Cool and Cold Gas within and surrounding Galaxies

Roland Szakacs

SIGILL: UNIVERSITAT: LUDOVIC: MAXIM:

München 2022

Cool and Cold Gas within and surrounding Galaxies

Roland Szakacs

Dissertation
an der Fakultät für Physik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Roland Szakacs
aus Wels, Österreich

München, den 05.12.2022

Erstgutachter: Prof. Dr. Volker Springel
Zweitgutachter: Prof. Dr. Daniel Grün
Tag der mündlichen Prüfung: 26.01.2023

The Road goes ever on and on
Down from the door where it began.
Now far ahead the Road has gone,
And I must follow, if I can,
Pursuing it with eager feet,
Until it joins some larger way
Where many paths and errands meet.
And whither then? I cannot say.

Bilbo Baggins in The Fellowship of the Ring by J.R.R Tolkien

To my child, Yves
and
To my parents, Emese and Arthur

This thesis has been carried out at the European Southern Observatory under the supervision of Dr. Céline Péroux and Dr. Martin Zwaan.

Table of Contents

List of Figures

List of Tables

Zusammenfassung

Galaxien sind Systeme, die sich durch eine Kombination aus internen Prozessen und ihrer Verbindung zu ihrer unmittelbaren Umgebung entwickeln. Mechanismen wie die Rückkopplung aktiver galaktischer Kerne (AGN), stellare Rückkopplung, wiederverwertung von Gas und Akkretion aus dem kosmischen Netz führen zu einer Umverteilung von Baryonen innerhalb von Galaxien und in ihrer Umgebung. Diese Prozesse interagieren im mehrphasigen zirkumgalaktischen Medium (CGM), definiert als das Gas, das Galaxien außerhalb der Scheibe oder des interstellaren Mediums (ISM), aber innerhalb des Virialradius der Galaxien umgibt. Das CGM ermöglicht somit die Verfolgung von Gasströmen, die im Zusammenhang mit der Entwicklung von Galaxien und der Sternentstehung von besonderer Bedeutung sind. Die Entfernung von Gas durch gewaltsame Rückkopplungsprozesse könnte zum Erlöschen der Sternentstehung in Galaxien führen, während die Akkretion für die Aufrechterhaltung der Sternentstehung entscheidend ist. Das akkretierte Gas vermischt sich mit der Materie innerhalb der Galaxien und kann anschließend abkühlen und zu Molekülwolken kollabieren, was zur Sternentstehung führt. Somit spielen das CGM, die kalte Gasphase, und ihr Vorläufer, die kühle Gasphase, eine Schlüsselrolle bei der Entwicklung von Galaxien im Laufe der kosmischen Zeit. Ziel dieser Arbeit ist es, die kalte und kühle Gasphase, die das molekulare, neutrale atomare und schwach ionisierte atomare Gas umfasst, in und rundum Galaxien über die kosmische Zeit hinweg zu studieren.

Mithilfe von mehrphasigen Beobachtungen der MUSE-ALMA Haloes Survey untersuchen wir das kalte molekulare Gas einer anhand von H I-Absorption selektierten Galaxie bei $z \sim 0.4$. Zusätzlich verfolgen wir Gasflüsse im CGM durch kinematische Analysen des niedrig ionisierten und molekularen Gases dieser Galaxie, welches in Emission und Absorption beobachtet wurde. Wir stellen fest, dass die Phasen des ionisierten und des molekularen Gases innerhalb der rotierenden Scheibe der Galaxie stark gekoppelt sind, und dass ein zwei komponentiges Absorptionssystem in 52 kpc Entfernung von der Galaxie mit einströmendem und mitrotierendem Gas vereinbar ist. Darüber hinaus zeigt eine umfassende Literaturzusammenstellung von Galaxien ($z < 1.1$) die einerseits anhand von H I-Absorption und andererseits anhand von Emission selektiert wurden, dass die Auswahl von Galaxien anhand H I-Absorption möglicherweise Objekte aufspürt, die über größere molekulare Gasreservoirs verfügen als andere Galaxien mit vergleichbaren Sternentstehungsraten. Dies führt zu Erschöpfungszeiten des molekularen Gases, die mehr als eine Größenordnung größer sein können als die mittleren Erschöpfungszeiten in Stichproben die anhand von Emission selektiert wurden.

Beobachtungen der kalten Gasphase über die kosmische Zeit hinweg deuten auf eine deutliche Zunahme der Massendichte des molekularen Gases in Richtung $z \sim 2-3$ hin. Diese Entwicklung impliziert eine damit einhergehende Änderung der globalen Verteilung der Säulendichte von molekularem Wasserstoff (H_2), die derzeit noch nicht erforscht ist. Um dieser Frage nachzugehen, verwenden wir einen Ansatz, der Beobachtungen (PHANGS-ALMA, SDSS) und Simulationen (Illustris-Project, GRIFFIN-Project) miteinander verbindet, um die Entwicklung der Säulendichten von molekularem Gas zu untersuchen. Diese Studie zeigt, dass die H_2-Säulendichte, die am meisten zur Gesamtmasse des molekularen Gases beiträgt, sich mit der Rotverschiebung entwickelt. Bei $z = 3$ ist mehr molekulares Gas in dichteren Regionen zu finden als bei $z = 0$. Dies steht im Einklang mit einer höheren Dichte der Sternentstehungsrate bei $z = 3$, da mehr molekulares Gas in dichteren Zuständen wahrscheinlich zu einer höheren globalen Sternentstehungsrate beiträgt. Vergleiche der Verteilung der H I- und H_2-Säulendichten deuten außerdem an, dass H I einen großen Anteil zur Gesamtmasse des kalten Gases im ISM von Galaxien beiträgt.

Künftig werden groß angelegte spektroskopische Absorptionslinien-Durchmusterungen neue Perspektiven für die Kartierung des CGM bieten. Solche großen Durchmusterungen, wie die bevorstehende VISTA/4MOST High-resolution Quasar Survey (4Hi-Q), erfordern neuartige Analysewerkzeuge, um diese Absorptionslinien in den Spektren effizient zu erkennen und zu lokalisieren. Zu diesem Zweck entwickeln wir ein solches Werkzeug, das auf Convolutional Neural Networks (CNN) basiert. In Ermangelung von Daten erstellen wir auf dem 4MOST Instrument basierende synthetische normalisierte Quasarspektren und injizieren Mg II ($\lambda 2796$, $\lambda 2803$) Absorber, die auf der TNG50-Simulation des Illustris-Projects basieren, in diese Spektren. Anhand dieser synthetischen Spektren trainieren wir ein CNN, um zu klassifizieren, ob ein Mg II-Absorber in den Spektren vorhanden ist und um deren Wellenlängenposition zu bestimmen. Das CNN hat eine hohe Klassifizierungs- (98 Prozent) und Lokalisierungsgenauigkeit (mittlerer absoluter Fehler von 6.9 Å) für eine Stichprobe mit gleichmäßig verteilten Absorber- und Rauscheigenschaften. Außerdem ist diese Methode um Größenordnungen schneller als herkömmliche Methoden. Diese Machbarkeitsstudie zeigt, dass Machine Learning ein praktikables Werkzeug für zukünftige groß angelegte Durchmusterungen ist.

In Zusammenfassung erforschen wir mit diesen drei verschiedenen Projekten das Innenleben des kosmischen Baryonenkreislaufs und des CGM auf einer Skala einzelner Galaxien bis hin zu großen statistischen Stichproben und verfolgen dabei einen Ansatz, der Beobachtungen und Simulationen miteinander verbindet. Außerdem, erproben wir die Verwendbarkeit von Machine Learning für die Analyse groß angelegter Durchmusterungen, um den Weg für die Ära der Big-Data-Astronomie zu ebnen.

Abstract

Galaxies are systems evolving through a combination of internal processes and their connection to their immediate surroundings. Mechanisms, such as Active Galactic Nuclei (AGN) feedback, stellar feedback, recycling of gas, and accretion from the cosmic web lead to a redistribution of baryons within and surrounding galaxies. These processes interact in the multi-phase Circumgalactic Medium (CGM), defined as the gas surrounding galaxies outside the disk or interstellar medium (ISM), but within the virial radius of galaxies. Thus, the CGM allows for the tracing of gas flows, which are particularly significant in the context of galaxy evolution and star formation. The removal of gas by violent feedback processes could lead to the quenching of star formation, while accretion is critical to sustaining star formation. The accreted gas mixes with the matter within galaxies and can subsequently cool down and then collapse into molecular clouds, leading to the formation of stars. Thus, the CGM, the cold gas phase, and its precursor, the cool gas phase, play key roles in how galaxies evolve across cosmic time. The goal of this thesis is to explore the cool and cold gas phases, encompassing molecular, neutral atomic, and low-ionized atomic gas within and surrounding galaxies across cosmic time.

Exploiting multi-phase observations of the MUSE-ALMA Haloes survey we study the cold molecular gas of a H I-absorption-selected galaxy. Additionally, we trace gas flows in the CGM by kinematically analyzing the low-ionized and molecular gas of a galaxy at $z \sim 0.4$ in emission and absorption. We find that the ionized and molecular gas phases are strongly coupled within the rotating disk of the galaxy and that a two-component absorption feature at 52 kpc distance from the galaxy is consistent with being inflowing and co-rotating gas. Further, a comprehensive literature compilation of H I-absorption and emission-selected galaxies at $z < 1.1$ reveals that H I-absorption selection of galaxies possibly traces objects that have large molecular gas reservoirs given their star formation rate. This leads to depletion times that can be over an order of magnitude larger than the median depletion times in samples of emission-selected galaxies.

Observations of the cold gas phase across cosmic time indicate a marked increase in the molecular gas mass density towards $z \sim 2-3$. This transformation implies an accompanied change in the global distribution of molecular hydrogen (H_2) column densities, which is currently not yet explored. To tackle this question, we use an approach that bridges observations (PHANGS-ALMA, SDSS) and simulations (Illustris Project, GRIFFIN Project) to explore this evolution of molecular gas column densities. This study reveals that the H_2 column density contributing most to the overall molecular gas mass evolves with redshift. We find that more molecular gas is found in denser regions at $z = 3$ compared to $z = 0$. This finding is consistent with a higher star formation rate density at $z = 3$, as more molecular gas in denser states is likely contributing to a higher global star formation rate. Further, comparisons of the distribution of H I and H_2 column densities lead to the conclusion that H I is an important contributor to the overall cold gas mass found in the ISM of galaxies at both redshifts.

Going forward, large-scale spectroscopic absorption-line surveys will offer new prospects for mapping the CGM surrounding galaxies. Such large surveys, as the upcoming VISTA/4MOST High-resolution Quasar survey (4Hi-Q), will require novel analysis tools to efficiently detect and localize these absorption lines in spectra. For this purpose, we develop such a tool based on Convolutional Neural Networks (CNN). In the absence of data, we produce synthetic normalized 4MOST high-resolution fibre quasar spectra and inject Mg II ($\lambda2796$, $\lambda2803$) absorbers based on the TNG50 simulation of the Illustris Project within them. Using these synthetic spectra, we train a CNN to classify whether an Mg II absorber is present within spectra and to localize them in wavelength space. The CNN has high classification (98 per cent) and localization (mean absolute error of 6.9 Å) accuracy for a sample with evenly distributed absorber and noise properties. Additionally, this method is orders of magnitude faster than traditional ones. This proof-of-concept study demonstrates that Machine Learning is a feasible tool for future large-scale surveys.

In summary, with these three distinct projects, we explore the inner workings of the cosmic baryon cycle and the CGM from the scale of individual galaxies to large statistical samples using an approach that bridges observations and simulations. Additionally, we explore the feasibility of novel methods for the analysis of large-scale surveys to pave the way for the era of big data astronomy.

Chapter 1

Introduction

Astronomy is deeply intertwined with the history of humankind. As the oldest natural science, the exploration of the night sky has fascinated cultures from all over the world. The first recordings of star charts and catalogues by Egyptian and Babylonian astronomers go back as far as 1500 BC. Humans have not only recorded the position of stars but also questioned the structure of our surroundings. Around 400 BC the Greek philosopher Democritus suggested that the Milky Way consists of stars. While many other philosophers and astronomers suggested similar theories, Galileo Galilei was the first to record his observations about the Milky Way consisting of countless stars with the revolutionary tool of the telescope in 1610. Around the same time, Johannes Kepler formulated the laws of planetary motion, paving the way for Isaac Newton's law of gravity that revolutionized the modern natural sciences half a century thereafter. Within one century the horizons of our understanding of our place in the Universe broadened even further. In 1750 Thomas Wright suggested that faint nebulae on the night sky are actually galaxies. It took almost two centuries to confirm this idea. Based on the work on Cepheids by Henrietta Leavitt, Edwin Hubble was able to calculate that the nebula now called the Andromeda galaxy, was beyond the bounds of the Milky Way. This ushered in the era of extragalactic astronomy. Since this discovery, scientists have explored not only the structures of galaxies themselves, but also how they form within the large-scale structure of the Universe using sophisticated spectroscopic instruments at large observatories, theoretical models, and numerical simulations.

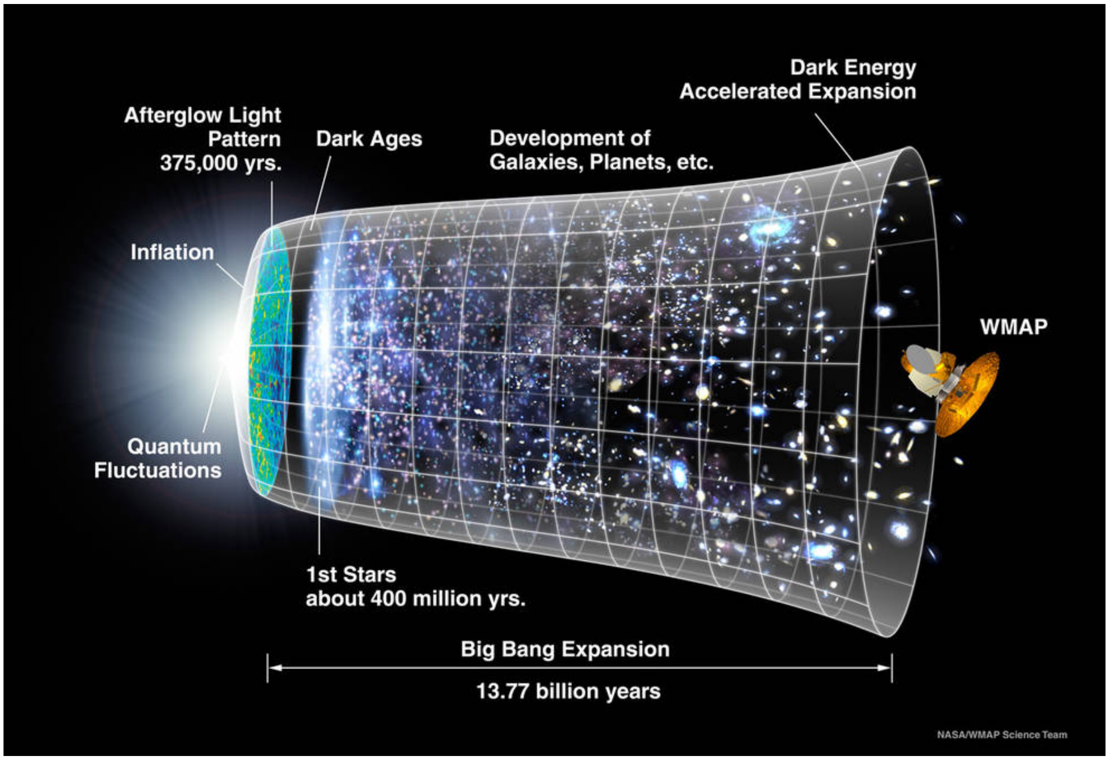

Figure 1.1: A sketch displaying the timeline for the expansion and evolution of the Universe. Beginning with a hot big bang quantum fluctuations and a rapid inflationary period provide the seeds for the formation of structures such as galaxies and cosmic filaments observed in the present Universe. Image Credit: NASA/WMAP Science Team

1.1 The Formation and Evolution of Galaxies

The basis of modern cosmology and astrophysics is built on the Λ Cold Dark Matter (ΛCDM) model. The model assumes that the Universe has four constituents [radiation (photons and neutrinos), ordinary matter (baryons and leptons), cold dark matter, and dark energy]. Additionally, it is assumed that the Universe is described by general relativity (Einstein, 1917), is statistically homogeneous and isotropic at large scales ($\gtrsim$ 100 Mpc), is flat, and described by the Friedmann-Lemaître-Robertson-Walker metric (for a current review see e.g. Perivolaropoulos & Skara, 2022). Starting from the hot big bang (Gamow, 1946) approximately 13.8 billion years ago (Planck Collaboration et al., 2020) the Universe rapidly expanded leading to a subsequent formation of structures such as planets, stars, galaxies, and cosmic filaments (see Fig. 1.1).

1.1.1 Structure Formation

The basis for structures in the Universe is thought to be laid by quantum fluctuations in the early, rapidly expanding Universe, leading to adiabatic, scale-invariant density perturbations (e.g. Guth, 1981; Linde, 1982; Hawking, 1982; Starobinsky, 1982). Evidence of these density fluctuations is imprinted in the cosmic microwave background (CMB, Alpher & Herman, 1948; Penzias & Wilson, 1965). The CMB consists of radiation from approximately 380 000 years after the big bang, the point of time when the Universe became transparent to photons as it cooled down enough for electrons to combine with protons to form hydrogen atoms (the era of recombination). These small density fluctuations have been observed in the form of temperature fluctuations within the CMB ($\Delta T/T \sim 10^{-5}$) by e.g. WMAP (Bennett et al., 2003) and Planck Satellites (Planck Collaboration et al., 2020).

An additional important factor to structure formation is the presence of dark matter. To explain the structures that we see in the present Universe, particles that do not interact with electromagnetic radiation are needed, as the small density fluctuations already need to grow before decoupling (e.g. Peebles, 1982). First observational evidence for dark matter was detected in the Coma Cluster by Zwicky (1933) who hypothesized that the galaxies must be held together by some sort of dark matter as they were moving too fast to be bound together by the observed baryonic matter. Subsequent observations of rotation curves of galaxies, which, contrary to theoretical predictions, flattened instead of declined towards the outskirts and observations of the orbits of satellite galaxies provided additional evidence for the existence of dark matter (e.g. Rubin & Ford, 1970; Ostriker et al., 1974; Einasto et al., 1974).

The combination of an early rapidly expanding Universe, small density perturbations, and dark matter led to the formation of the structures we observe in the present Universe. The small density perturbations develop over-dense regions of dark matter that hierarchically cluster due to the force of gravity to form progressively larger systems collapsing into dark matter haloes. Subsequently, as baryonic matter cools down it settles within the gravitational potential wells of these haloes leading to the formation of the first galaxies that assemble within groups and clusters and evolve across cosmic time (White & Rees, 1978).

(Magneto-)hydrodynamical cosmological simulations such as simulations by the Illustris Project (e.g. Vogelsberger et al., 2014a; Springel et al., 2018), EAGLE Project (e.g. Crain et al., 2015; Schaye et al., 2015) and the SIMBA project (e.g. Davé et al., 2019) have played a key part in furthering our understanding of structure formation in the Universe. These simulations evolve from given initial conditions at high redshifts to the current Universe. They include a large range of physical processes such as gas radiative effects, primordial and metal-line cooling, heating by the meta-galactic background radiation field, star formation, supernova feedback, the formation and merging of supermassive black holes and feedback of those black holes. Such models have successfully reproduced key observables such as the low-redshift two-point galaxy correlation function describing the distribution of galaxies in the Universe (e.g. Springel et al., 2018) and the stellar mass function across various redshifts (e.g. Schaye et al., 2015; Pillepich et al., 2018a; Davé et al., 2019). Thus, these simulations provide important tools to explore the accuracy of theoretical models of structure formation and subsequently galaxy formation and evolution. Further, they can provide important tools to aid the planning of future observations by providing predictions of e.g. the flux of emission lines (e.g. Augustin et al., 2019) by post-processing these simulations by using spectral synthesis codes such as CLOUDY (Ferland et al., 2017).

1.1.2 Galaxy Evolution

In the picture of ΛCDM and hierarchical clustering, galaxy mergers are one of the fundamental drivers of galaxy evolution (e.g. White & Rees, 1978; Blumenthal et al., 1984). Especially at higher redshifts high merger fractions are observed for galaxies (e.g. Conselice, 2003; Man et al., 2012; Mortlock et al., 2013) demonstrating that mergers are a common event in our Universe and lead to the growth of galaxies. In particular, minor mergers are likely important events for the formation of classical bulges, a spherical central concentration of stars within the center of disk galaxies (e.g. Brooks & Christensen, 2016). Further, the violent process of major mergers also leads to the formation of elliptical galaxies (e.g. Toomre, 1977; Joseph & Wright, 1985; Schweizer, 1989). Contrary to spiral galaxies, which are typically blue and star-forming, elliptical galaxies are redder and more massive and display little ongoing star formation (Holmberg, 1958).

At lower redshifts ($z < 1$) the reacceleration of the Universe (Riess et al., 1998) leads to an increasing isolation and thus decrease of the merger rate in galaxies. In isolated disk galaxies, internal dynamical processes can lead to a shift in mass profiles and metallicity distributions over large timescales (e.g. Sellwood, 2014). Nonetheless, in denser environments such as groups and clusters of galaxies, external processes still lead to significant changes within galaxy properties as they might lose a significant amount of gas by ram-pressure stripping (e.g. Peng et al., 2010), harassment (e.g. Moore et al., 1996) and strangulation (e.g. Peng et al., 2015).

It is evident that galaxies evolve in connection with their surroundings. Not only mergers are important for the growth of galaxies, but gas accretion through the intergalactic medium is an additional important factor (e.g. Conselice et al., 2013). Additionally, galaxies can also expel gas through violent internal processes such as AGN and stellar feedback. Evidence of these feedback processes have not only been observed with large telescopes (e.g. Shull et al., 2014; Ginolfi et al., 2020), but also have been explored in cosmological simulations as outlined in the previous section and also high-resolution (zoom-) simulations of galaxies and galactic haloes [e.g. GRIFFIN Project (e.g. Hu et al., 2014a; Steinwandel et al., 2020), AURIGA (e.g Grand et al., 2017; van de Voort et al., 2019)]. This gas expelled from galaxies and accreted from the cosmic web interacts within the surroundings of galaxies - the so-called Circumgalactic Medium.

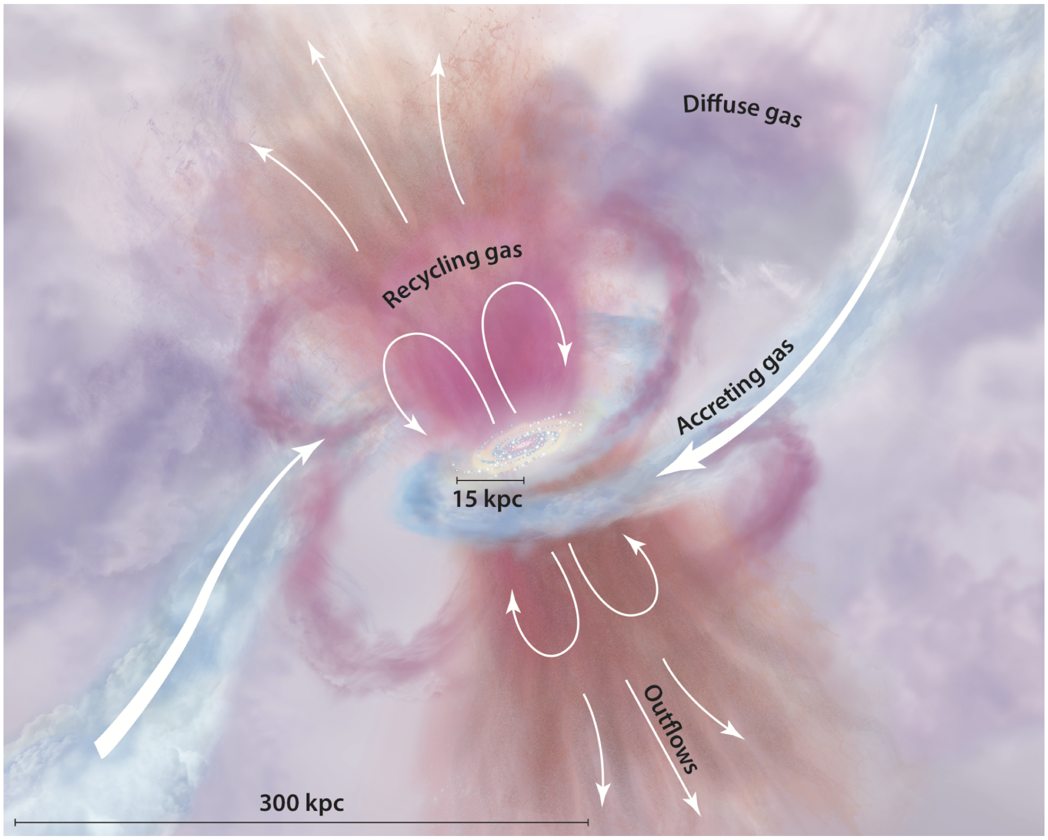

Figure 1.2: A sketch of the Circumgalactic Medium (CGM) surrounding galaxies displaying the rich dynamics of the multi-phase gas. The CGM acts as a transition zone between the ISM and IGM. The sketch displays outflows in brown, accreting gas in blue, recycling gas in pink, and the diffuse halo in purple (Tumlinson et al., 2017).

1.2 The Circumgalactic Medium

The Circumgalactic Medium (CGM) is the transition zone between the Interstellar Medium (ISM) and the Intergalactic Medium (IGM). It is loosely defined as the baryonic matter outside the ISM, however still within the virial radius of a galaxy (Tumlinson et al., 2017).

The CGM is a system of rich dynamics (see Fig. 1.2). Both observations and simulations have found evidence for gas accreting onto (e.g. Fumagalli et al., 2011; Rubin et al., 2012; Martin et al., 2012; Hafen et al., 2017; Zabl et al., 2019; Szakacs et al., 2021), being expelled from (e.g. Veilleux et al., 2005; Rubin et al., 2012; Bordoloi et al., 2014; Ginolfi et al., 2020; Nelson et al., 2019b; Costa et al., 2022) and gas being recycled by outflows raining back onto galaxies (e.g. Christensen et al., 2016; Fraternali, 2017; Bish et al., 2019).

The CGM also displays rich multi-phase gas with a large range of temperatures and ionization states. It consists of a cold gas phase ($T \leq 10^4$ K) traced by low ions of metal species (e.g. Na I, Ca II, ...), neutral atomic and molecular hydrogen, and dust (e.g. Zheng et al., 2020; Li et al., 2021; Szakacs et al., 2021; Damle et al., 2022), a cool gas phase ($T \sim 10^{4-5}$ K) traced by other low ions of metal species (e.g. Mg II, Si II, Si III, N II, ...; Steidel et al., 2010; Werk et al., 2014; Szakacs et al., 2021; Augustin et al., 2021), a warm gas phase ($T \sim 10^{5-6}$ K) traced through metal ions such as C IV, N V, O VI and Ne VII (e.g. Tumlinson et al., 2011; Bordoloi et al., 2014; Werk et al., 2016) and a hot phase ($T > 10^6$ K) traced by e.g. observations in the X-ray regime (e.g. Anderson & Bregman, 2010; Bregman et al., 2018; Nicastro et al., 2018). Compiled mass density profiles from these multi-phase observations show that the CGM is dominated by the cool gas phase, followed by the hot, warm, and cold phases (Tumlinson et al., 2017).

Studies of the CGM are closely related to current key questions regarding galaxy evolution. Observed star-forming galaxies typically have short depletion time scales ($\tau_{\rm dep} = M_{\rm gas}/{\rm SFR}$) in the range of 1-3 Gyrs depending on redshift and stellar mass (e.g. Saintonge et al., 2011; Scoville et al., 2017; Aravena et al., 2019; Tacconi et al., 2020). However, the population of star-forming galaxies persists for a longer time than these depletion times. Subsequently, galaxies need to accrete additional fuel for star formation, which has to be provided by the IGM and CGM. A related puzzle is the bimodality of the specific star formation rate (sSFR = $\mathrm{SFR/M_*}$) as a function of $\mathrm{M_*}$ (e.g. Schiminovich et al., 2010). In this bimodality, star-forming and passive galaxies are found. It is unclear what leads to this quenching of star formation in galaxies. Possible solutions are the shutting off of IGM accretion or a hot CGM that is not cooling and subsequently is prevented from entering the ISM. For satellite galaxies, ram-pressure stripping by the halo of the central galaxy is another possibility (Tumlinson et al., 2011). Finally, a large fraction of baryons predicted by the ΛCDM model is not accounted for by observed galaxies (missing baryons problem, see e.g. Fukugita et al., 1998; Bregman, 2007; Shull et al., 2012). These baryons might be found in a hot and diffuse gas within the CGM or even completely outside of galaxy haloes and further observations are needed to solve the missing baryons problem (Tumlinson et al., 2017; Comparat et al., 2022). In the upcoming years, future X-ray missions such as the Lynx X-ray observatory (Gaskin et al., 2019), the X-ray Imaging and Spectroscopy Mission (XRISM, XRISM Science Team, 2020) and the Athena X-ray Observatory (Barret, 2022) will provide observations to tackle these questions.

While the CGM is a highly interesting medium to study due to its complex structure, rich dynamics and given its importance in regards to galaxy evolution, the low surface brightness makes it difficult the explore. Current instruments cannot observe the CGM in emission in most cases (Frank et al., 2012; Augustin et al., 2019; Corlies et al., 2020a). Thus, the vast majority of the studies outlined above trace the CGM in absorption instead.

1.2.1 Exploring the CGM in Absorption

Absorption-line studies provide one of the most powerful tools to explore the CGM. This technique exploits bright background sources, in most cases quasars, to search for intervening systems along the line of sight of the background source. The gas in the intervening systems causes absorption in the spectrum of the background source (see Fig. 1.3). This method offers the advantage of the sensitivity of observations being independent of redshift,

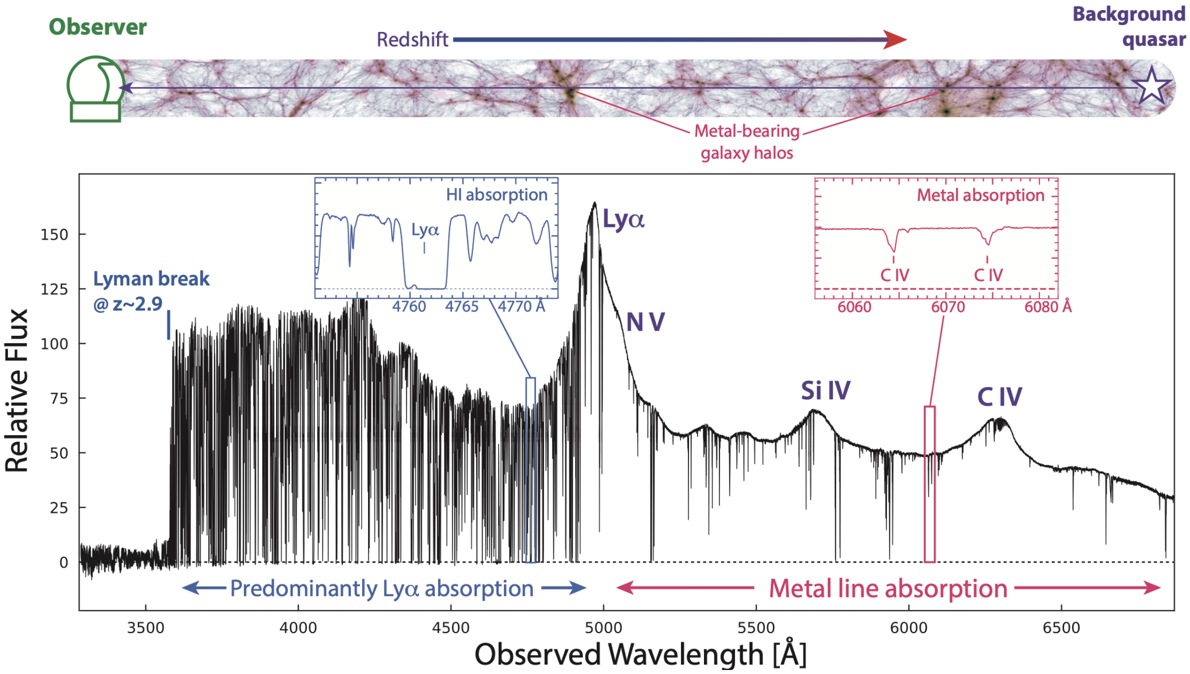

Figure 1.3: A sketch of a typical setup for absorption-line studies of the CGM. A bright background source (in most cases quasars) provides pencil-beam sightlines used for the search of intervening gas. These intervening systems are imprinted through absorption lines in the spectrum of the background source. (Péroux & Howk, 2020)

allowing for probing of the low-density CGM gas even at early epochs (e.g. Tripp et al., 1998; Péroux & Howk, 2020). Nonetheless, it also comes with the disadvantage of only probing pencil beams along the line of sight. However, this limitation can be minimized by obtaining large samples of absorption-line systems to study the CGM in a statistical manner.

Lyman-α absorbers provide an effective way of exploring the atomic gas phase of the CGM. They are typically categorized into three distinct types, depending on the observed H I column density (see e.g. Péroux & Howk, 2020). The strongest absorbers, Damped Lyman-α Absorbers (DLAs; $N_{\mathrm{H\,I}} \geq 10^{20.3}\mathrm{cm}^{-2}$; Wolfe et al., 1986), trace the mostly neutral cold and cool phase of the CGM and parts of the ISM (e.g. Wolfe et al., 1986; Péroux & Howk, 2020). Simulations indicate that DLAs have the highest covering fractions within approximately 20 kpc of the galaxy haloes (e.g. van de Voort et al., 2019; Nelson et al., 2020). Sub-Damped Lyman-α absorbers (sub-DLAs; $N_{\mathrm{H\,I}} = 10^{19.0-20.3}\,\mathrm{cm}^{-2}$; Péroux et al., 2003) trace gas consisting of a mixture of ionized and neutral gas further away from the core in galaxy haloes. Lyman-Limit-Systems (LLS; $N_{\mathrm{H\,I}} = 10^{17.2-19.0}\,\mathrm{cm}^{-2}$; Tytler, 1982) trace mostly ionized systems with low neutral fractions of $\mathrm{H}^0/\mathrm{H} = 10^{-3}$ (e.g. Fumagalli et al., 2016) and trace gas in large extended regions surrounding galaxies. Due to this higher volume filling factor, the detections of lower column density systems are more numerous compared to the high column density DLAs. Below these column densities, the low-density gas of the IGM is traced by the Lyman-α forest (see e.g. Weinberg et al., 2003).

In recent years, powerful instruments such as the Multi Unit Spectroscopic Explorer (MUSE) on the Very Large Telescope (used to trace the ionized gas phase in galaxies; Bacon et al., 2010) and the Atacama Large Millimeter/Submillimeter Array (ALMA; exploited for molecular gas observations; Wootten & Thompson, 2009) have revolutionized the field. These instruments offer Integral Field Spectroscopy (IFS) capabilities, thus providing a spectrum for each pixel of an image. The combination of spectroscopic absorption-line and IFU observations have proven to be a powerful tool to connect the properties of galaxies traced in emission with the CGM traced in absorption. Surveys such as MUSEQuBES (Muzahid et al., 2020, 2021), MEGAFLOW (e.g. Schroetter et al., 2016) and MUSE-ALMA Haloes (e.g. Péroux et al., 2022) have connected the CGM traced in absorption with the ISM traced in emission. Among other results, these studies find that H I-absorption-selected galaxies are preferentially found in groups (Hamanowicz et al., 2020), a higher metal enrichment of the CGM along the minor axis compared to the major axis of galaxies due to outflowing gas (Wendt et al., 2021) and traced inflowing, outflowing and co-rotating gas in the CGM (e.g. Rahmani et al., 2018a,b; Schroetter et al., 2019; Szakacs et al., 2021). Thus, it is observationally confirmed that the CGM plays an important role in the cycling of baryons through dynamical processes such as accreting, outflowing, and recycling gas.

1.3 The Cosmic Baryon Cycle

The amount of mass in baryons in the Universe (4 per cent) is well established from measurements of anisotropies in the Cosmic Microwave Background (CMB) (Planck Collaboration et al., 2020) and primordial nucleosynthesis (Cooke et al., 2018). This baryonic matter cycles through various phases across cosmic time. From collapsing within dark matter haloes, to cooling down and forming molecular clouds and subsequently stars that create metals, to stars exploding into supernovae and enriching their surroundings with their produced metals. The processes of this cosmic baryon cycle heavily impact the galaxy populations across cosmic time, thus exploring how star formation, cold gas, and metals evolve on a global scale is key to deciphering the many questions still open concerning galaxy evolution.

1.3.1 An Evolving Star Formation Rate Density across Cosmic Time

Measurements of galaxies in the ultraviolet (UV) and infrared (IR) regime have revealed an evolving star formation rate density (SFRD) across cosmic time (see Fig. 1.4, Madau & Dickinson, 2014). Beginning from the early Universe, the SFRD constantly rises up until cosmic noon ($z \sim 2$) where it peaks and subsequently drops towards $z = 0$.

These measurements are based on converting UV and IR luminosity densities to star formation rate densities using conversion factors, leading to various uncertainties. Firstly,

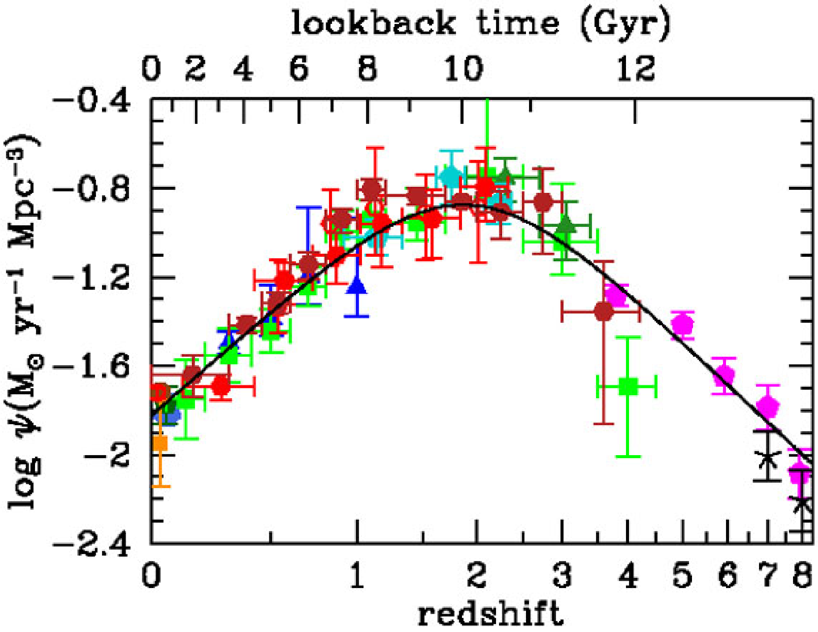

Figure 1.4: A compilation of measurements in ultraviolet (UV) and infrared (IR) wavelengths for the star formation rate density (SFRD) across cosmic time (Madau & Dickinson, 2014). UV measurements are in green, pink, black, and blue. IR measurements are in red and orange. Beginning from the early Universe, the SFRD rises until $z = 2$ where it peaks and drops towards $z = 0$.

the UV luminosity itself depends highly on the extinction correction due to dust attenuation. Next, both the UV and IR conversion factors are dependent on the stellar initial mass function (IMF) used, which in itself is uncertain. Further, the UV conversion factor is dependent on the recent star formation and metal enrichment history. Finally, the IR conversion factor additionally assumes that all IR luminosity is due to recent star formation, however, old stars and active galactic nuclei can also contribute to the IR luminosity (Madau & Dickinson, 2014).

Even though there are uncertainties that require further exploration, it is well established that the cold gas phase is crucial for star formation itself. Within the cold neutral atomic gas phase molecular clouds form, subsequently leading to the formation of stars within these clouds. (e.g. Schruba et al., 2011). Exploring the evolution of both the neutral atomic and molecular gas phase is critical for our understanding of the evolving star formation rate density in the Universe.

1.3.2 The Evolution of Neutral Atomic Gas across Cosmic Time

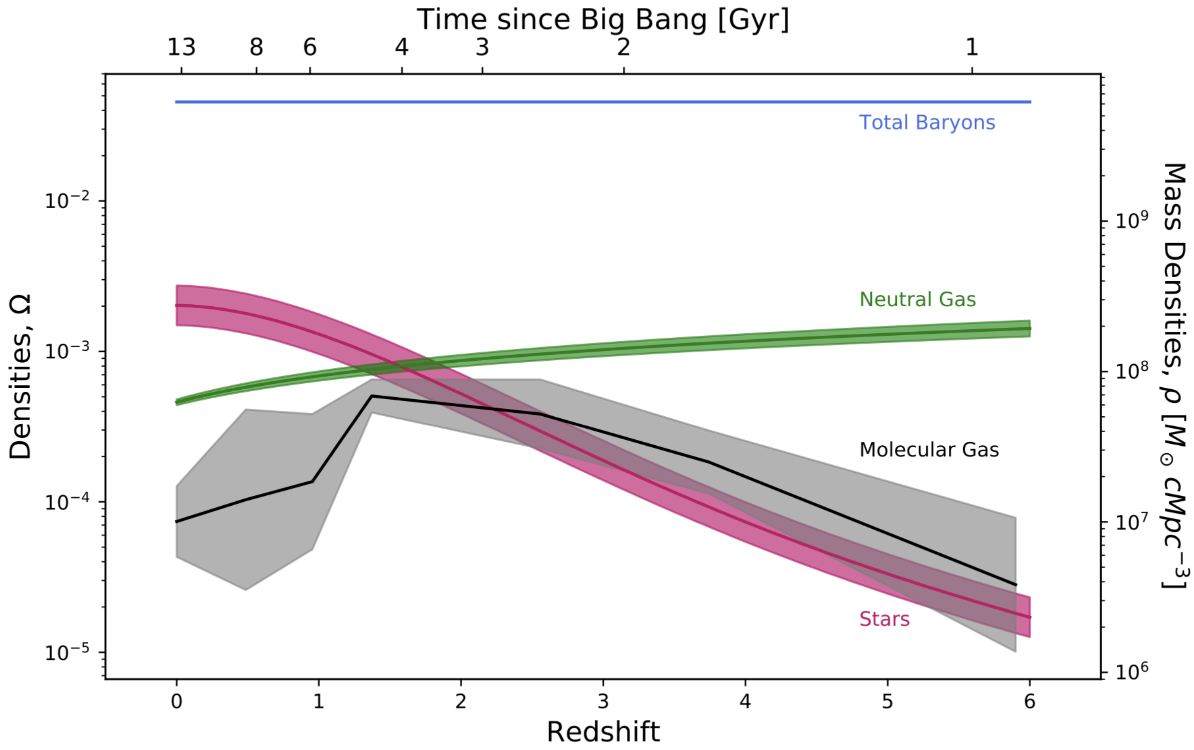

Figure 1.5: The evolution of neutral gas (green), molecular gas (gray), and stellar (red) mass densities in the Universe across cosmic time based on a compilation of observations (Péroux & Howk, 2020). The neutral gas phase displays a steady decline in the mass density towards $z = 0$. The molecular mass density rises with declining redshift up until $z \sim 2 - 3$ and subsequently drops rapidly towards $z = 0$. The evolution of the molecular gas mass density mimics the evolution of the star formation rate density across cosmic time in Fig. 1.4.

The neutral gas phase in the Universe is dominated by atomic Hydrogen (H I). It is the gas phase second closest to the star formation process (with the molecular gas phase being the closest), thus understanding its evolution is important to study the mechanisms of star formation within galaxies. Depending on redshift, mainly two observational techniques have been used to study the evolution of H I.

Below $z \leq 1.0$ the H I 21-cm transition is observable in direct emission or by stacking observations with current and future radio telescopes [e.g. VLA (Heeschen, 1981), MeerKAT (Jonas & MeerKAT Team, 2016), ASKAP (Hotan et al., 2021)]. Numerous efforts in recent years have contributed to deriving the H I mass density by measuring the H I 21-cm line of galaxies (e.g. Delhaize et al., 2013; Rhee et al., 2016; Jones et al., 2018b; Bera et al., 2019; Chowdhury et al., 2020, 2021) within this redshift regime. At redshifts above $z > 1$, the H I 21-cm transition seen in emission is below the sensitivity limit of current telescopes. Thus, alternative approaches have been used. The most commonly used approach is the search of Lyman-α absorption within quasar spectra (see Section 1.2.1 and see e.g. Zafar et al., 2013; Rao et al., 2017; Noterdaeme et al., 2012). However, above $z \geq 5$ the Lyman-α forest becomes opaque due to the high H I neutral fractions. Thus, at these redshifts, the continuum of the background quasar is mostly absorbed, making the search for Lyman-α absorption difficult. First efforts to overcome this are offered by alternative tracers of neutral gas. E.g. Becker et al. (2019) have successfully used OI absorption in background quasars as a tracer up to $z = 6.5$ and Heintz et al. (2021, 2022) have explored using the bright [C II] 158μm emission to trace the H I content of galaxies up to $z = 7$.

There is little evolution in the density of neutral gas ($\Omega_{\rm neutral\ gas}$) between the local universe and $z = 5$ (see the green line in Fig. 1.5 based on a comprehensive compilation of observations by Péroux & Howk, 2020). As the redshift decreases, so does $\Omega_{\rm neutral\ gas}$ following a two-parameter power-law (Péroux & Howk, 2020):

$$\Omega_{\rm neutral\ gas}(z) = [(4.6 \pm 0.2) \times 10^{-4}](1+z)^{0.57\pm0.04} \ . \tag{1.1}$$

This evolution of the neutral gas mass density compared to the evolution of the star formation history (see Section 1.3.1) indicates that there is no direct connection between the overall abundance of neutral gas and the star formation efficiency in the Universe. Clearly, exploring a gas phase closer to star formation is needed to better understand what causes the sharp decline of the star formation rate in the Universe towards $z = 2$.

1.3.3 The Evolution of Molecular Gas across Cosmic Time

As gas within the neutral atomic gas phase cools down it collapses into cold ($T \sim 10-40$K), dense ($n_{\rm H_2} \sim 10^2 - 10^5$ cm^{-3}) massive ($M \sim 10^4 - 10^{6.5} M_\odot$) molecular clouds (Giant Molecular Clouds (GMCs); e.g. McKee & Ostriker, 2007; Kennicutt & Evans, 2012; Tacconi et al., 2020). Molecules within these clouds are shielded from the interstellar UV radiation through a layer of surrounding atomic gas, dust and by self-shielding of molecular hydrogen itself (e.g. McKee & Ostriker, 2007; Krumholz et al., 2008). Under these conditions the Jeans criterion is reached within the dense cores of GMCs, leading to the formation of stars. Thus, the molecular phase is the gas phase closest to star formation and understanding its evolution across cosmic time is key to understanding the star formation history of the Universe itself.

Molecular gas is most abundantly found in the form of molecular Hydrogen (H_2). However, at the cool temperatures within molecular clouds the first accessible rotational level ($T_{\rm ex} \sim 500$ K) cannot be excited. Thus, the second most abundant molecule, CO, is typically used as a tracer for molecular gas. CO offers the advantage that it is both bright and has low excitation temperatures of $T_{\rm ex} \sim 5$ K (e.g. Bolatto et al., 2013). In practice, the integrated CO(1-0) line luminosity is converted to a molecular gas mass using various (metallicity-dependent) conversion factors from literature (e.g. Leroy et al., 2011; Genzel et al., 2012; Bolatto et al., 2013; Teng et al., 2022).

Particularly the instrument ALMA has revolutionized molecular gas studies. ALMA is an interferometry instrument with 66 antennas and operates in the millimeter/sub-millimeter regime. With its IFS capabilities, it provides spatially resolved spectra at spatial resolutions of down to 20 milliarcseconds and spectral resolutions of up to $R = \lambda/\Delta\lambda = 30$ 000 000. These capabilities have led to resolving the molecular gas in local galaxies down to scales of molecular clouds with programs such as the Physics at High Angular resolution in Nearby GalaxieS (PHANGS) ALMA survey (e.g. Leroy et al., 2021; Sun et al., 2022) and the ALMA Comprehensive High-resolution Extragalactic Molecular Inventory (ALCHEMI; Martín et al., 2021).

Deep surveys of molecular gas in galaxies have given first estimates of the molecular gas density ($\Omega_{\text{molecular gas}}$) across cosmic time (e.g. Decarli et al., 2016, 2019; Walter et al., 2016; Riechers et al., 2019; Saintonge et al., 2017). A compilation of such surveys (red line in Fig. 1.5, Péroux & Howk, 2020) shows a strong evolution in the mass density across cosmic time. $\Omega_{\text{molecular gas}}$ rises until cosmic noon (z $\sim$ 2-3) and subsequently shows a steep drop towards $z = 0$. This evolution of $\Omega_{\text{molecular gas}}$ is mirrored by the evolution of the star formation rate density described in 1.3.1, indicating the strong connection between the molecular gas mass density and star formation across cosmic time.

As displayed by the large errors in Fig. 1.5, the uncertainties on the molecular gas densities across cosmic time are still large and more observations are the key to improving the constraints. One survey tackling this question is the ALMACAL survey (Zwaan et al., 2022). By exploiting ALMA calibration data ($\sim$ 2500 hours integration time) it aims to blindly survey the Universe for molecular gas emission. A first pilot survey with a subset of the data has given first constraints on the molecular gas density at low redshifts (Hamanowicz et al., 2022) and will be extended to the full sample in the future.

1.3.4 The Evolution of Metallicity across Cosmic Time

Absorption-line studies (see Section 1.2.1) are a key method to trace the evolution of metallicity in the Universe. Absorption-based metallicity estimates offer the advantage of not requiring a local excitation source. Hence, they are less sensitive to density and temperature compared to emission-based estimates and probe both low- and high-metallicity excitation regions (Péroux & Howk, 2020).

The total expected amount of mass in metals produced in the Universe can be derived by scaling the stellar mass density with estimated yields of long-lived stars and stellar remnants (Péroux & Howk, 2020). The blue line in Fig. 1.6 displays the rising amount of expected metals with lower redshift ($\sim$ two orders of magnitude between $z = 5$ and 0), as the stellar mass density increases (see Fig. 1.5). Towards the local Universe, the curve of expected metals flattens due to the decrease of star formation.

A compilation of DLA measurements by Péroux & Howk (2020) displays a high contribution of metals within the neutral gas phase across cosmic time (green dots in Fig.

