Measuring gold molecular gas across cosmic time

Frias Castillo, M.

1 | Introduction

1.1. The Interstellar Medium

The interstellar medium (ISM) is comprised of all the gas, dust and cosmic rays (CRs) that exist in the space between stars within a galaxy. It is a complex and dynamic environment, and the wealth of physical and chemical processes that take place in it make it a rich source of information that is challenging to model. It is characterized by different phases, each with distinct physical properties, densities and temperatures. These phases often coexist and interact within galaxies through physical processes such as cooling or shock waves, contributing to the complex dynamics of the interstellar medium. The cold and warm neutral medium (CNM, WNM) primarily consists of neutral atomic hydrogen (HI) and have temperatures of $\sim$100 and 6000 K and densities of $\sim 10^{1-2}$ and 10^{-1-0} cm^{-3}, respectively. The hot ionised medium (HIM), with temperatures in excess of 100000 K but much lower densities compared to the CNM and the WNM, is composed of highly ionised gas, including helium (He) and other elements. Last but not least, we have (giant) molecular clouds (GMCs). These regions have very low temperatures ($\sim$10-30 K) and high gas densities of $\sim 10^{2-6}$ cm^{-3}. They have sizes ranging from a few parsecs to about 100 pc (e.g., Larson 1981; Solomon et al. 1987; McKee & Ostriker 2007), and are large reservoirs of molecular hydrogen (H_2) and other molecules such as carbon monoxide and water. They are also composed of dust, small solid particles – mostly less than a few microns in size – mixed in with the gas. Molecular clouds can become gravitationally unstable, fragmenting into smaller, denser clumps and ultimately forming a new generation of stars. Observations show strong correlations between star formation rates or star–formation surface densities (on galactic or subgalactic scales, respectively) and the H_2 mass or surface density (Schmidt 1959; Kennicutt 1998; Bigiel et al. 2008; Leroy et al. 2008, 2013), supporting the long–standing picture that cold molecular gas is the main driver of star formation in galaxies. Overall, the physical processes involved in star formation are intimately linked to the formation and growth of galaxies, and dictate the chemical composition of subsequent generations of stars and the dynamics and overall structure of the ISM. Measuring star formation and understanding its driving mechanisms, as well as their evolution through cosmic time, is therefore one of the most critical and complex fields of study in astrophysics.

1.1.1. Star Formation

A multitude of multiwavelength surveys of large samples of galaxies have provided a remarkable census of star formation rates and stellar masses over the last two decades. Two major results arise from these studies. The first is the evolution of the cosmic star formation rate density of the Universe, shown in Fig.1.1. Star formation steadily rose from the epoch of reionisation (cosmic dawn) until it peaked at $z = 1 - 3$, when the majority of the Universe's stars formed. Subsequently, it has gradually declined by about a factor of 8 up to the present day. This picture has been possible thanks to measurements of ultraviolet (UV) and far–infrared (FIR) emission: the former directly traces the direct emission coming from young stars, while the latter allows us to see the starlight that has been absorbed and then re–emitted in dust–enshrouded regions. In practice, this means that the CSFRD is only well-constrained out to $z \sim 3$ – while UV measurements extend out to $z \sim$10, FIR data is severely limited above $z \sim 3 - 4$ and it is unclear exactly how much dust–obscured star formation is currently unaccounted for. Therefore, our current picture of the history of cosmic star formation remains incomplete.

The second major result uncovered from these large surveys is the existence of a tight, nearly linear relation in the star formation–stellar mass plane commonly referred to as the main sequence (MS, e.g., Brinchmann et al. 2004; Daddi et al. 2007; Elbaz et al. 2007, 2011; Noeske et al. 2007). This relation shows a small scatter ($\sim$0.3 dex) and appears to be already in place at least at $z \sim 4-5$ (Speagle et al. 2014; Schreiber et al. 2015; Leslie et al. 2020). In addition, the normalisation of the MS increases with redshift, meaning that galaxies were forming more stars in the past, in line with the CSFRD. In fact, most of the star formation between $0< z < 3$ takes place in galaxies that fall on the MS, and are commonly referred to as star–forming galaxies (SFGs). At all redshifts, however, there are outliers to this relation, both above and below the main sequence.

Galaxies above the MS are known as starburst galaxies, and have star formation rates that may reach several order of magnitudes above those expected from their stellar mass. This starburst activity may be triggered by galaxy mergers and interactions (Sanders & Mirabel 1996; Engel et al. 2010), or from the steady accretion of cold gas from the intergalactic medium (CGM) that leads to the fragmentation of the galactic disks into smaller clumps of higher–density gas, where star formation takes place (Dekel et al. 2009; Inoue et al. 2016). These periods of intense star formation are typically short-lived, and most massive galaxies are expected to undergo one or more such periods at some point in their evolution. Conversely, galaxies below the MS are known as quiescent galaxies, and are usually massive, with little star formation left. They typically have old stellar populations, and have now been detected beyond $z \sim$3 (e.g., Valentino et al. 2020c; Alberts et al. 2023; Nanayakkara et al. 2024). Several mechanisms have been invoked to explain their low star formation, and involve processes that either remove or heat up the gas in galaxies, such as feedback from active galactic nuclei (AGNs), stellar feedback and gravitational effects. However, such mechanisms and their interplay are still not well understood and continue to be an area of active research (Man &

