Edinburgh Research Explorer

Tracing Molecular Gas Mass in z ⩾ 6 Galaxies with [C ii]

Tracing Molecular Gas Mass in $z \simeq 6$ Galaxies with [C II]

D. Vizgan[1,2,3,4], T. R. Greve[3,4,5] , K. P. Olsen[6] , A. Zanella[7] , D. Narayanan[3,8] , R. Davé[9,10] , G. E. Magdis[3,4,11] ,
G. Popping[12] , F. Valentino[3,11] , and K. E. Heintz[3,11,13]

[1] Astronomy Department and Van Vleck Observatory, Wesleyan University, 96 Foss Hill Drive, Middletown, CT 06459, USA
[2] Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 West Green Street, Urbana, IL 61801, USA
[3] The Cosmic Dawn Center, Denmark
[4] National Space Institute, DTU Space, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark
[5] Department of Physics and Astronomy, University College London, Gower Street, WC1E6BT London, UK
[6] Department of Astronomy and Steward Observatory, University of Arizona, Tucson, AZ 85721, USA
[7] Istituto Nazionale di Astrofisica (INAF), Vicolo dell'Osservatorio 5, I-35122 Padova, Italy
[8] Department of Astronomy, University of Florida, Gainesville, FL 32611, USA
[9] Institute for Astronomy, University of Edinburgh, Edinburgh EH9 3HJ, UK
[10] University of the Western Cape, Bellville, Cape Town 7535, South Africa
[11] Niels Bohr Institute, University of Copenhagen, Lyngbyvej 2, DK-2100 Copenhagen, Denmark
[12] European Southern Observatory, D-85478 Garching bei München, Germany
[13] Centre for Astrophysics and Cosmology, Science Institute, University of Iceland, Dunhagi 5, 107 Reykjavík, Iceland

Abstract

We investigate the fine-structure [C II] line at 158 μm as a molecular gas tracer by analyzing the relationship between molecular gas mass (M_{mol}) and [C II] line luminosity ($L_{[C II]}$) in 11,125 $z \simeq 6$ star-forming, main-sequence galaxies from the SIMBA simulations, with line emission modeled by the Simulator of Galaxy Millimeter/ Submillimeter Emission. Though most ($\sim$50%–100%) of the gas mass in our simulations is ionized, the bulk ($>$50%) of the [C II] emission comes from the molecular phase. We find a sublinear (slope 0.78 ± 0.01) $\log L_{[C II]} - \log M_{mol}$ relation, in contrast with the linear relation derived from observational samples of more massive, metal-rich galaxies at $z \lesssim 6$. We derive a median [C II]-to-M_{mol} conversion factor of $\alpha_{[C II]} \simeq 18\, M_{\odot}/L_{\odot}$. This is lower than the average value of $\simeq 30\, M_{\odot}/L_{\odot}$ derived from observations, which we attribute to lower gas-phase metallicities in our simulations. Thus, a lower, luminosity-dependent conversion factor must be applied when inferring molecular gas masses from [C II] observations of low-mass galaxies. For our simulations, [C II] is a better tracer of the molecular gas than CO $J = 1$–0, especially at the lowest metallicities, where much of the gas is *CO-dark*. We find that $L_{[C II]}$ is more tightly correlated with M_{mol} than with star formation rate (SFR), and both the $\log L_{[C II]} - \log M_{mol}$ and $\log L_{[C II]} - \log$ SFR relations arise from the Kennicutt–Schmidt relation. Our findings suggest that $L_{[C II]}$ is a promising tracer of the molecular gas at the earliest cosmic epochs.

