-removed power spectrum for a set of random choices of the hyperparameters. This random search could help to test the **choice of the hyperparameters** on a larger scale, but is computationally very expensive.

Lastly, we would like to note that the *Planck* collaboration has an uncontested expertise in deriving and working with angular power spectra. Therefore, we use the power spectra computed in this work as a first estimate, but highlight that our focus is the accurate and robust removal of foreground emission. This is based on our experience with the ISM and Galactic H I in particular. By combining this knowledge with the cosmological community, this work can lead to the advancement of both fields.

5.7 Summary

5.7.1 Conclusion

We have presented new large-scale measurements of the CIB anisotropies, using recently published data from the HI4PI Survey and the *Planck* satellite. Starting on well-studied fields, we have verified the GLM approach, which uses the full spectral information of the H I data cube to remove the FIR foreground emission. On top of that, the CMB fluctuations were removed from the maps to avoid a leakage into the CIB images. We have demonstrated the advantage of the GLM over the simple, H I column density-based approach. Moreover, we find slightly ($\sim 10\%$) smaller amplitudes for the CIB fluctuations than Planck Collaboration (2014 XXX), which is in line with the findings of Planck Collaboration (2016 XLVIII). For the IRIS data at 3000 GHz, our results on the CIB mean level show tension with respect to the *Planck* results which requires further investigation. Therefore, we constrained ourselves to the *Planck* data in the following steps of the analysis.

By applying the foreground-removal technique for an aperture around each line of sight at high Galactic latitudes, we have produced maps of the CIB fluctuations for 20% of the full sky. This is an improvement of a factor of two with respect to previous studies. For this large-scale analysis, we have verified our results by showing that the derived CIB anisotropies do not correlate with other emission processes such as the CMB or residual Galactic dust emission. Finally, we derived the unbiased, binned angular power spectrum of the CIB for 353, 545, and 857 GHz and compared it to previous studies and to a simple set of simulations. Within the uncertainties, we find that our results agree very well with the other studies. As expected, our derived uncertainties increase towards the resolution limit of our analysis. For the larger scales, however, we are able to very strongly constrain the CIB power spectrum due to our unprecedented sky coverage.

5.7.2 Outlook

First and foremost, the derived angular power spectra will be used to constrain cosmological halo models and therefore help to understand the galaxies that emit the FIR emission which we observe as the CIB. This helps to understand the reionisation history of the Universe and the SFH in dark matter halos across cosmic time. We will not only develop these models ourselves, but the CIB images and the angular power spectra will be made publicly available.

To improve the quality and the sky coverage of the foreground-removal and thus the CIB images, we have discussed several points that need to be addressed in a future study. The implementation of the algorithm needs to be improved and approximative methods need to be developed to apply a fully Bayesian framework to the data. Moreover, the combination with foreground-removal techniques which utilise the statistical spatial properties of the Galactic thermal dust and the CIB will verify our results. Here, we are particularly looking forward to the public release of the CIB maps from Planck Collaboration (2016 XLVIII). On top of that, the framework of information field theory has been demonstrated to yield promising results for the CMB component separation and could potentially improve the foreground removal for the CIB as well. Lastly, we note that future models will have to account for gas at higher densities to access lower Galactic latitudes. Therefore, we need to include a correction for optical thick H I and a model for the CO-dark and -bright molecular gas. This emphasises the importance of the collaboration between the ISM community that aims to understand the properties and transitions of the Galactic ISM phases and the cosmological community, which seeks to remove all this foreground emission to study the distant Universe.

CHAPTER 6

Conclusions

6.1 Results from the detailed study of IVCs and HVCs

In Chapters 3 and 4, we have investigated two H I halo cloud systems, their properties, and their role in the evolution of the Milky Way (MW) Galaxy. These studies made use of data from the *Planck* satellite and the recently finished EBHIS data, allowing the access to higher spatial resolutions and higher sensitivities.

In Chapter 3, we analysed a system that consists of a IVC and a HVC. For these clouds, we find several hints towards an inelastic collision between the two. First, the HVC exhibits a clear head-tail structure in the H I data and is continuously connected to the IVC in velocity space, creating a velocity bridge. These two phenomena have been observed for other HVCs and IVCs as well and are strong indicators for a collision between the two. Furthermore, we find FIR excess emission in the IVC at the suggested impact region of the HVC. This could results from the pressure-induced reduction of the dust formation time or from sweeping up the existing dust. Furthermore, we have shown that the IVC contains significant amount of molecular hydrogen, traced by the FIR excess emission, which is in line with previous detections of CO. Lastly, we have noted that the apparent distance discrepancy between the two objects could be explained by an overestimation of the HVC distance. We also note that the indicators for an interaction might be misleading and have another origin, but that does not seem very plausible. In the bigger picture, this means that we can locally observe the refuelling of the MW Galaxy, which is required to sustain the continuous star formation. The replenishment with pristine gas is one of the key aspects of Galaxy evolution, and the mixing of high- and intermediate-velocity gas at the disk-halo interface appears to be one of the most important mechanisms, for which this work adds an important piece of the puzzle.

We expanded our methodological repertoire and improved our understanding of the H I and FIR data throughout the progress of this thesis, leading to content published in Chapter 4. Here, we investigated the potential dust content of a cold, compact HVC, which could be a powerful tool of its origin and properties. The most severe systematic in this process is the removal of foreground dust emission, associated with the local MW gas. Moreover, the uncertainties must be accurately estimated to allow a significant determination of the HVC's dust content. We have presented a new method to remove the foreground dust emission by using the full spectral information of the H I data cube. This GLM technique was validated on simulated data and internally via cross validation, showing that it is superior to a simple approach that uses the H I column density as a template for the FIR foreground. After removing the foreground emission for the selected HVC, we employed a simple Bayesian model to determine its dust emissivity and found that it does not emit any significant amount of FIR emission. The 3σ upper limit to

the dust emissivity is more than an order of magnitude than typical values for Galactic dust, which means that the correlation of atomic neutral hydrogen and dust grains is not a suitable tool to investigate the potential dust content of HVCs. Whether the FIR deficiency stems from the missing heating through nearby stars or from the low metallicity and therefore low dust content could not be concluded, based on the present results.

6.2 Results from the large-scale analysis of the CIB

The GLM modelling technique that was developed for the search of dust in HVCs has led to the conclusion that the CIB anisotropies are the dominant source of uncertainty in this very low signal-to-noise regime. We greatly expanded this study of the CIB in Chapter 5, using the recently published HI4PI full-sky data of H I and the latest *Planck* data release. After expanding our foreground removal technique, allowing to apply it automatically to large fractions of the sky, we started the analysis on individual, smaller fields that have been studied in previous studies. Thereby, we validate our approach and compare it to the standard, H I column density approach. We have shown that our results are in agreement with the previous *Planck* studies, with some tension for the IRIS data that might be due to a different data reduction scheme.

For the following large-scale analysis, we constrained ourselves to using the 353, 545, and 857 GHz data and used the HI4PI data to remove the Galactic FIR foreground for the Galactic poles, covering 20% of the full sky. Furthermore, we generate an estimate of the deconvolved, unbiased angular power spectra for these frequencies and show that our results are consistent the previously derived angular power spectra. The large advantage of the approach and the data presented here is the study of the largest scales that have not been accessed before, and the present work serves as a pilot study for future work.

6.3 Outlook

For the future, we aim to continue the research on the multiphase modelling, based on the latest high-quality observational data sets. The advanced understanding of the interplay of the different ISM phases is also the key to the extragalactic sky, allowing the study of cosmological background light such as the CIB and CMB.

More specifically, we will develop the models that are compared to the CIB measurements in order to understand the reionisation of the Universe and the star-formation history in dark matter halos. To improve the foreground modelling that is required to derive the CIB images, we need to expand our models to a fully Bayesian framework that will be mathematically and computationally challenging. Moreover, the combination with CIB separation algorithms which use the statistical spatial information will help to independently verify our approach. Here, we look forward to employing the statistical framework of information field theory.

Lastly, we point out the interdisciplinary nature of this work. Without a proper understanding of the different ISM phases in the MW and their interplay, we will not be able to accurately account for the foreground component in extragalactic studies. Here, observations of different molecular species such as CO, C^+, and OH can help to model the high-density regions in the Galactic plane. New facilities such as ALMA, APEX and SOFIA offer a great opportunity to improve our understanding of the ISM at the smallest scales.

Bibliography

Alibés, A., J. Labay and R. Canal (2001), *Galactic chemical abundance evolution in the solar neighborhood up to the iron peak*, A&A **370** 1103, eprint: arXiv:astro-ph/0012505 (cit. on pp. 10, 28).

Allen, R. J., D. E. Hogg and P. D. Engelke (2015), *The Structure of Dark Molecular Gas in the Galaxy. I. A Pilot Survey for 18 cm OH Emission Toward l=105 deg, b=+1deg*, AJ **149**, 123 123, arXiv: 1502.00657 (cit. on pp. 14, 84).

Amblard, A. et al. (2011), *Submillimetre galaxies reside in dark matter haloes with masses greater than* $3x10^11$ *solar masses*, Nature **470** 510, arXiv: 1101.1080 (cit. on p. 15).

Arnal, E. M. et al. (2000), *A high sensitivity Hi survey of the sky at delta larger than -25 deg*, A&AS **142** 35 (cit. on p. 17).

Bajaja, E. et al. (2005), *A high sensitivity HI survey of the sky at delta larger than -25 deg. Final data release*, A&A **440** 767, eprint: astro-ph/0504136 (cit. on p. 17).

Barger, K. A. et al. (2012), *Present-day Galactic Evolution: Low-metallicity, Warm, Ionized Gas Inflow Associated with High-velocity Cloud Complex A*, ApJ **761**, 145 145, arXiv: 1211.1973 [astro-ph.GA] (cit. on p. 28).

Barnes, D. G., F. H. Briggs and M. R. Calabretta (2005), *Postcorrelation ripple removal and radio frequency interference rejection for Parkes Telescope survey data*, Radio Science **40**, RS5S13 RS5S13 (cit. on p. 20).

Barriault, L., G. Joncas and R. Plume (2011), *Multiwavelength observations of cirrus clouds in the North Celestial Loop: physical parameters of molecular sites*, MNRAS **416** 1250 (cit. on p. 38).

Barriault, L. et al. (2010), *Multiwavelength observations of cirrus clouds in the North Celestial Loop: the transition from atomic to molecular gas*, MNRAS **406** 2713 (cit. on p. 38).

Ben Bekhti, N. et al. (2012), *An absorption-selected survey of neutral gas in the Milky Way halo. New results based on a large sample of Ca ii, Na i, and H i spectra towards QSOs*, A&A **542**, A110 A110, arXiv: 1203.5603 [astro-ph.GA] (cit. on pp. 7, 9, 28).

Benjamin, R. A. et al. (1996), *The Distance to an X-Ray Shadowing Molecular Cloud in Ursa Major*, ApJ **464** 836 (cit. on pp. 7, 29, 30, 32, 40).

Bennett, C. L. et al. (2003), *The Microwave Anisotropy Probe Mission*, ApJ **583** 1, eprint: astro-ph/0301158 (cit. on pp. 22, 61, 67).

Bennett, C. L. et al. (2013), *Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Final Maps and Results*, ApJS **208**, 20 20, arXiv: 1212.5225 (cit. on p. 22).

Boggess, N. W. et al. (1992), *The COBE mission - Its design and performance two years after launch*, ApJ **397** 420 (cit. on pp. 22, 60).

Bolatto, A. D., M. Wolfire and A. K. Leroy (2013), *The CO-to-H_2 Conversion Factor*, ARA&A **51** 207, arXiv: 1301.3498 (cit. on p. 5).