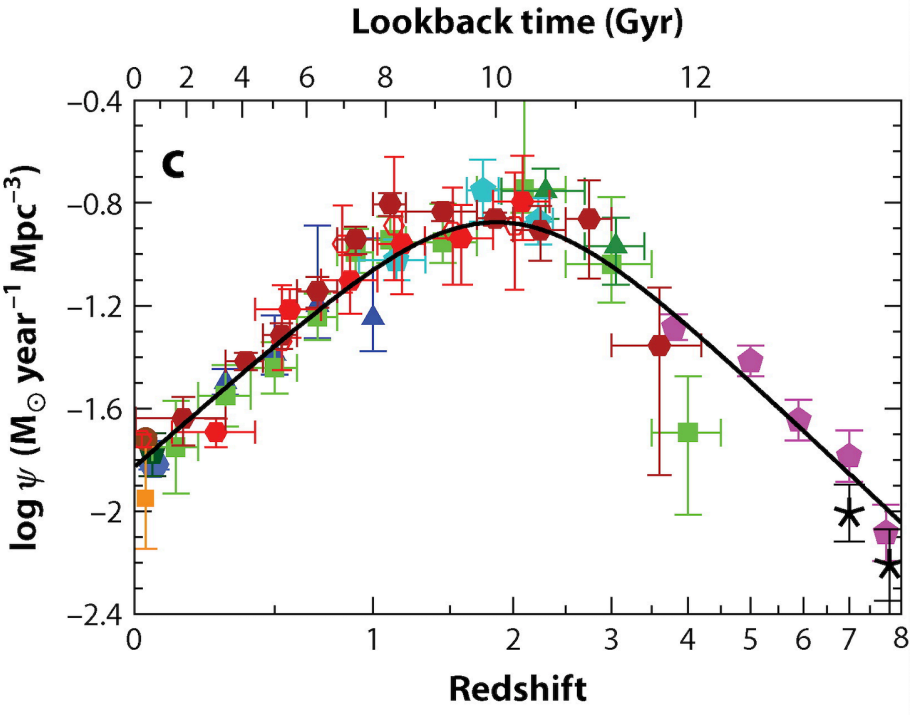

Figure 1.1: The cosmic star formation rate density (CSFRD) as a function of redshift. The datapoints show measurements from surveys at different wavelengths, and the black line is the best fit. The star formation rate peaked roughly 10 billion years ago, and has since steadily decreased until the present day. Above $z \sim 3$, the CSFRD might be underestimated as UV measurements cannot account for dust–obscured star formation. Figure adopted from Madau & Dickinson (2014).

Belli 2018). Generally, while the MS is considered an average trajectory for galaxy evolution, galaxies will vary their position over time, rising above or dipping below the main sequence multiple times throughout their lifetime.

It is clear that the star formation rate of a galaxy is a crucial parameter, as it determines its past build-up of stellar mass and its future evolution. Surveys in the local universe have shown that star formation is primarily driven by the availability of cold molecular gas – the direct fuel from which stars form. As such, great efforts have been dedicated in the past decade towards measuring the cold gas content in large numbers of galaxies across the MS plane and across time, with the aim to constrain the evolution of the cosmic molecular gas density in a similar fashion as was done for the CSFRD in Fig.1.1.

1.2. Cold Molecular Gas

The bulk of the molecular gas mass in galaxies is composed of molecular hydrogen, H_2. This molecule lacks a permanent dipole moment, which means that its lowest rotational transitions are both forbidden and have upper-level energies of $E/k \sim$500 and 1000 K above the ground state. These lines are therefore only excited in gas with temperatures above 100 K, significantly higher than those found in GMCs (10–50 K), which in practice means that most of the H_2 is invisible in emission line studies. Fortunately, the ISM also contains heavier elements that form molecules which can be studied through emission.

1.2.1. Tracers of cold molecular gas

1.2.1.1. Carbon Monoxide

Carbon monoxide, and in particular $^{12}CO^{16}O$ (hereafter CO), is the second most abundant molecule in the ISM after H_2. CO has a weak permanent dipole and has low excitation requirements, with an upper level energy of $E/k = 5.5$ K for the first excited state, CO $J = 1 \rightarrow 0$ (hereafter CO(1–0)). Along with its critical density of $n_{\rm crit}$=10^3 cm^{-3}, CO(1–0) is thus easily excited under average conditions in cold molecular clouds. Further, this transition lies at $\nu_0 = 115.27$ GHz, and is easily observed from the ground at $z = 0$. For these reasons, astronomers have historically used CO to measure cold molecular gas masses.

Converting the integrated CO luminosity to a molecular gas mass is non–trivial, and requires a mass–to–light conversion factor. In the local universe, this factor is $X_{\rm CO}$, the ratio of column density (H_2 molecules per cm^2, N_{H_2}) to CO integrated line intensity (K km s^{-1}), such that:

$$N(H_2) = X_{\rm CO} W(CO(1-0)). \tag{1.1}$$

At high redshift, spatially integrated quantities are usually measured, and the total molecular gas mass is then given by

$$M_{\rm H_2} = \alpha_{\rm CO}\ L'_{\rm CO(1-0)}, \tag{1.2}$$

where M_{gas} is in units of solar mass, $M_\odot$, $L'_{CO(1-0)}$ is expressed in units of K km s^{-1} pc^2. In both cases, it is necessary to include a correction for the contribution of heavier elements, which amounts to ∼36% of the total budget of molecular gas mass, according to standard cosmological abundances.

Numerous methods have been used in the Milky Way to calibrate the value of the conversion factor in molecular clouds, such as gamma–ray emission originating from the interaction of cosmic rays with H_2 molecules, emission from optically thin tracers like dust or CO isotopologues, and using the virial theorem (see Bolatto et al. 2013, for a review). The value of α_{CO} is likely a function of local ISM conditions, such as pressure, density, temperature and metallicity (Narayanan et al. 2012; Bolatto et al. 2013; Accurso et al. 2017a) and remains the main source of uncertainty when determining total molecular gas masses.

At higher redshifts, where CO(1–0) is often redshifted to frequency ranges inaccessible to current facilities, the higher-J rotational transitions of CO are usually employed to derive total molecular gas masses. When multiple CO lines are available, it is possible to build a CO spectral line energy distribution (CO SLED, also known as CO rotational ladder), which describes the relative strengths of CO emission lines (the gas excitation) and depends on the gas density and temperature. When only higher–J CO lines are available, it is possible to use excitation correction factors, i.e., the ratio between the two CO rotational transitions, to convert back to the CO(1–0) transition to which α_{CO} is calibrated. These are often taken from modelling the CO SLEDs of individual galaxies, or built in a statistical way using observations of multiple CO transitions in different sources (e.g., Bothwell et al. 2013; Boogaard et al. 2020; Birkin et al. 2021). The assumption of excitation correction factors thus adds further uncertainty to the derived molecular gas masses.

The final caveat of using CO at high redshifts to trace molecular gas comes from the increasing temperature of the CMB. This results in a decrease in the contrast of the intrinsic line emission with the CMB, which can have a pronounced impact on the derived molecular gas masses and their spatial distributions, particularly at low gas kinetic temperatures, and beyond $z \geq 4$ (da Cunha et al. 2013; Zhang et al. 2016).

