

# One H<sub>2</sub> molecule per ten million H atoms reveals sub-parsec-scale cold overdensities at $z \sim 4$ <sup>★</sup>

P. Noterdaeme<sup>1,★★</sup>, S. Balashev<sup>2,★★</sup>, T. Berg<sup>3,4</sup>, S. Cristiani<sup>5,6,7</sup>, R. Cuellar<sup>8</sup>, G. Cupani<sup>5</sup>, S. Di Stefano<sup>5</sup>, V. D’Odorico<sup>5</sup>, C. Fian<sup>5</sup>, B. Godard<sup>9,10</sup>, S. López<sup>8</sup>, D. Milakovic<sup>5,6</sup>, A. Trost<sup>5</sup>, and L. Welsh<sup>5,11</sup>

<sup>1</sup> Institut d’Astrophysique de Paris, CNRS-SU, UMR 7095, 98bis bd Arago, 75014 Paris, France

<sup>2</sup> Ioffe Institute, Polytekhnicheskaya 26, 194021 Saint-Petersburg, Russia

<sup>3</sup> NRC Herzberg Astronomy and Astrophysics Research Centre, 5071 West Saanich Road, Victoria, BC V9E 2E7, Canada

<sup>4</sup> Department of Physics and Astronomy, Camosun College, 3100 Foul Bay Road, Victoria, BC V8P 5J2, Canada

<sup>5</sup> INAF – Osservatorio Astronomico di Trieste, Via G. B. Tiepolo, 11, 34143 Trieste, Italy

<sup>6</sup> IFPU – Institute for Fundamental Physics of the Universe, Via Beirut 2, 34151 Trieste, Italy

<sup>7</sup> National Institute for Nuclear Physics, Via Valerio 2, 34127 Trieste, Italy

<sup>8</sup> Departamento de Astronomía, Universidad de Chile, Casilla 36-D, Santiago 7550000, Chile

<sup>9</sup> LERMA, Observatoire de Paris, Université PSL, Sorbonne Université, 75014 Paris, France

<sup>10</sup> Laboratoire de Physique de l’École Normale Supérieure, PSL, CNRS, SU, Université de Paris, 75005 Paris, France

<sup>11</sup> Centre for Extragalactic Astronomy, Durham University, South Road, Durham DH1 3LE, UK

Received 18 December 2025 / Accepted 23 January 2026

## ABSTRACT

We present the detection and analysis of H<sub>2</sub> absorption at  $z = 4.24$  towards the bright quasar J0007–5705, which was observed with the Very Large Telescope as part of the ESPRESSO QUasar Absorption Line Survey (EQUALS). The high resolving power of  $R \approx 120\,000$  enables the identification of extremely weak H<sub>2</sub> lines in several rotational levels at a total column density of  $N(\text{H}_2) \approx 2 \times 10^{14} \text{ cm}^{-2}$ , which is among the lowest ever measured in quasar absorption systems. Remarkably, this constitutes the highest redshift H<sub>2</sub> detection to date. Two velocity components are resolved that are separated by only  $3 \text{ km s}^{-1}$ : a narrow ( $b \sim 1.7 \text{ km s}^{-1}$ ) and a broader ( $b \approx 6.2 \text{ km s}^{-1}$ ) component. Modelling the rotational population of H<sub>2</sub> yields a density of  $\log n_{\text{H}}/\text{cm}^{-3} \sim 2.8$  and temperature of  $\sim 40 \text{ K}$  (typical of the cold neutral medium) for the narrow component and  $\log n_{\text{H}}/\text{cm}^{-3} \sim 1.4$ ,  $T \sim 600 \text{ K}$  for the warmer, more turbulent component under a moderate ultraviolet (UV) field, suggesting at least a several-megaparsec distance from the quasar. This system reveals the existence of tiny (down to  $\sim 0.01 \text{ pc}$ ), cold overdensities in the neutral medium. Their detection among only seven damped Lyman- $\alpha$  systems in EQUALS suggests that they may be widespread yet usually remain undetected. H<sub>2</sub> provides an exceptionally sensitive probe of these structures: even a minute molecular fraction produces measurable Lyman-Werner absorption lines along the extremely narrow optical beam –the size of the quasar’s accretion disc– when observed at sufficiently high spectral resolution. High-resolution spectroscopy on extremely large telescopes may routinely detect and resolve such structures in the distant Universe, where 21-cm absorption traces the collective contribution of many cold cloudlets towards larger radio background sources.

**Key words.** ISM: molecules – ISM: structure – quasars: absorption lines

## 1. Introduction

Molecular hydrogen (H<sub>2</sub>), the simplest and by far the most abundant molecule in the Universe, plays a central role in tracing the physical conditions of the interstellar and circumgalactic media. Owing to its specific formation, destruction, and excitation processes, it provides sensitive diagnostics of gas density, temperature, and radiation field via its rotational and vibrational populations (e.g. Jura 1975; Bialy et al. 2017). In addition, H<sub>2</sub> can reveal dynamical effects such as the dissipation of turbulent energy in vortices and shocks (e.g. Godard et al. 2009; Lesaffre et al. 2020) and phase-transition processes such as condensation and evaporation of cold-neutral-medium (CNM) clouds (Valdivia et al. 2016; Bellomi et al. 2020).

Infrared H<sub>2</sub> emission lines are typically detected in dense warm gas illuminated by massive stars (Habart et al. 2003), but they can also arise from diffuse and translucent gas, likely due

to the reprocessing of mechanical or ultraviolet (UV) energy (Villa-Vélez et al. 2024). Nevertheless, the most sensitive way to detect H<sub>2</sub> remains through electronic (Lyman and Werner) absorption lines against background sources. Since these lines lie in the far-UV, observations of H<sub>2</sub> in our Galaxy and up to intermediate redshifts require space-based instruments. Pioneering studies with Copernicus (e.g. Savage et al. 1977) and later with the Far Ultraviolet Spectroscopic Explorer (e.g. Shull et al. 2000; Gillmon et al. 2006) provided a detailed view of molecular gas in the Milky Way and nearby galaxies, while the Cosmic Origins Spectrograph (COS) onboard the Hubble Space Telescope (HST) extended these studies to  $z \sim 0.7$  (e.g. Muzahid et al. 2015).

At higher redshifts, the Lyman-Werner bands of H<sub>2</sub> shift into the optical window accessible from the ground (Levshakov & Varshalovich 1985). In particular, the Ultraviolet and Visual Echelle Spectrograph (UVES) and X-shooter on the Very Large Telescope (VLT) have revealed the presence of H<sub>2</sub> in a fraction of damped Ly $\alpha$  systems (DLAs) in quasar spectra at  $z \sim 2\text{--}3$  (e.g. Ledoux et al. 2003; Noterdaeme et al. 2008; Balashev et al. 2019). Building on the idea proposed by Stern et al. (2021) that the inner circumgalactic medium in the

\* Based on observations collected at the European Southern Observatory under programme ID 112.25NR.