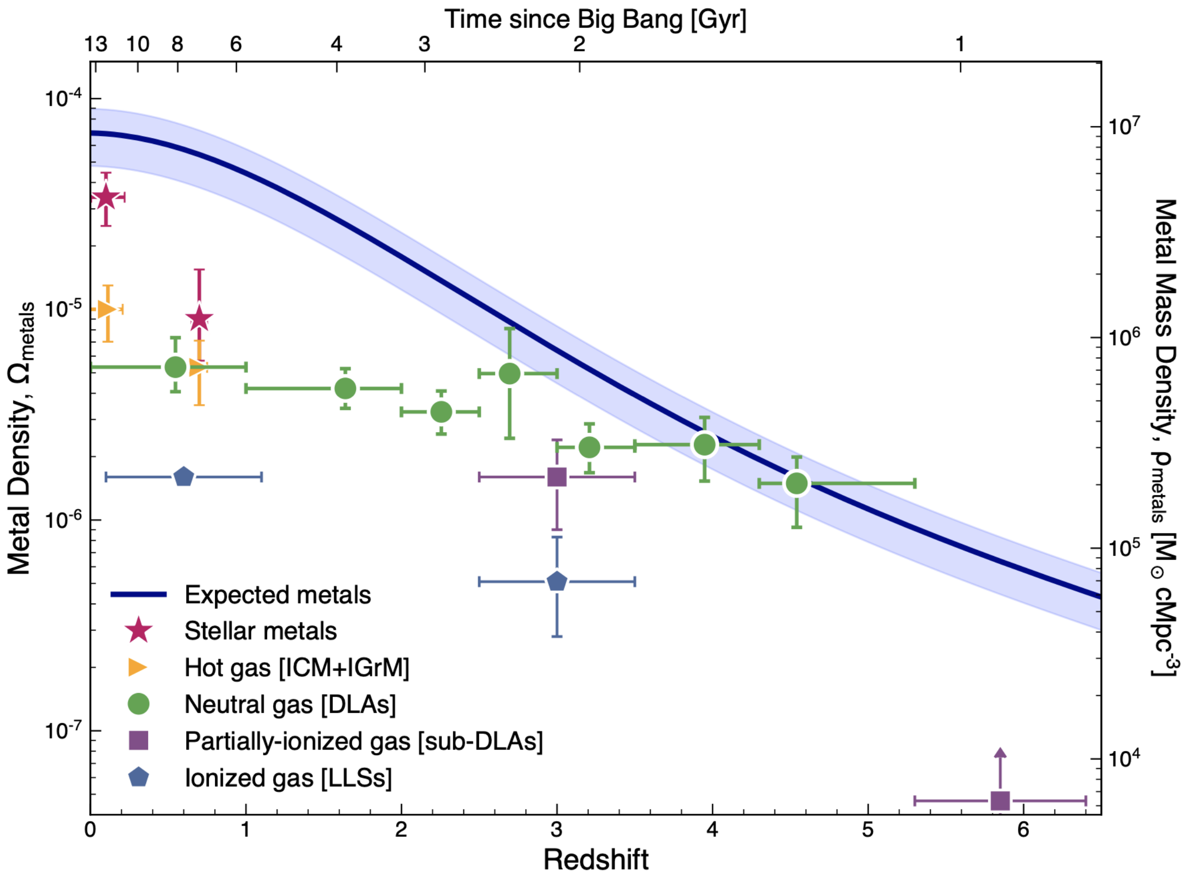

Figure 1.6: A compilation of the expected amount of metals in the Universe across cosmic time and the phase in which those are found by the currently most robust measurements by Péroux & Howk (2020). At redshifts above $z \geq 2.5$ almost all metals are found within the neutral gas phase. Below those redshifts contributions by stars and hot gas become increasingly important.

1.6). Current measurements indicate that above $z \geq 2.5$ the majority of metals are found within the neutral gas phase. At redshifts below $z \leq 2.5$ the contribution by the neutral gas phase decreases, while an increasing amount of metals is found within stars and hot gas (Péroux & Howk, 2020). The currently most robust measurements of lower density absorbers (Becker et al., 2011; Fumagalli et al., 2016; Lehner et al., 2019) indicate a smaller amount of metals within the (partially-)ionized gas phase at $z \sim 6$, $z \sim 3$ and $z \leq 1$.

Fig. 1.6 displays a clear lack of robust measurements to fully constrain the census of metals across cosmic time. However, upcoming large-scale surveys by e.g. VISTA/4MOST (de Jong et al., 2019), DESI (DESI Collaboration et al., 2016) and WHT/WEAVE (Dalton et al., 2012) will fill these gaps with increasing efficiency. However, given the massive amount of data, such surveys require specialized analysis methodologies. Techniques such as machine learning need to be explored in astronomical contexts to deal with the advent of big data astronomy.

1.4 Machine Learning

Machine learning has proven to be a useful tool for a wide range of applications. Among others, these include detecting features in images (e.g. Krizhevsky et al., 2012), forecasting time series (e.g. Lara-Benítez et al., 2021) and even for creating systems that play the traditional Chinese game Go so well that it beat world champions (Silver et al., 2016). A commonly used definition of the term Machine Learning is given by Tom Mitchell: ”A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E.” (Mitchell, 1997). Thus, machine learning includes methods that take advantage of data to improve the performance on set tasks. These methods can be categorized within distinct types (see e.g. Murphy, 2013). Supervised learning leverages labeled data. Meaning that the machine is given an input and an additional output that it is meant to reproduce. Unsupervised learning uses unlabeled data to e.g. cluster data into specific regions on a latent space. Semi-supervised learning provides a combined method of both supervised and unsupervised learning in case only a small amount of labeled data is available. Finally, reinforcement learning trains a machine on how to respond to given signals by providing rewards or penalties depending on the performed action. While there are numerous flavours of machine learning methods, the next section will focus on the description of neural networks in the supervised learning paradigm, as this is the relevant approach applied in this work.

1.4.1 Neural Networks

Neural networks are a machine learning approach that is inspired by the principles of neurobiology. An in-depth review of the basic principles of neural networks is provided in Gurney (1997) and the next two paragraphs provide a short summary based on this book. The human brain consists of neurons that are connected with each other and communicate through electrical signals. Synapses, electrochemical junctions located on the branches (dendrites) of a neuron’s cell body (soma), allow for these connections between the neurons. Each neuron is embedded in the overarching network via these connections to other neurons to transmit the aforementioned signals. If incoming signals exceed a certain threshold, the neuron generates a voltage impulse that is transmitted to other neurons through fibres called axons. The incoming signals can have both an inhibitory effect and an excitatory effect, with the latter promoting and the former preventing the firing of a neuron. Thus, each neuron’s processing ability is based on the type and strength of synaptic connections with other neurons.

Neural networks take these processes and recreate them artificially as they contain nodes

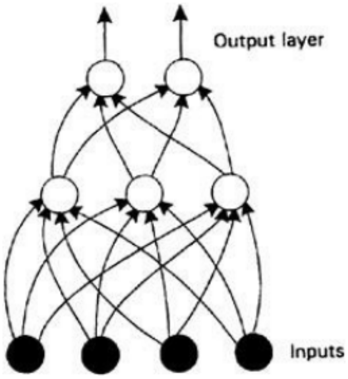

Figure 1.7: Sketch of a simple feed-forward neural network containing an input layer, a layer containing three neurons in the middle, and an output layer with two neurons. Each subsequent layer has connections to the neurons of the following layer. (Gurney, 1997)

that are rudimental emulations of neurons. The synapses are represented in terms of single numbers or weights. Inputs are multiplied by these weights and are sent to the analogue of the cell body, which sums up the weighted signals and inputs the values to an activation function (e.g. a threshold function returning 0 or 1 depending on the summed signals). For supervised learning methods, these weights are optimized to minimize a metric appropriate for a given task (e.g. the mean squared error for regression tasks or binary cross-entropy for classification tasks) by the backpropagation algorithm. In particular, this algorithm exploits the chain rule for derivatives combined with stochastic gradient descent. The gradients of the used metric (i.e. error function) are calculated with respect to the weights of the neural network starting from the final (i.e. output) layer. Going backwards in the network structure, the gradient of each layer is calculated including partial computations of the gradient from previous layers. The weights of the neural network are then updated based on these gradients. The iteration of this process leads to the optimization of the weights of the neural network minimizing the error function. In the most simple form, a neural network consists of stacked layers of neurons having connections to subsequent layers (see Fig. 1.7).

1.4.2 Convolutional Neural Networks

Convolutional neural networks are based on the organization of the visual cortex in animals. The basic idea is that the visual cortex consists of simple and complex cells (Hubel & Wiesel, 1959). Simple cells are good at detecting edges of particular orientations. Complex cells are similar to simple cells in the sense that they are also good at detecting edges of particular orientations, however, they are able to do this with spatial invariance. Hubel & Wiesel (1962) proposed that complex cells achieve this by summing the output of several simple cells. Inspired by this idea, Fukushima (1980) created the mathematical neocognitron model to recognize patterns in images invariant of their position. Based on these ideas, the first modern convolutional Neural network was developed to recognize hand-written digits by Lecun et al. (1998).

Specifically, CNNs consist of convolution, pooling, and fully connected layers (see e.g. the reviews on CNNs by LeCun et al., 2015; Yamashita et al., 2018). Convolutional layers extract features from input signals (e.g. images or spectra) by calculating element-wise products of given kernels and the input. Subsequently, a non-linear activation function is applied to the outputs of the linear convolution operation to allow for outputs that vary non-linearly with inputs. Pooling layers reduce the dimensionality of the feature maps calculated by the convolutional layers, allowing for a shift-invariance of the feature detection by e.g. calculating the maximum or average values in patches of the feature maps. Finally, fully connected layers map the extracted features to the desired output (e.g. classification or regression). The combination of these layers leads to a neural network that can extract desired features without being affected by small shifts and distortions of these features. This and other types of machine learning methods have been applied in astronomical contexts with increasing success.

1.4.3 Machine Learning as a Tool in Astronomy

Astronomical datasets are rapidly increasing in size, especially due to large-scale surveys such as SDSS (York et al., 2000), Gaia (Gaia Collaboration et al., 2016), the eROSITA all-sky survey (Predehl et al., 2021), and a large number of upcoming surveys, in particular VISTA/4MOST (de Jong et al., 2019), DESI (DESI Collaboration et al., 2016), WHT/WEAVE (Dalton et al., 2012), LSST (Ivezic et al., 2008) and SKA (Dewdney et al., 2009) surveys. This massive increase in data leads to a novel need for efficient and automated tools that can analyze or even infer physical knowledge from these large datasets.

Supervised learning methods, in the form of e.g. random forests (e.g. Ho, 1995) and convolutional neural networks (e.g. LeCun et al., 2015), have been used to identify causality in astronomical data (Bluck et al., 2020, 2022; Baker et al., 2022), to identify galaxy mergers (Rose et al., 2022), photometric quasar identification and redshift estimation (Li et al., 2022; He & Li, 2022), identifying Lyman-α and metal absorption lines in quasar spectra (Parks et al., 2018; Zhao et al., 2019; Wang et al., 2022; Xia et al., 2022), for photometric source detection, classification and parametrization (Shi et al., 2022) and many more applications. Unsupervised learning methods applied through e.g. autoencoders (Liou et al., 2014), self-organizing maps (Wehrens & Buydens, 2007) and others have been useful to e.g. classify galaxy morphology (Galvin et al., 2020; Fielding et al., 2022), denoising astronomical data (Shen et al., 2017; Gheller & Vazza, 2022) and for photometric membership assignment in stellar clusters (Krone-Martins & Moitinho, 2014).

These studies provide evidence that machine learning will play an important role in the era of big data astronomy. While still in their infancy and the viability of these approaches still needs to be explored for their usage in many applications, these methods often provide similar to better accuracy than traditional methods and are additionally computationally efficient (i.e. faster) once the algorithms have been trained.

1.5 The Goal of this Thesis

The overarching goal of this thesis is to study the cool and cold gas phases within and surrounding galaxies. These gas phases encompass the molecular gas and its precursors in the form of neutral and low-ionized atomic gas. As molecular gas is the closest gas phase to star formation, deciphering the properties and dynamics of this phase and its progenitors is of key importance to further our understanding of star formation within galaxies and subsequently the evolution of galaxies in general. The thesis can be divided into distinct questions all serving this overarching goal:

- How does the kinematics of the molecular gas phase compare to the ionized gas phase in galaxies? (Chapter 2)
- Does the circumgalactic medium of galaxies detected in absorption trace inflowing, outflowing, or co-rotating gas? (Chapter 2)
- What are the molecular gas properties of H I-absorption-selected galaxies and how do they differ compared to emission-selected samples? (Chapter 2)
- How does the distribution of molecular gas column densities evolve across cosmic time? (Chapter 3)
- Which column densities contribute the most to the overall molecular gas mass density across cosmic time? (Chapter 3)
- How does the evolution of molecular hydrogen column densities compare to the evolution of neutral hydrogen column densities? (Chapter 3)
- What is the shape of the molecular gas column density distributions of individual main-sequence star-forming galaxies? (Chapter 3)
- Is machine learning a viable alternative to traditional methods for the detection and localization of metal absorbers in background quasar spectra of upcoming large-scale surveys? (Chapter 4)

We address these questions on the scales of individual galaxies, up to large statistical samples. In Chapter 2 we explore the kinematics and properties of the neutral, molecular, and ionized gas phase within the ISM and CGM of H I-absorption selected galaxies by exploiting IFU observations by VLT/MUSE, ALMA and spectroscopic observations by HST/COS. In Chapter 3, we study the evolution of the distribution of H_2 column densities using an approach that bridges observations (ALMA and SDSS) and (cosmological magneto-)hydrodynamical simulations (TNG50 and TNG100 by the Illustris Project and an isolated dwarf galaxy simulation by the GRIFFIN project). Finally, in Chapter 4, we create synthetic normalized quasar spectra mimicking upcoming VISTA/4MOST observations with injected Mg II (λ2796, λ2803) absorbers derived from TNG50 to train a convolutional neural network for the detection and localization of Mg II absorption-line systems in millions of spectra of future data sets.

Chapter 2

MUSE-ALMA Haloes VI: Coupling Atomic, Ionized & Molecular Gas Kinematics of Galaxies

The content of this chapter is based on the published article Szakacs et al., (2021), MNRAS, 505, 4746.

2.1 Abstract

We present results of MUSE-ALMA Haloes, an ongoing study of the Circumgalactic Medium (CGM) of galaxies ($z \leq 1.4$). Using multi-phase observations we probe the neutral, ionized, and molecular gas in a sub-sample containing six absorbers and nine associated galaxies in the redshift range $z \sim 0.3-0.75$. Here, we give an in-depth analysis of the newly CO-detected galaxy Q2131-G1 ($z = 0.42974$), while providing stringent mass and depletion time limits for the non-detected galaxies. Q2131-G1 is associated with an absorber with column densities of $\log(N_{\mathrm{H\,I}}/\mathrm{cm}^{-2}) \sim 19.5$ and $\log(N_{\mathrm{H_2}}/\mathrm{cm}^{-2}) \sim 16.5$, has a star formation rate of $\mathrm{SFR} = 2.00 \pm 0.20\ \mathrm{M_{\odot}yr^{-1}}$, a dark matter fraction of $f_{\mathrm{DM}}(r_{1/2}) = 0.24 - 0.54$ and a molecular gas mass of $M_{\mathrm{mol}} = 3.52^{+3.95}_{-0.31} \times 10^{9}\ \mathrm{M_{\odot}}$ resulting in a depletion time of $\tau_{\mathrm{dep}} < 4.15$ Gyr. Kinematic modelling of both the CO (3–2) and [O III] $\lambda5008$ emission lines of Q2131-G1 shows that the molecular and ionized gas phases are well aligned directionally and that the maximum rotation velocities closely match. These two gas phases within the disk are strongly coupled. The metallicity, kinematics, and orientation of the atomic and molecular gas traced by a two-component absorption feature are consistent with being part of the extended rotating disk with a well-separated additional component associated with infalling gas. Compared to emission-selected samples, we find that H I-selected galaxies have high molecular gas masses given their low star formation rate. We consequently derive high depletion times for these objects.

2.2 Introduction

One of the most puzzling questions in galaxy evolution is how galaxies sustain their star formation. Due to the short depletion timescales that have been observed, it is evident that galaxies have to accrete gas from an external source in order to maintain their continuity on the main sequence (e.g. Scoville et al., 2017). The inflowing gas is accreted from the intergalactic medium (IGM). While it is a challenging task to observe the accretion process due to the low density of the extragalactic gas, a number of inflows have been observed over the last few years (e.g. Rubin et al., 2012; Martin et al., 2012; Turner et al., 2017; Zabl et al., 2019). Metal enriched gas is also expelled from galaxies due to AGN feedback (Shull et al., 2014) or stellar feedback (e.g. Ginolfi et al., 2020). A fraction of the expelled gas is returned through galactic fountains (Fraternali, 2017; Bish et al., 2019), where cooled down gas rains onto the galactic disk, while some other part is returned to the IGM through galactic winds driven by AGN and stellar feedback processes.

The inflowing and outflowing gas interacts in a zone called the circumgalactic medium (CGM), which is loosely defined as the gas surrounding galaxies outside of the disk or ISM, but within the virial radius (Tumlinson et al., 2017). While it can be a challenging endeavour to observe the CGM directly, due to the low surface brightness of the gas (e.g. Frank et al., 2012; Corlies et al., 2020b; Augustin et al., 2019), observations and simulations indicate that the CGM is a multi-phase medium. The hot phase of the CGM has been observed through X-ray observations (e.g. Anderson & Bregman, 2010; Anderson et al., 2013; Bregman et al., 2018; Nicastro et al., 2018) and Ly-α emission (e.g. Cantalupo et al., 2014; Wisotzki et al., 2016, 2018; Umehata et al., 2019). The cooler gas in the CGM can be probed by studying absorption lines in quasar (QSO) spectra, which offer the advantage of the sensitivity being independent of redshift (e.g. Tripp et al., 1998). This cooler low-density gas has been detected through the absorption lines of various metal species and Hydrogen (e.g. Steidel et al., 2010; Rudie et al., 2012; Werk et al., 2013; Turner et al., 2014). Hydrodynamic simulations strengthen the picture of a multi-phase CGM, by finding a mixture of cooler ($T \sim 10^4$ K) and hotter ($T \sim 10^{5.5} - 10^6$ K) gas within the virial radius of simulated galaxies (e.g. Stinson et al., 2012; Suresh et al., 2017; Nelson et al., 2020).

An important aspect to understand how galaxies sustain their star formation is to connect the CGM gas probed by absorption with the galaxies associated with the absorbers. Narrow-band imaging and long-slit spectroscopic studies have searched for nebular emissions from H I-selected galaxies (e.g. Kulkarni et al., 2000, 2001) and have, in part, been successful in the past (e.g. Chen et al., 2005; Fynbo et al., 2010). Further, integral field spectroscopy (IFS) combined with long-slit spectroscopy follow-ups have made it possible to not only associate galaxies with strong H I-absorbers (e.g. Bouché et al., 2007; Péroux et al., 2011a,c, 2017; Rudie et al., 2017), but to also study the star formation rate, metallicity of the emission line gas and kinematics of the ionized gas (e.g. Bouché et al., 2012; Péroux et al., 2017; Rahmani et al., 2018a,b; Hamanowicz et al., 2020). The findings, among others, include a correlation between the SFR of the associated galaxy and the equivalent width of the absorption, indicating a physical connection between starbursts and gas seen in absorption (Bouché et al., 2007). Péroux et al. (2011a) find that in the majority of the cases the metallicity of the absorption is lower than that of the associated galaxy. The number of studies associating absorption features found in the spectra of quasars with the physical properties of absorber host candidates is low. An additional issue remains: associating galaxies with absorbers that are in complex group environments as studied in this and a previous MUSE-ALMA Haloes publication (Hamanowicz et al., 2020). The authors suggest that galaxies found in these environments would benefit from associating the kinematics of the galaxies with the absorber in order to distinguish which galaxies/environments the absorption is tracing (e.g. see Rahmani et al., 2018a). Therefore, obtaining more observations of absorber - absorber host systems play a key part in furthering the understanding of the medium surrounding galaxies.

Searches at the radio / sub-mm wavelengths with instruments like the Atacama Large Millimeter Array (ALMA) have enabled the community to study the mass, depletion time, and kinematics of the molecular gas in galaxies associated with absorbers (e.g. Neeleman et al., 2016, 2018; Klitsch et al., 2018, 2019b; Augustin et al., 2018; Møller et al., 2017; Kanekar et al., 2018, 2020; Péroux et al., 2019; Freundlich et al., 2021). One of the findings that these H I-selected galaxies have in common is that the molecular gas masses of these galaxies are high for their given SFR, leading to depletion times that are up to multiple factors larger than the averages found in emission-selected galaxies. Further observations and constructing statistically significant samples, like the ones obtained in the MUSE-ALMA Haloes project, are needed in order to study a possible correlation.

Obtaining spatially-resolved multi-phase data of galaxies has furthered our understanding of how the ionized and molecular gas phases relate to each other. Kinematic studies have revealed that the two phases mostly align well spatially (e.g. Übler et al., 2018; Møller et al., 2017; Klitsch et al., 2018; Loiacono et al., 2019; Péroux et al., 2019; Molina et al., 2019, 2020). Further kinematic studies by the EDGE-CALIFA survey (Levy et al., 2018) have shown that 75% of the galaxies in their sample have higher maximum rotational velocities for the molecular gas while the remaining 25% have similar maximum rotational velocities to the ionized gas. Péroux et al. (2019) on the other hand did indeed find a case where the rotational velocity of the molecular gas was significantly lower than for the ionized gas in a galaxy associated with a strong H I-absorber. The number of galaxies observed in both the molecular and ionized gas phase is still low and studies of these gas phases are a key point in furthering our understanding of gas flows within and surrounding galaxies.

Another aspect of using IFS-based multi-wavelength observations is that these data make it possible to estimate the dark matter fractions in the inner parts of galaxies. A widely accepted notion is that dark matter dominates the outskirts of galaxies, however, the distribution of matter in the central parts of galaxies is still debated. Studies like the DiskMass survey (Martinsson et al., 2013) have observed 30 spiral galaxies at the current epoch and found the central dark matter fractions to be mostly in the range of 0.5-0.9. Studies of higher redshift galaxies find lower central dark matter fractions in both observations and simulations (e.g. Übler et al., 2018, 2020; Genzel et al., 2017, 2020). Price et al. (2020) report a decrease of the dark matter fraction toward higher redshifts, attributed to various intertwined effects of galaxy mass–size growth, gas fraction, and halo growth and evolution. Therefore obtaining further samples of central dark matter fractions is an important aspect of understanding the reasons for the differences in the central dark matter fractions over different epochs.

The studied MUSE-ALMA Haloes sub-sample includes six absorbers and nine associated galaxies in the redshift range $z \sim 0.3-0.75$. In this publication, we present the results from new ALMA observations of the fields Q2131-1207, Q1232-0224, Q0152-2001, Q1211-1030, Q1130-1449 each of which contain a strong H I-absorber at $z \sim 0.4$ and in the case of Q1232-0224 an additional one at $z \sim 0.75$. While we analyze and provide information on all fields, the focus of this publication lies on the CO-detected galaxy Q2131-G1 in the field Q2131-1207 [first reported in Bergeron (1986) and further analyzed in Guillemin & Bergeron (1997) and Kacprzak et al. (2015)].

The paper is organized as follows: Section 2.3 presents the observational set-up and data reduction and imaging process. Section 2.4 describes the molecular properties of the galaxies associated with the strong H I-absorbers while describing both the physical and morpho-kinematical properties and providing limits for non-detections. In Section 2.5 we discuss our findings and put them into context with previous observations. Finally, Section 2.6 gives a summary of the findings. Throughout this paper we adopt an $H_0 = 70\text{km s}^{-1}\text{Mpc}^{-1}, \Omega_\text{M} = 0.3, \text{and } \Omega_\Lambda = 0.7$ cosmology.

2.3 Observations

We follow a multi-wavelength approach to study the gas and associated galaxies in this study, combining VLT/MUSE, HST, and ALMA observations. The observations and corresponding data processing/imaging are presented in this section.

2.3.1 Optical Campaign

VLT/MUSE Observations

In this work, we study five fields containing quasar absorbers (Q2131-1207, Q1232-0224, Q0152-2001, Q1211+1030, and Q1130-1449). These fields are a subset of the full MUSE-ALMA Haloes sample which have ALMA follow-up observations targeting redshifts of $z \sim 0.4$ and $z \sim 0.75$. That sample has been observed using VLT/MUSE in period 96 under programme ESO 96.A-0303 (PI: C. Peroux). All fields were observed in nominal mode (4800-9400 Å) under good seeing conditions (<0.85 arcsec). The first four fields were observed for 1-2 hours per target, while Q1130-1449 was observed significantly deeper (12×1200 s). The observations and data reduction method for the 5 quasar fields is described in depth in Péroux et al. (2019) and Hamanowicz et al. (2020). In short, the ESO MUSE reduction pipeline v2.2 (Weilbacher et al., 2016) was used. Bias, flat, and wavelength calibration was applied in addition to line spread functions and illumination correction frames to each individual exposure. These astrometry solutions and the correction for geometry and flux calibrations were then applied. Each of the individual exposures was combined including field rotation. Instead of the pipeline sky subtraction method, the sky emission lines were removed using a Principal Component Analysis algorithm (Husemann et al., 2016). Additionally, the MUSE observations for the fields have been discussed in depth in the following publications: Q0152-020 (Rahmani et al., 2017; Rahmani et al., 2018c; Hamanowicz et al., 2020); Q1130-1449 (Péroux et al., 2019; Hamanowicz et al., 2020); Q2131-1207 (Péroux et al., 2017; Hamanowicz et al., 2020); Q1232-0224, Q1211-1030 (Hamanowicz et al., 2020).

HST Observations

We select fields that show strong H I-absorption column densities in quasar spectra. The column densities are based on literature and were derived using data from the Faint Object Spectrograph (FOS) and Cosmic Origins Spectrograph (COS) on HST [see Boissé et al. (1998) for details about the Q2131z039$_{\rm H\,I}$ absorber; Lane et al. (1998) for Q1130z031$_{\rm H\,I}$; Rao et al. (2006) for Q1232z075$_{\rm Mg\,II}$; Muzahid et al. (2016) for Q2131z043$_{\rm H\,I}$ and 1211z039$_{\rm H\,I}$; Rahmani et al. (2018b) for Q0152z038$_{\rm H\,I}$].

Readily available and reduced archival HST imaging is used for observations of the stellar continuum. The exposure times for the five fields range from 10 to 50 minutes. Observations of Q2131-1207 (PI: Maccheto, ID:5143), Q1232-0224 (PI: Bergeron, ID:5351), Q0152-2001 (PI: Steidel, ID:6557), and Q1211+1030 (PI: Bergeron, ID:5351) use the Wide Field Planetary Camera 2 (WPFC2) in the F702W filter. The observation of Q1130-1449 (PI: Bielby, ID: 14594) uses the Wide Field Camera 3 (WFC3) in filter IR-F140W.

Further archival HST data, obtained with the Cosmic Origins Spectrograph (COS) on HST, are used for studying the H_2 absorption lines of the absorber associated with Q2131-G1. Specifically, we use these spectra to study the position of the H_2 absorption line in velocity space. The H_2 absorption has been extensively studied in Muzahid et al. (2016). We use two observations with a wavelength range of 1140-1800Å , which consist of G130M (exposure time: 77 minutes) and G160M (exposure time: 120 minutes) FUV grating integrations at a medium resolution of $R \sim$ 20,000 (corresponding to a Full Width Half Maximum (FWHM) of $\sim$ 18 km s^{-1}). (PI: Churchill, ID: 13398). Due to the Lyman-limit break of the absorber ($z = 0.43$) there is no recorded QSO flux at wavelengths below 1310Å.

Additionally, we have an ongoing HST multi-band photometry program of 40 orbits (PI: Péroux, ID: 15939). This program will allow us to study the morphology and stellar masses of galaxies associated with H I-absorbers in the MUSE-ALMA Haloes survey.

This program will allow us to study the morphology and stellar masses of 200 $z < 1.2$ galaxies associated with H I and Mg II absorbers (including our current sample) in more detail at a later stage of the MUSE-ALMA Haloes project.

2.3.2 ALMA Observations

Observation Details

The fields Q2131-1207, Q1232-0224, Q0152-2001, and Q1211+1030 were observed with ALMA in Band 6 to cover the CO(3–2) lines of galaxies associated with absorbers found at $z \sim 0.4$ (programme 2017.1.00571.S, PI: C. Peroux). Given the field of view (FOV) of ALMA in band 6 we target a subset of galaxies previously observed with MUSE with impact parameters ranging from 8 to 82 kpc. All of the fields have one spectral window that was centred on the redshifted CO(3–2) frequency of 345.796 GHz with a high spectral resolution mode. This results in 3840 channels, each with a 1.129 MHz width. Additional three other spectral windows are also included for these observations in a low spectral resolution mode (31.250 MHz). The CO(3–2) line of one of the galaxies in the field Q1232-0224 ($z = 0.7566$) is expected to be in one of the low-resolution spectral windows. We also include the previously studied field Q1130-1449 in our analysis. Details concerning this observation can be found in Péroux et al. (2019).

A table with the quasar coordinates, observation dates, exposure times, angular resolution, used calibrators, precipitable water vapour (PWV) and antenna configurations for the different observed fields can be found in Table A.1 in the Appendix (Section A.1).

Data Reduction and Imaging

In this section, we describe the image processing of the fields Q2131-1207, Q1232-0224, Q0152-2001, and Q1211-1449 observed with ALMA. The fields are imaged and, when possible, self-calibrated using the Common Astronomy Software Applications package (CASA, McMullin et al., 2007) version 5.6.2-3.

As a starting point for all imaging and calibration, the pipeline-calibrated *uv*-datasets as delivered by ALMA-ARC are used. When multiple measurement sets (MS) are provided due to multiple observations, we combine them using the `concat` task. Using these combined measurement sets we reconstruct an initial continuum image of the field by using the task `tclean`. Depending on the synthesized beam size, we use different pixel sizes for the imaging (0.18" for Q0152-2001, 0.17" for Q1211-1030, 0.2" for Q2131-G1, and 0.22" for Q1232-0224). For all datasets we use `tclean` with a Briggs weighting scheme with the robust parameter set to 1.0, a standard gridder and a hogbom deconvolver.

In the case of Q0152-2001 and Q1211-1030, we follow up `tclean` with the task `uvcontsub` in order to subtract the central quasar in the field. As a final step, we use the continuum-subtracted uv-dataset and the task `tclean` with the same parameters as for the continuum images and a spectral binning of 50 km s^{-1}.

Both the quasars in Q2131-1207 and Q1232-0224 are bright at mm-wavelengths, allowing us to perform self-calibration. Therefore, after creating the initial model and continuum image mentioned above, we calculate the temporal gains using the task `gaincal` with gaintype `G` (which determines the gains for each polarization and spectral window) using a solution interval of 35 s for Q2131-1207 and 70 s for Q1232-0224. For both calibrations, we check that the solutions show a smooth evolution over time and that the solutions have an acceptable signal-to-noise ratio (SNR) > 10. Then we apply the solutions to the measurement sets using the task `applycal` in linear interpolation mode and create an updated sky model and continuum image using `tclean`. Following the phase calibration, we proceeded with a second round of amplitude calibration using gaintype G and a solution interval of 105 s for Q2131-1207 and 70 s for Q1232-0224. Following this, we create another updated sky model and continuum image using `tclean`. Then we follow up with the continuum subtraction using `uvcontsub` with order 3 for Q2131-1207 and 2 for Q1232-0224. We use the continuum-subtracted dataset to create a data cube using `tclean` with the same parameters as for the continuum images and a spectral binning of 50 km s^{-1}. As the final step, we produce a cube corrected for the primary beam using the `impbcor` task. The final RMS for the cubes where self-calibration was feasible is $\sim 1.5 \times 10^{-4}$ Jy. The cubes where no self-calibration was possible have an RMS $\sim 2.8 \times 10^{-4}$ Jy.

2.4 Molecular Gas Properties of the Galaxies Associated with the Absorbers

We target nine galaxies in the redshift range $z = 0.31 - 0.76$. Out of those nine galaxies we detect four: the previously detected galaxies Q1130-G2, Q1130-G4, and Q1130-G6 (presented in Péroux et al., 2019) and the newly CO-detected galaxy Q2131-G1. We provide an analysis of the physical and morpho-kinematical properties of Q2131-G1 in this section. Additionally, we provide stringent limits on the molecular gas content of undetected galaxies. All the calculated physical properties of the targeted galaxies can be found in Table 2.1 and the morpho-kinematical properties are listed in Table 2.2.

2.4.1 Properties of the CO-detected Galaxy (Q2131-G1, $z = 0.42974$)

In this section, we describe the physical and morpho-kinematical properties of the CO-detected galaxy Q2131-G1 and the galaxy-gas (absorber) connection.

Molecular Gas Mass and Depletion Time

We study the molecular gas properties of the CO-detected galaxy Q2131-G1. We create an integrated flux map using the CASA task `immoments` and set the threshold of pixels to be counted above $\sim 2\sigma$ of the created cube. This integrated flux map yields an observed CO(3–2) flux of $S_{\rm CO} = (0.36 \pm 0.02)$ Jy km s^{-1}. We derive the CO(1-0) luminosity by first calculating $L'_{\rm CO(3\text{–}2)}$ using $\rm S_{CO}$ and the prescription by Solomon et al. (1992). Then we use the $L'_{\rm CO(3\text{–}2)}$ to $L'_{\rm CO(1\text{-}0)}$ conversion factor from Fixsen et al. (1999): $L'_{\rm CO(3\text{–}2)}/L'_{\rm CO(1\text{-}0)} = 0.27$ and obtain a CO(1-0) luminosity of $L'_{\rm CO(1\text{-}0)} = (1.42 \pm 0.08) \times 10^{9}$K km s^{-1}pc^{2}. We choose the Milky Way spectral line energy distribution conversion factor due to the rather low redshift of the galaxy (z=0.42974). We note that absorption-selected systems may preferentially select interacting galaxies, which have more excited CO SLEDs than isolated

Table 2.1: **Physical properties of absorption-selected galaxies.**
Row 1 (red) - absorber: (1) reference name of the absorber used in this paper, (2) redshift of the absorber, (3) H I column density of the absorber, (4) metallicity of the absorber
Row 2 (green) - galaxy: (1) reference name of the galaxy used in this paper, (2) impact parameter in kpc and arcseconds, (3) star formation rate measured from the [O II] emission line (not dust corrected), (4/5) lower/upper metallicity 12+log(O/H) [not dust corrected, both metallicity branches derived by Hamanowicz et al. (2020) are displayed following (Kobulnicky et al., 1999)]. If only one metallicity branch is reported in literature, the upper column is left blank. (6) Observed frequency of the CO emission line
Row 3 (blue) - galaxy: (1) Redshift of the galaxy associated with the absorber, (2) observed CO flux density, (3) CO velocity width, (4) CO(1-0) Luminosity, (5) molecular gas mass, (6) depletion timescale of the galaxy.

Literature references: [a]) Hamanowicz et al. (2020) , [b]) Muzahid et al. (2016), [c]) Boissé et al. (1998), [d]) Rao et al. (2006), [e]) Rahmani et al. (2018b), [f]) Lane et al. (1998), [g]) Péroux et al. (2019).

[1] We note that Kanekar et al. (2009) reports a higher metallicity of $[\mathrm{Z/H}]_{\mathrm{abs}} = -0.90 \pm 0.11$ for this absorber.

Absorber ID	$z_{\mathbf{abs}}$ [a]	$\log(N_{\mathbf{H\,I,abs}})$ [cm^{-2}]	$[\mathbf{Fe/H}]_{\mathbf{abs}}$		
Galaxy	b [a] [kpc / ”]	SFR$_{[\mathrm{O\,II}]}$ [a] [M$_\odot$yr^{-1}]	$12+\log(\mathrm{O/H})_{\mathrm{l}}$ [a]	$12+\log(\mathrm{O/H})_{\mathrm{u}}$ [a]	f_{CO} [GHz]
z_{gal}	S_{CO} [Jy km s^{-1}]	FWHM$_{\mathrm{CO}}$ [km s^{-1}]	$L_{\mathrm{CO(1\text{-}0)}}$ [10^9 K km s^{-1} pc^2]	M_{mol} [10^9M$_\odot$]	τ_{dep} [Gyr]
Q2131z043$_{\mathbf{H\,I}}$	**0.43**	**19.5 ± 0.15** [b]	**> -0.96** [a]		
Q2131-G1	52 / 9.2	2.00 ± 0.2	8.98 ± 0.02	-	241.866
0.42974	0.36 ± 0.02	184 ± 50	1.42 ± 0.08	$3.52^{+3.95}_{-0.31}$	< 4.15
Q2131-G2	61 / 10.7	0.20 ± 0.1	8.32 ± 0.16	-	241.697
0.4307 [a]	< 0.068	-	< 0.27	< 3.64	< 36.37
Q1232z039$_{\mathbf{H\,I}}$	**0.3950**	**20.75 ± 0.07** [c]	**< -1.31** [c]		
Q1232-G1	8 / 1.5	0.67 ± 0.09	8.02 ± 0.06	8.66 ± 0.04	247.829
0.3953 [a]	< 0.070	-	< 0.24	< 6.09	< 8.02
Q1232z075$_{\mathbf{Mg\,II}}$	**0.7572**	**$18.36^{+0.09}_{-0.08}$** [d]	**> -1.48** [d]		
Q1232-G2	68 / 9.1	2.58 ± 0.23	8.19 ± 0.19	8.54 ± 0.19	262.462
0.7566 [a]	< 0.12	-	< 0.83	< 18.31	< 7.80
Q0152z038$_{\mathbf{H\,I}}$	**0.3887**	**<18.8** [e]	**> -1.36** [a]		
Q0152-G1	60 / 11.5	1.04 ± 0.03	8.65 ± 0.09	-	250.105
0.3826 [a]	< 0.17	-	< 0.53	< 2.80	< 2.78
Q1211z039$_{\mathbf{H\,I}}$	**0.3929**	**19.46 ± 0.08** [b]	**> -1.05** [a]		
Q1211-G1	37 / 6.8	4.71 ± 0.08	8.16 ± 0.01	8.48 ± 0.01	248.274
0.3928 [a]	< 0.15	-	< 0.49	< 6.78	< 1.47
Q1130z031$_{\mathbf{H\,I}}$	**0.3127**	**21.71 ± 0.07** [f]	**-1.94 ± 0.08** [g] [1]		
Q1130-G2	44 / 9.5	0.44 ± 0.3	8.77 ± 0.05	-	263.4
0.3127 [a]	0.63 ± 0.01 [g]	250 ± 50 [g]	3.1 ± 0.1 [g]	$11.03^{+1.44}_{-1.27}$	25^{+21}_{-20}
Q1130-G4	82 / 17.7	> 0.40	< 8.65	-	263.44
0.3126 [a]	0.42 ± 0.03 [g]	535 ± 50 [g]	2.1 ± 0.1 [g]	> 8.88	$\gtrsim 22.19$
Q1130-G6	98 / 21.3	1.14 ± 0.7	8.94 ± 0.16	-	263.67
0.3115 [a]	0.20 ± 0.01 [g]	205 ± 50 [g]	1.0 ± 0.1 [g]	$2.65^{+1.20}_{-0.82}$	$2.3^{+1.4}_{-1.1}$

galaxies making the used SLED a first order approximation for Q2131-G1 (Klitsch et al., 2019a). The molecular mass is calculated by using the geometric mean of the Bolatto et al. (2013) and Genzel et al. (2012) $\alpha_{\mathrm{CO}}(\mathrm{Z})$ prescription. This conversion factor is a good approximation for galaxies that are not significantly below solar metallicity and therefore appropriate for Q2131-G1 [12+log(O/H) = 8.98±0.02 (Péroux et al., 2017), also see Genzel et al. (2015) for a more detailed description of this averaged conversion factor]. We note that Muzahid et al. (2016) derived a lower metallicity, closer to the solar metallicity, for Q2131-G1 (12+log(O/H) = 8.68 ± 0.09). This discrepancy can be explained by the use of the N2-index, which is known to saturate at solar metallicities (Pettini & Pagel, 2004). We elect to use the R_{23} based metallicity by Péroux et al. (2017), but note that the metallicity is based on emission line fluxes that have not been dust-corrected and therefore possibly overestimate the metallicity. We therefore base the conversion factor on 12+log(O/H) = 8.98 ± 0.02, but include the lower metallicity in the error calculation and compute $\alpha_{\mathrm{CO}} = 2.48^{+2.50}_{-0.08}\ \mathrm{M}_{\odot}(\mathrm{K\ km/s\ pc})^{-1}$. The molecular mass is $M_{\mathrm{mol}} = 3.52^{+3.95}_{-0.31} \times 10^9 \mathrm{M}_{\odot}$. The calculated molecular mass is consistent with the mass limit of $M_{\mathrm{mol}} \leq 7 \times 10^9 \mathrm{M}_{\odot}$ using $L'_{\mathrm{CO(2\text{-}1)}} \leq 3.8 \times 10^9$ K km s^{-1}pc^2 (Klitsch et al., 2021).

Using the non-dust corrected star formation rate (SFR) derived by Hamanowicz et al. (2020) ($\mathrm{SFR}_{[\mathrm{O\,II}]} = 2.00 \pm 0.2 \mathrm{M}_{\odot}\mathrm{yr}^{-1}$) we calculate the limit on the depletion time using:

$$\tau_{\mathrm{dep}} < \frac{\mathrm{M}_{\mathrm{mol,max}}}{\mathrm{SFR}_{[\mathrm{O\,II,min}]}}\mathrm{yr} \tag{2.1}$$

The depletion time for Q2131-G1 is $\tau_{\mathrm{dep}} < 4.15$ Gyr.

Stellar Mass

In this section, we estimate the stellar mass of Q2131-G1. The stellar mass is derived from the Mass-Metallicity-Relation (MZR) (Tremonti et al., 2004). This relation is based on $\sim$ 53000 galaxies at $z \sim 0.1$ from the Sloan Digital Sky Survey (SDSS) sample and holds for $8.5 < \log(M_{\star}/\mathrm{M}_{\odot}) < 11.5$. Using the metallicity derived by Péroux et al. (2017) [12+log(O/H) = 8.98 ± 0.02] we get two solutions: $\log(M_{\star}/\mathrm{M}_{\odot}) = 10.1 \pm 0.1$ and $\log(M_{\star}/\mathrm{M}_{\odot}) = 12.9 \pm 0.1$. This relation does not hold for the second solution, as that stellar mass would be outside of the valid range. We attempt to break this degeneracy by applying the Tully-Fisher relation (linking the stellar mass with the maximum rotation velocity of the galaxy) (Tully & Fisher, 1977). We used the relation by Puech et al. (2008), derived from a sample of $z \sim 0.6$ galaxies using kinematics from the [O II] line. Using $V_{\mathrm{max}} = 200 \pm 3$km s^{-1} [as derived by the kinematical analysis of the [O III] λ5008 line in Péroux et al. (2017)] we estimate the stellar mass of G2131-G1 to be $\log(M_{\star}/\mathrm{M}_{\odot}) = 10.54 \pm 0.71$. This stellar mass is consistent with the lower stellar mass derived from the MZR. For further calculations, we decide to use the stellar mass derived from the MZR, but take into account the value derived by the Tully-Fisher relation and by the MZR using the (Muzahid et al., 2016) metallicity [12+log(O/H) = 8.68 ± 0.09, $\log(M_{\star}/\mathrm{M}_{\odot}) = 9.1^{+0.3}_{-0.2}$] in the error calculations: $\log(M_{\star}/\mathrm{M}_{\odot}) = 10.1^{+0.5}_{-1.0}$

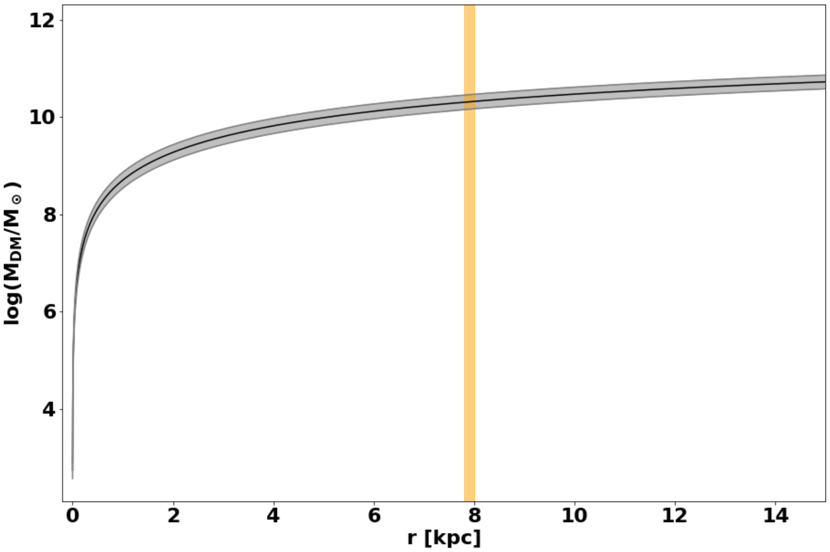

Figure 2.1: The cumulative mass of the dark matter within Q2131-G1 is derived assuming an NFW profile. The shaded regions show the profile for the minimum/maximum derived dark matter mass M_{200}. The vertical orange line marks the [O III] λ5008 half-light radius. The dark matter fraction within the half-light radius is in the range of $f_{\rm DM} = 0.24 - 0.53$. Therefore, we find the central regions of this galaxy to be baryon dominated.

Dark Matter Fraction

Current studies have shown a declining dark matter fraction with increasing redshift (e.g. Genzel et al., 2020; Price et al., 2020). We constrain the dark matter contribution to the galaxy within the half-light radius. We create an NFW-profile (Navarro et al., 1997) based on the halo mass estimate (see section 2.4.1) and compute the corresponding cumulative mass curve. We note that this is a first order approximation of the dark matter fraction within the central region of G2131-G1.

The halo mass estimate is based on abundance matching [e.g Behroozi et al. (2010), Moster et al. (2010) and Moster et al. (2018)]. We use the prescription provided in Genzel et al. (2020) [equation A13 in Genzel et al. (2020), provided in a priv. comm. with B. Moster] based on the galaxy halo pairs from Moster et al. (2018) to fit a halo mass - galaxy mass relation. This relation is appropriate for $z > 0.5$ and provides an estimate of the halo mass derived from the stellar mass. Using the stellar mass of $\log(M_\star/{\rm M}_\odot) = 10.1^{+0.5}_{-1.0}$ we compute a halo mass of $\log(M_{200}/{\rm M}_\odot) = 11.6 \pm 0.5$. This halo mass is consistent with the halo mass derived by Péroux et al. (2017) assuming a spherical virialized collapse model by Mo & White (2002) [$\log(M_{200}/{\rm M}_\odot) = 12.46^{+0.03}_{-0.04}$]. The corresponding radius (r_{200}), within which the mean mass density is ~ 200 times the critical density of the Universe, is calculated using:

$$r_{200} = \left[\frac{M_{200}}{\frac{4}{3}\,\pi\,200\,\rho_{crit}} \right]^{\frac{1}{3}} \quad , \tag{2.2}$$

with ρ_{crit} being:

$$\rho_{\rm crit} = \frac{3\,H^2(z)}{8\,\pi\,G} \quad , \tag{2.3}$$

and using:

$$H(z) = H_0 \sqrt{\Omega_M\,(1+z)^3 + \Omega_\Lambda} \quad . \tag{2.4}$$

Using equations 2.2, 2.3 and 2.4 we compute: $H(z) = 87.9$ km s^{-1}Mpc^{-1}, $\rho_{\rm crit} = 437.61\ h^2\ {\rm M_\odot kpc^{-3}}$ and $r_{200} = 133^{+54}_{-36}$ kpc.

In order to fully describe the NFW mass-profile we compute the concentration parameter (c) which we compute using the redshift dependent NFW concentration-mass relation from Dutton & Macciò (2014):

$$\log(c) = a + b \times \log(M_{200}/[10^{12} h^{-1} {\rm M_\odot}]) \quad , \tag{2.5}$$

with:

$$a = 0.520 + (0.905 - 0.520) \times \exp(-0.617 \times z^{1.21}) \tag{2.6}$$
$$b = -0.101 + 0.026 \times z \quad , \tag{2.7}$$

and with δ_c being:

$$\delta_c = \frac{200}{3} \times \frac{c^3}{\ln(1+c) - \frac{c}{1+c}} \quad . \tag{2.8}$$

Using our derived M_{200} we find $c = 7.5^{+0.7}_{-0.7}$ and $\delta_c = 22430^{+4924}_{-4364}$.
We calculate the NFW mass-profile using:

$$M_{\rm DM}(r) = 4\pi\rho_0 r_{\rm s}^3 \times \left[\ln\left(1 + \frac{r}{r_{\rm s}}\right) - \frac{\frac{r}{r_{\rm s}}}{1 + \frac{r}{r_{\rm s}}} \right] \quad . \tag{2.9}$$

with $\rho_0 = \delta_{\rm c}\,\rho_{\rm crit}$ and $r_{\rm s} = \frac{r_{200}}{c}$.
The resulting mass profile is shown in Fig. 2.1. The dark matter mass is in the range of $\log(M_{\rm DM}(r_{1/2})/\ {\rm M_\odot}) = 10.16 - 10.46$ at the [O III] $\lambda 5008$ emission half-light radius $r_{1/2} = 7.9 \pm 0.1$ kpc.

We calculate the dynamical mass within $r_{1/2}$ using (Epinat et al., 2009):

$$M_{\rm dyn}(r_{1/2}) = \frac{V^2(r_{1/2})\,r_{1/2}}{G} \quad , \tag{2.10}$$

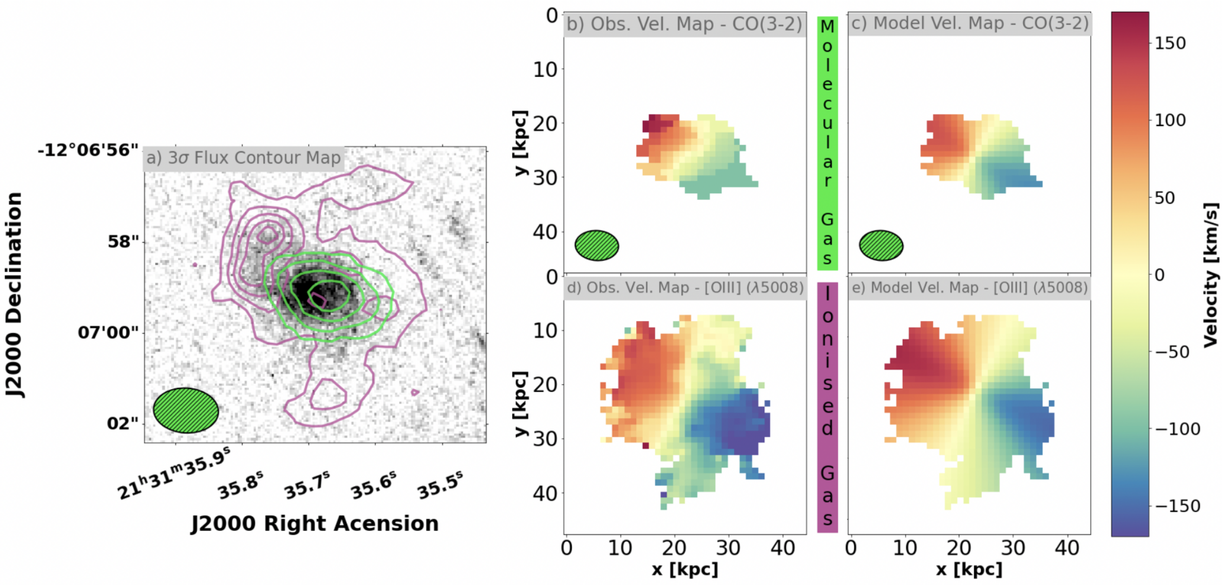

Figure 2.2: Contour plot and velocity maps of Q2131-G1. The ionized gas contour plot and velocity maps are based on [O III] λ5008. The molecular gas contour plot and velocity maps are based on CO (3–2). a) HST image of Q2131-G1 (detector: PC, filter: F702W, Kacprzak et al., 2015) overlayed with contour plots of the [O III] λ5008 (purple) and CO(3–2) flux using 3σ steps. b) Observed velocity map of the molecular gas c) Model velocity map of the molecular gas d) Observed velocity map of the ionized gas. e) Model velocity map of the ionized gas. In the contour plot a region of high [O III] λ5008 flux is visible where no CO (3–2) is being observed above the 3σ threshold. The direction of rotation for both the ionized and molecular gas are closely correlated and both gas phases show similar maximum rotational velocities.

with $V(r_{1/2})$ being computed using an arctan velocity profile with the fit parameters derived by GalPak^{3D} using the [O III] λ5008 emission line (Bouché et al., 2015):

$$V(r_{1/2}) = V_{\max}\,\frac{2}{\pi}\arctan(\frac{r_{1/2}}{r_{\rm t}}) \quad , \tag{2.11}$$

with $r_{\rm t} = 1.51$ kpc being the turnover radius. The velocity at the half-light radius is therefore $V(r_{1/2}) = 176 \pm 3$ km s^{-1} and the dynamical mass at $r_{1/2}$ is $M_{\rm dyn}(r_{1/2}) = 10.75 \pm 0.3$. Using the dynamical mass and the dark matter mass within the half-light we compute the dark matter fraction within the half-light radius to be $f_{\rm DM} = 0.24 - 0.54$.