1.2.1.2. Atomic Carbon

Additional tracers of the molecular gas mass are the atomic carbon fine structure lines, particularly its ground state transition [C I]($^3P_2$–$^3P_1$) (hereafter [C I]), at ν_0 =492 GHz. Early, simple one-dimensional modelling of photodissociation regions (PDRs) predicted the [C I] emission in the ISM to arise only from the surface of molecular clouds, in a region between CO and [C II] (Tielens & Hollenbach 1985). However, subsequent theoretical (Papadopoulos et al. 2004; Tomassetti et al. 2014) and observational work on nearby galaxies have suggested that [C I] is distributed across the ISM and its emission is concomitant with that of CO(1–0) (Keene et al. 1985, 1997; Ojha et al. 2001; Ikeda et al. 2002; Pérez-Beaupuits et al. 2015; Jiao et al. 2019), thus making [C I] a promising candidate for tracing their molecular gas content (Dunne et al. 2022). [C I] has the advantage of

being brighter than CO(1–0), and optically thin (Ikeda et al. 2002; Weiß et al. 2003; Harrington et al. 2021), thus being both easier to detect and able to probe high column density environments. Further, its excitation temperature of 24 K is sufficiently low to trace the bulk of the cold gas content in galaxies.

It is therefore no surprise that [C I] has been gaining attention in recent years for its potential at high redshift. Numerous studies have confirmed that its emission is correlated with that of CO in a variety of galaxy populations, from star–forming galaxies (Popping et al. 2017; Bourne et al. 2019; Valentino et al. 2018, 2020a) to starburst and quasar (QSO) hosts up to $z \sim 4$ (Walter et al. 2011; Alaghband-Zadeh et al. 2013; Bothwell et al. 2017; Yang et al. 2017; Andreani et al. 2018; Nesvadba et al. 2019; Gururajan et al. 2023). While these studies rely mostly on comparison against $J \geq 2$ CO emission lines, the advent of more sensitive facilities will enable direct comparison against CO(1–0) with both spatially resolved and unresolved scales in the nearby future.

The application of this tracer is currently limited due to a lack of a systematic calibration of the [C I]-to-H_2 conversion factor across various types of galaxies and redshifts. When available, this calibration is reliant on estimates of total molecular gas masses obtained from CO observations, and thus suffers from the same uncertainties inherent to α_{CO}.

1.2.1.3. Dust

Another way to estimate the mass of cold molecular gas present in the ISM is through the far–infrared and submillimiter emission from dust grains. The Rayleigh–Jeans (RJ) tail of the dust emission at long wavelengths is optically thin, and thus directly traces the cold dust mass. Once a dust mass is derived, it is possible to obtain an estimate of the cold gas mass under the assumption of a dust emissivity per unit mass and a gas–to–dust ratio, which can vary between sources and is dependent on metallicity (e.g., Rémy-Ruyer et al. 2014).

The main advantage of this method is its efficiency, as the continuum emission of large samples of galaxies can be obtained much faster than line observations. However, it suffers from several drawbacks. In the first place, this method was calibrated using CO observations, and thus suffers from the same limitations. It also assumes a single gas–to–dust ratio (which may not be applicable for sub–solar metallicities), as well as a single temperature for the bulk of the cold dust, without redshift evolution. Finally, the method was calibrated at rest–frame 850μm, and thus high–redshift observations at lower rest–frame wavelengths need extrapolations that can introduce systematic errors in the derived gas masses. Although this is a promising method, there have been few studies directly comparing results with CO(1–0) in the same samples of galaxies. A study by Kaasinen et al. (2019) in a few $z \sim 2$ SFGs found that both estimates agreed within a factor of two. More studies on larger samples of galaxies of different metallicities and star formation rates are necessary to better establish its reliability, particularly at high redshift.

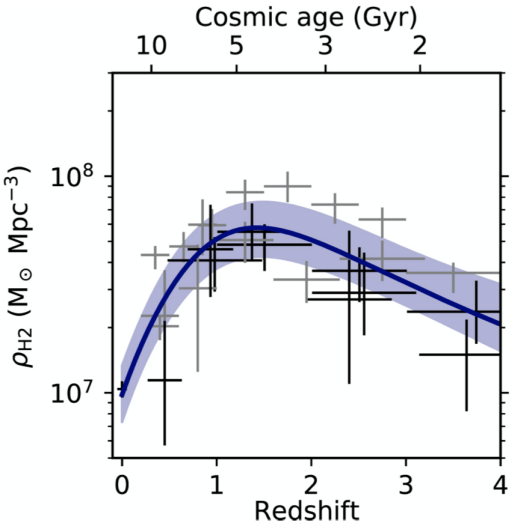

Figure 1.2: Cosmic molecular gas mass density as a function of redshift derived from a combination of low–and mid–J CO and dust continuum measurements. The amount of molecular gas mass available to form stars peaked at $z \sim$1.5. The evolutionary trend is reminiscent to that of the CSFRD, showing that molecular gas is the main driver of the star formation in galaxies. Figure adopted from Walter et al. (2020).

1.2.2. Current results

Improvements in the sensitivity and frequency coverage of current facilities have enabled the use of all these tracers to study the molecular gas emission over large cosmic volumes (see review by Hodge & da Cunha 2020). Recent "blind" CO spectroscopic surveys have successfully detected the continuum and CO/[C I]/[C II] emission lines of galaxies in an unbiased and complete manner (e.g., Walter et al. 2014; Aravena et al. 2016; Decarli et al. 2016; Walter et al. 2016; Aravena et al. 2020; Boogaard et al. 2023). This is crucial, as pre-selecting the targets at other wavelengths may introduce biases in our understanding of galaxy evolution by limiting our studies to specific subsets of the more general population.