1. Introduction

Molecular gas and star formation are observed to be tightly correlated (e.g., Kennicutt 1998; Kennicutt & Evans 2012; Bolatto et al. 2013) in our Galaxy, as well as in local and high-redshift galaxies, with the possible exception of star formation in extremely metal-poor environments where stars may form directly from atomic gas inflows (Krumholz 2012; Michałowski et al. 2015). Unfortunately, the dominant component of molecular gas, H_2, is invisible under the typical conditions prevailing in the molecular interstellar medium (ISM). Due to its lack of a permanent dipole moment, H_2 only emits from warm ($>$100 K) gas (Draine 2011), and its two lowest rotational transitions have upper energy levels of 510 and 1015 K. The low ($<$10–50 K) temperatures typical of the giant molecular clouds (Cox 2005), which make up the bulk of the star formation fuel, are unable to excite H_2 and are therefore "H_2-dark" to observers. Our inability to directly trace the distribution and amount of molecular gas in galaxies is a long-standing challenge for extragalactic studies and has led to the development of a number

of indirect H_2-tracing methods (Carilli & Walter 2013; Li et al. 2018), which we briefly summarize below.

The $J = 1$–0 rotational transition line of carbon monoxide (CO) is the most rigorously calibrated and most regularly used tracer of molecular gas mass (e.g., Dickman et al. 1986; Solomon & Barrett 1991; Downes & Solomon 1998; Narayanan et al. 2012; Bolatto et al. 2013). The utility of the CO line comes from the fact that it is relatively bright and easily excited in molecular ISM conditions.

To convert the observed CO (1–0) line luminosity into a molecular gas mass estimate, a CO-to-H_2 conversion factor, α_{CO}, is needed. The conversion factor is calibrated for giant molecular clouds (GMCs) in our Galaxy and nearby galaxies (e.g., Bolatto et al. 2013). A galaxy-averaged value for α_{CO} has been determined for local IR-luminous galaxies and found to be $\sim$4$\times$ lower than the Galactic value (Downes & Solomon 1998), albeit with significant scatter (Papadopoulos et al. 2012). For high-z galaxies, we do not have strong constraints on α_{CO} (Tacconi et al. 2013, 2018; Valentino et al. 2020a); both observations and simulations suggest that α_{CO} depends on the gas temperature, velocity dispersion, and metallicity (Tacconi et al. 2008; Ivison et al. 2011). In regions of high gas temperature and velocity dispersion, the optically thick CO (1–0) emission line emits more luminosity per unit H_2 column

density, which leads to lower conversion factors as seen in the local IR-luminous galaxies. In low-metallicity galaxies, α_{CO} scales inversely with metallicity since in such environments CO is photodissociated in a larger fraction of the molecular gas, which is thus referred to as "CO-dark" gas (e.g., Wilson 1995; Arimoto et al. 1996; Wolfire et al. 2010; Genzel et al. 2012; Narayanan et al. 2012). Intense cosmic-ray (CR) ionization rates that come along with extreme star formation activity can destroy CO deep within molecular clouds, thereby further degrading its H_2-tracing capabilities (Bisbas et al. 2015, 2017). Furthermore, at high redshifts, the cosmic microwave background (CMB) becomes brighter than the low-J CO emission from cold molecular gas; for example, da Cunha et al. (2013) showed that at $z \gtrsim 3$ the CMB could significantly affect both CO (1–0) and CO (2–1) line fluxes, and neglecting these effects can lead to nonnegligible underestimations of the derived luminosities and, therefore, the gas masses (see also Tunnard & Greve 2016; Zhang et al. 2016).