Borrill, J. (1999), *Power spectrum estimators for large CMB datasets*, Phys. Rev. D **59**.2, 027302 027302, eprint: astro-ph/9712121 (cit. on p. 67).

Boulanger, F. et al. (1996), *The dust/gas correlation at high Galactic latitude.*, A&A **312** 256 (cit. on pp. 13, 33, 36, 64).

Bracco, A. et al. (2011), *Herschel-ATLAS: statistical properties of Galactic cirrus in the GAMA-9 Hour Science Demonstration Phase Field*, MNRAS **412** 1151, arXiv: 1011.0725 (cit. on pp. 61, 82).

Braun, R. (2012), *Cosmological Evolution of Atomic Gas and Implications for 21 cm H I Absorption*, ApJ **749**, 87 87, arXiv: 1202.1840 [astro-ph.CO] (cit. on pp. 34, 38).

Braun, R. and W. B. Burton (2000), *High-resolution imaging of compact high-velocity clouds*, A&A **354** 853, eprint: astro-ph/9912417 (cit. on p. 44).

Braun, R. and D. A. Thilker (2004), *The WSRT wide-field H I survey. II. Local Group features*, A&A **417** 421, eprint: astro-ph/0312323 (cit. on p. 9).

Brüns, C., J. Kerp and A. Pagels (2001), *Deep H {I} observations of the compact high-velocity cloud {HVC 125+41-207}*, A&A **370** L26, eprint: astro-ph/0103119 (cit. on p. 44).

Brüns, C. et al. (2000), *The head-tail structure of high-velocity clouds. A survey of the northern sky*, A&A **357** 120 (cit. on pp. 9, 10, 32, 40).

Calabretta, M. R. and E. W. Greisen (2002), *Representations of celestial coordinates in FITS*, A&A **395** 1077, eprint: astro-ph/0207413 (cit. on p. 30).

Cardoso, J.-F. et al. (2008), *Component Separation With Flexible Models. Application to Multichannel Astrophysical Observations*, IEEE Journal of Selected Topics in Signal Processing **2** 735 (cit. on p. 63).

Chanial, P. (2003), Promotion: Paris-XI Univ. (cit. on p. 61).

Chomiuk, L. and M. S. Povich (2011), *Toward a Unification of Star Formation Rate Determinations in the Milky Way and Other Galaxies*, AJ **142**, 197 197, arXiv: 1110.4105 [astro-ph.GA] (cit. on pp. 10, 27).

Comeron, F. and J. Torra (1994), *The origin of the Gould Belt by the impact of a high velocity cloud on the galactic disk*, A&A **281** 35 (cit. on pp. 11, 28).

Cooray, A. and R. Sheth (2002), *Halo models of large scale structure*, Phys. Rep. **372** 1, eprint: astro-ph/0206508 (cit. on p. 77).

Cox, D. P. (2005), *The Three-Phase Interstellar Medium Revisited*, ARA&A **43** 337 (cit. on p. 7).

D'Agostini, G. (2003), *Bayesian inference in processing experimental data: principles and basic applications*, Reports on Progress in Physics **66** 1383, eprint: physics/0304102 (cit. on p. 47).

Dame, T. M., D. Hartmann and P. Thaddeus (2001), *The Milky Way in Molecular Clouds: A New Complete CO Survey*, ApJ **547** 792, eprint: arXiv:astro-ph/0009217 (cit. on pp. 14, 23, 82).

Danly, L., C. E. Albert and K. D. Kuntz (1993), *A Determination of the Distance to the High-Velocity Cloud Complex M*, ApJ **416** L29 (cit. on p. 28).

de Souza, R. S. et al. (2014), *The Overlooked Potential of Generalized Linear Models in Astronomy - I: Binomial Regression and Numerical Simulations*, ArXiv e-prints, arXiv: 1409.7696 [astro-ph.IM] (cit. on pp. 46, 64, 65).

Delabrouille, J. et al. (2009), *A full sky, low foreground, high resolution CMB map from WMAP*, A&A **493** 835, arXiv: 0807.0773 (cit. on p. 61).

Desert, F. X., D. Bazell and F. Boulanger (1988), *An all-sky search for molecular cirrus clouds*, ApJ **334** 815 (cit. on pp. 13, 33, 36, 43).

Diehl, R. et al. (2006), *Radioactive ^{26}Al from massive stars in the Galaxy*, Nature **439** 45, eprint: astro-ph/0601015 (cit. on p. 10).

Dole, H. et al. (2004), *Far-infrared Source Counts at 70 and 160 Microns in Spitzer Deep Surveys*, ApJS **154** 87, eprint: astro-ph/0406021 (cit. on p. 60).

Dole, H. et al. (2004b), *Confusion of Extragalactic Sources in the Mid- and Far-Infrared: Spitzer and Beyond*, ApJS **154** 93 (cit. on p. 60).

Dole, H. et al. (2006), *The cosmic infrared background resolved by Spitzer. Contributions of mid-infrared galaxies to the far-infrared background*, A&A **451** 417, eprint: astro-ph/0603208 (cit. on p. 15).
Draine, B. T. (2011), *Physics of the Interstellar and Intergalactic Medium* (cit. on pp. 3–5, 33).
Draine, B. T. and F. Bertoldi (1999), "Heating the gas in photodissociation regions", *The Universe as Seen by ISO*, ed. by P. Cox and M. Kessler, vol. 427, ESA Special Publication 553, eprint: astro-ph/9902379 (cit. on p. 4).
Draine, B. T. and A. Li (2007), *Infrared Emission from Interstellar Dust. IV. The Silicate-Graphite-PAH Model in the Post-Spitzer Era*, ApJ **657** 810, eprint: arXiv:astro-ph/0608003 (cit. on p. 78).
Draine, B. T. et al. (2007), *Dust Masses, PAH Abundances, and Starlight Intensities in the SINGS Galaxy Sample*, ApJ **663** 866, eprint: astro-ph/0703213 (cit. on p. 38).
Ewen, H. I. and E. M. Purcell (1951), *Observation of a Line in the Galactic Radio Spectrum: Radiation from Galactic Hydrogen at 1,420 Mc./sec.*, Nature **168** 356 (cit. on p. 3).
Field, G. B. (1965), *Thermal Instability.*, ApJ **142** 531 (cit. on p. 4).
Field, G. B., D. W. Goldsmith and H. J. Habing (1969), *Cosmic-Ray Heating of the Interstellar Gas*, ApJ **155** L149 (cit. on p. 6).
Flöer, L., B. Winkel and J. Kerp (2010), "RFI mitigation for the Effelsberg Bonn HI Survey (EBHIS)", *RFI Mitigation Workshop*, arXiv: 1007.2328 [astro-ph.IM] (cit. on p. 20).
Foreman-Mackey, D. et al. (2013), *emcee: The MCMC Hammer*, PASP **125** 306, arXiv: 1202.3665 [astro-ph.IM] (cit. on p. 49).
Fox, A. J., B. D. Savage and B. P. Wakker (2006), *A Survey of O VI, C III, and H I in Highly Ionized High-Velocity Clouds*, ApJS **165** 229, eprint: astro-ph/0604091 (cit. on pp. 9, 28).
Fox, A. J. et al. (2004), *Highly Ionized Gas Surrounding High-Velocity Cloud Complex C*, ApJ **602** 738, eprint: arXiv:astro-ph/0310769 (cit. on pp. 7, 9, 29, 33, 44).
Fox, A. J. et al. (2016), *On the Metallicity and Origin of the Smith High-velocity Cloud*, ApJ **816**, L11 L11, arXiv: 1512.04957 (cit. on p. 11).
Fraternali, F. (2014), "How can star formation be sustained?", *Setting the scene for Gaia and LAMOST*, ed. by S. Feltzing et al., vol. 298, IAU Symposium 228, arXiv: 1310.2956 (cit. on p. 10).
Fraternali, F. and J. J. Binney (2008), *Accretion of gas on to nearby spiral galaxies*, MNRAS **386** 935, arXiv: 0802.0496 (cit. on p. 28).
Fraternali, F. and M. Tomassetti (2012), *Estimating gas accretion in disc galaxies using the Kennicutt-Schmidt law*, MNRAS **426** 2166, arXiv: 1207.0093 (cit. on p. 10).
Fraternali, F. et al. (2013), *Ionized Absorbers as Evidence for Supernova-driven Cooling of the Lower Galactic Corona*, ApJ **764**, L21 L21, arXiv: 1301.4998 [astro-ph.GA] (cit. on pp. 7, 8, 13, 28).
Fraternali, F. et al. (2015), *Galactic hail: the origin of the high-velocity cloud complex C*, MNRAS **447** L70, arXiv: 1411.4050 (cit. on pp. 12, 13).
Fukui, Y. et al. (2014), *Optically thick HI dominant in the local interstellar medium; an alternative interpretation to "dark gas"*, ArXiv e-prints, arXiv: 1403.0999 [astro-ph.GA] (cit. on pp. 14, 85).
Gardiner, L. T. (1999), "N-body Simulations of the Magellanic Stream", *Stromlo Workshop on High-Velocity Clouds*, ed. by B. K. Gibson and M. E. Putman, vol. 166, Astronomical Society of the Pacific Conference Series 292 (cit. on p. 8).
Gentile, G. et al. (2013), *HALOGAS: Extraplanar gas in NGC 3198*, A&A **554**, A125 A125, arXiv: 1304.4232 (cit. on p. 11).
Gillmon, K. et al. (2006), *A FUSE Survey of Interstellar Molecular Hydrogen toward High-Latitude AGNs*, ApJ **636** 891, eprint: astro-ph/0507581 (cit. on p. 56).
Gladders, M. D. et al. (1998), *The Distance to the Draco Intermediate-Velocity Cloud*, ApJ **507** L161 (cit. on p. 7).

Glover, S. C. O. (2003), *Comparing Gas-Phase and Grain-catalyzed H_2 Formation*, ApJ **584** 331, eprint: astro-ph/0210493 (cit. on p. 5).

Goodman, J. and J. Weare (2010), *Ensemble samplers with affine invariance.*, English, Commun. Appl. Math. Comput. Sci. **5** 65 (cit. on p. 49).

Górski, K. M. et al. (2005), *HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere*, ApJ **622** 759, eprint: arXiv:astro-ph/0409513 (cit. on pp. 24, 25, 29, 57).

Grenier, I. A., J.-M. Casandjian and R. Terrier (2005), *Unveiling Extensive Clouds of Dark Gas in the Solar Neighborhood*, Science **307** 1292 (cit. on p. 14).

Guillard, P. et al. (2009), *H_2 formation and excitation in the Stephan's Quintet galaxy-wide collision*, A&A **502** 515, arXiv: 0904.4239 [astro-ph.CO] (cit. on pp. 32, 40, 44, 56).

Haffner, L. M. et al. (2003), *The Wisconsin Hα Mapper Northern Sky Survey*, ApJS **149** 405, eprint: astro-ph/0309117 (cit. on p. 37).

Hartmann, D. and W. B. Burton (1997), *Atlas of Galactic Neutral Hydrogen* (cit. on p. 17).

Hartmann, D., L. Magnani and P. Thaddeus (1998), *A Survey of High-Latitude Molecular Gas in the Northern Galactic Hemisphere*, ApJ **492** 205 (cit. on p. 7).

Haud, U. (2000), *Gaussian decomposition of the Leiden/Dwingeloo survey. I. Decomposition algorithm*, A&A **364** 83 (cit. on p. 20).

Hauser, M. G. and E. Dwek (2001), *The Cosmic Infrared Background: Measurements and Implications*, ARA&A **39** 249, eprint: astro-ph/0105539 (cit. on pp. 15, 60).