By scanning a region of the sky at a chosen frequency, these surveys can detect emission lines, and therefore molecular gas, at various redshifts, providing a flux–limited census of the gas content across redshift. Notable examples of such blind surveys are the ASPECS/VLASPECS (Walter et al. 2016; Decarli et al. 2019; Riechers et al. 2020) and the JVLA COLDz survey (Riechers et al. 2019), conducted with ALMA and the JVLA, which have begun to unveil the cold gas density evolution at high redshift. Similar constraints derived from dust emission show a consistent evolution (Scoville et al. 2017; Liu et al. 2019; Magnelli et al. 2020; Wang et al. 2022). The mass of molecular gas in galaxies slowly increased since early cosmic epochs, peaked around $z \sim 1-3$, and then decreased by a factor of $\sim$6 to the present day (Fig. 1.2). This evolution is in quantitative agreement

with the evolution of the CSFRD (Fig. 1.1), and so molecular gas appears to be the main factor driving the history of cosmic star formation. These studies still require the assumption of excitation correction factors (blind surveys detect different CO transitions at different redshifts) and a constant $\alpha_{\rm CO}$ that does not evolve with redshift.

The latest generations of hydrodynamical simulations (e.g., Vogelsberger et al. 2014; Schaye et al. 2015; Davé et al. 2017) and semi–analytical models (e.g., Lagos et al. 2015) have, in parallel to observations, made great efforts in recent years to predict the evolution of the cold gas content of the Universe. While theoretical work can qualitatively reproduce the observed behaviour of the cold gas density evolution, there is still significant tension between the predicted molecular gas fractions and the integrated molecular gas volume densities, as models and simulations tend to underpredict the amount of molecular gas in galaxies, particularly at high redshift (Davé et al. 2019; Popping et al. 2019a; Tacconi et al. 2020). Further, they predict most of the molecular gas to be associated with low mass galaxies, in disagreement with observations, and still struggle to reproduce the most gas–rich systems (e.g., Lagos et al. 2015).

1.3. Submillimeter Galaxies

Identified during the first wide–field extragalactic surveys with single-dish submillimeter telescopes, submillimeter galaxies (SMGs, Smail et al. 1997; Barger et al. 1998; Hughes et al. 1998) are characterised by their bright long–wavelength dust continuum emission, boasting flux densities $\geq$1 mJy at 850μm. They have traditionally been considered the high–redshift analogues of local ULIRGs due to their large infrared luminosities ($L_{\rm IR}$ $\geq 10^{12-13} L_{\odot}$, Swinbank et al. (2014); Dudzevičiūtė et al. (2020)), although they are $\sim$1000 times more populous (Simpson et al. 2014; Zavala et al. 2021). Large multiwavelength studies have slowly revealed the physical properties of SMGs over the last two decades. Their redshift distribution peaks at $z \sim 2$ (with a tail towards $z \geq 5$, Greve et al. 2005), where they contribute $\sim$20% and 30–50% to the volume–averaged cosmic star formation rate density and the stellar mass density, respectively. SMGs have stellar masses of 10^{10-11} $M_{\odot}$ (Hainline et al. 2011; Michałowski et al. 2010; da Cunha et al. 2015) and high star–formation rates 10^{2-3} $M_{\odot}$ yr^{-1} (Magnelli et al. 2012b; Swinbank et al. 2014; Dudzevičiūtė et al. 2020). They have copious amounts of dust ($M_{\rm dust}$ $\geq 10^{8}$ $M_{\odot}$) distributed in compact regions of 2–3 kpc (Ikarashi et al. 2015; Simpson et al. 2015; Hodge et al. 2016, 2019; Gullberg et al. 2019). The dust absorbs most of the optical/UV light and re-radiates it in the infrared, often rendering SMGs faint or nearly invisible at rest–frame optical wavelengths (Swinbank et al. 2014; Danielson et al. 2017; Ikarashi et al. 2022; Manning et al. 2022; Smail et al. 2023).

Although they were initially considered to be mostly composed of merger–triggered starbursts, large statistical studies of the SMG population have actually shown that they are distributed across the SFR-M_{*} plane, populating both the normal star–forming and the starburst regimes (e.g., da Cunha et al. 2015; Miettinen et al. 2017; Franco et al. 2020; Lim et al. 2020; Cardona-Torres et al. 2023).

Their high star formation rates could be due to several causes, from major mergers (Engel et al. 2010; Alaghband-Zadeh et al. 2012; McAlpine et al. 2019) and minor interactions with other galaxies, to cold gas accretion from the intergalactic medium (Dekel et al. 2009; Hodge et al. 2012; Inoue et al. 2016).

Several lines of evidence suggest an evolutionary pathway in which some SMGs evolve into local, massive, elliptical galaxies (Eales et al. 1999; Lilly et al. 1999; Swinbank et al. 2006; Simpson et al. 2017) after a period of unobscured QSO activity (see Fig.1.3 Sanders & Mirabel 1996; Alexander et al. 2005; Hopkins et al. 2008a,b). This is supported by comparable co-moving number densities (Simpson et al. 2014; Toft et al. 2014), dark matter halo masses and clustering properties (Hickox et al. 2012; García-Vergara et al. 2020) and stellar ages (Simpson et al. 2014; Dudzevičiūtė et al. 2020), among others. Unobscured QSOs are therefore predicted to be the intermediate stage in the evolutionary sequence between the merger of two gas–rich SMGs to gas–depleted, quenched galaxies, as AGN feedback shuts down star formation by either heating up or blowing out the molecular gas in the host galaxy. Interestingly, unobscured QSOs have been found to reside in host galaxies with ongoing star formation (e.g. Gürkan et al. 2015; Harris et al. 2016; Netzer et al. 2016; Pitchford et al. 2016; Stanley et al. 2017), and gas fractions that can be at times indistinguishable from normal, MS galaxies (e.g., Rodighiero et al. 2019; Valentino et al. 2021), challenging the long–established paradigm of massive galaxy evolution.

Historically, cosmological simulations of galaxy formation have had difficulties in reproducing the abundances and high SFRs of the SMG population (Baugh et al. 2005; Swinbank et al. 2008; Davé et al. 2010; Hayward et al. 2013). This can be due to a number of factors, such as a top-heavy initial mass function (IMF), blending of multiple galaxies into a single sub-mm source in low–resolution observations, and the inability of hydrodynamical simulations to adequately resolve starbursts. The challenges that this population poses to simulations makes them particularly useful for constraining theoretical models of galaxy formation, and there have been promising results from the latest generation of hydrodynamical simulations and semi–analytical models (e.g., McAlpine et al. 2019; Lagos et al. 2020; Lovell et al. 2021).