\*\* Corresponding authors: noterdaeme@iap.fr, s.balashev@gmail.com

early Universe may contain a larger neutral fraction than is typically inferred at low redshifts, recent simulations with detailed HI–H<sub>2</sub> transition models suggest that such gas can produce DLA-level HI columns while maintaining low molecular fractions (Gurman et al. 2025). This could explain the low detection rate of H<sub>2</sub> in most systems at  $z \sim 2$ –3, although some sight lines with higher atomic and molecular content likely intersect galactic discs (Ranjan et al. 2018).

Pushing the observations to  $z \sim 4$  is key to probing the neutral phase, but it introduces additional challenges: the increasing density of the Ly $\alpha$  forest and the dimming of background sources and their declining number all hamper absorption-line studies. Fortunately, the recent discovery of bright quasars at  $z > 3$  (e.g. Boutsia et al. 2020; Wolf et al. 2020) has enabled a high-resolution ( $R \sim 120\,000$ ) legacy survey of 23  $z \sim 4$  quasars with the Echelle Spectrograph for Rocky Exoplanets and Stable Spectroscopic Observations (ESPRESSO, Pepe et al. 2021) on the VLT: the QUasar Absorption Line Survey (EQUALS, Berg et al. 2025).

Here, we report the detection of H<sub>2</sub> absorption in a strong DLA at  $z_{\text{abs}} \approx 4.24$  towards J000736.56–570151.8 (hereafter J0007–5705;  $z_Q \approx 4.27$ <sup>1</sup>). This marks the highest redshift at which H<sub>2</sub> absorption has been detected to date. Although only slightly beyond the  $z_{\text{abs}} = 4.22$  absorber reported by Ledoux et al. (2006), its much lower column density would likely have remained undetected without the sensitivity and resolution of our observations. These are presented together with the data reduction in Sect. 2. The analysis is described in Sect. 3, and the results are presented in Sect. 4. We discuss the results and conclude in Sect. 5.

## 2. Observations and data reduction

J0007–5705 was observed seven times in October 2023, with each observation consisting of a 3053 s on-target ESPRESSO exposure. All exposures were obtained in dark time, and under clear or better conditions with a seeing of  $<0.9''$  and an airmass of  $<1.5$ ; we adopted a  $4 \times 2$  binning in SINGLEHR mode. We reduced the data using ESO ESPRESSO pipeline v3.3.10. Individual exposures were shifted to the barycentric frame and combined into a single 1D spectrum using Astrocook (Cupani et al. 2020). Atmospheric H<sub>2</sub>O and O<sub>2</sub> absorption was modelled in each individual observed exposure using Molecfit (Smette et al. 2015), and the resulting transmission spectra were then combined in the same way as the science data to produce an effective model of the atmospheric transmission in the final spectrum.

## 3. Analysis

We performed multi-component Voigt profile fitting to derive the column densities ( $N$ ), redshifts ( $z$ ), and Doppler parameters ( $b$ ) of the HI metals and H<sub>2</sub> absorption features. Telluric absorption lines were taken into account by including the effective atmospheric transmission in the overall fitted model.

### 3.1. Neutral hydrogen

We initially reconstructed the unabsorbed quasar continuum using a spline function with nodes placed in absorption-free

<sup>1</sup>  $z_Q = 4.26$  was originally derived from the shallow Si IV and C IV emission in the Irénée du Pont telescope discovery spectrum. The presence of N V absorption at  $z \approx 4.27$  suggests a redshift at or slightly above this value (see Cuellar et al. 2025).

**Table 1.** Results from fitting singly ionised metal lines.

$z$	$b$ (km s <sup>-1</sup> )	log $N/\text{cm}^{-2}$		
		S II	Fe II	C II*
4.241769	13.0 <sup>+1.0</sup> <sub>-2.0</sub>	13.59 <sup>+0.04</sup> <sub>-0.04</sub>		11.96 <sup>+0.27</sup> <sub>-0.28</sub>
4.242737	1.75 <sup>+0.23</sup> <sub>-0.22</sub>	13.38 <sup>+0.08</sup> <sub>-0.08</sub>	12.56 <sup>+0.24</sup> <sub>-0.24</sub>	12.92 <sup>+0.06</sup> <sub>-0.05</sub>
4.242755	6.34 <sup>+0.52</sup> <sub>-0.52</sub>	13.61 <sup>+0.09</sup> <sub>-0.10</sub>	13.66 <sup>+0.05</sup> <sub>-0.05</sub>	12.78 <sup>+0.12</sup> <sub>-0.13</sub>
4.242861	2.91 <sup>+0.25</sup> <sub>-0.25</sub>	13.60 <sup>+0.06</sup> <sub>-0.07</sub>	13.29 <sup>+0.09</sup> <sub>-0.10</sub>	12.86 <sup>+0.05</sup> <sub>-0.06</sub>
4.243244	4.08 <sup>+0.31</sup> <sub>-0.30</sub>	13.99 <sup>+0.05</sup> <sub>-0.06</sub>	13.97 <sup>+0.06</sup> <sub>-0.05</sub>	12.75 <sup>+0.06</sup> <sub>-0.06</sub>
4.243349	3.37 <sup>+0.14</sup> <sub>-0.15</sub>	14.25 <sup>+0.04</sup> <sub>-0.03</sub>	14.26 <sup>+0.03</sup> <sub>-0.04</sub>	12.99 <sup>+0.04</sup> <sub>-0.04</sub>
4.243554	8.20 <sup>+0.65</sup> <sub>-0.65</sub>	13.51 <sup>+0.04</sup> <sub>-0.05</sub>	13.60 <sup>+0.04</sup> <sub>-0.03</sub>	12.27 <sup>+0.10</sup> <sub>-0.10</sub>
Total		14.65 ± 0.02	14.58 ± 0.03	13.60 ± 0.03

regions. Over the Ly- $\alpha$  emission line –affected by the DLA– we positioned the spline anchor points manually, guided by a matched quasar composite spectrum. The continuum was then refined by simultaneously fitting the Lyman series lines along with a Chebyshev polynomial. From this procedure, we measured the total neutral hydrogen column density to be  $\log N(\text{HI})/\text{cm}^{-2} = 21.36 \pm 0.05$ . The effective Doppler parameter and redshift are also well constrained from the sharp edges of the high-order Lyman series to be  $b_{\text{HI}}^{\text{eff}} = 27 \pm 1 \text{ km s}^{-1}$  and  $z_{\text{HI}}^{\text{eff}} = 4.242708$  (see Fig. 1).

