



Far-ultraviolet Dust Extinction and Molecular Hydrogen in the Diffuse Milky Way Interstellar Medium

Dries Van De Putte¹ , Stefan I. B. Cartledge², Karl D. Gordon^{1,3} , Geoffrey C. Clayton⁴ , and Julia Roman-Duval¹

¹Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD, 21218, USA; dvandeputte@stsci.edu

²Dept. of Physical Sciences, MacEwan University, 10700-104 Avenue, Edmonton, AB T5J 4S2, Canada

³Sterrenkundig Observatorium, Universiteit Gent, Gent, Belgium

⁴Department of Physics & Astronomy, Louisiana State University, Baton Rouge, LA 70803, USA

Received 2022 July 15; revised 2022 September 23; accepted 2022 October 9; published 2023 February 9

Abstract

We aim to compare variations in the full-UV dust extinction curve (912–3000 Å), with the H I/H₂/total H content along diffuse Milky Way sightlines, to investigate possible connections between ISM conditions and dust properties. We combine an existing sample of 75 UV extinction curves based on IUE and FUSE data, with atomic and molecular column densities measured through UV absorption. The H₂ column density data are based on existing Lyman–Werner absorption band models from earlier work on the extinction curves. Literature values for the H I column density were compiled, and improved for 23 stars by fitting a Ly α profile to archived spectra. We discover a strong correlation between the H₂ column and the far-UV extinction, and the underlying cause is a linear relationship between H₂ and the strength of the far-UV rise feature. This extinction does not scale with H I, and the total H column scales best with $A(V)$ instead. The carrier of the far-UV rise therefore coincides with molecular gas, and further connections are shown by comparing the UV extinction features to the molecular fraction. Variations in the gas-to-extinction ratio $N(\text{H})/A(V)$ correlate with the UV-to-optical extinction ratio, and we speculate this could be due to coagulation or shattering effects. Based on the H₂ temperature, the strongest far-UV rise strengths are found to appear in colder and denser sightlines.

Unified Astronomy Thesaurus concepts: [Interstellar dust extinction \(837\)](#); [Interstellar absorption \(831\)](#); [Diffuse molecular clouds \(381\)](#); [Ultraviolet spectroscopy \(2284\)](#); [Interstellar medium \(847\)](#); [Neutral hydrogen clouds \(1099\)](#); [Dust composition \(2271\)](#)

Supporting material: machine-readable tables

1. Introduction

The presence of dust grains in the interstellar medium (ISM) has several effects, through which the state of the medium is influenced. Dust provides shielding by absorbing the UV photons in the Lyman and Werner bands (912–1100 Å), which dissociate H₂ (Sternberg et al. 2014). If there are more small grains that absorb strongly in the far-UV (FUV), then this shielding will be more efficient per dust mass unit. Dust also facilitates the formation of H₂, on the surfaces of grains (Hollenbach & Salpeter 1971; Wakelam et al. 2017). The H₂ formation rate is proportional to the available dust surface area, which is why this process is more efficient when a larger fraction of the dust content consists of smaller grains because of their higher surface area-to-volume ratio (Snow 1983). Because both the shielding and the formation efficiency of H₂ are enhanced by the presence of small grains, and the extinction curve of dust in the UV depends on the grain size distribution, there could be an observable connection between UV dust extinction and the H I/H₂ balance of the gas.

Extinction curves are the key observable providing information about the optical properties and size distribution of dust. Processes such as shattering, coagulation, and accretion can change the size distribution of dust grains, which in turn results in a variety of extinction curve shapes, as demonstrated in dust evolution models

of galaxies (Ormel et al. 2009; Hirashita 2012; Asano et al. 2014; Hirashita & Aoyama 2019; Aoyama et al. 2020). One of the key questions in the study of dust extinction curves, is the origin of two prominent UV features: a broad bump at 2175 Å in the near-UV (NUV bump), and a steepening of the curve at FUV wavelengths (FUV rise), which continues all the way to 912 Å (Fitzpatrick & Massa 1986, 1988). The strength and rate of occurrence of these features has been observed to be different between the Milky Way, Large Magellanic Cloud (LMC), and Small Magellanic Cloud (SMC; Gordon et al. 2003).

When the average grain size increases, the extinction curve becomes flatter toward shorter wavelengths. A commonly used quantity to measure the slope of the extinction curve, and make conclusions about changes in the average grain size, is the total-to-selective extinction ratio $R(V) = A(V)/E(B - V)$. The relationship between characteristics of the size distribution of dust models and the corresponding $R(V)$ is demonstrated in, e.g., Siebenmorgen et al. (2018). Observations in the mid-IR of individual molecular clouds (e.g., Chapman et al. 2009; Steinacker et al. 2010) find that $R(V)$ is greater in the inner regions with high extinction, suggesting that larger grains are formed more efficiently in the colder and denser parts of clouds. UV extinction curve observations of the Taurus molecular cloud TMC-1 show that the 2175 Å extinction bump is prominent in diffuse regions, but becomes minimal inside the dense clump (Whittet et al. 2004). The sightlines studied in this work are diffuse, with $A(V) < 3$, which means that are not observing deep into such clouds, but rather the edges, where photodissociation regions (PDRs) occur (Hollenbach & Tielens 1999). In



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

earlier work (Van De Putte et al. 2020), we measured the extinction in different locations of the IC 63 PDR using broadband photometry of its background stars, and found that $R(V)$ is larger near the illuminated edge. It is unclear if the observed $R(V)$ variations in this PDR originate from accretion processes similar to those assumed for dense molecular clouds. At larger scales, for a combined sample LMC and SMC sightlines, correlations between the gas-to-extinction ratio and the steepness of the curve were found (Gordon et al. 2003, Figure 9). In this work we aim to examine if similar dust evolution effects occur in the diffuse Milky Way ISM, by investigating $R(V)$ and other extinction curve shape parameters.

In addition to tracing the grain size with $R(V)$, the extinction curve also contains information about the grain composition. While it is still unclear what the exact carriers of the UV bump and rise features are, most models assume that they are associated with carbonaceous grains, as opposed to silicate grains. For example, in recent work by Zuo et al. (2021), models based on Weingartner & Draine (2001) are fitted to a sample of extinction curves that partially overlaps with the one used in this work. The silicate component of their model has a flatter extinction curve, while the carbonaceous component has the UV features. Under those assumptions, the strength of the bump and rise determine the amount of carbon in the dust. In the SMC on the other hand, the bump is generally weak, while the FUV rise is still strong. Depletion measurements have shown that some SMC dust has very little Si (Tchernyshyov et al. 2015), and that the relative amount of dust consisting of polycyclic aromatic hydrocarbons (PAHs) is significantly lower than that in the Milky Way (Jenkins & Wallerstein 2017). This indicates that the grains causing the bump and rise exhibit different behaviors depending on the environment, despite the fact that both features are thought to be carbonaceous in nature. A more subtle clue about the nature of the grains is the bump width, which has been found to vary depending on the environment, being narrower in the diffuse medium and broader in dark clouds. These changes are typically attributed to coatings forming on the grains (Cardelli & Clayton 1991; Valencic et al. 2004). By comparing the UV extinction features and the H_2 content of the gas, we aim to learn about the relationship between carbonaceous grains and molecular gas.

A secondary goal is to search for direct evidence for dust growth in the ISM, a process that is found necessary to explain the evolution of the dust content of galaxies (Zhukovska et al. 2008; Draine 2009; Boyer et al. 2012; Rowlands et al. 2014; Zhukovska et al. 2016). The discrepancy between the observed amount of dust, and the dust production and destruction rates, can be solved if dust grains grow in mass and size by accreting gas-phase metals onto their surfaces (McKinnon et al. 2016). Direct evidence for dust growth, i.e., atoms moving from the gas to the solid phase, has been found via depletion measurements. This type of measurement consists of determining which fraction of each metal is locked up in dust grains, based on measurements of the gas-phase abundances of those metals (Savage & Sembach 1996). The degree of depletion across different metals is mostly tied to a single parameter called F_* (Jenkins 2009), and this parameter correlates well with the average number density along the sightline $n(H)$, but less so with the molecular fraction $f(H_2)$. The gas-to-dust ratio is also a tracer for this type of growth, and this metric has been found to be different between the diffuse and the dense ISM in the Magellanic Clouds (Roman-Duval et al. 2014, 2017, 2021),

providing evidence that grain growth occurs mostly in high-density regions of galaxies.

The variations in the extinction likely influence the way the ISM is affected by a UV radiation field. In numerical models of PDRs the depth of the $H\text{I}$ - H_2 transition is commonly expressed in terms of the optical depth $A(V)$ (Röllig et al. 2007). However, due to changes in the grain size distribution, the UV-to-optical extinction ratio can differ between sightlines, between PDRs, or even between different locations in the same PDR. Figure 3 of Mathis (1990) shows how the penetration of UV radiation changes for different $R(V)$, at constant $A(V)$. Since UV radiation and its attenuation by dust and gas are the main drivers of the physics in a PDR, the outcome can be very different for clouds with similar $A(V)$ but different dust populations. Radiative transfer simulations of PDRs show that a spatially varying extinction law can produce substantially different locations of the $H\text{I}$ - H_2 transition (Goicoechea & Le Bourlot 2007).

Coming back to the main goal of this work, we aim to study how interstellar dust influences and responds to the different environments in the diffuse ISM of the Milky Way, with a focus on H_2 . To probe the relationship between details of the UV extinction curves, the H_2 content, and the gas-to-dust ratio, we need a sample of sightlines with measurements of the necessary quantities: the extinction at every UV wavelength $A(\lambda)$, and the column densities $N(H\text{I})$ and $N(H_2)$. The sample of 75 diffuse Milky Way sightlines by Gordon et al. (2009, hereafter **GCC09**) is ideal for this. It is currently the largest sample with extinction curves measured using FUSE FUV data (912–1200 Å), and covers a large range of $R(V)$ values ~ 2.6 –5.5. Preliminary work, investigating FUV extinction properties in the Galaxy and the Magellanic Clouds with International Ultraviolet Explorer (IUE) data (~ 1170 –3200 Å), includes 78 Milky Way sightlines by Fitzpatrick & Massa (1990, hereafter **FM90**), and several LMC sightlines (Misselt et al. 1999) and SMC sightlines (Gordon & Clayton 1998). Later, FUSE data were used to extend extinction curves to shorter wavelengths. Sofia et al. (2005) studied the FUV extinction curves along nine Milky Way sightlines and Cartledge et al. (2005) presented a study of nine FUV extinction curves in the Magellanic Clouds. The **GCC09** paper extended the sample of Milky Way sightlines with FUSE data to 75 stars, and used a more mature FUSE calibration, a better correction for the H_2 absorption, and a more complete set of comparison stars. These improvements in data quality have exposed the need for improved extinction curve parameterizations. In **GCC09**, the reformalization of the CCM family of curves (Cardelli et al. 1989) was initiated, concentrating on the FUV spectrum of stars observed using FUSE. In this work, we effectively extend the original analysis. While **GCC09** focused on correlations between extinction curve properties, we will investigate correlations between those properties and quantities related to the gas.

The number of previous observational studies including both the details of H_2 and the extinction curves is rather limited. A correlation between the molecular fraction and the bump width was found by Rachford et al. (2002), using a sample of 23 sightlines. This study also found a less significant correlation between the strength of the FUV rise and the molecular fraction. Our work significantly improves the statistical significance for the latter correlation. Another study investigated the relation between $R(V)$ and H_2 , for 38 sightlines (Rachford et al. 2009). They did not find significant evidence for a general trend where the molecular fraction is

Table 1
Molecular Hydrogen Column Density Data

Star	$N(J=0)$	$N(1)$	$N(2)$	$N(3)$	$N(4)$	$N(5)$	$N(6)$	$N(7)$
Reddened Stars								
BD+35 4258	19.10 ± 0.10	19.35 ± 0.10	17.45 ± 0.35	17.25 ± 0.40	15.50 ± 0.10	15.20 ± 0.06	...	...
BD+53 2820	19.55 ± 0.25	19.80 ± 0.25	17.45 ± 0.50	17.05 ± 0.50	15.25 ± 0.25	14.65 ± 0.15	...	...
	19.00 ± 0.50	19.00 ± 0.50	15.40 ± 0.25	15.40 ± 0.25	14.60 ± 0.25	14.30 ± 0.25	...	...
BD+56 524	19.65 ± 0.40	20.40 ± 0.40	18.55 ± 0.25	16.45 ± 0.25	15.65 ± 0.25	15.00 ± 0.25	...	...
	20.45 ± 0.40	19.85 ± 0.40	18.75 ± 0.25	16.65 ± 0.25	14.85 ± 0.25	14.65 ± 0.25	...	...
HD 001383	19.80 ± 0.25	19.85 ± 0.25	17.20 ± 0.25	16.55 ± 0.25	15.10 ± 0.25	14.75 ± 0.25	...	...
	19.60 ± 0.25	19.85 ± 0.25	17.00 ± 0.25	16.50 ± 0.25	15.10 ± 0.25	14.60 ± 0.25	...	...
	19.00 ± 0.25	19.35 ± 0.25	16.20 ± 0.25	15.15 ± 0.25	13.90 ± 0.25	13.60 ± 0.25	...	...
HD 013268	19.98 ± 0.10	20.22 ± 0.10	17.02 ± 0.50	16.57 ± 0.40	15.67 ± 0.10	15.19 ± 0.08	...	...
HD 014250	20.15 ± 0.25	20.40 ± 0.25	18.70 ± 0.15	18.05 ± 0.25	15.85 ± 0.15	14.90 ± 0.15	...	...
	19.55 ± 0.50	19.40 ± 0.25	16.00 ± 0.50	16.00 ± 0.25	15.05 ± 0.15	14.70 ± 0.15	...	...
HD 014434	20.10 ± 0.25	20.30 ± 0.15	18.00 ± 0.50	16.90 ± 0.15	15.40 ± 0.10	14.80 ± 0.15	...	...
	18.80 ± 0.25	19.00 ± 0.25	17.80 ± 0.50	16.80 ± 0.25	15.30 ± 0.15	14.85 ± 0.15	...	...
HD 015558	20.39 ± 0.09	20.58 ± 0.09	18.15 ± 0.65	16.96 ± 1.17	15.70 ± 0.52	15.26 ± 0.22	...	...
HD 017505	20.35 ± 0.10	20.59 ± 0.10	18.04 ± 0.40	17.20 ± 0.60	15.82 ± 0.20	15.47 ± 0.20	14.68 ± 0.13	14.40 ± 0.13
HD 023060	20.34 ± 0.09	20.13 ± 0.09	18.42 ± 0.05	17.10 ± 0.40	15.58 ± 0.25	14.74 ± 0.15	...	...
HD 027778 ^a	20.60 ± 0.10	20.10 ± 0.10	18.54 ± 0.08	17.47 ± 0.17	15.32 ± 0.42	14.69 ± 0.15	...	...
HD 037903 ^a	20.54 ± 0.09	20.50 ± 0.12	19.60 ± 0.09	18.56 ± 0.07	17.08 ± 0.46	15.61 ± 0.11	14.99 ± 0.35	14.49 ± 0.30
HD 038087 ^a	20.37 ± 0.09	20.30 ± 0.09	19.10 ± 0.13	18.21 ± 0.10	16.41 ± 0.65	15.84 ± 0.50	14.89 ± 0.30	14.61 ± 0.20
HD 045314	20.23 ± 0.07	20.25 ± 0.09	18.19 ± 0.12	17.32 ± 0.34	15.12 ± 0.17	14.34 ± 0.05	...	...
HD 046056 ^a	20.44 ± 0.10	20.38 ± 0.10	18.29 ± 0.10	17.37 ± 0.30	15.64 ± 0.17	15.02 ± 0.13	...	...
HD 046150	20.24 ± 0.08	20.24 ± 0.12	17.67 ± 0.40	16.90 ± 0.50	15.41 ± 0.09	14.90 ± 0.05	...	...
HD 046202 ^a	20.36 ± 0.09	20.29 ± 0.09	18.20 ± 0.22	17.54 ± 0.37	15.51 ± 0.24	15.10 ± 0.14	...	...
HD 047129	20.01 ± 0.09	20.14 ± 0.09	17.99 ± 0.17	17.11 ± 0.36	15.51 ± 0.13	14.90 ± 0.05	...	...
HD 047240	20.03 ± 0.12	20.06 ± 0.12	18.14 ± 0.07	17.75 ± 0.11	15.83 ± 0.14	15.60 ± 0.12	14.51 ± 0.09	14.43 ± 0.10
HD 047417	20.20 ± 0.08	20.10 ± 0.15	17.85 ± 0.22	17.00 ± 0.43	15.30 ± 0.06	14.70 ± 0.05	...	...
HD 062542 ^a	20.75 ± 0.25	20.13 ± 0.25	19.01 ± 0.25	16.29 ± 0.25	15.69 ± 0.25	...	...	...
HD 073882 ^a	21.00 ± 0.09	20.45 ± 0.09	18.87 ± 0.05	18.05 ± 0.11	15.81 ± 0.09	15.43 ± 0.06	...	...
HD 091651	18.75 ± 0.15	18.80 ± 0.15	17.00 ± 0.42	17.00 ± 0.40	15.45 ± 0.08	15.05 ± 0.08	14.30 ± 0.25	14.00 ± 0.50
HD 093222	19.50 ± 0.10	19.40 ± 0.15	18.20 ± 0.25	17.75 ± 0.30	15.10 ± 0.15	14.40 ± 0.15	...	...
HD 093250	19.90 ± 0.25	20.12 ± 0.15	17.35 ± 0.50	17.38 ± 0.40	16.10 ± 0.20	16.06 ± 0.30	15.30 ± 0.30	15.40 ± 0.30
HD 096675 ^a	20.63 ± 0.15	20.35 ± 0.10	18.55 ± 0.07	16.85 ± 0.38	15.45 ± 0.14	14.50 ± 0.17	...	...
HD 096715	20.48 ± 0.08	20.15 ± 0.10	17.55 ± 0.46	16.75 ± 0.62	15.55 ± 0.18	14.80 ± 0.16	...	...
HD 099872	20.31 ± 0.15	20.18 ± 0.15	18.48 ± 0.20	16.05 ± 1.20	14.90 ± 0.85	14.42 ± 0.80	...	...
HD 099890	19.15 ± 0.15	19.30 ± 0.10	16.65 ± 0.60	16.40 ± 0.55	14.93 ± 0.09	14.45 ± 0.06	...	...
HD 100213	20.15 ± 0.12	20.15 ± 0.12	16.85 ± 1.60	15.80 ± 2.10	14.90 ± 0.60	14.45 ± 0.25	...	...
HD 101190	20.22 ± 0.10	20.02 ± 0.10	18.23 ± 0.65	16.20 ± 0.51	14.72 ± 0.08	14.29 ± 0.08	...	...
HD 101205	20.00 ± 0.23	19.90 ± 0.12	18.30 ± 0.20	17.70 ± 0.42	15.13 ± 0.46	14.76 ± 0.25	...	...
HD 103779	19.35 ± 0.25	19.70 ± 0.10	16.15 ± 0.55	15.80 ± 0.35	14.60 ± 0.15	14.30 ± 0.18	...	...
HD 122879	19.90 ± 0.10	20.10 ± 0.10	17.38 ± 0.15	16.83 ± 0.45	15.30 ± 0.40	14.70 ± 0.25	...	...
HD 124979	20.08 ± 0.10	20.18 ± 0.10	17.85 ± 0.20	16.91 ± 0.45	15.38 ± 0.10	14.70 ± 0.10	...	...
HD 147888 ^a	20.37 ± 0.09	19.76 ± 0.12	18.58 ± 0.10	17.06 ± 0.25	15.70 ± 0.44	15.24 ± 0.38	14.38 ± 0.15	...
HD 148422	19.85 ± 0.10	19.80 ± 0.10	17.55 ± 0.30	16.85 ± 0.30	15.35 ± 0.25	14.75 ± 0.10	...	...
	15.30 ± 0.30	15.50 ± 0.40	14.85 ± 0.15	14.80 ± 0.15	13.75 ± 0.25	14.30 ± 0.10	...	...
HD 149404 ^a	20.60 ± 0.10	20.35 ± 0.10	18.50 ± 0.40	16.84 ± 0.15	15.94 ± 0.15	15.34 ± 0.10	14.00 ± 0.25	...
HD 151805	20.10 ± 0.10	20.01 ± 0.08	18.22 ± 0.18	16.90 ± 0.55	15.63 ± 0.37	15.06 ± 0.10	...	...
HD 152233	20.05 ± 0.15	19.95 ± 0.15	18.40 ± 0.10	17.55 ± 0.25	16.50 ± 0.25	15.50 ± 0.10	...	...
	17.30 ± 1.50	16.70 ± 1.00	15.00 ± 0.30	15.20 ± 0.15	14.65 ± 0.25	14.10 ± 0.15	...	...
HD 152234	20.30 ± 0.10	20.07 ± 0.10	18.25 ± 0.20	17.60 ± 0.25	15.90 ± 0.15	15.27 ± 0.25	...	...
HD 152248	20.00 ± 0.10	19.90 ± 0.10	18.30 ± 0.15	17.90 ± 0.25	15.70 ± 0.25	15.00 ± 0.25	...	...
HD 152249	19.95 ± 0.45	19.70 ± 0.15	18.00 ± 0.10	16.70 ± 0.15	14.85 ± 0.12	14.00 ± 0.25	...	...
	19.45 ± 0.25	19.65 ± 0.15	17.10 ± 0.10	16.65 ± 0.10	15.45 ± 0.12	14.65 ± 0.10	...	...
	18.60 ± 1.00	18.60 ± 1.00	14.90 ± 0.10	14.90 ± 0.10	14.25 ± 0.10	14.05 ± 0.10	...	...
HD 152723	20.00 ± 0.15	20.00 ± 0.10	16.35 ± 0.30	16.00 ± 0.60	15.10 ± 0.25	14.80 ± 0.10	...	...
HD 157857	20.30 ± 0.10	20.40 ± 0.10	18.00 ± 0.20	16.60 ± 0.20	15.05 ± 0.10	14.65 ± 0.10	...	...
HD 160993	19.14 ± 0.06	19.20 ± 0.06	17.90 ± 0.05	17.68 ± 0.05	15.63 ± 0.09	14.74 ± 0.07	...	...
HD 163522	19.20 ± 0.20	19.25 ± 0.20	17.60 ± 0.60	16.20 ± 0.60	14.80 ± 0.10	14.25 ± 0.10	...	...
	18.20 ± 0.50	18.50 ± 0.20	15.10 ± 0.20	15.00 ± 0.10	14.55 ± 0.10	13.75 ± 0.10	...	...
HD 164816	19.66 ± 0.06	19.50 ± 0.08	17.56 ± 0.10	17.04 ± 0.25	15.08 ± 0.50	14.32 ± 0.10	...	...
HD 164906	19.98 ± 0.10	19.89 ± 0.10	18.30 ± 0.10	17.15 ± 0.35	15.05 ± 0.10	14.30 ± 0.30	...	...
HD 165052	19.75 ± 0.15	19.95 ± 0.15	17.35 ± 0.60	16.80 ± 0.20	14.95 ± 0.10	14.50 ± 0.10	...	...
HD 167402	19.95 ± 0.10	19.75 ± 0.15	18.20 ± 0.10	17.85 ± 0.40	15.10 ± 0.10	14.35 ± 0.15	...	...
HD 167771	20.40 ± 0.15	20.35 ± 0.10	18.27 ± 0.25	16.12 ± 0.15	15.20 ± 0.10	14.65 ± 0.10	...	...
HD 168076 ^a	20.45 ± 0.10	20.30 ± 0.10	18.30 ± 0.30	17.20 ± 0.50	16.40 ± 0.30	15.85 ± 0.15	14.75 ± 0.15	14.60 ± 0.15
HD 168941	19.98 ± 0.10	19.82 ± 0.10	17.76 ± 0.50	17.20 ± 0.15	15.19 ± 0.10	14.56 ± 0.10	...	...
HD 178487	20.24 ± 0.10	20.15 ± 0.15	18.21 ± 0.08	17.83 ± 0.13	15.67 ± 0.10	14.85 ± 0.11	...	...
HD 179406 ^a	20.46 ± 0.08	20.25 ± 0.08	18.10 ± 0.15	17.05 ± 0.30	15.20 ± 0.10	14.45 ± 0.10	...	...
HD 179407	20.03 ± 0.10	19.99 ± 0.10	17.80 ± 0.32	17.01 ± 0.66	15.37 ± 0.21	14.78 ± 0.10	...	...
HD 185418 ^a	20.29 ± 0.09	20.48 ± 0.09	18.32 ± 0.05	17.34 ± 0.13	15.35 ± 0.08	14.53 ± 0.50	...	...
HD 188001	19.77 ± 0.10	19.97 ± 0.10	17.89 ± 0.18	16.89 ± 0.70	15.24 ± 0.54	14.78 ± 0.30	...	...
HD 190603	20.35 ± 0.10	20.40 ± 0.10	18.40 ± 0.10	17.58 ± 0.12	15.88 ± 0.10	15.41 ± 0.10	...	...