1.3.1. Molecular Gas in SMGs

SMGs were among the first systems to be targeted in CO(1–0) at high redshift. They host some of the most intense starbursts in the Universe. As such, not only are they a key test of our star formation and galaxy evolution models, but they are the most obvious target to detect the faint CO(1–0) line, since their large dust masses indicate that they must also possess an equally massive reservoir of cold gas. This presents itself as a great advantage, since the CO emission can then be detected, and in some cases even resolved, without the prohibitive time investments that would be required for less massive galaxies.

Early searches for molecular gas emission in SMGs revealed massive gas reservoirs of $M_{gas} \geq 10^{10-11}$ $M_{\odot}$ (Greve et al. 2005; Tacconi et al. 2006, 2008; Engel et al. 2010; Bothwell et al. 2013), implying gas consumption timescales of only

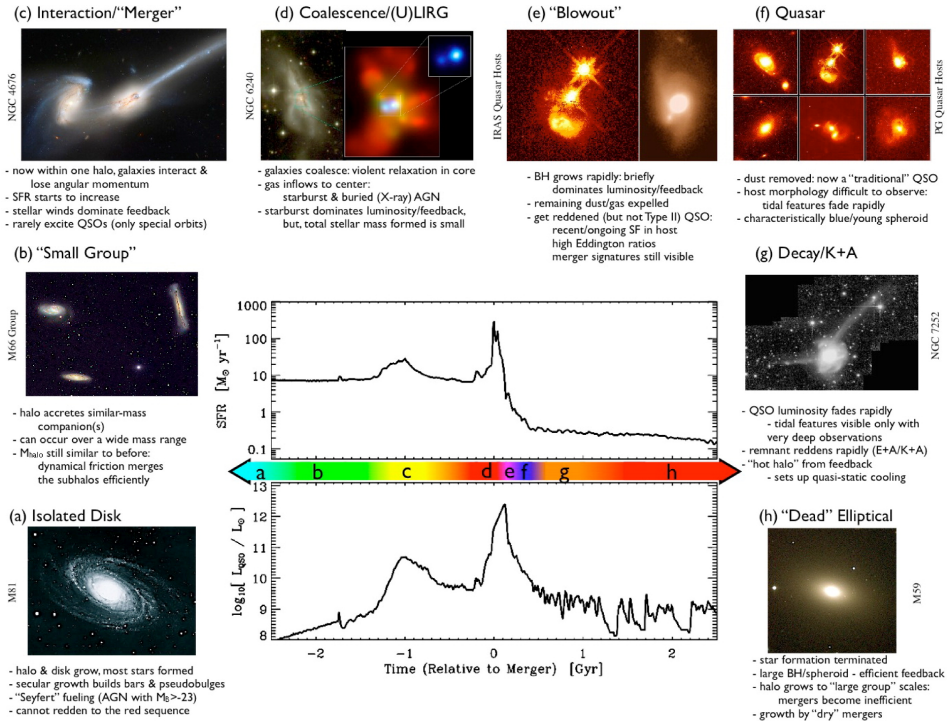

Figure 1.3: Canonical picture of massive galaxy evolution: the SMG–phase of a galaxy is caused by a major, gas–rich merger. Subsequently, gas is funnelled into the center of the galaxy, triggering the central AGN, which eventually shuts down star formation as it heats or expels the gas from the host galaxy. Ultimately, SMGs are believed to evolve into the present–day, massive, quiescent galaxies. Figure taken from Hopkins et al. (2008a).

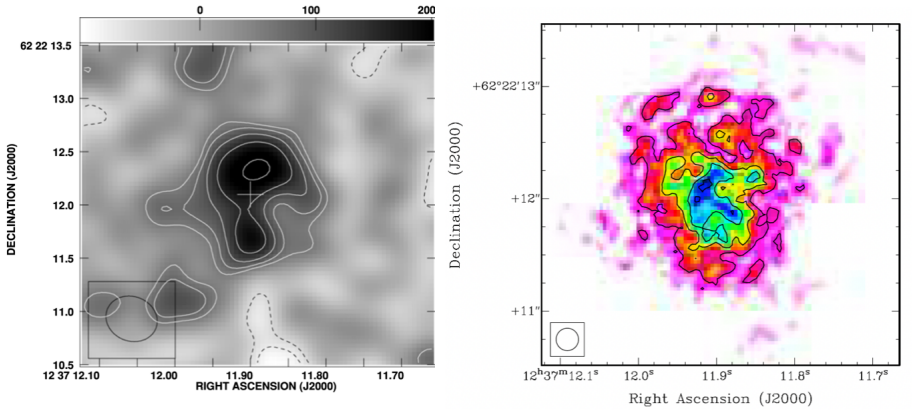

Figure 1.4: Mapping of the molecular gas in the SMG GN20 as traced by CO(2–1). The image on the right, taken from Hodge et al. (2012), is one of the highest resolution studies (0.19″) of a low–J CO emission line at high redshift, and required a total of 120 h on the JVLA. It also provided kinematic information, which the then existing CO(2–1) observations (shown on the left at 0.45″, figure taken from Carilli et al. (2010)) did not have. Resolving the gas reservoir traced by low–J emission at high redshift remains a technical challenge and is currently restricted to the most luminous SMGs.

a few hundred Myr. Subsequent efforts aimed to resolve these gas reservoirs to study their sizes and kinematics, although based on the brighter mid–J CO emission lines. These studies often resolved SMGs into multiple components with disturbed gas reservoirs, indicative of ongoing major mergers, and compact sizes of a few kpc (Tacconi et al. 2008; Engel et al. 2010). Contrary to these early observations, later spatially resolved studies have often found signatures of disc-like rotation as well (e.g., Hodge et al. 2012; Calistro Rivera et al. 2018; Amvrosiadis et al. 2023). Unfortunately, spatially resolved observations of low–J CO emission in high–redshift SMGs are limited. One of the best examples is GN20, whose gas reservoir was mapped on scales of 1.3 kpc (0.19″) after 120 h of observations targeting CO(2–1) with the JVLA (see Fig.1.4).