Heald, G. et al. (2011), *The Westerbork Hydrogen Accretion in LOcal GAlaxieS (HALOGAS) survey. I. Survey description and pilot observations*, A&A **526**, A118 A118, arXiv: 1012.0816 (cit. on p. 11).

Heitsch, F. and M. E. Putman (2009), *The Fate of High-Velocity Clouds: Warm or Cold Cosmic Rain?*, ApJ **698** 1485, arXiv: 0904.1995 [astro-ph.GA] (cit. on pp. 11, 39).

Henley, D. B., R. L. Shelton and K. Kwak (2014), *The Origin of the X-Ray Emission from the High-velocity Cloud MS30.7-81.4-118*, ApJ **791**, 41 41, arXiv: 1406.6363 (cit. on p. 12).

Hernandez, A. K. et al. (2013), *A Low-metallicity Molecular Cloud in the Lower Galactic Halo*, ApJ **777**, 19 19, arXiv: 1308.6313 [astro-ph.GA] (cit. on pp. 30, 39).

HI4PI collaboration (2016, submitted), *HI4PI: All-sky HI column density map from EBHIS and GASS* (cit. on pp. 21, 62, 84).

Hivon, E. et al. (2002), *MASTER of the Cosmic Microwave Background Anisotropy Power Spectrum: A Fast Method for Statistical Analysis of Large and Complex Cosmic Microwave Background Data Sets*, ApJ **567** 2, eprint: astro-ph/0105302 (cit. on pp. 61, 67, 69).

Houck, J. C. and J. N. Bregman (1990), *Low-temperature Galactic fountains*, ApJ **352** 506 (cit. on pp. 8, 12, 28).

J. A. Nelder, R. W. M. W. (1972), *Generalized Linear Models*, Journal of the Royal Statistical Society. Series A (General) **135**.3 370, ISSN: 00359238, URL: http://www.jstor.org/stable/2344614 (cit. on pp. 64, 65).

Jeffreys, H. (1946), *An Invariant Form for the Prior Probability in Estimation Problems*, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences **186**.1007 453, ISSN: 0080-4630 (cit. on p. 49).

Kalberla, P. M. W. and L. Dedes (2008), *Global properties of the H I distribution in the outer Milky Way. Planar and extra-planar gas*, A&A **487** 951, arXiv: 0804.4831 (cit. on pp. 6–8).

Kalberla, P. M. W. and U. Haud (2006), *Global properties of the HI high velocity sky. A statistical investigation based on the LAB survey*, A&A **455** 481, eprint: astro-ph/0604606 (cit. on p. 8).

– (2015), *GASS: The Parkes Galactic All-Sky Survey. Update: improved correction for instrumental effects and new data release*, A&A **578**, A78 A78, arXiv: 1505.01011 (cit. on pp. 19, 20, 62).

Kalberla, P. M. W. and J. Kerp (2009), *The Hi Distribution of the Milky Way*, ARA&A **47** 27 (cit. on pp. 6, 7, 11, 29).

Kalberla, P. M. W., U. Mebold and W. Reich (1980a), *Time variable 21 CM lines and the stray radiation problem*, A&A **82** 275 (cit. on p. 20).

Kalberla, P. M. W., U. Mebold and L. Velden (1980b), *Time variable 21 CM lines in high galactic latitudes*, A&AS **39** 337 (cit. on p. 20).

Kalberla, P. M. W. et al. (2005), *The Leiden/Argentine/Bonn (LAB) Survey of Galactic HI. Final data release of the combined LDS and IAR surveys with improved stray-radiation corrections*, A&A **440** 775, eprint: arXiv:astro-ph/0504140 (cit. on pp. 8, 17, 62).

Kalberla, P. M. W. et al. (2007), *Dark matter in the Milky Way. II. The HI gas distribution as a tracer of the gravitational potential*, A&A **469** 511, arXiv: 0704.3925 (cit. on p. 7).

Kalberla, P. M. W. et al. (2010), *GASS: the Parkes Galactic all-sky survey. II. Stray-radiation correction and second data release*, A&A **521**, A17 A17, arXiv: 1007.0686 (cit. on pp. 14, 20, 62).

Kalberla, P. M. W. et al. (2016), *Cold Milky Way HI Gas in Filaments*, ApJ **821**, 117 117, arXiv: 1602.07604 (cit. on pp. 5, 82).

Kamphuis, P. et al. (2013), *HALOGAS observations of NGC 5023 and UGC 2082: modelling of non-cylindrically symmetric gas distributions in edge-on galaxies*, MNRAS **434** 2069, arXiv: 1306.5312 (cit. on p. 11).

Kaufmann, T. et al. (2006), *Cooling flows within galactic haloes: the kinematics and properties of infalling multiphase gas*, MNRAS **370** 1612, eprint: astro-ph/0507296 (cit. on p. 8).

Kelsall, T. et al. (1998), *The COBE Diffuse Infrared Background Experiment Search for the Cosmic Infrared Background. II. Model of the Interplanetary Dust Cloud*, ApJ **508** 44, eprint: astro-ph/9806250 (cit. on p. 23).

Kerp, J. et al. (1999), *A search for soft X-ray emission associated with prominent high-velocity-cloud complexes*, A&A **342** 213, eprint: arXiv:astro-ph/9810307 (cit. on pp. 12, 28).

Kerp, J. et al. (2011), *The Effelsberg Bonn H I Survey (EBHIS)*, Astronomische Nachrichten **332** 637, arXiv: 1104.1185 [astro-ph.GA] (cit. on pp. 17, 29, 44, 62).

Klein, B. et al. (2006), *The APEX digital Fast Fourier Transform Spectrometer*, A&A **454** L29 (cit. on p. 17).

Knox, L. et al. (2001), *Probing Early Structure Formation with Far-Infrared Background Correlations*, ApJ **550** 7, eprint: astro-ph/0009151 (cit. on p. 60).

Lagache, G., H. Dole and J.-L. Puget (2003), *Modelling infrared galaxy evolution using a phenomenological approach*, MNRAS **338** 555, eprint: astro-ph/0209115 (cit. on p. 60).

Lagache, G., J.-L. Puget and H. Dole (2005), *Dusty Infrared Galaxies: Sources of the Cosmic Infrared Background*, ARA&A **43** 727, eprint: astro-ph/0507298 (cit. on pp. 15, 60, 61).

Lagache, G. et al. (1999), *First detection of the warm ionised medium dust emission. Implication for the cosmic far-infrared background*, A&A **344** 322, eprint: arXiv:astro-ph/9901059 (cit. on p. 13).

Lagache, G. et al. (2000), *Evidence for dust emission in the Warm Ionised Medium sing WHAM data*, A&A **354** 247, eprint: astro-ph/9911355 (cit. on pp. 13, 38).

Langer, W. D. et al. (2014), *A Herschel [C ii] Galactic plane survey. II. CO-dark H_2 in clouds*, A&A **561**, A122 A122, arXiv: 1312.3320 (cit. on p. 84).

Lee, M.-Y. et al. (2014), *The CO-to-H2 Conversion Factor across the Perseus Molecular Cloud*, ApJ **784**, 80 80, arXiv: 1401.5117 (cit. on p. 14).

Lee, M.-Y. et al. (2015), *Cold and Warm Atomic Gas around the Perseus Molecular Cloud. II. The Impact of High Optical Depth on the HI Column Density Distribution and Its Implication for the HI-to-H_2 Transition*, ApJ **809**, 56 56, arXiv: 1504.07405 (cit. on pp. 14, 84, 85).

Leitner, S. N. and A. V. Kravtsov (2011), *Fuel Efficient Galaxies: Sustaining Star Formation with Stellar Mass Loss*, ApJ **734**, 48 48, arXiv: 1011.1252 [astro-ph.GA] (cit. on pp. 10, 27).

Lenz, D., L. Flöer and J. Kerp (2016), *Dust in a compact, cold, high-velocity cloud: A new approach to removing foreground emission*, A&A **586**, A121 A121, arXiv: 1512.08769 (cit. on pp. 9, 14, 16, 43, 57, 65).

Lenz, D. et al. (2015), *Far-infrared excess emission as a tracer of disk-halo interaction*, A&A **573**, A83 A83, arXiv: 1412.2529 (cit. on pp. 9, 13, 14, 16, 27, 42, 82).

Lockman, F. J. (2002), *Discovery of a Population of H I Clouds in the Galactic Halo*, ApJ **580** L47, eprint: astro-ph/0210424 (cit. on p. 17).

– (2003), *High-Velocity Cloud Complex H: A Satellite of the Milky Way in a Retrograde Orbit?*, ApJ **591** L33, eprint: astro-ph/0305408 (cit. on p. 12).

Lockman, F. J. et al. (2008), *The Smith Cloud: A High-Velocity Cloud Colliding with the Milky Way*, ApJ **679** L21, arXiv: 0804.4155 (cit. on pp. 11, 28).

MacRobert, T. and I. Sneddon (1967), *Spherical Harmonics: An Elementary Treatise on Harmonic Functions, with Applications*, International series of monographs in pure and applied mathematics, Pergamon Press, URL: https://books.google.de/books?id=Ja8NAQAAIAAJ (cit. on p. 67).

Madau, P., L. Pozzetti and M. Dickinson (1998), *The Star Formation History of Field Galaxies*, ApJ **498** 106, eprint: astro-ph/9708220 (cit. on p. 15).

Madsen, H. and P. Thyregod (2010), *Introduction to General and Generalized Linear Models*, Chapman & Hall/CRC Texts in Statistical Science, Taylor & Francis, ISBN: 9781420091557, URL: http://books.google.de/books?id=JhuDNgAACAAJ (cit. on pp. 46, 64, 65).

Magnani, L., D. Hartmann and B. G. Speck (1996), *A Catalog of Molecular Gas at High Galactic Latitudes*, ApJS **106** 447 (cit. on pp. 7, 82).

Magnani, L. and A. J. Smith (2010), *Intermediate-velocity Molecular Gas at High Northern Galactic Latitudes*, ApJ **722** 1685 (cit. on pp. 7, 29, 40).

Marasco, A., F. Fraternali and J. J. Binney (2012), *Supernova-driven gas accretion in the Milky Way*, MNRAS **419** 1107, arXiv: 1109.1010 [astro-ph.GA] (cit. on p. 12).

Marasco, A., F. Marinacci and F. Fraternali (2013), *On the origin of the warm-hot absorbers in the Milky Way's halo*, MNRAS **433** 1634, arXiv: 1305.2964 (cit. on pp. 12, 13, 28, 39).

Martin, P. G. et al. (2015), *GHIGLS: H I Mapping at Intermediate Galactic Latitude Using the Green Bank Telescope*, ApJ **809**, 153 153, arXiv: 1504.07723 (cit. on pp. 62, 64, 73, 74).

McClure-Griffiths, N. M. et al. (2009), *Gass: The Parkes Galactic All-Sky Survey. I. Survey Description, Goals, and Initial Data Release*, ApJS **181** 398, arXiv: 0901.1159 [astro-ph.GA] (cit. on pp. 20, 62).

McKee, C. F. and J. P. Ostriker (1977), *A theory of the interstellar medium - Three components regulated by supernova explosions in an inhomogeneous substrate*, ApJ **218** 148 (cit. on pp. 4, 6).

Melioli, C., F. Brighenti and A. D'Ercole (2015), *Galactic fountains and outflows in star-forming dwarf galaxies: interstellar medium expulsion and chemical enrichment*, MNRAS **446** 299, arXiv: 1409.7625 (cit. on p. 12).

Miville-Deschênes, M.-A. and G. Lagache (2005), *IRIS: A New Generation of IRAS Maps*, ApJS **157** 302, eprint: arXiv:astro-ph/0412216 (cit. on pp. 19, 22, 24, 29, 63).