Morphological and Kinematical Properties

We first study the morphological properties of Q2131-G1 based on the flux observed with HST, MUSE, and ALMA. The HST image (filter: F702W) of Q2131-G1 with overlayed contours from the observed [O III] λ5008 and CO(3–2) observed flux maps can be found in

Fig. 2.2 a). CO(3–2) has a compact and elliptical morphology in the centre of the galaxy with an extent $\sim$ 20 kpc. We stress that these higher-z observations would not resolve small-scale clumps as observed in the PHANGS-ALMA survey (Schinnerer et al., 2019). The ionized gas shows a greater extent of $\sim$ 40 kpc and has a shape that indicates spiral arms or possible tidal tails (Péroux et al., 2017). There is a region of a [O III] λ5008 flux maximum (at $\sim 21^{\rm h}31^{\rm m}35.77^{\rm s}$, $-12°06'57.8''$) where CO(3–2) is not detected above the 3σ threshold. This region coincides with a spiral structure in Q2131-G1 and therefore most likely is a region of active star formation. A large fraction of the molecular gas in this region is possibly already depleted due to the star formation process, leading to a CO flux density below the 3σ threshold. The stellar continuum observed by HST extends beyond the molecular gas emission above the 3σ threshold.

We study the kinematics of both the ionized and molecular gas of the detected galaxy using the 3D fitting algorithm GalPak$^{\rm 3D}$ (Bouché et al., 2015). The algorithm assumes a disk parametric model with 10 free (but also optionally fixable) parameters and probes the parameter space by implementing a Monte Carlo Markov Chain approach with non-traditional sampling laws. The algorithm provides stable results if the signal-to-noise (SNR) per spaxel of the brightest spaxel in the cube is SNR>3. Additionally, the half-light radius has to satisfy the condition $r_{1/2}/\mathrm{FWHM} > 0.75$ in order for the algorithm to converge, with the FWHM being the Full Width Half Maximum of the Point Spread Function (PSF). The $r_{1/2}$/FWHM ratio of Q2131-G1 is below that condition $r_{1/2}/\mathrm{FWHM} \sim 0.5$, but the algorithm nonetheless fully converges as we assessed from the MCMC chain. In order to be consistent with the ionized gas kinematic model used in (Péroux et al., 2017) we also use the exponential flux profile and an arctan velocity profile as assumptions for the disk model. We also ensured that the ALMA cube is in the same reference frame as the MUSE cube (BARY). We additionally create two models with an exponential and tanh velocity profile, which yield different results, in order to take the differences in models into account for the error calculation of the derived properties. The observations are well reproduced by a rotating disk, as can be assessed from the low residuals in the flux (Fig. A.2) and velocity residual maps (Fig. A.3) in the Appendix (Section A.1).

The morpho-kinematical properties of the ionized gas of Q2131-G1 derived from the [O III] λ5008 and Hβ line in the MUSE observations are described in Péroux et al. (2017). The authors report the following: The maximum circular velocity is well constrained at $V_{\rm max} = 200 \pm 3$ km s^{-1}, the half-light radius is found to be $r_{1/2} = 7.9 \pm 0.1$ kpc, the derived position angle is PA $= 65 \pm 1°$ and the inclination is $i_{\rm CO} = 60.5 \pm 1.2$. Based on the derived flux, velocity, and dispersion maps, Péroux et al. (2017) argue that the galaxy is a large rotating disc, with a velocity gradient along the major axis and a dispersion peak at the centre of the galaxy. Using this approach we create velocity maps of both the ionized and molecular gas. The observable [b) - ALMA CO(3–2), (d) - MUSE [O III] λ5008)] and model [c) - ALMA CO(3–2), (e) - MUSE λ5008] velocity maps are shown in Fig. 2.2. We find that the rotational velocities for both the ionized and molecular gas are closely correlated. This is also the case for the model maximum velocities of both components ($V_{\rm max,[O\,III]} = 200 \pm 3$ km s^{-1} and $V_{\rm max,CO} = 195^{+4}_{-30}$ km s^{-1}). Both of the model velocities are consistent with the observed velocities of both components ($V_{\rm max\text{-}obs,[O\,III]} \sim 205$ km s^{-1}

Table 2.2: **Morpho-kinematic properties of galaxies detected in both [O III] and CO(1-0) / CO(3–2).**
Row 1 (red) - properties derived from [O III]: (1) reference name of the galaxy used in this paper, (2) half-light radius, (3) inclination, (4) position angle, (5) maximum velocity, (6) dynamical mass, (7) halo mass.
Row 2 (green) - properties derived from CO(1-0) / CO(3–2): (1) half-light radius, (2) inclination, (3) position angle, (4) maximum velocity.
Literature references: The values for Q2131-G1 [O III] are taken from Péroux et al. (2017) and the values for Q1130-G2 [O III]/CO, Q1130-G4 [O III]/CO are taken from Péroux et al. (2019).

Galaxy	$r_{1/2,[\text{O III}]}$ [kpc]	$i_{[\text{O III}]}$ [deg]	$\text{PA}_{[\text{O III}]}$ [deg]	$V_{\text{max},[\text{O III}]}$ [km s^{-1}]	$\log(M_{\text{dyn},[\text{O III}]})$ [M$_\odot$]	$\log(M_{\text{h},[\text{O III}]})$ [M$_\odot$]
	$r_{1/2,\text{CO}}$ [kpc]	i_{CO} [deg]	PA_{CO} [deg]	$V_{\text{max,CO}}$ [km s^{-1}]		
Q2131-G1	7.9 ± 0.1	60.5 ± 1.2	65 ± 1	200 ± 3	10.87 ± 0.03	11.7 ± 0.1
	$3.7^{+0.5}_{-0.1}$	47^{+10}_{-1}	59 ± 2	195^{+4}_{-30}		
Q1130-G2	14 ± 2	77 ± 2	131 ± 2	264 ± 14	11.3 ± 0.2	12.9 ± 0.1
	2 ± 1	76 ± 3	117 ± 2	134 ± 14		
Q1130-G4	9 ± 2	54 ± 2	86 ± 2	231 ± 12	11.1 ± 0.2	12.7 ± 0.1
	6 ± 1	82 ± 4	84 ± 2	290 ± 19		

and $V_{\text{max-obs,CO}} \sim 190$ km s^{-1}).

The derived inclination of the molecular and ionized gas in Q2131-G1 are $i_{[\text{O III}]} = 60.5 \pm 1.2°$ and $i_{\text{CO}} = 47^{+10°}_{-1}$. The position angles (PA) are $\text{PA}_{[\text{O III}]} = 65 \pm 1°$ and $\text{PA}_{\text{CO}} = 59 \pm 2°$. We conclude that the gas phases in Q2131-G1 are aligned directionally.

While the two models converge in terms of morpho-kinematical properties, they differ in redshifts (CO (3–2): $z_{\text{CO}} = 0.42974 \pm 0.00001$, [O III] λ5008: $z_{[\text{O III}]} = 0.42914 \pm 0.00001$, H$\beta$: $z_{\text{H}\beta} = 0.42950 \pm 0.00001$). The other [O III] line in the spectrum is too weak and the [O II] line is disregarded due to its doublet nature. We attribute this discrepancy to a combination of the wavelength calibration uncertainty of MUSE, which translates to a velocity uncertainty of ~ 25 km s^{-1}, and an underestimate of the errors provided by GalPak$^{3\text{D}}$. The ALMA frequency accuracy is set by the system electronics and is much better than the corresponding channel width of the cube (50 km s^{-1}). We therefore use the redshift derived from the CO (3–2) model as a zero-point in the analysis of the absorber and gas kinematics. We include the value of $z_{\text{H}\beta}$ and other uncertainties mentioned above to estimate an error of ± 100 km s^{-1} (~ 25 km s^{-1} MUSE velocity uncertainty + ~ 75 km s^{-1} kinematical modelling uncertainty) for the kinematic zero-point of the [O III] emission line in the following study of the absorber and gas kinematics. For the CO (3–2) zero-point we estimate an error of ~ 75km s^{-1} (kinematical modelling uncertainty).

Galaxy - Gas Connection

Kinematical studies of the gas in the galaxies seen in emission and probed by the quasar sightlines allow us to probe what galaxy/environment the absorbing gas is tracing. We use an approach based on the model rotation curve obtained by GalPak^{3D} to tackle this question.

We extrapolate the rotation curves of Q2131-G1 for both MUSE and ALMA data to the line-of-sight (LOS) towards the quasar to relate it to the gas traced by the H_2 and Mg II absorber. The corresponding plots can be found in Fig. 2.3 where we additionally show the normalized absorption and emission lines with the zero-point of velocity at the redshift of CO (3–2) derived by GalPak^{3D} ($z_0 = 0.42974$). We find the extrapolated velocities of the molecular and ionized of Q2131-G1 between ~ -130 and -135 km s^{-1} and ~ -255 and -275 km s^{-1}. The absorption features, with column densities of $\log(N_{\rm H\,I}/{\rm cm}^{-2}) = 19.5 \pm 0.15$ and $\log(N_{\rm H_2}/{\rm cm}^{-2}) = 16.36 \pm 0.08$ (Muzahid et al., 2016) are found between ~ -60 and $+60$ km s^{-1} from the zero-point.

A limit on the CO absorption column density of the absorber Q2131z043$_{\rm H\,I}$ associated with the galaxy Q2131-G1 is calculated following (Mangum & Shirley, 2015), using an excitation temperature equal to the CMB temperature at the redshift, a 5σ level from the spectrum at the expected position and frequency of the CO(3–2) absorption line as the detection threshold and a FWHM of 40 km s^{-1} and derive $\log(N_{\rm CO}/{\rm cm}^{-2}) < 14.6$. Using the mean ratio of $N_{\rm CO}/N_{\rm H_2} = 3 \times 10^{-6}$ (Burgh et al., 2007) we derive $\log(N_{\rm H_2}/{\rm cm}^{-2}) < 20.1$. This limit is consistent with the value observed from UV wavelength absorption by Muzahid et al. (2016).

Studies of the absorption and emission metallicity connect the absorber to its host. Using a metallicity gradient based on a sample of galaxy-absorber pairs (-0.022 ± 0.004 dex/kpc, Christensen et al., 2014), we extrapolate the metallicity of Q2131-G1 to the LOS towards the quasar. We take into account the observed flattening of the Oxygen metallicity gradient beyond $2 \times r_{1/2}$ (Sánchez-Menguiano et al., 2016) and assume that there is no change in the metallicity of the galaxy between $2 \times r_{1/2} = 15.8 \pm 0.2$ kpc and the impact parameter $b = 52$ kpc. We use 12+log(O/H) = 8.98 ± 0.02 by Péroux et al. (2017) as the metallicity of the galaxy, including the value by Muzahid et al. (2016) [12+log(O/H) = 8.68 ± 0.09] in the error calculation. The extrapolated metallicity of Q2131-G1 at the impact parameter ($b = 52$ kpc) is $Z_{\rm em} = -0.06^{+0.09}_{-0.62}$. We additionally use an alternative metallicity gradient of $0.1/r_{1/2}$ (which in the case of Q2131-G1 translates to 0.01266 ± 0.00016 dex/kpc) derived by the CALIFA survey (Sánchez et al., 2014) and find the extrapolated metallicity of Q2131-G1 at the impact parameter to be $Z_{\rm em} = 0.09^{+0.02}_{-0.48}$. Literature provides metallicity measurements using various species: $[{\rm Fe/H}]_{\rm abs} > -0.96$ from Hamanowicz et al. (2020), $[{\rm O/H}]_{\rm abs} = -0.26 \pm 0.19$ using ionisation modelling from Muzahid et al. (2016), the ionisation corrected metallicity of $[{\rm S/H}]_{\rm abs} > -0.72$ [originally reported as $[{\rm S/H}]_{\rm abs} > -0.40$ assuming $\log(N_{\rm H\,I,abs}/{\rm cm}^{-2}) = 19.18$ instead of 19.5) by Som et al. (2015)]. The global dust-free metallicity is $[{\rm X/H}]_{\rm abs} = -0.54 \pm 0.18$ (Péroux et al., 2017).We find that both of the extrapolated metallicities are consistent with each other and consistent with the metallicity derived by Péroux et al. (2017).

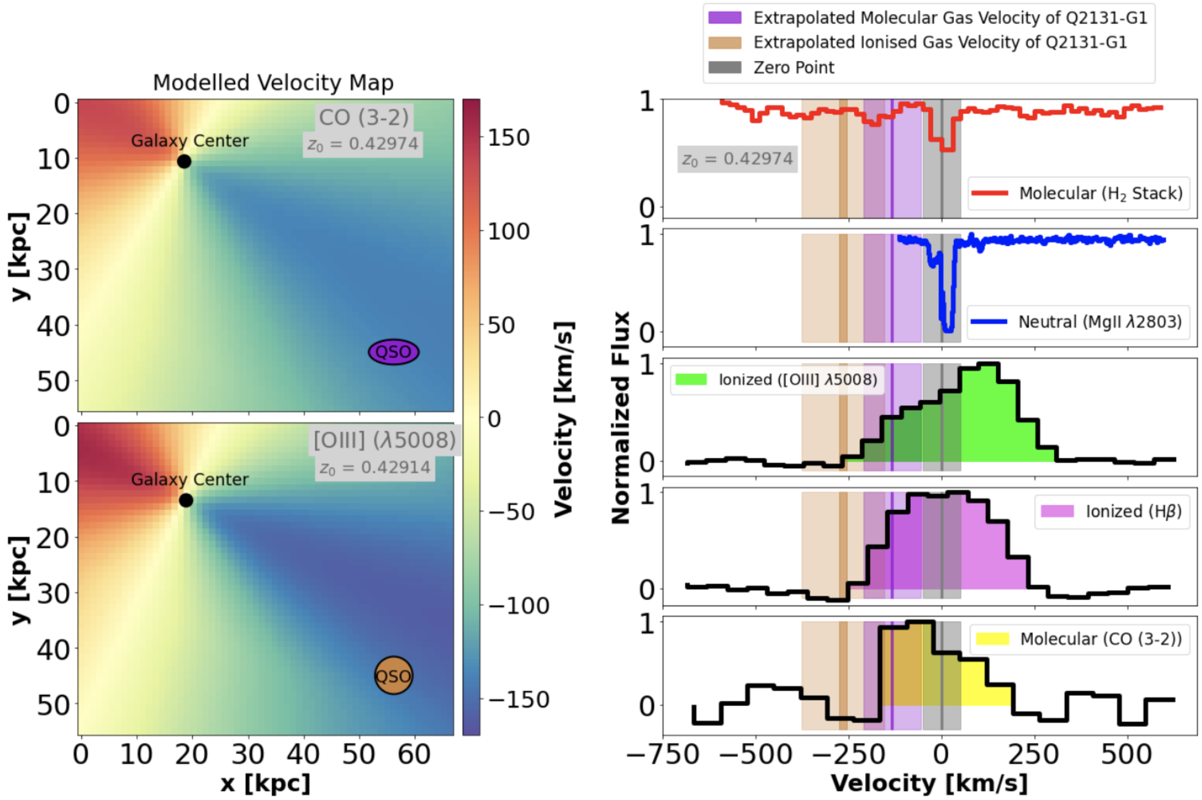

Figure 2.3: Extrapolated model velocity maps of Q2131-G1 and normalized flux of the Mg II λ2803 and stacked H_2 absorption line and [O III] λ5008, Hβ, CO(3–2) emission lines. The velocity zero point of the spectra is set to the redshift of CO (3–2) derived from the kinematic study ($z_0 = 0.42974$) (displayed as the gray shaded area). The magenta and brown bars display the extrapolated velocities of the molecular and ionized gas of Q2131-G1 respectively. The shaded magenta and brown bars display the errors of the extrapolated velocities (75 km s^{-1} for the molecular gas, 100 km s^{-1} for the ionized gas). Extrapolating the model velocity maps derived from GalPak^{3D} to the line of sight toward the quasar shows that at the position of the quasar the molecular and ionized gas of Q2131-G1 are located between ~ -130 and -135 km s^{-1} and ~ -255 and -275 km s^{-1} respectively, while the absorption features are found between ~ -60 and $+60$ km s^{-1}. We thus conclude that the two-component absorption features are consistent with in part an extended rotating disk of Q2131-G1 and in part gas falling onto Q2131-G1.

2.4.2 Limits from Non-Detections

For the fields Q1232-0224, Q0152-2001, and Q1211-1030 with no CO-detected counterparts in emission to the galaxies observed with MUSE and HST, we derive limits on the molecular mass and depletion times.

For each cube we consider an ellipsoidal area with the minor axis, position angle, and FWHM of the synthesized beam centred around the expected position of the galaxy with the frequency range being set to ± 100 km s^{-1} centred around the redshifted frequency of the CO(3–2) emission line. We then assume the emission spectrum to be a Gaussian with an amplitude set to the RMS of the ellipsoidal area and a FWHM of 200 km s^{-1}. The flux limit is then the area under this line within the 5σ range.

We calculate the mass limits and depletion times following the same prescription as described in section 2.4.1 [namely following Solomon et al. (1992), Fixsen et al. (1999) and Genzel et al. (2015)] and Equation 2.1. The molecular gas mass limits use an α_{CO} conversion factor based on the lowest measured metallicity of the galaxy to provide conservative limits of both the molecular gas mass and depletion time.

The results for the CO flux, luminosity, mass, and depletion time limits for Q0152-G1, Q1211-G1, and Q1232-G1 can be found in Table 2.1. The CO(1–0) limits on the luminosity L_{CO} are of the order $L_{\mathrm{CO}} \sim 10^8$ K kms^{-1}pc^2, which fits the sensitivity estimates based on the ALMA sensitivity calculator calculated for our observations at $z \sim 0.4$. Our limits are more stringent than similar observations studying the molecular gas in objects associated with absorbers [e.g. MEGAFLOW by Freundlich et al. (2021), targeting galaxies around Mg II absorbers, or Kanekar et al. (2018, 2020)], which are sensitive to luminosities $L_{\mathrm{CO}} >\sim 10^9$ K kms^{-1}pc^2. The molecular gas mass limits are in the range of $\mathrm{M_{mol}} \sim (2.8 - 18.3) \times 10^9\ \mathrm{M_\odot}$ and the depletion time limits are in the range of $\tau_{\mathrm{dep}} \sim 1.4 - 37$ Gyr.

2.5 Discussion

The multi-wavelength approach in this work allows us to closely study the different gas phases within and around H I-selected galaxies. HST spectroscopy provides neutral and molecular gas information through absorption while MUSE and ALMA observations enable us to study the ionized and molecular gas content through emission. In this section we provide a detailed discussion of the observed properties and how they compare to current observations.

2.5.1 Strongly Coupled Gas Phases within a Rotating Disk

Recent observations of the ionized and molecular gas phases in galaxies between redshifts $z \sim 0.1 - 1.4$ have found that both phases mostly align well directionally (e.g. Übler et al., 2018; Møller et al., 2017; Klitsch et al., 2018; Loiacono et al., 2019; Péroux et al., 2019; Molina et al., 2019, 2020). Similarly, we find that Q2131-G1 is well constrained by a disk model and that the ionized and molecular gas phases are aligned well directionally with similar inclinations and position angles.

We also find a similar maximum rotational velocity ($\mathrm{V_{max}} \sim 200\,\mathrm{km\ s^{-1}}$) of the molecular and ionized gas within Q2131-G1. This is consistent with the EDGE-CALIFA survey (Levy et al., 2018), where ionized and molecular gas kinematics (traced by Hα) were compared in local galaxies. While the survey does find that for the majority of galaxies the rotational velocity measured from the molecular gas is higher than that from the ionized gas, there are cases where similar rotational velocities for both phases have been observed.

Due to the good alignment of the ionized and molecular gas phases, both directionally and rotationally, we find that the two gas phases are strongly coupled within Q2131-G1.

2.5.2 Identifying the Disk Tilt

Kinematic modelling provides the inclinations of both gas phases, but these values are degenerate without knowing the tilt of the disk. A proposed solution to breaking the degeneracy of the disk tilt is to use the rotation curve and the winding direction of spiral arms (Martin et al., 2019). Based on the likely assumption that in a self-gravitating, collisionless system only trailing spiral patterns are long-lived (Carlberg & Freedman, 1985) and most spiral patterns therefore lag behind the direction of rotation with increasing radius, depending on the winding rotation, one can infer a positive or negative sign of the inclination. The winding rotation of the spiral arms in Q2131-G1 observed in the HST image are opposite to the direction of rotation of the galaxy and the inclination therefore has a negative sign (see Fig. 2.4).

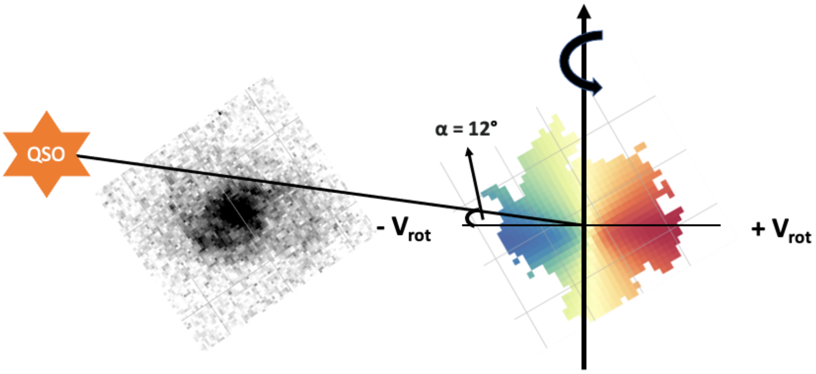

Figure 2.4: Sketch of the QSO - galaxy plane for identifying the disk tilt. The galaxy is rotated in order to align the major axis with the x-axis in the sketch. The spiral arms of Q2131-G1 wind in the opposite direction of the galaxies rotation and we conclude that the inclination has a negative sign.

2.5.3 Gas Probed in Absorption Connected to a Rotating Disk and Infalling Gas

Previous authors state that an individual absorber is sometimes associated with multiple galaxies (Hamanowicz et al., 2020). In particular, in the field Q2131-1207 four galaxies are found at the same redshift and physically close to the absorber, indicating that Q2131-G1, found at $b = 52$ kpc, is part of a group environment. Kinematical studies of the gas phases and the absorption features help alleviate these ambiguities by studying how the different components relate in velocity space (see e.g. Rahmani et al., 2017).

To relate the gas probed in absorption with the absorber host we extrapolate the model rotation curve towards the sightline of the quasar in section 2.4.1 (see Fig. 2.3). We find that the velocity of the ionized and molecular gas of Q2131-G1 at the point of the quasar sightline are blueshifted compared to the systemic redshift. A two-component absorption is found between ~ -60 and 60 km s^{-1}. Due to the low azimuthal angle ($12^\circ \pm 1^\circ$) of Q2131-G1 (Péroux et al., 2017) and simulations indicating that outflowing gas preferentially leaves the galaxy in a conical shape along its minor axis (Brook et al., 2011; Péroux et al., 2020), we assume an outflow scenario to be unlikely for both absorption components.

The weaker component is rotating in the same direction as the galaxy at less negative velocities. Further, the extrapolated metallicities of Q2131-G1 ($Z_{\rm em} = -0.06^{+0.09}_{-0.62}$ and $Z_{\rm em} = 0.09^{+0.02}_{-0.48}$, depending on the metallicity gradient used) at the LOS towards the quasar indicate a connection between the gas probed in absorption and emission as it is consistent with the absorber metallicity ($[{\rm X/H}]_{\rm abs} = -0.54 \pm 0.18$). The extrapolated velocities and metallicities of the galaxy and the weaker absorption component are therefore consistent with being part of an extended rotating disk.

The stronger absorption component is redshifted compared to the systemic redshift. Gas rotating with the disk of the galaxy is expected to have blueshifted velocities, making the stronger absorption component inconsistent with being part of the extended rotating disk. Further, the low azimuthal angle makes it a likely inflow (e.g. Bordoloi et al., 2011; Stewart et al., 2011; Shen et al., 2012). The metallicity difference between the Q2131-G1 and the absorber lies in the infalling section of the galaxy to gas metallicity versus azimuthal angle plot seen in Péroux et al. (2016) (figure 8 in the publication). Based on the metallicity difference and the geometry and orientation arguments, the stronger component is consistent with being gas falling onto Q2131-G1. We note that current data does not exclude that the gas could potentially also be falling onto Q2131-G2. The H_2 column density of the absorber also poses the question if and how it is possible to have a considerable molecular gas phase, with temperatures down to 10 K, in infalling gas.

We thus conclude that the two-component absorption features are consistent with in part an extended rotating disk of Q2131-G1 and in part gas falling onto Q2131-G1.

2.5.4 Specifics of H I-selected Systems

Previous studies of H I-selected systems have observed gas depletion times that are a few times longer than what is typically found in surveys of emission-selected galaxies (see especially Kanekar et al., 2018). This poses the question of whether the H I-selection preferentially selects galaxies that have large gas reservoirs for their SFR. We compare the detected galaxy Q2131-G1 with two current emission-selected molecular gas surveys, namely xCOLD GASS (e.g. Saintonge et al., 2017) and the PHIBSS 1 & 2 surveys (e.g. Tacconi et al., 2018) of galaxies at redshift $z < 1.1$. We additionally contrast with previously published H I-selected galaxies where molecular masses, stellar masses, and SFR have been measured. We use a metallicity-dependent $\alpha_{\rm CO}$ conversion factor for the comparison sample if metallicity information is provided (namely Genzel et al., 2015; Bolatto et al., 2013; Papadopoulos et al., 2012). Otherwise we use $\alpha_{\rm CO} = 4.3 {\rm M}_\odot ({\rm K\ km/s\ pc})^{-1}$ from

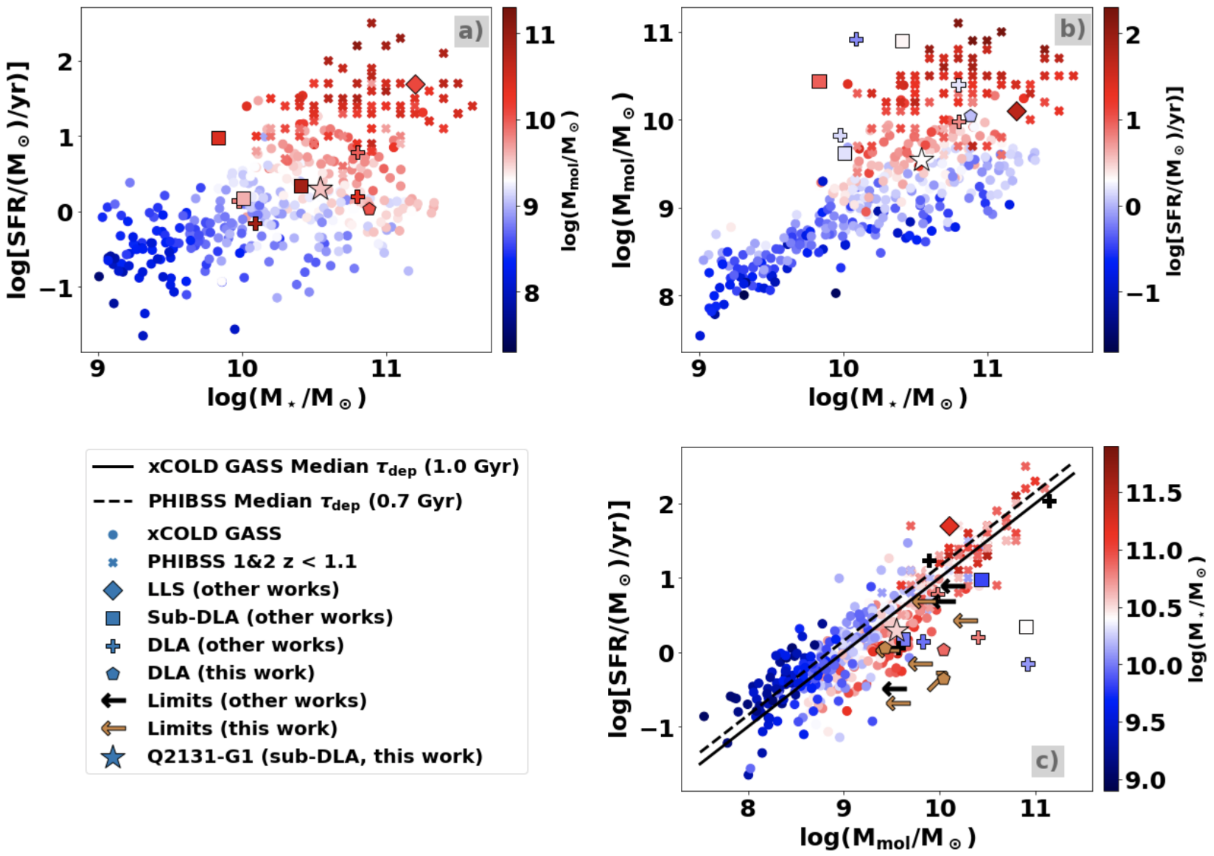

Figure 2.5: Star formation rate (SFR), molecular gas mass ($M_{\rm mol}$) and depletion time ($\tau_{\rm dep}$) plotted for H I-selected galaxies, the xCOLD GASS, and the PHIBSS 1&2 survey (at $z <$ 1.1). H I-selected galaxies with molecular gas mass limits as well as galaxies without stellar mass data are additionally plotted in the SFR - $M_{\rm mol}$ plot and have a black (other works) / brown (this work) color. The median depletion time for xCOLD GASS ($\tau_{\rm dep} = 1.0$ Gyr, Saintonge et al., 2017) and PHIBSS ($\tau_{\rm dep} = 0.7$ Gyr, Tacconi et al., 2018) are plotted as a black (dashed) line. Q2131-G1 lies within the $M_{\rm mol}$ - SFR, $M_\star$ - $M_{\rm mol}$ and $M_\star$ - SFR planes of the xCOLD GASS and PHIBSS 1&2 surveys. We note that the SFR of Q2131-G1 is not dust-corrected and therefore should be considered as a lower limit. The molecular gas mass for the majority of sub-DLAs and DLAs for their given SFR is found to be higher than for emission-selected samples. This leads to depletion times in H I-selected galaxies that are up to multiple factors higher than for emission-selected galaxies. This implies that selection based on strong H I-absorbers traces objects that have large gas reservoirs (at given SFR). Literature references: xCOLD GASS: Saintonge et al. (2017); PHIBSS Tacconi et al. (2018); LLS (other works): Klitsch et al. (2018); sub-DLA (other works): Kanekar et al. (2018); Neeleman et al. (2016); DLA (other works): Kanekar et al. (2018); Møller et al. (2017); Neeleman et al. (2018); Limits (other works): Klitsch et al. (2018); Kanekar et al. (2018).

(Bolatto et al., 2013) or in the case of Klitsch et al. (2018) $\alpha_{\mathrm{CO}} = 0.6\mathrm{M}_{\odot}(\mathrm{K\ km/s\ pc})^{-1}$ from Papadopoulos et al. (2012) is used because there is evidence that this galaxy is a luminous infrared galaxy (LIRG). In the case of the MUSE-ALMA Haloes sample the SFR is not dust corrected (with exception of the field Q1130-1449), therefore the SFR can be considered as a lower limit.

Fig. 2.5 shows the SFR, molecular mass, and depletion times of the emission and H I-selected galaxies. Q2131-G1 (star symbol) is comparable to the galaxies of the mass-selected xCOLD GASS and PHIBSS 1 & 2 galaxies, as it lies within the M_{mol} - SFR, $M_{\star}$ - M_{mol} and $M_{\star}$ - SFR planes. While Q2131-G1 fits well in the $\mathrm{M}_{\mathrm{mol}}$ - SFR plane, it is on the lower side of the derived SFR of comparable molecular masses, comparable to other galaxies associated with sub-DLAs. The deviations from the $\mathrm{M}_{\mathrm{mol}}$ - SFR plane are especially drastic in the case of galaxies associated with DLAs, which implies that H I-selection traces objects that have large gas reservoirs (at given SFR). Similarly the depletion timescale of Q2131-G1, Q1130-G2, and Q1130-G6 are an order of $\sim 2-53$ larger than the median for emission-selected galaxies in the xCOLD GASS and PHIBSS survey with $\tau_{\mathrm{dep,med}} \approx 1.0$ Gyr and ≈ 0.7 Gyr respectively.

Studying this trend is limited due to the low number of molecular gas and star formation rate observations of H I-selected galaxies. Further studies will test whether H I-selection preferentially selects galaxies that have large molecular gas reservoirs for their given SFR.

2.5.5 Connecting Galaxy Properties with Gas Properties

One key objective in studying absorption-selected galaxies is associating absorbers with potential absorber hosts and connecting absorber properties to the low-density gas found by absorption. We compare the derived molecular gas mass of Q2131-G1 and the H I column density of the associated absorber $\mathrm{Q2131z043}_{\mathrm{H\,I}}$ with previously detected H I absorbers and associated absorber hosts detected in CO in Fig. 2.6. In order to provide a fair comparison, we use the same conversion factors as described in section 2.5.4.

Molecular gas in H I-selected systems is found in systems with H I column densities between $\log(N_{\mathrm{H\,I}}/\mathrm{cm}^{-2}) \sim 18-22$, from Lyman-limit systems (Klitsch et al., 2018), to sub-DLAs (this work; Neeleman et al., 2016; Kanekar et al., 2018) and DLAs (Møller et al., 2017; Neeleman et al., 2018; Kanekar et al., 2018; Péroux et al., 2019) (see Fig. 2.6). The molecular masses detected span over a large range of $\log(M_{\mathrm{mol}}/\mathrm{M}_{\odot}) \sim 9.5-11.3$. The lower end of this range is typically for the detection limit of the observations. It is interesting to note that H I-selection can be associated with such large molecular gas reservoirs, but no correlation between the H I-absorption column density and the absorber host molecular mass is seen.

The most similar counterpart to Q2131-G1 is the galaxy associated with the absorber at redshift $z = 0.101$ in the quasar spectrum of PKS 0439-433 (Neeleman et al., 2016). While the absorber metallicity in PKS 0439-433 is higher ([S/H]= 0.1, Som et al., 2015), both absorbers show a H I-column density of $\log(N_{\mathrm{H\,I}}/\mathrm{cm}^{-2}) \sim 19.5$ and the associated galaxies have closely matching molecular masses of $\log(M_{\mathrm{mol}}/\mathrm{M}_{\odot}) \sim 9.6$. Additionally, both absorber systems have H_2 absorption features with H_2 column densities of $\log(N_{\mathrm{H}_2}/\mathrm{cm}^{-2}) \sim$

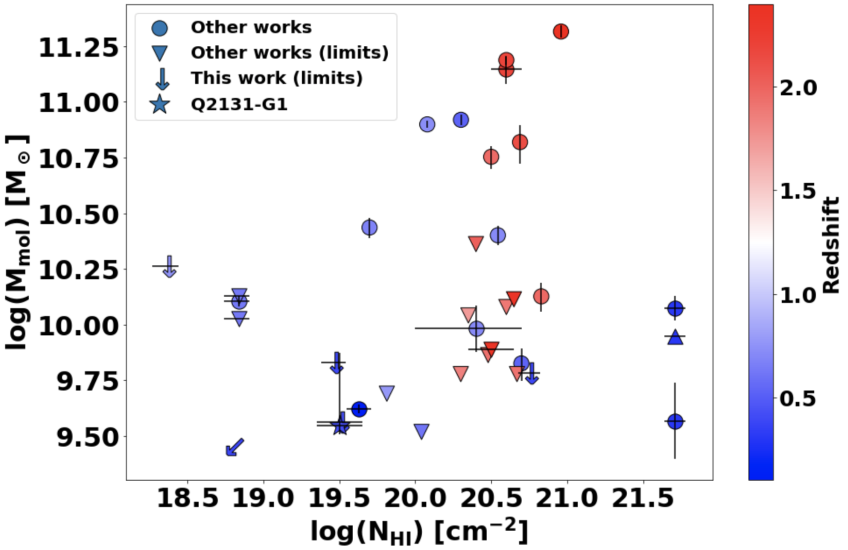

Figure 2.6: Absorber H I column density plotted against the molecular mass (limits) of absorber hosts by various published works (Neeleman et al., 2016; Møller et al., 2017; Augustin et al., 2018; Kanekar et al., 2018; Klitsch et al., 2018; Neeleman et al., 2018; Péroux et al., 2019; Kanekar et al., 2020) Q2131-G1 is on the lower side of previously detected molecular masses in H I-absorption selected galaxies.

16.5. The calculated limit on the CO column density [$\log(N_{\rm CO}/{\rm cm}^{-2}) < 14.6$] in Section 2.4.1 and the subsequently derived limit on the H2 column density [$\log(N_{\rm H_2}/{\rm cm}^{-2}) < 20.1$] is consistent with the detected H_2 column density. The galaxy in Neeleman et al. (2016) does have a lower impact parameter of ~ 20 kpc than Q2131-G1 ($b = 52$ kpc), but the absorption features cannot be kinematically associated with the rotating disk of the absorber host or infalling gas and is likely part of the CGM of the galaxy. While we might probe different environments, the similarity of the molecular masses in the absorber hosts and the H I/H_2 column densities of the absorbers indicate a connection of these parameters. The other galaxies in this sample either lack observations of possible H_2 absorption features, or have not been detected at all. This is partly due to the low detection rates of H_2 in quasars [$\sim$ 16 per cent for high-z absorbers (Noterdaeme et al., 2008), $\sim$ 50 per cent for low-z absorbers (Muzahid et al., 2015)]. Nonetheless, future studies of molecular gas in both absorbers and absorber hosts, combined with kinematic studies that help to associate these systems are essential for studying a possible connection between the high-density molecular gas found in galaxies and the low-density molecular gas found in absorbers.

2.5.6 A Dark Matter Fraction Evolving with Redshift

Current extragalactic surveys of the dark matter fraction in the central regions of galaxies provide evidence for a dark matter fraction evolution with redshift, with the dark matter fraction declining for higher redshifts (e.g. Genzel et al., 2020; Price et al., 2020). A possible explanation for this evolution of the dark matter fraction over different redshifts is given by the IllustrisTNG (TNG) simulations (Lovell et al., 2018). The authors find that the evolution is due to the more centrally concentrated baryonic mass at higher redshift galaxies. They also show that this evolution is highly aperture dependent. Using a fixed physical aperture for all galaxies, in their case 5 kpc, leads to a dark matter fraction that is almost constant over time. Using the stellar half-mass radius instead reveals the evolution of the dark matter fraction with redshift. This is especially evident for galaxies in the $10^{11}\,\mathrm{M}_\odot$ stellar mass regime, which are highly concentrated at high redshifts. At a fixed stellar mass these galaxies show a substantial increase in size, leading to smaller half-mass radii at higher redshifts.

Observationally, various surveys provide some constraints on the dark matter fraction at different redshifts. The DiskMass survey of local galaxies finds that the central dark matter fractions are in the range of 0.5-0.9 within 2.2 times the disc scale radius, which corresponds to ~ 1.6 times the half-light radius (Martinsson et al., 2013). The SWELLS survey (Barnabè et al., 2012; Dutton et al., 2013; Courteau & Dutton, 2015) finds lower dark matter fractions in the range of 0.1 - 0.4 using the same aperture. This discrepancy is most likely due to the SWELLS galaxies having larger bulge components than the DiskMass survey. Galaxies in the redshift range z = 0.6 - 1.2 show a median of $f_{\mathrm{DM}} \sim 0.3$, while galaxies in the redshift range $z = 1.2 - 2.5$ have a median of $f_{\mathrm{DM}} \sim 0.12$ within the half-light radius (Genzel et al., 2020). For higher redshifts the dark matter fraction goes as low as $f_{\mathrm{DM}} = 0.05$ (Price et al., 2020) within the half-light radius.

In Q2131-G1 we find a dark matter fraction within the half-light radius of $f_{\mathrm{DM}} = 0.24 - 0.54$. We therefore find that the central regions of this galaxy are dominated by baryons. Compared to the surveys and simulations, Q2131-G1 fits well between the dark matter fractions found in the DiskMass survey and is also consistent with the median of galaxies observed in the redshift range 0.6-1.2. It is also within the range of galaxies observed within the SWELLS survey. The dark matter fraction is comparable to the one found in galaxies in TNG at the stellar mass $M_\star = 10^{10.5}\,\mathrm{M}_\odot$ at redshift $z = 2$. While the redshift of these galaxies in TNG is higher than of Q2131-G1, the galaxies within the SWELLS survey fit into the same regime of galaxies found in TNG. Therefore, this discrepancy could partly also be due to Q2131-G1 possibly having a significant bulge component.

2.5.7 CO Detection Rate of the MUSE-ALMA Haloes Survey

We target the CO(3–2) line of nine galaxies associated with six absorbers with ALMA and detect four of them (~ 45 per cent detection rate). All of the non-detected galaxies have metallicities below 12 + log(O/H) ~ 8.65. Four of the non-detected galaxies have sub-solar metallicities of $12 = \log(\mathrm{O/H}) < 8.32$, but have higher molecular gas mass limits than the detected galaxy Q2131-G1. Molecular gas in galaxies with sub-solar gas phase metallicity is shown to be deficient in CO, due to the CO molecule being photo-dissociated at larger fractions compared to higher metallicity galaxies (Wolfire et al., 2010; Bolatto et al., 2013). This in turn leads to a lower observed CO flux density and longer integration times are needed for observing low metallicity galaxies in CO. We therefore attribute these non-detections to the low metallicities of the galaxies.

2.6 Conclusions

In this paper, we present MUSE and new ALMA observations of the fields Q2131-1207, Q0152-2001, Q0152-2001, Q1211+1030 with LLS, sub-DLAs, and DLAs at $z \sim 0.4$ and $z \sim 0.75$. We also include the previously published field Q1130-1449 with 3 CO-detected galaxies ($z \sim 0.3$) in our analysis (Péroux et al., 2019). We detect one counterpart (Q2131-G1) of a previously detected (HST & MUSE) galaxy with ALMA observing the CO(3–2) emission line in the field Q2131-1207. We analyze the morphological, kinematical, and physical properties of Q2131-G1 with a focus on the molecular gas content. For the non-detections we provide limits on the molecular gas mass and depletion time.

The findings can be summarised as follows:

- The ionized gas phase in Q2131-G1 has a shape indicating spiral arms and possible tidal tails from previous interactions and an extent of ∼ 40 kpc. The molecular gas is found in a more compact and elliptical morphology of smaller extent (∼ 20 kpc). The extent of the stellar continuum is in between the ionized and molecular gas phases.

- Using the sophisticated 3D forward modelling tool GalPak$^{\mathrm{3D}}$ we study the kinematics of the ionized and molecular gas phase of Q2131-G1. We assume a disk model with an exponential flux profile and a tanh rotation curve for both gas phases and find that the gas phases align well directionally with similar inclinations [$i_{[\mathrm{O\,III}]} = (60.5 \pm 1.2)^{\circ}$, $i_{\mathrm{CO}} = (47^{+10}_{-1})^{\circ}$] and position angles [$PA_{[\mathrm{O\,III}]} = (65 \pm 1)^{\circ}$, $PA_{\mathrm{CO}} = (59 \pm 2)^{\circ}$]. The maximum rotational velocity is equal for both gas phases ($V_{\mathrm{max}} \sim 200\,\mathrm{km\ s^{-1}}$). This is consistent with findings by the EDGE-CALIFA survey (Levy et al., 2018), where a fraction of 25 per cent of their sample contained galaxies with equal maximum velocities. We therefore conclude that the ionized and molecular gas phases are strongly coupled within Q2131-G1.

- The absorber shows a neutral and molecular absorption two-component profile, with the weaker component blueshifted and the stronger component redshifted compared to the systemic redshift derived from the kinematic model of the CO emission. Extrapolating the model velocity maps towards the line of sight of the quasar shows that the weaker absorption component is consistent with being part of the extended rotating disk of Q2131-G1. Thanks to metallicity, geometry, and orientation arguments, we find that the stronger component is consistent with being gas falling onto Q2131-G1. The considerable amount of molecular gas traced by the absorber poses the question of the presence of a molecular cold phase in infalling gas.

- The molecular gas mass ($M_{\mathrm{mol}} = 3.52^{+3.95}_{-0.31} \times 10^{9}\ \mathrm{M}_{\odot}$) is on the low end of previously detected H I-selected galaxies. A similar counterpart, associated with the absorber at redshift $z = 0.101$ in the quasar spectrum of PKS 0429-433 (Neeleman et al., 2016), interestingly has a similar molecular mass and shows roughly the same H I column density [$\log(N_{\rm H\,I}/{\rm cm}^{-2}) \sim 19.5$] and H_2 column density [$\log(N_{\rm H_2}/{\rm cm}^{-2}) \sim 16.5$]. While the absorption was attributed to the CGM of the galaxy and not being part of the rotating disk or infalling gas, the similarities of these properties are striking. We conclude that future studies of molecular gas in both absorbers and absorber hosts are essential to studying a possible connection of these properties.

- We compute a dark matter fraction within the half-light radius of $f_{\rm DM} = 0.24 - 0.54$, showing that the inner parts of the galaxy are baryon dominated. The dark matter fraction fits between the dark matter fraction of the DiskMass survey ($f_{\rm DM} \sim 0.5-0.9$, Martinsson et al., 2013) and the median dark matter fractions observed in the redshift range $z = 0.6$-1.2 ($f_{\rm DM} \sim 0.3$, Genzel et al., 2020), providing a further indicator for a redshift evolution of the dark matter fraction.

- The depletion times (including upper limits) of our sample are in the range of ($\tau_{\rm dep} \sim 1.4 - 37$ Gyr). The depletion times of the CO-detected galaxies Q2131-G1, Q1130, and Q1130-G6 are an order of $\sim 2-53$ times larger than the median depletion time for emission-selected galaxies in the xCOLD GASS (Saintonge et al., 2017) and PHIBSS (Tacconi et al., 2018) samples. This result is consistent with previously detected H I-selected galaxies which also showed higher depletion times compared to emission-selected samples. The high depletion times are a consequence of the high molecular gas masses of H I-selected galaxies for their low SFR. We therefore conclude that H I-selected galaxies possibly preferentially select galaxies that have large molecular gas reservoirs for their low SFR, while a complete picture of the H I-selected population should be obtained by following up the non-detected galaxies for further studies of this possible selection bias.

- The 5 non-detected galaxies all have metallicities below $12 = \log({\rm O/H}) \sim 8.65$. Four of the non-detected galaxies have low sub-solar metallicities of $12 = \log({\rm O/H}) < 8.32$, but have higher molecular gas mass limits than the detected galaxy Q2131-G1. Combined with the evidence that CO is photo-dissociated at larger fractions in low metallicity galaxies compared to higher metallicity galaxies (Wolfire et al., 2010; Bolatto et al., 2013) and therefore having a lower CO flux density leads to the conclusion that one should account for a higher integration time when observing CO in sub-solar metallicity galaxies.

Chapter 3

The Column Densities of Molecular Gas across Cosmic Time: Bridging Observations and Simulations

The content of this chapter is based on the published article Szakacs et al., (2022), MNRAS, 512, 4736.

3.1 Abstract

Observations of the cosmic evolution of different gas phases across time indicate a marked increase in the molecular gas mass density towards $z \sim 2-3$. Such a transformation implies an accompanied change in the global distribution of molecular hydrogen column densities ($N_{\mathrm{H_2}}$). Using observations by PHANGS-ALMA/SDSS and simulations by GRIFFIN/IllustrisTNG we explore the evolution of this $\mathrm{H_2}$ column density distribution function [$f(N_{\mathrm{H_2}})$]. The $\mathrm{H_2}$ (and H I) column density maps for TNG50 and TNG100 are derived in post-processing and are made available through the IllustrisTNG online API. The shape and normalization of $f(N_{\mathrm{H_2}})$ of individual main-sequence star-forming galaxies are correlated with the star formation rate (SFR), stellar mass (M_*), and $\mathrm{H_2}$ mass ($M_{\mathrm{H_2}}$) in both observations and simulations. TNG100, combined with $\mathrm{H_2}$ post-processing models, broadly reproduces observations, albeit with differences in slope and normalization. Also, an analytically modelled $f(N)$, based on exponential gas disks, matches well with the simulations. The GRIFFIN simulation gives first indications that the slope of $f(N_{\mathrm{H_2}})$ might not majorly differ when including non-equilibrium chemistry in simulations. The $f(N_{\mathrm{H_2}})$ by TNG100 implies that higher molecular gas column densities are reached at $z = 3$ than at $z = 0$. Further, denser regions contribute more to the molecular mass density at $z = 3$. Finally, $\mathrm{H_2}$ starts dominating compared to H I only at column densities above $\log(N_{\mathrm{H_2}}/\mathrm{cm^{-2}}) \sim 21.8 - 22$ at both redshifts. These results imply that neutral atomic gas is an important contributor to the overall cold gas mass found in the ISM of galaxies including at densities typical for molecular clouds at $z = 0$ and $z = 3$.