Despite the lack of CO(1–0) in most cases, detecting several CO lines in the same target allows modeling of the (global) CO SLEDs, which has provided much insight into the average molecular gas properties in SMGs such as density and temperature. In local thermodynamic equilibrium (LTE), the excitation temperature is equal to the kinetic temperature ($T_{\rm kin}$) of the gas, and it determines the population of the molecular levels (and therefore the emission coming from a given CO transition) through the Boltzmann distribution. If the measured intensity is less than given by $T_{\rm kin}$, such as in low density environments, the excitation is said to be 'sub–thermal'.

The most commonly used methods to model the gas excitation are the large velocity gradient (LVG) method (e.g., de Jong et al. 1975; Goldreich & Scoville 1976; van der Tak et al. 2007a; Weiss et al. 2007) and photo–dominated regions (PDR) models (Tielens & Hollenbach 1985; Hollenbach et al. 1991; Kaufman et al.

1999; Bisbas et al. 2012). The former calculates how the levels of CO are populated through collisional excitation with H_2 for a given $T_{\rm kin}$, H_2 density, CO abundance and velocity gradient. These codes take into account the CMB temperature, and calculate the optical depths and therefore the resulting line intensities. With a sufficiently well–sampled line emission ladder (this could be done for any molecule, not just CO), it is possible to put constraints on the temperature and density of the gas. PDR codes similarly constrain the temperature and density of the gas, but they also take into account the UV radiation field. The main limitation of PDR codes is that they are based on one–dimensional infinite slabs, rather than a confined volume. In both cases, the line emission is assumed to arise from the same volume, which may not always hold true, as different CO levels need different densities and temperatures to be excited. On average, the typical density of SMGs is $\log(n_{\rm H_2}\ [{\rm cm^{-3}}])$=3–6, and temperatures span a wide range from a few tens up to hundreds of Kelvin (e.g., Alaghband-Zadeh et al. 2012; Bothwell et al. 2017; Cañameras et al. 2018; Valentino et al. 2020a; Harrington et al. 2021).

The first studies of CO(1–0) in SMGs revealed two surprising facts. First, the gas reservoirs traced by CO(1–0) emission had sizes of $\geq$10 kpc, more extended than seen from the dust continuum or higher CO transitions (Ivison et al. 2010; Riechers et al. 2011a,c; Hodge et al. 2012). Second, there seemed to be an excess of CO(1–0) emission, which could not be accounted for with simple extrapolations from mid–J CO emission lines using average excitation ratios. These studies revealed the diverse range of CO SLEDs in SMGs (see Fig.1.5), with some displaying extreme excitation conditions (e.g., Riechers et al. 2013), and others containing large volumes of sub–thermally excited gas (e.g., Hainline et al. 2006; Carilli et al. 2010; Ivison et al. 2011; Riechers et al. 2011c). It soon became apparent that mid–J CO transitions provided a limited and, in some cases, biased view of the total molecular gas reservoirs at high redshift, leading to underestimations of their total mass and extent. These constraints came however from sources that were usually gravitationally lensed or of an extreme nature (e.g., Harrington et al. 2021). Therefore, obtaining measurements of the low–excitation CO transitions in large samples of SMGs is crucial for fully characterising the gas reservoirs of these extreme, high–redshift systems. Unfortunately, this task has proved particularly challenging due to the lack of spectroscopic redshifts for most known SMGs.

1.3.2. Towards large spectroscopically–confirmed SMG samples

The redshift of a galaxy is a measure of how far away that galaxy is, and is a crucial parameter to put that galaxy in the broader context of hierarchical galaxy assembly. There are two types of redshifts: photometric and spectroscopic. Photometric redshifts are more easily obtained –they require identification of the optical/MIR/radio counterparts of the SMGs, followed by a fit to their photometry (which together forms the spectral energy distribution, or SED, of a galaxy) using empirical or synthetic galaxy spectra. Spectroscopic redshifts, on the other hand, are derived by detecting and identifying an emission line, either in the optical (e.g., Lyα, Hα, [O II], [N II], [S II]) or in the sub-mm regime (e.g., CO, [C I], [C II]).

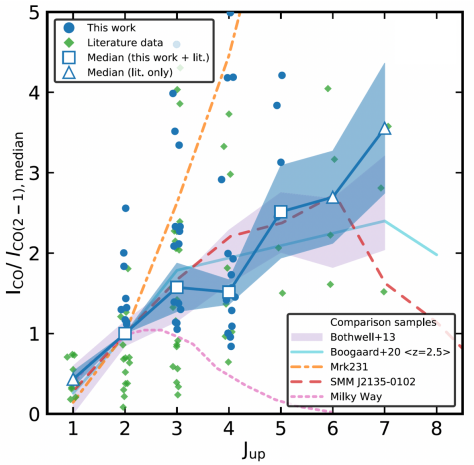

Figure 1.5: At high redshift, it is common to detect only one or two CO emission lines per galaxy. Statistical CO SLEDs are then constructed by combining all the measurements available, with different galaxies contributing to different transitions. This figure, taken from Birkin et al. (2021), shows the statistical CO SLED for high–redshift SMGs. A huge variety in excitation conditions is evident, as well as the scarcity of CO(1–0) measurements needed to properly constrain the gas excitation.

Obtaining spectroscopic redshifts requires more observing time, and therefore most SMGs detected to-date only have photometric redshifts (Hodge & da Cunha 2020). These are often too inaccurate for dedicated follow-up observations of CO emission lines, which has hampered efforts to study the gas mass and properties of large samples of SMGs in a statistical way.

This has changed in recent years thanks to spectral line scans, which have become easier to conduct due to the increased bandwidths of the receivers operating at mm and sub-mm facilities like ALMA and NOEMA. This technique involves setting up the receivers to observe a continuous range in frequency in order to detect and identify one or more emission lines, usually mid–J CO or [C I]. These lines are bright in SMGs, and their detection offers both a robust spectroscopic redshift and insights into their gas content and kinematics. Several surveys targeting well–studied cosmological fields (e.g., Birkin et al. 2021; Chen et al. 2022; Cox et al. 2023) and/or gravitationally lensed galaxies (e.g., Cañameras et al. 2015; Reuter et al. 2020) have slowly increased the number of SMGs with spectroscopic redshifts. Now, we are finally in a position to observe CO(1–0) on large enough samples of SMGs.