Miville-Deschênes, M.-A. et al. (2005b), *The First Detection of Dust Emission in a High-Velocity Cloud*, ApJ **631** L57, eprint: astro-ph/0508154 (cit. on pp. 9, 14, 44, 45, 55).

Miville-Deschênes, M.-A. et al. (2007), *Statistical properties of dust far-infrared emission*, A&A **469** 595, arXiv: 0704.2175 (cit. on p. 82).

Mizuno, A. and Y. Fukui (2004), "Physical properties of molecular clouds as revealed by NANTEN CO survey: from the galactic center to the galactic warp", *Milky Way Surveys: The Structure and Evolution*

of our Galaxy, ed. by D. Clemens, R. Shah and T. Brainerd, vol. 317, Astronomical Society of the Pacific Conference Series 59 (cit. on p. 82).

Moos, G. et al. (2001), *Anisotropy of Quasiparticle Lifetimes and the Role of Disorder in Graphite from Ultrafast Time-Resolved Photoemission Spectroscopy*, Physical Review Letters **87**.26, 267402 267402 (cit. on p. 5).

Muller, C. A., J. H. Oort and E. Raimond (1963), *Hydrogène neutre dans la couronne galactique?*, Academie des Sciences Paris Comptes Rendus **257** 1661 (cit. on pp. 8, 11, 28, 43).

Narayanan, D. et al. (2012), *A general model for the CO-H_2 conversion factor in galaxies with applications to the star formation law*, MNRAS **421** 3127, arXiv: 1110.3791 [astro-ph.GA] (cit. on p. 38).

Neugebauer, G. et al. (1984), *The Infrared Astronomical Satellite (IRAS) mission*, ApJ **278** L1 (cit. on pp. 13, 24).

Nidever, D. L., S. R. Majewski and W. Butler Burton (2008), *The Origin of the Magellanic Stream and Its Leading Arm*, ApJ **679**, 432-459 432 (cit. on pp. 8, 9).

Nidever, D. L. et al. (2010), *The 200 deg Long Magellanic Stream System*, ApJ **723** 1618, arXiv: 1009.0001 (cit. on p. 9).

Oppermann, N. and T. A. Enßlin (2014), "Bayesian CMB foreground separation with a correlated log-normal model", *Statistical Challenges in 21st Century Cosmology*, ed. by A. Heavens, J.-L. Starck and A. Krone-Martins, vol. 306, IAU Symposium 16, arXiv: 1407.0941 [astro-ph.IM] (cit. on p. 61).

Partridge, R. B. and P. J. E. Peebles (1967), *Are Young Galaxies Visible? II. The Integrated Background*, ApJ **148** 377 (cit. on p. 15).

Patil, A., D. Huard and C. J. Fonnesbeck (2010), *PyMC: Bayesian Stochastic Modelling in Python*, Journal of Statistical Software **35**.4 1, ISSN: 1548-7660, URL: http://www.jstatsoft.org/v35/i04 (cit. on p. 34).

Pearson, K. (1895), *Note on regression and inheritance in the case of two parents*, Proceedings of the Royal Society of London **58**.347-352 240 (cit. on p. 74).

Pedregosa, F. et al. (2011), *Scikit-learn: Machine Learning in Python*, Journal of Machine Learning Research **12** 2825 (cit. on p. 46).

Peek, J. E. G., M. E. Putman and J. Sommer-Larsen (2008), *Ongoing Galactic Accretion: Simulations and Observations of Condensed Gas in Hot Halos*, ApJ **674** 227, arXiv: 0705.0357 (cit. on pp. 8, 28).

Peek, J. E. G. et al. (2007), *Reconstructing Deconstruction: High-Velocity Cloud Distance through Disruption Morphology*, ApJ **656** 907, eprint: astro-ph/0610429 (cit. on pp. 9, 17).

Peek, J. E. G. et al. (2009), *Low-Velocity Halo Clouds*, ApJ **692** 827, arXiv: 0810.4142 (cit. on pp. 9, 37, 44, 55, 65, 79).

Picard, R. R. and R. D. Cook (1984), *Cross-Validation of Regression Models*, Journal of the American Statistical Association **79**.387 575, ISSN: 01621459, URL: http://dx.doi.org/10.2307/2288403 (cit. on pp. 53, 66).

Pietz, J. et al. (1996), *H I Velocity bridges: signature of HVC interaction with the Galactic halo*, A&A **308** L37 (cit. on pp. 9, 12, 28, 32).

Pineda, J. L. et al. (2013), *A Herschel [C ii] Galactic plane survey. I. The global distribution of ISM gas components*, A&A **554**, A103 A103, arXiv: 1304.7770 (cit. on p. 84).

Planck Collaboration (2011 I), *Planck early results. I. The Planck mission*, A&A **536**, A1 A1, arXiv: 1101.2022 [astro-ph.IM] (cit. on p. 22).

– (2011 XIX), *Planck early results. XIX. All-sky temperature and dust optical depth from Planck and IRAS. Constraints on the "dark gas" in our Galaxy*, A&A **536**, A19 A19, arXiv: 1101.2029 [astro-ph.GA] (cit. on pp. 14, 15, 85).

Planck Collaboration (2011 XVIII), *Planck early results. XVIII. The power spectrum of cosmic infrared background anisotropies*, A&A **536**, A18 A18, arXiv: 1101.2028 (cit. on pp. 15, 51, 52, 55, 60, 61, 82).
– (2011 XXIV), *Planck early results. XXIV. Dust in the diffuse interstellar medium and the Galactic halo*, A&A **536**, A24 A24, arXiv: 1101.2036 [astro-ph.GA] (cit. on pp. 9, 12–14, 30, 34, 35, 38, 39, 44, 45, 49, 55, 65, 66).
– (2013 VI), *Planck 2013 results. VI. High Frequency Instrument data processing*, ArXiv e-prints, arXiv: 1303.5067 [astro-ph.CO] (cit. on pp. 19, 29).
– (2013 XI), *Planck 2013 results. XI. All-sky model of thermal dust emission*, ArXiv e-prints, arXiv: 1312.1300 [astro-ph.GA] (cit. on pp. 33, 34, 36).
– (2013 XIII), *Planck 2013 results. XIII. Galactic CO emission*, ArXiv e-prints, arXiv: 1303.5073 [astro-ph.GA] (cit. on p. 82).
– (2014 XII), *Planck 2013 results. XII. Diffuse component separation*, A&A **571**, A12 A12, arXiv: 1303.5072 (cit. on p. 61).
– (2014 XV), *Planck 2013 results. XV. CMB power spectra and likelihood*, A&A **571**, A15 A15, arXiv: 1303.5075 (cit. on p. 23).
– (2014 XVII), *Planck intermediate results. XVII. Emission of dust in the diffuse interstellar medium from the far-infrared to microwave frequencies*, A&A **566**, A55 A55, arXiv: 1312.5446 (cit. on pp. 14, 23, 44, 45, 61).
– (2014 XVIII), *Planck 2013 results. XVIII. The gravitational lensing-infrared background correlation*, A&A **571**, A18 A18, arXiv: 1303.5078 (cit. on pp. 15, 60, 79).
– (2014 XXX), *Planck 2013 results. XXX. Cosmic infrared background measurements and implications for star formation*, A&A **571**, A30 A30, arXiv: 1309.0382 (cit. on pp. 9, 15, 16, 23, 44, 51, 55, 60, 61, 63–65, 67, 73–75, 77, 79, 80, 82, 84, 86, 102, 116).
– (2015 I), *Planck 2015 results. I. Overview of products and scientific results*, ArXiv e-prints, arXiv: 1502.01582 (cit. on p. 63).
– (2015 IX), *Planck 2015 results. IX. Diffuse component separation: CMB maps*, ArXiv e-prints, arXiv: 1502.05956 (cit. on pp. 15, 23, 60, 63, 75, 102).
– (2015 VIII), *Planck 2015 results. VIII. High Frequency Instrument data processing: Calibration and maps*, ArXiv e-prints, arXiv: 1502.01587 (cit. on pp. 23, 44, 45, 63, 70, 79).
– (2015 X), *Planck 2015 results. X. Diffuse component separation: Foreground maps*, ArXiv e-prints, arXiv: 1502.01588 (cit. on pp. 22, 23, 60, 82).
– (2015 XI), *Planck 2015 results. XI. CMB power spectra, likelihoods, and robustness of parameters*, ArXiv e-prints, arXiv: 1507.02704 (cit. on p. 23).
– (2015 XXVI), *Planck 2015 results. XXVI. The Second Planck Catalogue of Compact Sources*, ArXiv e-prints, arXiv: 1507.02058 (cit. on p. 63).
– (2016 XLVIII), *Planck intermediate results. XLVIII. Disentangling Galactic dust emission and cosmic infrared background anisotropies*, ArXiv e-prints, arXiv: 1605.09387 (cit. on pp. 16, 61, 62, 74, 85, 86).
Ponthieu, N., J. Grain and G. Lagache (2011), *POKER: estimating the power spectrum of diffuse emission with complex masks and at high angular resolution*, A&A **535**, A90 A90, arXiv: 1111.0766 [astro-ph.CO] (cit. on pp. 61, 67).
Puget, J.-L. et al. (1996), *Tentative detection of a cosmic far-infrared background with COBE.*, A&A **308** L5 (cit. on p. 36).
Putman, M. E., J. E. G. Peek and M. R. Joung (2012), *Gaseous Galaxy Halos*, ARA&A **50** 491, arXiv: 1207.4837 [astro-ph.GA] (cit. on pp. 7, 8, 11, 27, 28, 43).

Reach, W. T., C. Heiles and J.-P. Bernard (2015), *Variations between Dust and Gas in the Diffuse Interstellar Medium*, ApJ **811**, 118 118, arXiv: 1508.07889 (cit. on p. 85).

Reach, W. T., W. F. Wall and N. Odegard (1998), *Infrared Excess and Molecular Clouds: A Comparison of New Surveys of Far-Infrared and H I 21 Centimeter Emission at High Galactic Latitudes*, ApJ **507** 507, eprint: arXiv:astro-ph/9802169 (cit. on pp. 13, 33, 34).

Reid, M. J. et al. (2014), *Trigonometric Parallaxes of High Mass Star Forming Regions: The Structure and Kinematics of the Milky Way*, ApJ **783**, 130 130, arXiv: 1401.5377 [astro-ph.GA] (cit. on p. 7).

Remazeilles, M. et al. (2016), *Sensitivity and foreground modelling for large-scale cosmic microwave background B-mode polarization satellite missions*, MNRAS **458** 2032, arXiv: 1509.04714 (cit. on pp. 60, 85).

Reynolds, R. J. (1991), "Ionized disk/halo gas - Insight from optical emission lines and pulsar dispersion measures", *The Interstellar Disk-Halo Connection in Galaxies*, ed. by H. Bloemen, vol. 144, IAU Symposium 67 (cit. on p. 9).

Richter, P. (2012), *Cold Gas Accretion by High-velocity Clouds and Their Connection to QSO Absorption-line Systems*, ApJ **750**, 165 165, arXiv: 1111.1717 [astro-ph.CO] (cit. on pp. 29, 40).

Richter, P. et al. (2001), *The Diversity of High- and Intermediate-Velocity Clouds: Complex C versus IV Arch*, ApJ **559** 318, eprint: arXiv:astro-ph/0105466 (cit. on pp. 7, 9, 11, 28).

Richter, P. et al. (2003), *A Far Ultraviolet Spectroscopic Explorer Survey of Molecular Hydrogen in Intermediate-Velocity Clouds in the Milky Way Halo*, ApJ **586** 230, eprint: arXiv:astro-ph/0211356 (cit. on p. 7).