Table 1
(Continued)

Star	$N(J=0)$	$N(1)$	$N(2)$	$N(3)$	$N(4)$	$N(5)$	$N(6)$	$N(7)$
HD 192639 ^a	20.33 ± 0.10	20.45 ± 0.10	18.41 ± 0.09	17.51 ± 0.17	15.77 ± 0.07	15.24 ± 0.07	14.44 ± 0.17	...
HD 197770	20.60 ± 0.20	20.59 ± 0.10	18.44 ± 0.10	17.58 ± 0.17	16.20 ± 0.71	15.15 ± 0.60	...	...
HD 198781	20.23 ± 0.09	20.11 ± 0.09	17.55 ± 0.73	16.30 ± 1.40	14.90 ± 0.11	14.35 ± 0.14	...	...
HD 199579 ^a	20.25 ± 0.10	20.20 ± 0.15	18.45 ± 0.15	17.10 ± 0.10	15.95 ± 0.05	15.55 ± 0.05	14.30 ± 0.05	14.35 ± 0.05
HD 200775	20.80 ± 0.10	20.60 ± 0.20	19.50 ± 0.15	18.00 ± 0.30	17.05 ± 0.15	15.50 ± 1.30	14.90 ± 1.20	...
HD 203938 ^a	20.75 ± 0.09	20.65 ± 0.09	19.10 ± 0.10	17.71 ± 0.20	16.16 ± 0.35	15.46 ± 0.13	...	...
HD 206267 ^a	20.62 ± 0.06	20.35 ± 0.09	17.94 ± 0.21	16.76 ± 0.58	15.26 ± 0.15	14.92 ± 0.08	...	...
HD 206773	20.00 ± 0.15	20.20 ± 0.10	18.20 ± 0.08	17.40 ± 0.18	15.40 ± 0.06	14.75 ± 0.05	...	...
HD 207198 ^a	20.60 ± 0.10	20.45 ± 0.10	18.40 ± 0.10	17.00 ± 0.15	15.70 ± 0.10	14.90 ± 0.05	...	...
HD 209339	19.85 ± 0.10	20.00 ± 0.10	18.00 ± 0.18	17.08 ± 0.24	15.31 ± 0.10	14.70 ± 0.05	...	...
HD 216898	20.60 ± 0.10	20.80 ± 0.10	18.82 ± 0.14	17.60 ± 0.34	15.97 ± 0.10	15.36 ± 0.15	...	...
HD 239729	20.92 ± 0.15	20.52 ± 0.15	18.72 ± 0.25	17.74 ± 0.30	15.36 ± 1.00	...	...	...
HD 326329	19.98 ± 0.06	19.96 ± 0.06	18.20 ± 0.25	17.70 ± 0.35	15.70 ± 0.10	15.15 ± 0.12	13.70 ± 0.60	...
HD 332407	20.04 ± 0.10	20.15 ± 0.10	17.76 ± 0.25	16.87 ± 0.35	15.61 ± 0.06	15.16 ± 0.07	...	...
Comparison Stars								
BD+32 270	18.28 ± 0.08	18.47 ± 0.05	16.48 ± 0.57	15.82 ± 0.20	14.60 ± 0.10	13.73 ± 0.10	...	...
BD+52 3210	19.40 ± 0.09	19.71 ± 0.09	16.75 ± 0.25	16.75 ± 0.25	15.65 ± 0.10	15.30 ± 0.10	...	...
HD 037332	15.50 ± 0.25	15.08 ± 0.25	14.58 ± 0.25	14.30 ± 0.25	...	...	...	...
HD 037525	15.40 ± 1.40	16.10 ± 1.20	15.20 ± 1.65	14.90 ± 2.80	...	...	...	...
HD 051013	13.40 ± 0.15	14.04 ± 0.15	13.60 ± 0.15	13.90 ± 0.15	...	...	...	...
HD 075309	19.74 ± 0.09	19.87 ± 0.09	17.63 ± 0.12	17.14 ± 0.26	15.18 ± 0.12	14.76 ± 0.07	...	...
HD 091824	19.51 ± 0.09	19.36 ± 0.09	17.80 ± 0.11	17.41 ± 0.18	15.25 ± 0.05	14.72 ± 0.05	...	...
HD 091983	19.97 ± 0.12	19.72 ± 0.09	17.36 ± 0.50	16.42 ± 0.45	15.00 ± 0.20	14.67 ± 0.15	...	...
HD 093028	19.10 ± 0.12	19.26 ± 0.12	17.77 ± 0.09	17.52 ± 0.13	14.92 ± 0.11	14.49 ± 0.05	...	...
HD 094493	19.85 ± 0.12	19.80 ± 0.09	18.04 ± 0.14	17.45 ± 0.26	15.15 ± 0.07	14.66 ± 0.05	...	...
HD 097471	19.65 ± 0.10	19.50 ± 0.10	17.58 ± 0.34	17.14 ± 0.43	15.41 ± 0.11	14.85 ± 0.05	...	...
HD 100276	19.27 ± 0.15	19.26 ± 0.10	17.56 ± 0.60	16.73 ± 0.60	15.23 ± 0.40	14.57 ± 0.10	...	...
	19.07 ± 0.10	19.27 ± 0.10	16.23 ± 0.15	15.72 ± 0.60	14.65 ± 0.25	14.17 ± 0.10	...	...
HD 104705	19.64 ± 0.10	19.66 ± 0.15	15.95 ± 0.20	15.85 ± 0.20	15.00 ± 0.20	14.38 ± 0.10	...	...
	16.48 ± 1.00	16.10 ± 1.00	15.95 ± 0.20	15.27 ± 0.20	14.10 ± 0.30	14.00 ± 0.20	...	...
HD 114444	18.20 ± 0.25	18.10 ± 0.25	16.00 ± 0.35	15.50 ± 0.50	14.30 ± 0.20	...	...	...
	19.40 ± 0.15	19.60 ± 0.15	16.10 ± 0.35	15.60 ± 0.50	14.50 ± 0.10	...	...	...
HD 116852	19.54 ± 0.10	19.45 ± 0.10	17.61 ± 0.05	17.37 ± 0.06	14.97 ± 0.05	14.11 ± 0.05	...	...
HD 172140	19.00 ± 0.15	18.80 ± 0.20	16.23 ± 0.10	15.82 ± 0.10	14.44 ± 0.10	14.01 ± 0.10	...	...
HD 210809	18.15 ± 0.75	18.50 ± 0.50	15.40 ± 0.20	16.00 ± 0.40	14.60 ± 0.20	14.15 ± 0.15	...	...
	19.00 ± 0.35	19.45 ± 0.20	15.25 ± 0.30	14.70 ± 0.30	15.70 ± 0.15	15.05 ± 0.10	...	...
	19.40 ± 0.35	18.65 ± 0.35	17.25 ± 1.00	17.35 ± 0.50	14.15 ± 0.15	13.70 ± 0.25	...	...
	18.10 ± 0.75	19.00 ± 0.35	15.30 ± 0.20	15.00 ± 0.20	14.75 ± 0.10	14.40 ± 0.10	...	...
HD 235874	19.20 ± 0.10	19.45 ± 0.10	17.70 ± 0.25	17.40 ± 0.25	15.72 ± 0.25	15.20 ± 0.25	...	...
Extra (no extinction curve)								
HD 003827	16.89 ± 0.15	17.47 ± 0.15	16.44 ± 0.15	16.07 ± 0.15	14.57 ± 0.08	14.17 ± 0.08	...	...
HD 022586	14.92 ± 0.10	15.20 ± 0.10	14.62 ± 0.20	14.35 ± 0.15	...	...	...	...
HD 177989	20.00 ± 0.10	19.65 ± 0.10	18.00 ± 0.30	17.05 ± 0.20	15.05 ± 0.10	14.47 ± 0.10	...	...
HD 210121 ^a	20.63 ± 0.15	20.13 ± 0.15	18.71 ± 0.30	18.10 ± 0.30	16.50 ± 0.90	14.80 ± 0.75	...	...
HD 210839 ^a	20.50 ± 0.10	20.40 ± 0.10	18.40 ± 0.05	17.20 ± 0.15	16.20 ± 0.10	15.70 ± 0.10	14.50 ± 0.05	14.20 ± 0.05
HD 220057	20.05 ± 0.10	19.90 ± 0.10	17.70 ± 0.10	16.10 ± 0.20	14.85 ± 0.10	14.35 ± 0.15	...	...
HD 239683	20.42 ± 0.30	20.48 ± 0.25	18.69 ± 0.25	17.75 ± 0.25	15.66 ± 0.25	14.88 ± 0.25	...	...
HD 303308	15.00 ± 0.50	15.00 ± 0.50	15.00 ± 0.50	14.40 ± 0.50	13.60 ± 0.50	13.30 ± 0.50	13.00 ± 0.50	12.00 ± 0.50
	16.80 ± 0.50	17.00 ± 0.50	15.80 ± 0.50	15.50 ± 0.50	16.00 ± 0.50	15.40 ± 0.50	15.40 ± 0.50	15.00 ± 0.50
	19.20 ± 0.25	19.35 ± 0.25	17.50 ± 0.50	16.80 ± 0.50	16.00 ± 0.50	15.90 ± 0.50	15.50 ± 0.25	15.60 ± 0.50
	19.25 ± 0.25	19.50 ± 0.25	17.40 ± 0.50	16.30 ± 0.50	16.00 ± 0.50	15.90 ± 0.25	14.80 ± 0.25	14.70 ± 0.25
	19.35 ± 0.25	19.60 ± 0.25	17.20 ± 0.50	15.60 ± 0.50	16.00 ± 0.50	15.30 ± 0.25	14.80 ± 0.50	14.60 ± 0.50
	17.00 ± 0.50	16.00 ± 0.50	14.50 ± 0.50	14.50 ± 0.50	13.50 ± 0.50	13.70 ± 0.50	13.00 ± 0.00	13.00 ± 0.50
HD 315021	19.55 ± 0.15	19.75 ± 0.15	17.60 ± 0.20	16.10 ± 0.50	14.45 ± 0.10	14.00 ± 0.15	...	...
	18.60 ± 0.20	18.65 ± 0.20	15.25 ± 0.20	14.90 ± 0.10	13.72 ± 0.20	13.80 ± 0.10	...	...

Notes. Table is divided into three sections, with 75 reddened stars and 18 comparison stars of the GCC09 sample, as well as nine extra sightlines for which extinction curves could not be derived by GCC09. All column densities are expressed in logarithmic form for absolute values per square centimeter. For models with multiple velocity components, one data line is given per component.

^a FUSE H₂ also measured by Rachford et al. (2002), Rachford et al. (2009).

(This table is available in machine-readable form.)

lower for high- $R(V)$ sightlines. With our sample, the number of sightlines is about double, and we examine if this conclusion holds. Works investigating other diatomic molecules based on UV absorption, have discovered relationships between the FUV extinction rise and CH (Jenniskens et al.

1992) as well as CO (Burgh et al. 2000). Given the linear relationship between CH and H₂ (Federman 1982), and the common use of CO as a tracer for H₂ (Scoville et al. 1987), there is a strong expectation to see a similar behavior of the dust extinction for H₂.

In Section 2, we describe data used for our investigation of the GCC09 sample 75 diffuse Milky Way sightlines. This includes the data provided by the GCC09 paper, the H_2 absorption models that were fit to the FUSE data to determine $N(H_2)$, and the $Ly\alpha$ absorption profile-fitting method used to determine $N(H\text{ I})$. We also describe how several derived quantities were calculated, and how we deal with the uncertainties and correlations between those quantities. In Section 3, we show the correlations that were found when comparing the extinction and the gas contents. The interpretation of these results is discussed in Section 4, and to conclude, the main results and their interpretation are summarized in Section 5, followed by a brief discussion of the future outlook.

2. Data

2.1. Sample

In GCC09, the original paper discussing this sample, it is described in detail how the sightlines were selected. In short they were selected from the set of all Galactic O and B stars observed by FUSE that also have sufficient observations to provide spectral coverage from the infrared (IR) through the FUV. This sample includes 75 reddened stars and 18 lightly reddened stars, the latter serving as comparison stars. The data we collected for these sightlines are summarized in two tables. Table 1 contains molecular hydrogen column densities for rotational levels up to $J = 7$. The level populations result from fitting an H_2 absorption model to FUSE spectra, as described in Section 2.2. These models were created in the context of the GCC09 paper, to remove H_2 absorption features from the FUV extinction curves. However, the resulting level populations and a detailed description of the fitting method went unpublished until now.

The total molecular and atomic hydrogen column densities, as well as several derived quantities for the sightlines, are summarized in Table 2. Sections 2.3 and 2.4 describe how these data were obtained.

2.2. H_2 Modeling

The complete set of FUSE data contains 110 sightlines. In GCC09, 75 of them have extinction curves, 18 were used as comparison stars, and for nine of them it was not possible to derive extinction curves. We publish the H_2 results for all of them, and have sorted the sightlines accordingly in Table 1, but we will not use the last nine sightlines in the rest of this work.