1.4. Instruments

Our current understanding of star formation and cold molecular gas across redshift would not be possible without the continuous development and improvement of radio–interferometers. These are collections of antennas spread over a large area

that work together as a single telescope. The signals received by each antenna are then combined to create an image of the observed object. This technique allows us to achieve much higher resolution than would otherwise be possible with a single antenna. Three facilities have been essential in carrying out the work presented in this thesis (see Fig.1.6).

1.4.1. Karl G. Jansky Very Large Array (JVLA)

The Karl G. Jansky Very Large Array (JVLA) was originally known simply as the VLA, but was renamed in honor of the pioneer radio astronomer Karl G. Jansky after it underwent a major upgrade in 2012. The facility has been operating since the 1970s, and it is situated on the Plains of San Agustin, in New Mexico, USA. The JVLA consists of 27 radio antennae of 25-m distributed in a Y-shaped configuration, and covers a continuous frequency range from 1 to 50 GHz. The antennae can be placed in four different configurations at various distances to provide varying angular resolutions. It has been instrumental for studies in various areas of radio/sub–mm astronomy, including radio mapping of galaxies, pulsar discoveries, star formation studies, and exploring the dynamics of the interstellar medium across cosmic time.

1.4.2. NOrthern Extended Millimeter Array (NOEMA)

NOEMA (Northern Extended Millimeter Array) is a radio interferometer facility located on the Plateau de Bure, in the French Alps. It superseded the Plateau de Bure Interferometer (PdBI), which started operations in the 1990s with 6 individual antennae. In an effort to enhance its capabilities and sensitivity, the PdBI was upgraded in 2012 to what is now NOEMA. It consists of 12 individual 15-meter antennae covering 80 to 370 GHz, and can cover baselines with a maximum separation of 1.7 km in the E-W direction and 368 m in the N-S direction. Achieving resolutions of up to $0.2''$ at 1.3 mm, it is essential for the study of the Northern Hemisphere at sub–mm wavelengths.

1.4.3. Atacama Large sub-Millimeter Array (ALMA)

The Atacama Large Millimeter/submillimeter Array (ALMA) is currently the world's largest and most powerful submillimeter array. It is located at an altitude of approximately 5,000 meters on the Chajnantor Plateau in the Atacama Desert. It started scientific observations in September 2011, and it consists of 66 antennae operating in 10 receiver bands covering from 35 GHz (Band 1) up to 950 GHz (Band 10). The antennae can be moved around to cover baselines ranging from 150m to 16km, achieving resolutions that range from $4.3''$ at 110 GHz in the most compact configuration to $0.043''$ at 110 GHz in the most extended one. With its sensitivity and large frequency coverage, ALMA has revolutionised our understanding of the cold ISM of galaxies, from the study of young protoplanetary disks, to imaging of molecular clouds in nearby galaxies and mapping the distribution of dust and molecular gas in high–redshift galaxies.

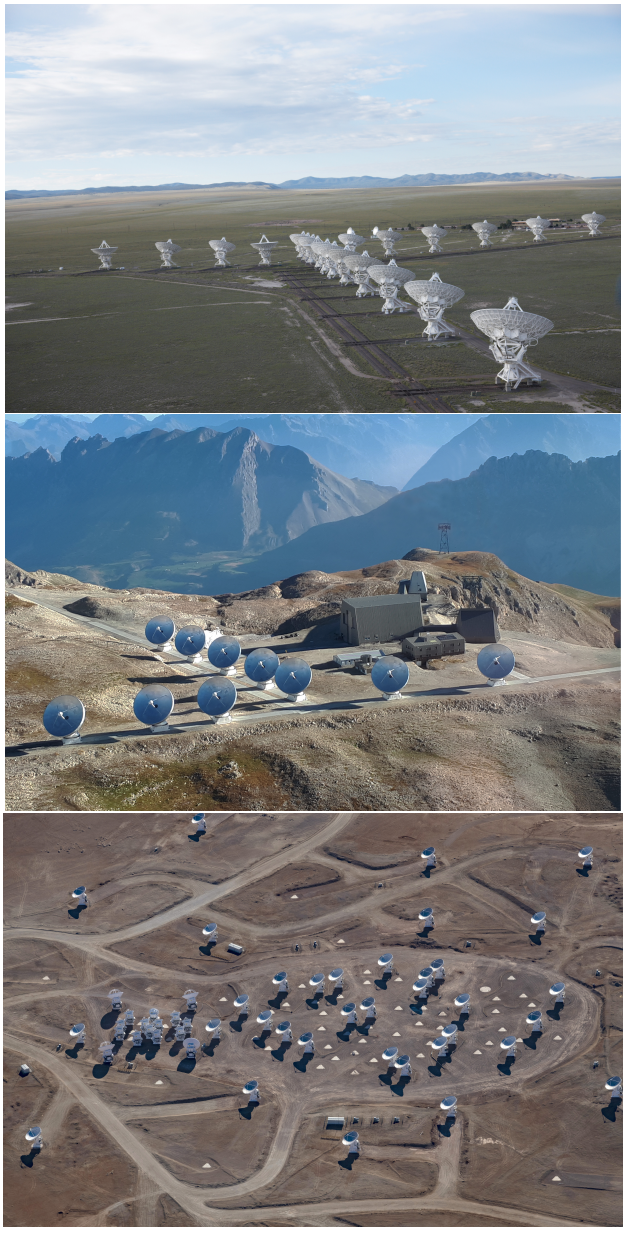

Figure 1.6: The JVLA (top), NOEMA (center) and ALMA (bottom) are the three most powerful facilities currently observing at mm/sub–mm wavelengths. Image credits: NRAO, IRAM, ESO.