Röhser, T. (2016), *The Physical Properties of Intermediate-Velocity Gas at High Galactic Latitudes*, Promotion: Rheinische Friedrich-Wilhelms Universität Bonn (cit. on p. 42).

Röhser, T. et al. (2014), *A dynamical transition from atomic to molecular intermediate-velocity clouds*, A&A **564**, A71 A71, arXiv: 1402.6850 [astro-ph.GA] (cit. on pp. 13, 40, 44, 56).

Röhser, T. et al. (2016, submitted), *An all-sky census of Galactic high-latitude molecular intermediate-velocity clouds*, A&A (cit. on p. 7).

Sancisi, R. et al. (2008), *Cold gas accretion in galaxies*, A&A Rev. **15** 189, arXiv: 0803.0109 (cit. on p. 28).

Saul, D. R., J. E. G. Peek and M. E. Putman (2014), *Dust-to-gas ratios of the GALFA-H I Compact Cloud Catalog*, MNRAS **441** 2266, arXiv: 1403.4617 [astro-ph.GA] (cit. on pp. 9, 13, 14, 44, 55).

Schiminovich, D. et al. (1997), "Imaging and Kinematics of Neutral Hydrogen in and around "Shell Galaxies"", *The Nature of Elliptical Galaxies; 2nd Stromlo Symposium*, ed. by M. Arnaboldi, G. S. Da Costa and P. Saha, vol. 116, Astronomical Society of the Pacific Conference Series 362 (cit. on p. 32).

Schönrich, R. and J. Binney (2009), *Chemical evolution with radial mixing*, MNRAS **396** 203, arXiv: 0809.3006 (cit. on pp. 10, 28).

Schwarz, U. J., B. P. Wakker and H. van Woerden (1995), *Distance and metallicity limits of high-velocity clouds.*, A&A **302** 364 (cit. on p. 40).

Serra, P. et al. (2014), *Cross-correlation of cosmic far-infrared background anisotropies with large scale structures*, A&A **570**, A98 A98, arXiv: 1404.1933 (cit. on pp. 60, 67, 71, 82).

Shannon, C. E. (1949), *Communication in the Presence of Noise*, IEEE Proceedings **37** 10 (cit. on p. 20).

Shapiro, P. R. and G. B. Field (1976), *Consequences of a New Hot Component of the Interstellar Medium*, ApJ **205** 762 (cit. on p. 28).

Stanimirović, S. et al. (2006), *First Results from the Arecibo Galactic H I Survey: The Disk/Halo Interface Region in the Outer Galaxy*, ApJ **653** 1210, eprint: astro-ph/0609137 (cit. on pp. 9, 17).

Stanimirović, S. et al. (2008), *The Many Streams of the Magellanic Stream*, ApJ **680**, 276-286 276, arXiv: 0802.1349 (cit. on p. 9).

Stanko, S., B. Klein and J. Kerp (2005), *A field programmable gate array spectrometer for radio astronomy. First light at the Effelsberg 100-m telescope*, A&A **436** 391 (cit. on p. 17).

Sunyaev, R. A. and Y. B. Zeldovich (1970), *Small-Scale Fluctuations of Relic Radiation*, Ap&SS **7** 3 (cit. on pp. 15, 22, 63).

Tang, N. et al. (2016), *Physical properties of CO-dark molecular gas traced by C^+*, ArXiv e-prints, arXiv: 1606.00949 (cit. on p. 84).

Tauber, J. A. et al. (2010), *Planck pre-launch status: The Planck mission*, A&A **520**, A1 A1 (cit. on p. 22).

The HDF Group (1997-NNNN), *Hierarchical Data Format, version 5*, http://www.hdfgroup.org/HDF5/ (cit. on p. 64).

Thom, C. et al. (2008), *An Accurate Distance to High-Velocity Cloud Complex C*, ApJ **684** 364, arXiv: 0712.0612 (cit. on pp. 9, 11, 29, 40, 41).

Tibshirani, R. (1996), *Regression Shrinkage and Selection via the Lasso*, English, Journal of the Royal Statistical Society. Series B (Methodological) **58**.1, ISSN: 00359246, URL: http://www.jstor.org/stable/2346178 (cit. on pp. 46, 66).

van de Hulst, H. C. (1945), *Radio Waves from Space: Origin of Radiowaves*, Ned. tijd. natuurkunde **11** 210 (cit. on p. 3).

van den Bergh, S. (1962), *The frequency of stars with different metal abundances.*, AJ **67** 486 (cit. on p. 10).

Venzmer, M. S., J. Kerp and P. M. W. Kalberla (2012), *The four leading arms of the Magellanic Cloud system. Evidence for interaction with Milky Way disk and halo*, A&A **547**, A12 A12 (cit. on pp. 9, 12).

Voges, W. et al. (1999), *The ROSAT all-sky survey bright source catalogue*, A&A **349** 389, eprint: astro-ph/9909315 (cit. on p. 28).

Wakker, B. P. (2001), *Distances and Metallicities of High- and Intermediate-Velocity Clouds*, ApJS **136** 463, eprint: arXiv:astro-ph/0102147 (cit. on pp. 7, 9, 29, 43, 46, 55).

Wakker, B. P. and F. Boulanger (1986), *A search for dust in high-velocity clouds*, A&A **170** 84 (cit. on pp. 14, 43, 55).

Wakker, B. P. and H. van Woerden (1997), *High-Velocity Clouds*, ARA&A **35** 217 (cit. on pp. 8, 12, 28, 29, 43).

Wakker, B. P. et al. (1999), *Accretion of low-metallicity gas by the Milky Way*, Nature **402** 388 (cit. on pp. 8, 9).

Wakker, B. P. et al. (2007), *Distances to Galactic High-Velocity Clouds: Complex C*, ApJ **670** L113, arXiv: 0710.3340 (cit. on pp. 9, 40).

Wakker, B. P. et al. (2008), *Distances to Galactic High-Velocity Clouds. I. Cohen Stream, Complex GCP, Cloud g1*, ApJ **672** 298, arXiv: 0709.1926 (cit. on pp. 7, 9, 11).

Wandelt, B. and E. Hivon (1999), "CMB anisotropy power spectrum statistics", *Evolution of Large Scale Structure : From Recombination to Garching*, ed. by A. J. Banday, R. K. Sheth and L. N. da Costa 43, eprint: astro-ph/9810248 (cit. on p. 69).

Weiß, A. et al. (1999), *A molecular cloud forming in the disk-halo interface*, A&A **344** 955 (cit. on pp. 33, 38–40, 82).

Williams, R. J. et al. (2012), *A Spitzer-MIPS Search for Dust in Compact High-velocity H I Clouds*, AJ **143**, 82 82, arXiv: 1108.5181 [astro-ph.GA] (cit. on p. 9).

Winkel, B., L. Flöer and A. Kraus (2012), *Efficient least-squares basket-weaving*, A&A **547**, A119 A119, arXiv: 1210.4005 [astro-ph.IM] (cit. on p. 20).

Winkel, B., A. Kraus and U. Bach (2012), *Unbiased flux calibration methods for spectral-line radio observations*, A&A **540**, A140 A140, arXiv: 1203.0741 [astro-ph.IM] (cit. on p. 20).

Winkel, B., D. Lenz and L. Flöer (2016b), *Cygrid: A fast Cython-powered convolution-based gridding module for Python*, ArXiv e-prints, arXiv: 1604.06667 [astro-ph.IM] (cit. on pp. 20, 24, 64).

Winkel, B. et al. (2010), *The Effelsberg-Bonn H I Survey: Data Reduction*, ApJS **188** 488, arXiv: 1005.4604 [astro-ph.IM] (cit. on pp. 17, 20, 29, 44, 62).

Winkel, B. et al. (2011), *The high-velocity cloud complex Galactic center negative as seen by EBHIS and GASS. I. Cloud catalog and global properties*, A&A **533**, A105 A105, arXiv: 1107.3399 [astro-ph.GA] (cit. on p. 9).

Winkel, B. et al. (2016a), *The Effelsberg-Bonn H i Survey: Milky Way gas. First data release*, A&A **585**, A41 A41, arXiv: 1512.05348 [astro-ph.IM] (cit. on pp. 17–20, 44, 62).

Wolfire, M. G., D. Hollenbach and C. F. McKee (2010), *The Dark Molecular Gas*, ApJ **716** 1191, arXiv: 1004.5401 [astro-ph.GA] (cit. on p. 14).

Wolfire, M. G. et al. (1995), *The Multiphase Structure of the Galactic Halo: High-Velocity Clouds in a Hot Corona*, ApJ **453** 673 (cit. on pp. 4–6, 9).

– (2003), *Neutral Atomic Phases of the Interstellar Medium in the Galaxy*, ApJ **587** 278, eprint: astro-ph/0207098 (cit. on p. 4).