The H_2 column densities in each rotational excitation level, the line width parameter b , and the heliocentric velocity of the profile center define the absorption model for each sightline. In general, an effort was made to keep the models as simple as possible. In the absence of clear evidence for multiple velocity components in the absorption, a single component is assumed. Single-component models are used for 92 of the 110 sightlines, two-component models are used for 14 sightlines, two sightlines are described by three-component models, one by a four-component model, and one is composed of six absorption components. To determine the number of components needed, lines of other gas species were used, in data from earlier work on elemental abundances (Cartledge et al. 2006) or information from similar papers such as that of Rachford et al. (2002). When other species significantly indicated multiple absorption components, the number is based on the strongest absorbing components. Otherwise, the modeling started with a single

component, and the number of components was increased until the continuum of the lines could be reconstructed sufficiently.

The procedure for constructing these molecular hydrogen models was patterned after a common approach applied to FUSE data (Rachford et al. 2001, 2002), which we have previously used in analyses of FUV extinction curves (Sofia et al. 2005; Cartledge et al. 2005; Gordon et al. 2009). In the context of the GCC09 paper, the FUSE observations from the archive were processed using CALFUSE (v3.0). The GCC09 paper did not provide the details about the reduction and H_2 modeling of these FUSE data, and instead referred to a future paper, which went unpublished. We therefore provide the details in this work.

In cases where the different observations or FUSE channels produced conflicting flux levels for a given wavelength (e.g., due to channel misalignment), the fluxes for these observations or channels were scaled by a single factor to match the maximum observed flux across the regions of spectral coverage overlap. Then, for each single channel of each observation, the calibrated spectrum was cross-correlated with either the LiF1A or LiF2B spectrum. The latter choice applied to the LiF1B and LiF2A spectra, whose wavelength coverage does not overlap with LiF1A. In this fashion, each spectrum could be shifted such that it conformed to the LiF1A wavelength solution across the LiF1A bandpass, and could then be averaged together across the entire FUSE wavelength window. Note that the solutions for each segment are technically not compatible with a simple linear shift; nevertheless, the quality of the data resulting from the averaging process suffices for the purpose of measuring molecular hydrogen column densities. A single FUSE spectrum was generated for each sightline by averaging together all observations of a given star on a wavelength-by-wavelength basis, weighted by exposure time.

The FUSE spectrum for a given sightline was then processed using a code that allows the user to concurrently identify absorption features due to the various rovibrational H_2 transitions ($J' \rightarrow 0$ and $v' \rightarrow 0$) in the FUSE 910–1190 Å bandpass. The equivalent widths were measured assuming a background stellar spectrum fit by a low-order polynomial. The order of the polynomial varies from sightline to sightline, depending on the stellar continuum, but was at most second order. For each segment of the H_2 fits, there was no concrete evidence for higher-order variations on the scale of the wavelength windows used. The data for each absorption component along the sightline are listed in Table 1, omitting lines that were indistinguishably blended with other absorption features. For the sightlines where multiple H_2 components were needed, the table has multiple rows associated with a single star name.

In cases where multiple components were evident, particular care was taken to separate the amount of absorption associated with each line-of-sight cloud using the profile-fitting code *fits6p* (Welty et al. 1991; Mar & Bailey 1995). In addition, this code was used on the spectra from every sightline to determine reasonable b values for each absorption component along a sightline using weak $J = 0-0$ transitions longward of 1110 Å. The errors on the equivalent widths are also provided by the *fits6p* code, after the determination of the b values, and these include contributions from both the continuum fitting and the noise in the data. Each table was then subjected to a curve-of-growth (CoG) analysis which generated a b value and $J = 0-7$ molecular hydrogen column density solution using the

Table 2
Total Gas Column Densities and Additional Quantities

Star	$\log_{10} N(\text{H})$ (cm^{-2})	$\log_{10} N(\text{H I})$ (cm^{-2})	$\log_{10} N(\text{H}_2)$ (cm^{-2})	d (kpc)	$f(\text{H}_2)$	$\log_{10} n(\text{H})$ (cm^{-3})	H I Reference
Reddened Stars							
BD+35d4258	21.26 $\pm$ 0.05	21.24 $\pm$ 0.05	19.55 $\pm$ 0.07	1.89 $\pm$ 0.19	0.04	0.31	6
BD+53d2820	21.39 $\pm$ 0.05	21.35 $\pm$ 0.05	20.08 $\pm$ 0.17	3.18 $\pm$ 0.41	0.10	0.25	6
BD+56d524	21.57 $\pm$ 0.16	21.38 $\pm$ 0.16	20.82 $\pm$ 0.28	2.19 $\pm$ 0.21	0.36	0.55	1
HD 001383	21.54 $\pm$ 0.04	21.46 $\pm$ 0.04	20.44 $\pm$ 0.13	3.04 $\pm$ 0.30	0.16	0.37	6
HD 013268 ^a	21.43 $\pm$ 0.04	21.34 $\pm$ 0.05	20.42 $\pm$ 0.08	2.80 $\pm$ 0.28	0.19	0.31	6
HD 014250	21.63 $\pm$ 0.13	21.52 $\pm$ 0.15	20.66 $\pm$ 0.18	1.20 $\pm$ 0.11	0.22	1.14	4
HD 014434 ^a	21.48 $\pm$ 0.08	21.37 $\pm$ 0.09	20.54 $\pm$ 0.14	2.98 $\pm$ 0.30	0.23	0.33	2
HD 015558	21.66 $\pm$ 0.14	21.52 $\pm$ 0.18	20.80 $\pm$ 0.07	2.02 $\pm$ 0.36	0.28	0.73	3
HD 017505	21.48 $\pm$ 0.10	21.26 $\pm$ 0.15	20.79 $\pm$ 0.07	2.56 $\pm$ 0.22	0.40	0.39	3
HD 023060	21.51 $\pm$ 0.06	21.40 $\pm$ 0.07	20.55 $\pm$ 0.07	0.53 $\pm$ 0.01	0.22	1.96	1
HD 027778	21.36 $\pm$ 0.08	21.10 $\pm$ 0.12	20.72 $\pm$ 0.08	0.22 $\pm$ 0.00	0.45	3.36	2
HD 037903	21.46 $\pm$ 0.06	21.16 $\pm$ 0.09	20.85 $\pm$ 0.07	0.40 $\pm$ 0.01	0.49	2.34	2
HD 038087	21.59 $\pm$ 0.12	21.48 $\pm$ 0.15	20.65 $\pm$ 0.06	0.34 $\pm$ 0.01	0.23	3.78	4
HD 045314 ^a	21.26 $\pm$ 0.06	21.05 $\pm$ 0.09	20.54 $\pm$ 0.06	0.80 $\pm$ 0.08	0.38	0.74	1
HD 046056	21.53 $\pm$ 0.10	21.38 $\pm$ 0.14	20.71 $\pm$ 0.07	1.46 $\pm$ 0.13	0.30	0.76	3
HD 046150 ^a	21.40 $\pm$ 0.09	21.26 $\pm$ 0.12	20.54 $\pm$ 0.08	1.47 $\pm$ 0.15	0.28	0.55	3
HD 046202	21.52 $\pm$ 0.10	21.39 $\pm$ 0.13	20.63 $\pm$ 0.07	1.30 $\pm$ 0.25	0.26	0.83	1
HD 047129	21.36 $\pm$ 0.07	21.26 $\pm$ 0.09	20.38 $\pm$ 0.07	1.45 $\pm$ 0.20	0.21	0.51	1
HD 047240	21.31 $\pm$ 0.10	21.20 $\pm$ 0.12	20.35 $\pm$ 0.09	2.01 $\pm$ 0.30	0.22	0.33	3
HD 047417 ^a	21.28 $\pm$ 0.07	21.13 $\pm$ 0.09	20.45 $\pm$ 0.09	1.23 $\pm$ 0.12	0.29	0.50	3
HD 062542	21.69 $\pm$ 0.11	21.54 $\pm$ 0.12	20.85 $\pm$ 0.21	0.39 $\pm$ 0.00	0.29	4.11	1
HD 073882 ^a	21.59 $\pm$ 0.07	21.11 $\pm$ 0.15	21.11 $\pm$ 0.07	0.83 $\pm$ 0.08	0.67	1.51	4
HD 091651 ^a	21.16 $\pm$ 0.06	21.15 $\pm$ 0.06	19.08 $\pm$ 0.11	3.98 $\pm$ 0.40	0.02	0.12	3
HD 093222 ^a	21.49 $\pm$ 0.03	21.47 $\pm$ 0.03	19.77 $\pm$ 0.09	2.47 $\pm$ 0.25	0.04	0.40	6
HD 093250 ^a	21.46 $\pm$ 0.14	21.39 $\pm$ 0.15	20.32 $\pm$ 0.14	2.20 $\pm$ 0.22	0.15	0.42	3
HD 093827	21.41 $\pm$ 0.18	21.34 $\pm$ 0.19	20.25 $\pm$ 0.10	3.60 $\pm$ 0.47	0.14	0.23	1
HD 096675	21.88 $\pm$ 0.09	21.80 $\pm$ 0.10	20.82 $\pm$ 0.11	0.16 $\pm$ 0.00	0.17	15.28	1
HD 096715 ^a	21.39 $\pm$ 0.11	21.20 $\pm$ 0.16	20.65 $\pm$ 0.06	3.15 $\pm$ 0.32	0.36	0.25	3
HD 099872b	21.47 $\pm$ 0.07	21.35 $\pm$ 0.08	20.55 $\pm$ 0.11	...	0.24	...	1
HD 099890 ^a	21.14 $\pm$ 0.05	21.12 $\pm$ 0.05	19.53 $\pm$ 0.09	3.22 $\pm$ 0.32	0.05	0.14	6
HD 100213 ^a	21.32 $\pm$ 0.06	21.18 $\pm$ 0.07	20.45 $\pm$ 0.09	2.15 $\pm$ 0.21	0.27	0.31	3
HD 101190 ^a	21.36 $\pm$ 0.03	21.24 $\pm$ 0.04	20.44 $\pm$ 0.07	2.09 $\pm$ 0.21	0.24	0.35	6
HD 101205 ^a	21.29 $\pm$ 0.06	21.20 $\pm$ 0.07	20.26 $\pm$ 0.15	1.64 $\pm$ 0.16	0.19	0.39	3
HD 103779 ^a	21.21 $\pm$ 0.03	21.17 $\pm$ 0.03	19.86 $\pm$ 0.11	3.99 $\pm$ 0.40	0.09	0.13	6
HD 122879	21.39 $\pm$ 0.04	21.31 $\pm$ 0.04	20.31 $\pm$ 0.08	2.23 $\pm$ 0.19	0.17	0.36	6
HD 124979 ^a	21.38 $\pm$ 0.05	21.27 $\pm$ 0.06	20.44 $\pm$ 0.07	3.09 $\pm$ 0.31	0.23	0.25	6
HD 147888	21.73 $\pm$ 0.12	21.68 $\pm$ 0.13	20.47 $\pm$ 0.08	0.09 $\pm$ 0.00	0.11	19.04	6
HD 148422 ^a	21.30 $\pm$ 0.05	21.24 $\pm$ 0.06	20.13 $\pm$ 0.07	8.26 $\pm$ 0.83	0.13	0.08	6
HD 149404	21.58 $\pm$ 0.10	21.40 $\pm$ 0.14	20.80 $\pm$ 0.07	1.27 $\pm$ 0.28	0.33	0.97	3
HD 151805	21.41 $\pm$ 0.03	21.33 $\pm$ 0.03	20.36 $\pm$ 0.07	1.59 $\pm$ 0.14	0.18	0.53	6
HD 152233 ^a	21.37 $\pm$ 0.09	21.29 $\pm$ 0.10	20.31 $\pm$ 0.11	1.52 $\pm$ 0.15	0.17	0.50	3
HD 152234	21.43 $\pm$ 0.10	21.32 $\pm$ 0.12	20.50 $\pm$ 0.08	1.53 $\pm$ -7.01	0.23	0.58	3
HD 152248 ^a	21.66 $\pm$ 0.14	21.62 $\pm$ 0.14	20.26 $\pm$ 0.07	1.61 $\pm$ 0.16	0.08	0.91	1
HD 152249	21.45 $\pm$ 0.06	21.38 $\pm$ 0.05	20.34 $\pm$ 0.25	1.95 $\pm$ 0.21	0.15	0.47	6
HD 152723 ^a	21.49 $\pm$ 0.12	21.43 $\pm$ 0.13	20.30 $\pm$ 0.09	1.94 $\pm$ 0.19	0.13	0.52	3
HD 157857 ^a	21.43 $\pm$ 0.07	21.26 $\pm$ 0.09	20.65 $\pm$ 0.07	2.56 $\pm$ 0.26	0.33	0.34	2
HD 160993	21.20 $\pm$ 0.10	21.18 $\pm$ 0.10	19.49 $\pm$ 0.04	3.11 $\pm$ 0.51	0.04	0.16	3
HD 163522	21.16 $\pm$ 0.05	21.14 $\pm$ 0.05	19.59 $\pm$ 0.14	3.87 $\pm$ 1.10	0.05	0.12	6
HD 164816 ^a	21.22 $\pm$ 0.12	21.18 $\pm$ 0.13	19.89 $\pm$ 0.05	1.08 $\pm$ 0.11	0.09	0.50	3
HD 164906	21.29 $\pm$ 0.08	21.20 $\pm$ 0.09	20.24 $\pm$ 0.07	1.19 $\pm$ 0.08	0.18	0.53	3
HD 165052 ^a	21.41 $\pm$ 0.09	21.36 $\pm$ 0.10	20.16 $\pm$ 0.11	1.37 $\pm$ 0.14	0.11	0.61	3
HD 167402 ^a	21.22 $\pm$ 0.03	21.13 $\pm$ 0.03	20.17 $\pm$ 0.09	7.61 $\pm$ 0.76	0.18	0.07	6
HD 167771 ^a	21.33 $\pm$ 0.08	21.08 $\pm$ 0.12	20.68 $\pm$ 0.10	1.40 $\pm$ 0.14	0.44	0.50	3
HD 168076 ^a	21.73 $\pm$ 0.21	21.65 $\pm$ 0.23	20.68 $\pm$ 0.07	1.97 $\pm$ 0.20	0.18	0.89	3
HD 168941 ^a	21.26 $\pm$ 0.04	21.18 $\pm$ 0.05	20.21 $\pm$ 0.07	3.72 $\pm$ 0.37	0.18	0.16	6
HD 178487 ^a	21.36 $\pm$ 0.06	21.22 $\pm$ 0.07	20.50 $\pm$ 0.09	5.22 $\pm$ 0.52	0.28	0.14	6
HD 179406 ^a	21.55 $\pm$ 0.08	21.42 $\pm$ 0.10	20.67 $\pm$ 0.06	0.21 $\pm$ 0.02	0.26	5.50	1
HD 179407 ^a	21.30 $\pm$ 0.06	21.20 $\pm$ 0.07	20.31 $\pm$ 0.07	7.72 $\pm$ 0.77	0.20	0.08	6
HD 185418 ^a	21.41 $\pm$ 0.03	21.19 $\pm$ 0.03	20.70 $\pm$ 0.07	0.78 $\pm$ 0.08	0.39	1.06	6
HD 188001	21.14 $\pm$ 0.07	21.03 $\pm$ 0.08	20.18 $\pm$ 0.08	1.77 $\pm$ 0.15	0.22	0.25	3
HD 190603	21.72 $\pm$ 0.10	21.63 $\pm$ 0.12	20.68 $\pm$ 0.07	2.71 $\pm$ 0.73	0.18	0.62	1
HD 192639 ^a	21.47 $\pm$ 0.06	21.29 $\pm$ 0.09	20.70 $\pm$ 0.07	2.14 $\pm$ 0.21	0.34	0.45	2

Table 2
(Continued)

Star	$\log_{10} N(\text{H})$ (cm^{-2})	$\log_{10} N(\text{H I})$ (cm^{-2})	$\log_{10} N(\text{H}_2)$ (cm^{-2})	d (kpc)	$f(\text{H}_2)$	$\log_{10} n(\text{H})$ (cm^{-3})	H I Reference
HD 197770	21.42 ± 0.12	21.01 ± 0.23	20.90 ± 0.12	0.87 ± 0.02	0.61	0.97	1
HD 198781	21.16 ± 0.04	20.93 ± 0.05	20.48 ± 0.07	0.91 ± 0.04	0.42	0.52	6
HD 199579 ^a	21.25 ± 0.08	21.04 ± 0.11	20.53 ± 0.09	0.92 ± 0.09	0.38	0.62	3
HD 200775	21.53 ± 0.07	21.10 ± 0.10	21.03 ± 0.10	0.36 ± 0.01	0.63	3.09	5
HD 203938	21.70 ± 0.10	21.48 ± 0.15	21.01 ± 0.07	0.22 ± 0.04	0.40	7.42	4
HD 206267 ^a	21.47 ± 0.03	21.22 ± 0.04	20.81 ± 0.05	0.73 ± 0.07	0.44	1.31	6
HD 206773 ^a	21.24 ± 0.04	21.09 ± 0.05	20.42 ± 0.09	0.69 ± 0.07	0.30	0.82	6
HD 207198 ^a	21.51 ± 0.05	21.28 ± 0.06	20.83 ± 0.07	1.04 ± 0.10	0.42	1.02	6
HD 209339 ^a	21.25 ± 0.09	21.16 ± 0.11	20.24 ± 0.08	1.08 ± 0.11	0.19	0.54	1
HD 216898 ^a	21.82 ± 0.19	21.66 ± 0.25	21.02 ± 0.07	0.91 ± 0.09	0.31	2.37	1
HD 239729	21.58 ± 0.10	21.15 ± 0.15	21.07 ± 0.12	0.93 ± 0.03	0.62	1.31	4
HD 326329	21.56 ± 0.12	21.51 ± 0.13	20.27 ± 0.04	1.40 ± 0.10	0.10	0.84	1
HD 332407 ^a	21.35 ± 0.11	21.24 ± 0.14	20.40 ± 0.07	2.56 ± 0.26	0.22	0.28	3
Comparison Stars							
BD+32d270	20.79 ± 0.09	20.78 ± 0.09	18.69 ± 0.04	1.89 ± 0.19	0.02	0.11	3
BD+52d3210	21.47 ± 0.11	21.45 ± 0.11	19.88 ± 0.07	3.18 ± 0.41	0.05	0.30	1
HD 037332	21.58 ± 0.15	21.58 ± 0.14	15.69 ± 0.19	2.19 ± 0.21	0.00	0.56	1
HD 037525	21.44 ± 0.12	21.44 ± 0.12	16.24 ± 1.50	3.04 ± 0.30	0.00	0.29	1
HD 051013	21.08 ± 0.14	21.08 ± 0.13	14.40 ± 0.09	1.61 ± 0.12	0.00	0.24	1
HD 075309	21.18 ± 0.04	21.10 ± 0.04	20.11 ± 0.07	1.20 ± 0.11	0.17	0.41	6
HD 091824 ^a	21.16 ± 0.03	21.12 ± 0.03	19.75 ± 0.07	2.75 ± 0.28	0.08	0.17	6
HD 091983	21.23 ± 0.04	21.15 ± 0.04	20.16 ± 0.09	2.02 ± 0.36	0.17	0.27	6
HD 093028 ^a	21.17 ± 0.13	21.15 ± 0.13	19.50 ± 0.09	2.97 ± 0.30	0.04	0.16	1
HD 094493	21.18 ± 0.09	21.10 ± 0.11	20.13 ± 0.08	0.53 ± 0.01	0.18	0.93	1
HD 097471 ^a	21.39 ± 0.11	21.36 ± 0.11	19.89 ± 0.07	2.78 ± 0.28	0.06	0.29	1
HD 100276 ^a	21.23 ± 0.08	21.19 ± 0.09	19.83 ± 0.06	2.94 ± 0.29	0.08	0.19	3
HD 104705 ^a	21.20 ± 0.04	21.15 ± 0.04	19.95 ± 0.10	4.18 ± 0.42	0.11	0.12	6
HD 114444	21.24 ± 0.09	21.20 ± 0.09	19.83 ± 0.11	0.81 ± 0.03	0.08	0.69	3
HD 116852 ^a	21.02 ± 0.03	20.96 ± 0.03	19.80 ± 0.07	4.88 ± 0.49	0.12	0.07	6
HD 172140 ^a	21.12 ± 0.08	21.11 ± 0.08	19.21 ± 0.13	6.54 ± 0.65	0.02	0.07	3
HD 210809 ^a	21.34 ± 0.04	21.31 ± 0.04	19.92 ± 0.17	3.88 ± 0.39	0.08	0.18	6
HD 235874	21.46 ± 0.11	21.45 ± 0.11	19.64 ± 0.08	1.45 ± 0.20	0.03	0.65	1

Notes. The sources for the $N(\text{H I})$ data are numbered in the rightmost column, and given above. Additional quantities shown: distance d , either photometric distance values from Shull et al. (2021), or values based on Gaia DR2 parallax. Derived molecular fraction $f(\text{H}_2)$. Log of density estimate $N(\text{H})/d$.