1.5. This Thesis

The upgrades in power and sensitivity of current facilities like ALMA and the JVLA now offer the potential to detect and even resolve the cold molecular gas reservoirs using the more elusive low–excitation CO transitions in large samples of galaxies at high redshift. At the same time, this will enable comparative studies between CO and other tracers, such as [C I] and dust. While the latter are easier to observe at high redshift, they currently lack a robust calibration to be used independently from CO observations. This thesis includes four studies of the gas content traced by low–J CO and [C I] emission lines in star–forming galaxies and quasars across a range of redshifts, and is organised as follows:

In **Chapter Two**, we present deep JVLA observations of CO(1–0) in the $z = 3.4$ unlensed SMM J13120+4242, an SMG lying at the very massive end of the main sequence. We resolve the molecular gas reservoir on $\sim$3 kpc scales, and find it shows a disturbed morphology and significant turbulent motion. The short depletion time and presence of a central AGN suggest that this galaxy is a late stage merger, in a transitional stage towards an unobscured QSO. This study underscores the unique capabilities of the VLA in resolving CO(1–0) at high redshift, and the importance of high–resolution kinematic studies for characterising the dynamical state of the cold ISM.

In **Chapter Three**, we observe CO(2–1) in a sample of ten FIR–bright, unobscured QSOs at $z = 1 - 1.5$, just after the peak epoch of QSO activity and massive galaxy assembly. We detect gas emission in 7 out of 10 QSOs, showing that they can retain their gas reservoirs despite the presence of a central AGN. We find that the QSO-hosts are mostly low mass, starburst galaxies undergoing a phase of intense, rapid growth that will quickly use up the molecular gas available and quench star formation. Cosmological simulations still struggle to reproduce the high star formation rates and gas fractions found in this population.

In **Chapter Four**, we present the first phase of a large JVLA legacy survey targeting CO(1–0) in a sample of $z = 2 - 5$ unlensed SMGs. The gas excitation shows a large scatter within our sample, and do not show a correlation with any physical properties of the galaxies. Comparison with semi–analytical models show the improvement of theoretical models, which are now capable of reproducing key parameters of the ISM of these massive, gas–rich SMGs. These results highlight the heterogeneous nature of the most massive star–forming galaxies at high–redshift, and the importance of CO(1–0) observations to robustly constrain the total molecular gas content and ISM properties of galaxies.

Finally, in **Chapter Five**, we build on the work from Chapter Four and combine the CO(1–0) data with [C I](1–0) and dust continuum observations in a subsample of 21 galaxies from the JVLA survey. This allows us to test the robustness of [C I] and dust as molecular gas tracers at high redshift when compared to the standard tracer, CO. Both CO(1–0) and [C I](1–0) appear to be well correlated out to $z \sim 5$, and the gas masses derived from them are consistent once we account for the assumptions in the conversion factors, supporting the use of [C I] as a tracer of molecular gas. We investigate the suppression of CO/[C I] flux due to the warmer CMB at high redshift. We find that, while the effect on both CO and [C I] can be

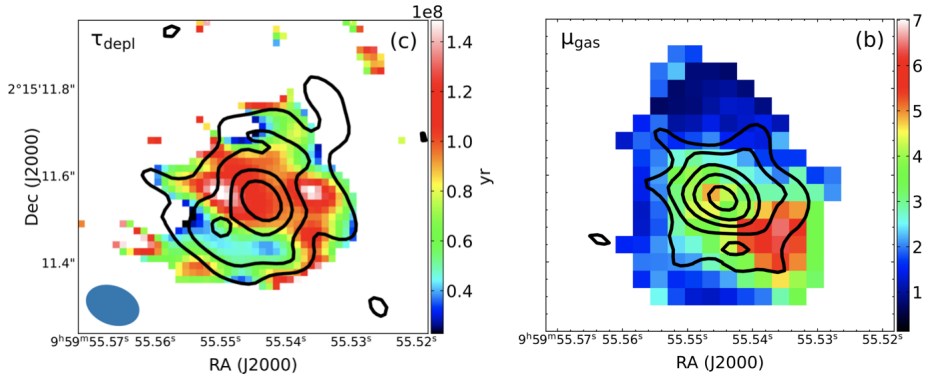

Figure 1.7: JWST is revealing the stellar distribution in high–redshift SMGs. The combination with high–resolution mapping of gas and dust with ALMA and the VLA will open a new window to the intricacies of the ISM in these gas–rich, highly–star forming galaxies. Figure adapted from Liu et al. (2023).

up to 70%, the relative effect on the ratio is only up to 30%, which is within the scatter of the current data.

1.6. Future Work

The next generation VLA (ngVLA) is poised to commence operations in the near future. The upgrade will provide an order of magnitude improvement in sensitivity and resolution compared to the current JVLA and ALMA, covering the frequency range of 1–116 GHz. Routine studies at sub-kiloparsec scale resolution will be conducted on cold gas reservoirs even at high redshift, unveiling the physical conditions of gas throughout cosmic history in unprecedented detail. It will also be possible to resolve sub–structure in the galaxies such as clumps, spiral arms and bars, key to understanding the mechanisms that impact star formation in galaxies. In combination with the higher frequency coverage with NOEMA, multi–line studies of large samples of galaxies are needed to constrain the ISM conditions in a wide range of galaxies. At the same time, comparative studies of resolved dust and gas (kinematics) will be necessary to probe the interplay of stars, gas and dust at high redshift, and the driving mechanisms of star formation at smaller scales, although they currently still require significant time investments. Further, the increased bandwidth and sensitivity of current and future facilities will increase survey speeds, revealing thousands of galaxies in blind surveys.

Leveraging its remarkable sensitivity, ALMA will continue to push the limits of our understanding of the cold ISM at high redshift. With the introduction of the new Band 1 and Band 2 receivers, which will cover a frequency range down to 35 GHz, ALMA will enable observations of CO(1–0) up to $z = 2$. This will directly enable surveys of CO(1–0) (and CO(2–1) at higher redshifts) in the southern sky.

The James Webb Space Telescope started operations in July 2022 and has al-

ready revolutionised the study of high–redshift galaxies, providing ground–breaking results of galaxy formation with the discovery of massive galaxies well above $z = 10$ (Curtis-Lake et al. 2023). Its frequency coverage and sensitivity are providing a window into the structure of dust–obscured galaxies at rest–frame optical and near infrared wavelengths. Sub-kiloparsec imaging of the stellar distribution in galaxies will for the first time enable a detailed comparison of the stars, gas and dust content in galaxies (see Fig. 1.7), and will constrain state-of-the art models and numerical simulations of galaxy formation.