APPENDIX A

CIB results on individual fields

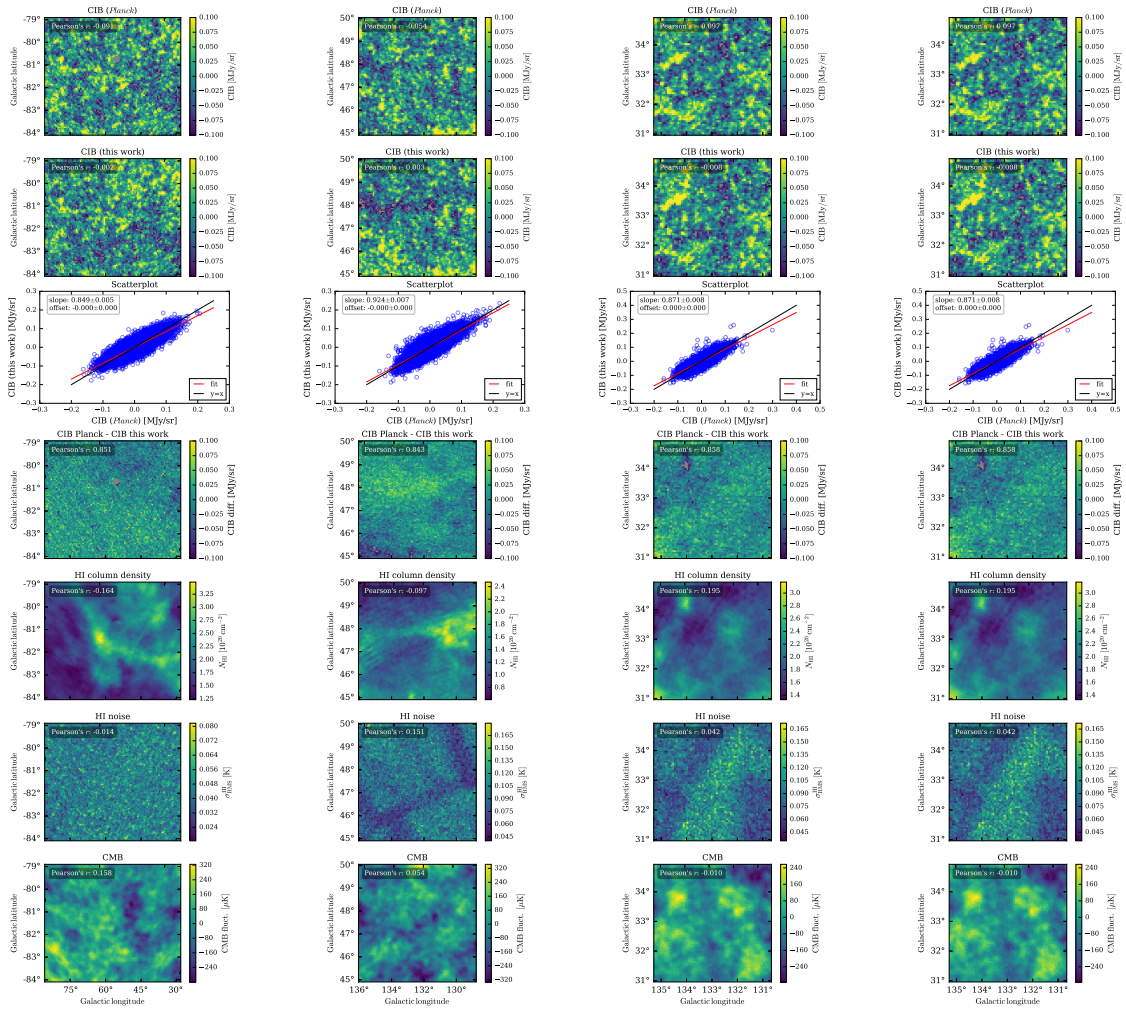


Figure A.1: Comparison of CIB reconstruction at 545 GHz between Planck Collaboration (2014 XXX) and this work for the fields MC (**column 1**), SP (**column 2**), SPC5 (**column 3**), and SPC5 (**column 4**). Row description from top to bottom: **(1)** CIB image derived by Planck Collaboration (2014 XXX), Pearson's r w.r.t to $N_{\mathrm{H\,I}}$. **(2)** CIB image derived in this work, using a simple linear relation between FIR intensity and H I column density. Pearson's r w.r.t to $N_{\mathrm{H\,I}}$. **(3)** Scatterplot between individual CIB values found by Planck Collaboration and this work. **(4)** Difference map between the the *Planck* results and this work. The Pearson's r indicates the similarity of the two CIB images (rows 1 and 2). **(5)** Total EBHIS H I column density. **(6)** EBHIS noise map. **(7)** CMB map, taken from Planck Collaboration (2015 IX). For the rows (5) to (7), the Pearons's r indicates the correlation between the individual map and the CIB difference map (row 4).

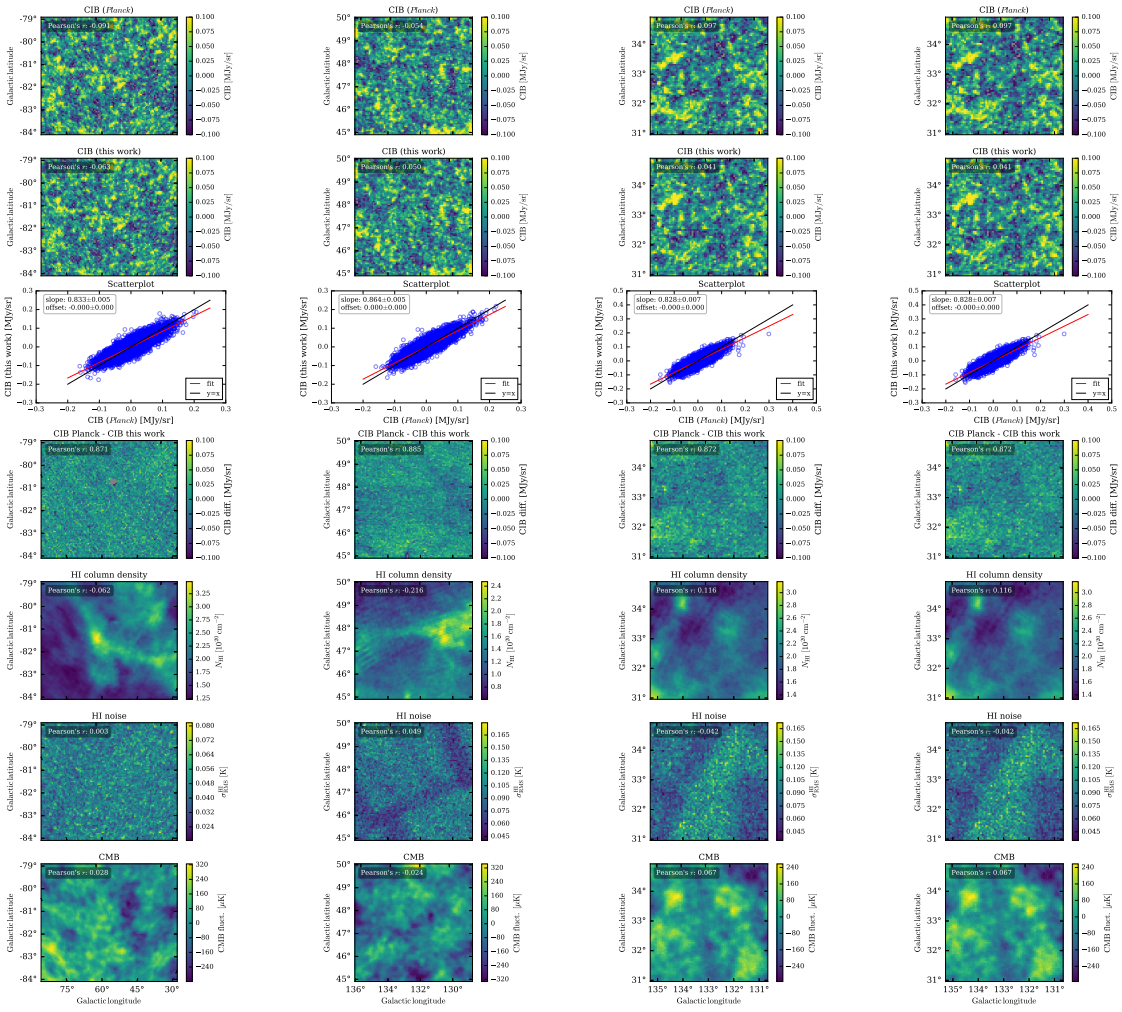


Figure A.2: Identical to Fig. A.1, but now using the GLM to remove the foreground.

APPENDIX B

Large scale CIB results

B.1 Tables of power spectra

l	C_b [Jy^2/sr]	ΔC_b [Jy^2/sr]
3.15e+01	9.97e+05	9.85e+05
9.55e+01	6.22e+03	7.41e+03
1.60e+02	8.19e+03	8.60e+02
2.24e+02	7.06e+03	3.16e+02
2.88e+02	5.02e+03	2.13e+01
3.52e+02	4.25e+03	5.47e+02
4.16e+02	3.82e+03	9.26e+02
4.80e+02	2.85e+03	4.50e+02
5.44e+02	2.39e+03	4.57e+02
6.08e+02	2.10e+03	5.17e+02
6.72e+02	1.92e+03	5.75e+02
7.36e+02	1.64e+03	4.91e+02
8.00e+02	1.43e+03	4.43e+02
8.64e+02	1.24e+03	3.91e+02
9.28e+02	1.29e+03	5.38e+02
9.92e+02	1.24e+03	5.57e+02

Table B.1: Unbiased, binned power spectrum for the 353 GHz data. l corresponds to the centre value of each l-bin and ΔC_b indicates 1-σ uncertainties.

l	C_b [Jy^2/sr]	ΔC_b [Jy^2/sr]
3.15e+01	9.20e+06	9.10e+06
9.55e+01	5.11e+04	5.27e+04
1.60e+02	5.92e+04	9.42e+03
2.24e+02	5.00e+04	1.51e+03
2.88e+02	3.38e+04	4.36e+03
3.52e+02	2.82e+04	1.19e+02
4.16e+02	2.48e+04	2.60e+03
4.80e+02	1.91e+04	8.73e+02
5.44e+02	1.58e+04	1.08e+03
6.08e+02	1.34e+04	1.35e+03
6.72e+02	1.23e+04	2.07e+03
7.36e+02	1.02e+04	1.43e+03
8.00e+02	9.00e+03	1.49e+03
8.64e+02	7.98e+03	1.48e+03
9.28e+02	8.07e+03	2.37e+03
9.92e+02	8.00e+03	2.78e+03

Table B.2: Unbiased, binned power spectrum for the 545 GHz data. l corresponds to the centre value of each l-bin and ΔC_b indicates 1-σ uncertainties.

l	C_b [Jy^2/sr]	ΔC_b [Jy^2/sr]
3.15e+01	6.62e+07	6.58e+07
9.55e+01	2.32e+05	1.52e+05
1.60e+02	2.44e+05	1.10e+04
2.24e+02	1.93e+05	2.13e+03
2.88e+02	1.26e+05	1.71e+04
3.52e+02	1.06e+05	1.58e+03
4.16e+02	9.39e+04	1.17e+04
4.80e+02	6.94e+04	1.80e+03
5.44e+02	6.03e+04	5.72e+03
6.08e+02	5.20e+04	7.20e+03
6.72e+02	4.51e+04	7.31e+03
7.36e+02	3.93e+04	6.82e+03
8.00e+02	3.60e+04	8.15e+03
8.64e+02	3.08e+04	6.63e+03
9.28e+02	3.33e+04	1.21e+04
9.92e+02	3.54e+04	1.58e+04

Table B.3: Unbiased, binned power spectrum for the 857 GHz data. l corresponds to the centre value of each l-bin and ΔC_b indicates 1-σ uncertainties.

B.2 Large-scale images

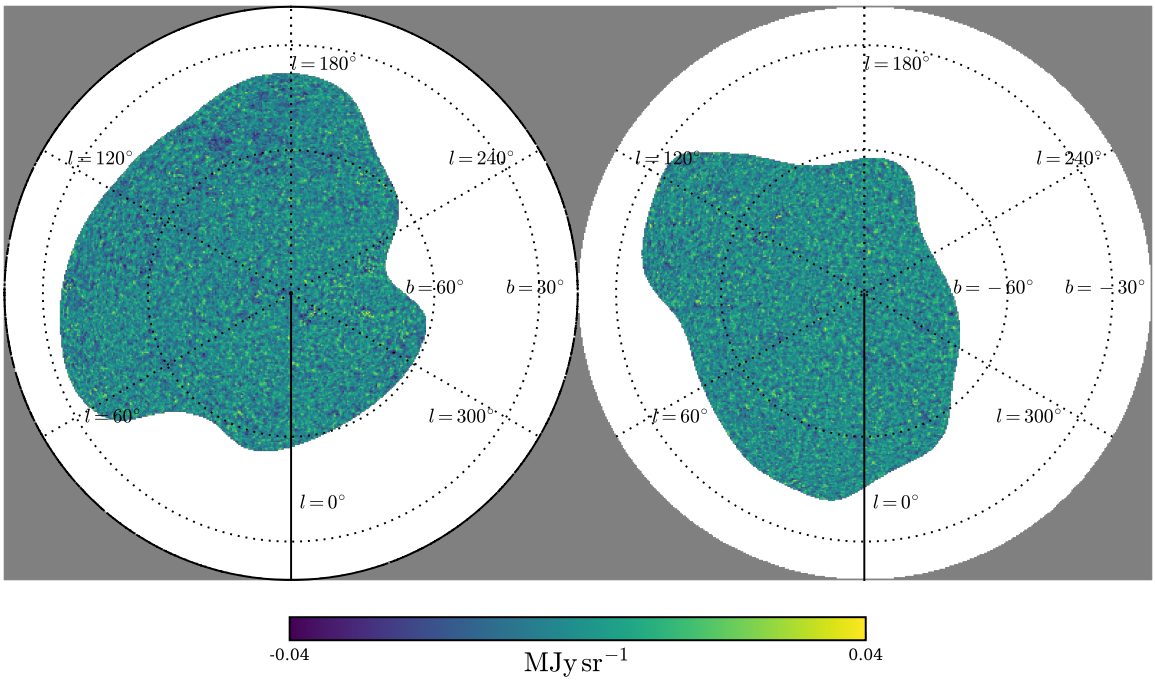


Figure B.1: Large-scale CIB in zenith equal-area projection for the 353 GHz data. The thick lines indicates l = 0, with l increasing clockwise. The Galactic poles are in the centre of the image (northern Galactic pole left, southern Galactic pole right). Concentric circles correspond to $|b|$ = 30° and 60°.