The fine-structure transitions of neutral carbon—[C I]($^3P_1-^3P_0$) and [C I]($^3P_2-^3P_1$)—have also been proposed as tracers of the molecular gas in galaxies (Gerin & Phillips 2000; Papadopoulos & Greve 2004; Narayanan & Krumholz 2017). Carbon is abundant ($\sim 10^{-5}$ relative to H_2), yet the [C I] lines have low to moderate optical depths ($\tau_{[C I]} \sim 0.1$–1). Both lines are easily excited in typical molecular gas conditions, which, together with its simple three-level partition function, makes it a promising tracer of H_2. [C I] surveys of molecular clouds in our Galaxy (Ikeda et al. 2002) showed a remarkable spatial and kinematical similarity in the [C I] and CO emission, suggesting that [C I] does indeed trace the bulk H_2 reservoir. Simulations of molecular clouds have shown that cloud inhomogeneity (resulting in deeper UV penetration and thus an increase in the [C I] layer; Offner et al. 2014) and turbulence (smoothing any initial [C I]/CO abundance gradient; Glover et al. 2015) are responsible for the observed [C I]–CO concomitant. Offner et al. (2014) further find that over a wide range of column densities [C I] is a comparable or superior tracer of the molecular gas, especially when an intense UV field is impinging on it. Another work on absorption lines in high-redshift galaxies suggests a universal metallicity-dependent [C I]-to-H_2 conversion factor (Heintz & Watson 2020). Recent work (Valentino et al. 2018, 2020b) has also investigated the utility of [C I] as a gas mass tracer in the high-z universe.

The far-IR (FIR)/submillimeter dust continuum emission has been used as an indirect tracer of the total gas content (molecular and atomic) of the Galactic ISM (Hildebrand 1983), of nearby galaxies (Guelin et al. 1993, 1995; Eales et al. 2012), and of high-z galaxies (Magdis et al. 2012; Scoville et al. 2014, 2016; Liang et al. 2018; Privon et al. 2018; Kaasinen et al. 2019). One approach has been to derive the dust mass from careful modeling of the FIR/submillimeter spectral energy distribution (SED) and apply a dust-to-gas mass ratio. The problem with this method is not only the uncertainties associated with the SED fitting but also a number of other factors, such as the dust temperature, the grain emissivity, and the dust-to-gas mass conversion factor. The latter depends on metallicity (Rémy-Ruyer et al. 2014; Li et al. 2019), which implies a redshift dependence via the mass–metallicity relation. Furthermore, this method requires flux measurements over a range of FIR/(sub)millimeter bands. To circumvent these problems, a direct calibration between the gas mass and the dust continuum emission at a single, broadband (sub)millimeter wavelength has been sought (e.g., Scoville et al. 2014). Such

calibrations have been established for optically thin continuum emission along the Rayleigh–Jeans tail, e.g., at 850 μm, and successfully applied to both local and high-z galaxies (Scoville et al. 2014, 2016). This method has the advantage over CO and [C I] in that it is easier to detect the continuum emission in a single broadband than it is to detect a single line. At redshifts $\gtrsim 4$–5, however, this method also runs into the limitations set by the increasing CMB temperature, which will not only couple thermally to the cold dust but also "drown out" a significant fraction of the dust continuum emission (e.g., da Cunha et al. 2013; Zhang et al. 2016).

In recent years, there has been a growing effort, from both the observation and simulation sides, to explore the 158 μm (1900.5 GHz) fine-structure transition of singly ionized carbon ([C II]) as a molecular gas mass tracer (Accurso et al. 2017; Hughes et al. 2017; Zanella et al. 2018; Dessauges-Zavadsky et al. 2020; Madden et al. 2020). The [C II]158 μm line is one of the strongest cooling lines of the ISM and can carry up to a few percent of the total FIR energy emitted from galaxies (e.g., Malhotra et al. 1997). Carbon is the fourth most abundant element and has an ionization potential of 11.3 eV, lower than that of neutral hydrogen. [C II], therefore, permeates much of the ISM. It is found in photodissociation regions (PDRs), diffuse ionized and atomic regions, and even molecular gas (e.g., Kaufman et al. 1999). Because the [C II] line arises in most phases of the ISM, it is important to disentangle the different contributions to interpret galaxy-wide, integrated [C II] observations correctly. How much of the [C II] emission comes from the molecular phase, and to what extent the line can be used as a molecular gas tracer, is of particular interest. Observations of [C II] in the Milky Way, for instance, show that $\sim 75\%$ of the [C II] emission comes from dense PDRs and CO-dark H_2 gas (Pineda et al. 2014).