3.2 Introduction

While the total amount of baryons in the Universe ($\Omega_{\mathrm{baryons}} = \rho_{\mathrm{baryons}}/\rho_{\mathrm{crit,0}} \sim 4$ per cent, where $\rho_{\mathrm{crit,0}}$ is the critical density of the Universe) is well established from measurements of anisotropies in the Cosmic Microwave Background (Planck Collaboration et al., 2016) and from primordial nucleosynthesis (Cooke et al., 2018), the contribution and evolution of different gas phases remain to be probed. Especially constraints on the evolution of the phases most closely linked to star formation, namely the neutral atomic and molecular gas phases, are limited. Nonetheless, recent observations have shown first indications of how these gas phases are evolving.

The neutral atomic gas phase shows little evolution with redshift, with its comoving baryonic mass density only slightly declining as the redshift decreases [$\rho_{\mathrm{neutral-gas}} \sim (1+z)^{0.57\pm0.04}$] (e.g. Wolfe et al., 2005; Noterdaeme et al., 2009; Crighton et al., 2015; Jones et al., 2018a; Péroux & Howk, 2020; Tacconi et al., 2020; Walter et al., 2020). This can be traced by the column density distribution function [$f(N_{\mathrm{H\,I}})$] across cosmic time, which describes the number of H I systems per unit column density per unit distance interval. $f(N_{\mathrm{H\,I}})$ quantifies the distribution of H I column densities on the sky and by integrating $f(N_{\mathrm{H\,I}})$ one can compute the comoving H I mass density. While $f(N_{\mathrm{H\,I}})$ of various shapes can result in the same $\rho_{\mathrm{neutral-gas}}$, H I-absorption in quasar spectra and emission-line measurements have revealed that the $f(N_{\mathrm{H\,I}})$ shows little to no evolution, either in shape or in normalization (e.g. Zwaan et al., 2005; Péroux et al., 2005; Zafar et al., 2013; Ho et al., 2021).

Observations calculating the comoving molecular mass density on the other hand have indicated a more radical evolution of the gas phase crucially needed for star formation. The comoving mass density of H_2 rises until cosmic noon ($z \sim 2-3$) where it peaks and drops towards $z = 0$ (e.g. Liu et al., 2019; Popping et al., 2019; Riechers et al., 2019; Péroux & Howk, 2020; Decarli et al., 2020; Tacconi et al., 2020; Walter et al., 2020). Given this evolution of the H_2 comoving mass density over cosmic time, changes in the normalization or shape of $f(N_{\mathrm{H_2}})$ can be expected.

Globally, the neutral atomic gas mass density is higher than that of the molecular phase (Péroux & Howk, 2020; Tacconi et al., 2020; Walter et al., 2020), but $f(N)$ helps reveal in which type of objects the neutral and molecular gas lies. H I-absorbers can be split into different categories from the Lyα forest for column densities $N_{\mathrm{H\,I}} \leq 1.6 \times 10^{17}\mathrm{cm}^{-2}$, to Lyman-limit systems (LLSs, $1.6 \times 10^{16} \leq N_{\mathrm{H\,I}} \leq 10^{19}\mathrm{cm}^{-2}$), to sub-damped Ly$\alpha$ absorbers (sub-DLAs, $10^{19} \leq N_{\mathrm{H\,I}} \leq 2 \times 10^{20}\mathrm{cm}^{-2}$), up to Damped Ly$\alpha$ absorbers (DLAs, $N_{\mathrm{H\,I}} \geq 2 \times 10^{20}\mathrm{cm}^{-2}$). The association between these systems and their origin is still challenging, but various works have kinematically associated LLSs, sub-DLAs, and DLAs to environments like parts of the extended rotating disks, inflows, and outflows of galaxies (e.g. Rahmani et al., 2018a,b; Zabl et al., 2020; Schroetter et al., 2019; Szakacs et al., 2021). H I emission-line studies on the other hand (e.g. Zwaan et al., 2005; Braun, 2012; French et al., 2021) can easily associate column densities with regions of galaxies like the interstellar medium (ISM) as the galaxies are completely imaged down to a given sensitivity instead of individual pencil beams. While the gas mass densities and $f(N)$

are global properties including multiple objects, comparing $f(N_{\mathrm{H\,I}})$ and $f(N_{\mathrm{H_2}})$ gives an indication in which regions of galaxies (e.g. the ISM, CGM, molecular clouds) neutral atomic or molecular gas dominates on average. The typical cold gas column densities for these regions are the following: Molecular Clouds: $N \geq 10^{20.8}\mathrm{cm}^{-2}$ (e.g. Spilker et al., 2021), ISM: $N \geq 10^{19}\mathrm{cm}^{-2}$, CGM: $N \sim 10^{14} - 10^{19}\mathrm{cm}^{-2}$ (e.g. van de Voort et al., 2019). Therefore, this helps us understand if neutral atomic gas is an important mass contributor in the ISM compared to molecular gas or if it is only substantial in the haloes surrounding galaxies.

Today's state-of-the-art cosmological simulations enable the study of physical processes of galaxy formation for both the dark matter and baryonic components of the Universe. The results of these simulations are compared to observables to learn how well the model fits. A limitation of these simulations is that due to their large volume, the scale at which these physical processes and observables can be resolved is limited so that sub-grid models are used. The advantage that these simulations offer is the large statistical sample, as thousands of galaxies are simulated. Recently, there have been considerable efforts in modelling the cold gas phase by post-processing these simulations (e.g. Lagos et al., 2015; Diemer et al., 2018; Popping et al., 2019). While properties of cold gas in these simulations show various levels of (dis)agreement with observations [e.g. a higher cosmic mass density of H I and H_2 in IllustrisTNG compared to observations at $z = 0$ (Diemer et al., 2019), tensions concerning the cosmic metal density evolution in neutral gas in EAGLE, IllustrisTNG and L-GALAXIES 2020 (Yates et al., 2021), the lower molecular mass as a function of stellar mass and number of H_2 rich galaxies in IllustrisTNG compared to the ASPECS survey (Popping et al., 2019)], other observables, like the H I column density distribution function have been accurately reproduced (Rahmati et al., 2013). Therefore, further studies and comparisons of these and similar observables, like the $f(N_{\mathrm{H_2}})$, are needed to improve the models and to design future observations.

The goal of this study is to probe the evolution of $f(N_{\mathrm{H_2}})$ across cosmic time. For this we compare data from observations on one hand and isolated and cosmological (magneto-)hydrodynamical simulations on the other hand. In the past, $f(N_{\mathrm{H_2}})$ has been studied using CO emission lines at low-z (Zwaan & Prochaska, 2006) and more recently by studying composite H_2 QSO absorption spectra at $z \sim 3$ (Balashev & Noterdaeme, 2018). High-resolution CO emission-line observations of local galaxies by the PHANGS-ALMA survey (Leroy et al., 2021) now enable us to derive $f(N_{\mathrm{H_2}})$ using emission lines from galaxies on scales of giant molecular clouds (GMCs). Further, state-of-the-art hydrodynamical simulations including non-equilibrium chemical networks tracking H_2 on-the-fly in high-resolution dwarf galaxy simulations (Hu et al., 2014b, 2016, 2017; Lahén et al., 2019, 2020a,b; Hislop et al., 2021) and post-processing the TNG100 cosmological magnetohydrodynamical simulation (Marinacci et al., 2018; Springel et al., 2018; Naiman et al., 2018; Nelson et al., 2018; Pillepich et al., 2018a) enable the bridging of observations and simulations. Finally, we aim to compare $f(N_{\mathrm{H\,I}})$ and $f(N_{\mathrm{H_2}})$ to provide indications of the regions of galaxies (e.g the ISM, CGM, molecular clouds) in which the molecular or neutral atomic gas phases dominate.

The paper is organized as follows: Section 3.3 describes the column density distribution function $f(N)$. Section 3.4 presents the observational setup as well as the simulations used for the analysis of $f(N_{\mathrm{H_2}})$. Section 3.5 describes the resolution dependence of $f(N_{\mathrm{H_2}})$. Section 3.6 presents the $f(N_{\mathrm{H_2}})$ of individual galaxies in the PHANGS-ALMA survey and their correlations with integrated physical properties of the galaxies. Section 3.7 presents the results of the key goal of this manuscript. We describe the redshift evolution of the $f(N_{\mathrm{H_2}})$ derived from both observations and simulations and study their differences and similarities across cosmic time. Further, we compare $f(N_{\mathrm{H_2}})$ with $f(N_{\mathrm{H\,I}})$ in order to explore at which densities neutral atomic gas dominates over molecular gas in and surrounding galaxies. In Section 3.8 we discuss our results from the previous sections. Finally, in Section 3.9 we give a summary of the findings. Throughout this paper we adopt an $H_0 = 67.74 \mathrm{km\ s^{-1} Mpc^{-1}}, \Omega_{\mathrm{M}} = 0.3089, \mathrm{and}\ \Omega_{\Lambda} = 0.6911$ cosmology.

3.3 Quantifying the Distribution of Column Densities Observed on the Sky

Column densities of different chemical species or different phases are not distributed uniformly on the sky as low-density gas is more frequent within our Universe. One way to quantify the distribution of column densities is the so-called column density distribution function $f(N_s)$. It is defined such that $f(N_s) dN_s dX$ is the number of systems with column densities between N_s and $N_s + dN_s$ over a distance interval dX, where s is the species one is studying (e.g. H I or H_2). While in the past $f(N_s)$ have been mostly studied using absorption systems, high-resolution data of emission lines in galaxies enable an alternative way of studying the column density distribution function. Using emission-line observations one can calculate the $f(N_s)$ as follows (e.g. Zwaan et al., 2005; Zwaan & Prochaska, 2006):

$$f(N_s) = \frac{c}{H_0} \frac{\sum_i \Phi(x_i) w(x_i) A_i(\log(N_{\mathrm{s}}))}{N_s \ln(10)\ \Delta\log(N_{\mathrm{s}})} \quad . \tag{3.1}$$

We bin the galaxies of our samples by their stellar mass, with a bin size of $\Delta\log(M_{*,i}/\mathrm{M_\odot})$ =0.2. $\Phi(M_{*,i})$ is the stellar mass function with $M_{*,i}$ being the central stellar mass value of the bin i the corresponding galaxy is in. $w(M_{*,i}) = \frac{1}{N_{\mathrm{gal,i}}}$ is a weighting function taking into account the varying number of galaxies across the range $\log(M_{*,i}/\mathrm{M_\odot})$ - $\Delta\log(M_{*,i}/\mathrm{M_\odot})/2$ to $\log(M_{*,i})$ + $\Delta\log(M_{*,i}/\mathrm{M_\odot})/2$ by calculating the reciprocal of the number of galaxies within the stellar mass bin i. $A_i(\log(N_s))$ is the area function describing the area corresponding to a column density in the range $\log(N_s)$ to $\log(N_s)$ + Δ $\log(N_s)$ for stellar mass bin i in Mpc^2. We use $\Delta\log(N_{\mathrm{H_2}}) = 0.1$ in our calculations of $f(N_{\mathrm{H_2}})$. Finally, we convert the number of systems per Mpc to that per unit redshift using c/H_0. The column densities in this work are in units of H_2 molecules $\mathrm{cm^{-2}}$ for molecular gas and H atoms $\mathrm{cm^{-2}}$ for neutral atomic gas. The space densities and areas are proper.

3.4 Bridging Observations and Simulations

In order to study the H_2 column density distribution function $[f(N_{\mathrm{H}_2})]$ at $z = 0$ and $z = 3$ we use an approach that bridges observations and simulations. We study how the state-of-the-art simulations compare to recent observations and explore if a similar evolution of the $f(N_{\mathrm{H}_2})$ can be seen in both of the approaches. On the observational side we use data from the PHANGS-ALMA survey[1] (see Section 3.4.1, Leroy et al., 2021) at $z = 0$ and data from Balashev & Noterdaeme (2018) (SDSS, see Section 3.4.2) at $z = 3$. On the simulation side we use TNG100 of the IllustrisTNG project[2] (see Section 3.4.3, Marinacci et al., 2018; Springel et al., 2018; Naiman et al., 2018; Nelson et al., 2018; Pillepich et al., 2018a) at both redshifts $z = 0$ and $z = 3$ and a high-resolution isolated dwarf galaxy simulation including a non-equilibrium chemical network from the GRIFFIN Project[3] (see Section 3.4.4 Lahén et al., 2019, 2020a,b) meant to represent a low-redshift dwarf galaxy.

3.4.1 Resolved Molecular Gas in Local Galaxies

State-of-the-art mm- and radio-telescopes like the Atacama Large Millimeter/submillimeter Array (ALMA) have enabled the astronomical community to study the coldest gas in the Universe with unprecedented spatial and spectral resolution. One of the surveys making use of these technological advances is the PHANGS-ALMA survey (Leroy et al., 2021). This survey is the first cloud-scale ($\sim$ 100 pc) survey aimed at studying the physics of molecular gas within the local galaxy population and targets galaxies that lie on or near the $z = 0$ main-sequence of star-forming galaxies with a stellar mass range of $10^9\ \mathrm{M}_\odot < \mathrm{M}_* < 10^{11}\ \mathrm{M}_\odot$. PHANGS-ALMA quantifies the physics of star formation and feedback at giant molecular cloud scales and further connects them to galaxy-scale properties and processes (Leroy et al., 2021). Further, additional state-of-the-art multi-wavelength data are provided by the PHANGS-MUSE (Emsellem et al., 2021) and PHANGS-HST surveys (Lee et al., 2021), which will study the ionized gas, stellar populations and characterize stellar clusters of the objects observed by the PHANGS-ALMA survey.

We make use of the highly resolved CO(2–1) data of the PHANGS-ALMA Survey (Leroy et al., 2021) in order to constrain the global and local $f(N_{\mathrm{H}_2})$ in the range $\log(N_{\mathrm{H}_2}/\mathrm{cm}^{-2}) \sim 19.5$ to 24 at z=0. We use a pixel-by-pixel analyzed sample consisting of 70 galaxies from Sun et al. (2020). The stellar mass distribution of the sample can be seen in Fig. 3.1. In summary, the CO(2–1) data were analyzed by Sun et al. (2020) as follows: The cubes were convolved to a common spatial resolution of 150 pc and 1 kpc. Then the data cubes were masked to only include voxels that contain emission detected with high confidence. Those cubes were finally integrated to create integrated intensity maps. The integrated maps were then used to derive the molecular gas surface density for each pixel

---

[1] sites.google.com/view/phangs/home

[2] tng-project.org

[3] mpa-garching.mpg.de/ naab/griffin-project/

as follows:[4]

$$\Sigma_{\rm mol} = \alpha_{\rm CO}\ R_{21}^{-1}\ I_{\rm CO} \quad , \tag{3.2}$$

here $R_{21} = 0.65$ is the CO(2–1)-to-CO(1–0) line ratio (Leroy et al., 2013; den Brok et al., 2021) and $\alpha_{\rm CO}$ is the metallicity-dependent CO-to-H_2 conversion factor taken as:

$$\alpha_{\rm CO} = 4.35\ Z'^{-1.6}\ {\rm M_\odot\ pc^{-2}\ (K\ km\ s^{-1})^{-1}} \quad , \tag{3.3}$$

where Z' is the local ISM metallicity in units of the solar value. The local Z' is estimated using the global stellar mass, effective radius, and the stellar mass metallicity relation by Sánchez et al. (2019) combined with a metallicity gradient (Sánchez et al., 2014). For more details see Sun et al. (2020). For the error calculation we additionally compute the surface density using the constant $\alpha_{\rm CO} = 4.3\ {\rm M_\odot\ pc^{-2}\ (K\ km\ s^{-1})^{-1}}$ of the Milky Way (Bolatto et al., 2013). Measurement uncertainties are omitted as they are negligible compared to the uncertainties of the different $\alpha_{\rm CO}$ conversion factors used.

With this sample we are able to constrain $f(N_{\rm H_2})$ at $z = 0$. We convert the derived surface densities to column densities using:

$$N_{\rm H_2} = \frac{\Sigma_{\rm H_2}}{\rm M_{H_2-molecule}} \quad , \tag{3.4}$$

with $\Sigma_{\rm H_2}$ in units of kg / cm^{-2}.

We then follow equation 3.1 to calculate $f(N_{\rm H_2})$ and use two stellar mass functions by Weigel et al. (2016) as our space density function. The first stellar mass function is that of the entire sample, and the second is one for late-type galaxies only as the PHANGS-ALMA sample mostly consists of late-type galaxies on the star-forming main-sequence [see Table 5 in Weigel et al. (2016) for the Schechter parameters].

Sun et al. (2018) estimates the 100 per cent completeness surface density limit for a sub-sample of galaxies in the PHANGS-ALMA sample to be $\log(\Sigma_{\rm H_2}/{\rm M_\odot pc^{-2}}) = 10$ - 100 at 120 pc resolution. This translates to a column density completeness limit of $\log(N_{\rm H_2}/{\rm cm^{-2}}) = 20.8$ - 21.8. At 150 pc resolution the completeness limit is expected to be lower. We therefore use a conservative estimate of $\log(N_{\rm H_2}/{\rm cm^{-2}}) = 21.6$ for 100 per cent completeness of the full PHANGS-ALMA sample.

3.4.2 Absorption Lines as a Probe for the H_2 Column Density Distribution at High Redshifts

At high redshifts it is currently challenging to observe H_2 directly or resolve CO emission lines in galaxies at spatial scales similar to the PHANGS-ALMA survey. Therefore one has to resort to another approach to study the H_2 column density distribution. H_2 imprints

---

[4] Surface density table for 150 pc can be found at canfar.net/storage/list/phangs/RELEASES/Sun_etal_2020b , datafileB1. 1 kpc table provided by authors of Pessa et al. (2021).

resonant electronic absorption bands in the UV and so studying absorption systems is a promising way of studying $f(N_{\mathrm{H}_2})$ at high redshifts. H_2 absorption lines are usually found within DLAs. It is time consuming to detect these H_2 absorbers, due to the low detection rate of $\leq$ 10 per cent. For these reasons Balashev & Noterdaeme (2018) use composite spectra of DLAs by Mas-Ribas et al. (2017), which are based on $\sim$ 27000 DLAs from SDSS (Noterdaeme et al., 2012) in order to detect the weak mean signature of H_2 at $z \sim 3$. Balashev & Noterdaeme (2018) revert to these composite H_2 spectra in order to fit a $f(N_{\mathrm{H}_2})$ in the range $\log(N_{\mathrm{H}_2}/\mathrm{cm}^{-2})$ = 18-22 on which in turn they fit the observed composite line profiles.

3.4.3 Cosmological Simulations providing Large Statistical Samples

Cosmological simulations provide large statistical samples for studies of galaxy evolution. One of these simulations is TNG100 of the IllustrisTNG project (Marinacci et al., 2018; Springel et al., 2018; Naiman et al., 2018; Nelson et al., 2018; Pillepich et al., 2018a). TNG100 is a state-of-the-art gravomagnetohydrodynamics (MHD) cosmological simulation including a comprehensive model for galaxy formation physics (Weinberger et al., 2017; Pillepich et al., 2018a) within a 75000 ckpc/h sized box using the AREPO code (Springel, 2010). IllustrisTNG aims to study the physical processes that drive galaxy formation and to study how galaxies evolve within large-scale structures.

We aim to exploit the large sample size of the TNG100 simulation in order to compare the observed $f(N_{\mathrm{H}_2})$ at $z = 0$ and $z = 3$. While TNG50 offers higher resolution, we choose TNG100 as our fiducial model due to two reasons: 1) The SMF, which is an important parameter in our calculations, is closer to observations for TNG100 than for TNG50. 2) To enable future comparisons with the EAGLE cosmological simulation (Schaye et al., 2015), as TNG100 is the closest in terms of resolution to the EAGLE 100 Mpc box simulation.

The molecular gas phase in current large-scale cosmological simulations is challenging to assess. Using chemical networks to track H_2 on-the-fly is computationally time consuming due to the complex physics involved and the high resolution needed in order for the H_2 mass fraction to converge within the forming molecular clouds ($\sim$ 0.12 pc, Seifried et al., 2017). In order to capture the unresolved physics, one has to revert to sub-grid models, which split the cold hydrogen component in the simulations into a neutral atomic and molecular component. We use the H_2 post-processing catalogs of Popping et al. (2019) for TNG100, for which three different models are available. The used models are by Blitz & Rosolowsky (2006); Gnedin & Kravtsov (2011); Krumholz (2013). The model by Blitz & Rosolowsky (2006) is a pressure-based empirical fit based on a sample of 14 local spiral and dwarf galaxies that have measured atomic, molecular, and stellar surface densities. Using this sample they find a nearly linear relation between the hydrostatic pressure and the ratio of molecular to atomic gas. Gnedin & Kravtsov (2011) designed a phenomenological model for the formation of molecular hydrogen, which is dependent on the gas density, dust-to-gas ratio and the far-UV radiation flux. This model was tested on cosmological simulations by Gnedin et al. (2009). Finally, the model by Krumholz (2013) is a column density, metallicity, and radiation field dependent relation for splitting the cold hydrogen component in simulations.

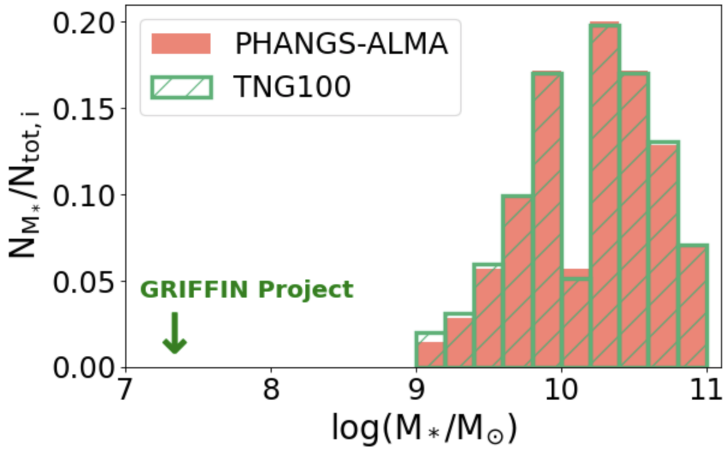

Figure 3.1: The stellar mass distribution of the PHANGS-ALMA survey (red) and the matching TNG100 sample (green-hatched). For each stellar mass bin we additionally only select galaxies in TNG100 with similar SFRs as found in the corresponding stellar mass bin in PHANGS-ALMA.

In our study, we select central galaxies at $z = 0$ and $z = 3$. We add the ability to generate post-processed H_2 column density (N_{H_2}) maps based on Popping et al. (2019) for TNG50 and TNG100 using the sub-grid models mentioned above, which split the cold hydrogen within those galaxies into atomic and molecular components, to the IllustrisTNG online API[5]. We use this functionality to create H_2 column density (N_{H_2}) maps at 150 pc and 1 kpc resolution. These maps are generated by projecting gas cells as adaptively sized SPH kernels. The kernel size parameter is set to $h_{\rm kernel} = 2.5 r_{\rm cell}$. With $r_{\rm cell}$ being the cell size determined by using the Voronoi cell volume: $r_{\rm cell} = (3V_{\rm cell}/4\pi)^{1/3}$. We use the same projection direction for every galaxy (z-axis of the simulation) and only consider gas cells gravitationally bound to the selected subhaloes within a fixed 200 × 200 kpc box. This method reproduces $f(N_{H_2})$ derived from the full box of TNG100 when using the same resolution for the derivation as in Klitsch et al. (2019b). Therefore, we do not expect this choice to affect $f(N_{H_2})$ as compared to using a box size dependent on halo properties. At $z = 0$ we selected a PHANGS-ALMA survey-like sample within TNG100. For this we select ~ 570 galaxies within a stellar mass of 10^9 and 10^{11} $M_\odot$. We match the PHANGS-ALMA sample stellar mass distribution (see Fig. 3.1). Further, for each 0.2 dex stellar mass bin we select galaxies with similar star formation rates as in the PHANGS-ALMA sample. We note that when selecting a sample of ~ 700 central galaxies with stellar masses between 10^9 and $10^{12.6}$ $M_\odot$ and no star-formation selection criterion we derive very similar

---

[5] tng-project.org/api/

results in TNG100 as for the PHANGS-ALMA-like sample, with the only differences being the slightly higher column densities reached in the larger sample (~ 0.2 dex higher) and a slightly higher normalization at column densities above $\log(N_{\mathrm{H_2}}/\mathrm{cm}^{-2}) \sim 21.5$, which is likely due to the larger size of the additional galaxies. At $z = 3$ we select ~ 550 galaxies with stellar masses between 10^9 and $10^{11.8}$ $\mathrm{M}_\odot$. Therefore this includes galaxies between the resolution limit of TNG100 up to the highest stellar mass limit of TNG100. We do not set any constraints on the SFR of the galaxies, as the observational $f(N_{\mathrm{H_2}})$ is based on $\mathrm{H_2}$ absorption line studies, where we do not have any SFR information.

Following equation 3.1 we use the derived $\mathrm{H_2}$ column density maps to calculate $f(N_{\mathrm{H_2}})$. For the space density function we use the Stellar Mass Function (SMF) of the simulation box itself. At $z = 0$ the Schechter parameters are: $\log(M^*/\mathrm{M}_\odot) = 11.27$, $\log(\Phi_1^*/h^3\ \mathrm{Mpc}^{-3})$ = -3.31, $\log(\Phi_2^*/h^3\ \mathrm{Mpc}^{-3})$ = -3.28, α_1 = -1.36, α_2 = -1.36. At $z = 3$ the double Schechter parameters are: $\log(M^*/\mathrm{M}_\odot) = 10.83$, $\log(\Phi_1^*/h^3\ \mathrm{Mpc}^{-3})$ = -3.84, $\log(\Phi_2^*/h^3\ \mathrm{Mpc}^{-3})$ = -3.58, α_1 = -0.29, α_2 = -1.64.

3.4.4 Molecular Gas in Highly Resolved Simulations of Individual Galaxies

An alternative approach to studying molecular gas in simulations is to use highly resolved simulations of individual isolated galaxies, which include non-equilibrium chemical networks that track $\mathrm{H_2}$ on the fly throughout the simulation. Although this is currently mostly limited to dwarf galaxies, the advantage of these simulations is a more accurate representation of $\mathrm{H_2}$ due to a non-equilibrium chemical network.

One of these simulations is the high-resolution isolated dwarf simulation from the GRIFFIN Project (Lahén et al., 2019, 2020a,b) with a stellar mass of $\log(M_*/\mathrm{M}_\odot) \sim 7.3$. The simulation is based on the smoothed particle hydrodynamics tree code GADGET-3 (Springel, 2005) with the gas dynamics modelled using the SPH implementation SPHGal (Hu et al., 2014b, 2016, 2017). The simulation resolves individual massive stars at subparsec resolutions and includes a non-equilibrium chemical network based on Nelson & Langer (1997); Glover & Mac Low (2007a); Glover & Clark (2012). The chemical network follows the abundances of six chemical species for cooling processes at low temperatures ($< 3 \times 10^3$ K, most importantly $\mathrm{H_2}$). Further, the simulation includes star formation, an interstellar radiation field, and stellar feedback prescriptions. A detailed discussion of the isolated dwarf simulation is given in Hu et al. (2016, 2017).

For the calculation of $f(N_{\mathrm{H_2}})$ we time- and inclination-average the isolated dwarf galaxy simulation. Therefore we produce $\mathrm{H_2}$ column density maps with all possible lines of sight and slightly varying total $\mathrm{H_2}$ masses using the analysis tool PYGAD (Röttgers et al., 2020). First we create $\mathrm{H_2}$ column density maps by using snapshots over a time range of ~ 300 Myrs. For each of these snapshots we create $\mathrm{H_2}$ column density maps at a resolution of 150 pc with inclinations between 0 and 90° in $\Delta\cos(i) = 0.05$ steps. We then follow equation 3.1 to calculate $f(N_{\mathrm{H_2}})$ by using these $\mathrm{H_2}$ column density maps and use the Weigel et al. (2016) SMF of the entire sample for the normalization of $f(N_{\mathrm{H_2}})$ following the prescription described in Section 3.3 (for the Schechter parameters see Table 5 in Weigel et al., 2016). The $f(N_{\mathrm{H}_2})$ is therefore calculated using a single stellar mass bin (as the stellar mass of the simulated dwarf galaxy does not evolve much over time). However, galaxies of this stellar mass are not represented in the PHANGS-ALMA and TNG100 sample, so we can not directly compare the $f(N_{\mathrm{H}_2})$ of similar galaxies.

3.5 A Resolution-dependent H_2 Column Density Distribution Function

We test how $f(N_{\mathrm{H}_2})$ depends on the resolution of the data used for its calculation. First we study how $f(N_{\mathrm{H}_2})$ depends on the resolution of the simulation by comparing $f(N_{\mathrm{H}_2})$ in TNG50 and TNG100 from the Illustris project. Then we compare how the resolution of the N_{H_2}-maps from both observations and simulations affects $f(N_{\mathrm{H}_2})$.

3.5.1 $f(N_{\mathrm{H}_2})$ - Dependence on the Resolution of Simulations

Here we compare the $f(N_{\mathrm{H}_2})$ derived from TNG100 with TNG50 (Pillepich et al., 2019; Nelson et al., 2019b) using the Gnedin & Kravtsov (2011) H_2 model. TNG50 has a box length of 51.7 Mpc and 2×2160^3 resolution elements, while TNG100 has a box length of 110.7 Mpc and 2×1820^3 resolution elements. Therefore TNG50 gives us an indication of how a higher resolution simulation affects $f(N_{\mathrm{H}_2})$.

In Fig. 3.2 (left) the $f(N_{\mathrm{H}_2})$ derived from TNG100 and TNG50 at $z = 0$ using a 150 pc resolution of the post-processed column density map are displayed. TNG50 extends to higher column densities compared to TNG100. The finer resolution reaches higher gas densities and in turn higher column densities. Further, at column densities above $\log(N_{\mathrm{H}_2}/\mathrm{cm}^{-2}) \sim 22$ the $f(N_{\mathrm{H}_2})$ in TNG50 initially displays a steep drop with a subsequent flattening of the $f(N_{\mathrm{H}_2})$. These differences indicate that the H_2 column densities are not converged in this region. Given these differences we would expect higher resolution simulations to reach even higher column densities, and possibly also affect the shape in the region above $\log(N_{\mathrm{H}_2}/\mathrm{cm}^{-2}) \sim 22$. We note that IllustrisTNG uses the sub-grid model of Springel & Hernquist (2003) for the star-forming ISM. Independent of the resolution, the sub-grid model begins star formation at ISM densities of 0.1 cm^{-3} preventing the simulation from resolving the cold gas phase and subsequently the formation of molecular clouds. Due to this the model itself is limited by the sub-grid ISM model and a higher resolution is only sensitive up to the limitations of the model. The resolution tests, however, indicate that the sub-grid model is not the limiting factor in terms of densities reached at the resolution of TNG100 since $f(N_{\mathrm{H}_2})$ is not converged at high column densities. However, modifying the model to treat the multiphase ISM more realistically will likely affect the results.

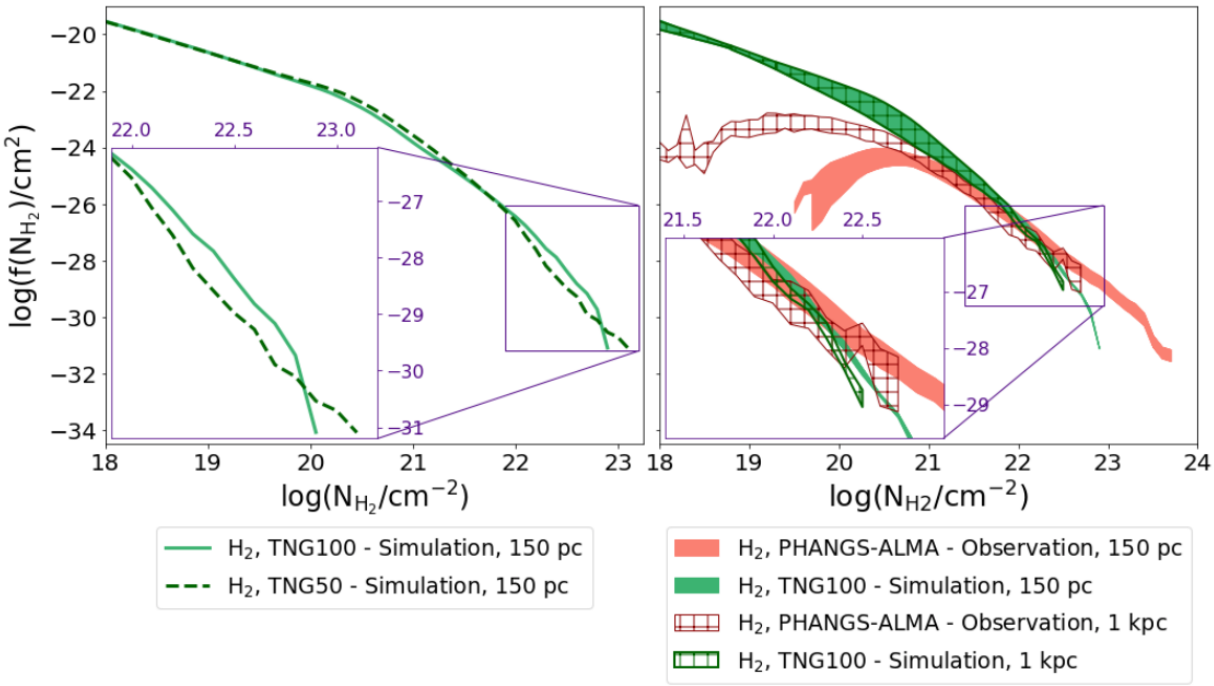

Figure 3.2: Resolution study for the molecular gas column density distribution function derived from simulations and observations. **Left:** $f(N_{\mathrm{H_2}})$ derived from TNG100 and TNG50. Higher column densities are reached in TNG50, as a higher resolution enables the simulation to reach higher gas densities. Especially at column densities above $\log(N_{\mathrm{H_2}}/\mathrm{cm}^{-2}) \sim 22$ the $f(N_{\mathrm{H_2}})$ differs. This indicates that H_2 in TNG100 is not converged for those column densities. **Right:** $f(N_{\mathrm{H_2}})$ derived from PHANGS-ALMA (red bands) and TNG100 (green bands) data at map resolutions of 150 pc (filled) and 1 kpc (hatched). The $f(N_{\mathrm{H_2}})$ both in observations and simulations show a map resolution dependence. In TNG100 this effect arises due to the averaging of column densities over a larger area. Very high column densities are usually detected on small scales (much smaller than 1 kpc), leading to a dilution of high column densities. In PHANGS-ALMA this effect is additionally combined with sensitivity and incompleteness specifics of the observations. The core of the distribution $[\log(N_{\mathrm{H_2}}/\mathrm{cm}^{-2}) = 21$ to $22]$ is robust to resolution effects in both simulations and observations. In general, high-resolution observations and simulations are needed to resolve column densities typically found in very dense environments like molecular clouds.

3.5.2 $f(N_{\mathrm{H_2}})$ - Dependence on the Resolution of $N_{\mathrm{H_2}}$ Maps

We compare how $f(N_{\mathrm{H_2}})$ depends on the resolution of observed and simulated H_2 column density ($N_{\mathrm{H_2}}$) maps. We calculate $f(N_{\mathrm{H_2}})$ using 150 pc and 1 kpc map resolution CO(2–1) data from the PHANGS-ALMA survey and a sample of galaxies from TNG100.

The $f(N_{\mathrm{H_2}})$ for these data sets and resolutions are displayed in Fig. 3.2 (right). Differences in both shape and the column density range are found between the different map resolutions in both TNG100 and the PHANGS-ALMA survey. The $f(N_{\mathrm{H}_2})$ derived from PHANGS-ALMA shows a more substantial map resolution dependence compared to TNG100. There are two factors that together cause this higher map resolution dependence of the PHANGS-ALMA data. 1.) Creating N_{H_2} maps using larger pixel sizes averages the column densities over larger regions. This leads to lower mean observed column densities, as very high column densities are usually detected at GMC scales, which are smaller than 1kpc (Leroy et al., 2021). This effect is especially apparent at the high column density end, as in the 1 kpc map resolution data column densities above $\log(N_{\mathrm{H}_2}/\mathrm{cm}^{-2}) \sim 22.8$ are diluted by this effect. 2.) Observational data are limited by their sensitivity and completeness. Coarser resolution data have a reduced noise and are thus more complete (see the comparison of native resolution vs. 150 pc map resolution data in PHANGS-ALMA, Leroy et al., 2021). Therefore the coarser map resolution observations are sensitive to lower column densities compared to finer resolution observations. This effect is especially significant below $\log(N_{\mathrm{H}_2}/\mathrm{cm}^{-2}) \sim 22.5$ in Fig. 3.2 (right). TNG100 does not suffer from these sensitivity and incompleteness effects at lower column densities and therefore the map resolution dependence is less drastic. TNG100 is only affected by beam smearing. The core of the distribution [$\log(N_{\mathrm{H}_2}/\mathrm{cm}^{-2}) = 21$ to 22] is robust to resolution effects in both simulations and observations.

3.6 Does the H_2 Column Density Distribution of Individual Galaxies depend on their Physical Properties?

We study the $f(N_{\mathrm{H}_2})$ of individual objects in the PHANGS-ALMA survey and TNG100 ($z = 0$) in order to explore how it depends on integrated physical properties of the galaxies. We calculate the individual $f(N_{\mathrm{H}_2})$ using Equation 3.1, but set the normalization parameters [$\Phi(x_i)$ and $w(x_i)$] equal to one. The individual column density distributions, colour-coded according to the integrated star formation rate (SFR), stellar mass (M_*) and H_2 mass (M_{H_2})[6] of the corresponding galaxy are displayed in Fig. 3.3. For these calculations we use column density maps with a resolution of 150 pc. The colour coding of the plots reveals a connection between $f(N_{\mathrm{H}_2})$ and the physical parameters mentioned. We note that the integrated molecular masses of galaxies in TNG100 are generally higher compared to PHANGS-ALMA. This is to be expected, as TNG100 probes the full disk and is not limited by observational sensitivity and incompleteness limits when compared to the PHANGS-ALMA sample. Further, Leroy et al. (2021) estimates that on average ∼ 30 per cent of molecular gas is missed by PHANGS-ALMA due to the limited field of view when compared to WISE3 luminosities. Additionally, TNG100 is possibly overestimating H_2 within the simulation at $z = 0$ (Diemer et al., 2019).

---

[6] SFR and M_* are taken from sites.google.com/view/phangs/sample. M_{H_2} is calculated by summing up the surface density of individual pixels multiplied by pixel area (table found in datafileB1 at canfar.net/storage/list/phangs/RELEASES/Sun_etal_2020b)

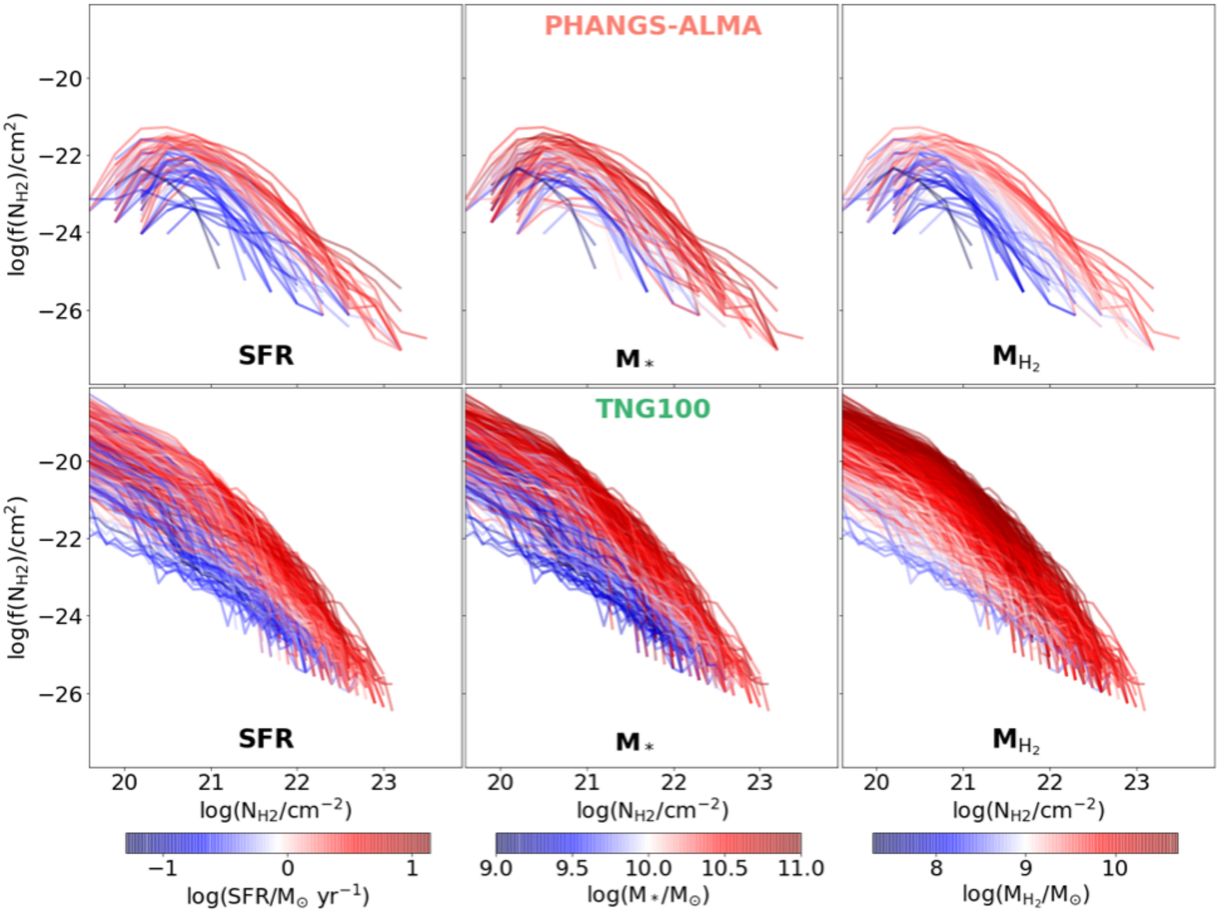

Figure 3.3: H_2 column density distributions [$f(N_{H_2})$] of individual PHANGS-ALMA and TNG100 ($z = 0$) galaxies. The plots display the correlation of the individual $f(N_{H_2})$ on the integrated SFR, M_* and M_{H_2} of the galaxies. The flattening and steep drop of the PHANGS-ALMA $f(N_{H_2})$ at lower column densities is due to the incompleteness and sensitivity limit of the observations. The individual $f(N_{H_2})$ in both PHANGS-ALMA and TNG100 have similar shapes. This is likely due to the sample consisting of main-sequence star-forming galaxies. These galaxies mostly have rotating disks and are hypothesized to have radially exponential gas profiles. The $f(N_{H_2})$ correlate with integrated physical parameters (SFR, M_* and M_{H_2}) of the galaxies. The higher these parameters, the larger the galaxies, leading to a higher normalization of $f(N_{H_2})$. Further, higher column densities are detected in more massive galaxies, implying that more dense gas is found in larger galaxies.

The $f(N_{H_2})$ of individual galaxies have very similar shapes in both observation and simulation. This is possibly related to the galaxies in the sample, which are main-sequence star-forming galaxies. These types of galaxies mostly have rotating disks and are hypothesized to have radially exponential gas profiles (Leroy et al., 2008; Stevens et al., 2019).

While the diskiness and exponential gas profiles of the galaxies under consideration still need to be established, the similar $f(N_{\mathrm{H}_2})$ could indeed stem from similar gas profiles within these galaxies. In Section 3.7 we further explore this possibility by comparing an analytical $f(N_{\mathrm{H}_2})$ model assuming radially exponential gas disks with simulated results.

While the shapes of the $f(N_{\mathrm{H}_2})$ are similar, the $f(N_{\mathrm{H}_2})$ also show a correlation with integrated physical parameters of the galaxies. The colour coding in Fig. 3.3 indicates that the $f(N_{\mathrm{H}_2})$ are correlated with the integrated SFR, M_*, and M_{H_2} of the galaxies. The higher the SFR, M_* and M_{H_2}, the more massive these galaxies are, leading to a higher normalization of $f(N_{\mathrm{H}_2})$. Further, higher column densities are detected in more massive galaxies, implying that more dense gas is formed in larger galaxies. The higher abundance of denser gas in more massive galaxies could also lead to higher star formation rates, as more gas is found at densities suitable for star formation (e.g. above $\log(N_{\mathrm{H}_2}/\mathrm{cm}^{-2}) \sim 21$, Clark & Glover, 2014). This is possibly related to the correlation between the SFR surface density and molecular gas surface densities in galaxies (e.g. Bigiel et al., 2008; Feldmann, 2020). We however note that the correlation we find is related to the integrated SFR of the galaxy and not the SFR surface density. In the Appendix (Section B.1) we explore these correlations using the PHANGS-ALMA sample and provide a way to approximate $f(N_{\mathrm{H}_2})$ given physical parameters. It, however, remains unclear which galaxy property is the governing parameter for the shape of $f(N_{\mathrm{H}_2})$, as SFR, M_* and $\mathrm{M}_{\mathrm{H}_2}$ all have similar correlation strengths with parameters of the gamma distribution used to fit $f(N_{\mathrm{H}_2})$ in the Appendix.

3.7 The Redshift Evolution of the H_2 Column Density Distribution in Simulations and Observations

We study $f(N_{\mathrm{H}_2})$ at $z = 0$ and $z = 3$ using both observations and simulations. First we study how recent observations compare to the state-of-the-art simulation TNG100 at both $z = 0$ and $z = 3$ and how the isolated dwarf galaxy simulation from the GRIFFIN Project fits into the column density distribution at $z = 0$. Then we discuss the evolution of $f(N_{\mathrm{H}_2})$ from $z = 0$ to $z = 3$. Finally, we examine how $f(N_{\mathrm{H}_2})$ compares to $f(N_{\mathrm{H\,I}})$ at $z = 0$ and $z = 3$ to explore in which regions of galaxies the neutral atomic gas is dominating over the molecular gas.

3.7.1 $f(N_{\mathrm{H}_2})$ at z=0

In Fig. 3.4 (left) the $f(N_{\mathrm{H}_2})$ from both observations, simulations and an analytical model at $z = 0$ are displayed. For TNG100 we plot a band (green band) encompassing the three post-processing methods described in Section 3.4.3. Note that below $\log(N_{\mathrm{H}_2}/\mathrm{cm}^{-2}) < 18$ the post-processing results for H_2 become unreliable as post-processing the simulations with different SPH kernel smoothing lengths leads to highly different results in that region. This region is represented by bands filled with the $\star$ symbol. The dwarf galaxy simulation $f(N_{\mathrm{H}_2})$ (green line) is based on the results from the on-the-fly chemical network included in

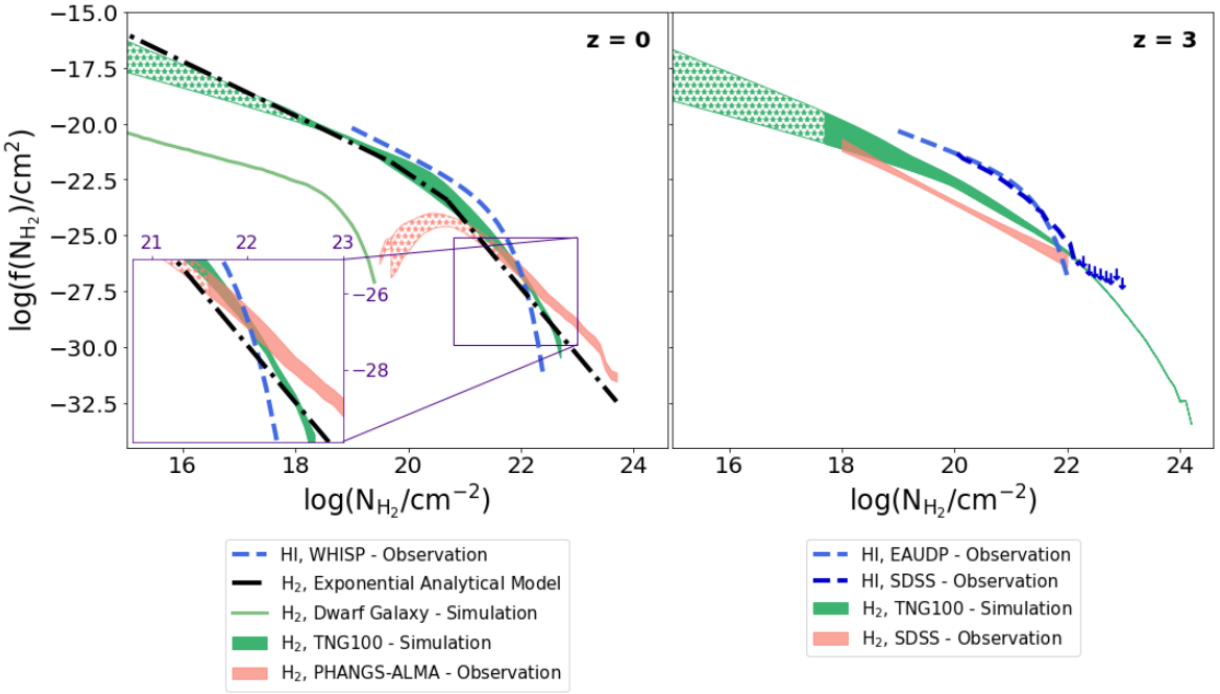

Figure 3.4: $f(N_{\mathrm{H\,I}})$ and $f(N_{\mathrm{H_2}})$ derived from both simulations and observations at $z = 0$ and $z = 3$. The column densities at which the $f(N_{\mathrm{H_2}})$ become unreliable due to incompleteness or simulation specifics are indicated by regions filled with the $\star$ symbol. **Left:** $f(N_s)$ of TNG100, the PHANGS-ALMA survey (red band), the WHISP sample (blue, Zwaan et al., 2005), an analytical model (black line, Zwaan, 2000) and the simulated dwarf galaxy from the GRIFFIN project (green line) at $z = 0$. The $f(N_{\mathrm{H_2}})$ by TNG100 broadly reproduces the observations by the PHANGS-ALMA survey. The analytical $f(N_{\mathrm{H_2}})$ based on the assumption of a radially exponential gas profile in galaxies is a good approximation for $f(N_{\mathrm{H_2}})$ for both observations and simulations. The $f(N_{\mathrm{H_2}})$ derived from a simulated dwarf galaxy including a non-equilibrium chemical network displays similar slopes compared to TNG100. **Right:** $f(N_s)$ of TNG100 (green band), SDSS [(red band, Balashev & Noterdaeme, 2018), (dark blue line, Ho et al., 2021)] and the EAUDP sample (dark blue line, Zafar et al., 2013) at $z = 3$. The $f(N_{\mathrm{H_2}})$ from TNG100 and the observational results based on the SDSS sample have matching slopes. The normalization between the two shows a ~ 1 dex difference, possibly arising due to differences in selection and environments probed.

the simulation. The red band encompasses the $f(N_{\mathrm{H_2}})$ from the PHANGS-ALMA survey using varying assumptions. It includes calculations using a stellar mass function based on the full sample and late-type galaxy only sample in Weigel et al. (2016). Further, we calculate the $f(N_{\mathrm{H_2}})$ with both a metallicity dependent α_{CO} (see Section 3.4.1) and a constant $\alpha_{\mathrm{CO}} = 4.3\ \mathrm{M_\odot}\ (\mathrm{K\ km\ s^{-1}\ pc^2})^{-1}$ (Bolatto et al., 2013). We note that the drop of the PHANGS $f(N_{\mathrm{H_2}})$ at column densities below $\log(N_{\mathrm{H_2}}/\mathrm{cm^{-2}}) \sim 21$ is not of physical origin, but due to the sensitivity and incompleteness limit of the observations, leading to the observations not probing the full disk. We also include an analytical model (black line) used to estimate $f(N)$ assuming radially exponential gas disks averaged over all possible inclinations (Zwaan, 2000) [7]. The analytical model is approximated by three linear functions.