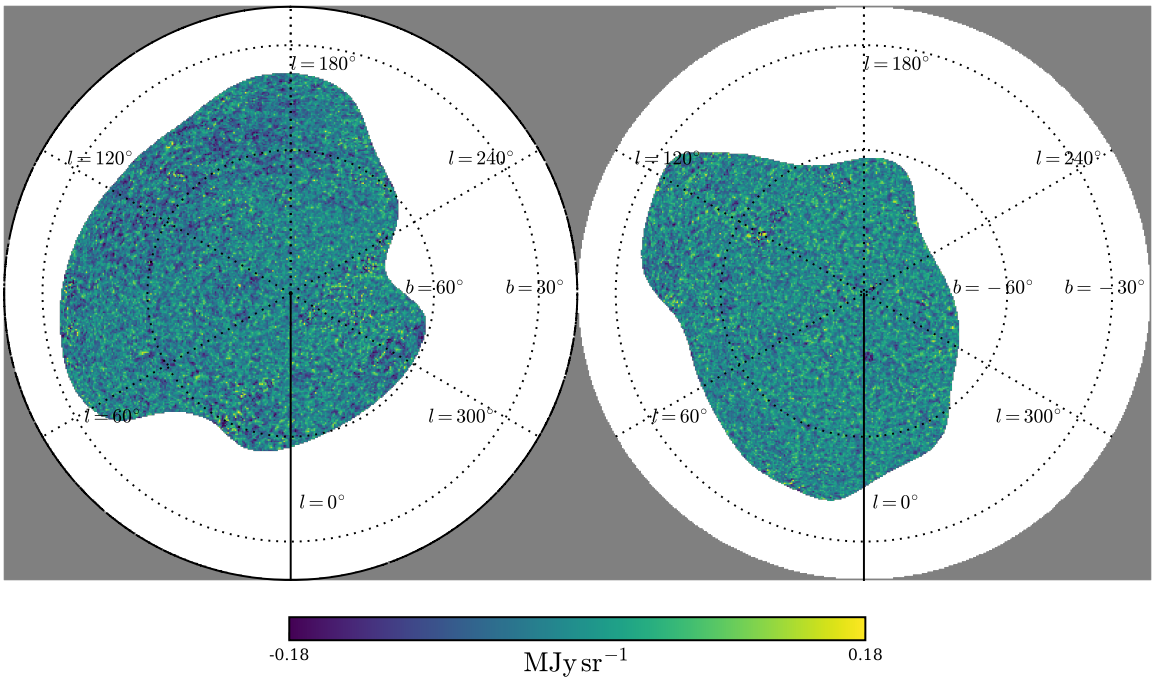
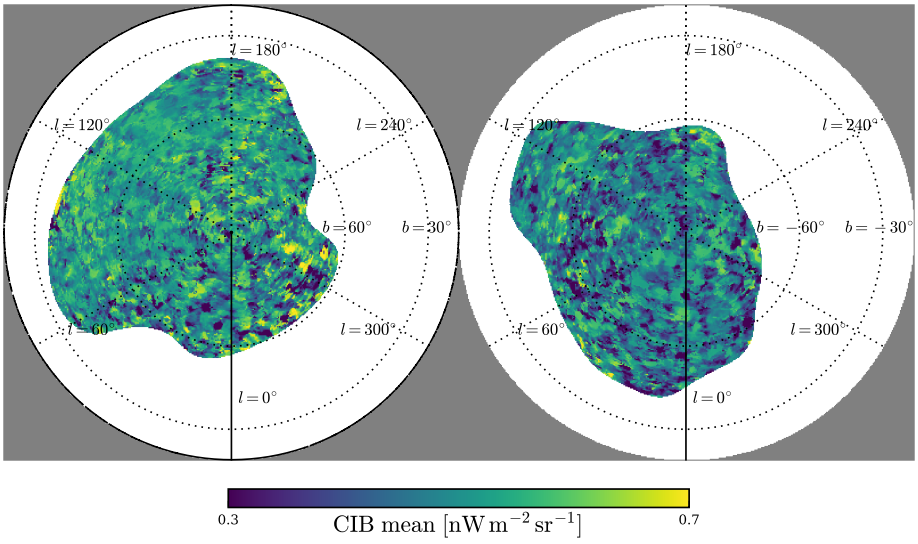
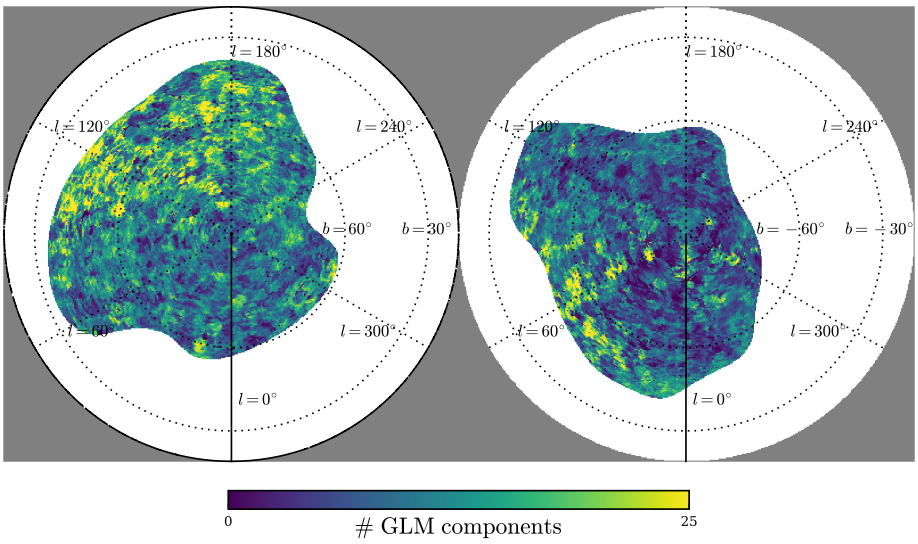


Figure B.2: Large-scale CIB in zenith equal-area projection for the 857 GHz data. The thick lines indicates $l = 0$, with l increasing clockwise. The Galactic poles are in the centre of the image (northern Galactic pole left, southern Galactic pole right). Concentric circles correspond to $|b| = 30°$ and $60°$.

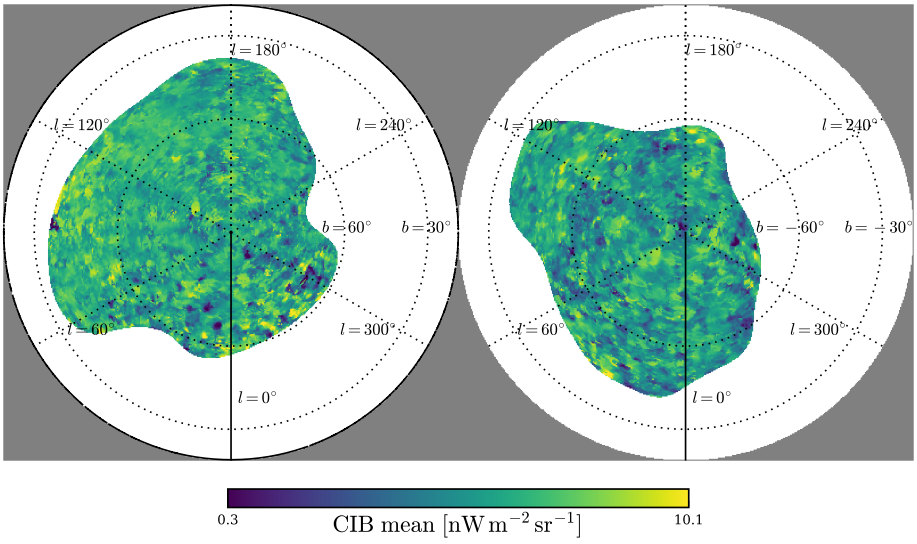


(a)

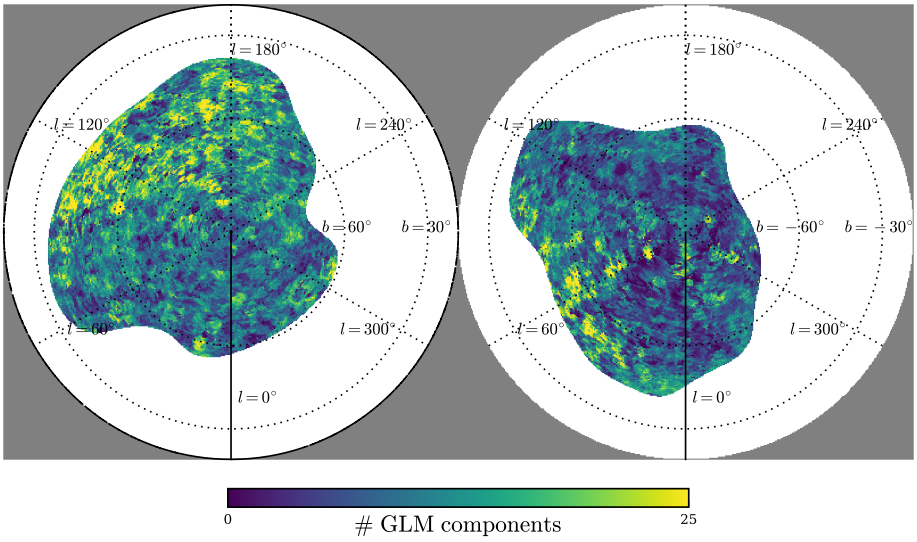


(b)

Figure B.3: **(a)** FIR offset at 353 GHz, derived with the model. While this offset is dominated by the CIB emission, other Galactic components additionally increase or decrease this quantity. **(b)** Number of GLM coefficients for each line of sight at 353 GHz. As expected, the number of components required to accurately describe the correlation between dust and gas increases towards the Galactic plane.



(a)



(b)

Figure B.4: **(a)** FIR offset at 857 GHz, derived with the model. While this offset is dominated by the CIB emission, other Galactic components additionally increase or decrease this quantity. **(b)** Number of GLM coefficients for each line of sight at 857 GHz. As expected, the number of components required to accurately describe the correlation between dust and gas increases towards the Galactic plane.

B.3 Scatterplots against uncorrelated components

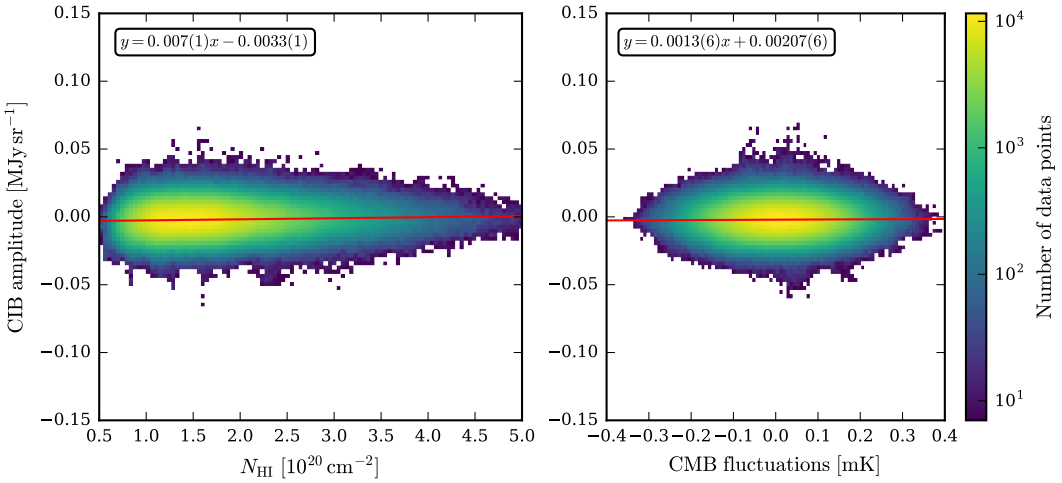


Figure B.5: **Left:** Scatterplot of total H I column density against the derived CIB at 353 GHz. **Right:** Scatterplot of CMB fluctuations against the derived CIB at 353 GHz. For both panels, the red line shows a linear fit to the data points. Fit parameters are shown in each panel. The entire sky coverage of 20% (see also Fig. 5.16) is used for this comparison.