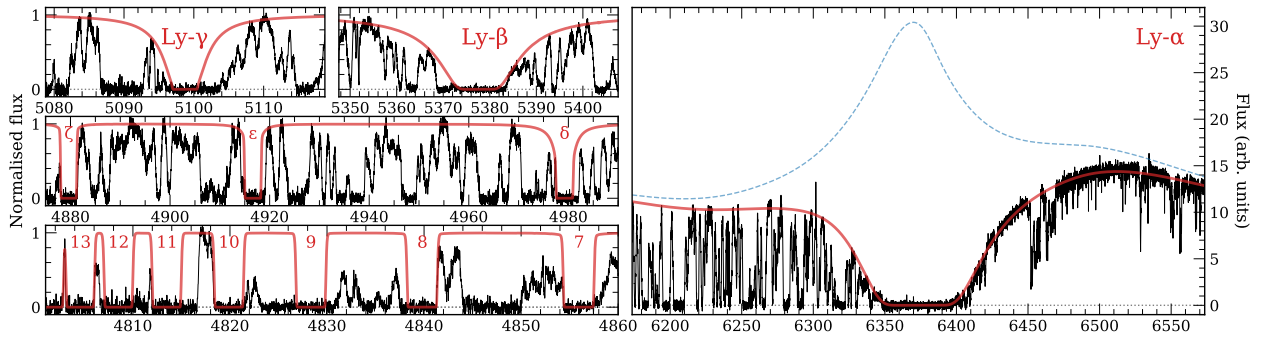
### 3.2. Metals

Singly ionised metal absorption lines show that the neutral gas is distributed over only about 50 km s<sup>-1</sup> (Fig. 2), with a shallow satellite component separated by about 60 km s<sup>-1</sup> from the rest. Atomic data from Morton (2003) were used to fit the metals, except for S II, for which Ritz wavelengths were taken from the NIST Atomic Spectra Database<sup>2</sup>. A mismatch with the data was evident for S II  $\lambda 1253$  when using the wavelength from Morton (2003, 1253.805 Å), as noted by Noterdaeme et al. (2007a). In contrast, the Ritz wavelengths provide an excellent match (with 1253.811 Å). While the profiles of O I, Si II and C II are strongly saturated and not fitted, those of S II, Fe II and C II in its first excited level are well reproduced with a seven component model (Table 1).

### 3.3. Molecular hydrogen

H<sub>2</sub> absorption lines were detected in the first four rotational levels (see Fig. 2). To account for both continuum uncertainties and frequent blending with intergalactic features, we simultaneously fitted intervening HI components and H<sub>2</sub> lines. Two components – one narrow and one broad – were required to model the H<sub>2</sub> absorption profile. During the fit, the redshifts ( $z$ ) and Doppler parameters ( $b$ ) were tied across rotational levels ( $J$ ). For the narrow component, the Doppler parameter is primarily constrained by the  $J = 1$  level, which shows numerous transitions spanning from optically thin to intermediate optical depths. In contrast, lines from higher  $J$  levels lie mostly in the optically thin regime, where the column densities are largely insensitive to the  $b$  value, and the observed line width merely reflects the instrumental line spread function. The broad component is in turn well resolved,

<sup>2</sup> <https://www.nist.gov/pml/atomic-spectra-database>



**Fig. 1.** H I Lyman series at  $z = 4.24$  towards J0007–5705. ESPRESSO data are shown in black, with the synthetic H I profile overplotted in red. For Ly- $\alpha$ , the reconstructed quasar emission profile is indicated by the dashed blue line.

and its Doppler parameter is directly constrained by the observed line widths. The optically thin  $J \neq 1$  lines justify our assumption of a shared  $b$  across levels, and, importantly, it allowed us to distinguish between kinematically and thermally distinct components rather than fitting a single, level-dependent effective profile. Allowing  $b$  to vary with level may actually cause the contributions of different gas phases to blend, obscuring the underlying structure (see Table 2).

## 4. Results

### 4.1. Chemical enrichment

From the total column densities of neutral gas,  $\log N(\text{H I}) = 21.36 \pm 0.05$ , and of volatile, singly ionised sulphur,  $\log N(\text{S II}) = 14.64 \pm 0.02$ , we inferred the average metallicity of the system to be  $Z \approx 0.01 Z_{\odot}$  and the depletion of iron to be  $[\text{Fe}/\text{S}] \approx -0.4$ . One component has stronger depletion with  $[\text{Fe}/\text{S}] \approx -1.2$  and coincides (within  $0.4 \text{ km s}^{-1}$ ) with the narrow  $\text{H}_2$  component. This suggests that the main  $\text{H}_2$ -formation route remains catalytic reactions on the surface of dust grains (e.g. Wakelam et al. 2017), even in this low-metallicity environment (see Glover 2003 for a discussion about the possible dominance of gas-phase reactions in environments with a low dust content).

### 4.2. Physical conditions

The abundance of  $\text{H}_2$  in the neutral gas depends on the formation-photodissociation equilibrium (Jura 1975):

$$R n n_{\text{HI}} = n_{\text{H}_2} D_0 S, \quad (1)$$

with  $n = n_{\text{HI}} + 2n_{\text{H}_2}$ ;  $D_0 = 5.8 \times 10^{-11} I_{\text{UV}} \text{ s}^{-1}$ , where  $I_{\text{UV}}$  is the UV field intensity in Draine units;  $S$  is the self-shielding function, which, given the low  $\text{H}_2$  column density, can be approximated here to unity; and  $R$  is the formation rate onto dust grains. When introducing the molecular fraction  $f_{\text{H}_2} = 2n_{\text{H}_2}/(2n_{\text{H}_2} + n_{\text{HI}})$ , Eq. (1) becomes

$$nR = D_0 \frac{f_{\text{H}_2}}{2(1 - f_{\text{H}_2})}. \quad (2)$$

Considering the typical density-to-UV ratio in high- $z$  CNM simulations,  $\langle n/I_{\text{UV}} \rangle = (N(\text{H I})/10^{21} \text{ cm}^{-2}) \text{ cm}^{-3}$  (Gurman et al. 2025), the observed  $\langle f_{\text{H}_2} \rangle \approx 1.7 \times 10^{-7}$  implies an average formation rate of  $\langle R \rangle \approx 2 \times 10^{-18} \text{ s}^{-1}$ . This is below the standard CNM value in the local ISM ( $R \approx 3 \times 10^{-17} \text{ s}^{-1}$ ; Jura 1975), but somewhat above expectations at the measured metallicity<sup>3</sup>. In other

<sup>3</sup>  $R < 0.4 \times 10^{-18} \text{ s}^{-1}$ , accounting for the sub-linear scaling of  $R$  with metallicity (e.g. Roman-Duval et al. 2022) and its  $\sqrt{T}$ -dependence.