^a Distance taken from Shull et al. (2021).

^b No Gaia parallax or photometric distance provided.

References. (1) This work; (2) Cartledge et al. (2006); (3) Diplas & Savage (1994); (4) FM90; (5) Buss et al. (1994); (6) Shull et al. (2021).

(This table is available in machine-readable form.)

multidimensional IDL fitting routine AMOEBA.⁵ Generally, however, the CoG analysis was limited to $J \geq 3$ because the $J \leq 2$ lines were saturated to the extent that they possessed broad damping wings such that the equivalent width measurement algorithm could not reliably distinguish from the stellar continuum. An example spectrum and the resulting H_2 model is shown in Figure 1. The H_2 columns for each of the rotational levels ($J = 0-7$) were summed to give the total column density for each sightline, as listed in Table 2. The table also distinguishes sightlines used as comparison stars in the extinction analysis from the more heavily reddened program stars (GCC09).

2.3. Atomic Hydrogen

Table 2 displays the column densities of total ($N(\text{H})$), atomic ($N(\text{H I})$), and molecular hydrogen ($N(\text{H}_2)$). Wherever it was possible, literature values for $N(\text{H I})$ were adopted, as indicated by the reference indices in the rightmost column of Table 2. For

23 of the sightlines, the available sources only listed $N(\text{H I})$ values inferred from the extinction, using the average $E(B - V)/N(\text{H})$ ratio. These values are not suitable for our analysis, as we need *independent* dust extinction and gas column density measurements to probe the $E(B - V)-N(\text{H})$ relationship and variations in $N(\text{H})/A(V)$.

More accurate $N(\text{H I})$ measurements were made using spectroscopic data and a continuum reconstruction method for the $\text{Ly}\alpha$ absorption line (Bohlin 1975; Diplas & Savage 1994; Gordon et al. 2003). If available, Hubble Space Telescope (HST) data with the Space Telescope Imaging Spectrograph (STIS) and the E140H or E140M echelle grating were used, as these were found to provide the best signal-to-noise ratio and resolution compared to the other options. These data can be found in MAST: doi:10.17909/5qq7-4c81. For most of the stars, the only available UV spectroscopy data were from IUE. For our analysis, we preferred high-resolution IUE data whenever they were available, as these spectra allow us to carefully exclude stellar features. In some cases, we used low-resolution IUE data, either because no other data were

⁵ <https://www.l3harrisgeospatial.com/docs/amoeba.html>

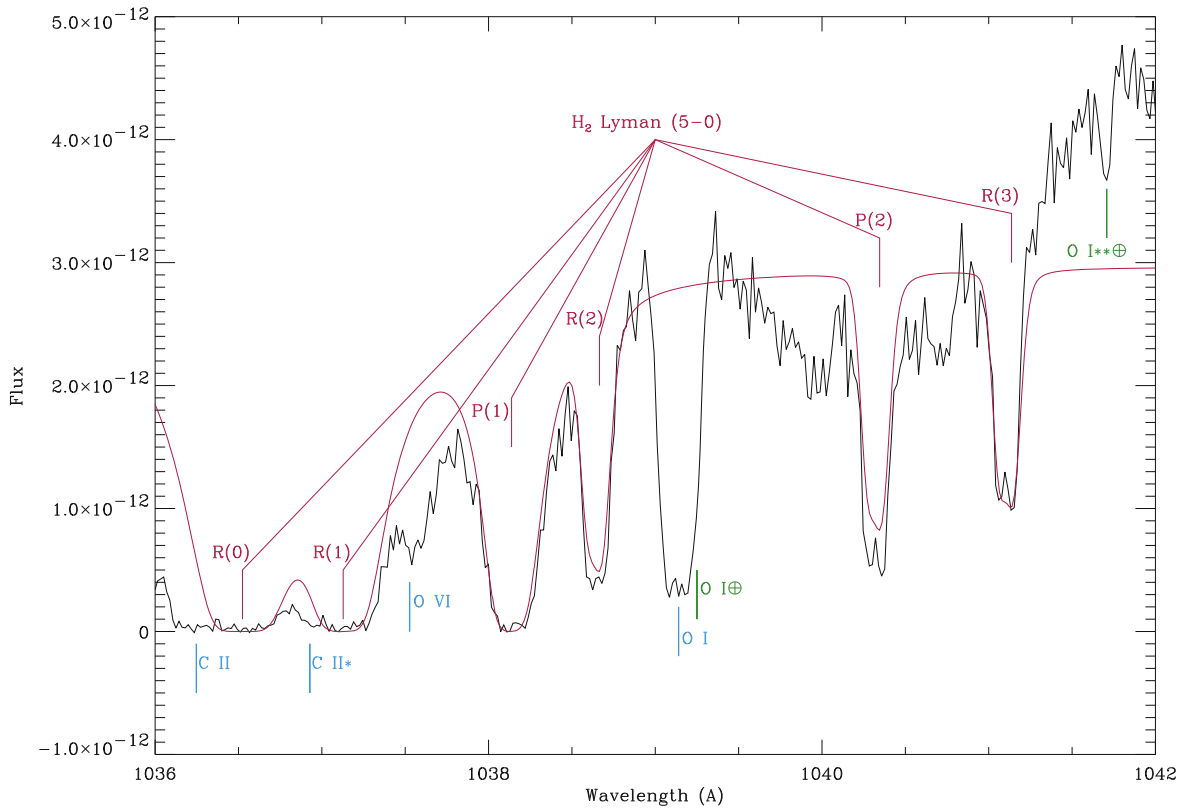


Figure 1. Example of the H_2 absorption line modeling. The flux units are $\text{erg cm}^{-2} \text{s}^{-1} \text{\AA}^{-1}$. The red line shows a transmission model for the Lyman band of H_2 , which consists of transitions from the electronic ground state X , to the excited electronic state B ($2p^1 \Sigma_u^+$). This example model includes contributions by H_2 molecules with rotational quantum numbers $J = 5-0$. The H_2 lines are labeled with the J of the lower state, and a branch name indicating whether J increases by 1 (R branch) or decreases by 1 (P branch) during the electronic transition, as constrained by the selection rules. The absorption lines by several atoms and ions have been indicated in blue for the interstellar gas, and in green with a “ $\oplus$ ” symbol for lines caused by the Earth’s upper atmosphere.

available, or because the high-resolution data were too noisy. If multiple observations or exposures of the same type were available, the individual spectra were resampled onto a common wavelength grid using nearest neighbor interpolation. These resampled spectra were then co-added by taking a weighted average at every point of the common wavelength grid. The weights used at wavelength λ_i , are $n_i/f_i \times t_{\text{exp}}$, with n_i the net count rate (counts per second), f_i the flux (flux unit per second), and t_{exp} the sum of the exposure times. In other words, the data are weighed with a factor of sensitivity $\times$ exposure time because n_i/f_i is the sensitivity (counts per flux unit). In the case of high-resolution spectra, an additional rebinning step was applied, onto a uniformly spaced wavelength grid with a bin width of $0.25 \text{ \AA}$. We found that this bin width provides a good balance between noise reduction and resolution, making it easier to visually inspect the spectra and select suitable wavelength windows, and to run the fitting.

The continuum reconstruction method starts with a linear model fit to estimate the continuum level. The data points used for this fit were chosen by visually inspecting plots of the spectra, and manually selecting suitable wavelength ranges far enough away from the $\text{Ly}\alpha$ line. The model for the $\text{Ly}\alpha$ line is given by $F_{\text{model}}(\lambda_i) = F_{\text{continuum}}(\lambda_i)\phi(\lambda_i)$, where the first factor is the linear continuum model, and $\phi(\lambda)$ is the absorption profile of the $\text{Ly}\alpha$ line as given by Diplás & Savage (1994). This model has one parameter, $\log N(\text{H I})$, which is optimized utilizing a least squares approach to minimize the difference between the observed and model flux for a set of carefully selected data points. The data points chosen for this are

preferably in the $\text{Ly}\alpha$ wings, and not contaminated by any large stellar or interstellar spectral features. The approach for choosing wavelength points, fitting the continuum, and fitting the $\text{Ly}\alpha$ line is analogous to the one presented in Roman-Duval et al. (2019), barring a correction for the line-of-sight velocity.

The noise on the model flux σ_c , is an estimate for the noise due to both observational and instrumental effects, and smaller spectral features, which were not excluded from the fitting range. It is assumed constant over the entire wavelength range, and is estimated by taking the standard deviation of $F_{\text{data}}(\lambda_i) - F_{\text{continuum}}(\lambda_i)$, where i runs over the same wavelength range that was used for the continuum fit. A demonstration of the wavelength range selection and the resulting absorption line model is shown in Figure 2.

The uncertainties shown for $\log N(\text{H I})$ in Table 2, are half of the distance between the 16th and 84th percentiles of the likelihood function $L(\log N(\text{H I})) \propto \exp \chi^2$. This value for the uncertainty depends on the noise estimate σ_c , but should be a reasonable estimate for most data points.

2.4. Sightline Properties

We extended the GCC09 data with accurate measurements of the atomic and molecular column densities $N(\text{H I})$ and $N(\text{H}_2)$, so that the total column $N(\text{H})$, the gas-to-dust ratio $N(\text{H})/A(V)$, and the molecular fraction $f(\text{H}_2)$ could be derived. The extinction curves themselves, and the procedures used to construct them from the raw data and hydrogen abundances, may be found in GCC09. The GCC09 data contains a set of parameters describing the shape of each

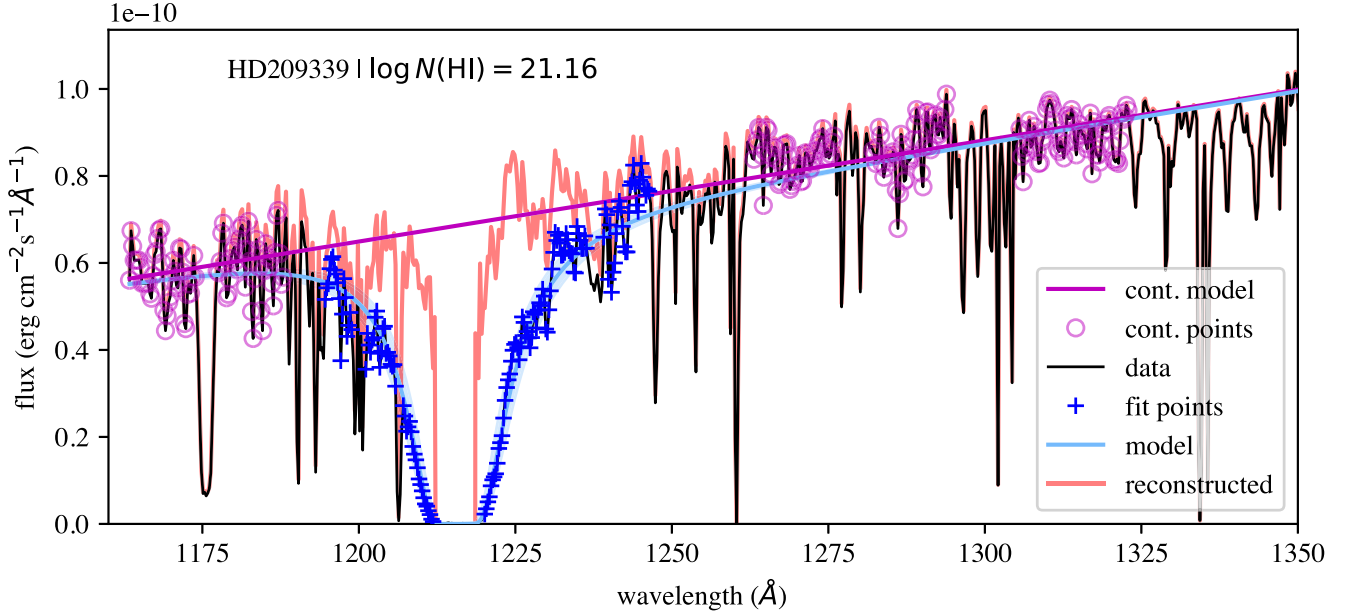


Figure 2. Example Ly α absorption profile fit. Black: ancillary data from HST/STIS. Magenta: data points selected to fit the continuum, and the resulting linear continuum model. Blue: data points selected for Ly α wings, and the resulting line profile. Red: data with Ly α profile divided out, which brings those data back to the continuum level. This is used as a visual aid to judge the quality of the fit.

extinction curve, together with the optical extinction quantities $A(V)$ and $R(V)$ as derived from broadband photometry. The extinction curve shape is represented by the six FM90 parameters, which are found by fitting the measured extinction for each wavelength $E(\lambda - V)/E(B - V)$ with the functional shape given by FM90. Note that we use slightly different parameters because GCC09 fitted the extinction curve model to $A(\lambda)/A(V)$, instead of the customary $E(\lambda - V)/E(B - V)$. Aside from this minor distinction, the model equation remains identical to Equation (2) of FM90.

We denote the FM90 parameters as $c_1^{A(V)}$, $c_2^{A(V)}$, $c_3^{A(V)}$, $c_4^{A(V)}$, x_0 , and γ . Each parameter corresponds to a certain aspect of the extinction curve. As illustrated in Figure 3, the parameters $c_1^{A(V)}$ and $c_2^{A(V)}$ describe the linear portion of the curve. The nonlinear features added to this linear base are the FUV rise ($c_4^{A(V)}$) and the 2175 Å bump ($c_3^{A(V)}$), for which x_0 is the central wavelength and γ is the width. The bump amplitude can be derived from these parameters, and is equal to $c_3^{A(V)}/\gamma^2$ (FM90), and the area of the bump is $\pi c_3^{A(V)}/2\gamma$. By evaluating the functional form of the extinction curve using the fitted FM90 parameters, we can obtain extinction ratios at specific UV wavelengths. We use this to focus on the UV-to-optical extinction ratio $A(1000)/A(V)$ in the range of the FUV rise, and $A(2175)/A(V)$ at the bump center (wavelength between parenthesis in Å), in addition to the optical quantity $R(V)$.

Next to the $N(\text{H I})$ and $N(\text{H}_2)$ data measured and collected in this work, Table 2 also lists a few derived quantities, such as the total gas column density and the molecular fraction. They are defined as $N(\text{H}) = N(\text{H I}) + 2N(\text{H}_2)$, and $f(\text{H}_2) = 2N(\text{H}_2)/N(\text{H})$. For completeness, we also calculated a set of rough estimates for the total number density $n(\text{H})$ (cm^{-3}), by dividing the calculated column densities by the distance d . Photometric distance values from Shull et al. (2021) were used for the 39 stars common with our sample. For the

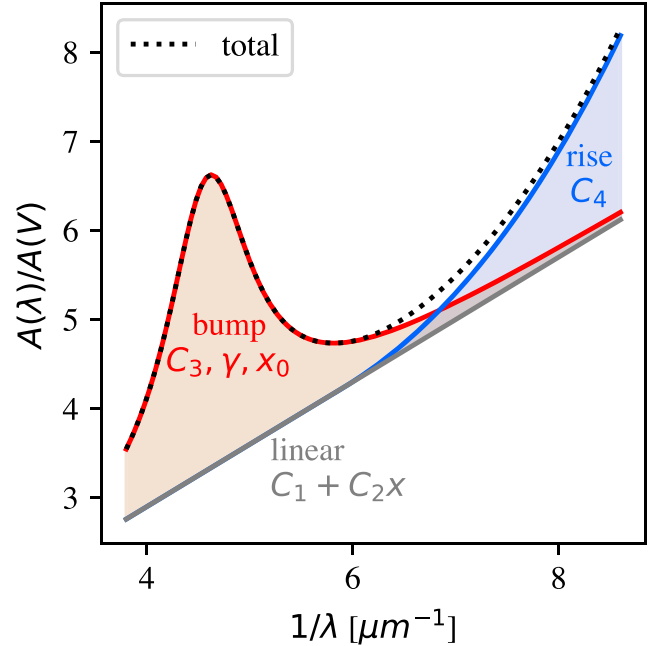


Figure 3. Sketch of the FM90 extinction curve model (their Equation (2)), with arbitrary parameters. The functions describing the 2175 Å bump and the FUV rise are added on top the linear part of the extinction model.

other stars, the distances were derived from parallax values provided by Gaia DR2 (Gaia Collaboration et al. 2018).

The rotational temperature T_{01} of H_2 is derived using the Boltzmann equation and the ratio of the column densities of H_2 in its first two rotational states $J = 0$ and $J = 1$ (Rachford et al. 2009). We use $N(0)$ and $N(1)$ from Table 1.

$$T_{01} = \frac{74.0}{\log N(0) - \log N(1) + 0.954}. \quad (1)$$

When the balance between the populations of these states is driven by thermal collisional transitions, then the rotational temperature is a good measure for the kinetic temperature of the gas where H_2 is present.

2.5. Uncertainties and Covariance

2.5.1. Covariance Estimation for Derived Quantities

In certain plots of the results presented in Section 3, relationships between two derived quantities are shown, some of which depend on a common factor. A common factor can introduce correlations between the quantities shown on the x and y axes. If the uncertainty in the common factor is large, the resulting correlations might affect our conclusions about the sample. To be clear, this section concerns correlations between different quantities for the same sightline due to the nature of the measurements, and not correlations between different sightlines.