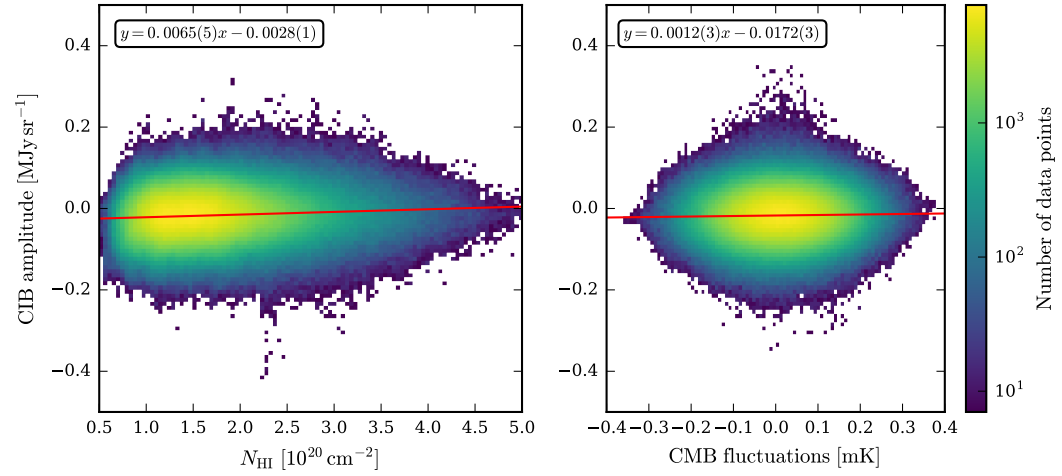


Figure B.6: **Left:** Scatterplot of total H I column density against the derived CIB at 857 GHz. **Right:** Scatterplot of CMB fluctuations against the derived CIB at 857 GHz. For both panels, the red line shows a linear fit to the data points. Fit parameters are shown in each panel. The entire sky coverage of 20% (see also Fig. 5.16) is used for this comparison.

B.4 Angular power spectra

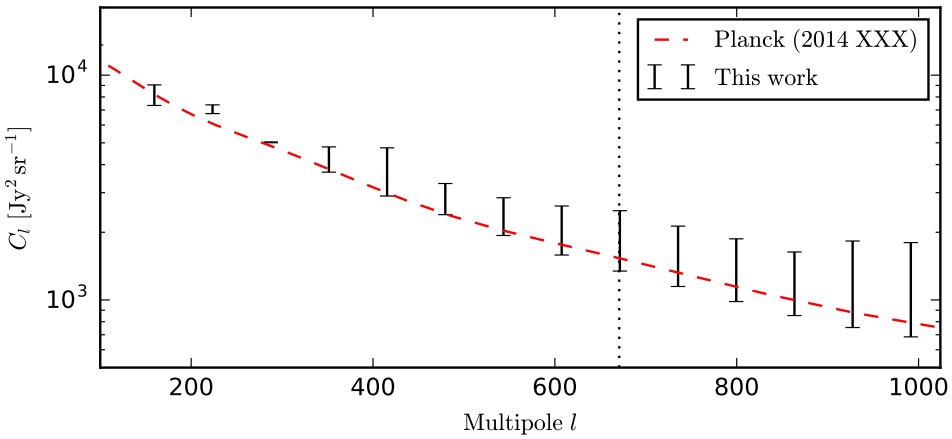


Figure B.7: Unbiased, binned CIB power spectrum at 353 GHz for 20% of the sky. The black error bars indicate the 1σ uncertainties, derived with 200 Monte Carlo estimates. The red line corresponds to the results from Planck Collaboration (2014 XXX) and the dashed red line indicates the approximate resolution of the data sets used in the present study. Due to the very complex and different approach to quantify the uncertainties in the Planck Collaboration (2014 XXX) results, we abstain from showing them as the two estimates for the errors would not be comparable.

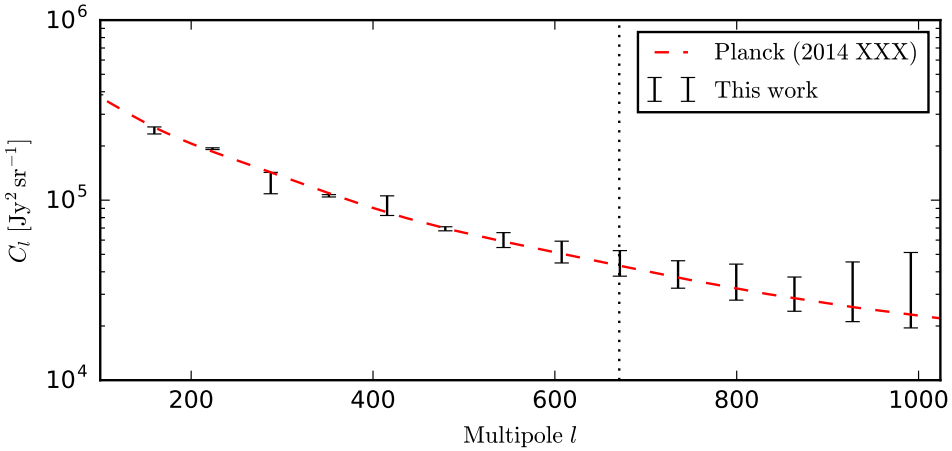


Figure B.8: Unbiased, binned CIB power spectrum at 857 GHz for 20% of the sky. The black error bars indicate the 1σ uncertainties, derived with 200 Monte Carlo estimates. The red line corresponds to the results from Planck Collaboration (2014 XXX) and the dashed red line indicates the approximate resolution of the data sets used in the present study. Due to the very complex and different approach to quantify the uncertainties in the Planck Collaboration (2014 XXX) results, we abstain from showing them as the two estimates for the errors would not be comparable.

APPENDIX C

Estimators in GLM reconstruction

We present the evaluation of the simulations for the spiky and flat GLM coefficients (Fig. C.1 and C.2, respectively). See Sect. 4.4.2 for a detailed description of the results and their implication.

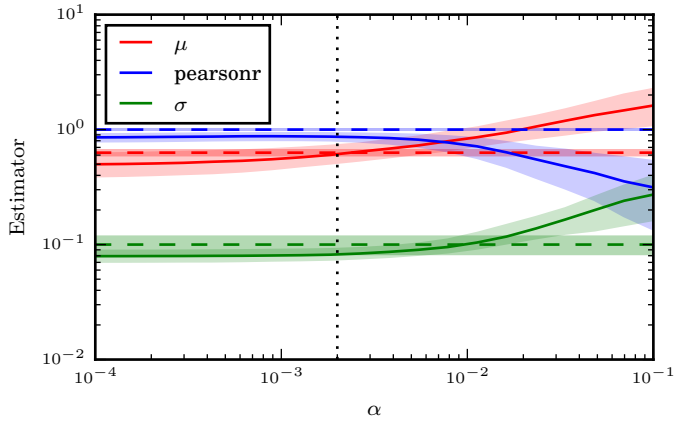


Figure C.1: Evaluation of the reconstructed CIB mean μ, CIB fluctuation amplitude σ and Pearson's r of input CIB image and reconstruction. This is based on the spiky input GLM coefficients (second panel from the top in Fig. 4.7). Dashed lines indicate the input, solid lines the reconstructed quantities. Contours correspond to 1σ uncertainties.

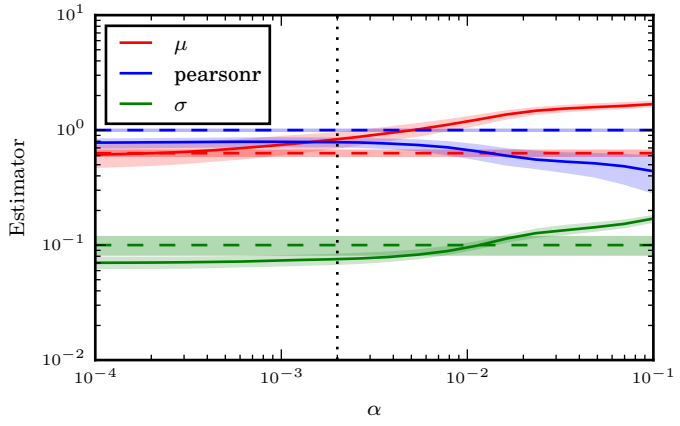


Figure C.2: Same as Fig. C.1, but for a flat spectrum of GLM coefficients (bottom panel in Fig. 4.7).

List of Figures

List of Tables

Acronyms

AGN Active Galactic Nuclei.

AIfA Argelander Institute for Astronomy.

API Application programming interface.

CIB Cosmic infrared background.

CMB Cosmic microwave background.

CNM Cold neutral medium.

COBE Cosmic Background Explorer.

DIRBE Diffuse Infrared Background Experiment.

DM Dark matter.

EBHIS Effelsberg-Bonn H I Survey.

EBL Extragalactic background light.

FIR Far-infrared.

FUSE Far Ultraviolet Spectroscopic Explorer.

FWHM Full width at half-maximum.

GALFA-HI Galactic Arecibo L-band Feed Array HI Survey.

GASS Galactic All Sky Survey.

GBT Green Bank Telescope.

GLM Generalised linear model.

HALOGAS Hydrogen Accretion in LOcal GAlaxieS.

HDF5 Hierarchical Data Format 5.

HEALPix Hierarchical Equal Area isoLatitude Pixelation of a sphere.

HFI High Frequency Instrument aboard the *Planck* satellite.

HIM Hot ionised medium.

HT Head-tail structure.

HVC High-velocity cloud.

IFT Information field theory.

IGM Intergalactic medium.

IR Infrared.

IRAS Infrared Astronomical Satellite.

IRIS Improved Reprocessing of the IRAS Survey.

ISM Interstellar Medium.

ISRF Interstellar radiation field.

IVC Intermediate-velocity cloud.

IVMC Intermediate-velocity molecular cloud.

LAB Leiden/Argentine/Bonn Survey.

LFI Low Frequency Instrument aboard the *Planck* satellite.

LMC Large Magellanic Cloud.

LSR Local standard of rest.

MASTER Monte Carlo Apodized Spherical Transform Estimator.

MCMC Markov chain Monte Carlo.

MM Molecular medium.

MW Milky Way Galaxy.

PDR Photodissociation region.

POKER P of k EstimatoR.

RFI Radio frequency interference.

RMS Root mean square.

RSS Residiual sum of squares.

SED Spectral energy distribution.

SFH Star formation history.

SFR Star formation rate.

SFRD Star formation rate density.

SMICA Spectral Matching Independent Component Analysis.

SNR Signal-to-noise ratio.

SR Stray radiation.

SZ Sunyaev-Zeldovich effect.

VB Velocity bridges.

WIM Warm ionised medium.

WMAP Wilkinson Microwave Anisotropy Probe.

WNM Warm neutral medium.

WSRT Westerbork-Synthesis Radio Telescope.