TNG100 broadly reproduces Observations

We study how TNG100 (green band, Fig. 3.4 left) reproduces the $f(N_{\mathrm{H_2}})$ observed by the PHANGS-ALMA survey (red band, Fig. 3.4 left). While there are differences in the $f(N_{\mathrm{H_2}})$, the observations are broadly reproduced by TNG100 in the column density ranges where simulation and observation specifics do not hinder a fair comparison.

One difference is that TNG100 does not reach as high a column density as observed by the PHANGS-ALMA survey. This is due to the limitations of the simulation. Given the resolution of TNG100 gas densities that can be reached at given redshifts are limited (see Section 3.5.1). Further, for the regions below $\log(N_{\mathrm{H_2}}/\mathrm{cm^{-2}}) \sim 21.6$, TNG100 $f(N_{\mathrm{H_2}})$ shows a higher normalization than observed with PHANGS-ALMA. This can be explained by the sensitivity and incompleteness limit of the observations below those column densities. In the region between $\log(N_{\mathrm{H_2}}/\mathrm{cm^{-2}}) \sim 21.6 - 22.2$ both $f(N_{\mathrm{H_2}})$ are overlapping, albeit with TNG having a steeper slope when approximated as a linear function in log space ($\beta_{\mathrm{PHANGS}} \sim 2.3 - 2.35$, $\beta_{\mathrm{TNG}} \sim 3.15 - 3.6$).

Given that the two $f(N_{\mathrm{H_2}})$ are based on vastly different methods of calculating the column densities (one being post-processed $\mathrm{H_2}$ from a cosmological magnetohydrodynamical simulation, and one being observations of CO(2–1), which are converted to $\mathrm{H_2}$) the similarity between the two $f(N_{\mathrm{H_2}})$ is remarkable. Nonetheless, further tests and studies are needed to explore the inconsistencies between observations and simulations, especially at the high column density end of $f(N_{\mathrm{H_2}})$. Higher resolution simulations would likely extend the $f(N_{\mathrm{H_2}})$ to higher column densities. Additionally, alternatives to the Springel & Hernquist (2003) sub-grid star formation prescription in order to resolve the cold gas phase in the simulations might be needed to reach the high column densities detected in observations. Including non-equilibrium chemistry in these simulations could also give more accurate representations of $\mathrm{H_2}$ in the simulations (e.g. Maio et al., 2022). Finally, deeper observations would enable fair comparisons of $f(N_{\mathrm{H_2}})$ below $\log(N_{\mathrm{H_2}}/\mathrm{cm^{-2}}) \sim 21.6$.

An Analytical Model closely matching TNG100

We compare the analytical model by Zwaan (2000) (black line, Fig. 3.4 left) with the results from TNG100 (green band, Fig. 3.4 left) at $z = 0$ to study how well the simulated predictions match the analytical model. We approximate the analytical model using three ad-hoc linear functions in the following $\mathrm{H_2}$ column density ranges: $\log(N_{\mathrm{H_2}}/\mathrm{cm^{-2}}) \leq 20$, $\log(N_{\mathrm{H_2}}/\mathrm{cm^{-2}}) = 20 - 21$, $\log(N_{\mathrm{H_2}}/\mathrm{cm^{-2}}) \geq 21$ and therefore use these regions for a

---

[7] With the N_0 parameter of the model, which determines the knee of the curve, set to $10^{20.7}$ cm^{-2}. For $\mathrm{H_2}$ this is an ad-hoc choice.

comparison. Generally, the analytical model assuming radially exponential gas disks in galaxies produces comparable results to calculating $f(N_{\mathrm{H}_2})$ using the post-processed H_2 column densities of galaxies within TNG100.

Below $\log(N_{\mathrm{H}_2}/\mathrm{cm}^{-2}) \leq 20$ the normalizations are comparable, but the TNG100 $f(N_{\mathrm{H}_2})$, depending on the post-processing prescription used, has a slightly lower slope compared to the analytical model ($\beta_{\mathrm{TNG}} \sim -0.75$ to -1.15, $\beta_{\mathrm{ana}} \sim -1.22$). In the range of $\log(N_{\mathrm{H}_2}/\mathrm{cm}^{-2}) = 20 - 21$ we find similar results, with the normalization matching, but a slightly higher slope in TNG100 ($\beta_{\mathrm{TNG}} \sim -1.94$ to -2.05, $\beta_{\mathrm{ana}} \sim -1.71$). Further, approximating the slope of TNG100 at $\log(N_{\mathrm{H}_2}/\mathrm{cm}^{-2}) \geq 21$ using a linear function in log space leads to similar results, with TNG100 producing higher slopes for $f(N_{\mathrm{H}_2})$ in this region ($\beta_{\mathrm{TNG}} \sim -3.4$ to -3.8, $\beta_{\mathrm{ana}} \sim -3.0$). Finally, the analytical model predicts slightly fewer systems in this column density range compared to TNG100. We note that Zwaan (2000) also proposes an analytical model based on Gaussian gas profiles. This model results in slopes of $\beta_{\mathrm{ana,gauss}} \sim -1$ at column densities below $\log(N_{\mathrm{H}_2}/\mathrm{cm}^{-2}) = 20.7$ and $\beta_{\mathrm{ana,gauss}} \sim -3$ at column densities above this threshold.

We conclude that while TNG100 produces $f(N_{\mathrm{H}_2})$ with slightly higher slopes and in some parts different normalizations, the $f(N_{\mathrm{H}_2})$ derived from the analytical model is still comparable and a good approximation. Since radially exponential gas disks are also a good approximation for disk galaxies in TNG100 (e.g. H I disks described in Stevens et al., 2019) it appears natural that an analytical model making the assumption of radially exponential gas disks yields similar results. While the results of an analytical model using exponential gas disks match predictions by TNG100 well, a Gaussian distribution within gas disks of galaxies yields similar results. Therefore, further studies of the distribution in gas disks and their relevance to $f(N_{\mathrm{H}_2})$ are required to fully understand how the gas disk distribution and $f(N_{\mathrm{H}_2})$ relate.

A Dwarf Galaxy Simulation producing similar Slopes compared to TNG100

We compare the $f(N_{\mathrm{H}_2})$ of the simulated dwarf galaxy from the GRIFFIN project, which includes a non-equilibrium chemical network tracking H_2 on the fly (green line, Fig. 3.4 left) with the results of TNG100 (green band, Fig. 3.4 left) at $z = 0$. This helps us understand the impact for $f(N_{\mathrm{H}_2})$ when running simulations at sub-pc resolution including a non-equilibrium chemical network in an isolated environment.

The $f(N_{\mathrm{H}_2})$ only probes one galaxy with a stellar mass of $\log(M_*/\mathrm{M}_\odot) \sim 7.3$[8]. This leads to several differences when compared to a sample of galaxies. Due to the limited mass and size, the dwarf galaxy in the simulation only reaches column densities up to $\log(N_{\mathrm{H}_2}/\mathrm{cm}^{-2}) \sim 19.5$. The slope of both $f(N_{\mathrm{H}_2})$ is consistent. For the dwarf simulation the logarithmic slope is $\beta_{\mathrm{dwarf}} \sim -0.7$ before the drop off at $\log(N_{\mathrm{H}_2}/\mathrm{cm}^{-2}) \sim 18$. The slope found in TNG100 at those column densities is $\beta_{\mathrm{TNG}} \sim -0.7$ to -1.1. It is surprising that the slope of $f(N_{\mathrm{H}_2})$ of a single galaxy is so similar to the slope of a large sample of galaxies with varying sizes, especially given the different methods for deriving molecular

---

[8] The stellar mass does not evolve much over the course of the simulation

gas in these simulations. While it is difficult to disentangle the effects that the different galaxy properties and derivation methods of molecular gas have on the slope of $f(N_{\mathrm{H_2}})$, this is possibly a first indication that the slope $f(N_{\mathrm{H_2}})$ is not affected by non-equilibrium chemistry. Especially since the slope of $f(N_{\mathrm{H_2}})$ for individual main-sequence star-forming galaxies in TNG100 is similar below $\log(N_{\mathrm{H_2}}/\mathrm{cm}^{-2}) \lesssim 20$ and not majorly affected by galaxy properties. In order to further our understanding of how non-equilibrium chemistry might affect $f(N_{\mathrm{H_2}})$ a larger sample size of highly resolved simulated galaxies spanning a wider range of stellar masses would be needed. Alternatively, running and comparing the dwarf galaxy simulation by GRIFFIN without non-equilibrium chemistry with the current GRIFFIN model would also help disentangling the effects that non-equilibrium chemistry and galaxy properties have on the slope of $f(N_{\mathrm{H_2}})$. This is an interesting avenue to explore in the future.

Which Column Densities contribute most to the H_2 Mass Density at $z = 0$?

As a final analysis of $f(N_{\mathrm{H_2}})$ at $z = 0$ we study which column densities contribute the most to the overall mass density (ρ_{mol}) in both TNG100 and the PHANGS-ALMA survey. Disentangling which column densities contribute the most to the mass density helps us understand in which regions of galaxies (e.g. the ISM, CGM, molecular clouds) most of the molecular gas is detected. Further, we can interpret if most of the gas is in regions suitable for star formation or not.

In Fig. 3.5 (left panel) we plot the mass densities as a function of H_2 column density. The red band corresponds to the PHANGS-ALMA results and the green band to the TNG100 results. For TNG100 the highest mass density contribution stems from column densities in the range $\log(N_{\mathrm{H_2}}/\mathrm{cm}^{-2}) \sim 20.5 - 20.7$. Therefore the majority of molecular gas in TNG100 is found at column densities typical for the ISM of galaxies, but below densities of molecular clouds (e.g. Spilker et al., 2021) as opposed to less dense and diffuse regions surrounding galaxies, like the CGM.

Using numerical models Clark & Glover (2014) predict that star formation is possible in regions where the mean area averaged column density exceeds $\log(N_{\mathrm{H_2}}/\mathrm{cm}^{-2}) \sim 21$. TNG100 predicts the mass density peak slightly below the star formation threshold advocated by Clark & Glover (2014) and therefore in a region not suitable for star formation. This fraction of the gas could be either in regions where the molecular gas has been depleted due to star formation or in regions that are possibly in the process of collapsing into denser regions.

In PHANGS-ALMA we find an overall flatter distribution of the H_2 mass densities in the regions where the observations are complete and when compared to TNG100. The highest contribution to the overall mass density is in the range of $\log(N_{\mathrm{H_2}}/\mathrm{cm}^{-2}) \sim 21.2 - 21.5$. This is at densities detected in the ISM and typical for molecular clouds. We note that for the 1 kpc resolution PHANGS-ALMA data the highest contribution shifts to $\log(N_{\mathrm{H_2}}/\mathrm{cm}^{-2}) \sim 21$. However, it is not trivial to quantify how much this is an effect of higher completeness at lower resolutions compared to averaging over a larger area.

The mass density peak in PHANGS-ALMA is detected at densities above the star formation threshold advocated by Clark & Glover (2014). While this is inconsistent with results by TNG100, we note that the observations of PHANGS-ALMA are incomplete in this region. It is therefore conceivable that deeper observations of molecular gas in these galaxies may shift the observed column density contributions to lower column densities.

In conclusion, when combining results by observations and simulations, the highest H_2 mass density contribution is found at column densities detected within the ISM of galaxies and partly in regions observed in local molecular clouds.

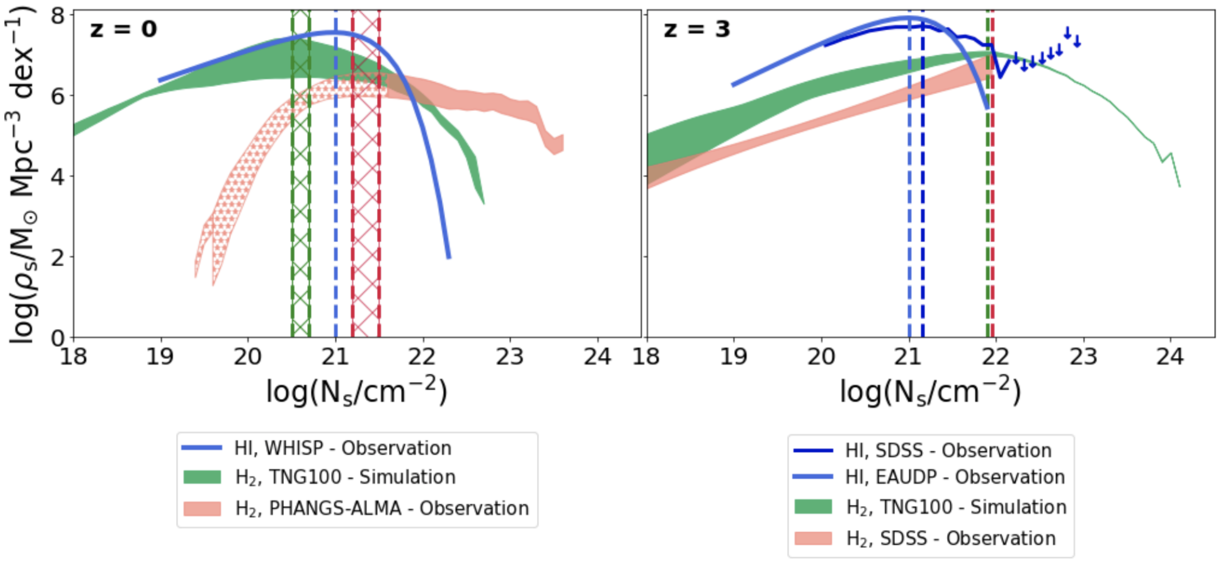

Figure 3.5: The mass density contribution per dex column density (ρ_s) of H I and H_2 derived from observations and simulations. Overall at $z = 3$ the highest mass contribution of H_2 comes from denser gas compared to $z = 0$ in both simulations and observations. Further, the mass density distributions suggest that H I dominates over H_2 at most column densities, making it an important contributor to the cold gas mass density of galaxies. H_2 starts dominating compared to H I at column densities above $\log(N_{H_2}/cm^{-2}) \sim 21.8 - 22$ at both redshifts. **Left:** Mass densities of H_2 and H I against column densities at $z = 0$ for TNG100 (green band), the PHANGS-ALMA survey (red band) and the WHISP sample (blue line, Zwaan et al., 2005). The H_2 highest mass density contribution can be constrained between $\log(N_{H_2}/cm^{-2}) \sim 20.5 - 21.5$ (column density regions typical for the ISM and, in part, molecular clouds) **Right:** Mass densities of H_2 and H I against column densities at $z = 3$ for TNG100 (green band), SDSS [(red band, Balashev & Noterdaeme, 2018), (dark blue line, Ho et al., 2021)] and the EAUDP sample (dark blue line, Zafar et al., 2013). The highest H_2 mass density contribution can be constrained at $\log(N_{H_2}/cm^{-2}) \sim 22$ (column densities typical for molecular clouds).

3.7.2 $f(N_{\mathrm{H_2}})$ at z=3

TNG100 broadly reproducing Observations

Here we compare the $f(N_{\mathrm{H_2}})$ based on composite SDSS H_2 absorption spectra (Balashev & Noterdaeme, 2018) to the results of TNG100 at $z = 3$. In Fig. 3.4 (right plot) we display these $f(N_{\mathrm{H_2}})$. The data from Balashev & Noterdaeme (2018) (red band) only includes column densities of $\log(N_{\mathrm{H_2}}/\mathrm{cm^{-2}}) = 18$ - 22 and we therefore can only compare the slopes before the steeper drop of $f(N_{\mathrm{H_2}})$ at higher column densities. Both $f(N_{\mathrm{H_2}})$ have similar slopes in this region ($\beta_{\mathrm{balashev}} \sim 1.13 - 1.45$, $\beta_{\mathrm{TNG}} \sim 0.88 - 1.47$). The normalization of the observed $f(N_{\mathrm{H_2}})$ is ~ 1 dex lower than that predicted by TNG100 in most column density regions.

This may be caused by the different methods of deriving $f(N_{\mathrm{H_2}})$. The observed $f(N_{\mathrm{H_2}})$ is based on absorption line studies of DLAs, which are typically observed at high impact parameters surrounding galaxies (e.g. Péroux et al., 2011b; Christensen et al., 2014; Krogager et al., 2017), while TNG100 relies on post-processed H_2 column density maps, which include all regions of galaxies. Further, these observations might be biased towards galaxies in group environments (Hamanowicz et al., 2020). Given this, while there are still inconsistencies between simulation and observation, the two $f(N_{\mathrm{H_2}})$ are remarkably close in slope. Further studies, including high spatial resolution molecular gas observations or post-processing TNG100 using ray-casting codes at typical impact parameters of absorption line systems, might help alleviate some of these inconsistencies and are an interesting avenue for future studies.

We stress that another model by Krogager & Noterdaeme (2020) using the fraction of cold gas absorption in strong H I selected absorbers derived by Balashev & Noterdaeme (2018) predicts an $f(N_{\mathrm{H_2}})$ with a knee at $\log(N_{\mathrm{H_2}}/\mathrm{cm^{-2}}) \sim 21$ and a highest column density of $\log(N_{\mathrm{H_2}}/\mathrm{cm^{-2}}) \sim 23$. While the model is also using the Blitz & Rosolowsky (2006) method for splitting the cold gas into a neutral and molecular fraction as for TNG100, the results are inconsistent with the predictions made by TNG100, which estimates the knee of $f(N_{\mathrm{H_2}})$ to occur at $\log(N_{\mathrm{H_2}}/\mathrm{cm^{-2}}) \sim 22$ and includes column densities beyond $\log(N_{\mathrm{H_2}}/\mathrm{cm^{-2}}) = 24$.

In conclusion, both observations and simulations have $f(N_{\mathrm{H_2}})$ with well-matching slopes in the overlapping regions. However, they differ in normalization by ~ 1 dex. Thus, the results by observations and simulations are in tension for the overlapping H_2 column density regions at $z = 3$.

Which Column Densities contribute most to the H_2 Mass Density at $z = 3$?

As a final analysis of $f(N_{\mathrm{H_2}})$ at $z = 3$ we study which column densities contribute the most to the overall mass density (ρ_{mol}) in both TNG100 and the $f(N_{\mathrm{H_2}})$ by Balashev & Noterdaeme (2018) derived from composite H_2 spectra.

In Fig 3.5 (right panel) we plot the mass densities for each H_2 column density. The peak of the H_2 mass density is not reached by the Balashev & Noterdaeme (2018) data (red band), meaning that we can only set a limit of $\log(N_{\mathrm{H_2}}/\mathrm{cm^{-2}}) \gtrsim 22$. This is in regions typically observed within molecular clouds. Further, it is well above the density threshold for star formation. The TNG100 results show that the highest mass density contribution is at densities of $\log(N_{\mathrm{H_2}}/\mathrm{cm}^{-2}) \sim 21.9$, so slightly below the limit that one can set with observations.

3.7.3 Denser Molecular Gas found at High Redshifts

In this section we study how $f(N_{\mathrm{H_2}})$ evolves from redshift $z = 0$ to $z = 3$. The $f(N_{\mathrm{H_2}})$ for both redshifts is shown in Fig. 3.4.

In TNG100 (green bands) the slopes below $\log(N_{\mathrm{H_2}}/\mathrm{cm}^{-2}) = 20$ are similar ($\beta_{\mathrm{TNG,z=0}} \sim 0.75 - 1.14$, $\beta_{\mathrm{TNG,z=3}} \sim 0.73 - 1.10$) and show little to no evolution. At column densities above that differences start to arise. At $z = 0$, TNG100 predicts molecular gas up to column densities of $\log(N_{\mathrm{H_2}}/\mathrm{cm}^{-2}) \sim 23$. At z=3 column densities beyond $\log(N_{\mathrm{H_2}}/\mathrm{cm}^{-2}) \gtrsim 24$ are reached in TNG100. This indicates that denser H_2 gas exists in the earlier Universe. It is the case that physical densities are intrinsically higher in the high-redshift versus low-redshift Universe. At the same time, this prediction from TNG100 could be affected by its finite numerical resolution. Further, there is a steeper drop off at high column densities at $z = 0$ compared to $z = 3$ in TNG100, where the $f(N_{\mathrm{H_2}})$ is flatter at high column densities. Due to limitations in the observations, we cannot make similar statements at the high column densities using observations. We, however, find that in the overlapping region $f(N_{\mathrm{H_2}})$ of both the SDSS sample (Balashev & Noterdaeme, 2018) and the PHANGS-ALMA survey are similar. The $f(N_{\mathrm{H_2}})$ of Balashev & Noterdaeme (2018) is a good continuation of the $f(N_{\mathrm{H_2}})$ found in the PHANGS-ALMA survey. We therefore expect larger differences in the $f(N_{\mathrm{H_2}})$ to arise at higher column densities. This would mean that the largest differences of $f(N_{\mathrm{H_2}})$ arise at the densest molecular regions in the Universe. Observations at $z = 3$ with higher column densities are needed to test if the predictions by TNG100 are correct.

Figure 3.5 shows that the column densities contributing the most to the molecular gas mass densities are shifting towards higher column densities at $z = 3$. When combining the results from observations and simulations we find the following: While at $z = 0$ the highest contribution is found at column densities of $\log(N_{\mathrm{H_2}}/\mathrm{cm}^{-2}) \sim 20.5 - 21.5$, at $z = 3$ it is found at column densities of $\log(N_{\mathrm{H_2}}/\mathrm{cm}^{-2}) \sim 21.9 - 22$. When assuming that the column density relates to the density of the gas, denser gas found at higher redshifts is in line with observations of the star formation rate across cosmic time, which is higher at $z = 3$ compared to $z = 0$ (Madau & Dickinson, 2014; Tacconi et al., 2020). The shape of the cosmic molecular mass density as a function of redshift is similar to the shape of the SFR density, making a coupling of these two quantities likely. Therefore, one would expect that more molecular gas found in denser regions leads to a higher global star formation rate in galaxies (Péroux & Howk, 2020). When assuming that the column density relates to the density of the gas, this is exactly what we observe when studying the column density distributions at $z = 0$ and $z = 3$.

3.7.4 Is H_2 dominating the Higher Column Densities?

We compare $f(N_{H_2})$ and $f(N_{H\,I})$ at $z = 0$ and $z = 3$ to study the column densities at which H_2 overtakes H I. In the following sections, we compare these derived densities with the combined results of $f(N_{H_2})$ derived from both observation and simulation. At $z = 0$ we compare the $f(N_{H_2})$ with $f(N_{H\,I})$ derived by Zwaan et al. (2005). At $z = 3$ we compare the $f(N_{H_2})$ with the $f(N_{H\,I})$ derived by Zafar et al. (2013) and Ho et al. (2021).

H I and H_2 Column Density Distributions at $z = 0$

The $f(N_{H\,I})$ at $z = 0$ from Zwaan et al. (2005) is based on H I 21-cm maps of 355 galaxies of the WHISP sample (van der Hulst et al., 2001). The WHISP sample covers galaxies of all Hubble types from S0 to Im and a considerable luminosity range and was selected using the Uppsala General Catalogue (UGC) of galaxies (Nilson, 1973). The median spatial resolution reached by these observations is ~ 1.4 kpc.

The $f(N_s)$ (left panel in Fig. 3.4) and ρ_s (left panel in Fig 3.5) at $z = 0$ show that H_2 starts to dominate the mass density at column densities above $\log(N_{H_2}/\mathrm{cm}^{-2}) \sim 21.8 - 22$ [9]. This is consistent with results from Schaye (2001), who predicted that H I clouds with $N_{H\,I} \gtrsim 10^{22}\mathrm{cm}^{-2}$ transform to molecular clouds before reaching higher column densities. Similar predictions have also been made more recently by Altay et al. (2011) and Bird et al. (2014) using (magneto-)hydrodynamical simulations.

These results imply that while molecular gas dominates the high column densities above $\log(N_s/\mathrm{cm}^{-2}) \gtrsim 22$, H I dominates the majority of the column density regions found within the interstellar medium (including column density regimes typical for molecular clouds), making neutral gas an important contributor to the cold gas mass found within galaxies at $z = 0$.

H I and H_2 Column Density Distributions at $z = 3$

The two $f(N_{H\,I})$ at $z \sim 3$ are based on H I-absorption systems (sub-DLAs and DLAs). The calculation therefore relies on pencil beam observations of H I-column densities as studying 21-cm H I in emission is not feasible at this redshift. The $f(N_{H\,I})$ by Ho et al. (2021) is based on the Sloan Digital Sky Survey Data Release 16 which was analyzed using Gaussian processes, where DLAs are detected using Bayesian model selection. While SDSS-DR16 includes redshifts between $z = 2$ and $z = 5$, we only use the results of the $z = 2.5 - 3$ integration for our comparison. The $f(N_{H\,I})$ by Zafar et al. (2013) is based on the ESO UVES advanced data products (EUADP) sample and includes measurements in the $z \sim 1.5 - 3.1$ range. The $f(N_{H\,I})$ of both samples show comparable results up to $\log(N_{H\,I}/\mathrm{cm}^{-2}) \sim 22$. Above this density SDSS results display a possible flattening of the $f(N_{H\,I})$. This flattening would be inconsistent with the predictions of the maximum $N_{H\,I}$

---

[9] We note that the TNG100 ρ_{H_2} band implies that at $z = 0$ the H_2 mass density is roughly equal in the $\log(N_{H_2}/\mathrm{cm}^{-2}) \sim 19 - 20.5$ column density region. We attribute this to a possible over-prediction of H_2 (and H I) in the simulation compared to observations at $z = 0$ (Diemer et al., 2019). Deeper observations are needed to quantify how high the contribution of molecular gas is at these densities.

by Schaye (2001), but the Gaussian process analysis shows that the $f(N_{\mathrm{H\,I}})$ in that region is also consistent with 0 and therefore not well constrained. We further note that while the SDSS-DR16 sample is larger than the EAUDP sample, the resolution is lower. The lower resolution could lead to blending at higher column densities, which would lead to measurements of column densities above $\log(N_{\mathrm{H\,I}}/\mathrm{cm}^{-2}) \sim 22$.

The $f(N_s)$ (right panel in Fig. 3.4) and ρ_s (right plot in Fig 3.5) at $z = 3$ show that H_2 starts to dominate the mass density at column densities between $\log(N_{\mathrm{H_2}}/\mathrm{cm}^{-2}) \sim 21.5 - 22$. As for $z = 0$, neutral gas is an important contributor to the global mass in a wide range of regions found in the ISM including higher-density regions typical of molecular clouds.

H I - An Important Contributor to the Cold Gas Mass of Galaxies

In conclusion, Figure 3.5 indicates that H I dominates over H_2 at most column densities. The H I column density contributing most to the overall mass density (blue vertical lines) has a higher mass contribution than H_2 at both redshifts. H I could therefore be an important contributor to the cold gas mass of galaxies at $z = 0$ and $z = 3$.

The column density contributing the most to the overall H I gas mass density is at $\log(N_{\mathrm{H\,I}}/\mathrm{cm}^{-2}) \sim 21$ for both redshifts. In contrary the highest contributing column density of H_2 evolves with redshift. It is $\log(N_{\mathrm{H_2}}/\mathrm{cm}^{-2}) \sim 22$ at $z = 3$ and less than $\log(N_{\mathrm{H_2}}/\mathrm{cm}^{-2} \sim 21.5$ at $z = 0$. We note that the molecular phase of the gas cycle is likely to be shorter than the neutral atomic phase as indicated by cold gas depletion time scales (Péroux & Howk, 2020). Therefore the molecular gas phase is more dynamic and variations in the gas densities are to be expected across cosmic time.

The H I column density contributing the most to the H I mass density is $\log(N_{\mathrm{H\,I}}/\mathrm{cm}^{-2}) \sim 21$. These high column densities are not found in diffuse gas (e.g. the CGM), but are typical of column densities found in the ISM.

At both $z = 0$ and $z = 3$ H_2 starts to dominate the mass density at column densities in the $\log(N_{\mathrm{H_2}}/\mathrm{cm}^{-2}) \sim 22$ range therefore showing little to no evolution of this observable. This is consistent with the predictions made by Schaye (2001) suggesting that little to no gas is found in the neutral phase at column densities above $\log(N_{\mathrm{H\,I}}/\mathrm{cm}^{-2}) \gtrsim 22$ due to the clouds turning molecular at those column densities.

3.8 Discussion

Given the evolution of the H_2 comoving mass density over cosmic time (e.g. Riechers et al., 2019; Péroux & Howk, 2020; Decarli et al., 2020), changes in the normalization or shape of $f(N_{\mathrm{H_2}})$ are expected. The $f(N_{\mathrm{H_2}})$ derived from both observations and simulations corroborate this hypothesis with various changes of the $f(N_{\mathrm{H_2}})$ across cosmic time. In general, the combined results of observations and simulations imply that molecular gas is more often found in systems of higher column densities at $z = 3$ when compared to $z = 0$. These changes in the $f(N_{\mathrm{H_2}})$ are in line with the higher comoving molecular mass densities detected at $z = 3$. Combined with the higher star formation rate density detected around cosmic noon (e.g. Madau & Dickinson, 2014) the results imply that the overall denser molecular gas at higher redshifts leads to a higher global star formation rate. While we study global properties in this work, these results are similar to findings of local observations of nearby star-forming galaxies where a correlation between the SFR surface density and H_2 surface density is well established (e.g. the molecular Schmidt law in Bigiel et al., 2008).

Rahmati et al. (2013) have demonstrated that observed $f(N_{\mathrm{H\,I}})$ can be accurately reproduced using the cosmological hydrodynamical simulation EAGLE (Schaye et al., 2015). Similarly, at $z = 3$, the cosmological simulation Illustris (Genel et al., 2014; Vogelsberger et al., 2014a,b; Sijacki et al., 2015) reproduces $f(N_{\mathrm{H\,I}})$ of observations (Noterdaeme et al., 2009; Zafar et al., 2013; Prochaska et al., 2010) accurately. However, there are still tensions between Illustris and observations below $z = 3$ (Bird et al., 2014). However, Villaescusa-Navarro et al. (2018) demonstrate that these tensions are not apparent in the successor of Illustris. Comparing results by TNG100 of the IllustrisTNG project with observations Villaescusa-Navarro et al. (2018) find that $f(N_{\mathrm{H\,I}})$ is accurately reproduced at $z \lesssim 5$.

While it has been demonstrated that $f(N_{\mathrm{H\,I}})$ is consistent with observations in different (magneto-)hydrodynamical cosmological simulations, there are still a number of inconsistencies for $f(N_{\mathrm{H_2}})$, despite the broad similarities of simulated and observed $f(N_{\mathrm{H_2}})$. At $z = 0$, Klitsch et al. (2019b) demonstrate that TNG100 predicts more low column density molecular gas compared to constraints by the ALMACAL survey (e.g. Oteo et al., 2016; Bonato et al., 2018; Klitsch et al., 2018) and, similarly to this work, does not reach the high column densities detected in observations (Zwaan & Prochaska, 2006). These shortcomings are, in part, due to TNG100 not resolving the cold gas phase of the ISM. These simulation specifics stem from limitations in resolution and sub-grid star formation models. Further, at $z = 0$ TNG100 might over-predict H_2 compared to observational findings (Diemer et al., 2019), especially when not taking observational apertures into account (Popping et al., 2019).

At $z = 3$, we find a ~ 1 dex difference in normalization for $f(N_{\mathrm{H_2}})$, which could arise due to the difference in selection and environments probed. The observational $f(N_{\mathrm{H_2}})$ at $z = 3$ is based on DLA studies. DLAs mostly trace the outskirts of galaxies (e.g. Péroux et al., 2011b; Christensen et al., 2014; Krogager et al., 2017) and are often associated with group environments (Hamanowicz et al., 2020), while in TNG100 the full disk with no constraints on the environment of the galaxies is probed. Therefore, further efforts, on both the observational and simulation side are needed. On the simulation side more accurate representations of the cold gas phase are needed, including different sub-grid models of star formation, higher resolution, and the inclusion of non-equilibrium chemistry. On the observational side we need better constraints of $f(N_{\mathrm{H_2}})$, especially at $z = 3$. Preferably, this could be achieved by a combination of high spatial resolution galaxy observations and a larger sample of H_2 absorption line systems at $z = 3$.

Non-equilibrium chemistry networks (e.g. Glover & Mac Low, 2007b; Glover & Clark, 2012; Gong et al., 2017), have recently been used to model the cold gas phase in simulations on the fly. Such models have been implemented in simulations of individual regions of galactic disks (e.g. Walch et al., 2015; Rathjen et al., 2021; Hu et al., 2021), isolated galaxies (e.g. Richings & Schaye, 2016; Hu et al., 2016; Lahén et al., 2019), and more recently in cosmological simulations (Maio et al., 2022). These studies have shown that non-equilibrium chemistry e.g. heavily influences the H_2 mass fraction at low metallicities (Hu et al., 2021), affects the chemical make-up of outflows (Richings & Schaye, 2016) and more accurately reproduce cosmological H_2 mass densities of observations (Maio et al., 2022). As a first attempt to study how and if non-equilibrium chemistry affects $f(N_{\mathrm{H_2}})$, we compare the time- and inclination-averaged $f(N_{\mathrm{H_2}})$ derived from a dwarf galaxy simulation by the GRIFFIN Project with the $f(N_{\mathrm{H_2}})$ derived by TNG100. The normalization of the dwarf galaxy $f(N_{\mathrm{H_2}})$ in the overlapping column density region is lower than for TNG100 $f(N_{\mathrm{H_2}})$ due to the highly different stellar masses that are probed. Interestingly, the slope of the $f(N_{\mathrm{H_2}})$ is similar, even though the samples and cold gas models are vastly different. We cannot disentangle the effects that non-equilibrium chemistry and the different samples have on $f(N_{\mathrm{H_2}})$ with our current study. Nonetheless, this could be a first indication that non-equilibrium chemistry might not affect the slope of $f(N_{\mathrm{H_2}})$, especially since the slope of $f(N_{\mathrm{H_2}})$ for individual main-sequence star-forming galaxies in TNG100 is similar below $\log(N_{\mathrm{H_2}}/\mathrm{cm}^{-2}) \lesssim 20$ and not majorly affected by galaxy properties. However, comparisons between simulation runs of the same galaxy with and without non-equilibrium chemistry could help understand if and how $f(N_{\mathrm{H_2}})$ is affected by non-equilibrium chemistry. Further, studies with larger samples, similar to Maio et al. (2022), are needed to further investigate how non-equilibrium chemistry might affect $f(N_{\mathrm{H_2}})$.

The global $f(N_{\mathrm{H_2}})$ and that of individual main-sequence star-forming galaxies give first indications that its shape could be related to the gas distribution within gas disks. Exponential gas distributions have not only been observed in disk galaxies (e.g. Leroy et al., 2008), but also reproduced in simulated ones (e.g. in TNG100, Stevens et al., 2019). An analytical model, based on exponential gas distribution in disks (Zwaan, 2000) broadly reproduces $f(N_{\mathrm{H_2}})$ of simulations and observations and is giving a first indication that these two distributions are related. Nonetheless, analytical models with e.g. Gaussian gas distributions in gas disks yield similar results. Therefore, it currently remains unclear how closely coupled the shape of $f(N_{\mathrm{H_2}})$ and the gas distribution in gas disks are. Further studies are needed for a complete understanding to confirm the hypothesis of this connection between these two observables.

At $z = 0$, observations have shown that neutral atomic hydrogen dominates the total mass of the neutral ISM, with $M_{\mathrm{H\,I}} \sim 2 - 10\ M_{\mathrm{mol}}$ (e.g. Saintonge et al., 2011; Saintonge & Catinella, 2022). In studies at higher redshifts, it is often assumed that the neutral atomic component can be omitted and H_2 is assumed to be the dominant gas component in galaxies (e.g. between z=0.4 and 4, Tacconi et al., 2018). In part, this is due to technical limitations, as the H I 21-cm emission line is not observable at higher redshifts with current instruments. Further, the molecular mass density peaks within this redshift range, while the neutral atomic mass density remains fairly constant across cosmic time, possibly making molecular gas an important contributor to the overall gas mass of galaxies within this redshift range (especially around cosmic noon). However, it still remains unclear what the contribution of the neutral atomic gas phase is to galaxies at higher redshifts.

Heintz et al. (2021) have given first indications of the contribution of H I at higher redshifts, by exploiting [C II] as a tracer for neutral atomic gas. The results indicate that at $z = 4-6$ the contribution of H I is substantial, with the H I mass being equal to the dynamical mass of galaxies. At $z \sim 2$ the contribution of H I is found to be less substantial, with the H I mass being between 0.2 - 1 dex lower than the dynamical mass of galaxies. Therefore, at $z \sim 2$, the contribution by molecular gas or the stellar component is possibly higher. Comparing $f(N_{\mathrm{H_2}})$ and $f(N_{\mathrm{H\,I}})$ we, however, find that H I is an important contributor to the overall cold gas mass found in the ISM of galaxies (see Section 3.5) at both redshift $z = 0$ and $z = 3$. We therefore caution from omitting the neutral atomic gas component in studies at these redshifts.

3.9 Conclusions

In this work, we study the H_2 column density distribution $[f(N_{\mathrm{H_2}})]$ at redshift $z = 0$ and $z = 3$ using observations and simulations. On the observational side we use data from the PHANGS-ALMA survey (Leroy et al., 2021) at $z = 0$ and from a H_2 absorption line study by Balashev & Noterdaeme (2018) at $z = 3$ based on SDSS data. On the simulation side we use data from TNG100 of the IllustrisTNG project (Marinacci et al., 2018; Springel et al., 2018; Naiman et al., 2018; Nelson et al., 2018; Pillepich et al., 2018a) at both redshift $z = 0$ and $z = 3$ and a high-resolution isolated dwarf galaxy simulation including a non-equilibrium chemical network by the GRIFFIN project (Lahén et al., 2019, 2020a,b) meant to represent a low-redshift dwarf galaxy.

In summary, our analysis includes the following studies:

- We study how the integrated properties of galaxies in the PHANGS-ALMA sample shape the $f(N_{\mathrm{H_2}})$ of individual objects.
- We contrast the $f(N_{\mathrm{H_2}})$ from observations and simulations to test how predictions made by TNG100 match observations.
- We study how well analytical models match results by TNG100.
- We compare results from a simulation including non-equilibrium chemistry (GRIFFIN Project) with results from the post-processed simulation TNG100.
- We study the evolution of $f(N_{\mathrm{H_2}})$ from $z = 3$ to $z = 0$.
- We explore which column densities contribute most to the overall H_2 and H I mass density at $z = 0$ and $z = 3$.
- We investigate how the $f(N_{\mathrm{H_2}})$ compare to $f(N_{\mathrm{H\,I}})$ based on the WHISP sample (Zwaan et al., 2005), EAUDP sample (Zafar et al., 2013) and SDSS data (Ho et al., 2021) to examine in which regions of galaxies molecular gas dominates over neutral atomic gas.

In conclusion our findings are the following:

- The shapes of the $f(N_{\mathrm{H_2}})$ of individual galaxies in the PHANGS-ALMA and the TNG100 sample at $z = 0$ are similar. This is possibly related to the galaxies in the sample. The sample consists of main-sequence star-forming galaxies, which typically have rotating disks and are hypothesized to have radially exponential gas profiles. (Leroy et al., 2008; Stevens et al., 2019). The radially exponential gas profiles could potentially be the cause of the similar $f(N_{\mathrm{H_2}})$ observed for individual galaxies. Further, the normalization of $f(N_{\mathrm{H_2}})$ and highest observed $\mathrm{H_2}$ column densities depend on the integrated star formation rate (SFR), stellar mass (M_*) and $\mathrm{H_2}$ mass ($M_{\mathrm{H_2}}$) of the galaxy. More massive galaxies lead to a higher normalization of the $f(N_{\mathrm{H_2}})$ of individual galaxies. The $f(N_{\mathrm{H_2}})$ indicates that more massive galaxies produce more dense gas.

- TNG100 broadly reproduces the $f(N_{\mathrm{H_2}})$ we observe at both $z = 0$ and $z = 3$, albeit with some key differences. At $z = 0$ TNG100 produces steeper slopes for the $f(N_{\mathrm{H_2}})$ compared to PHANGS-ALMA. Further, observations detect column densities up to $\log(N_{\mathrm{H_2}}/cm^{-2}) \sim 24$ at $z = 0$. Such high column densities are not present in TNG100 at that redshift. This is potentially due to resolution effects and the star formation sub-grid interstellar medium model, both of which could inhibit the formation of high column densities of cold gas phases. At $z = 3$ the normalization of the $f(N_{\mathrm{H_2}})$ is higher in the simulations compared to observations for the majority of the regions. This is likely due to the different environments probed by SDSS observations. Nonetheless, the slopes $f(N_{\mathrm{H_2}})$ in TNG100 and from observations are in good agreement at $z = 3$.

- The dwarf galaxy simulation from the GRIFFIN project produces similar slopes as TNG100 for $f(N_{\mathrm{H_2}})$ in the overlapping column density region. It is surprising that the slope of $f(N_{\mathrm{H_2}})$ of a single simulated galaxy including a non-equilibrium chemistry network is so similar to the slope of a large sample of galaxies where $\mathrm{H_2}$ was derived using post-processing prescriptions. This could be a first indication that non-equilibrium chemistry might not majorly affect the slope of $f(N_{\mathrm{H_2}})$. However, further studies are needed to understand how and if non-equilibrium chemistry affects $f(N_{\mathrm{H_2}})$.

- The slopes of $f(N_{\mathrm{H_2}})$ below $\log(N_{\mathrm{H_2}}/\mathrm{cm}^{-2}) \sim 20$ show little to no evolution from $z = 3$ to $z = 0$. As indicated by the $f(N_{\mathrm{H_2}})$ derived from TNG100, we expect an evolution of the $f(N_{\mathrm{H_2}})$ to arise at higher column densities.

- The mass density distributions of the neutral atomic and molecular gas phase indicate that H I dominates over $\mathrm{H_2}$ at most column densities and shows that H I could be an important contributor to the cold gas mass of galaxies at $z = 0$ and $z = 3$.

- The H_2 column density contributing most to the overall molecular gas density evolves with redshift. When combining data from observations and simulations, we find that the shift is from $\log(N_{\mathrm{H_2}}/\mathrm{cm}^{-2}) \sim 20.5 - 21.5$ at $z = 0$ to $\log(N_{\mathrm{H_2}}/\mathrm{cm}^{-2}) \sim 21.9 - 22$ at $z = 3$. We therefore find that more gas in denser regions is found at $z = 3$ compared to $z = 0$. These results are in line with observations of the star formation rate across cosmic time, which is higher at $z = 3$ compared to $z = 0$. The shape of the cosmic molecular mass density as a function of redshift is similar to the shape of the SFR density, making a coupling of these two quantities likely. Therefore, one would expect that more molecular gas found in denser regions leads to a higher global star formation rate in galaxies (Péroux & Howk, 2020). When assuming that the column density relates to the density of the gas, this is exactly what we observe when studying the column density distributions at both redshifts.

- Contrary to H_2, the column density contributing most to the H I gas mass density [$\log(N_{\mathrm{H\,I}}/\mathrm{cm}^{-2}) \sim 21$] does not evolve with redshift. Given that the molecular phase of the gas cycle is likely to be shorter than the neutral atomic gas phase, more variations in the molecular gas densities are expected. The highest column density contribution of H I is therefore found in regions of the ISM, and not in more diffuse regions like the CGM.

- H_2 starts dominating compared to H I at column densities above $\log(N_{\mathrm{H_2}}/\mathrm{cm}^{-2}) \sim 21.8 - 22$ at both redshifts. This is consistent with results by Schaye (2001), who predicted that H I clouds with $\log(N_{\mathrm{H\,I}}/\mathrm{cm}^{-2}) \gtrsim 22$ do not occur due to the clouds turning molecular before reaching higher column densities. Further, this implies that neutral gas is an important contributor to the overall gas mass found in the ISM of galaxies, including column density regions typical for molecular clouds.

- In order to further constrain the evolution of $f(N_{\mathrm{H_2}})$ additional observations and simulations are needed: At $z = 0$ deeper observations are needed to constrain the low-density end of $f(N_{\mathrm{H_2}})$. At $z = 3$ high spatial resolution molecular gas observations of galaxies would enable the study of the high column density end of $f(N_{\mathrm{H_2}})$ and also probe more central regions of galaxies compared to absorption line studies. On the simulational side, efforts on resolving the cold gas phase within simulations are needed to constrain the high column density end of $f(N_{\mathrm{H_2}})$ at $z = 0$. This may necessitate higher-resolution simulations together with physical models for interstellar medium gas which aim to resolve the coldest phases. Further, the use of non-equilibrium chemical networks could provide a more accurate representation of the cold gas phase (e.g. Maio et al., 2022).

Chapter 4

The 4MOST High-resolution Quasar survey (4Hi-Q): Identifying and Localizing Mg II Metal Absorbers with Machine Learning

The content of this chapter is based on an article submitted to MNRAS.

4.1 Abstract

The upcoming 4Hi-Q survey on the VISTA/4MOST multi-object spectrograph will offer new prospects of using a massive sample of $\sim$ 1 million background quasars to map the circumgalactic metal content of foreground galaxies, as traced by metal absorption at high spectral resolution (R = 20,000). Such large surveys require specialized analysis methodologies. In the absence of early data, we instead produce synthetic 4MOST high-resolution fibre quasar spectra. To do so, we use the TNG50 cosmological magnetohydrodynamical simulation, combining photo-ionization post-processing and ray tracing, to capture Mg II (λ2796, λ2803) absorbers. We then use this sample to train a Convolutional Neural Network (CNN) which searches for, and estimates the redshift of, Mg II absorbers within these spectra. For a test sample of quasar spectra with uniformly distributed properties ($\lambda_{\mathrm{Mg\,II},2796}$, $\mathrm{EW}^{\mathrm{rest}}_{\mathrm{Mg\,II},2796} = 0.05 - 5.15$ Å, SNR $= 3 - 50$), the algorithm has a robust classification accuracy of 98.6 per cent and a mean wavelength accuracy of 6.9 Å. For high signal-to-noise spectra (SNR > 20), the algorithm robustly detects and localizes Mg II absorbers down to equivalent widths of $\mathrm{EW}^{\mathrm{rest}}_{\mathrm{Mg\,II},2796} = 0.05$ Å. For the lowest SNR spectra (SNR $= 3$), the CNN reliably recovers and localizes $\mathrm{EW}^{\mathrm{rest}}_{\mathrm{Mg\,II},2796} \geq 0.75$ Å absorbers. This is more than sufficient for subsequent Voigt profile fitting to characterize detected Mg II absorbers. We make the code publicly available through GitHub. Our work provides a

proof-of-concept for future analyses of quasar spectra datasets numbering in the millions, soon to be delivered by the next generation of surveys.

4.2 Introduction

Measurements of anisotropies in the Cosmic Microwave Background (Planck Collaboration et al., 2020) and from primordial nucleosynthesis (Cooke et al., 2018) have established a clear picture of the basic constituents of the present Universe: 73 per cent dark energy, 23 per cent dark matter, and 4 per cent baryons. Across cosmic time, baryons accumulate within dark matter haloes and form the large-scale structure, galaxies, and stars of the Universe. However, a large fraction of the baryonic matter (~ 90 per cent) is expected to be in the form of low-density gas (e.g. Péroux & Howk, 2020), which is difficult to observe in emission with current instruments (e.g. Frank et al., 2012; Augustin et al., 2019; Corlies et al., 2020a).

Part of this low-density gas is attributed to the circumgalactic medium (CGM), which is loosely defined as the gas surrounding galaxies outside the disk or interstellar medium, but within the virial radius (e.g. Tumlinson et al., 2017). The CGM is a multi-phase medium with rich dynamics, as gas expelled from galaxies due to Active Galactic Nuclei (AGN) feedback (e.g. Shull et al., 2014) and stellar feedback (e.g. Ginolfi et al., 2020) interacts with gas being accreted from the cosmic web (e.g. Rubin et al., 2012; Martin et al., 2012; Turner et al., 2017; Zabl et al., 2019; Szakacs et al., 2021). This feedback-driven redistribution of baryons occurs to large scales, up to many times the virial radii of haloes, imprinting signatures of astrophysical feedback processes out to the closure radius (Ayromlou et al., 2022).

Absorption lines close in projected separation, and in frequency space, of foreground galaxies detected in background quasar (QSO) spectra are a powerful tool to study the CGM and other low surface brightness regions of the Universe. Their detection sensitivity is independent of redshift (e.g. Tripp et al., 1998). This method has allowed for the study of various metal species as well as atomic and molecular hydrogen (e.g. Ledoux et al., 2003; Noterdaeme et al., 2008; Steidel et al., 2010; Rudie et al., 2012; Werk et al., 2013; Turner et al., 2014). Additionally, absorption enables the study of the metallicity evolution of the Universe. Contrary to emission-based metallicity estimates, absorption line-based metallicity estimates are independent of excitation conditions, largely insensitive to density or temperature and require no local source of excitation. Thus, absorption-line metallicity estimates probe both low- and high-excitation gas (Péroux & Howk, 2020).

One of the most extensively studied absorption lines is the Mg II doublet ($\lambda 2796, \lambda 2803$). The doublet traces cool gas ($T \sim 10^4$ K) at low ionization states. Because of its distinct doublet feature, Mg II has been used extensively in a great number of spectroscopic surveys. Especially in the last two decades, Mg II absorption surveys have constrained the physical properties of large samples of galaxies, across a wide range of luminosities and morphologies (e.g. Lanzetta & Bowen, 1990; Nestor et al., 2005; Narayanan et al., 2007; Seyffert et al., 2013; Anand et al., 2022). To find Mg II absorbers in QSO spectra, traditional approaches use convolution-based template matching and significance thresholding (e.g. Zhu & Ménard, 2013; Anand et al., 2021). While these methods have proven highly successful, they are computationally demanding, and require heuristic parameter optimization. With upcoming massive spectroscopic surveys and the subsequent increase in data volume, new approaches need to be explored which are more computationally efficient and more accurate.

To this end, several studies have recently turned to machine learning (ML), more specifically to convolutional neural networks (CNN), to detect absorption-line systems of various species within QSO spectra. These initial investigations show promising results. The model by Zhao et al. (2019) can classify the presence or absence of Mg II absorbers with $\mathrm{EW}_{\mathrm{Mg\,II},2796} \geq 0.3$ in SDSS DR12 (Alam et al., 2015) QSO spectra with an accuracy of 94 per cent. Similar approaches for the detection of Ca II (Xia et al., 2022) and Lyman-α absorbers (Parks et al., 2018; Wang et al., 2022) have clearly demonstrated the value of CNNs in this context. In addition to their accuracy, ML-based approaches are more efficient than classical approaches. They can be orders of magnitude faster when processing a given set of quasar spectra. Thus, we are motivated to explore these techniques further, in preparation for future large-scale absorption-line data, including surveys with DESI (DESI Collaboration et al., 2016), WHT/WEAVE (Dalton et al., 2012) and VISTA/4MOST (de Jong et al., 2019).