To take these correlations into account, an estimate for the covariance is needed. We calculate the covariance between two derived quantities using the standard, first order expansion method. For three stochastic variables x , y , z , and two functions $f(x, z)$ and $g(y, z)$, which both have a dependency on z , the covariance for a data point x_i , y_i , z_i can be estimated using

$$\text{cov}(f(x_i, z_i), g(y_i, z_i)) = \sigma_{z,i}^2 \frac{\partial f}{\partial z}(x_i, z_i) \frac{\partial g}{\partial z}(y_i, z_i). \quad (2)$$

As a practical example, we explain the situation when $N(\text{H})/A(V)$ and $1/R(V)$ are compared. In the work by GCC09, $A(V)$ was derived by extrapolating photometric measurements of $E(J - V)$, $E(H - V)$, and $E(K - V)$. We therefore consider $A(V)$ and $E(B - V)$ to be independent, while the $R(V)$ measurement is not independent because it was derived using $1/R(V) = E(B - V)/A(V)$. The common factor between $N(\text{H})/A(V)$ and $1/R(V)$ is $A(V)$, so the covariance is given by

$$\text{cov}\left(\frac{N(\text{H})}{A(V)}, R(V)^{-1}\right) = N(\text{H})E(B - V) \frac{\sigma_{A(V)}^2}{A(V)^4}. \quad (3)$$

In this example, $N(\text{H})$ is independently measured from $A(V)$ and $E(B - V)$, so it does not contribute a correlation term. Technically, $A(V)$ and $E(B - V)$ are correlated because $E(B - V) = A(B) - A(V)$, but since the uncertainty is dominated by $A(B)$, we consider them to be uncorrelated for this example.

In certain plots of this work, the standard deviations and covariances calculated via this method are illustrated using ellipses. The length and orientation of the axes of each ellipse was determined by calculating the eigenvalues and eigenvectors of the covariance matrix for each point, given by

$$C_i = \begin{pmatrix} \sigma_{x,i}^2 & \text{cov}(x_i, y_i) \\ \text{cov}(x_i, y_i) & \sigma_{y,i}^2 \end{pmatrix}. \quad (4)$$

where the elements were obtained using the standard error propagation methods described above.

2.5.2. Linear Fitting with Covariance

We report the best-fitting slope and intercept of a linear model for certain parameter pairs. In all cases, the measurement

uncertainties are substantial for both x and y , and in one case, they are correlated. A simple weighted regression with error bars in the y direction cannot be used, as it would ignore the uncertainties in the x direction, and the xy correlations. To include these aspects of the uncertainty in the calculation of the best-fitting line, a likelihood function was set up that is based on the perpendicular distance of each data point from the linear model. From the xy covariance, an uncertainty in the direction perpendicular to the line is derived by performing a projection onto the normal of the line. This perpendicular uncertainty determines how many standard deviations a data point deviates from the linear model. The reasoning and equations behind this likelihood function are given in Hogg et al. (2010) and Robotham & Obreschkow (2015).

The best-fitting line is obtained by maximizing this 2D likelihood function, $L(m, b)$, where m and b are the slope and intercept of the line. To estimate the uncertainty on these parameters, the likelihood function was evaluated on a grid of points around the solution, and then used as a probability density function to draw 2000 random (m, b) pairs. The uncertainties reported in this work are the standard deviations of these sampled m and b values.

2.5.3. Sample Correlation Coefficient and Significance

For every scatter plot in this work, quantify to what degree the observed quantities are correlated, and how significant the correlation is. For this reason, we report values for the Pearson correlation coefficient r , and a level of significance expressed as a number times a certain σ . The paragraphs below clarify how we calculate these indicators while including the general xy uncertainties described above.

Uncorrelated Uncertainties—The significance level of a certain value for r is defined as the probability that an uncorrelated data set produces a value of r that is as large as the one observed, solely because of statistical fluctuations. In a previous study of the correlations between $f(\text{H}_2)$ and the FM90 parameters by Rachford et al. (2002) the significance was expressed as a numerical multiplier times a certain σ , e.g., “3.4 σ .” When determining the significance that a correlation coefficient is nonzero, this σ relates to the r distribution under the null hypothesis that the underlying physical data are fully uncorrelated.

The commonly used analytical equations for the significance of r only work when the underlying distribution of the data is bivariate normal. Because the sightlines in our sample were selected only if a high-quality extinction curve could be obtained, the distribution for the measured quantities is non-normal. Therefore, we perform a random permutation test instead, a resampling technique commonly used for dealing with non-normal data. Permutation testing is a general technique for testing hypotheses that relate to certain orderings or associations of the data points. The main principle is choosing a test statistic (such as r), and seeing how it behaves when the labels of the data are scrambled. See the book by Good (1994) or blog posts (e.g., <https://towardsdatascience.com/bootstrapping-vs-permutation-testing-a30237795970>) for use cases and examples. In this case, a nonzero value for r is expected if the measurements x_i and y_i each belong to the same sightline, and the underlying physics cause the quantities x and y to be correlated. Scrambling the labels turns the set of dependent measurements x_i , y_i into one of independent measurements x_j , y_i , thereby breaking the correlation, resulting

in a population that has $\rho=0$ by construction, or an r distribution centered around 0 when this population is sampled.

For each pair of parameters displayed in one of the scatter plots, we take the effect of the uncertainties on r into account by redrawing the scrambled data sets many times in a Monte Carlo fashion. Each of the 75 data points is redrawn from a bivariate normal distribution, of which the parameters are defined according to the estimated covariance matrix for that point. A sample of 6000 r values is created by calculating the regular Pearson correlation coefficient for each realization of the data set generated this way. This results in a sample of coefficients r_0 , centered around zero, and the sample standard deviation $s(r_0) = \sqrt{\sum_i r_{0,i}^2}$ is the value we use to express the confidence as a number of σ . The displayed significance level in the corner of each plot is then $r_{\text{measured}}/s(r_0)$.

Correlated Uncertainties—Even when the underlying physical data are uncorrelated, a strong correlation in the measurement errors of x and y will lead to an r_0 distribution that is no longer centered around zero. We call this bias an *induced* correlation, as it is intrinsic to the correlated measurements, and not correlations in the underlying physics. To adjust the significance level, we need to generate a biased sample of r_0 values, so that the expression for the significance becomes $[r_{\text{measured}} - \text{median}(r_0)]/s(r_0)$.

To generate this sample, it is crucial that the simulated (correlated) measurement noise is applied *after* the scrambling step described in the method above. This generates a mock observation of the sample, where the measurement noise induces a nonzero correlation into the inherently uncorrelated (scrambled) sample. Generating an appropriate shift for each data point, to simulate the measurement noise, is not trivial, because the scrambling makes it unclear what should happen with the covariance matrix for each data point. As an approximation, we randomly assign the covariance matrices to the data points, and then draw the simulated measurement noise according to the corresponding bivariate normal distributions. The shift in the average of r_0 , induced by adding this mock data for the correlated measurement noise, is demonstrated in Figure 4.

3. Results

3.1. Gas Column Versus Dust Extinction

3.1.1. Total Column—Optical Extinction

The correlations between the gas column densities $N(\text{H})$, $N(\text{H I})$, and $N(\text{H}_2)$, and the dust extinction in the V band $A(V)$ and in the UV $A(1000)$, are shown in Figure 5. The latter extinction was calculated by evaluating the FM90 extinction curve model $A(\lambda)/A(V)$ at 1000 Å. This wavelength was chosen because it is in the UV absorption/dissociation band of H_2 . Since we are taking this value from a fitted, smooth extinction model, the data for all wavelengths near 1000 Å is taken into account. Choosing any value between 912 and 1150 Å will result in the same conclusions.

The total H column $N(\text{H})$ correlates best with the V-band extinction $A(V)$. It is interesting to note that Butler & Salim (2021) found that $A(\sim 3000)$ correlates better with $N(\text{H})$, but this does not significantly affect the discussion and conclusions in this work. A measurement of the average gas-to-extinction ratio for the sample is obtained by fitting a linear model to the $N(\text{H})$ - $A(V)$ relationship. Because the uncertainties on the data are substantial for both the $N(\text{H})$ and $A(V)$ axes, we use the

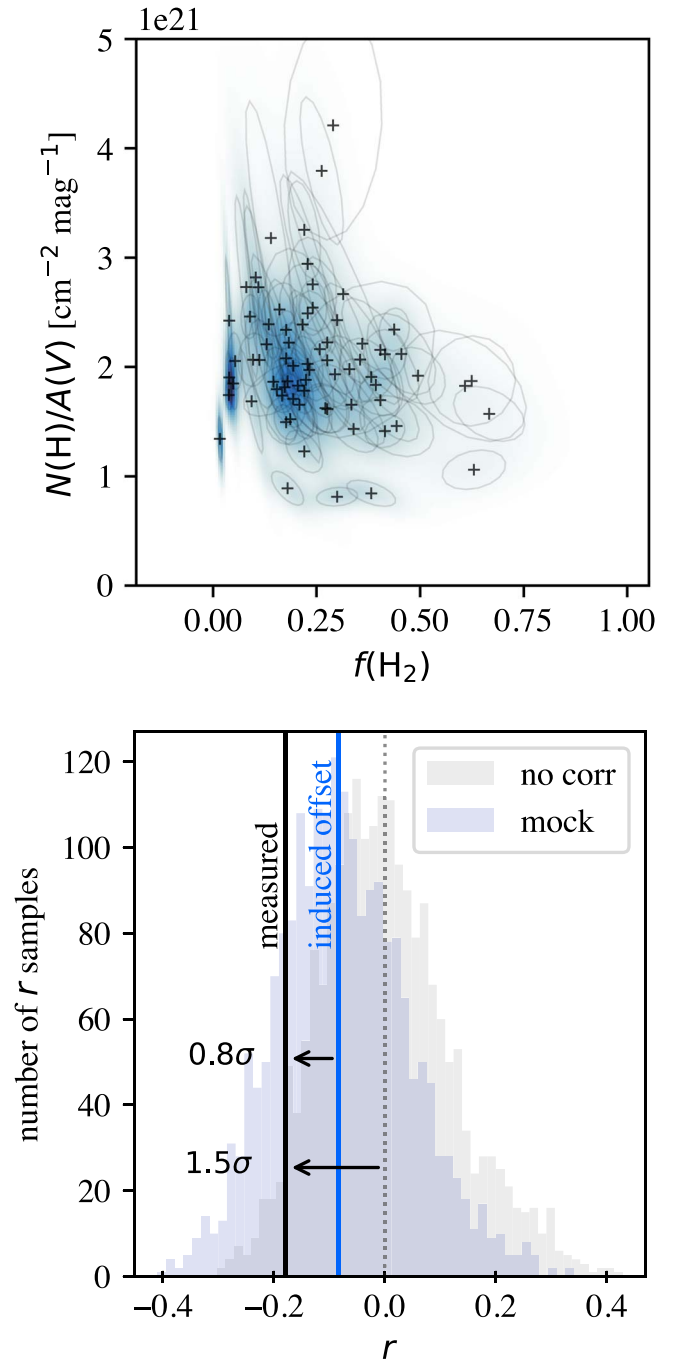


Figure 4. Demonstration of our method for dealing with correlated uncertainties. The top panel shows a scatter plot between $N(\text{H})/A(V)$ and $f(\text{H}_2)$, and the ellipses and density map visualize the correlations due to the common $N(\text{H I})$ and $N(\text{H}_2)$ factors in both quantities. The gray and blue histograms show the sample of correlation coefficients r before and after applying mocked correlated measurement errors to the scrambled samples. The two arrows and the displayed σ values, show how the significance level of the measured correlation changes, when the offset induced by the correlated measurement uncertainties is taken into account.

least squares linear model described in Section 2.5.2. The results for the linear fits performed throughout this paper are summarized in Table 3.

For $N(\text{H})$ versus $A(V)$ we find a slope of about $(1.6 \pm 0.1) \text{ cm}^{-2} \text{ mag}^{-1}$, which is significantly lower than the average $N(\text{H})/A(V)$ ratio obtained from observations of X-ray sources, which is around $2 \times 10^{21} \text{ cm}^{-2} \text{ mag}^{-1}$ (Reina & Tarengi 1973; Gorenstein 1975; Predehl & Schmitt 1995;

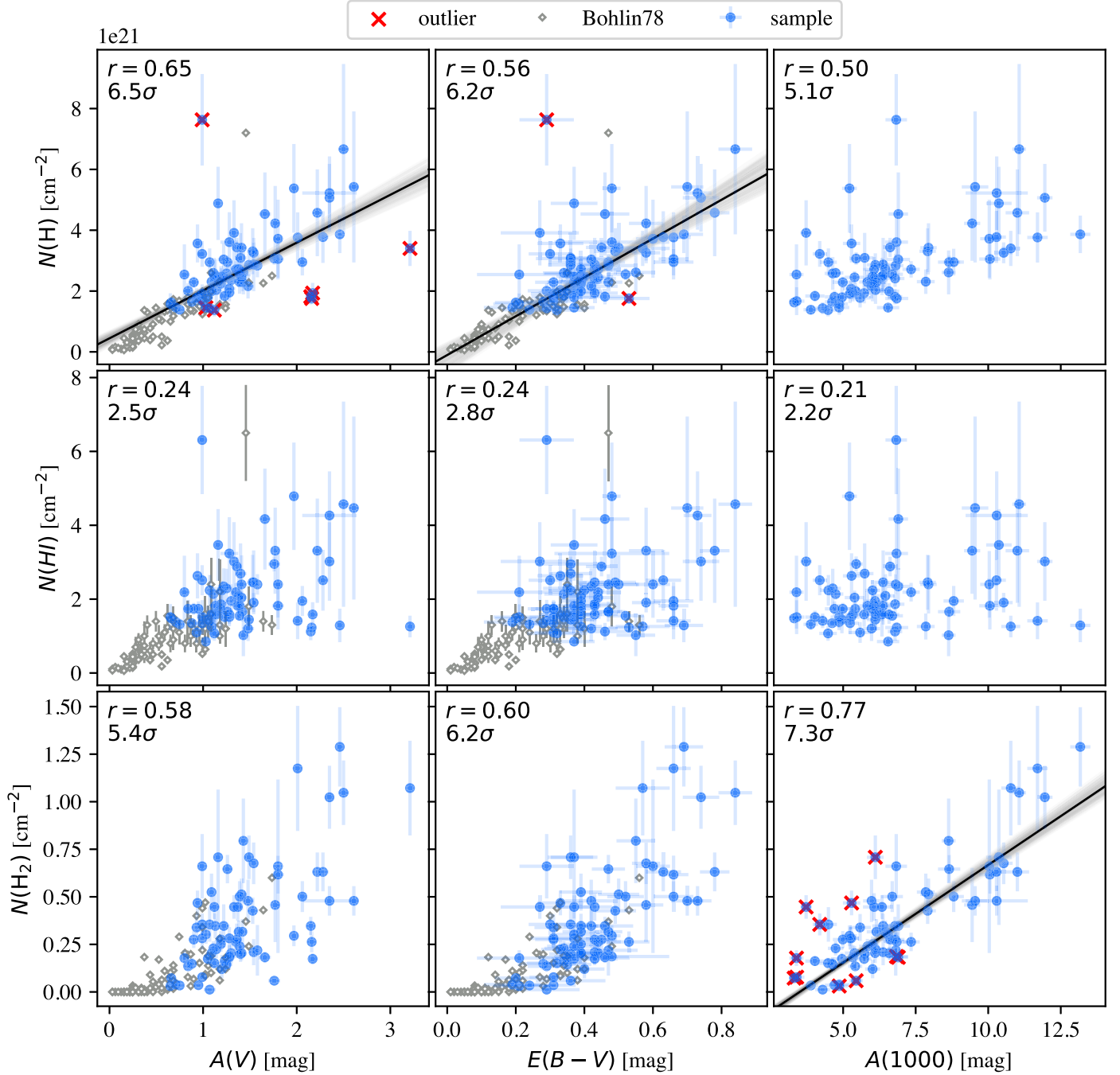


Figure 5. Total hydrogen, atomic hydrogen, and molecular hydrogen column densities vs. dust column, measured via the extinction $A(V)$, $E(B - V)$ and $A(1000)$. Our sample of 75 stars is shown using blue circles, and the error bars represent the 1σ uncertainty on the measurements. The original *Copernicus* sightlines by Bohlin et al. (1978) are shown for reference, using the gray diamonds. The black lines illustrate the fit results from the method described in Section 2.5.2, with the gray auras showing the uncertainty on each fit. The numbers in each corner show the Pearson correlation coefficient of the data and its significance, as explained in Section 2.5.3. Data that deviate more than 3σ from the fitted line, are indicated with a red “x.” They have been excluded from the fit for $N(\text{H})$ vs. $A(V)$, but not from the fit for $N(\text{H}_2)$ vs. $A(1000)$. In the latter plot, they are only highlighted to show the intrinsic width of the $N(\text{H}_2)$ vs. $A(1000)$ relationship.

Güver & Özel 2009; Zhu et al. 2017). Examining the work by Zhu et al. (2017), the X-ray data mainly cover the $N(\text{H})$ range from 2×10^{21} to $2 \times 10^{22} \text{ cm}^{-2}$, or $A(V) = 1$ to 10, while our data are in the $N(\text{H})$ range between 2×10^{21} and $6 \times 10^{21} \text{ cm}^{-2}$. According to the above data, the $N(\text{H})/A(V)$ ratio seems to be systematically lower in the more diffuse column density regime covered by this work, with $A(V) < 3$, compared to the average over the entire $A(V)$ range from 1–10. Applying our fitting method to $N(\text{H})$ versus $E(B - V)$, yields a

value that is slightly higher than, but consistent with the classic gas-reddening ratio $N(\text{H I})/E(B - V) = 5.8 \times 10^{21} \text{ cm}^{-2}$ from Bohlin et al. (1978), and similar to the values reported by others since then (Rachford et al. 2002, 2009; Shull et al. 2021).

3.1.2. H_2 Column—UV Extinction

We expect to see a relationship between $N(\text{H}_2)$ or $f(\text{H}_2)$ and the extinction measured at wavelengths within the H_2

Table 3Slope m and Intercept b of the Best-fitting Linear Model, for Several Parameter Pairs (x, y) Displayed in Figures 5, –7

x -axis	y -axis	Slope m ($10^{20} \text{ cm}^{-2} \text{ mag}^{-1}$)	Intercept b (10^{20} cm^{-2})
$A(V)$	$N(\text{H})$	15.73 ± 0.98	4.5 ± 1.1
$E(B - V)$	$N(\text{H})$	63.8 ± 5.8	-1.0 ± 2.3
$A(1000)$	$N(\text{H}_2)$	1.027 ± 0.041	-3.60 ± 0.22
$A(1000)_{\text{rise}}$	$N(\text{H}_2)$	2.20 ± 0.09	-0.95 ± 0.12
$A(2175)_{\text{bump}}$	$N(\text{H}_2)$	3.33 ± 0.17	-1.76 ± 0.20
x -axis	y -axis	Slope m ($10^{20} \text{ cm}^{-2} \text{ mag}^{-1}$)	Intercept b ($10^{20} \text{ cm}^{-2} \text{ mag}^{-1}$)
$1/R(V)$	$N(\text{H})/A(V)$	101.3 ± 8.6	-12.6 ± 2.6
$A(1000)/A(V)$	$N(\text{H})/A(V)$	3.19 ± 0.22	2.54 ± 0.97
$A(2175)/A(V)$	$N(\text{H})/A(V)$	9.43 ± 0.67	-10.7 ± 2.0

dissociation band. Therefore, we have calculated $A(1000)$ from the parameterized extinction curves, and show its correlation with the gas contents in the third column of Figure 5.