**Table 2.** Results from fitting  $\text{H}_2$  lines.

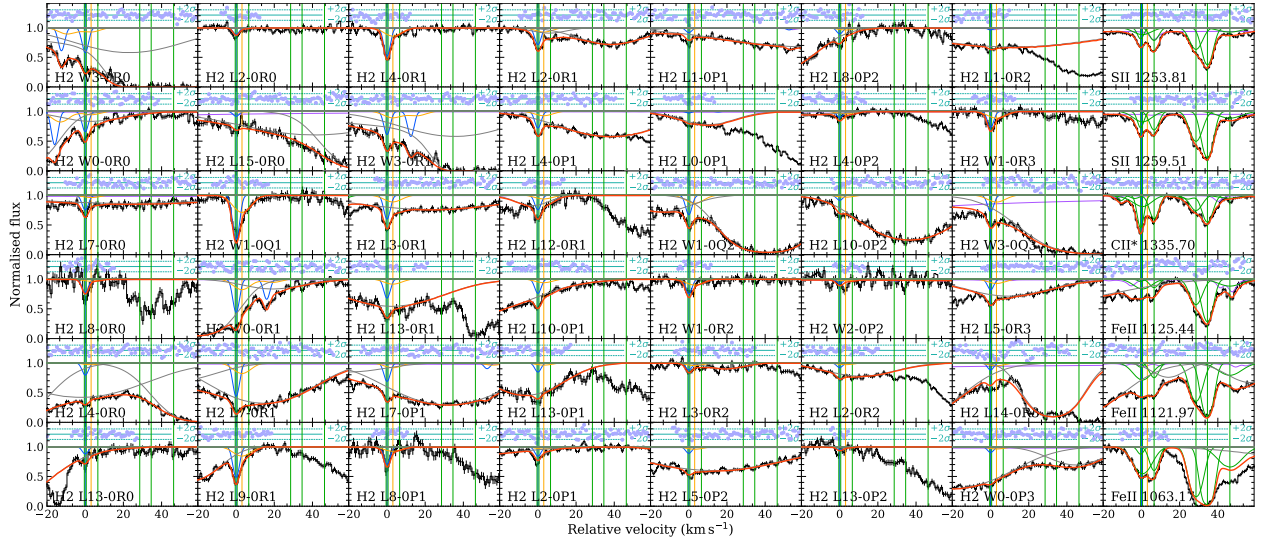
Component:	Narrow	Broad
$z$	4.242745	4.242799
$b$ ( $\text{km s}^{-1}$ )	$1.7 \pm 0.1$	$6.2 \pm 0.3$
$\log N(\text{H}_2, J = 0)$	$13.17 \pm 0.04$	$12.75 \pm 0.14$
$\log N(\text{H}_2, J = 1)$	$13.81 \pm 0.02$	$13.60 \pm 0.04$
$\log N(\text{H}_2, J = 2)$	$13.23 \pm 0.05$	$13.24 \pm 0.08$
$\log N(\text{H}_2, J = 3)$	$13.28 \pm 0.05$	$13.30 \pm 0.07$
Total	$14.06 \pm 0.02$	$13.92 \pm 0.03$

words, if average densities from the simulations are assumed, reproducing the  $\text{H}_2$  constraints requires adopting a formation rate higher than that implied by the mean metal content. This suggests that localised density peaks contribute to  $\text{H}_2$  production. We note that moderately elevated dust abundances assumptions were also used in the fiducial model of Gurman et al. (2025) to reproduce global  $\text{H}_2$  detection rates in DLAs.

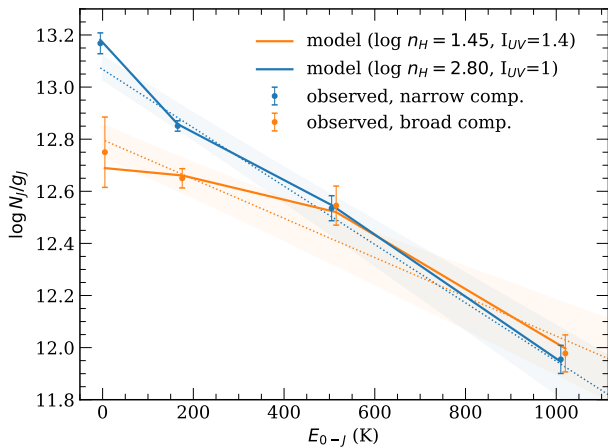
Next, we investigated the regions where  $\text{H}_2$  is effectively formed. From the Doppler parameters of the narrow and broad components ( $b = 1.7$  and  $6.2 \text{ km s}^{-1}$ ), assuming pure thermal contribution, it is possible to derive strict upper limits on the gas temperatures to be  $\sim 350 \text{ K}$  and  $\sim 4600 \text{ K}$ . In other words, the narrow component arises from cold medium, while the broad one could in principle arise from a warmer medium.

Key information about the gas temperature comes from the population of  $\text{H}_2$  across different rotational levels (Fig. 3). For the narrow component, we measure  $T_{01} \approx 233 \pm 33 \text{ K}$ , with the distribution across levels ( $J = 0-3$ ) reasonably fitted by a single excitation temperature of  $T_{\text{ex}} \sim 390 \text{ K}$ . While  $T_{01}$  is generally considered a good proxy for the kinetic temperature of the gas in the high-column-density regime, where  $\text{H}_2$  is self-shielded in the low rotational levels, it more likely represents an upper limit here, since  $\text{H}_2$  formation tends to overpopulate  $J = 1$  relatively to  $J = 0$ . The broad component, in turn, exhibits a flatter excitation diagram with  $T_{\text{ex}} \sim 600 \pm 100 \text{ K}$  and a  $J = 1/J = 0$  ratio consistent with 3 –the ratio of statistical weights– i.e., the value expected when the ortho-para ratio is set by  $\text{H}_2$  formation and destruction processes (Abgrall et al. 1992). This directly indicates that the number density in the broad component is lower than that in the narrow one, where collisional excitation is likely at play. In contrast, in the broad component, the rotational populations do not directly trace the gas kinetic temperature.

To further constrain the physical conditions, we investigated the rotational excitation of  $\text{H}_2$  in constant-density models with



**Fig. 2.** Voigt-profile fit to H<sub>2</sub> and metal lines towards J0007–5705. The origin is set at  $z_{\text{H}_2} = 4.242745$ . The observed normalised spectrum is shown in black, with the total synthetic absorption in red. H<sub>2</sub> components are shown in blue (narrow) and orange (wide) while metal components are shown in green. The transmission from the DLA H I lines, as well as telluric lines (visible in e.g. Si II $\lambda$ 1253 and Fe II $\lambda$ 1125 panels), is shown in purple. Additional components in grey, fitted jointly with H<sub>2</sub> and metal lines, represent intervening Lyman- $\alpha$  forest lines (the  $v \sim 20 \text{ km s}^{-1}$  component in the H<sub>2</sub>W3-0R0 panel is actually Ly- $\beta$  from a  $z = 3.838$  system, also visible in Ly- $\alpha$  at the blue edge of the Fe II $\lambda$ 1121 panel).



**Fig. 3.** Rotational  $J = 0\text{--}3$  population of H<sub>2</sub> for both components. Points are artificially shifted by  $\pm 5 \text{ K}$  along the  $x$ -axis for clarity. Dashed lines show single-temperature fits (with shaded uncertainties), and solid lines show our best fiducial Cloudy models.