The goal of this paper is to develop an approach that is specific to an upcoming high-resolution QSO survey, which is part of the 4MOST project on the 4-m VISTA telescope. The manuscript is organized as follows: Section 4.3 presents a short overview of the 4MOST project and the 4Hi-Q survey. Section 4.4 details the construction of mock 4Hi-Q spectra with Mg II absorbers based on the TNG50 simulation. Section 4.5 focuses on the machine learning model and training, while Section 4.6 summarises the results of our analysis. In Section 4.7, we provide a discussion of these results in a broader context and conclude in Section 4.8. We adopt an $\mathrm{H}_0 = 68$ km s^{-1} Mpc^{-1}, $h = 0.68$, $\Omega_M = 0.3$, and $\Omega_\Lambda = 0.7$ cosmology throughout.

4.3 The 4MOST High-resolution Quasar survey (4Hi-Q)

In the last two decades, large statistical samples of QSO absorbers have enabled breakthroughs in our understanding of galaxy formation and evolution. Large-scale surveys have brought such studies into a new era (e.g Noterdaeme et al., 2012; Bird et al., 2017). Ambitious endeavours with 2.5-m class telescopes - the Sloan Digital Sky Survey (SDSS, e.g. Blanton et al., 2017) in the northern hemisphere and the 2dF QSO survey (e.g. Shanks et al., 2000) in the southern hemisphere - advanced the field significantly, primarily because they produced homogeneous low-resolution spectra samples for one million QSOs. The 4Hi-Q (4MOST High-resolution Quasar survey) project is based on the next generation of such dedicated spectroscopic surveys on 4-m class telescopes, which will provide

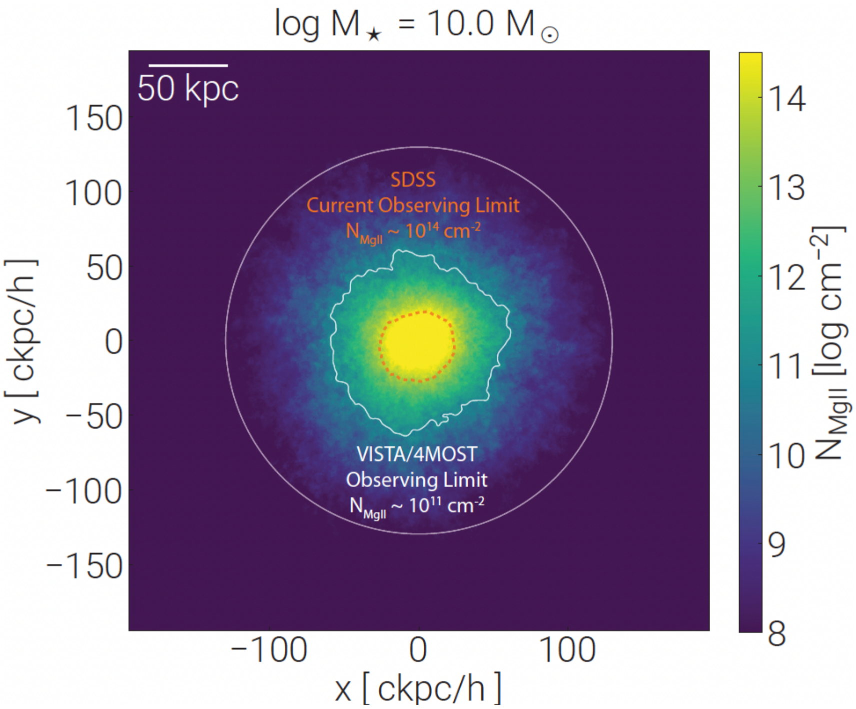

Figure 4.1: Simulation prediction for the average circumgalactic metal distribution around $M_\star = 10^{10}$ $\mathrm{M}_\odot$ galaxies at $z = 0.5$. Here we show a mean stack from the TNG50 simulation of 200 galaxies at this mass, which are similar to those targeted by the 4Hi-Q survey. The orange contour illustrates the Mg II column density detection limit currently accessible with SDSS (Anand et al., 2021), while the white contour corresponds to the Mg II column density limit within reach of the VISTA/4MOST survey. The white circle shows the virial radius r_{200}. The 4Hi-Q project will provide three orders of magnitude improvement in the Mg II column density probed throughout the extended circumgalactic medium of galaxies thanks to its large multiplexing capability and R=20,000 high-spectral resolution.

a new wealth of medium and high-resolution QSO spectra in extremely large numbers. In particular, the combination of VISTA/4MOST multiplexing capabilities (812 out of 2436 total fibres) and high spectral resolution (R=$\lambda/\Delta\lambda$ = 18000 - 21000) of the 4MOST high-resolution spectrograph will enable the construction of a unique long-lasting legacy sample of QSO spectra. The start of observations is foreseen for 2024, lasting for 5 years. The 4Hi-Q project will use data of $\sim$ 1 million background QSOs from an approved 2.8 million fibre-hour VISTA/4MOST open-time (community) survey (PI: Péroux) to search for metal [e.g. Mg II (λ2796, λ2803), C IV (λ1548, λ1550)] and Lyman-α absorption-line systems. While individual absorption measurements are limited to a pencil-beam along the line of sight and hence sample a small section of the host galaxy, large samples allow us to statistically measure the mean properties of the CGM of galaxies by combining many sightlines.

Fig. 4.1 illustrates the three orders of magnitude gain in Mg II column density which will be reached with the 4Hi-Q survey, in comparison to current SDSS sensitivities (see e.g. Anand et al., 2021). Therefore, VISTA/4MOST will probe the CGM of galaxies at larger scales than SDSS. Importantly, what makes the 4Hi-Q survey unique, is a well-studied population of over 1.5 million foreground galaxies (Driver et al., 2019; Richard et al., 2019), AGN (Merloni et al., 2019) and groups and clusters (Finoguenov et al., 2019) to be observed with the low-resolution fibres ($R = \lambda/\Delta\lambda = 4000 - 7500$) of VISTA/4MOST in the same fields at a redshift concomitant with the Mg II absorbers. Clearly, such surveys will require novel and targeted approaches to analyze their massive data outputs, in order to detect the expected hundreds of thousands of intervening absorbers.

4.4 Constructing the Training and Test Sets

4.4.1 Mg II Absorbers in TNG50 Simulations

We use the TNG50 simulation (Pillepich et al., 2019; Nelson et al., 2019b) of the IllustrisTNG project to create synthetic Mg II absorption profiles. The TNG project[1] (Naiman et al., 2018; Pillepich et al., 2018c; Springel et al., 2018; Nelson et al., 2018; Marinacci et al., 2018) is a large-volume cosmological gravo-magnetohydrodynamics (MHD) simulation incorporating a comprehensive model for galaxy formation physics. TNG uses the AREPO code (Springel, 2010) which self-consistently evolves a cosmological mixture of dark matter, gas, stars, and black holes as prescribed by self-gravity coupled to ideal, continuum MHD (Pakmor et al., 2011; Pakmor & Springel, 2013).

The physical processes included in the simulations are, broadly: gas radiative effects, including primordial and metal-line cooling, plus heating from a meta-galactic background radiation field (Faucher-Giguère et al., 2009); star formation within the cold component of a two-phase interstellar medium model (Springel & Hernquist, 2003); the evolution of stellar populations and subsequent chemical enrichment, including Supernovae Ia, II, and AGB stars (independently tracking the ten elements H, He, C, N, O, Ne, Mg, Si, Fe, and Eu); galactic-scale outflows generated by supernova feedback (Pillepich et al., 2018b); the formation and mergers of supermassive black holes (SMBHs) and their accretion of neighbouring gas (Springel et al., 2005; Di Matteo et al., 2005); SMBH feedback that operates in a dual mode with a thermal 'quasar' mode for high accretion rates and a kinetic 'wind' mode for low accretion rates (Weinberger et al., 2017; Pillepich et al., 2021). TNG50 includes 2×2160^3 resolution elements (gas plus dark matter) in a ~ 50 Mpc (comoving) box, giving a baryon mass resolution of 8.5×10^4 M$_\odot$. All data from TNG are publicly released (Nelson et al., 2019a).

Recent studies have demonstrated that the TNG50 volume is particularly suited for circumgalactic medium studies as it produces sufficiently high covering fractions of extended, cold gas, as inferred by observations. Quantitative comparisons of predicted low-ionization Mg II column densities, around massive galaxies at intermediate redshifts, reveal reasonable

---

[1] `www.tng-project.org`

agreement with observations (Nelson et al., 2020). Further, the diversity and kinematics of observed strong Mg II absorbers ($\mathrm{EW}^{\mathrm{rest}}_{\mathrm{Mg\,II},2796} \geq 0.5$Å) are reflected in mock Mg II absorber spectra based on TNG50 (DeFelippis et al., 2021), and in the overall diversity of the properties of CGM gas around the large galaxy population (Ramesh et al., 2022). Analysis of extended Lyman-α and Mg II haloes, tracing the CGM in emission, has also shown promising consistency with MUSE data (Byrohl et al., 2021; Nelson et al., 2021).

To compute Mg II we take the total magnesium mass per cell as tracked during the simulation, and use CLOUDY (Ferland et al., 2017) to calculate the ionization state assuming both collisional and photo-ionization (following the modeling approach of Nelson et al., 2020). We then ray-trace through the simulated gas distribution to create synthetic absorption spectra, akin to those in real observations (Nelson, in prep). This is similar in spirit to several other techniques for creating absorption spectra from hydrodynamical simulations, e.g. SPECWIZARD (Theuns et al., 1998; Schaye et al., 2003), TRIDENT (Hummels et al., 2017) and PYGAD (Röttgers et al., 2020).

We use three discrete snapshots from TNG50 at redshifts $z = 0.5, 0.7, 1.0$. In each case, we generate $N = 10^6$ random sightlines and propagate each for a total distance equal to the simulation box length of 35 cMpc/h. Some will intersect galaxies and cold gas, generating observable equivalent widths of Mg II absorption, while many will not. The simulated Mg II equivalent widths of the sample used in this work range from $\mathrm{EW}^{\mathrm{rest}}_{\mathrm{Mg\,II},2796} = 0.05$ to 5.15 Å, and provide physically motivated wavelength separations, doublet ratios, and other detailed spectral characteristics.

4.4.2 Synthetic 4Hi-Q Quasar Spectra

We create ∼ 680,000 normalized synthetic QSO spectra for the training of a convolutional neural network (CNN). Approximately 510,000 of these spectra include Mg II absorbers, while ∼170,000 do not. These mock spectra are 4Hi-Q-like, meaning that they are created with the 4MOST High-Resolution fibres technical specifications. As part of the 4MOST project, all data will be calibrated to a so-called Level 1 (L1). This pipeline will remove the instrumental signatures, identify the sky lines and calibrate the raw data. It will produce all L1 data products, including the science-ready, calibrated one-dimensional spectra, their associated variances, and bad pixel masks as well as any other associated information. For these reasons, we produce mock quasar spectra free of these instrumental and unwanted astronomical signatures.

First, roughly 97,000 normalized QSO spectra are created. They span a wavelength range composed of spectral windows (Blue Arm: 392.6 nm $\leq \lambda \leq$ 435.5 nm, Green Arm: 516 nm $\leq \lambda \leq$ 573 nm, Red Arm: 610 nm $\leq \lambda \leq$ 679 nm) with spectral gaps between these windows and a spectral resolution of R=$\lambda/\Delta\lambda$=20,000.

Second, we insert Mg II absorption-line systems into ∼ 72,000 of these normalized QSO spectra. The absorbers are randomly drawn from the simulation-based sightlines described in Section 4.4.1. While randomly drawn, the Mg II absorbers are inserted such that they are equally distributed in wavelength $\lambda_{\mathrm{Mg\,II},2796}$ and equivalent width $\mathrm{EW}^{\mathrm{rest}}_{\mathrm{Mg\,II},2796}$ as illustrated

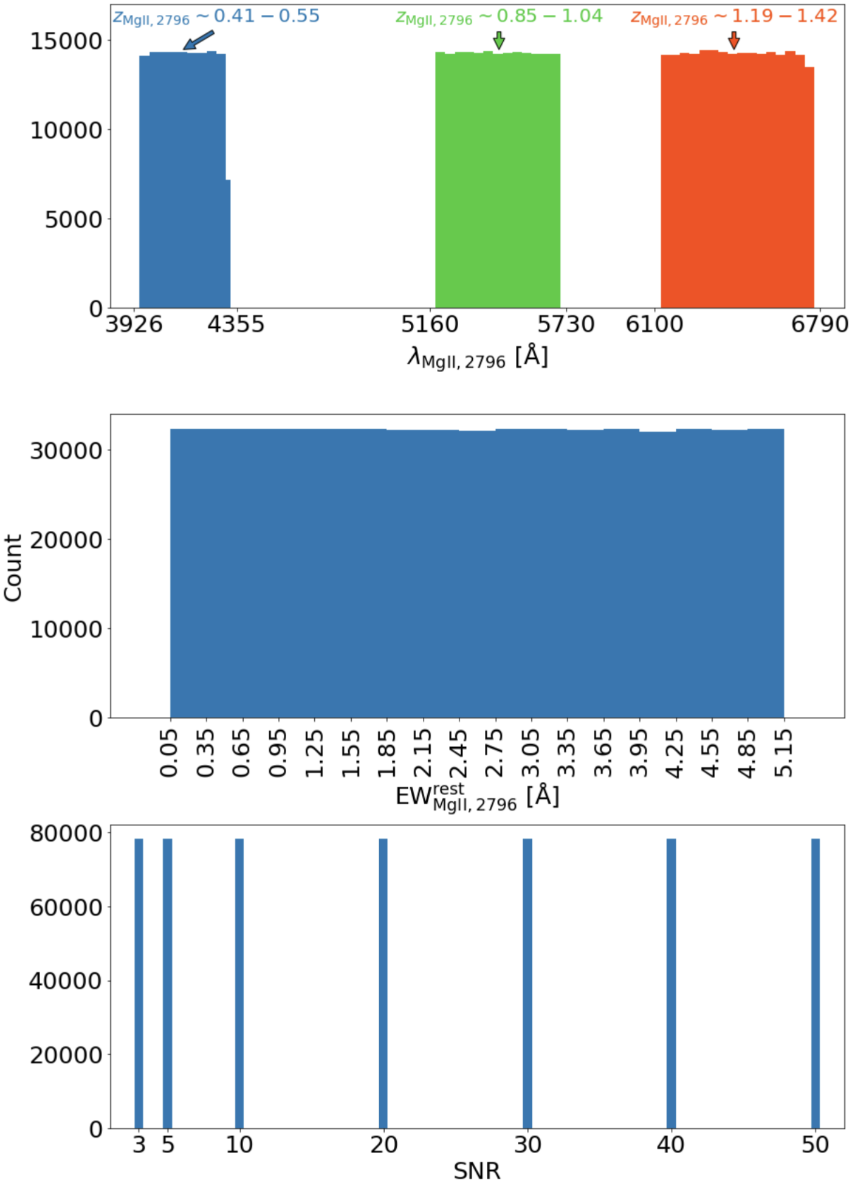

Figure 4.2: The distribution of Mg II absorber wavelength ($\lambda_{\rm Mg\,II,2796}$), equivalent width ($\rm EW^{rest}_{Mg\,II,2796}$) and SNR of our fiducial synthetic spectra sample used for training. We include only the spectra that contain a Mg II absorber. For training, we intentionally synthesize flat distributions for each of these three parameters to avoid any biases in the machine learning model.

in Fig. 4.2. Specifically, for given wavelength bins of 40 Å we randomly draw an equal number of absorbers from each 0.3 Å $\rm EW^{rest}_{Mg\,II,2796}$ bin and inject them at random positions within the wavelength bins. We note that we do not inject absorbers within 25 Å of the edges of the spectral windows, to avoid only including partial features of the Mg II doublet.

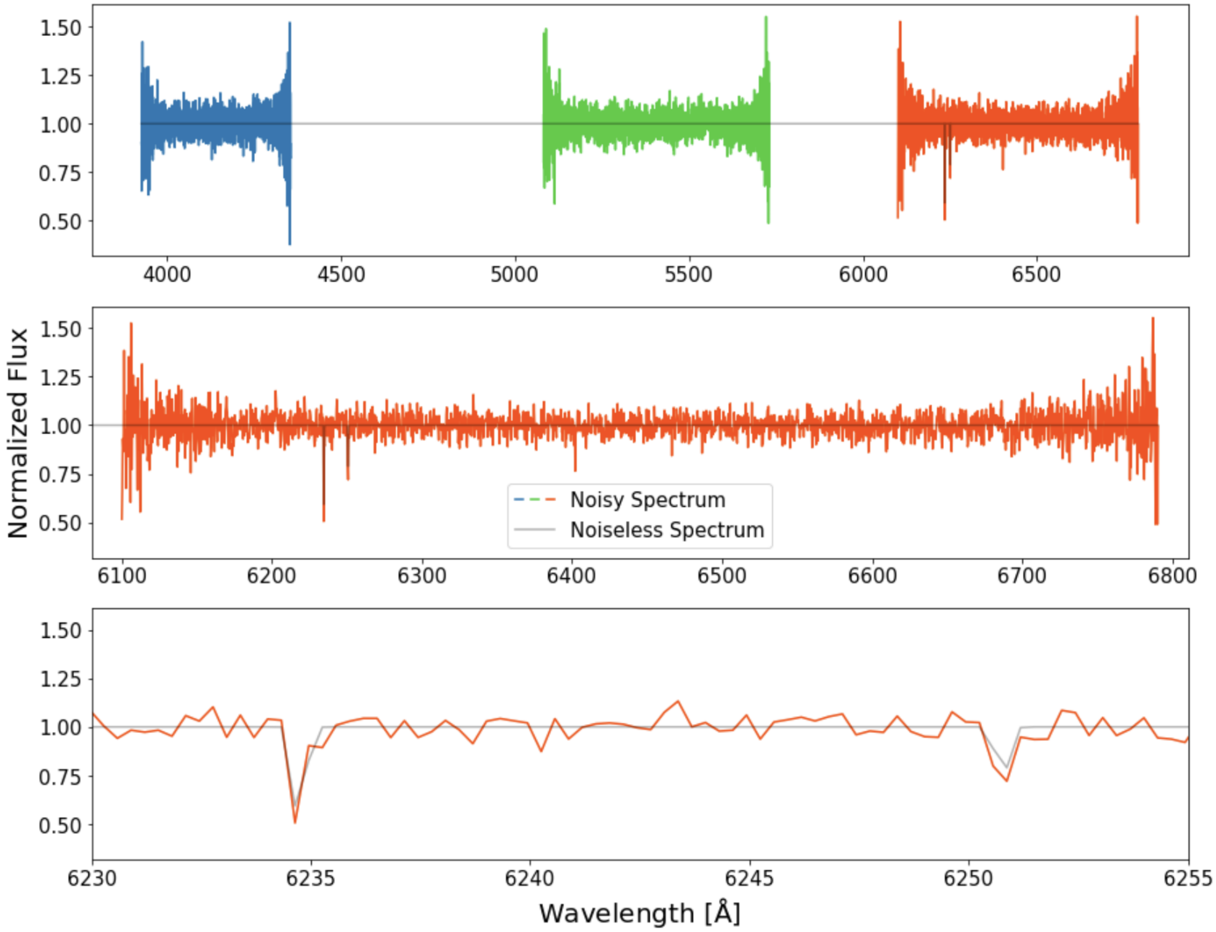

Figure 4.3: An example of our mock normalized QSO spectra with SNR = 20, a Mg II absorber at $\lambda_{\mathrm{Mg\,II},2796} = 6234$ Å and equivalent width of $\mathrm{EW}^{\mathrm{rest}}_{\mathrm{Mg\,II},2796} = 0.08$Å. The mock spectra include the spectral gaps that characterize 4MOST high-resolution fibre spectra, with decreasing SNR towards the edges of spectral windows. **Top:** Full spectrum including spectral gaps between the three spectrographs. **Middle:** Red arm of the spectrum shown above. **Bottom:** Zoom in towards the Mg II absorber in the normalized QSO spectrum displaying the Mg II doublet feature (λ2796, λ2803).

Third, we add Gaussian noise to all spectra to create spectra with 7 discrete SNR values: 3, 5, 10, 20, 30, 40, 50. Similar to Sloan Digital Sky Survey (SDSS) spectra, we expect a decreasing signal-to-noise ratio (SNR) at the edge of the spectral windows for 4Hi-Q spectra due to the specifics of the instrumental response.[2] Thus, in the absence of early data, we base the estimated SNR decrease on the properties of SDSS spectra. We take a random sample of 10,000 SDSS Data Release 16 (Ahumada et al., 2020) normalized

---

[2] We note that our quoted SNR values correspond to the SNR within the centre of the spectral windows, and do not reflect these edge effects.

QSO spectra, calculate the SNR within the central 670 pixels and calculate the SNR ratio between the centre and edges in bins of 25 pixels.[3] The resulting increase in noise towards the edges of the spectral windows is apparent in the synthetic spectrum of Fig. 4.3.

These three steps lead to a final synthetic normalized QSO spectra training sample of ∼ 680,000. We also create an additional sample for testing the CNN including ∼730,000 spectra with the same distribution of properties outlined above, however including a 50-50 split of spectra with and without Mg II absorbers.

4.5 Machine Learning Model and Training

In this section, we describe the deep learning model used and detail the choice of architecture, hyperparameter optimization, and the chosen method of training the neural network. The deep learning architecture is implemented using PYTHON 3.10.4 (Van Rossum & Drake, 2009) and the open-source machine learning libraries KERAS 2.9.0 (Chollet et al., 2015) and TENSORFLOW 2.9.1 (Abadi et al., 2015). The training and testing of the deep learning models were performed on an NVIDIA TESLA V100 GPU with 16 Gigabytes of memory. The CNN and python codes related to this paper are publicly available on GitHub[4].

Our main goal is to classify the presence or absence of Mg II absorbers in spectra and to localize them in wavelength space. Thus, the network is designed to produce the following outputs:

- **Classification:**
    - 0: No intervening Mg II absorber detected in the spectrum
    - 1: Intervening Mg II absorber detected in spectrum
- **Localization:**
    - Observed wavelength of the Mg II absorber ($\lambda_{\mathrm{Mg\,II},2796}$) in Å.

When training this type of CNN, the wavelength labels for cases with no Mg II absorbers need to have a real value as well. One cannot set an invalid i.e. NaN value, as the training loss will then also be NaN, and the optimization of the network will fail. In the spirit of Parks et al. (2018), we use a central value for the wavelength label in these cases, as this approach worked well in the case of Lyman-α absorption detection and localization within QSO spectra. Thus, we set $\lambda_{\mathrm{Mg\,II},2796} = 5358$Å. This corresponds to the mean $\lambda_{\mathrm{Mg\,II},2796}$ of the synthetic spectra sample containing Mg II absorbers.

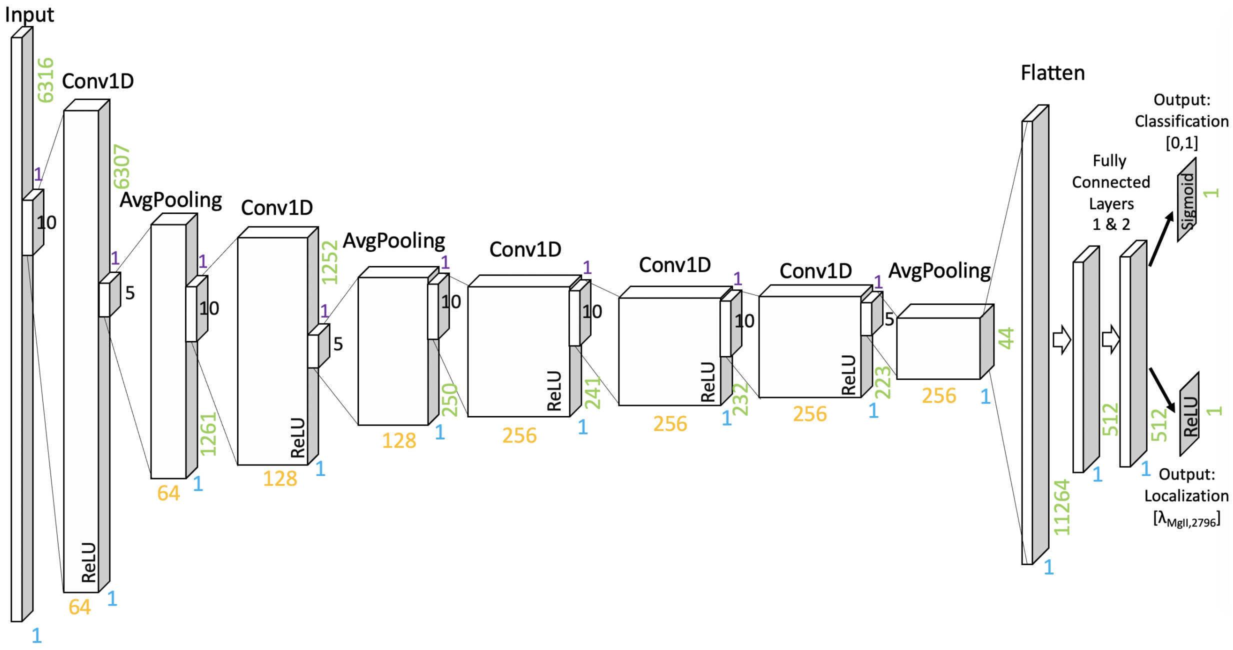

Figure 4.4: The CNN architecture used in this work. Our layer structure is similar to the AlexNet (Krizhevsky et al., 2012) structure, however, the dimensions and filters were modified and optimized to work with the 4MOST mock spectra. The input is an array containing the flux values of the spectrum with a length of 6316. There are two outputs 1) Classification (using binary cross-entropy as the loss function): 0 - no intervening Mg II absorber in the spectrum, 1 - intervening Mg II absorber in the spectrum; 2) Localization (using the mean absolute error as the loss function): Observed wavelength of the Mg II absorber ($\lambda_{\mathrm{Mg\,II},2796}$) in Å.

4.5.1 Convolutional Neural Network Architecture

We use a convolutional neural network (CNN) model [see e.g. LeCun et al. (2015) and Yamashita et al. (2018) for in-depth reviews on this topic]. CNNs are often associated with detecting features in images. However, recent studies have shown that they are useful for the analysis of QSO spectra and features within them (e.g. Parks et al., 2018; Busca & Balland, 2018; Zhao et al., 2019; Wang et al., 2022). In short, this type of network takes advantage of the fact that local groups of values in e.g. images, or in this case spectra, are often correlated.

Typically, this advantage is exploited through three layers within the CNN models: (i) convolutional layers, (ii) pooling layers, and (iii) fully connected layers. Convolutional layers perform discrete convolutions of their input using set filter (or kernel) sizes. These layers serve to detect local connections of features from previous layers. After the convo-

---

[3] Due to the higher resolution of 4Hi-Q spectra versus SDSS, that we rescale the SNR modulation from 100 to 420 pixels.

[4] github.com/astroland93/qso-mag2net

lutional layer, a non-linear activation function is applied (e.g. ReLU, sigmoid) to allow for outputs that vary non-linearly for the given inputs. Pooling layers down-sample the data by, depending on the type of pooling layer used, calculating the e.g. maximum or average values in patches of feature maps output by the convolutional layers. This allows for a shift-invariance of the feature detection. Finally, fully connected layers connect all inputs from the previous layer to all activation units of the fully connected layer. Subsequently, a non-linear activation function is applied. Thus, fully connected layers compile all the data extracted from previous layers to provide desired outputs (e.g. classification or regression). The combination of these layers leads to a neural network that can extract desired features without being affected by small shifts and distortions of these features.

We use a CNN structure resembling an AlexNet in terms of layer structure (Krizhevsky et al., 2012). AlexNet was developed as an image classification network. Specifically, it was designed to work with two-dimensional images including three color channels (Red, Green, Blue). We modify the network to work with one-dimensional data by changing the input, filter, and pooling dimensions of the network.

Our modified version of the network is shown in Fig. 4.4. The CNN takes an input spectrum of 6316 pixels, which is fed through a series of convolutional average pooling layers. We use filter sizes of 10 for the convolutional layers and use a pooling size of 5 for the pooling layers. After each convolutional layer, the ReLU non-linear activation function (Fukushima, 1975) is applied. Subsequently, the features derived after the last average pooling layer are flattened to 1 dimension and 2 fully connected layers leading to our two final fully connected output layers for the classification of the spectrum and the localization of the Mg II absorption feature. The classification output layer uses the sigmoid non-linear activation function, while the localization output layer uses ReLU.

4.5.2 Training the Convolutional Neural Network

In this section, we outline the training method and parameters used for the deep learning model. We describe the modification of the training set before training, the loss functions used, the optimizer used, and how the learning rate was chosen.

Before training the model we remove the spectral gaps between the windows in the synthetic normalized QSO spectra. This is done for two reasons: 1) the spectral gaps do not include any important information that the network needs to learn, and 2) removing the gaps decreases the input size and thus the time needed for the training of the network. Hence, this makes the multi-task model more efficient.

A multi-task learning model, such as the CNN used here, has two different outputs that often cannot be optimized by a single loss function. In these cases a combined loss function is preferred. For the classification task we use the cross-entropy loss function:

$$L_{\text{class}} = \sum_{i}^{N} -y_{\text{class,i}} \log(\hat{\mathrm{y}}_{\text{class,i}}) - (1 - \mathrm{y}_{\text{class,i}}) \log(1 - \hat{\mathrm{y}}_{\text{class,i}}) \ , \tag{4.1}$$

where y_{class} is the ground truth of the classification and $\hat{y}_{\text{class}}$ is the CNN prediction for the label. $\hat{y}_{\text{class}}$ can be in the range $[0, 1]$ and we adopt the definition that $\hat{y}_{\text{class}} \geq 0.5$ is

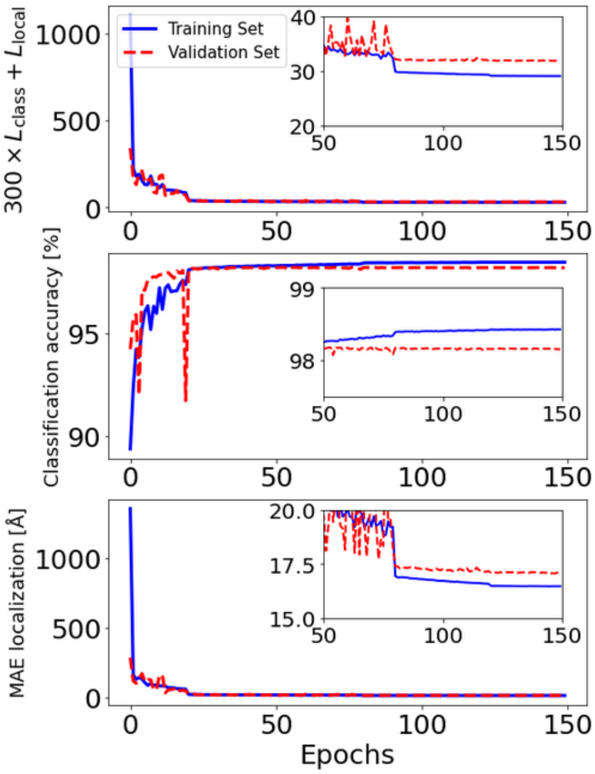

Figure 4.5: The training history of our optimized CNN. **Top:** Training history of the combined loss function. **Middle:** Training history displaying the classification accuracy metric. **Bottom:** Training history of the localization mean absolute error (MAE). While the validation set (red) metrics are below the training set (blue), the difference is small and the values for both training and validation sets are converged.

a prediction for a spectrum with a Mg II absorber, while $\hat{y}_{\text{class}} < 0.5$ indicates a spectrum without one. For the localization task, we use the mean absolute error (MAE) as the loss function:

$$L_{\text{local}} = \frac{\sum_i^N |y_{\text{local}} - \hat{y}_{\text{local}}|}{N} \ . \tag{4.2}$$

Finally, the multi-task learning model uses the sum of these two functions as its final loss function:

$$L_{\text{model}} = 300 \times L_{\text{class}} + L_{\text{local}} \ , \tag{4.3}$$

with the weight of the classification loss function (L_{class}) set to 300. This weighting is needed as the final values of the binary cross entropy loss function (L_{class}) is ~ 300 times lower than that of the localization loss function (L_{local}). Without this weighting, the CNN would put a priority on optimizing the localization loss and might neglect to optimize the classification loss.

To find the optimal parameters of the CNN, we use the Adam (Adaptive Moment Estimation) algorithm (Kingma & Ba, 2014) with the default exponential decay rates and stability constant of the KERAS library ($\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-7}$). We also implement a learning rate scheduler that additionally decreases the learning rate by an order of magnitude at set epochs: Epoch ≤ 19: LR $= 10^{-2}$, $19 <$ Epoch ≤ 79: LR $= 10^{-3}$, $79 <$ Epoch ≤ 119: LR $= 10^{-4}$, $119 <$ Epoch ≤ 150: LR $= 10^{-5}$). This was an ad-hoc choice after manually testing different types of decaying learning rates for this work (exponential decay, smaller and larger learning rates at different epochs). Given the large training set ($\sim$ 680,000 spectra), we use a data generator that individually loads datasets with a batch size of 500 into memory instead of loading the whole dataset at once. Finally, we train the CNN for 150 epochs. The training history, namely the decrease and convergence of the loss functions, for our final model is shown in Fig. 4.5.

4.5.3 Hyperparameter Optimization

Optimally, the full parameter space of hyperparameters and their various combinations should be explored simultaneously. However, given the large amount of time needed to train one model with the training sample ($\sim$ 15 hours on one V100 GPU), we split the hyperparameter optimization into two parts. First, we explored if our large fiducial model can be reduced without any significant loss in accuracy. Then, we optimized the size of the kernels in the convolutional layers and pooling layers using Bayesian (see Snoek et al., 2012) and random optimization methods.

We began our hyperparameter optimization with a fiducial model which has an excessively large width for each layer:

- Convolutional Layer - 1:
  - Filters: 128
  - Filter size: 10
- Convolutional Layer - 2:
  - Filters: 256
  - Filter size: 10
- Convolutional Layer - 3,5,6:
  - Filters: 512
  - Filter size: 10
- Fully Connected Layer size: 1024
- Average Pooling Layer size: 5

Using this fiducial model, we first explored reducing the width of the individual layers by reducing the number of filters for each layer and the size of the fully connected layers by one-half. Doing this once led to no loss in accuracy. Further, however, the accuracy of the localization task slightly degraded. Thus, we continued optimizing the filter sizes with the network width depicted in Fig. 4.4.

The parameter space probed for the filter sizes of each convolutional block was $\{5, 10, 15, 20\}$. The pooling layer kernel was varied using a size of either 3, 5, or 7. Instead of training for 150 epochs, we trained for 60 epochs for efficiency and thus modified the decrease of our learning rate accordingly (Epoch $\leq$ 5: LR $= 10^{-2}$, 5 $<$ Epoch $\leq$ 25: LR $=$ 10^{-3}, 25 $<$ Epoch $\leq$ 40: LR $= 10^{-4}$, 40 $<$ Epoch $\leq$ 60: LR $= 10^{-5}$). Otherwise, we use the same training parameters as explained in Section 4.5.2. When training for 60 epochs the CNN was sufficiently converged to appreciate the differences of the results for different hyperparameters.

We applied both the Bayesian and random optimization toolkits provided by KERAS to probe the available parameter space. The Bayesian optimization used 35 different trials, with twelve initial random parameter combinations, and 23 parameter combinations where Bayesian optimization was applied. The random optimization used 20 different random parameter combinations. Both of these methods did not find a better parameter combination than our initial model within the parameter space explored, which is somewhat surprising. However, given the computational intensity of a more extensive parameter optimization, and the proof-of-concept nature of our work, we choose to adopt our initial network. This network already achieves its principal goal of detecting and localizing the Mg II absorbers with the needed accuracy for subsequent Voigt profile fitting.

4.5.4 Alternative CNN Architectures

To test whether other model architectures could provide better results, and try a number of possibilities. In short, none was more accurate than our fiducial choice. We give a short description of these tests here.

We explored an alternative CNN model resembling a 1D version of a residual network architecture (He et al., 2015). However, the advantage of residual networks, which is the possibility to create much deeper neural networks, was not needed in this case. In particular, we found that more than one residual block led to no improvement of the network. At the same time, this architecture resulted in a worse localization accuracy, by a factor of ~ 2. While there is a possibility that this accuracy could be improved by further optimizing this type of architecture, we found that our fiducial architecture works better in our initial tests and also suited our accuracy needs in both classification and localization of Mg II absorbers.

Another possibility we explored was using two individual fully connected layers, instead of a combined one for each output after the first fully connected layer of the network. The accuracy for both classification and localization was slightly lower for both cases (~ 1 per cent lower for classification, ~ 2 Å for localization). Given this, we decided to use a combined second fully connected layer.

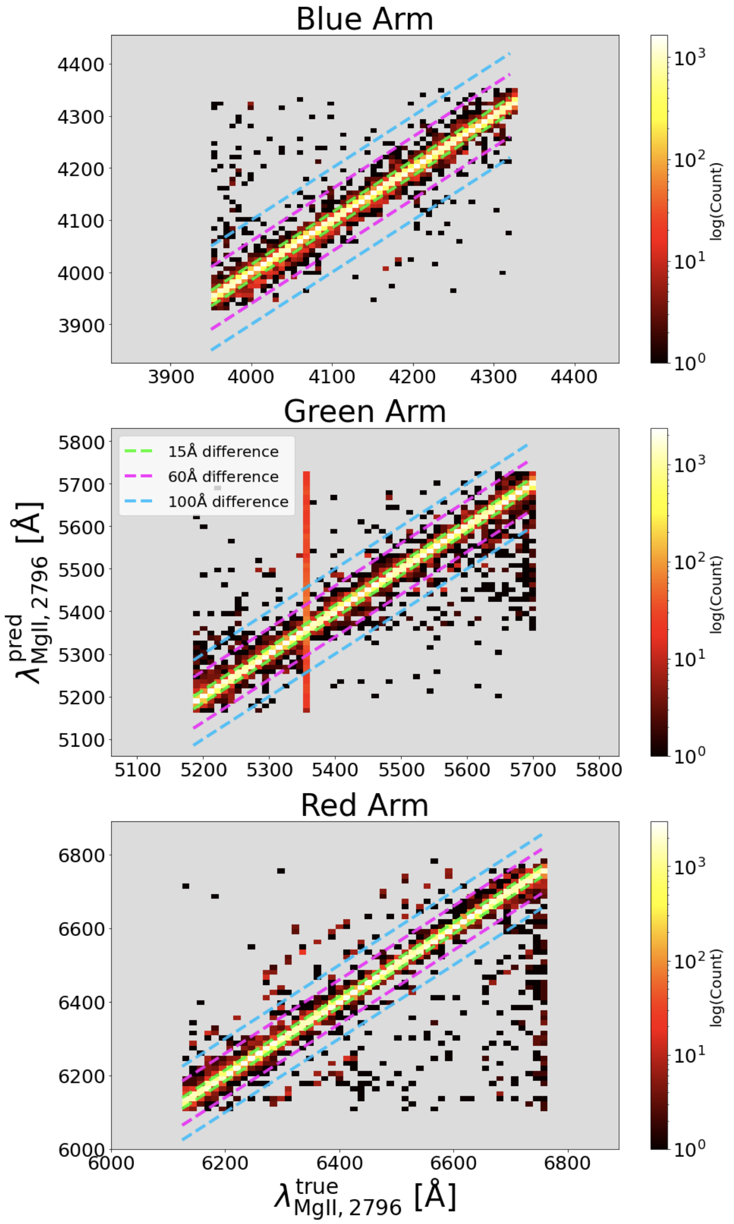

Figure 4.6: True observed Mg II absorber wavelength $\lambda^{\mathrm{true}}_{\mathrm{Mg\,II},2796}$ against predicted Mg II absorber wavelength $\lambda^{\mathrm{pred}}_{\mathrm{Mg\,II},2796}$ for spectra classified as containing a Mg II absorber by the CNN within the full test sample. **Top:** Blue Arm of 4MOST. **Middle:** Green Arm of 4MOST. A distinct line is visible at $\lambda_{\mathrm{Mg\,II},2796,\mathrm{true}} = 5358$ Å. This is caused by false positives, as $\lambda_{\mathrm{Mg\,II},2796}$ is set to 5358 Å for spectra not containing Mg II absorbers. **Bottom:** Red Arm of 4MOST. The majority of the predictions are within 15 Å of the true wavelength (i.e. within the green dotted line), which is fully sufficient to perform subsequent Voigt profile fitting.

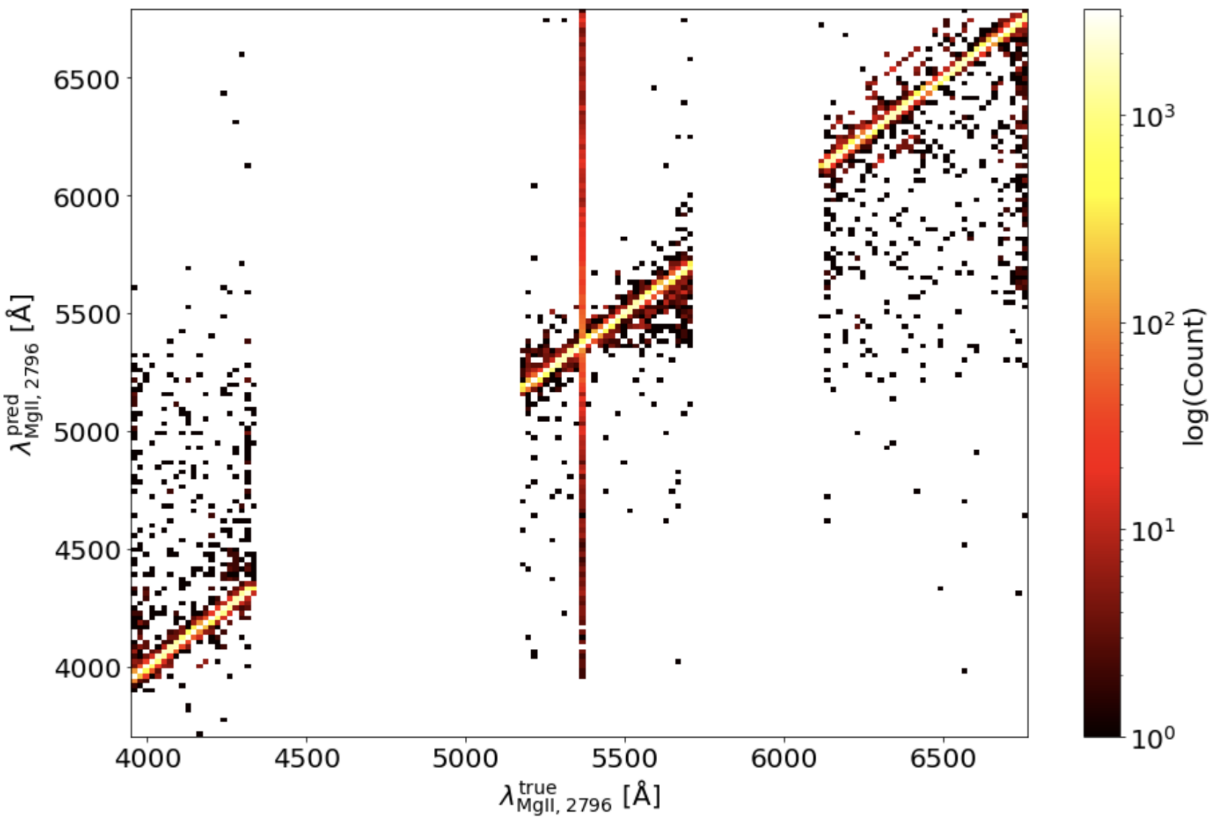

Figure 4.7: True observed Mg II absorber wavelength $\lambda^{\rm true}_{\rm Mg\,II,2796}$ against predicted Mg II absorber wavelength $\lambda^{\rm pred}_{\rm Mg\,II,2796}$ for spectra in the test set classified as containing a Mg II absorber. The majority of the predictions are within 15 Å of the true value. A distinct line is visible at $\lambda_{\rm Mg\,II,2796,true} = 5358$ Å. This is caused by false positives, as $\lambda_{\rm Mg\,II,2796}$ is set to 5358 Å for spectra not containing Mg II absorbers.

4.6 Results

We first test the ability of our CNN-based machine learning model to correctly identify if a Mg II absorber is present in a given spectrum, as well as its ability to estimate the Mg II absorber wavelength. Our test set has the same uniform statistical properties in terms of SNR, Mg II absorber wavelength ($\lambda_{\rm Mg\,II,2796}$), and Mg II absorber equivalent width ($\mathrm{EW}^{\rm rest}_{\rm Mg\,II,2796}$) distribution as the training set. In this case, we find high accuracy for classification as well as localization tasks. For ~ 98.6 per cent of the spectra, the CNN correctly identified whether a Mg II absorber is contained within the spectrum. In terms of localization, the MAE of the wavelength prediction, if the spectrum is classified as containing a Mg II absorber, is ~ 6.9 Å for the full test sample. This corresponds to a redshift MAE of $\Delta z \sim \pm 0.0025$.

In practice, it is important that the CNN provides an accurate localization when the spectrum is classified as containing an absorber. This allows for the subsequent selection of a region in which the Mg II absorption-line profile can be fitted to derive its properties. Fig. 4.6 shows the predicted versus ground truth wavelength positions of Mg II absorption for our fiducial test case. The vast majority of predictions are within ~ 15Å. This is more than sufficient for a subsequent Voigt profile fit to obtain the physical properties of the absorber.

Although the network localizes the Mg II absorber accurately in the majority of cases, there are also outliers. In Fig. 4.7 the localization predictions for the full test sample are shown, for cases where the CNN predicted that the spectrum contains an absorber. A distinct vertical line at $\lambda_{\mathrm{Mg\,II},2796} = 5358$Å is apparent: this is the ad-hoc $\lambda_{\mathrm{Mg\,II},2796}$ value set for spectra without Mg II absorbers. Hence, the vertical line corresponds to false positives, where for example the network incorrectly identified a noise feature as a Mg II absorber.

Approximately 0.4 per cent of the predicted Mg II wavelengths fall into spectral gaps. While there are no true $\lambda_{\mathrm{Mg\,II},2796}$ labels within the spectral gaps, the CNN returns $\lambda_{\mathrm{Mg\,II},2796}$ values within the full wavelength range ($\lambda_{\mathrm{Mg\,II},2796}$ = 3950 - 6930 Å). To overcome this limitation we tested suppressing the spectral gaps by remapping the true wavelength onto a continuous scale. However, the result does not reduce the number of outliers and has the side-effect of introducing additional errors at the spectral window edges. Thus, we opt to avoid remapping and keep the observed wavelength values as the output of the localization task.

4.6.1 Accurate Mg II Absorber Detection down to SNR=3

Correct classification depends sensitively on both the SNR of the spectrum and the $\mathrm{EW}^{\mathrm{rest}}_{\mathrm{Mg\,II},2796}$ of the absorber. This is apparent in Fig. 4.8, where we show the confusion matrix of the classification task normalized by the true values for different $\mathrm{EW}^{\mathrm{rest}}_{\mathrm{Mg\,II},2796}$ bins and SNRs. The first row of the figure displays the true positive rate, which can be understood as the completeness of finding Mg II absorbers. The second row displays the false negative rate, which is the inverse of the completeness. The third row shows the false positive rate. Thus, it displays the percentage of spectra where a noise feature was identified as an absorber even though no absorber is contained within the spectrum. The fourth row displays the true negative rate, i.e. the fraction of spectra for which the CNN correctly identified the non-existence of a Mg II absorber within the spectrum.

We can set a reliability threshold for our network. For each SNR, we consider the results to be reliable if the completeness is > 95 per cent for a $\mathrm{EW}^{\mathrm{rest}}_{\mathrm{Mg\,II},2796}$ bin. As the SNR increases, the $\mathrm{EW}^{\mathrm{rest}}_{\mathrm{Mg\,II},2796}$ values where the threshold is met decrease. This is to be expected, as the lower the SNR, the more difficult it is for the network to detect weaker Mg II absorption-line systems. The thresholds for different SNRs are given Table 4.1 (fourth column). There, we also provide benchmarks of the CNN for different SNRs, including all spectra, and only including spectra with Mg II absorbers above the outlined thresholds. As one would expect, the classification accuracy for sub-sets including all spectra of specific SNR increases as the SNR increases. For the sub-sets including only spectra with Mg II absorbers above the thresholds the completeness of Mg II absorber detection is ≥ 99.4 throughout. The false positive rate is low (< 4 per cent) for all SNRs, and has a small

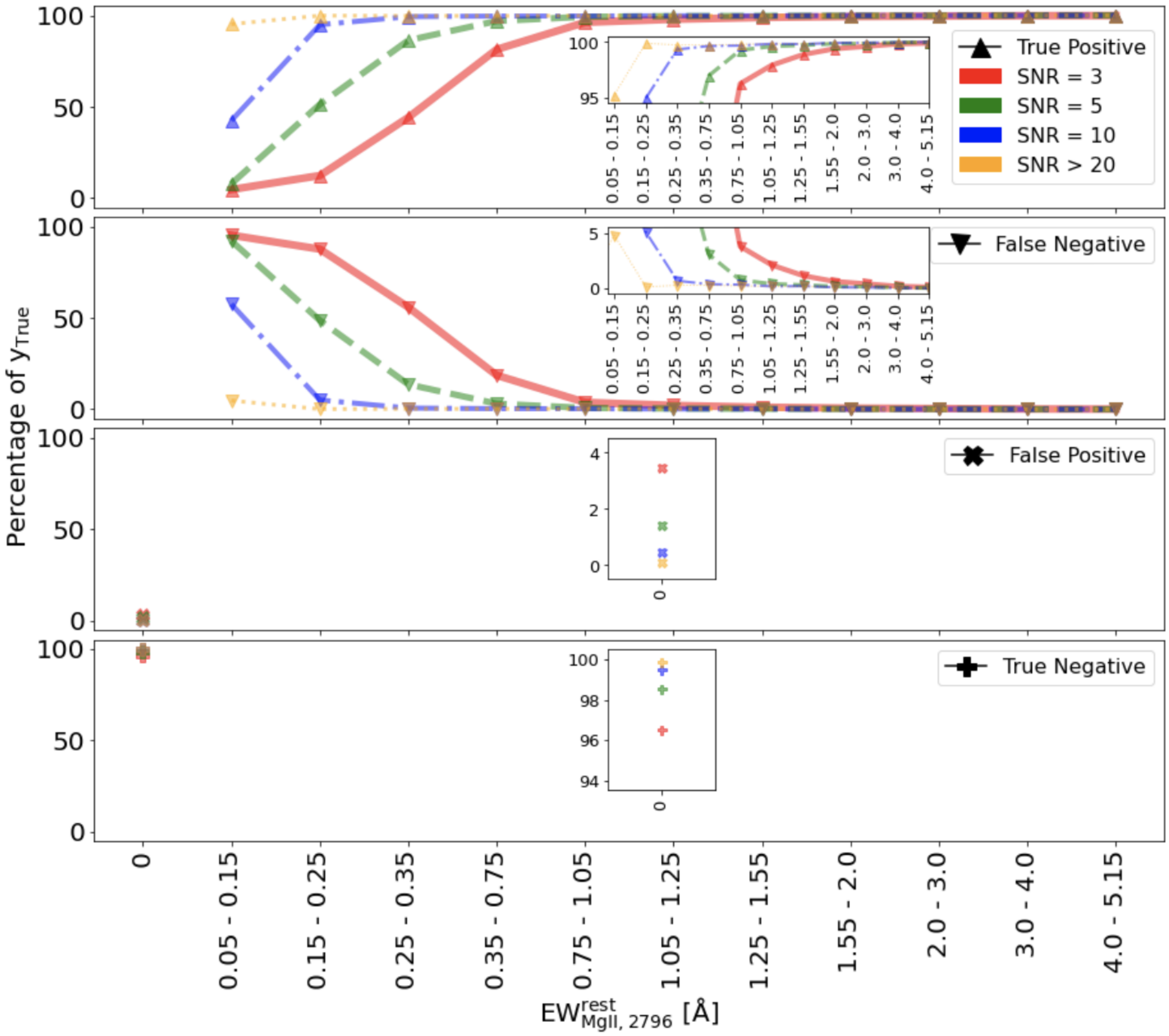

Figure 4.8: Confusion Matrix of the classification task of the CNN for spectra of different SNR and binned in $EW^{rest}_{Mg\,II,2796}$. **First row:** True positive rate. This plot describes the completeness of the detections. The completeness rises as SNR and $EW^{rest}_{Mg\,II,2796}$ increase. **Second Row:** False negative rate. The inverse of the completeness plot in the first row. The false negative rate decreases as SNR and $EW^{rest}_{Mg\,II,2796}$ increase. **Third Row:** False positive rate. This plot displays the percentage of spectra where spectra without Mg II absorbers were wrongly classified as spectra with Mg II absorbers. Thus, the CNN classified a noise feature as a Mg II absorber. There is a weak dependence on SNR, with SNR=3 spectra being a clear outlier. **Fourth Row:** True negative rate. The inverse of the third row. Thus, spectra without Mg II absorbers that are correctly classified as not containing Mg II absorbers.

dependence on SNR. As the SNR increases, the false positive rate also decreases. Thus, with more noise there is a higher chance that the CNN identifies a random noise feature as a Mg II absorber.