Unlike $A(V)$ and the total gas density, the relationship between $A(V)$ and $N(\text{H}_2)$ exhibits a much larger scatter. However, in the lower right panel of Figure 5, a well-defined relationship between $A(1000)$ and $N(\text{H}_2)$ is discovered, with a strong linear correlation coefficient (~ 0.8). Fitting a slope using the same method yields the result in Table 3, and it can be derived that the fitted line intersects the x -axis at $A(1000) = 3.4$. This value can be interpreted as a minimum magnitude of UV extinction, required to observe a significant amount of H_2 along the sightline, although the marked outliers in Figure 5 make it clear that there is still a significant spread on the H_2 column density. For sightlines with $A(1000) \leq 7$, many data points deviate more than 3σ from the best-fitting line, showing that there is a significant amount of scatter on this relationship. For $A(1000) \geq 7$, the error bars on $N(\text{H}_2)$ are too large to discern if this scatter persists.

Small grains absorb primarily at short wavelengths, while larger grains also absorb efficiently at longer wavelengths such as in the V band. We will refer to the FUV-absorbing population as *small grains* in the rest of this work. As $A(1000)$ does not correlate with $N(\text{H I})$, these small grains appear to be present only in the molecular regime, where they contribute to a large fraction of the FUV extinction. The relationship between $N(\text{H}_2)$ and $A(V)$ has more scatter because the larger grains also appear in the H I gas.

The FM90 extinction curve model gives us a way to separate the contribution to $A(1000)$ by the FUV rise specifically, from the total extinction. The *absolute* contribution by the FUV rise is given by $c_4^{A(V)} F(x) A(V)$, with $F(x)$ as defined in FM90. By using this quantity in Figure 6, we find a correlation with $N(\text{H}_2)$ that is even stronger, at almost 0.9. On the other hand, the FUV rise term contributes to at most 40% of the total FUV extinction. After subtracting A_{rise} , the residual extinction shown as the *rest* in Figure 6, still has a noticeable correlation of 0.63, but it is also the main source of the scatter on the $A(1000) - N(\text{H}_2)$ relationship. Most notably, the FUV rise contribution is consistent with a relationship that intersects the origin, while the residual contribution is not. For both components, no correlation with $N(\text{H I})$ was observed. These findings impose a strong constraint on the nature of the FUV rise carrier: in the diffuse Galactic ISM, it only resides in molecular gas.

An analogous approach was taken to investigate the absolute contribution of the NUV bump to $A(2175)$. For this component, the results are not as clear cut. In fact, the *rest* term exhibits a stronger correlation with $N(\text{H}_2)$ than the *bump* term $A_{\text{bump}} = c_3^{A(V)} \gamma^{-2} A(V)$: the coefficients are 0.75 and 0.56, respectively. Given this difference in behavior between the FUV rise and NUV bump, it is unlikely that both of these features are produced by the same carrier. However, it remains possible that the two carriers have a common origin in terms of physical conditions and formation processes, considering that the behavior with respect to $N(\text{H}_2)$ is somewhat similar at low column density. At higher column densities, it seems that the NUV bump amplitude stops increasing, while that of the FUV rise keeps following the trend. Further interpretation and possible causes for the observed relationships are discussed in Section 4.1.

3.2. Gas-to-dust Ratio versus Dust Extinction Ratios

The $N(\text{H})/A(V)$ data have a weighted average of $1.6 \times 10^{21} \text{ cm}^{-2} \text{ mag}^{-1}$, which is equal to the fitted $N(\text{H}) - A(V)$ slope in Section 3.1.1. There are variations in this ratio, but for most of the quantities we measured, we find no clear trend with the variations in $N(\text{H})/A(V)$. Most notably, $N(\text{H})/A(V)$ does not seem to show any systematic variation with the gas properties $f(\text{H}_2)$ or T_{01} . The top panel of Figure 4, which demonstrated our method to determine the significance of the observed correlations, shows this result for $N(\text{H})/A(V)$ versus $f(\text{H}_2)$. For the rest of these negative results, we do not show plots in this paper.

There is no systematic variation of $N(\text{H})/A(V)$ with the column density $N(\text{H})$ or number density $n(\text{H})$. This result was to be expected, as earlier studies of the dust-to-gas ratio showed that most of the variations occur below $\log N(\text{H}) = 21.5 [\text{cm}^{-2}]$ (Roman-Duval et al. 2021). Above this density, most of the metals have been converted into dust mass, and the dust-to-gas ratio as a function of column density forms a plateau (C. J. R. Clark et al. 2023, in preparation). There are also no significant correlations of $N(\text{H})/A(V)$ with the dust columns $A(V)$ or $A(1000)$, or the FM90 bump parameters γ and $c_3^{A(V)}$, or the FUV rise parameter $c_4^{A(V)}$. However, we did discover a relationship between the gas-to-dust ratio and certain quantities related to the extinction curve shape.

The top row of Figure 7 compares $N(\text{H})/A(V)$ with three relative measures of the extinction: $1/R(V) = E(B - V)/A(V)$, $A(1000)/A(V)$, and $A(2175)/A(V)$. There is a positive correlation between $N(\text{H})/A(V)$ and both $A(1000)/A(V)$ and $A(2175)/A(V)$ at around the 3σ level, meaning that sightlines with a stronger UV extinction compared to the optical, also have a lower amount of optical extinction per gas unit. For the optical quantity $1/R(V)$, there seems to be a similar trend when judged with the human eye, but the error bars are too large to have a significant correlation according to our metric. In Figure 7, the four points corresponding to HD 045314, HD 164906, HD 200775, and HD 206773 have been indicated with red squares because they have a markedly low $N(\text{H})/A(V)$ ratio. The best-fitting slope for $N(\text{H})/A(V)$ was calculated relative to the three parameters $1/R(V)$, $A(1000)/A(V)$, and $A(2175)/A(V)$, and each time the resulting line goes right through this group of four points. Even when these low- $N(\text{H})/A(V)$ points are excluded from the data used for fitting, the fitted line still passes through that group. In other words, their $1/R(V)$ values deviate from the Galactic average of $1/3.1 = 0.32$, in a manner that is consistent

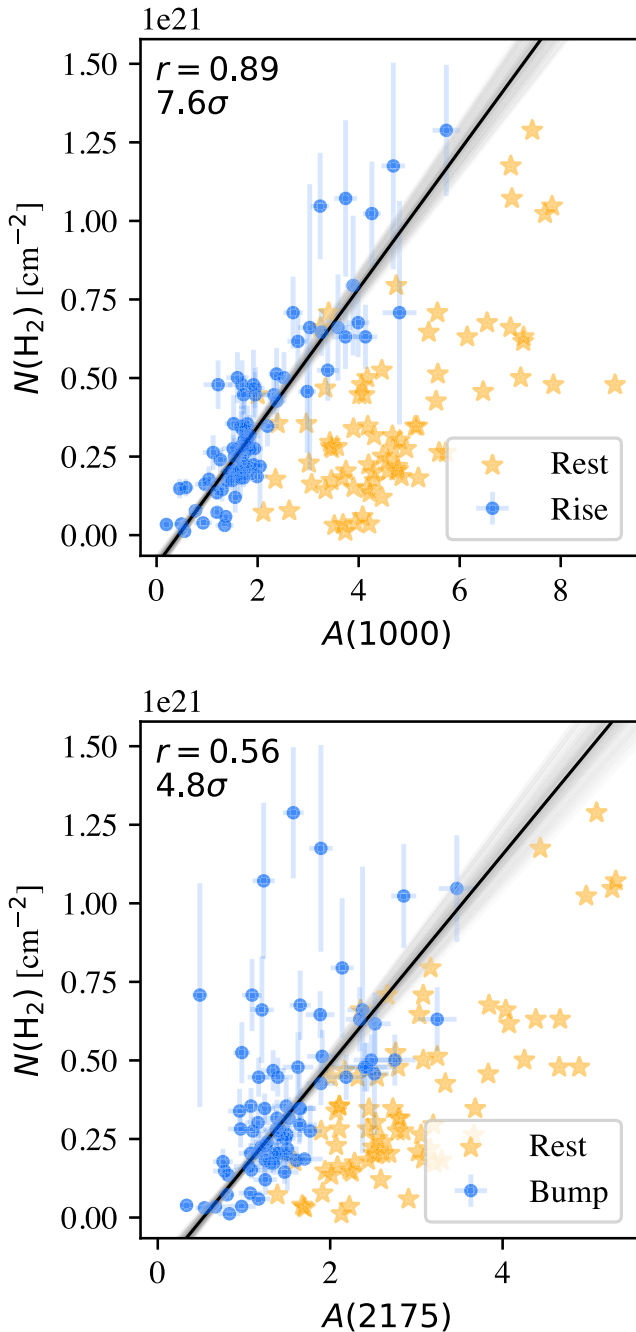


Figure 6. Upper panel: a different view of the $N(\text{H}_2)$ – $A(1000)$ correlation, where we separately show the contribution of the FUV rise $A_{\text{rise}} = c_4^{A(V)} F(x) A(V)$ (blue circles) and the remainder of the extinction $A_{\text{rest}} = A(1000) - A_{\text{rise}}$ (orange stars). The observed $N(\text{H}_2)$ correlation of 0.89 is even higher than the one observed for the total value of $A(1000)$. The remaining $A(1000)$ extinction still correlates with $N(\text{H}_2)$ but with a coefficient of 0.63. The FUV rise contribution approaches zero when $N(\text{H}_2)$ is small while the rest does not. Lower panel: analogous view of the extinction at 2175 Å, separating the contribution of the bump from the rest. While the bump contribution still nears zero at small $N(\text{H}_2)$, it has a weaker correlation of only 0.56. The rest contribution correlates the strongest in this case, with $r = 0.75$. The fit results of both panels are shown in Table 3.

with a trend fitted to the rest of the sample. The ratios $A(1000)/A(V)$ and $A(2175)/A(V)$ show similar behavior, which is expected due to their connection with $1/R(V)$, as shown in Figure 7 of GCC09. The observed trend could be interpreted as the effect coagulation, which increases the amount

of dust contributing to $A(V)$ (decreasing $N(\text{H})/A(V)$), while reducing the relative amount of small grains (decreasing $A(1000)/A(V)$).

The average $N(\text{H})/A(V)$ of our sample is lower than the Galactic average (see Section 3.1.1), because of the sample selection, and the corresponding $N(\text{H})/A(V)$ distribution. Instead of being statistically representative of the entire population of Milky Way sightlines, the sample focuses on including a broad range of $R(V)$ values. By evaluating the fit for $N(\text{H})/A(V)$ versus $1/R(V)$ at the Galactic average $1/R(V) = 0.32$, we can adjust the average $N(\text{H})/A(V)$ of our sample for the difference in $R(V)$, and obtain an estimate for the average of $N(\text{H})/A(V)$ as if the average $R(V)$ were equal to 3.1. With the values shown in Table 3, the result is $\langle N(\text{H})/A(V) \rangle = 0.32m + b = (1.99 \pm 0.05) \times 10^{21} \text{ cm}^{-2} \text{ mag}^{-1}$, and the error on this result was calculated by taking the standard deviation of $0.32m + b$, over a sample of (m, b) pairs that was drawn as explained in Section 2.5.2. This value is consistent with the literature value of $\sim 2 \times 10^{21} \text{ cm}^{-2} \text{ mag}^{-1}$ (Bohlin et al. 1978; Diplaz & Savage 1994; Zhu et al. 2017), showing that the difference in $R(V)$ between our sample and the total Galactic sightline population also explains the difference in $N(\text{H})/A(V)$.

3.3. H_2 Fraction versus Extinction Curve Shape

The sightlines have a wide variety of molecular fractions, ranging from 0.02 to nearly 0.7, with most above 0.1. The bottom row of Figure 7 shows how the molecular fraction $f(\text{H}_2)$ relates to $1/R(V)$ and two UV-to-optical extinction ratios. The plot for $A(2175)/A(V)$ has a structure similar to the one with $1/R(V)$, while the cloud of data points look somewhat different for $A(1000)$. This is expected as GCC09 already showed that $A(2175)/A(V)$ strongly correlates with $1/R(V)$, while the $R(V)$ -dependent relationship for $A(1000)$ shows more scatter. Earlier works that examined $f(\text{H}_2)$ versus $R(V)$ found a possible correlation, but it was of limited significance (Cardelli 1988; Rachford et al. 2002, 2009). For example, in Cardelli (1988), $N(\text{H}_2)$ was measured for a number of diffuse ($A(V) < 2$) sightlines, and $f(\text{H}_2)$ was found to decrease going from $R(V) \sim 3$ to $R(V) \sim 4$. For our data, the calculated correlation coefficient for $f(\text{H}_2)$ is very weak for all three extinction ratios because there are many points that deviate from the main cluster of points centered around $1/R(V) = 0.3$. It should also be noted that the four low- $N(\text{H})/A(V)$ outliers, again indicated with red squares, lie distinctly separate from the main group of points, with a wide range of $f(\text{H}_2)$ values. By excluding these points from the data, the correlation coefficient increases to 0.3, with a significance of 2.9 for $f(\text{H}_2)$ versus $1/R(V)$ and 2.3 for both UV-to-optical extinction ratios. Whether this correlation is observed, therefore depends on the subset of the ISM being probed.

In addition to the total extinction in Figure 5, and the extinction ratios in Figure 7, we also compare the specific features from the FM90 parameterization with the molecular contents. In Figure 8, we present plots comparing $f(\text{H}_2)$ to the six FM90 parameters: $c_1^{A(V)}$, $c_2^{A(V)}$, $c_3^{A(V)}$, $c_4^{A(V)}$, x_0 , and γ . The parameters describing the linear aspect of the extinction curve, $c_1^{A(V)}$ and $c_2^{A(V)}$, do not show any trends with $f(\text{H}_2)$, nor does the center of the FUV bump x_0 . The main trends in $f(\text{H}_2)$ relate to the parameters for the magnitude and width of the 2175 Å bump ($c_3^{A(V)}$, γ) and the strength of the FUV rise ($c_4^{A(V)}$). Considering the results for the column density presented in Section 3.1, these results are just a different view of the same

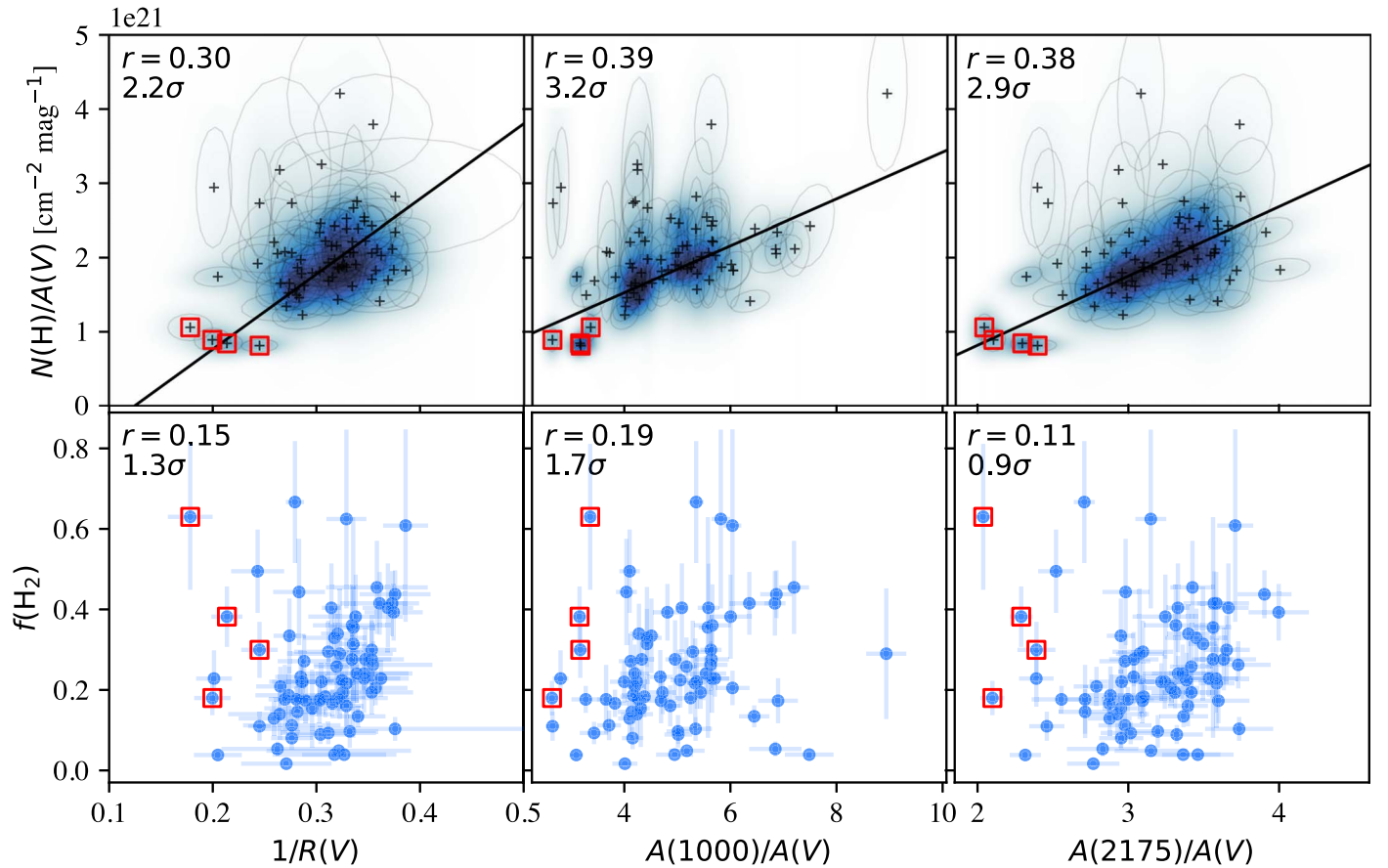


Figure 7. Gas-to dust ratio and molecular fraction vs. three extinction ratios. The four low- $N(\text{H})/A(V)$ sightlines toward HD 045314, HD 164906, HD 200775, and HD 206773 have been indicated with red squares. HD 096675 has an exceptionally high value for $N(\text{H})/A(V)$, and has been excluded from these plots and the analysis. Due to the common factor $A(V)$ for the quantities shown on the axes in the top row, the uncertainties are correlated in x and y . The shapes of the correlated uncertainties are illustrated by the ellipses, and the color map shows where the data are concentrated, by showing each point as a 2D Gaussian probability density. The correlations in the uncertainty are generally small, and do not strongly affect the conclusions. The scrambling and resampling test described in Section 2.5.3 estimates that the correlation coefficient is boosted by about 0.03 by the correlated x and y measurement uncertainties.

relationship. However, this view allows us to compare these results to previous work. Similar results were found by Rachford et al. (2002), who found a correlation at the 2.6σ level for c_4 , and one of 3.7σ with γ . In this work, the correlation with $c_4^{A(V)}$ is the more significant one at 3.3σ , while the one with γ is at 2.7σ .