Cloudy v25 (see Chatzikos et al. 2023). We assumed a plane-parallel geometry for the two H<sub>2</sub>-bearing clouds illuminated by a radiation field. As a starting point, we adopted observational constraints whenever possible and standard values when no constraints were available. We included a fixed metagalactic background from Khaire & Srianand (2019) at  $z = 4.24$  and an additional Draine-like UV field, which was treated as a varying parameter. H-ionising photons were removed, as our focus is on the neutral gas. We adopted a Galactic cosmic-ray (CR) ionisation rate of  $2 \times 10^{-16} \text{ s}^{-1}$  per H<sup>0</sup> (Indriolo et al. 2007) and scaled the metal and dust abundances to 1% the local ISM values, according to the mean DLA metallicity.

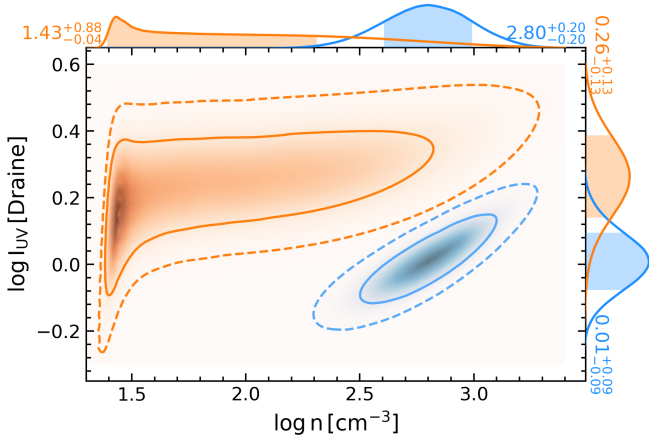
Under these assumptions, we created model grids varying only the gas density and the strength of the UV field. Our fiducial model best reproduces the observed H<sub>2</sub> excitation in both components for densities of  $\log n_{\text{H}} \approx 2.8$  and 1.4 for the narrow and broad components, respectively, under a UV field compa-

table with the Galactic level for both components (see Figs. 3 and 4). The modelled gas temperatures, 40 and 560 K, correspond to similar thermal pressure ( $P_{\text{th}}/k_{\text{B}} \sim 2 \times 10^4 \text{ K cm}^{-3}$ ) in the two components and are consistent with the limits implied by the Doppler parameters. Although we did not attempt to reproduce the metal column densities (as the association between metals and H<sub>2</sub> components is uncertain and the depletion pattern is not well constrained), the predicted Si II and C II\* columns are reasonably close to (below) the observed values, providing additional confidence in the model.

We also tested models with CR ionisation rates varied by a factor of ten and with either baseline or tenfold-higher metal and grain abundances. These provide poorer fits to the H<sub>2</sub> excitation and metal column densities, especially when assuming identical UV field and CR rates for both components. A more realistic model, where the cold component has a higher gas-phase metallicity – consistent with stronger iron depletion – but still a low grain abundance (as dust scales sub-linearly with metals), improves the match to the C II\* and Si II columns without changing the derived density or UV field. Since C II\* excitation scales as  $nT^{0.35}$  (Balashev et al. 2022), the derived temperatures and densities imply a C II\*/C II ratio about four times higher in the narrow component, in agreement with the observations.

#### 4.3. Length scales

The inferred densities imply length scales of  $L \approx 0.01 \text{ pc}$  (narrow H<sub>2</sub> component) to a few parsecs (broad component), pointing to a highly fragmented neutral medium, consistent with lower-redshift studies of self-shielded H<sub>2</sub> systems (e.g. Jorgenson et al. 2009; Balashev et al. 2011). Here, H<sub>2</sub> is merely a tracer of cold gas, yet it reveals that the latter is structured down to very small scales. The fact that the larger cloud is also the more turbulent one is consistent with the trend expected from Larson’s scaling ( $b_{\text{turb}} \propto L^{0.38}$ , Larson 1981), providing a self-consistent link between cloud sizes and the underlying turbulent dynamics of the gas. The coexistence of these narrow and broad H<sub>2</sub> components likely reflects the underlying  $n_{\text{H}}$  distribution shaped by turbulence and self-gravity. Such a close physical



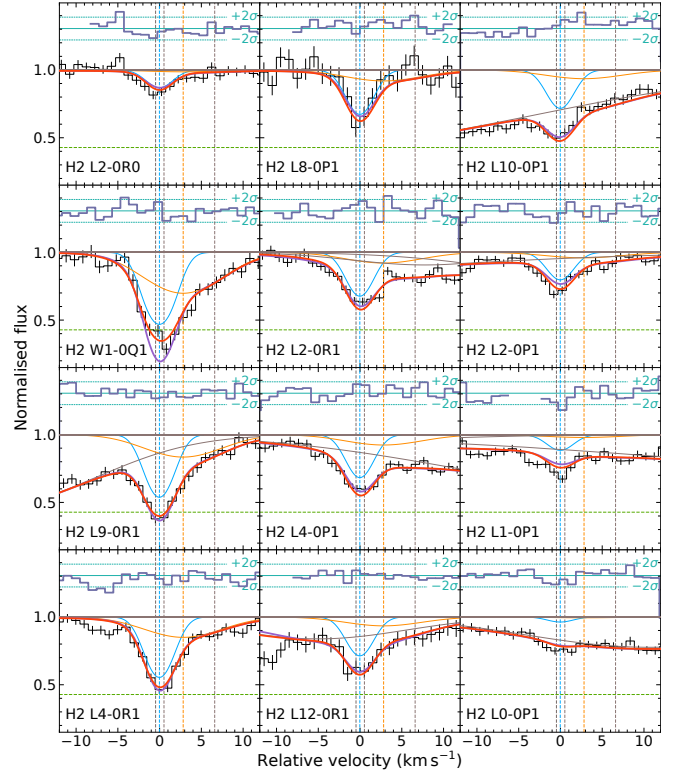
**Fig. 4.** Posterior distributions of the UV field and number density in the broad (orange) and narrow (blue) components. The likelihood was derived by comparing the observed  $H_2$   $J = 0$  to  $J = 3$  column densities with those from a grid of Cloudy models and converted into 2D posterior distribution assuming independent flat priors in logarithmic space. The 1D marginalised posteriors are shown along the top and right axes.

association could also explain the previously reported increase in the  $b$  parameter when the absorption is considered as a whole (e.g. Jenkins & Peimbert 1997; Noterdaeme et al. 2007a; Tchernyshyov 2022).

Motivated by the very small inferred size of the cold cloud, we investigated the possibility of partial coverage. This occurs when the absorbing cloud does not fully cover the background source, which is expected when the cloud’s characteristic size is smaller than, or comparable to, that of the emitter at the absorption wavelength. This effect is most readily identified through a non-zero residual flux at the bottom of saturated lines, as observed, for example, in strong  $H_2$  absorption against the parsec-scale broad emission-line regions of quasars (e.g. Balashev et al. 2011).