In summary, the CNN has a high detection completeness above $\mathrm{EW}^{\mathrm{rest}}_{\mathrm{Mg\,II},2796}$ thresholds between 0.05 and 0.75 Å for all SNR. Above $\mathrm{EW}^{\mathrm{rest}}_{\mathrm{Mg\,II},2796} \geq 0.75$ Å the CNN provides accurate results for the entire SNR range probed (SNR = 3-50).

Table 4.1: **Benchmarks of the CNN for the full test sample, and different SNR.** Row 1 (red): (1) The SNR of the benchmarked spectra. (2/3): Results for all spectra with the specified SNR.
Row 2 (green): (1) The threshold Mg II absorber rest equivalent widths ($\mathrm{EW}^{\mathrm{rest}}_{\mathrm{Mg\,II},2796}$) above which the completeness is at least 95 per cent, for a given SNR. (2/3) Results for spectra including Mg II absorbers above the threshold $\mathrm{EW}^{\mathrm{rest,thresh}}_{\mathrm{Mg\,II},2796}$, specified in row 2, column 1. The wavelength accuracy is always the mean absolute error for spectra classified by the CNN as containing a Mg II absorber.

SNR	Classification Accuracy All $\mathrm{EW}^{\mathrm{rest}}_{\mathrm{Mg\,II},2796}$ [%]	Wavelength Accuracy All $\mathrm{EW}^{\mathrm{rest}}_{\mathrm{Mg\,II},2796}$ [Å]	
	Equivalent Width Threshold $\mathrm{EW}^{\mathrm{rest,thresh}}_{\mathrm{Mg\,II},2796}$ [Å]	Completeness $\geq \mathrm{EW}^{\mathrm{rest,thresh}}_{\mathrm{Mg\,II},2796}$ [%]	Wavelength Accuracy $\geq \mathrm{EW}^{\mathrm{rest,thresh}}_{\mathrm{Mg\,II},2796}$ [Å]
3-50	98.6	6.9	
	-	-	-
3	94.7	26.7	
	0.75	99.4	7.6
5	97.3	10.8	
	0.35	99.6	4.1
10	98.8	4.9	
	0.15	99.8	2.4
20-50	99.8	1.2	
	0.05	99.8	1.6

4.6.2 Accurate Estimations of Absorber Location down to SNR=3

The mean absolute error (MAE) of the wavelength predictions in specific $\mathrm{EW}^{\mathrm{rest}}_{\mathrm{Mg\,II},2796}$ bins is tied to the completeness of the Mg II absorber detection. This is apparent in Fig. 4.9, where the MAE of the wavelength predictions for different $\mathrm{EW}^{\mathrm{rest}}_{\mathrm{Mg\,II},2796}$ bins and SNR is shown, for spectra that were classified as containing a Mg II absorber.

The high errors below the $\mathrm{EW}^{\mathrm{rest,thresh}}_{\mathrm{Mg\,II},2796}$ threshold for each SNR are driven by the CNN misinterpreting a noise feature as a Mg II absorber. Below the thresholds a large part of the Mg II absorbers are not identified by the CNN, as seen by the increasing amount of false negatives below the thresholds in Fig. 4.8. Thus, with decreasing $\mathrm{EW}^{\mathrm{rest}}_{\mathrm{Mg\,II},2796}$, fewer

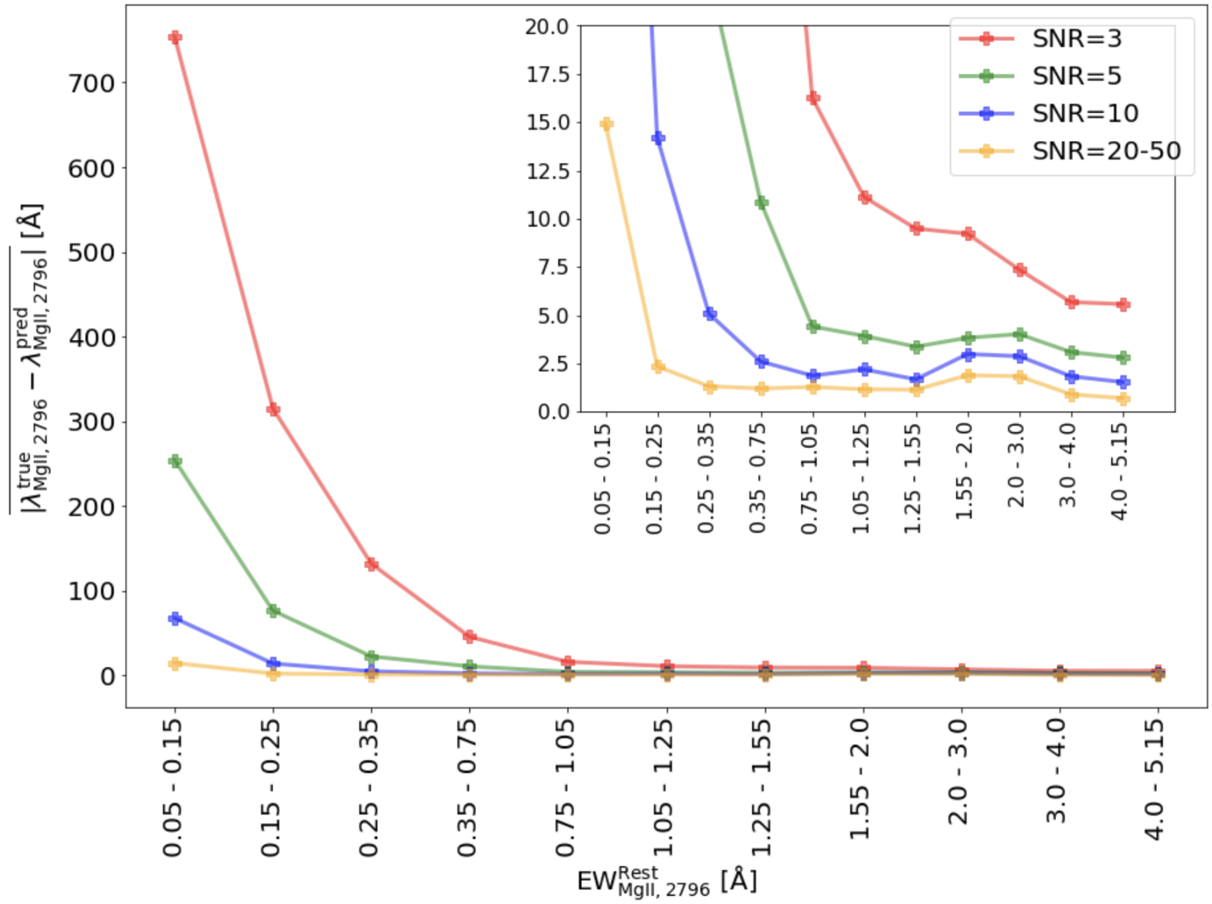

Figure 4.9: The mean absolute error for the localization output ($\lambda_{\text{Mg II},2796}$) of the CNN for various $\text{EW}^{\text{rest}}_{\text{Mg II},2796}$ and SNR. The large errors below the $\text{EW}^{\text{rest}}_{\text{Mg II},2796}$ thresholds for different SNR detailed in Table 4.1 are driven by the CNN misinterpreting noise features as Mg II absorbers. Above these thresholds, the wavelength accuracy increases with higher equivalent widths of the absorbers for all SNRs. The achieved accuracy is fully sufficient for subsequent Voigt profile fitting.

Mg II absorbers are identified correctly, as they disappear below the noise. In extreme cases this leads to only incorrectly classifying noise features as Mg II absorbers, leading to a high MAE in low $\text{EW}^{\text{rest}}_{\text{Mg II},2796}$ bins.

Above the $\text{EW}^{\text{rest}}_{\text{Mg II},2796}$ threshold we reach mean accuracies between 1 and 16 Å, depending on the SNR of the spectrum and $\text{EW}^{\text{rest}}_{\text{Mg II},2796}$ of the Mg II absorbers. The wavelength accuracy for low SNR spectra is lower than for high SNR spectra. Strong noise features lead to a possible shift of several Angstroms in the predicted localization, and thus a higher MAE. Nonetheless, our accuracy above the threshold is always high enough to enable subsequent Voigt profile fitting of Mg II absorbers.

A summary of the wavelength accuracy, for the full samples, and for samples only including absorbers above the threshold, is given in Table 4.1 (columns 3 and 6).

4.7 Discussion

Machine learning is a useful tool in many applications. However, its performance and reliability should be carefully evaluated.

4.7.1 Accuracy and Efficiency versus Traditional Methods

Traditional approaches often use convolution-based filter matching to detect Mg II candidates above a given SNR threshold (e.g. Zhu & Ménard, 2013; Anand et al., 2021). Direct comparisons with traditional methods are difficult, as this work relies on a different sample and uses idealized mock spectra with higher resolution compared to the methods outlined in Zhu & Ménard (2013) and Anand et al. (2021) that are benchmarked on SDSS spectra. However, we find that our model detects Mg II absorbers within our sample with at least the same level of completeness as traditional methods do for SDSS samples. Traditional methods typically have a completeness between 80 and 95 per cent (see e.g. Fig. 7 in both Zhu & Ménard, 2013; Anand et al., 2021) for $\mathrm{EW}^{\mathrm{rest}}_{\mathrm{Mg\,II},2796} \geq 1.0$ Å depending on the sample. The CNN-based approach reaches a higher completeness in our sample in this $\mathrm{EW}^{\mathrm{rest}}_{\mathrm{Mg\,II},2796}$ parameter space, with a completeness > 95 per cent for all SNR ≥ 3. Compared to traditional methods, the completeness drops steeply below $\mathrm{EW}^{\mathrm{rest}}_{\mathrm{Mg\,II},2796} < 0.75$ Å instead of $\mathrm{EW}^{\mathrm{rest}}_{\mathrm{Mg\,II},2796} < 1.0$ Å for lower SNR spectra in our sample. However, we note that SDSS also includes QSO spectra with SNR < 3. Further work is needed to determine if this difference arises due to our idealized spectra, the differences in samples (and subsequently SNR), or the method itself.

The CNN-based approach has a clear advantage in terms of computational efficiency. While the training of the CNN takes a significant amount of time, subsequent evaluation of the trained network is essentially free. The CNN can classify and localize Mg II absorbers within 10,000 spectra in a matter of seconds. Thus, implementing a CNN within a survey pipeline enables real-time data introspection and scientific-level output, even for $\sim$ million spectra datasets.

These results reinforce the findings concerning the feasibility of the CNN approach by Zhao et al. (2019) based on SDSS quasar spectra. With their CNN, they drew the same statistical results as the traditional approach by Zhu & Ménard (2013), however with a significantly higher computational efficiency. However, Zhao et al. (2019) only classified whether a QSO spectrum included a Mg II absorber or not, and with a different architecture than the one we explore herein. They did not include the localization aspect.

4.7.2 Future Work

Our investigation is a proof of concept for the feasibility of using deep learning to detect and localize the Mg II doublet (λ2796, λ2803) absorption-line systems in normalized QSO spectra with a significant increase in computational efficiency compared to traditional methods. There are several possible improvements for the future.

First, we only consider the case of one Mg II doublet within each spectrum. Thus, our CNN is not able to provide a prediction if multiple Mg II absorbers exist within one spectrum. To address this issue we could train the network to only look at sub-sections within the spectrum, with a sliding window size in which multiple Mg II absorbers are unlikely (e.g. similar to an approach for DLA detection in Parks et al., 2018). Alternatively, we could train the network including spectra with multiple Mg II absorbers, including simultaneous output for several absorbers within a spectrum. Finally, we could mask each detected absorption line system after its identification in a spectrum, and then run another iteration of the CNN. This final approach would allow the network architecture to remain essentially unchanged from its current form.

In addition to multiple absorbers, our spectra do not include other metal absorption lines. Many are commonly detected within QSO spectra, including C IV ($\lambda1548$, $\lambda1550$), Si IV ($\lambda1393$, $\lambda1402$), and Fe II ($\lambda2382$, $\lambda2600$). To identify these species, we would clearly need to include the corresponding transitions in our mock spectra. Additional absorbing species, i.e. at the same redshift as Mg II, could significantly increase the accuracy of identifying low equivalent width absorbers. Multiple metal absorbers in spectra could trace the same intervening gas and thus provide additional information through the intrinsic wavelength spacing between different species. Some metal lines might also have a higher equivalent width than others, making their detection easier. Multi-species joint inference could boost the performance of the CNN. We note that including different species could also potentially lead to a decrease in accuracy due to the chance of the CNN confusing the different absorption lines. Hence, this needs to be carefully evaluated.

Beyond the properties of the absorbers, the quasar spectra themselves can be improved. In this work, we use idealized, normalized QSO spectra. Thus, they do not include possible artifacts related to inaccurate normalization. To improve this aspect, we can either include continuum normalization-related errors or train the network directly on non-normalized spectra. This would increase the parameter space needed for the training set, as different QSO parameters would have to be taken into account. However, our first tests on a subset of the parameter space, and other works based on SDSS spectra (e.g. Zhao et al., 2019; Xia et al., 2022), show that this is a viable method. This effectively incorporates the continuum estimation process into the CNN itself.

Finally, our method currently identifies and localizes, Mg II absorbers. The model could be extended to simultaneously measure the equivalent width $\mathrm{EW}^{\mathrm{rest}}_{\mathrm{Mg\,II},2796}$ and column density ($N_{\mathrm{Mg\,II}}$). This would prevent the need for the second step of Voigt profile fitting, and this approach has been used to measure Lyman-α absorber column densities (Parks et al., 2018). If implemented with a method such as conditional invertible neural networks (cINNs), the full posterior distribution i.e. uncertainties on these parameters could simultaneously be constrained (see Eisert et al., 2022).

4.8 Summary

In preparation for the upcoming VISTA/4MOST community survey 4Hi-Q, we explore the feasibility of a machine learning approach to detect and localize Mg II absorption-line systems in synthetic, 4MOST-like high-resolution QSO spectra. Using the TNG50 cosmological simulation TNG50 we create millions of mock Mg II absorption profiles by combining a post-processing photo- plus collisional ionization calculation with a geometrical ray-tracing step.

We then use these synthetic Mg II absorbers, with uniform distributions in $\mathrm{EW}^{\mathrm{rest}}_{\mathrm{Mg\,II},2796}$ and $\lambda_{\mathrm{Mg\,II},2796}$, to create R=$\lambda/\Delta\lambda$=20,000 mock, continuum normalized QSO spectra. These cover the parameter space of $\mathrm{EW}^{\mathrm{rest}}_{\mathrm{Mg\,II},2796} = 0.05-5.15$Å. We add noise corresponding to expected signal-to-noise levels, from SNR = 3 to SNR = 50.

We design a convolution neural network (CNN) model to simultaneously identify, and measure the wavelength of, Mg II absorbers. For training, we construct a sample that consists of $\sim 680,000$ spectra (∼510,000 with Mg II absorbers, and ∼170,000 without).

After a hyper-parameter optimization step, we test our final trained model on a test sample that has a 50-50 split of spectra with, and without, Mg II absorbers, as well as a flat distribution of $\mathrm{EW}^{\mathrm{rest}}_{\mathrm{Mg\,II},2796}$, $\lambda_{\mathrm{Mg\,II},2796}$ and SNR. Our best trained model achieves a 98.6 per cent global classification accuracy, correctly identifying whether a Mg II absorber is present in a spectrum for the majority of spectra. It localizes Mg II absorbers with a mean absolute error of 6.9 Å for spectra classified as containing a Mg II absorber. This is fully sufficient for subsequent Voigt profile fitting.

The Mg II absorber detection completeness and localization accuracy of our method depend strongly on the SNR of the spectrum and on the $\mathrm{EW}^{\mathrm{rest}}_{\mathrm{Mg\,II},2796}$ of the absorber. We determine a $\mathrm{EW}^{\mathrm{rest}}_{\mathrm{Mg\,II},2796}$ threshold above which our method gives reliable predictions, defined as 95 per cent detection completeness. For SNR = 3 spectra, this is $\mathrm{EW}^{\mathrm{rest,thresh}}_{\mathrm{Mg\,II},2796} \geq 0.75$ Å, with a corresponding completeness of 99.4 per cent and a localization mean absolute error (MAE) of 7.6 Å. For the highest quality spectra SNR $= 20 - 50$, this improves to $\mathrm{EW}^{\mathrm{rest,thresh}}_{\mathrm{Mg\,II},2796} \geq 0.05$ Å, with a corresponding detection completeness of 99.8 per cent and a localization MAE of 1.6 Å (see Table 4.1).

In addition to its high classification and localization accuracy, one key advantage of our CNN-based technique is speed. The computational efficiency of the detection of Mg II absorbers with this approach is significantly higher compared to traditional methods. Although the initial training step is expensive ($\sim$ 15 hours on one NVIDIA TESLA V100 GPU), subsequent evaluation is essentially free: the network can process ∼10,000 spectra in seconds.

As a result, we propose that CNNs are a practical and feasible tool to detect and localize Mg II absorption-line systems in idealized 4MOST-like high-resolution spectra with high accuracy. Future work, in terms of the realism of our mock spectra, and the functionality of the model, will prepare it to be a production-quality tool for the start of 4MOST observations in 2024.

Chapter 5

Summary and Conclusions

Galaxies are systems evolving through a combination of internal processes and their connection to their immediate surroundings. Violent processes, such as feedback by Active Galactic Nuclei (AGN) and stellar feedback, expel gas from galaxies. Depending on the strength of these mechanisms, this gas can either be removed from the galaxy's halo or can stay within it and subsequently might be recycled through re-accretion. Additionally, gas reservoirs within galaxies are being replenished by the accretion of additional matter from the cosmic web. The combination of these activities leads to a redistribution of baryons within and surrounding galaxies. In particular, these processes interact in the Circumgalactic Medium (CGM), defined as the gas surrounding galaxies outside the disk or interstellar medium (ISM), but within the virial radius of galaxies. This multi-phase medium is of key importance for studies of galaxy evolution, as it allows for the tracing of gas flows. In particular, these gas flows are important in the context of star formation. The removal of gas by the aforementioned violent feedback processes could lead to the quenching of star formation in galaxies, while accretion is critical to sustaining star formation. The accreted gas mixes with the matter within galaxies and can subsequently cool down and then collapse into molecular clouds that are a necessary ingredient for the formation of stars within galaxies. Thus, the CGM, the cold gas phase, and its precursor, the cool gas phase, play key roles in how galaxies evolve across cosmic time. In this thesis we studied this cool and cold gas phase, encompassing the molecular, neutral atomic, and low-ionized atomic gas within and surrounding galaxies, applying an approach that bridges observations and simulations. Below, we reiterate some of the main findings of the three thesis projects summarized at the end of the last three chapters.

To trace inflowing, outflowing, and co-rotating gas within the CGM of galaxies, we undertook an in-depth analysis of a newly CO-detected and H I-absorption-selected galaxy within the MUSE-ALMA Haloes survey (*MUSE-ALMA Haloes: Coupling Atomic, Ionized, and Molecular Gas Kinematics of Galaxies*, Chapter 2). The MUSE-ALMA Haloes project is an ongoing study of the CGM of galaxies ($z \leq 1.4$). The project exploits multi-phase observations of H I-absorption-selected galaxies to explore the properties of galaxies (traced in emission) and their surrounding CGM (traced in absorption).

Combining multi-phase observations we explored the kinematics of the ionized and molecular gas and disentangled the galaxy-absorber connection. Kinematic modeling of the ionized and molecular gas phases revealed that the molecular [traced by CO(3–2)] and ionized (traced by [O III] $\lambda5008$) gas phases are strongly coupled within a rotating disk, as they align well directionally and have similar rotational curves and maximum rotational velocities. Kinematic studies of the absorber associated with the galaxy showed that the two-component absorption feature detected in H_2 and Mg II ($\lambda2803$) is consistent with being infalling and co-rotating gas within the CGM of the galaxy.

We compared the molecular gas properties and depletion times of H I-selected and emission-selected galaxies by compiling literature measurements of both samples at $z < 1.1$, including the newly CO-detected galaxy by the MUSE-ALMA Haloes project. This compilation indicates that selection based on H I absorbers traces objects with large molecular gas reservoirs (at given star formation rates). Subsequently, we observed depletion times that can be over an order of magnitude larger than the median depletion times in samples of emission-selected galaxies.

In the second project (*The Column Densities of Molecular Gas across Cosmic Time: Bridging Observations and Simulations*, Chapter 3), we explored the evolution of the distribution of H_2 column densities between $z = 0$ and $z = 3$ by exploiting large statistical samples from both observations (PHANGS-ALMA, SDSS) and simulations (Illustris project, GRIFFIN Project). By bridging the results of observations and simulations, we find a clear evolution of molecular gas column densities. At $z = 3$, higher molecular gas column densities are more numerous compared to $z = 0$. Subsequently, the peak of the molecular gas mass density contribution shifts to higher column densities. These results are consistent with a higher star formation rate density at $z = 3$ compared to $z = 0$, as one would expect that more molecular gas found at higher densities in the Universe leads to a higher global star formation rate.

Comparisons of the neutral atomic (H I) and molecular hydrogen (H_2) column density distribution functions revealed that H_2 starts dominating over H I at the same column density [$\log(N_{H_2}/\mathrm{cm}^{-2}) \sim 21.8 - 22$] at both of the studied redshifts. This finding is consistent with theoretical predictions of the column density thresholds at which clouds turn molecular. Contrasting the mass density contributions of H I and H_2 at different column densities additionally showed that while H I contributes little to the overall cold gas mass density at high column densities, it contributes a significant amount at column densities below column densities where H_2 dominates. This implies that H I is an important contributor to the overall cold gas mass found in the ISM of galaxies at both redshifts.

On the scale of individual galaxies, the shape of the column density distribution function within our sample of both observed and simulated galaxies is similar. This is likely caused by the sample consisting of main-sequence star-forming galaxies that are hypothesized to have radially exponential gas profiles. However, we find that higher molecular gas column densities are detected in more massive galaxies and thus in galaxies with higher star formation rates.

As the size of data sets in astronomy increase with current and future large-scale surveys, the need for fast and accurate data analysis tools increases. In preparation for the upcoming 4MOST High-resolution Quasar survey (4Hi-Q), we explored the feasibility of using machine learning approaches for the detection and localization of Mg II ($\lambda2795$, $\lambda2803$) absorbers in normalized high-resolution quasar spectra (*4Hi-Q: Identifying and Localizing Mg* II *Metal Absorbers with Machine Learning*, Chapter 4). In the absence of data, we produced synthetic normalized VISTA/4MOST high-resolution quasar spectra with injected Mg II absorbers based on the TNG50 simulation of the Illustris project for the training of a convolutional neural network (CNN).

The trained CNN is highly accurate in both classifying if an absorber is contained within the spectrum (98 per cent accuracy) and in localizing detected absorbers with a mean wavelength accuracy of 6.9 Å for a sample with evenly distributed properties [wavelength position of the Mg II absorber ($\lambda_{\mathrm{Mg\,II},2796}$), Mg II rest equivalent width ($\mathrm{EW}^{\mathrm{rest}}_{\mathrm{Mg\,II},2796} = 0.05$ - 5.15 Å) and SNR of the spectrum (SNR=3-50)]. As expected, the algorithm shows a strong dependence on the SNR of the spectrum and $\mathrm{EW}^{\mathrm{rest}}_{\mathrm{Mg\,II},2796}$, with higher global classification and localization accuracies for higher SNR spectra and larger $\mathrm{EW}^{\mathrm{rest}}_{\mathrm{Mg\,II},2796}$.

This demonstrated that CNNs are a reliable tool to detect and localize Mg II absorbers in quasar spectra. In addition, they are orders of magnitude faster in the analysis of spectra than traditional methods, making them especially useful for large datasets. However, as the third thesis project was a proof of concept study, a number of future improvements in terms of realism of the mock spectra (e.g. including more realistic noise, adding other absorbers) and in terms of functionality of the model (e.g. estimating $\mathrm{EW}^{\mathrm{rest}}_{\mathrm{Mg\,II},2796}$) can be included in the future.

In conclusion, in this thesis we explored the cool and cold gas within and surrounding galaxies on the scales of individual galaxies up to large statistical samples by exploiting both observations and simulations. All of these projects served the goal of furthering our understanding of how galaxies interact with their immediate surroundings and how the cool and cold gas phases evolve across cosmic time by answering the questions outlined in Section 1.5. Thus, with this thesis we were able to show both observational and simulational evidence of the inner workings of the cosmic baryon cycle by tracing gas flows and the evolution of the molecular gas phase. Additionally, we provided a proof-of-concept for a novel analysis method that will improve the efficiency of studying the CGM with future large-scale absorption-line surveys.

Chapter 6

Outlook

Current and upcoming instruments and novel analysis methods will push the boundaries of cold gas and circumgalactic medium (CGM) studies. Especially at high redshifts, constraining the neutral atomic gas and molecular gas content of individual galaxies is a challenging endeavour, due to the faintness of the typically used tracers such as the H I 21-cm and CO transitions. However, alternative tracers, particularly the [C II] 158μm emission line, provide exciting opportunities to overcome these limitations (e.g. Heintz et al., 2021, 2022; Zanella et al., 2018; Madden et al., 2020; Vizgan et al., 2022a,b). While early results of [C II] as a tracer of the cold gas phase at high redshifts are promising, additional studies are needed to better constrain which gas phase is traced by it.

Similarly, studies of the galaxy-CGM connection at high redshifts are limited. The [C II] line provides promising prospects to explore this connection in the early Universe as it allows for efficient high-resolution observations that enable kinematic studies of the ISM of galaxies associated with absorbers (e.g. Neeleman et al., 2020). Additionally, joint [C II] 158μm - [O III] 88μm measurements can be exploited to explore the physical conditions of these galaxies at high redshifts (Vallini et al., 2021), further disentangling the galaxy-absorber connection.

Given this, the [C II] 158μm and [O III] 88μm emission lines are promising tools to advance the studies of cool and cold gas and the circumgalactic medium at high redshifts. In this section, we will give an outlook of two possible projects that aim to exploit these emission lines for this purpose.

6.1 Exploring [C II] 158μm as a Tracer for Molecular and Neutral Atomic Gas

The neutral atomic and molecular gas phases are challenging to detect in emission at early epochs, due to the weakness of associated H I 21-cm and H_2 emission lines. Hence, various alternative tracers of molecular gas have been used in the past [e.g. CO (Bolatto et al., 2013); CI (Valentino et al., 2018)]. Recent studies have suggested using [C II] as a tracer of either H I (Heintz et al., 2021) or H_2 (Zanella et al., 2018; Madden et al., 2020) at higher

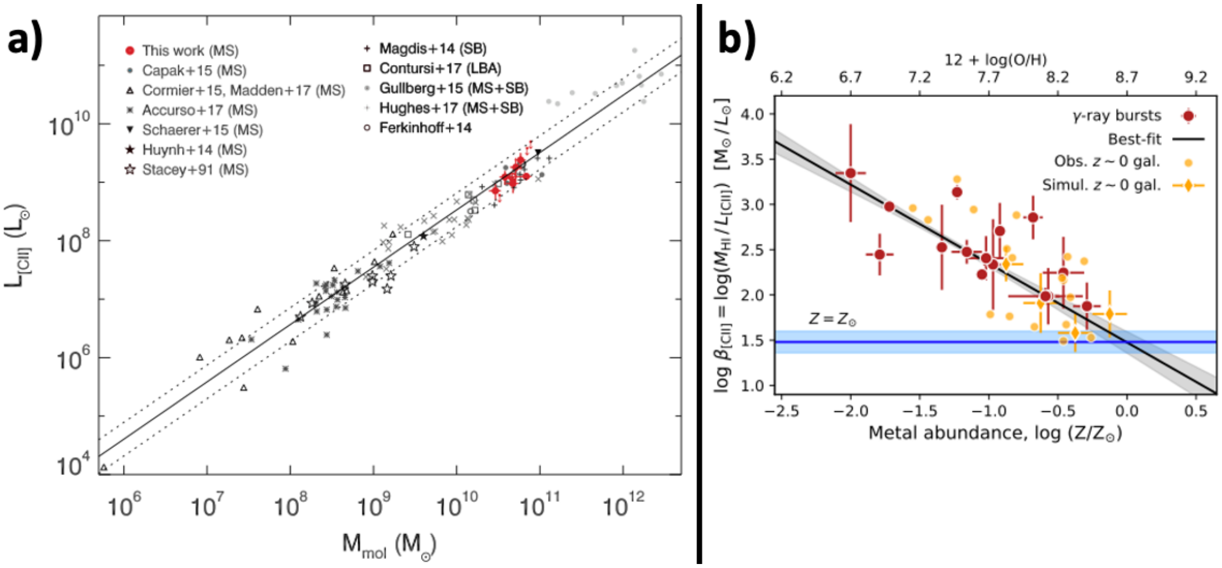

Figure 6.1: **a)** $L_{[C\,II]}$-M_{mol} relation for intermediate main sequence galaxies from low to high redshifts (Zanella et al., 2018). **b)** Metallicity dependent [C II]-to-H I conversion factor calibrated on [C II] and H_2 absorption lines in Gamma Ray bursts (Heintz et al., 2021)

redshifts (see Fig. 6.1). These studies are still in their infancy leading to large uncertainties regarding the conversion factors that need to be explored. Understanding how well [C II] traces these gas phases would provide exciting possibilities to explore the molecular and neutral atomic content of galaxies at the epoch of reionization in an efficient manner, as [C II] is brighter than other suggested tracers.

[C II] as a tracer for molecular and neutral atomic gas can be explored by taking an approach bridging observations and simulations. On the observational side, a wealth of archival data of overlapping H I (e.g. VLA), CO (e.g. ALMA), and [C II] (e.g. SOFIA, Herschel) emission-line observations of local galaxies are available. With these readily available, the local [C II]-to-H I and -H_2 conversion factors can be constrained. These observations can be contrasted with a large statistical sample from simulations by the Illustris project (e.g. Pillepich et al., 2018a) by creating post-processed [C II] and CO maps using prescriptions such as SÍGAME (Olsen et al., 2021) or CLOUDY (Ferland et al., 2017; Ramos Padilla et al., 2021) and readily available H I and H_2 maps (Popping et al., 2019; Szakacs et al., 2022). Further, these simulations can be used to explore first constraints of the conversion factor at high redshifts. Finally, current and upcoming simulations including additional physics such as non-equilibrium chemistry and radiation hydrodynamics will also offer exciting prospects for these studies in the future by a more accurate representation of [C II] and the cold gas phase [e.g. SPHINX^{20} (Katz et al., 2022) and COLDSIM (Maio et al., 2022)].

6.2 Connecting the High Redshift Interstellar- and Circumgalactic Medium

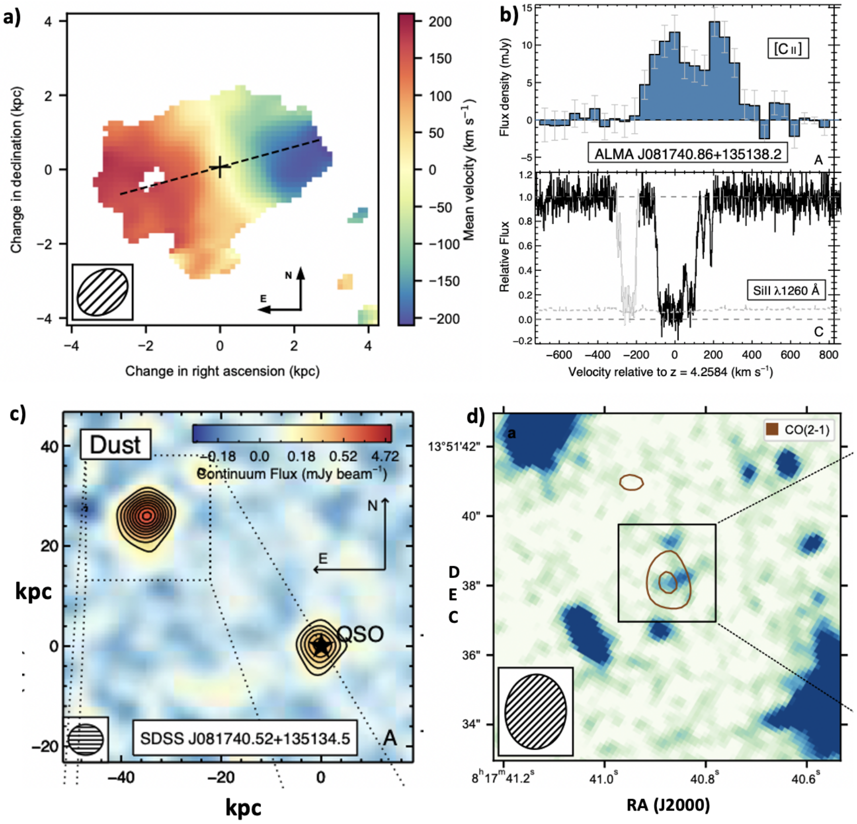

Figure 6.2: Selection of available archival ancillary data for DLA0817g at $z = 4.2601$ (Neeleman et al., 2017, 2020): a) Observed velocity map of DLA0817g based on [C II] 158 μm emission. b) [C II] 158 μm emission line spectrum of DLA0817g and absorption profile of Si II of the associated absorber. The agreement in redshift and width of the absorption and emission lines indicate that the [C II] 158 μm emission is from the DLA host galaxy. c) Dust continuum of DLA0817g and the QSO in the quasar field J081740.52+135134.5. The impact parameter of the galaxy-absorber pair is $b = 42$ kpc. d) HST F160W WFC3 rest-frame near-UV emission of DLA0817g overlayed with CO(2–1) contours obtained by JVLA.

Studying galaxy-absorber pairs has mostly been limited to low to intermediate redshifts (e.g. Schroetter et al., 2021; Szakacs et al., 2021), but recently ALMA has shown its capability to advance these studies to higher redshifts (e.g. Neeleman et al., 2019, 2020) by observations of the [C II] 158μm emission line. Future observations of this kind will provide key constraints on the relationship between galaxy-absorber pairs through kinematic studies (similar to the studies outlined in Chapter 2). Additionally, joint observations of the

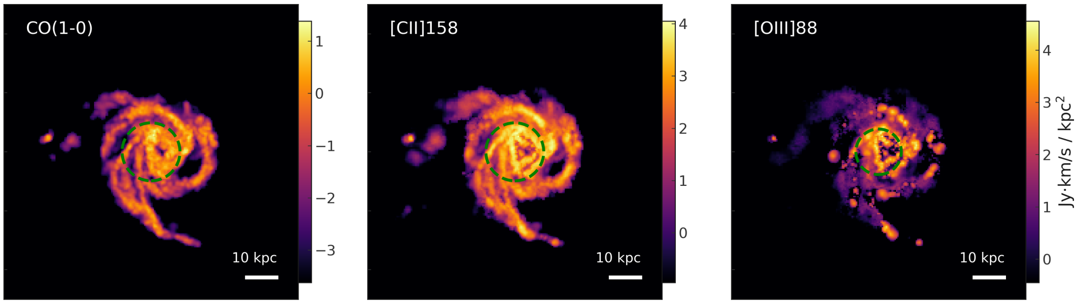

Figure 6.3: Example of a moment 0 map of CO(1-0), [C II] and [O III] of a simulated galaxy post-processed using SÍGAME (Olsen et al., 2017, 2021). Using these simulations one can contrast emission line ratios, gas density, metallicity, covering fraction of CO(2-1), [C II] 158 μm and [O III] 88 μm and kinematics of the cold and ionized gas phase from observations and simulations to further explore the physical conditions of the ISM/CGM at $z = 4$ and subsequently better constrain models of galaxy formation and evolution.

[C II] 158μm and [O III] 88μm emissions lines have demonstrated a huge diagnostic potential as they yield a complementary view of the ISM at early epochs by tracing both ionized gas in H II regions through [O III] and neutral atomic and molecular gas through [C II]. Early galaxies (e.g. Hashimoto et al., 2019; Harikane et al., 2020) have larger $L_{[\mathrm{O\,III}]}/L_{[\mathrm{C\,II}]}$ ratios when compared with local dwarf galaxies (e.g. Cormier et al., 2015). This is likely due to the highly different physical conditions in these early galaxies (Vallini et al., 2021). However, studies between the EoR and local analogues are limited, thus exploring galaxies within this redshift gap can provide additional constraints on the evolution of the ISM. Additionally, exploiting the $L_{[\mathrm{O\,III}]}/L_{[\mathrm{C\,II}]}$ ratios to derive the physical conditions within the ISM in combination with absorption line studies can aid in disentangling the galaxy-CGM connection at higher redshifts.

For this purpose, [O III] 88μm observations of galaxies associated with Damped Lyman-α (DLAs) at $z \sim 4$ offer ideal laboratories to further explore the diagnostic potential of joint [C II]-[O III] observations. The galaxy observed by Neeleman et al. (2017) offers the perfect pilot laboratory for such studies. It is in itself a highly interesting object, as it is one of the highest redshift galaxies with a confirmed cold, dusty rotating disk with a high rotational velocity of ~ 270 km/s and a large molecular gas reservoir of $M_{\mathrm{mol}} \sim 8.8\mathrm{M}_{\odot}$. Additionally, a wealth of ancillary data is already available [JVLA (CO(2–1)), ALMA ([C II]), HST (1.6 μm stellar emission), Keck/ESI + Keck/HIRES (absorption lines), see Fig. 6.2]. This object would thus serve as a unique laboratory to explore the physical conditions of the CGM and ISM at high redshifts by:

1. Deriving physical conditions of the ISM (gas density, gas metallicity) using a Markov Chain Monte Carlo (MCMC) approach (Vallini et al., 2021) and exploring their connection to the physical conditions of the CGM (H I / metal column densities, gas metallicity).

2. Studying the [O III]/[C II] emission line ratios, contrasting the line kinematics, and comparing these to other low and high redshift galaxies.
3. Exploiting this special object to compare its emission line properties (emission line ratios, kinematics) and physical conditions (gas density, metallicity, covering fractions) of the ISM with analogue quantities obtained through post-processed cosmological simulations (e.g. TNG100 by the Illustris Project and the post-processing tool SÍGAME; see Fig 6.3; Pillepich et al., 2018a; Olsen et al., 2021).

This project can then subsequently be expanded to a larger sample of [C II] emitting galaxies associated with H I absorbers at similar redshifts (e.g. Neeleman et al., 2017, 2019, 2020).

Appendix A

Appendix to Chapter 2

A.1 Observation Details

We provide further information about the ALMA observations of the MUSE-ALMA Haloes sample used in this work and provide additional information about the QSOs and observed galaxies. The additional information can be seen in table A.1.

A.2 Q2131-G1 - Kinematic Modelling Residuals and Model Flux Map

The model molecular gas flux map and residuals of the galaxy Q2131-G1 derived from GalPak^{3D} can be seen in Figures A.1 and A.2. The modelled disk reproduces the observations well, as can be seen by the low residuals.

Table A.1: **Properties of the quasars and galaxies in the MUSE-ALMA Haloes sample.**
Row 1 (red): (1) reference name of the QSO used in this paper (2) full name of the QSO, (3) right ascension of the QSO, (4) declination of the QSO, (5) QSO redshift.
Row 2 (green): (1) dates for the observations of the field, (2) exposure time of the observation, (3) angular resolution of the observation, (4) calibrators used for the observation, (5) precipitable water vapour (PWV) of the observation, (6) ALMA antenna configuration used for the observation.
Row 3 (blue): (1) reference name of the galaxy used in this paper, (2) redshifted frequency of the observed CO line, (3) right ascension of the galaxy, (4) declination of the galaxy, (5) redshift of the galaxy, (6) Detection in CO (yes/no)
Literature references: [a]) Hamanowicz et al. (2020)

QSO name	**QSO alt. name**	**RA$_{\rm QSO}$ [hh:mm:ss]**	**DEC$_{\rm QSO}$ [dd:mm:ss]**	**$z_{\rm QSO}$ [a]**	
Observation Dates	**$T_{\rm exp}$ [hrs]**	**θ ["]**	**Calibrators**	**PWV [mm]**	**Ant. Config.**
Galaxy	$f_{\rm CO}$ [GHz]	RA$_{\rm gal}$ [hh:mm:ss]	DEC$_{\rm gal}$ [dd:mm:ss]	$z_{\rm gal}$	Detected in CO
Q2131-1207	**Q2128-123**	**21:31:35**	**-12:07:04.8**	**0.43**	
4, 5, 7 Jun 2018	**2.0**	**1.02**	**J2148+0657, J2158-1501**	**0.65 - 2.8**	**C43-1**
Q2131-G1	241.866	21:31:35.636	-12:07:00.177	0.42974	yes
Q2131-G2	241.697	21:31:35.775	-12:07:11.558	0.4307 [a]	no
Q1232-0224	**1229-021**	**12:32:00**	**-02:24:04.6**	**1.05**	
26, 28 Jun 2018	**2.15**	**1.02**	**J1218-0119, J1229+0203**	**1.2 - 2.2**	**C43-1**
Q1232-G1	247.829	12:31:59.943	-02:24:05.275	0.3953 [a]	no
Q1232-G2	262.462	12:31:59.727	-02:24:12.20	0.7566 [a]	no
Q0152-2001	**UM 675**	**01:52:27**	**-20:01:07.1**	**2.06**	
2, 11, 12 Jul 2018	**1.5**	**0.96**	**J0006-0623, J0151-1732**	**0.65 - 2.6**	**C43-1**
Q0152-G1	250.105	01:52:27.827	-20:01:13.991	0.3826 [a]	no
Q1211-1030	**1209+107**	**12:11:41**	**+10:30:02.8**	**2.19**	
23 Aug 2018	**0.75**	**0.73**	**J1229+0203, J1222+0413**	**0.9 - 1.2**	**C43-3**
Q1211-G1	248.274	12:11:40.899	10:30:06.990	0.3928 [a]	no
Q1130-1449	**1127-145**	**11:30:07**	**-14:49:27.7**	**1.19**	
4, 8, 15 Dec 2016	**3.6**	**1.13**	**J1058+0133, J1139-1350**	**1.5 - 5.4**	**C40-3**
Q1130-G2	263.4	11:30:07.66	-14:49:23.41	0.3127 [a]	yes
Q1130-G4	263.44	11:30:07.62	-14:49:11.44	0.3126 [a]	yes
Q1130-G6	263.67	11:30:08.53	-14:49:28.54	0.3115 [a]	yes

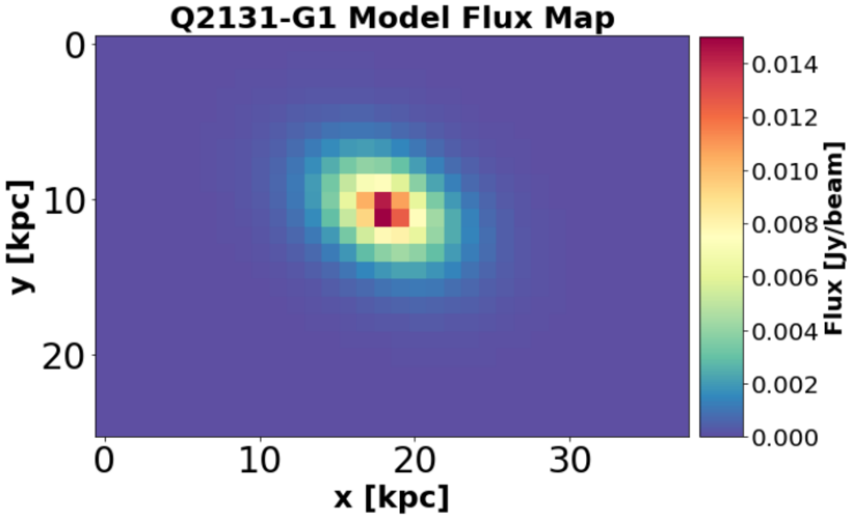

Figure A.1: Flux map of Q2131-G1 modelled in 3D-space with GalPak^{3D}.

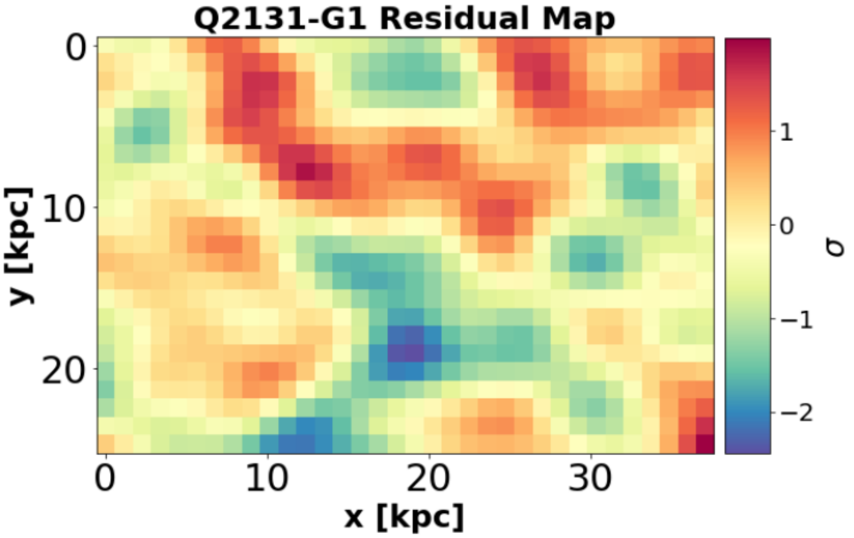

Figure A.2: Residual flux map of Q2131-G1 between the modelled and observed fluxes. The low residuals show that the disk model reproduces the observations well. The colorbar displays data - model normalized by the pixel noise σ.

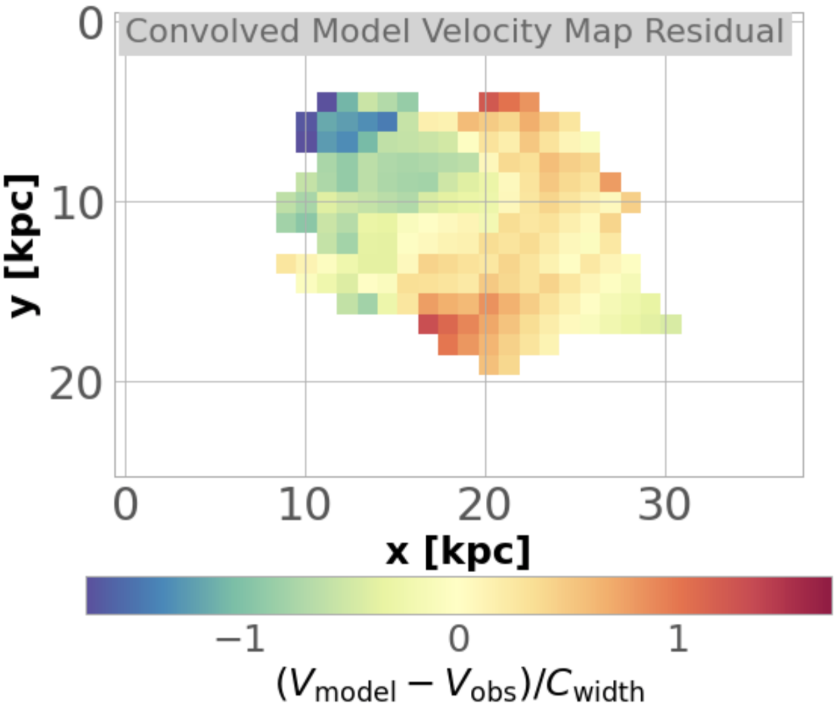

Figure A.3: Convolved model velocity map residual. The colorbar is the residual divided by the spectral resolution of the cube ($C_{\mathrm{width}} = 50\mathrm{km\ s}^{-1}$). The low residuals across the galaxy indicate that the disk model with an arctan velocity profile reproduces the observations robustly.

Appendix B

Appendix to Chapter 3

B.1 $f(N_{\mathrm{H_2}})$ Dependence on Physical Properties

We explore how the integrated properties of galaxies in the PHANGS-ALMA sample shape the $f(N_{\mathrm{H_2}})$ of individual objects. Namely we study the dependence of the $f(N_{\mathrm{H_2}})$ on the star formation rate (SFR), stellar mass (M_*), and $\mathrm{H_2}$ mass ($M_{\mathrm{H_2}}$).