As in the other figures, the red squares in Figure 8 indicate the same four low- $N(\text{H})/A(V)$ sightlines, which were found grouped together in the $N(\text{H})/A(V)$ versus $1/R(V)$ plot of Figure 7. In this case, it is interesting that their $c_3^{A(V)}$ and $c_4^{A(V)}$ parameters are on the lower side of the sample, while their molecular fractions are very different. So while the dust in these four sightlines seems to have evolved in a similar way, the observed molecular fraction does not seem to play a role in this phenomenon.

3.4. Density and Rotational Temperature

We investigated a variety of relationships between the extinction properties, the rotational temperature of H_2 , and the density estimates $n(\text{H}) = N(\text{H})/d$ as calculated according to Section 2.4. Rachford et al. (2009) already showed that at higher temperatures, the medium along Milky Way sightlines generally is of lower density and molecular fraction, but with a lot of scatter. We confirm these findings in our sample. Sightlines with low T_{01} values also seem to have a wide range

of molecular fractions, with most of the points between 0.1 and 0.7.

In the left column of Figure 9, we show how $c_4^{A(V)}$ behaves under different gas conditions, using these estimated quantities. The most striking part is that the points with $c_4^{A(V)} > 0.15$ all have H_2 temperatures on the low end (40–80 K), and densities on the high end ($1\text{--}10 \text{ cm}^{-3}$), relative to the rest of the sample. Meanwhile, low values of $c_4^{A(V)}$ occur at a wide variety of rotational temperatures (60–120 K) and number densities ($0.05\text{--}10 \text{ cm}^{-3}$).

3.5. Depletion

The data by Jenkins (2009) have 31 sightlines in common with this work, of which 29 have a measurement of F_* , a quantity that describes the overall depletion of metals in the gas, or how much metals are locked up in dust grains. We find no obvious correlation of F_* with $N(\text{H})/A(V)$, nor with $1/R(V)$ or $A(1000)/A(V)$. Our number density measurements for these targets confirm the linear relationship between $\log n(\text{H})$ and F_* , which was already shown in Jenkins (2009). This means that the exchange of metals between the dust and the gas does not play a major role in influencing the variations in $N(\text{H})/A(V)$ or the extinction curve shape. Instead, processes such as coagulation and shattering are more

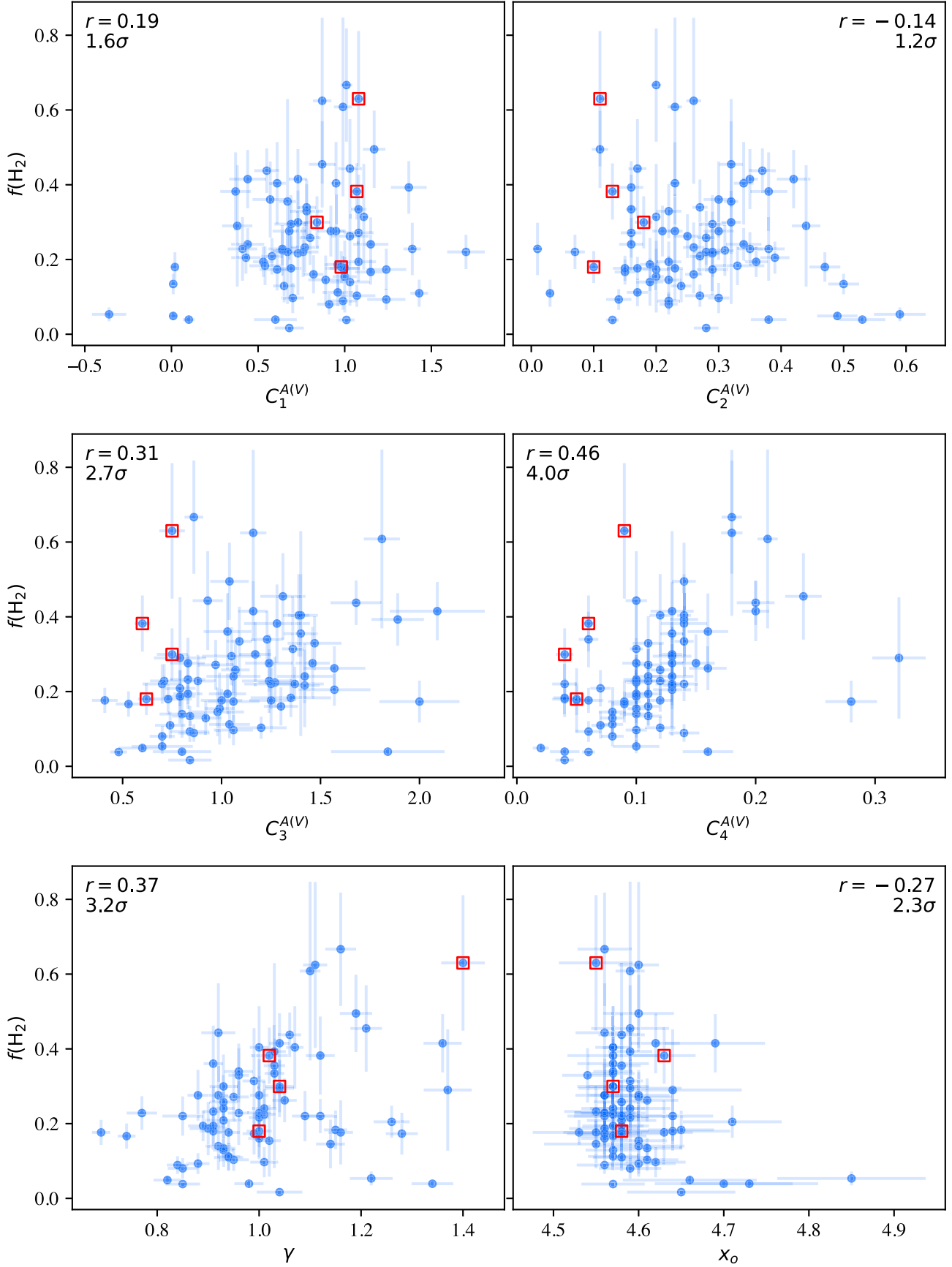


Figure 8. FM90 parameters of the extinction curves vs. molecular fraction $f(\text{H}_2)$. The numbers shown are analogous to those in Figure 5. The four low- $N(\text{H})/A(V)$ sightlines toward HD 045314, HD 164906, HD 200775, and HD 206773 have been indicated with red squares, showing that the NUV and FUV feature strengths $c_3^{A(V)}$ and $c_4^{A(V)}$ are weak.

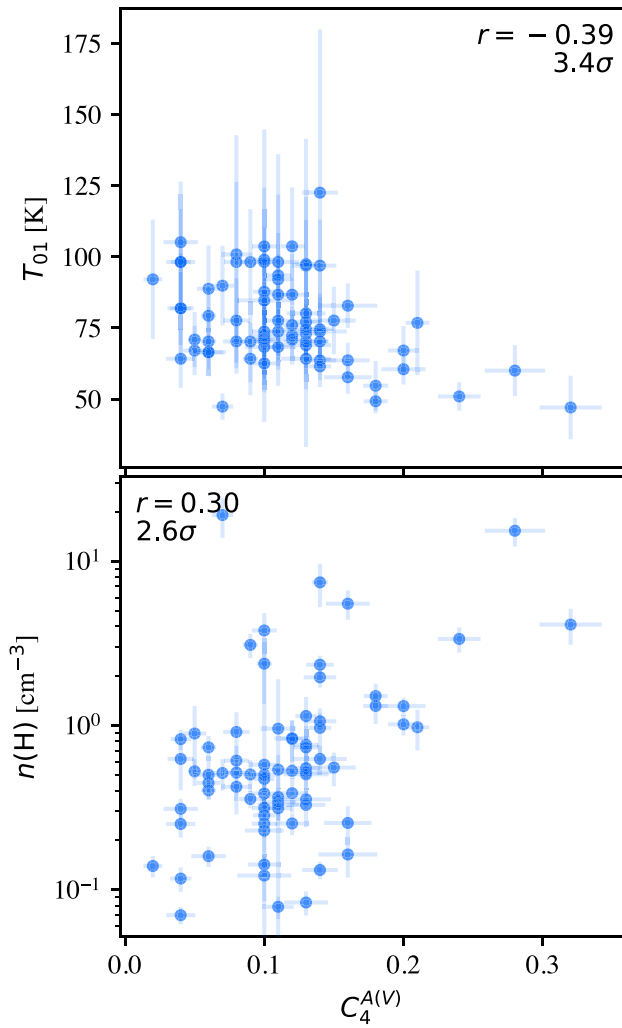


Figure 9. Left: rotational H_2 temperature T_{01} and average number density $n(H)$, compared to the FUV rise strength $c_4^{A(V)}$. The sightlines that have a strong FUV rise ($c_4^{A(V)} \gtrsim 0.2$) have low H_2 temperatures and high densities.

likely explanations, where only the size distribution changes, leaving the dust mass and composition mostly unchanged.

4. Discussion

4.1. Relationship between Small Grains and H_2

We found that $N(H_2)$ has a strong linear correlation with the absolute extinction $A(1000)$, and specifically with the contribution of the FUV rise term to $A(1000)$. This means that wherever there is H_2 , the carrier of the FUV rise feature is also present. On the other hand there is no obvious relationship between $f(H_2)$ and the UV-to-optical extinction ratio, but the FUV rise $c_4^{A(V)}$ and 2175 Å bump width γ do show a significant positive correlation. The rest of this discussion will refer to the carriers of the UV extinction in question as *small grains*. This can be justified by, for example, the models by Siebenmorgen et al. (2018), where the c_4 parameter correlates strongly with the mass ratio of small-to-large grains.

Given that our findings mostly concern the bump and rise features, the small grains in question likely have a carbonaceous composition. This could include very small solid-state grains (hydrogenated amorphous carbon) and large molecules such as PAHs, both of which are viable carriers for these dust

observables (Jones et al. 2013). As shown by laboratory measurements (e.g., Joblin et al. 1992; Steglich et al. 2010) and theoretical calculations (Mallocci et al. 2004), the absorption cross section of PAHs typically exhibits several features in the range of the 2175 Å bump. However, the extinction curve can become smooth if the size distribution of PAHs is sufficiently broad (Steglich et al. 2012). In the FUV, these cross sections exhibit a similar behavior as the FUV rise.

In the next two parts of the discussion, we speculate about two scenarios that could explain the relationship between small grains and H_2 .

4.1.1. Small Grains Enhance H_2 Shielding/Formation

The first scenario is that the presence of small grains could help the formation and/or survival of H_2 . A certain level of shielding is needed for H_2 to survive exposure to dissociating FUV photons. PDR models with different (spatially varying) grain size distributions were compared by Goicoechea & Le Boulrot (2007), and the resulting H_2 transitions occurred between $A(V) \sim 0.1$ and $A(V) \sim 1$, depending on the number of small grains in the model. At large $1/R(V)$ or large UV-to-optical extinction ratios, the enhanced UV extinction could lower the minimal $A(V)$ required for H_2 to survive. Most entries in our sample have $A(V) < 2$ along the entire sightline, so the individual clouds intersected by these sightlines have an $A(V)$ of the order 1 or lower. Therefore, a sufficient quantity of small, FUV-absorbing grains might be needed to allow the H_2 transition to occur in these diffuse clouds. However, Figure 6 showed that the underlying cause of the observed relationship is specifically the absolute extinction contribution by the FUV rise, as opposed to the total FUV extinction. If the shielding by dust was the main driver for the observed H_2 relation, one would expect the total dust extinction to be the key quantity determining $N(H_2)$. Therefore, while dust shielding in the FUV might be important in the ISM, we do not expect it to be the most important effect driving the observed H_2 correlation.

Another other option for the scenario where dust enhances the H_2 abundance is that these small dust grains significantly enhance the formation of H_2 . The dominant formation pathway for molecular hydrogen is on the surface of grains, and an increased number of small grains could provide a substantial boost to the available surface area. Small grains provide the most surface area because of their much larger number densities. The available surface area provided by a certain grain size scales with $a^2 n(a)$, where a is the grain size and $n(a)$ is the number density of grains, which typically scales with $a^{-3.5}$ (Mathis et al. 1977). The surface area A provided by grains of size a then scales as $A(a) \sim a^{-1.5}$. Now consider a grain evolution process that redistributes the mass, changing the size of a collection of dust grains with a factor of C . The number density will then scale with C^{-3} , while the surface area of a single grain will scale as C^2 . The surface area provided by all the grains will hence scale as C^{-1} (Snow 1983). So shrinking the average dust grain size while conserving the mass or volume results in more surface area. As an example, Lipshtat & Biham (2005) have shown that a broad size distribution of dust grains enhances the production of H_2 .

Besides the surface area effect, the efficiency of H_2 formation on certain types of dust grains could play a role. The grain temperature affects various processes that determine the efficiency, such as the diffusion rate and evaporation time of atoms and molecules on the grain surface. The review by

Wakelam et al. (2017) explains the possible effects in depth, and Section 4.1 of this review provides an overview of which processes are the most important for several regimes in the diffuse ISM. For small grains in particular, the temperature undergoes random fluctuations, jumping between high and low temperatures as the grains absorb UV photons and emit in the IR. This can allow small grains to have a high formation efficiency, even when the average temperature is high (Bron et al. 2014).

Small grains can also consist of large molecules with specific surface properties. Considering that the FUV rise is the most important term in the $N(\text{H}_2)$ correlation, the carriers of this feature could have surface properties that make them the dominant contributor to H_2 formation. PAHs have been found to be an important H_2 formation pathway, at temperatures $\gtrsim 200$ K and in dense PDRs (Le Page et al. 2009; Boschman et al. 2015; Castellanos et al. 2018). There are various theoretical and experimental results confirming that PAHs can be superhydrogenated when H atoms stick to them, and become catalysts for H_2 formation (Rauls & Herring’s 2008; Mennella et al. 2012; Thrower et al. 2012). Hydrogenated amorphous carbon grains could also play a role, through a photolysis process that forms H_2 (Alata et al. 2014; Jones & Habart 2015). Both PAHs and small hydrogenated amorphous carbon grains have optical properties that make them possible candidates for the NUV bump and FUV rise (Joblin et al. 1992; Jones et al. 2013).

4.1.2. Small Grains Survive Only in H_2 -dominated Environment

The second scenario is that the presence of molecular gas, or the environment in which it appears, promotes the abundance of the small in grains question. This could be either because those shielded environments prevent photodestruction of the grains, or because the density and temperature conditions are more favorable for the formation processes of the grains.

Maíz Apellániz & Barbá (2018) found that grains are larger in more strongly irradiated regions of the ISM, by comparing R_{5495} (similar to $R(V)$) in H II regions and hot bubbles with molecular sightlines. They suggest that selective destruction of small dust grains by EUV photons or thermal sputtering by atoms or ions causes increased values of R_{5495} in hotter regions. In dense molecular regions on the other hand, $R(V)$ has also been found to be higher due to coagulation causing a growth in big grains and a removal of small grains (Chokshi et al. 1993; Ormel et al. 2009). For example, in TMC-1, such processes have resulted in a extinction curve without 2175 Å bump, showing that the small grains carrying this feature have been removed (Whittet et al. 2004). The fact that the average grain size as traced by $R(V)$ is larger in both hot bubbles and dense molecular clouds, means that the relative abundance of small grains must reach a peak somewhere between those two regimes. Perhaps this is in the diffuse neutral regime that we are focusing on.

The plots including T_{01} and $n(\text{H})$ in Figure 9 showed us that the highest values of $c_4^{A(V)}$ ($\gtrsim 0.2$) only occur for sightlines of low temperature ($T_{01} \lesssim 75$ K) and higher density ($N(\text{H})/d \gtrsim 1 \text{ cm}^{-3}$) in this sample, while the results discussed in Section 3.1 showed that the FUV rise carrier coincides with molecular gas. This could mean that the subpopulation of small grains carrying the FUV feature, preferably survives in colder, denser regions of the diffuse molecular ISM. Similar behavior has been suggested for PAHs by Hensley et al. (2022), who

found a correlation between the cold neutral medium fraction and the fraction of dust in PAHs, in an all-sky map of the cold neutral medium. It is plausible that PAHs are at least partly responsible for the UV extinction features. One indication is that IR emission features associated with PAHs correlate with the strength of the dust extinction bump in the NUV, between sightlines (Jenniskens & Desert 1993; Blasberger et al. 2017; Massa et al. 2022) and between averages over the Milky Way, LMC, and SMC (Vermeij et al. 2002). Additionally, there are indications that volatile hydrocarbon dust is responsible for at least part of the FUV extinction, linking the FUV extinction to the life cycle of evaporation, condensation, and possible photodestruction of such macromolecules (Buss et al. 1994). The PAH fraction has also been found to correlate with molecular gas in the SMC (Sandstrom et al. 2010). However, sightlines in the SMC typically have a very weak or nonexistent bump, while some still have an FUV rise (Gordon et al. 2003). Therefore, if one assumes that PAHs always cause a substantial NUV bump, then there should exist another small grain component that only contributes to the FUV rise.

4.1.3. Simulation Suggestions

To determine if scenario #1 is viable, one would need to quantify how strongly the observed differences in the dust properties affect the H_2 fraction of the gas. We suggest that PDR models could be used for this purpose, as long as they have a flexible enough dust model, one that can represent the variations in $R(V)$ and the FUV rise. Both the dust extinction affecting the radiative transfer in the model, and the models for physical processes, such as H_2 formation on dust grain surfaces and heating by the photoelectric on dust grains, should depend on the same underlying grain size distribution. If one finds that certain dust properties are required to form a substantial amount of H_2 , within the constraint typical for our sample of $A(V) \lesssim 2$, then the simulations would support scenario #1. On the other hand, if these dust variations only cause a minor boost in the H_2 fraction, then the FUV rise versus $N(\text{H}_2)$ relation likely has another cause.

Scenario #2 is harder to model, as this requires knowledge of the destruction processes of small grains, which need to be simulated simultaneously with H_2 formation and dissociation. The galaxy evolution simulation by Romano et al. (2022) includes a variable grain size distribution together with H_2 formation and dissociation, and shows a jump in the molecular fraction as a function of their metric for the small-to-large grain ratio. An initial step would be to calculate the properties along randomly selected sightlines of this model, and see if the same statistical properties are observed compared to our observations of Milky Way sightlines. In that way our results could provide constraints on this type of simulation.