In the present case, the 0.01-pc length scale inferred for the cold component is comparable to the light-day scale of quasar accretion discs (e.g. Yu et al. 2020), against which the  $H_2$  absorption is observed. However, the  $H_2$  lines are optically thin, so any partial-coverage signature is expected to be subtle. Nevertheless, introducing the covering fraction as a free parameter for the narrow  $H_2$  component yields a covering factor of  $C_f \sim 60\%$  and a slight improvement in the fit, most noticeably for a few  $H_2$  lines (Fig. 5). This provides an independent, transverse indication of the small cloud size. While the formal preference of the model with partial covering factor is high ( $\Delta AIC \approx 45$ , where AIC is the Akaike information criterion), the improvement is modest in a visual sense<sup>4</sup>. Higher signal-to-noise (S/N) data will be required to robustly confirm partial coverage and to fully disentangle it from possible modelling systematics that could arise from component decomposition, line spread function, local continuum placement, or local deviation in noise properties.

In this scenario, the inferred column densities in all rotational levels increase by  $\approx 0.3$  dex (Table 3), with no noticeable impact on the derived physical conditions. The inferred cloud size becomes about twice as large, and the predicted metal column densities increase accordingly, bringing them closer to the observed values.



**Fig. 5.** Comparison of best-fit model including partial coverage for the narrow component (red) and assuming full coverage (violet). Only a subset of  $H_2$  lines is shown, focusing on transitions where profile differences are most pronounced. Graphical elements follow Fig. 2. The horizontal dashed green line depicts  $(1 - C_f)$ , the fraction of the background source uncovered by the narrow  $H_2$  component.

**Table 3.** Results from fitting  $H_2$  lines allowing for partial coverage for the narrow component.

Component	Narrow	Broad
$z$	4.2427455(1)	4.242795(7)
$b$ (km s <sup>-1</sup> )	$1.59 \pm 0.12$	$6.2 \pm 0.4$
$\log N(H_2, J = 0)$	$13.49 \pm 0.06$	$12.67 \pm 0.23$
$\log N(H_2, J = 1)$	$14.20 \pm 0.06$	$13.63 \pm 0.04$
$\log N(H_2, J = 2)$	$13.52 \pm 0.06$	$13.25 \pm 0.09$
$\log N(H_2, J = 3)$	$13.59 \pm 0.07$	$13.31 \pm 0.09$
$C_f$	$0.57 \pm 0.04$	

## 5. Discussions and conclusion

As the inferred incident UV radiation field is close to the Galactic value ( $I_{UV} \approx 1$ ), the absorber likely lies far from the quasar influence. Given the quasar’s brightness ( $i = 17.2$ ) and redshift ( $z \approx 4.27$ ), its UV luminosity ( $L_\nu \sim 3 \times 10^{32} \text{ erg s}^{-1} \text{ Hz}^{-1}$ ) would produce a Galactic-level field only beyond  $\sim 2$  Mpc, which is consistent with the  $\sim 1500 \text{ km s}^{-1}$  velocity offset between the quasar and the DLA ( $\sim 3$  Mpc for pure Hubble flow). This illustrates that proximity in velocity space does not necessarily imply physical proximity. Conversely, the inferred UV field can be explained by the quasar alone, without requiring additional sources. This, in turn, suggests that the absorbing gas may reside far away from any luminous galaxy.

The physical conditions inferred for the  $H_2$ -bearing gas are broadly consistent with those reported for other  $H_2$ -

<sup>4</sup> We checked that it is not only driven by the strongest transition, W1-0Q1; see Fig. 5.

bearing DLAs (e.g. Petitjean et al. 2000; Cui et al. 2005; Noterdaeme et al. 2007b; Jorgenson et al. 2010; Albornoz Vásquez et al. 2014; Noterdaeme et al. 2017; Balashev et al. 2019; Klimenko & Balashev 2020; Balashev & Kosenko 2024), as well as for diffuse molecular gas observed along high-latitude lines of sight through the Milky Way disc and halo (Richter et al. 2003; Gillmon et al. 2006; Wakker 2006). However, the gas densities inferred here – particularly for the narrow component – lie towards the upper end of the range typically reported in previous studies. This relatively high density allows trace amounts of H<sub>2</sub> to form in a low-metallicity environment without self-shielding. In this optically thin regime, the gas density determines both the kinetic temperature and the H<sub>2</sub> excitation. By contrast, the broader component has a density closer to those commonly inferred in the literature, although such densities are usually associated with much higher molecular columns; its temperature is also substantially higher. What distinguishes the present system, therefore, is not extreme physical conditions per se, but the presence of dense, although almost fully atomic, gas.

Even in the cold clump, the molecular fraction remains indeed diminute ( $\log f_{\text{H}_2} < -5$ ), highlighting H<sub>2</sub> as a sensitive overdensity tracer in low-metallicity neutral gas. Here, it could trace a very compact structure, with a density of  $\log n_{\text{H}} \sim 2.8$  over a length scale of only  $\sim 0.01$  pc. Yet, the detection of such an absorber among only seven DLAs in the EQUALS sample suggests that these overdensities may be relatively common. This system also likely reflects key evolutionary effects: a higher average gas density at these redshifts and a lower metallicity. The latter reshapes the thermal balance, restricting the formation of cold gas to higher densities because of the reduced metal-line cooling.

A second example is found at  $z = 2.97$  in the EQUALS spectrum of J0204–3251 (Appendix A). Despite its lower S/N and more uncertain measurements, this system presents similar properties: a relatively low H<sub>2</sub> column density distributed over two closely spaced components and a broader component ( $b \sim 2.6$  km s<sup>-1</sup>) more highly excited than the narrow one ( $b \sim 0.8$  km s<sup>-1</sup>). Cold overdensities could thus actually be frequent enough in the neutral medium at high redshift to collectively build a significant covering fraction.

Individual CNM cloudlets would likely remain undetectable in 21-cm absorption, even with the Square Kilometer Array, because their physical sizes are much smaller than the extent of radio sources. A sufficiently large covering factor could in principle make their collective 21-cm imprint detectable, despite their low volume-filling factor, but the signal could remain shallow, as clouds may spread over a relatively broad velocity interval. In the optical, such compact structures are also challenging to identify at lower spectral resolution or S/N than those achieved here. Their detection in this system was only possible thanks to the achieved resolving power combined with the brightness of the background quasar. Even the broader  $\sim 4$ -pc component would have remained hidden without the narrow feature flagging the presence of cold gas. ESPRESSO’s sensitivity to narrow absorption lines thus provides a rare probe of the small-scale density structure of the neutral medium at high redshift. Looking ahead, using weak H<sub>2</sub> as a routine tracer of CNM cloudlets in low metallicity gas at high redshift (potentially leading to star formation before reaching high molecular fraction; Krumholz 2012; Glover & Clark 2012) will require much larger collecting areas, which will be provided by ANDES on the ELT.

Re-observing known H<sub>2</sub>-bearing systems at higher spectral resolution will also help isolate multi-phase structures and uncover weak H<sub>2</sub> components that are otherwise difficult to detect.

*Acknowledgements.* We thank the referee for useful comments and suggestions. We thank IFPU in Trieste for hospitality. SL acknowledges support by FONDECYT grant 1231187.