The colour coding in Fig. 3.3 already displays the dependence of the $f(N_{\mathrm{H_2}})$ on the three parameters mentioned above. In order to quantify this relationship, we fit a gamma distribution of the form:

$$f(N_{\mathrm{H_2}}) = \frac{f^*}{N^*} \left(\frac{N_{\mathrm{H_2}}}{N^*}\right)^{-\beta} \mathrm{e}^{-\frac{\mathrm{N_{H_2}}}{\mathrm{N^*}}} \quad , \tag{B.1}$$

to the computed individual $f(N_{\mathrm{H_2}})$ of the 150 pc resolution PHANGS-ALMA sample. Note that there is no physical motivation for fitting a gamma distribution to the individual $f(N_{\mathrm{H_2}})$, it simply provides good fits of the $f(N_{\mathrm{H_2}})$ for a minimal number of parameters.

The individual $f(N_{\mathrm{H_2}})$ is largely determined by the parameter N^* as the second free parameter f^* correlates with N^* (slope: -0.74 ± 0.09, intercept: 14.3 ± 2.0, Pearson-r: 0.7, p-value [calculated using a Kolmogorov-Smirnov test] <0.05) and β in turn correlates with f^* (slope: -0.55 ± 0.06, intercept 0.11 ± 0.13, Pearson-r:0.73, p-value <0.05). Fig. B.1 displays the dependence of f^* on N^* and β on f^*.

As already indicated in Fig. 3.3, $f(N_{\mathrm{H_2}})$ depends on the physical parameters of the galaxies within the PHANGS-ALMA sample. This is quantified in Fig. B.2, where we show the relationship of the free parameter N^* of the gamma distribution with SFR, M_* and $\Sigma_{\mathrm{H_2}}$. As N^* largely determines the $f(N_{\mathrm{H_2}})$ of a galaxy, it is implied that these physical properties of a galaxy affect the $f(N_{\mathrm{H_2}})$ of a galaxy. The three studied properties of the galaxies show the following correlation and fit parameters (in log space): SFR - N^*: slope: 0.90 ± 0.15 intercept: 21.90 ± 0.07, Pearson-r = 0.59, p-value < 0.05; M_* - N^*: slope: 1.21 ± 0.05 intercept: -5.5 ± 1.1 Pearson-r = 0.62, p-value < 0.05 and $M_{\mathrm{H_2}}$ - N^*: slope: 0.79 ± 0.11, intercept: 14.9 ± 1.0 Pearson-r: 0.64, p-value < 0.05.

Using the SFR, M_* or $M_{\mathrm{H_2}}$ of a galaxy one could approximate its $f(N_{\mathrm{H_2}})$ using these

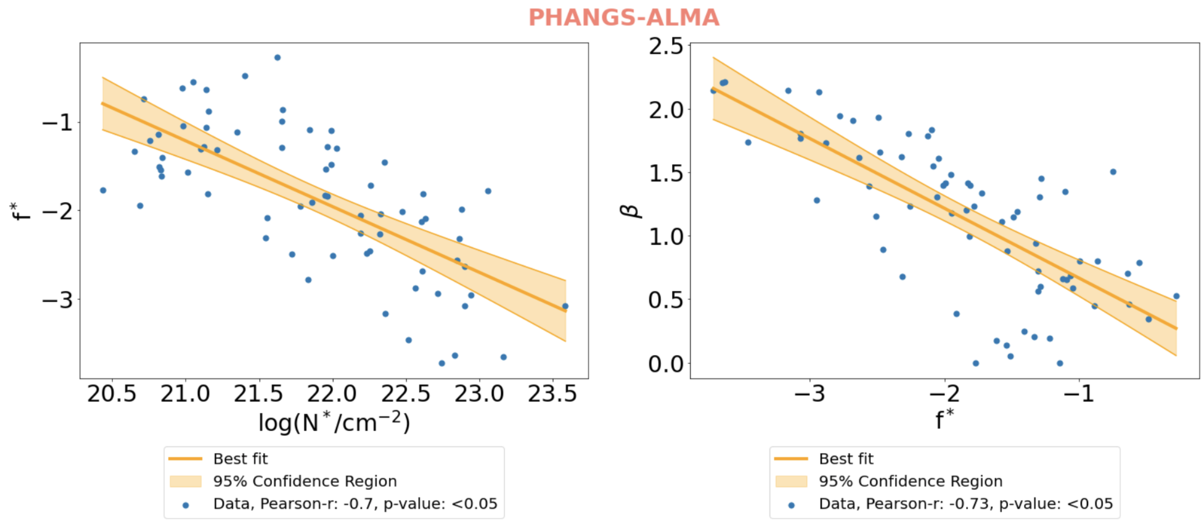

Figure B.1: Correlations between the free parameter f^* and the free parameter N^* and between f^* and the slope β of the gamma distribution fits on the $f(N_{H_2})$ in the PHANGS-ALMA sample. The blue dots indicate the fit values of individual galaxies, the orange line the best fit, and the orange band the 95% confidence region of the fit. Both samples show strong correlations with Pearson-rs of ~ 0.7 and p-values < 0.05.

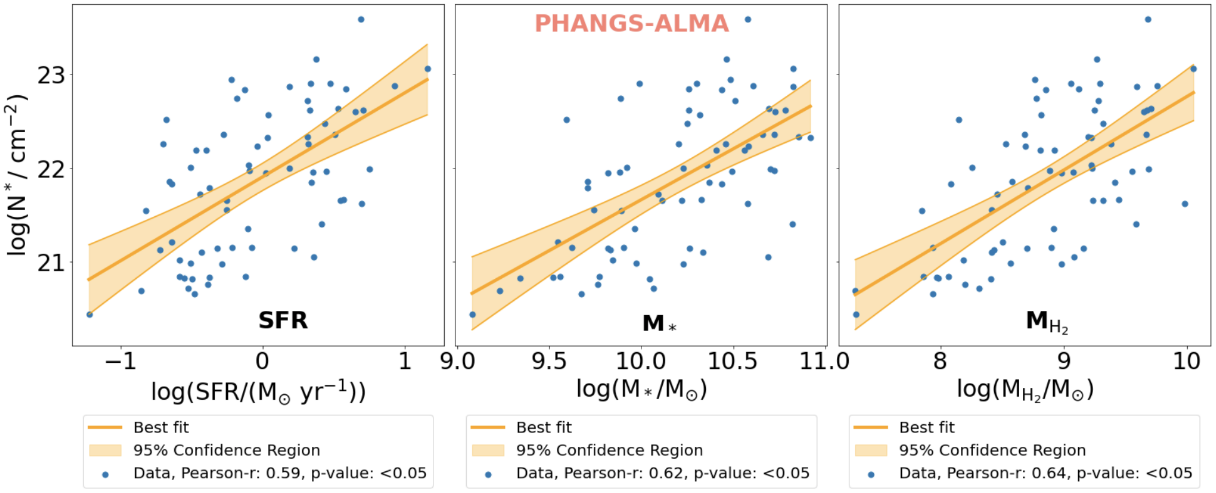

Figure B.2: Dependence on different physical properties of the free parameter N^* of the gamma distribution fits on the $f(N_{H_2})$ within the PHANGS-ALMA sample. The blue dots indicate the fit values and properties of individual objects, the orange line the best fit, and the orange band the 95% confidence region of the fit. The physical properties (SFR, M_* and M_{H_2}) of the galaxies correlate with the free parameter N^*, albeit with a significant scatter.

correlations. We note that our tests have shown that while these fits approximate the global $f(N_{\mathrm{H}_2})$ well when using the PHANGS-ALMA sample, they often fail for individual galaxies because they are degenerate. We therefore caution from using these fits to predict individual $f(N_{\mathrm{H}_2})$.

Bibliography

Abadi M., et al., 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, `https://www.tensorflow.org/`

Ahumada R., et al., 2020, ApJS, 249, 3

Alam S., et al., 2015, ApJS, 219, 12

Alpher R. A., Herman R., 1948, Nature, 162, 774

Altay G., Theuns T., Schaye J., Crighton N. H. M., Dalla Vecchia C., 2011, ApJ, 737, L37

Anand A., Nelson D., Kauffmann G., 2021, MNRAS, 504, 65

Anand A., Kauffmann G., Nelson D., 2022, MNRAS, 513, 3210

Anderson M. E., Bregman J. N., 2010, ApJ, 714, 320

Anderson M. E., Bregman J. N., Dai X., 2013, ApJ, 762, 106

Aravena M., et al., 2019, ApJ, 882, 136

Augustin R., et al., 2018, MNRAS, 478, 3120

Augustin R., et al., 2019, MNRAS, 489, 2417

Augustin R., Péroux C., Hamanowicz A., Kulkarni V., Rahmani H., Zanella A., 2021, MNRAS, 505, 6195

Ayromlou M., Nelson D., Pillepich A., 2022, arXiv e-prints, p. arXiv:2211.07659

Bacon R., et al., 2010, in McLean I. S., Ramsay S. K., Takami H., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 7735, Ground-based and Airborne Instrumentation for Astronomy III. p. 773508, doi:10.1117/12.856027

Baker W. M., et al., 2022, arXiv e-prints, p. arXiv:2210.03755

Balashev S. A., Noterdaeme P., 2018, MNRAS, 478, L7

Barnabè M., et al., 2012, MNRAS, 423, 1073

Barret D., 2022, in 44th COSPAR Scientific Assembly. Held 16-24 July. p. 2316

Becker G. D., Sargent W. L. W., Rauch M., Calverley A. P., 2011, ApJ, 735, 93

Becker G. D., et al., 2019, ApJ, 883, 163

Behroozi P. S., Conroy C., Wechsler R. H., 2010, ApJ, 717, 379

Bennett C. L., et al., 2003, ApJS, 148, 1

Bera A., Kanekar N., Chengalur J. N., Bagla J. S., 2019, ApJ, 882, L7

Bergeron J., 1986, A&A, 155, L8

Bigiel F., Leroy A., Walter F., Brinks E., de Blok W. J. G., Madore B., Thornley M. D., 2008, AJ, 136, 2846

Bird S., Vogelsberger M., Haehnelt M., Sijacki D., Genel S., Torrey P., Springel V., Hernquist L., 2014, MNRAS, 445, 2313

Bird S., Garnett R., Ho S., 2017, MNRAS, 466, 2111

Bish H. V., Werk J. K., Prochaska J. X., Rubin K. H. R., Zheng Y., O'Meara J. M., Deason A. J., 2019, ApJ, 882, 76

Blanton M. R., et al., 2017, AJ, 154, 28

Blitz L., Rosolowsky E., 2006, ApJ, 650, 933

Bluck A. F. L., Maiolino R., Sánchez S. F., Ellison S. L., Thorp M. D., Piotrowska J. M., Teimoorinia H., Bundy K. A., 2020, MNRAS, 492, 96

Bluck A. F. L., Maiolino R., Brownson S., Conselice C. J., Ellison S. L., Piotrowska J. M., Thorp M. D., 2022, A&A, 659, A160

Blumenthal G. R., Faber S. M., Primack J. R., Rees M. J., 1984, Nature, 311, 517

Boissé P., Le Brun V., Bergeron J., Deharveng J.-M., 1998, A&A, 333, 841

Bolatto A. D., Wolfire M., Leroy A. K., 2013, ARA&A, 51, 207

Bonato M., et al., 2018, MNRAS, 478, 1512

Bordoloi R., et al., 2011, ApJ, 743, 10

Bordoloi R., et al., 2014, ApJ, 794, 130

Bouché N., Murphy M. T., Péroux C., Davies R., Eisenhauer F., Förster Schreiber N. M., Tacconi L., 2007, ApJ, 669, L5

Bouché N., et al., 2012, MNRAS, 419, 2

Bouché N., Carfantan H., Schroetter I., Michel-Dansac L., Contini T., 2015, GalPaK 3D: Galaxy parameters and kinematics extraction from 3D data (ascl:1501.014)

Braun R., 2012, ApJ, 749, 87

Bregman J. N., 2007, ARA&A, 45, 221

Bregman J. N., Anderson M. E., Miller M. J., Hodges-Kluck E., Dai X., Li J.-T., Li Y., Qu Z., 2018, ApJ, 862, 3

Brook C. B., et al., 2011, MNRAS, 415, 1051

Brooks A., Christensen C., 2016, in Laurikainen E., Peletier R., Gadotti D., eds, Astrophysics and Space Science Library Vol. 418, Galactic Bulges. p. 317 (`arXiv:1511.04095`), doi:10.1007/978-3-319-19378-6˙12

Burgh E. B., France K., McCandliss S. R., 2007, ApJ, 658, 446

Busca N., Balland C., 2018, arXiv e-prints, p. arXiv:1808.09955

Byrohl C., et al., 2021, MNRAS, 506, 5129

Cantalupo S., Arrigoni-Battaia F., Prochaska J. X., Hennawi J. F., Madau P., 2014, Nature, 506, 63

Carlberg R. G., Freedman W. L., 1985, ApJ, 298, 486

Chen H.-W., Kennicutt Robert C. J., Rauch M., 2005, ApJ, 620, 703

Chollet F., et al., 2015, Keras, `https://keras.io`

Chowdhury A., Kanekar N., Chengalur J. N., Sethi S., Dwarakanath K. S., 2020, Nature, 586, 369

Chowdhury A., Kanekar N., Das B., Dwarakanath K. S., Sethi S., 2021, ApJ, 913, L24

Christensen L., Møller P., Fynbo J. P. U., Zafar T., 2014, MNRAS, 445, 225

Christensen C. R., Davé R., Governato F., Pontzen A., Brooks A., Munshi F., Quinn T., Wadsley J., 2016, ApJ, 824, 57

Clark P. C., Glover S. C. O., 2014, MNRAS, 444, 2396

Comparat J., et al., 2022, A&A, 666, A156

Conselice C. J., 2003, ApJS, 147, 1

Conselice C. J., Mortlock A., Bluck A. F. L., Grützbauch R., Duncan K., 2013, MNRAS, 430, 1051

Cooke R. J., Pettini M., Steidel C. C., 2018, ApJ, 855, 102

Corlies L., Peeples M. S., Tumlinson J., O'Shea B. W., Lehner N., Howk J. C., O'Meara J. M., Smith B. D., 2020a, ApJ, 896, 125

Corlies L., Peeples M. S., Tumlinson J., O'Shea B. W., Lehner N., Howk J. C., O'Meara J. M., Smith B. D., 2020b, ApJ, 896, 125

Cormier D., et al., 2015, A&A, 578, A53

Costa T., Arrigoni Battaia F., Farina E. P., Keating L. C., Rosdahl J., Kimm T., 2022, MNRAS, 517, 1767

Courteau S., Dutton A. A., 2015, ApJ, 801, L20

Crain R. A., et al., 2015, MNRAS, 450, 1937

Crighton N. H. M., et al., 2015, MNRAS, 452, 217

DESI Collaboration et al., 2016, arXiv e-prints, p. arXiv:1611.00036

Dalton G., et al., 2012, in McLean I. S., Ramsay S. K., Takami H., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV. p. 84460P, doi:10.1117/12.925950

Damle M., et al., 2022, MNRAS, 512, 3717

Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, MNRAS, 486, 2827

DeFelippis D., Bouché N. F., Genel S., Bryan G. L., Nelson D., Marinacci F., Hernquist L., 2021, ApJ, 923, 56

Decarli R., et al., 2016, ApJ, 833, 69

Decarli R., et al., 2019, ApJ, 882, 138

Decarli R., et al., 2020, ApJ, 902, 110

Delhaize J., Meyer M. J., Staveley-Smith L., Boyle B. J., 2013, MNRAS, 433, 1398

Dewdney P. E., Hall P. J., Schilizzi R. T., Lazio T. J. L. W., 2009, IEEE Proceedings, 97, 1482

Di Matteo T., Springel V., Hernquist L., 2005, Nature, 433, 604

Diemer B., et al., 2018, ApJS, 238, 33

Diemer B., et al., 2019, MNRAS, 487, 1529

Driver S. P., et al., 2019, The Messenger, 175, 46

Dutton A. A., Macciò A. V., 2014, MNRAS, 441, 3359

Dutton A. A., et al., 2013, MNRAS, 428, 3183

Einasto J., Saar E., Kaasik A., Chernin A. D., 1974, Nature, 252, 111

Einstein A., 1917, Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften, pp 142–152

Eisert L., Pillepich A., Nelson D., Klessen R. S., Huertas-Company M., Rodriguez-Gomez V., 2022, arXiv e-prints, p. arXiv:2202.06967

Emsellem E., et al., 2021, arXiv e-prints, p. arXiv:2110.03708

Epinat B., et al., 2009, A&A, 504, 789

Faucher-Giguère C.-A., Lidz A., Zaldarriaga M., Hernquist L., 2009, ApJ, 703, 1416

Feldmann R., 2020, Communications Physics, 3, 226

Ferland G. J., et al., 2017, Rev. Mex. Astron. Astrofis., 53, 385

Fielding E., Nyirenda C. N., Vaccari M., 2022, in 2022 International Conference on Electrical. p. 1 (`arXiv:2206.06165`), doi:10.1109/ICECET55527.2022.9872611

Finoguenov A., et al., 2019, The Messenger, 175, 39

Fixsen D. J., Bennett C. L., Mather J. C., 1999, ApJ, 526, 207

Frank S., et al., 2012, MNRAS, 420, 1731

Fraternali F., 2017, Gas Accretion via Condensation and Fountains. p. 323, doi:10.1007/978-3-319-52512-9˙14

French D. M., et al., 2021, arXiv e-prints, p. arXiv:2108.07419

Freundlich J., Bouché N. F., Contini T., Daddi E., Zabl J., Schroetter I., Boogaard L., Richard J., 2021, MNRAS, 501, 1900

Fukugita M., Hogan C. J., Peebles P. J. E., 1998, ApJ, 503, 518

Fukushima K., 1975, Biological Cybernetics, 20, 121

Fukushima K., 1980, Biological Cybernetics, 36, 193

Fumagalli M., Prochaska J. X., Kasen D., Dekel A., Ceverino D., Primack J. R., 2011, MNRAS, 418, 1796

Fumagalli M., O'Meara J. M., Prochaska J. X., 2016, MNRAS, 455, 4100

Fynbo J. P. U., et al., 2010, MNRAS, 408, 2128

Gaia Collaboration et al., 2016, A&A, 595, A1

Galvin T. J., et al., 2020, MNRAS, 497, 2730

Gamow G., 1946, Physical Review, 70, 572

Gaskin J. A., et al., 2019, Journal of Astronomical Telescopes, Instruments, and Systems, 5, 021001

Genel S., et al., 2014, MNRAS, 445, 175

Genzel R., et al., 2012, ApJ, 746, 69

Genzel R., et al., 2015, ApJ, 800, 20

Genzel R., et al., 2017, Nature, 543, 397

Genzel R., et al., 2020, ApJ, 902, 98

Gheller C., Vazza F., 2022, MNRAS, 509, 990

Ginolfi M., et al., 2020, A&A, 633, A90

Glover S. C. O., Clark P. C., 2012, MNRAS, 421, 116

Glover S. C. O., Mac Low M.-M., 2007a, ApJS, 169, 239

Glover S. C. O., Mac Low M.-M., 2007b, ApJ, 659, 1317

Gnedin N. Y., Kravtsov A. V., 2011, ApJ, 728, 88

Gnedin N. Y., Tassis K., Kravtsov A. V., 2009, ApJ, 697, 55

Gong M., Ostriker E. C., Wolfire M. G., 2017, ApJ, 843, 38

Grand R. J. J., et al., 2017, MNRAS, 467, 179

Guillemin P., Bergeron J., 1997, A&A, 328, 499

Gurney K., 1997, An Introduction to Neural Networks. An Introduction to Neural Networks, Taylor & Francis, `https://books.google.de/books?id=HOsvllRMMP8C`

Guth A. H., 1981, Phys. Rev. D, 23, 347

Hafen Z., et al., 2017, MNRAS, 469, 2292

Hamanowicz A., et al., 2020, MNRAS, 492, 2347

Hamanowicz A., et al., 2022, arXiv e-prints, p. arXiv:2211.00066

Harikane Y., et al., 2020, ApJ, 896, 93

Hashimoto T., Inoue A. K., Tamura Y., Matsuo H., Mawatari K., Yamaguchi Y., 2019, PASJ, 71, 109

Hawking S. W., 1982, Physics Letters B, 115, 295

He Z., Li N., 2022, Research in Astronomy and Astrophysics, 22, 095021

He K., Zhang X., Ren S., Sun J., 2015, arXiv e-prints, p. arXiv:1512.03385

Heeschen D. S., 1981, in Burbidge G., Hewitt A., eds, , Telescopes for the 1980s, Annual Reviews Monograph. pp 1–61

Heintz K. E., Watson D., Oesch P. A., Narayanan D., Madden S. C., 2021, ApJ, 922, 147

Heintz K. E., et al., 2022, ApJ, 934, L27

Hislop J. M., Naab T., Steinwandel U. P., Lahén N., Irodotou D., Johansson P. H., Walch S., 2021, arXiv e-prints, p. arXiv:2109.08160

Ho T. K., 1995, in Proceedings of 3rd international conference on document analysis and recognition. pp 278–282

Ho M.-F., Bird S., Garnett R., 2021, MNRAS, 507, 704

Holmberg E., 1958, Meddelanden fran Lunds Astronomiska Observatorium Serie II, 136, 1

Hotan A. W., et al., 2021, Publ. Astron. Soc. Australia, 38, e009

Hu C.-Y., Naab T., Walch S., Moster B. P., Oser L., 2014a, MNRAS, 443, 1173

Hu C.-Y., Naab T., Walch S., Moster B. P., Oser L., 2014b, MNRAS, 443, 1173

Hu C.-Y., Naab T., Walch S., Glover S. C. O., Clark P. C., 2016, MNRAS, 458, 3528

Hu C.-Y., Naab T., Glover S. C. O., Walch S., Clark P. C., 2017, MNRAS, 471, 2151

Hu C.-Y., Sternberg A., van Dishoeck E. F., 2021, arXiv e-prints, p. arXiv:2103.03889

Hubel D., Wiesel T., 1959, J Physiol., 148, 574

Hubel D., Wiesel T., 1962, J Physiol., 160, 54

Hummels C. B., Smith B. D., Silvia D. W., 2017, ApJ, 847, 59

Husemann B., Bennert V. N., Scharwächter J., Woo J. H., Choudhury O. S., 2016, MNRAS, 455, 1905

Ivezic Z., et al., 2008, Serbian Astronomical Journal, 176, 1

Jonas J., MeerKAT Team 2016, in MeerKAT Science: On the Pathway to the SKA. p. 1

Jones M. G., Haynes M. P., Giovanelli R., Moorman C., 2018a, MNRAS, 477, 2

Jones M. G., et al., 2018b, A&A, 609, A17

Joseph R. D., Wright G. S., 1985, MNRAS, 214, 87

Kacprzak G. G., Muzahid S., Churchill C. W., Nielsen N. M., Charlton J. C., 2015, ApJ, 815, 22

Kanekar N., Smette A., Briggs F. H., Chengalur J. N., 2009, ApJ, 705, L40

Kanekar N., et al., 2018, ApJ, 856, L23

Kanekar N., Prochaska J. X., Neeleman M., Christensen L., Møller P., Zwaan M. A., Fynbo J. P. U., Dessauges-Zavadsky M., 2020, ApJ, 901, L5

Katz H., et al., 2022, MNRAS,

Kennicutt R. C., Evans N. J., 2012, ARA&A, 50, 531

Kingma D. P., Ba J., 2014, arXiv e-prints, p. arXiv:1412.6980

Klitsch A., Péroux C., Zwaan M. A., Smail I., Oteo I., Biggs A. D., Popping G., Swinbank A. M., 2018, MNRAS, 475, 492

Klitsch A., et al., 2019a, MNRAS, 482, L65

Klitsch A., et al., 2019b, MNRAS, 490, 1220

Klitsch A., Péroux C., Zwaan M. A., De Cia A., Ledoux C., Lopez S., 2021, MNRAS, 506, 514

Kobulnicky H. A., Kennicutt Robert C. J., Pizagno J. L., 1999, ApJ, 514, 544

Krizhevsky A., Sutskever I., Hinton G. E., 2012, in Advances in neural information processing systems. pp 1097–1105

Krogager J.-K., Noterdaeme P., 2020, A&A, 644, L6

Krogager J. K., Møller P., Fynbo J. P. U., Noterdaeme P., 2017, MNRAS, 469, 2959

Krone-Martins A., Moitinho A., 2014, A&A, 561, A57

Krumholz M. R., 2013, MNRAS, 436, 2747

Krumholz M. R., McKee C. F., Tumlinson J., 2008, ApJ, 689, 865

Kulkarni V. P., Hill J. M., Schneider G., Weymann R. J., Storrie-Lombardi L. J., Rieke M. J., Thompson R. I., Jannuzi B. T., 2000, ApJ, 536, 36

Kulkarni V. P., Hill J. M., Schneider G., Weymann R. J., Storrie-Lombardi L. J., Rieke M. J., Thompson R. I., Jannuzi B. T., 2001, ApJ, 551, 37

Lagos C. d. P., et al., 2015, MNRAS, 452, 3815

Lahén N., Naab T., Johansson P. H., Elmegreen B., Hu C.-Y., Walch S., 2019, ApJ, 879, L18

Lahén N., Naab T., Johansson P. H., Elmegreen B., Hu C.-Y., Walch S., Steinwandel U. P., Moster B. P., 2020a, ApJ, 891, 2

Lahén N., Naab T., Johansson P. H., Elmegreen B., Hu C.-Y., Walch S., 2020b, ApJ, 904, 71

Lane W. M., Briggs F. H., Turnshek D. A., Rao S. M., 1998, in American Astronomical Society Meeting Abstracts. p. 04.09

Lanzetta K. M., Bowen D., 1990, ApJ, 357, 321

Lara-Benítez P., Carranza-García M., Riquelme J. C., 2021, International Journal of Neural Systems, 31, 2130001

LeCun Y., Bengio Y., Hinton G., 2015, nature, 521, 436

Lecun Y., Bottou L., Bengio Y., Haffner P., 1998, Proceedings of the IEEE, 86, 2278

Ledoux C., Petitjean P., Srianand R., 2003, MNRAS, 346, 209

Lee J. C., et al., 2021, arXiv e-prints, p. arXiv:2101.02855

Lehner N., Wotta C. B., Howk J. C., O'Meara J. M., Oppenheimer B. D., Cooksey K. L., 2019, ApJ, 887, 5

Leroy A. K., Walter F., Brinks E., Bigiel F., de Blok W. J. G., Madore B., Thornley M. D., 2008, AJ, 136, 2782

Leroy A. K., et al., 2011, ApJ, 737, 12

Leroy A. K., et al., 2013, AJ, 146, 19

Leroy A. K., et al., 2021, arXiv e-prints, p. arXiv:2104.07739

Levy R. C., et al., 2018, ApJ, 860, 92

Li J., et al., 2021, ApJ, 922, L29

Li C., et al., 2022, MNRAS, 509, 2289

Linde A. D., 1982, Physics Letters B, 116, 335

Liou C.-Y., Cheng W.-C., Liou J.-W., Liou D.-R., 2014, Neurocomputing, 139, 84

Liu D., et al., 2019, ApJ, 887, 235

Loiacono F., Talia M., Fraternali F., Cimatti A., Di Teodoro E. M., Caminha G. B., 2019, MNRAS, 489, 681

Lovell M. R., et al., 2018, MNRAS, 481, 1950

Madau P., Dickinson M., 2014, ARA&A, 52, 415

Madden S. C., et al., 2020, A&A, 643, A141

Maio U., Péroux C., Ciardi B., 2022, A&A, 657, A47

Man A. W. S., Toft S., Zirm A. W., Wuyts S., van der Wel A., 2012, ApJ, 744, 85

Mangum J. G., Shirley Y. L., 2015, PASP, 127, 266

Marinacci F., et al., 2018, MNRAS, 480, 5113

Martin C. L., Shapley A. E., Coil A. L., Kornei K. A., Bundy K., Weiner B. J., Noeske K. G., Schiminovich D., 2012, ApJ, 760, 127

Martin C. L., Ho S. H., Kacprzak G. G., Churchill C. W., 2019, ApJ, 878, 84

Martín S., et al., 2021, A&A, 656, A46

Martinsson T. P. K., Verheijen M. A. W., Westfall K. B., Bershady M. A., Andersen D. R., Swaters R. A., 2013, A&A, 557, A131

Mas-Ribas L., et al., 2017, ApJ, 846, 4

McKee C. F., Ostriker E. C., 2007, ARA&A, 45, 565

McMullin J. P., Waters B., Schiebel D., Young W., Golap K., 2007, in Shaw R. A., Hill F., Bell D. J., eds, Astronomical Society of the Pacific Conference Series Vol. 376, Astronomical Data Analysis Software and Systems XVI. p. 127

Merloni A., et al., 2019, The Messenger, 175, 42

Mitchell T. M., 1997, Machine learning, International Edition. McGraw-Hill Series in Computer Science, McGraw-Hill, `https://www.worldcat.org/oclc/61321007`

Mo H. J., White S. D. M., 2002, MNRAS, 336, 112

Molina J., Ibar E., Smail I., Swinbank A. M., Villard E., Escala A., Sobral D., Hughes T. M., 2019, MNRAS, 487, 4856

Molina J., et al., 2020, A&A, 643, A78

Moore B., Katz N., Lake G., Dressler A., Oemler A., 1996, Nature, 379, 613

Mortlock A., et al., 2013, MNRAS, 433, 1185

Moster B. P., Somerville R. S., Maulbetsch C., van den Bosch F. C., Macciò A. V., Naab T., Oser L., 2010, ApJ, 710, 903

Moster B. P., Naab T., White S. D. M., 2018, MNRAS, 477, 1822

Murphy K., 2013, Machine Learning: A Probabilistic Perspective. Adaptive computation and machine learning series, Mit Press, `/bib/murphy/murphy2012machine/Machine%20Learning%20A%20Probabilistic%20Perspective.pdf,http://books.google.com/books?id=0In7ugAACAAJ,http://www.cs.ubc.ca/~murphyk/MLbook/,/bib/murphy/murphy2012machine/Solutions%20Manual.pdf`

Muzahid S., Srianand R., Charlton J., 2015, MNRAS, 448, 2840

Muzahid S., Kacprzak G. G., Charlton J. C., Churchill C. W., 2016, ApJ, 823, 66

Muzahid S., et al., 2020, MNRAS, 496, 1013

Muzahid S., et al., 2021, MNRAS, 508, 5612

Møller P., et al., 2017, MNRAS, 474, 4039

Naiman J. P., et al., 2018, MNRAS, 477, 1206

Narayanan A., Misawa T., Charlton J. C., Kim T.-S., 2007, ApJ, 660, 1093

Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493

Neeleman M., et al., 2016, ApJ, 820, L39

Neeleman M., Kanekar N., Prochaska J. X., Rafelski M., Carilli C. L., Wolfe A. M., 2017, Science, 355, 1285

Neeleman M., Kanekar N., Prochaska J. X., Christensen L., Dessauges-Zavadsky M., Fynbo J. P. U., Møller P., Zwaan M. A., 2018, ApJ, 856, L12

Neeleman M., Kanekar N., Prochaska J. X., Rafelski M. A., Carilli C. L., 2019, ApJ, 870, L19

Neeleman M., Prochaska J. X., Kanekar N., Rafelski M., 2020, Nature, 581, 269

Nelson R. P., Langer W. D., 1997, ApJ, 482, 796

Nelson D., et al., 2018, MNRAS, 475, 624

Nelson D., et al., 2019a, Computational Astrophysics and Cosmology, 6, 2

Nelson D., et al., 2019b, MNRAS, 490, 3234

Nelson D., et al., 2020, MNRAS,

Nelson D., Byrohl C., Peroux C., Rubin K. H. R., Burchett J. N., 2021, MNRAS, 507, 4445

Nestor D. B., Turnshek D. A., Rao S. M., 2005, ApJ, 628, 637

Nicastro F., et al., 2018, Nature, 558, 406

Nilson P., 1973, Uppsala general catalogue of galaxies. Uppsala Astron. Obs.

Noterdaeme P., Ledoux C., Petitjean P., Srianand R., 2008, A&A, 481, 327

Noterdaeme P., Petitjean P., Ledoux C., Srianand R., 2009, A&A, 505, 1087

Noterdaeme P., et al., 2012, A&A, 547, L1

Olsen K., Greve T. R., Narayanan D., Thompson R., Davé R., Niebla Rios L., Stawinski S., 2017, ApJ, 846, 105

Olsen K. P., et al., 2021, ApJ, 922, 88

Ostriker J. P., Peebles P. J. E., Yahil A., 1974, ApJ, 193, L1

Oteo I., Zwaan M. A., Ivison R. J., Smail I., Biggs A. D., 2016, ApJ, 822, 36

Pakmor R., Springel V., 2013, MNRAS, 432, 176

Pakmor R., Bauer A., Springel V., 2011, MNRAS, 418, 1392

Papadopoulos P. P., van der Werf P., Xilouris E., Isaak K. G., Gao Y., 2012, ApJ, 751, 10

Parks D., Prochaska J. X., Dong S., Cai Z., 2018, MNRAS, 476, 1151

Peebles P. J. E., 1982, ApJ, 263, L1

Peng Y.-j., et al., 2010, ApJ, 721, 193

Peng Y., Maiolino R., Cochrane R., 2015, Nature, 521, 192

Penzias A. A., Wilson R. W., 1965, ApJ, 142, 419

Perivolaropoulos L., Skara F., 2022, New Astron. Rev., 95, 101659

Péroux C., Howk J. C., 2020, ARA&A, 58, 363

Péroux C., McMahon R. G., Storrie-Lombardi L. J., Irwin M. J., 2003, MNRAS, 346, 1103

Péroux C., Dessauges-Zavadsky M., D'Odorico S., Sun Kim T., McMahon R. G., 2005, MNRAS, 363, 479

Péroux C., Bouché N., Kulkarni V. P., York D. G., Vladilo G., 2011b, MNRAS, 410, 2237

Péroux C., Bouché N., Kulkarni V. P., York D. G., Vladilo G., 2011a, MNRAS, 410, 2237

Péroux C., Bouché N., Kulkarni V. P., York D. G., Vladilo G., 2011c, MNRAS, 410, 2251

Péroux C., et al., 2016, MNRAS, 457, 903

Péroux C., et al., 2017, MNRAS, 464, 2053

Péroux C., et al., 2019, MNRAS, 485, 1595

Péroux C., Nelson D., van de Voort F., Pillepich A., Marinacci F., Vogelsberger M., Hernquist L., 2020, MNRAS, 499, 2462

Péroux C., et al., 2022, MNRAS, 516, 5618

Pessa I., et al., 2021, A&A, 650, A134

Pettini M., Pagel B. E. J., 2004, MNRAS, 348, L59

Pillepich A., et al., 2018a, MNRAS, 473, 4077

Pillepich A., et al., 2018b, MNRAS, 473, 4077

Pillepich A., et al., 2018c, MNRAS, 475, 648

Pillepich A., et al., 2019, MNRAS, 490, 3196

Pillepich A., Nelson D., Truong N., Weinberger R., Martin-Navarro I., Springel V., Faber S. M., Hernquist L., 2021, MNRAS, 508, 4667

Planck Collaboration et al., 2016, A&A, 594, A13

Planck Collaboration et al., 2020, A&A, 641, A1

Popping G., et al., 2019, ApJ, 882, 137

Predehl P., et al., 2021, A&A, 647, A1

Price S. H., et al., 2020, ApJ, 894, 91

Prochaska J. X., O'Meara J. M., Worseck G., 2010, ApJ, 718, 392

Puech M., et al., 2008, A&A, 484, 173

Rahmani H., et al., 2017, MNRAS, 474, 254

Rahmani H., et al., 2018a, MNRAS, 474, 254

Rahmani H., et al., 2018b, MNRAS, 480, 5046

Rahmani H., et al., 2018c, MNRAS, 480, 5046

Rahmati A., Pawlik A. H., Raičević M., Schaye J., 2013, MNRAS, 430, 2427

Ramesh R., Nelson D., Pillepich A., 2022, arXiv e-prints, p. arXiv:2211.00020

Ramos Padilla A. F., Wang L., Ploeckinger S., van der Tak F. F. S., Trager S. C., 2021, A&A, 645, A133

Rao S. M., Turnshek D. A., Nestor D. B., 2006, ApJ, 636, 610

Rao S. M., Turnshek D. A., Sardane G. M., Monier E. M., 2017, MNRAS, 471, 3428

Rathjen T.-E., et al., 2021, MNRAS, 504, 1039

Rhee J., Lah P., Chengalur J. N., Briggs F. H., Colless M., 2016, MNRAS, 460, 2675

Richard J., et al., 2019, The Messenger, 175, 50

Richings A. J., Schaye J., 2016, MNRAS, 458, 270

Riechers D. A., et al., 2019, ApJ, 872, 7

Riess A. G., et al., 1998, AJ, 116, 1009

Rose C., et al., 2022, arXiv e-prints, p. arXiv:2208.11164

Röttgers B., Naab T., Cernetic M., Davé R., Kauffmann G., Borthakur S., Foidl H., 2020, MNRAS, 496, 152

Rubin V. C., Ford W. Kent J., 1970, ApJ, 159, 379

Rubin K. H. R., Prochaska J. X., Koo D. C., Phillips A. C., 2012, ApJ, 747, L26

Rudie G. C., et al., 2012, ApJ, 750, 67

Rudie G. C., Newman A. B., Murphy M. T., 2017, ApJ, 843, 98

Saintonge A., Catinella B., 2022, arXiv e-prints, p. arXiv:2202.00690

Saintonge A., et al., 2011, MNRAS, 415, 32

Saintonge A., et al., 2017, ApJS, 233, 22

Sánchez-Menguiano L., et al., 2016, A&A, 587, A70

Sánchez S. F., et al., 2014, A&A, 563, A49

Sánchez S. F., et al., 2019, MNRAS, 484, 3042

Schaye J., 2001, ApJ, 562, L95

Schaye J., Aguirre A., Kim T.-S., Theuns T., Rauch M., Sargent W. L. W., 2003, ApJ, 596, 768

Schaye J., et al., 2015, MNRAS, 446, 521

Schiminovich D., et al., 2010, MNRAS, 408, 919

Schinnerer E., et al., 2019, ApJ, 887, 49

Schroetter I., et al., 2016, ApJ, 833, 39

Schroetter I., et al., 2019, MNRAS, 490, 4368

Schroetter I., et al., 2021, MNRAS, 506, 1355

Schruba A., et al., 2011, AJ, 142, 37

Schweizer F., 1989, Nature, 338, 119

Scoville N., et al., 2017, ApJ, 837, 150

Seifried D., et al., 2017, MNRAS, 472, 4797

Sellwood J. A., 2014, Reviews of Modern Physics, 86, 1

Seyffert E. N., Cooksey K. L., Simcoe R. A., O'Meara J. M., Kao M. M., Prochaska J. X., 2013, ApJ, 779, 161

Shanks T., Boyle B. J., Croom S., Loaring N., Miller L., Smith R. J., 2000, in Mazure A., Le Fèvre O., Le Brun V., eds, Astronomical Society of the Pacific Conference Series Vol. 200, Clustering at High Redshift. p. 57 (`arXiv:astro-ph/0003206`)

Shen S., Madau P., Aguirre A., Guedes J., Mayer L., Wadsley J., 2012, ApJ, 760, 50

Shen H., George D., Huerta E. A., Zhao Z., 2017, arXiv e-prints, p. arXiv:1711.09919

Shi J.-H., Qiu B., Luo A. L., He Z.-D., Kong X., Jiang X., 2022, MNRAS, 516, 264

Shull J. M., Smith B. D., Danforth C. W., 2012, ApJ, 759, 23

Shull J. M., Danforth C. W., Tilton E. M., 2014, ApJ, 796, 49

Sijacki D., Vogelsberger M., Genel S., Springel V., Torrey P., Snyder G. F., Nelson D., Hernquist L., 2015, MNRAS, 452, 575

Silver D., et al., 2016, Nature, 529, 484

Snoek J., Larochelle H., Adams R. P., 2012, arXiv e-prints, p. arXiv:1206.2944

Solomon P. M., Downes D., Radford S. J. E., 1992, ApJ, 398, L29

Som D., Kulkarni V. P., Meiring J., York D. G., Péroux C., Lauroesch J. T., Aller M. C., Khare P., 2015, ApJ, 806, 25

Spilker A., Kainulainen J., Orkisz J., 2021, arXiv e-prints, p. arXiv:2108.04518

Springel V., 2005, MNRAS, 364, 1105

Springel V., 2010, MNRAS, 401, 791

Springel V., Hernquist L., 2003, MNRAS, 339, 289

Springel V., Di Matteo T., Hernquist L., 2005, MNRAS, 361, 776

Springel V., et al., 2018, MNRAS, 475, 676

Starobinsky A. A., 1982, Physics Letters B, 117, 175

Steidel C. C., Erb D. K., Shapley A. E., Pettini M., Reddy N., Bogosavljević M., Rudie G. C., Rakic O., 2010, ApJ, 717, 289

Steinwandel U. P., Moster B. P., Naab T., Hu C.-Y., Walch S., 2020, MNRAS, 495, 1035

Stevens A. R. H., Diemer B., Lagos C. d. P., Nelson D., Obreschkow D., Wang J., Marinacci F., 2019, Monthly Notices of the Royal Astronomical Society, 490, 96–113

Stewart K. R., Kaufmann T., Bullock J. S., Barton E. J., Maller A. H., Diemand J., Wadsley J., 2011, ApJ, 738, 39

Stinson G. S., et al., 2012, MNRAS, 425, 1270

Sun J., et al., 2018, ApJ, 860, 172

Sun J., et al., 2020, ApJ, 901, L8

Sun J., et al., 2022, AJ, 164, 43

Suresh J., Rubin K. H. R., Kannan R., Werk J. K., Hernquist L., Vogelsberger M., 2017, MNRAS, 465, 2966

Szakacs R., et al., 2021, MNRAS, 505, 4746

Szakacs R., Péroux C., Zwaan M. A., Nelson D., Schinnerer E., Lahén N., Weng S., Fresco A. Y., 2022, MNRAS, 512, 4736

Tacconi L. J., et al., 2018, ApJ, 853, 179

Tacconi L. J., Genzel R., Sternberg A., 2020, ARA&A, 58, 157

Teng Y.-H., et al., 2022, ApJ, 925, 72

Theuns T., Leonard A., Efstathiou G., 1998, MNRAS, 297, L49

Toomre A., 1977, in Tinsley B. M., Larson Richard B. Gehret D. C., eds, Evolution of Galaxies and Stellar Populations. p. 401

Tremonti C. A., et al., 2004, ApJ, 613, 898

Tripp T. M., Lu L., Savage B. D., 1998, ApJ, 508, 200

Tully R. B., Fisher J. R., 1977, A&A, 500, 105

Tumlinson J., et al., 2011, ApJ, 733, 111

Tumlinson J., Peeples M. S., Werk J. K., 2017, ARA&A, 55, 389

Turner M. L., Schaye J., Steidel C. C., Rudie G. C., Strom A. L., 2014, MNRAS, 445, 794

Turner M. L., Schaye J., Crain R. A., Rudie G., Steidel C. C., Strom A., Theuns T., 2017, MNRAS, 471, 690

Tytler D., 1982, Nature, 298, 427

Übler H., et al., 2018, ApJ, 854, L24

Übler H., et al., 2020, MNRAS,

Umehata H., et al., 2019, Science, 366, 97

Valentino F., et al., 2018, ApJ, 869, 27

Vallini L., Ferrara A., Pallottini A., Carniani S., Gallerani S., 2021, MNRAS, 505, 5543

Van Rossum G., Drake F. L., 2009, Python 3 Reference Manual. CreateSpace, Scotts Valley, CA

Veilleux S., Cecil G., Bland-Hawthorn J., 2005, ARA&A, 43, 769

Villaescusa-Navarro F., et al., 2018, ApJ, 866, 135

Vizgan D., et al., 2022a, ApJ, 929, 92

Vizgan D., Heintz K. E., Greve T. R., Narayanan D., Davé R., Olsen K. P., Popping G., Watson D., 2022b, ApJ, 939, L1

Vogelsberger M., et al., 2014a, MNRAS, 444, 1518

Vogelsberger M., et al., 2014b, Nature, 509, 177

Walch S., et al., 2015, MNRAS, 454, 238

Walter F., et al., 2016, ApJ, 833, 67

Walter F., et al., 2020, ApJ, 902, 111

Wang B., et al., 2022, ApJS, 259, 28

Wehrens R., Buydens L. M. C., 2007, Journal of Statistical Software, 21, 1

Weigel A. K., Schawinski K., Bruderer C., 2016, MNRAS, 459, 2150

Weilbacher P. M., Streicher O., Palsa R., 2016, MUSE-DRP: MUSE Data Reduction Pipeline (ascl:1610.004)

Weinberg D. H., Davé R., Katz N., Kollmeier J. A., 2003, in Holt S. H., Reynolds C. S., eds, American Institute of Physics Conference Series Vol. 666, The Emergence of Cosmic Structure. pp 157–169 (`arXiv:astro-ph/0301186`), doi:10.1063/1.1581786

Weinberger R., et al., 2017, MNRAS, 465, 3291

Wendt M., Bouché N. F., Zabl J., Schroetter I., Muzahid S., 2021, MNRAS, 502, 3733

Werk J. K., Prochaska J. X., Thom C., Tumlinson J., Tripp T. M., O'Meara J. M., Peeples M. S., 2013, ApJS, 204, 17

Werk J. K., et al., 2014, ApJ, 792, 8

Werk J. K., et al., 2016, ApJ, 833, 54

White S. D. M., Rees M. J., 1978, MNRAS, 183, 341

Wisotzki L., et al., 2016, A&A, 587, A98

Wisotzki L., et al., 2018, Nature, 562, 229

Wolfe A. M., Turnshek D. A., Smith H. E., Cohen R. D., 1986, ApJS, 61, 249

Wolfe A. M., Gawiser E., Prochaska J. X., 2005, ARA&A, 43, 861

Wolfire M. G., Hollenbach D., McKee C. F., 2010, ApJ, 716, 1191–1207

Wootten A., Thompson A. R., 2009, IEEE Proceedings, 97, 1463

XRISM Science Team 2020, arXiv e-prints, p. arXiv:2003.04962

Xia I., Ge J., Willis K., Zhao Y., 2022, arXiv e-prints, p. arXiv:2210.04896

Yamashita R., Nishio M., Do R., Togashi K., 2018, Insights Imaging., 9, 611

Yates R. M., Péroux C., Nelson D., 2021, MNRAS, 508, 3535

York D. G., et al., 2000, AJ, 120, 1579

Zabl J., et al., 2019, MNRAS, 485, 1961

Zabl J., et al., 2020, MNRAS, 492, 4576

Zafar T., Péroux C., Popping A., Milliard B., Deharveng J. M., Frank S., 2013, A&A, 556, A141

Zanella A., et al., 2018, MNRAS, 481, 1976

Zhao Y., Ge J., Yuan X., Zhao T., Wang C., Li X., 2019, MNRAS, 487, 801

Zheng Y., et al., 2020, ApJ, 896, 143

Zhu G., Ménard B., 2013, ApJ, 770, 130

Zwaan M. A., 2000, PhD thesis, -

Zwaan M. A., Prochaska J. X., 2006, ApJ, 643, 675

Zwaan M. A., van der Hulst J. M., Briggs F. H., Verheijen M. A. W., Ryan-Weber E. V., 2005, MNRAS, 364, 1467

Zwaan M., et al., 2022, The Messenger, 186, 10

Zwicky F., 1933, Helvetica Physica Acta, 6, 110

de Jong R. S., et al., 2019, The Messenger, 175, 3

den Brok J. S., et al., 2021, MNRAS, 504, 3221

van de Voort F., Springel V., Mandelker N., van den Bosch F. C., Pakmor R., 2019, MNRAS, 482, L85

van der Hulst J. M., van Albada T. S., Sancisi R., 2001, in Hibbard J. E., Rupen M., van Gorkom J. H., eds, Astronomical Society of the Pacific Conference Series Vol. 240, Gas and Galaxy Evolution. p. 451

Acknowledgements

> *You can trust us to stick to you through thick and thin- to the bitter end.*
> *And you can trust us to keep any secret of yours- closer than you keep it yourself.*
> *But you cannot trust us to let you face trouble alone and go off without a word.*
>
> - Merry Brandybuck in The Fellowship of the Ring by J.R.R. Tolkien

After roughly three years, I have arrived at the end of my Ph.D. journey. It was a fun and exciting, but also challenging journey. During this time, I have learned a lot about many areas of astrophysics and a lot about myself as well. This resulting thesis, would not have been possible without the support of so many people. I will try to be exhaustive with my acknowledgments, but I am very worried that I might forget someone in this stressful time just before the submission of the thesis. So if you read this, and you are not mentioned even though you should be, I hope that you can forgive me for my mistake. Let me know if this happened and I will print a version of this thesis with you included in the acknowledgments and, depending on where you are, ship it to you or hand it to you personally as an apology[1].

First and foremost I would like to thank my advisors at ESO: Céline and Martin. I want to thank you for your support, encouragement, helpful advice, and kindness. You have taught me so many things during these last years while also being patient with me. You helped me mature as a researcher by teaching me how to write papers, how to give good talks, and how to write proposals and applications. Additionally, you also taught me a lot about life in general through our chats over these last years. I am truly happy that you two were my advisors.

Thanks also go out to Prof. Dr. Volker Springel. Thank you for overseeing my Ph.D. as my supervisor at LMU, for your valuable insights, and for your support.

I would also like to thank the members of my thesis committee for overseeing my progress during these last three years. Thank you for the great atmosphere during our meetings, and your advice and support.

I also want to thank all the members of the doctoral examination commission. Thank you for taking time out of your surely busy schedules to be part of it.

---

[1] This offer is not valid in cases were I specifically refrain from stating names.

Thank you to everyone who was proofreading this thesis trying to scour out the even most well-hidden typo (even though some are likely still hiding in here): Patrica - the awesome sister, Jannik - the grammar shaman, Malwin - the fast corrector, and Simon - the watchful eye, Akash - the great typo terminator.

I want to thank Nicola for being my ESO fellow mentor. You supported me throughout this time and it was great to chat with you about science, academia, and life. The same thanks go out to Michele who I feel was my second unofficial fellow mentor. Thank you for your encouragement and support.

Thanks go out to ESO and the ESO community in general. Thanks for giving me the opportunity to do my Ph.D. at such a wonderful place, for making me feel welcome, and for the great discussions. Special thanks to Nelma, one of the most helpful people I have ever met in my life. There was no issue she could not solve. I also want to thank Giacomo. Thank you for your support and compassion in times of need.

Further, thanks go out to the ESO and Garching student community. Especially Avinash, Pierre, Aish, Alex, Simon, Jianhang, Dominika, Alejandra, and many more. Thank you for all the fun times and chats. Especially thanks to Avinash for the almost daily chats over coffee. I will miss those.

I also want to thank all of my friends and my partner who not only supported and encouraged me during these last three years, but also throughout my life. You say you want names? I plead the fifth. You all know who you are and I love all of you. Thank you for all the fun times, trips, advice, support, chats and so much more. You all enrich my life in so many ways and I cannot imagine one without you.

Special thanks go out to my family. You believed in my abilities even in times when I did not do so myself, encouraged me, and supported me even through the most difficult times in my life. Thank you for all of your love and compassion throughout the years. I could not have done any of this without you. I love you, and I am truly grateful for all of you.