4.2. Dust Evolution

We compared $N(\text{H})/A(V)$ and $f(\text{H}_2)$ to investigate if the phase of the medium plays a role in the variations of $N(\text{H})/A(V)$, but there is a null result for any correlation between $N(\text{H})/A(V)$ and $f(\text{H}_2)$. In other words, we find no evidence that dust grows preferentially in more molecular regions of the diffuse Milky Way ISM, at least not within our parameter space and when using $N(\text{H})/A(V)$ as a tracer for the gas-to-dust ratio. This is within expectations, as far-IR studies (C. J. R. Clark et al. 2023, in preparation, Roman-Duval et al. 2017)

and depletions (Roman-Duval et al. 2021) suggest that the dust-to-gas ratio reaches a plateau at $N(\text{H}) \geq 10^{21} \text{ cm}^{-2}$, as most of the available metals in the gas have already been locked up in dust at that column density. Most of the dust-to-gas ratio variations happen at lower column densities, which we are not probing with our sightlines.

The remaining variations in $N(\text{H})/A(V)$ do correlate positively with the UV-to-optical extinction ratio however, according to our results shown in Figure 7. In other words, increases in the amount of dust absorbing in the V band (relative to the gas) go along with relatively fewer small grains. Especially notable are four low- $N(\text{H})/A(V)$ sightlines, highlighted on the scatter plots using red squares. They have a wide range of $f(\text{H}_2)$ values, but similarly low $N(\text{H})/A(V)$, $R(V)$, $A(1000)$, and $A(2175)/A(V)$ values. These four sightlines seem to have undergone a more drastic form of dust growth, which has increased $A(V)$ relative to the gas, while flattening the FUV rise, as seen from their highlighted position in the $c_4^{A(V)}$ panel of in Figure 8. The value of $A(1000)$ for these four sightlines is not an outlier of the $N(\text{H}_2)$ relationship shown in Figure 5. Meanwhile, $A(1000)/A(V)$ seems to be near the minimum of the range of values in the sample (Figure 7), and the same goes for the FUV bump and rise parameters (Figure 8), which means those features become small relative to the total extinction. More data points with $1/R(V)$ between 0.2 and 0.3 could shed further light on whether the four low- $N(\text{H})/A(V)$ sightlines are just anomalies, or if there is a global trend across the entire $1/R(V)$ range. Since one of these four sightlines is HD 200775, a star behind the nebula NGC 7023 which contains a PDR, the increased $R(V)$ observed for this sightline might be related to the changes in $R(V)$ observed in the IC 63 PDR (Van De Putte et al. 2020), where $R(V)$ increases near the dense part of the PDR front.

It seems unlikely that the changes in the UV-to-optical extinction ratio are driven by the destruction/survival balance of the small grains, as that would not affect $N(\text{H})/A(V)$ because small grains have only a limited amount of absorption in the V band. According to theoretical and numerical models of dust evolution and extinction curves in galaxies (Hirashita 2012; Asano et al. 2013; Hou et al. 2017; Hirashita & Aoyama 2019; Aoyama et al. 2020), extinction curves are initially flat because stellar dust production results in predominantly large grains. The FUV rise and bump become more prominent under the effects of shattering in the diffuse ISM, which produces small grains at the expense of large grains, and by the accretion of gas-phase metals in the dense ISM, which primarily enhances small grains. Coagulation then depletes the small grain population as they stick to the larger grains, increasing the dust optical depth at longer wavelengths. These simulations have been combined with observations of three nearby galaxies, to find that regions with high metallicity and molecular fraction are affected by both accretion and coagulation (Relaño et al. 2020). For our diffuse Milky Way sample, we observe that a decrease in $N(\text{H})/A(V)$ (increase in big grains per gas unit) is coupled with a decrease in $A(1000)/A(V)$ and $A(2175)/A(V)$ (decrease in small grains), so these trends likely originate from the balance between coagulation and shattering effects.

5. Conclusions

We investigated the correlations between the H I and H_2 gas contents and dust extinction in the diffuse Milky Way ISM, by

combining the GCC09 extinction data with accurate measurements of the H I and H_2 column densities. Using distances from Gaia DR2, and H_2 excitation data, several derived quantities were calculated, most notably the average number density $n(\text{H})$ and the rotational temperature T_{01} of H_2 . Correlations between pairs of quantities were explored, while taking into account both the errors in x and y , and correlations between the measurement uncertainties of derived quantities that depend on a common factor. The following notable trends were found:

1. The molecular column $N(\text{H}_2)$ correlates very strongly with $A(1000)$, but not with $A(V)$, indicating that a portion of the carriers of the FUV extinction coincide with H_2 along these diffuse sightlines. This correlation is caused by a linear relationship between the contribution of the FUV rise term to $A(1000)$, showing that the carriers of the FUV rise feature appear primarily in molecular gas. This explicitly confirms the FUV rise versus H_2 relationship previously suggested by Jenniskens et al. (1992) based on CH. We discussed two scenarios: either the FUV rise carriers are crucial for H_2 formation, or they coincide with H_2 because they require similar physical conditions to form and survive.
2. The connection between small grains and H_2 and is further highlighted by a positive correlation between $f(\text{H}_2)$, and the FM90 parameters for the UV bump width γ and the FUV rise strength $c_4^{A(V)}$.
3. The highest values of $c_4^{A(V)}$ coincide with the lowest values for the estimated number density $n(\text{H})$, and the H_2 rotational temperature T_{01} . This implies that the FUV rise carriers have a higher abundance in denser and colder regions of the diffuse molecular gas.
4. The total hydrogen column $N(\text{H})$ shows a strong linear trend with the optical dust extinction measurements $A(V)$ and $E(B - V)$, as expected from earlier literature. Deviations from the main trend in $N(\text{H})/A(V)$ are partially explained by a positive correlation with parameters that indicate the small-to-large grain ratio: $1/R(V)$, $A(1000)/A(V)$, and $A(2175)/A(V)$. The effects of coagulation are a possible interpretation for the observed reduction in small grains and enhanced extinction at optical wavelengths.

The connection between the FUV rise carrier and H_2 , which we consider the main conclusion of this work, raises new questions. The first open question is which specific type or class of dust particles, most likely carbonaceous, shows such a strong preference for H_2 -dominated environments. To solve this question, different candidates that absorb mostly in the FUV could be considered, and their viability could be evaluated according to several criteria, including but not limited to: (1) They are destroyed in atomic or ionized regions. (2) They survive in the diffuse molecular regime, and also satisfy the observed relationships with CH (Jenniskens et al. 1992) and CO (Burgh et al. 2000). (3) Their optical depth, cross section, number density, and elemental composition need to be compatible with constraints based on depletions.

The second question is the exact mechanism through which the observed FUV rise- $N(\text{H}_2)$ correlation occurs. As suggested in Section 4.1.3, we believe results from theoretical PDR models might be able to disentangle whether the extra extinction and H_2 formation by the FUV rise particles enhance

the H_2 content, or if this grain population and H_2 simply trace the same physical conditions in the diffuse ISM.

This study was supported by NASA ADP grant NAG5-13033. Based on observations made with the NASA-CNES-CSA Far Ultraviolet Spectroscopic Explorer. FUSE is operated for NASA by the Johns Hopkins University under NASA contract NAS5-32985. The IUE and HST data used in this paper were obtained from the Mikulski Archive for Space Telescopes (MAST) at the Space Telescope Science Institute. The specific observations analyzed can be accessed via doi:[10.17909/5qq7-4c81](https://doi.org/10.17909/5qq7-4c81). STScI is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS526555. Support to MAST for these data is provided by the NASA Office of Space Science via grant NAG57584 and by other grants and contracts. This research has made use of the SIMBAD database, operated at CDS (Strasbourg, France).

Software: The analysis and plotting code written specifically for this work is available on the GitHub profile of the author (https://github.com/drvdputt/dust_fuse_h2). A permanent version is on Zenodo: Van De Putte (2022). Packages: astropy (Astropy Collaboration et al. 2013, 2018), SciPy (Virtanen et al. 2020).

ORCID iDs

Dries Van De Putte  <https://orcid.org/0000-0002-5895-8268>

Karl D. Gordon  <https://orcid.org/0000-0001-5340-6774>

Geoffrey C. Clayton  <https://orcid.org/0000-0002-0141-7436>

Julia Roman-Duval  <https://orcid.org/0000-0001-6326-7069>

References

- Alata, I., Cruz-Díaz, G. A., Muñoz Caro, G. M., & Dartois, E. 2014, *A&A*, **569**, A119
- Aoyama, S., Hirashita, H., & Nagamine, K. 2020, *MNRAS*, **491**, 3844
- Asano, R. S., Takeuchi, T. T., Hirashita, H., & Nozawa, T. 2013, *MNRAS*, **432**, 637
- Asano, R. S., Takeuchi, T. T., Hirashita, H., & Nozawa, T. 2014, *MNRAS*, **440**, 134
- Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, *AJ*, **156**, 123
- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, **558**, A33
- Blasberger, A., Behar, E., Perets, H. B., Brosch, N., & Tielens, A. G. G. M. 2017, *ApJ*, **836**, 173
- Bohlin, R. C. 1975, *ApJ*, **200**, 402
- Bohlin, R. C., Savage, B. D., & Drake, J. F. 1978, *ApJ*, **224**, 132
- Boschman, L., Cazaux, S., Spaans, M., Hoekstra, R., & Schlathölter, T. 2015, *A&A*, **579**, A72
- Boyer, M. L., Srinivasan, S., Riebel, D., et al. 2012, *ApJ*, **748**, 40
- Bron, E., Le Bourlot, J., & Le Petit, F. 2014, *A&A*, **569**, A100
- Burgh, E. B., McCandliss, S. R., Andersson, B. G., & Feldman, P. D. 2000, *ApJ*, **541**, 250
- Buss, R. H. J., Allen, M., McCandliss, S., et al. 1994, *ApJ*, **430**, 630
- Butler, R. E., & Salim, S. 2021, *ApJ*, **911**, 40
- Cardelli, J. A. 1988, *ApJ*, **335**, 177
- Cardelli, J. A., & Clayton, G. C. 1991, *AJ*, **101**, 1021
- Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, *ApJ*, **345**, 245
- Cartledge, S. I. B., Clayton, G. C., Gordon, K. D., et al. 2005, *ApJ*, **630**, 355
- Cartledge, S. I. B., Lauroesch, J. T., Meyer, D. M., & Sofia, U. J. 2006, *ApJ*, **641**, 327
- Castellanos, P., Candian, A., Andrews, H., & Tielens, A. G. G. M. 2018, *A&A*, **616**, A167
- Chapman, N. L., Mundy, L. G., Lai, S.-P., & Evans, N. J. I. 2009, *ApJ*, **690**, 496
- Chokshi, A., Tielens, A. G. G. M., & Hollenbach, D. 1993, *ApJ*, **407**, 806
- Diplas, A., & Savage, B. D. 1994, *ApJ*, **427**, 274
- Draine, B. T. 2009, in ASP Conf. Ser. 414, Cosmic Dust—Near and Far, ed. T. Henning, V. Grün, & J. Steinacker (San Francisco, CA: ASP), 453
- Federman, S. R. 1982, *ApJ*, **257**, 125
- Fitzpatrick, E. L., & Massa, D. 1986, *ApJ*, **307**, 286
- Fitzpatrick, E. L., & Massa, D. 1988, *ApJ*, **328**, 734
- Fitzpatrick, E. L., & Massa, D. 1990, *ApJS*, **72**, 163
- Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, *A&A*, **616**, A1
- Goicoechea, J. R., & Le Bourlot, J. 2007, *A&A*, **467**, 1
- Good, P. 1994, *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses* (New York: Springer)
- Gordon, K. D., Cartledge, S., & Clayton, G. C. 2009, *ApJ*, **705**, 1320
- Gordon, K. D., & Clayton, G. C. 1998, *ApJ*, **500**, 816
- Gordon, K. D., Clayton, G. C., Misselt, K. A., Landolt, A. U., & Wolff, M. J. 2003, *ApJ*, **594**, 279
- Gorenstein, P. 1975, *ApJ*, **198**, 95
- Güver, T., & Özel, F. 2009, *MNRAS*, **400**, 2050
- Hensley, B. S., Murray, C. E., & Dodici, M. 2022, *ApJ*, **929**, 23
- Hirashita, H. 2012, *MNRAS*, **422**, 1263
- Hirashita, H., & Aoyama, S. 2019, *MNRAS*, **482**, 2555
- Hogg, D. W., Bovy, J., & Lang, D. 2010, arXiv:1008.4686
- Hollenbach, D., & Salpeter, E. E. 1971, *ApJ*, **163**, 155
- Hollenbach, D. J., & Tielens, A. G. G. M. 1999, *RvMP*, **71**, 173
- Hou, K.-C., Hirashita, H., Nagamine, K., Aoyama, S., & Shimizu, I. 2017, *MNRAS*, **469**, 870
- Jenkins, E. B. 2009, *ApJ*, **700**, 1299
- Jenkins, E. B., & Wallerstein, G. 2017, *ApJ*, **838**, 85
- Jenniskens, P., & Desert, F. X. 1993, *A&A*, **275**, 549
- Jenniskens, P., Ehrenfreund, P., & Desert, F. X. 1992, *A&A*, **265**, L1
- Joblin, C., Leger, A., & Martin, P. 1992, *ApJL*, **393**, L79
- Jones, A. P., Fanciullo, L., Köhler, M., et al. 2013, *A&A*, **558**, A62
- Jones, A. P., & Habart, E. 2015, *A&A*, **581**, A92
- Le Page, V., Snow, T. P., & Bierbaum, V. M. 2009, *ApJ*, **704**, 274
- Lipshtat, A., & Biham, O. 2005, *MNRAS*, **362**, 666
- Maíz Apellániz, J., & Barbá, R. H. 2018, *A&A*, **613**, A9
- Mallocci, G., Mulas, G., & Joblin, C. 2004, *A&A*, **426**, 105
- Mar, D. P., & Bailey, G. 1995, *PASA*, **12**, 239
- Massa, D., Gordon, K. D., & Fitzpatrick, E. L. 2022, *ApJ*, **925**, 19
- Mathis, J. S. 1990, *ARA&A*, **28**, 37
- Mathis, J. S., Rumpl, W., & Nordsieck, K. H. 1977, *ApJ*, **217**, 425
- McKinnon, R., Torrey, P., & Vogelsberger, M. 2016, *MNRAS*, **457**, 3775
- Misselt, K. A., Clayton, G. C., & Gordon, K. D. 1999, *ApJ*, **515**, 128
- Mennella, V., Hornekar, L., Thrower, J., & Accolla, M. 2012, *ApJL*, **745**, L2
- Ormel, C. W., Paszun, D., Dominik, C., & Tielens, A. G. G. M. 2009, *A&A*, **502**, 845
- Predehl, P., & Schmitt, J. H. M. M. 1995, *A&A*, **293**, 889
- Rachford, B. L., Snow, T. P., Destree, J. D., et al. 2009, *ApJS*, **180**, 125
- Rachford, B. L., Snow, T. P., Tumlinson, J., et al. 2001, *ApJ*, **555**, 839
- Rachford, B. L., Snow, T. P., Tumlinson, J., et al. 2002, *ApJ*, **577**, 221
- Rauls, E., & Hornekar, L. 2008, *ApJ*, **679**, 531
- Reina, C., & Tarengi, M. 1973, *A&A*, **26**, 257
- Relaño, M., Lisenfeld, U., Hou, K. C., et al. 2020, *A&A*, **636**, A18
- Robotham, A. S. G., & Obreschkow, D. 2015, *PASA*, **32**, e033
- Röllig, M., Abel, N. P., Bell, T., et al. 2007, *A&A*, **467**, 187
- Roman-Duval, J., Bot, C., Chastenot, J., & Gordon, K. 2017, *ApJ*, **841**, 72
- Roman-Duval, J., Gordon, K. D., Meixner, M., et al. 2014, *ApJ*, **797**, 86
- Roman-Duval, J., Jenkins, E. B., Tchernyshyov, K., et al. 2021, *ApJ*, **910**, 95
- Roman-Duval, J., Jenkins, E. B., Williams, B., et al. 2019, *ApJ*, **871**, 151
- Romano, L. E. C., Nagamine, K., & Hirashita, H. 2022, *MNRAS*, **514**, 1461
- Rowlands, K., Gomez, H. L., Dunne, L., et al. 2014, *MNRAS*, **441**, 1040
- Sandstrom, K. M., Bolatto, A. D., Draine, B. T., Bot, C., & Stanimirović, S. 2010, *ApJ*, **715**, 701
- Savage, B. D., & Sembach, K. R. 1996, *ARA&A*, **34**, 279
- Scoville, N. Z., Yun, M. S., Clemens, D. P., Sanders, D. B., & Waller, W. H. 1987, *ApJS*, **63**, 821
- Shull, J. M., Danforth, C. W., & Anderson, K. L. 2021, *ApJ*, **911**, 55
- Siebenmorgen, R., Voshchinnikov, N. V., Bagnulo, S., et al. 2018, *A&A*, **611**, A5
- Snow, T. P. 1983, *ApJL*, **269**, L57
- Sofia, U. J., Wolff, M. J., Rachford, B., et al. 2005, *ApJ*, **625**, 167
- Steglich, M., Carpentier, Y., Jäger, C., et al. 2012, *A&A*, **540**, A110
- Steglich, M., Jäger, C., Rouillé, G., et al. 2010, *ApJL*, **712**, L16
- Steinacker, J., Pagani, L., Bacmann, A., & Guieu, S. 2010, *A&A*, **511**, A9
- Sternberg, A., Le Petit, F., Roueff, E., & Le Bourlot, J. 2014, *ApJ*, **790**, 10
- Tchernyshyov, K., Meixner, M., Seale, J., et al. 2015, *ApJ*, **811**, 78
- Thrower, J. D., Jørgensen, B., Friis, E. E., et al. 2012, *ApJ*, **752**, 3

- Valencic, L. A., Clayton, G. C., & Gordon, K. D. 2004, *ApJ*, **616**, 912
- Van De Putte, D. 2022, drvdputt/dust_fuse_h2, v1.0, Zenodo, doi:[10.5281/zenodo.7108380](https://doi.org/10.5281/zenodo.7108380)
- Van De Putte, D., Gordon, K. D., Roman-Duval, J., et al. 2020, *ApJ*, **888**, 22
- Vermeij, R., Peeters, E., Tielens, A. G. G. M., & van der Hulst, J. M. 2002, *A&A*, **382**, 1042
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *NatMe*, **17**, 261
- Wakelam, V., Bron, E., Cazaux, S., et al. 2017, *MolAs*, **9**, 1
- Weingartner, J. C., & Draine, B. T. 2001, *ApJ*, **548**, 296
- Welty, D. E., Hobbs, L. M., & York, D. G. 1991, *ApJS*, **75**, 425
- Whittet, D. C. B., Shenoy, S. S., Clayton, G. C., & Gordon, K. D. 2004, *ApJ*, **602**, 291
- Zhu, H., Tian, W., Li, A., & Zhang, M. 2017, *MNRAS*, **471**, 3494
- Zhukovska, S., Dobbs, C., Jenkins, E. B., & Klessen, R. S. 2016, *ApJ*, **831**, 147
- Zhukovska, S., Gail, H. P., & Tieloff, M. 2008, *A&A*, **479**, 453
- Zuo, W., Li, A., & Zhao, G. 2021, *ApJS*, **257**, 63