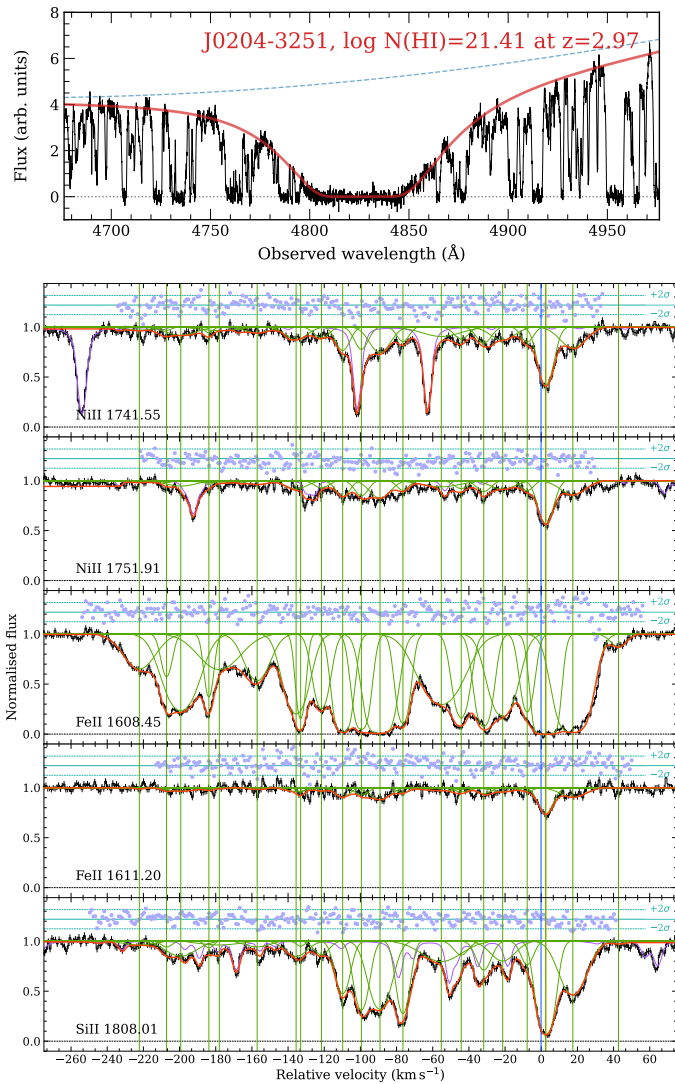
## References

- Abgrall, H., Le Bourlot, J., Pineau Des Forets, G., et al. 1992, *A&A*, 253, 525  
 Albornoz Vásquez, D., Rahmani, H., Noterdaeme, P., et al. 2014, *A&A*, 562, A88  
 Balashev, S. A., & Kosenko, D. N. 2024, *MNRAS*, 527, 12109  
 Balashev, S. A., Petitjean, P., Ivanchik, A. V., et al. 2011, *MNRAS*, 418, 357  
 Balashev, S. A., Klimenko, V. V., Noterdaeme, P., et al. 2019, *MNRAS*, 490, 2668  
 Balashev, S. A., Telikova, K. N., & Noterdaeme, P. 2022, *MNRAS*, 509, L26  
 Bellomi, E., Godard, B., Hennebelle, P., et al. 2020, *A&A*, 643, A36  
 Berg, T., D’Odorico, V., Boera, E., et al. 2025, *The Messenger*, 195, 23  
 Bialy, S., Burkhart, B., & Sternberg, A. 2017, *ApJ*, 843, 92  
 Boutsia, K., Grazian, A., Calderone, G., et al. 2020, *ApJS*, 250, 26  
 Chatzikos, M., Bianchi, S., Camilloni, F., et al. 2023, *Rev. Mexicana Astron. Astrofis.*, 59, 327  
 Cuellar, R., Noterdaeme, P., Balashev, S., et al. 2025, *A&A*, 694, A294  
 Cui, J., Bechtold, J., Ge, J., & Meyer, D. M. 2005, *ApJ*, 633, 649  
 Cupani, G., D’Odorico, V., Cristiani, S., et al. 2020, *SPIE Conf. Ser.*, 11452, 114521U  
 Gillmon, K., Shull, J. M., Tumlinson, J., & Danforth, C. 2006, *ApJ*, 636, 891  
 Glover, S. C. O. 2003, *ApJ*, 584, 331  
 Glover, S. C. O., & Clark, P. C. 2012, *MNRAS*, 421, 9  
 Godard, B., Falgarone, E., & Pineau Des Forêts, G. 2009, *A&A*, 495, 847  
 Gurman, A., Sternberg, A., Bialy, S., Cochrane, R. K., & Stern, J. 2025, *ApJ*, 995, 116  
 Habart, E., Boulanger, F., Verstraete, L., et al. 2003, *A&A*, 397, 623  
 Indriolo, N., Geballe, T. R., Oka, T., & McCall, B. J. 2007, *ApJ*, 671, 1736  
 Jenkins, E. B., & Peimbert, A. 1997, *ApJ*, 477, 265  
 Jorgenson, R. A., Wolfe, A. M., Prochaska, J. X., & Carswell, R. F. 2009, *ApJ*, 704, 247  
 Jorgenson, R. A., Wolfe, A. M., & Prochaska, J. X. 2010, *ApJ*, 722, 460  
 Jura, M. 1975, *ApJ*, 197, 575  
 Khaire, V., & Srianand, R. 2019, *MNRAS*, 484, 4174  
 Klimenko, V. V., & Balashev, S. A. 2020, *MNRAS*, 498, 1531  
 Krumholz, M. R. 2012, *ApJ*, 759, 9  
 Larson, R. B. 1981, *MNRAS*, 194, 809  
 Ledoux, C., Petitjean, P., & Srianand, R. 2003, *MNRAS*, 346, 209  
 Ledoux, C., Petitjean, P., & Srianand, R. 2006, *ApJ*, 640, L25  
 Lesaffre, P., Todorov, P., Levrier, F., et al. 2020, *MNRAS*, 495, 816  
 Levshakov, S. A., & Varshalovich, D. A. 1985, *MNRAS*, 212, 517  
 Morton, D. C. 2003, *ApJS*, 149, 205  
 Muzahid, S., Srianand, R., & Charlton, J. 2015, *MNRAS*, 448, 2840  
 Noterdaeme, P., Ledoux, C., Petitjean, P., et al. 2007a, *A&A*, 474, 393  
 Noterdaeme, P., Petitjean, P., Srianand, R., Ledoux, C., & Le Petit, F. 2007b, *A&A*, 469, 425  
 Noterdaeme, P., Ledoux, C., Petitjean, P., & Srianand, R. 2008, *A&A*, 481, 327  
 Noterdaeme, P., Krogager, J.-K., Balashev, S., et al. 2017, *A&A*, 597, A82  
 Pepe, F., Cristiani, S., Rebolo, R., et al. 2021, *A&A*, 645, A96  
 Petitjean, P., Srianand, R., & Ledoux, C. 2000, *A&A*, 364, L26  
 Ranjan, A., Noterdaeme, P., Krogager, J. K., et al. 2018, *A&A*, 618, A184  
 Richter, P., Wakker, B. P., Savage, B. D., & Sembach, K. R. 2003, *ApJ*, 586, 230  
 Roman-Duval, J., Jenkins, E. B., Tchernyshyov, K., et al. 2022, *ApJ*, 935, 105  
 Savage, B. D., Bohlin, R. C., Drake, J. F., & Budich, W. 1977, *ApJ*, 216, 291  
 Shull, J. M., Tumlinson, J., Jenkins, E. B., et al. 2000, *ApJ*, 538, L73  
 Smette, A., Sana, H., Noll, S., et al. 2015, *A&A*, 576, A77  
 Stern, J., Sternberg, A., Faucher-Giguère, C.-A., et al. 2021, *MNRAS*, 507, 2869  
 Tchernyshyov, K. 2022, *ApJ*, 931, 78  
 Valdivia, V., Hennebelle, P., Gérin, M., & Lesaffre, P. 2016, *A&A*, 587, A76  
 Villa-Vélez, J. A., Godard, B., Guillard, P., & Pineau des Forêts, G. 2024, *A&A*, 688, A96  
 Wakelam, V., Bron, E., Cazaux, S., et al. 2017, *Mol. Astrophys.*, 9, 1  
 Wakker, B. P. 2006, *ApJS*, 163, 282  
 Wolf, C., Hon, W. J., Bian, F., et al. 2020, *MNRAS*, 491, 1970  
 Yu, Z., Martini, P., Davis, T. M., et al. 2020, *ApJS*, 246, 16

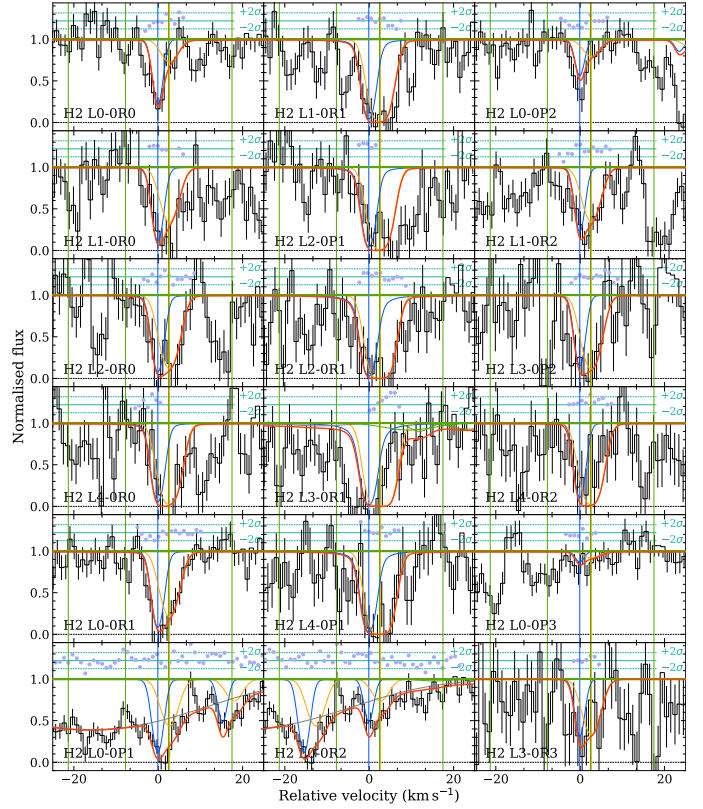
## Appendix A: H<sub>2</sub> at $z \sim 3$ towards J0204-3251

The quasar J0204-3251 ( $z_Q \approx 3.8$ ) has also been observed with ESPRESSO as part of the EQUALS survey. It features a strong intervening DLA at  $z_{\text{abs}} \approx 2.97$  with  $\log N(\text{H I}) = 21.41 \pm 0.01$ . The metal-line profile is complex, with numerous components spanning  $\sim 250 \text{ km s}^{-1}$ , and indicates an overall metallicity of  $[\text{Si}/\text{H}] \approx -0.86$  (Fig. A.1). Because silicon is mildly depleted, this value likely represents a lower limit to the true metallicity; Zn II lines are not covered and S II lines fall within the Ly $\alpha$  forest.

Although this spectrum has a lower S/N (particularly in the blue) and covers fewer H<sub>2</sub> bands than the main system discussed in this paper, the higher H<sub>2</sub> column density enabled the detection of two closely spaced components ( $\Delta v \approx 2.8 \text{ km s}^{-1}$ ) from  $J = 0$  to  $J = 3$  (Fig. A.2 and Table A.1). The corresponding excitation diagram is shown in Fig. A.3.



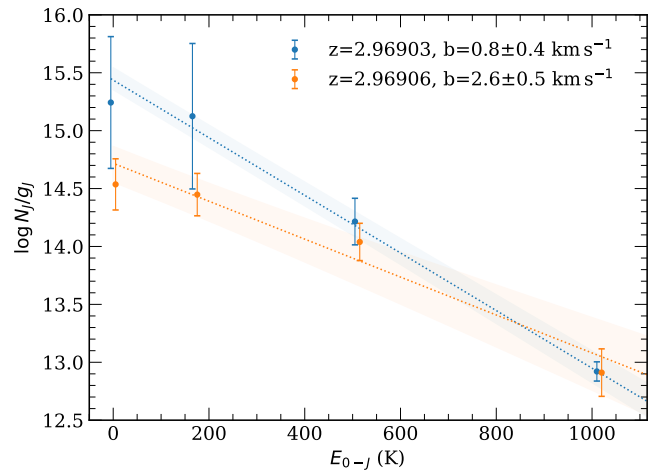
**Fig. A.1.** Fit to H I Ly- $\alpha$  and low-ionisation metals lines in the  $z = 2.97$  DLA system towards J0204-3251. The zero of the velocity scale corresponds to the strongest H<sub>2</sub> component. Colours are as per Fig. 2.



**Fig. A.2.** H<sub>2</sub> absorption lines towards J0204-3251.

**Table A.1.** Results from fitting H<sub>2</sub> lines towards J0204-3251.

$z$	2.969028	2.969063
$b$ (km s <sup>-1</sup> )	$0.8 \pm 0.4$	$2.6 \pm 0.5$
$\log N(\text{H}_2, J = 0)$	$15.2 \pm 0.6$	$14.5 \pm 0.2$
$\log N(\text{H}_2, J = 1)$	$16.1 \pm 0.6$	$15.4 \pm 0.2$
$\log N(\text{H}_2, J = 2)$	$14.9 \pm 0.2$	$14.7 \pm 0.2$
$\log N(\text{H}_2, J = 3)$	$14.2 \pm 0.1$	$14.2 \pm 0.2$
total	$16.2 \pm 0.5$	$15.6 \pm 0.1$



**Fig. A.3.** H<sub>2</sub> excitation diagram at  $z = 2.97$  towards J0204-3251.