Because of its association with PDRs, [C II] has been viewed as a tracer of the star formation rate (SFR; e.g., De Looze et al. 2014; Magdis et al. 2014; Herrera-Camus et al. 2015; Schaerer et al. 2020). At low redshift, metal-poor dwarf galaxies and star-forming galaxies display slightly offset but tight loglinear relations between the [C II] luminosity and the sum of total obscured and unobscured SFR (De Looze et al. 2014). However, compact starbursts with high SFR surface densities and intense UV radiation fields tend to exhibit a "[C II]/FIR deficit" by as much as an order of magnitude compared to the local [C II]–SFR relations (e.g., Díaz-Santos et al. 2013; Magdis et al. 2014; Narayanan & Krumholz 2017). The [C II]/FIR deficit is typically observed for local ultraluminous IR galaxies (ULIRGs; $L_{IR} \gtrsim 10^{12} L_\odot$) and high-$z$ starbursts/mergers, while normal star-forming galaxies at high redshifts are in line with the local [C II]–SFR relationship (e.g., Carniani et al. 2018). Schaerer et al. (2020) also found that there is little evolution in the [C II]–SFR relationship for main-sequence galaxies both locally and in the high-redshift (i.e., $4 < z < 8$) universe, although the scatter is larger at higher redshift.

By extension, this explains why [C II] traces molecular gas since the Kennicutt–Schmidt (KS) law shows a strong linear relationship between SFR and molecular gas mass. However, this relationship is not unique; from observations, normal galaxies seem to differ from starbursts/mergers in that the latter seem to deplete their gas reservoirs around 10 times faster. As pointed out by Zanella et al. (2018), if one assumes that [C II] traces molecular gas mass, the [C II]/IR relationship becomes M_{mol}/SFR, which is the gas depletion timescale of a galaxy. In

this case, the observed "FIR deficit" of starbursts reflects their $10\times$ shorter gas depletion timescales compared to normal star-forming galaxies. Theoretically, a strong connection between [C II] luminosity and molecular gas mass is also expected from the fact that the low critical density of carbon (11.3 eV) allows for the excitation of [C II] in dense, molecular regions by collisions with H_2 molecules. Indeed, simulations of high-redshift normal galaxies (e.g., Vallini et al. 2015; Popping et al. 2019; Leung et al. 2020) found that most [C II] emission comes from dense PDRs associated with regions of molecular gas; though Olsen et al. (2017) found that most (∼64%) of [C II] emission arose from ionized gas clouds in their simulations, they amended this result in an erratum, after accounting for the CMB in CLOUDY, and found that about 50% of [C II] emission arose from molecular gas clouds (see Olsen et al. 2018).

The use of [C II] as a molecular gas mass tracer has been implemented on a sample of 10 main-sequence galaxies at $z \sim 2$ by Zanella et al. (2018, hereafter Z18). The authors combine Band 9 Atacama Large Millimeter/submillimeter Array (ALMA) observations of [C II] emission with existing multiwavelength observations to derive an average [C II] mass-to-luminosity conversion factor $\alpha = 31\ M_\odot/L_\odot$ with an uncertainty of 0.3 dex. The molecular gas mass was derived from the integrated KS relation using the SFR measured from pixel-by-pixel SED fitting, while an independent and consistent gas mass estimate was derived from the dust continuum. The resulting conversion factor was found to be mostly invariant with galaxies' redshift, depletion time, and gas-phase metallicity. At lower redshifts [C II] has also been demonstrated to trace molecular gas mass (Madden et al. 2020). Notably, [C II] has also been shown to trace atomic hydrogen in galaxies until the epoch of reionization (Heintz et al. 2021).

The aim of this paper is to investigate whether [C II] is still a reliable tracer of the molecular gas mass for lower-mass and lower-metallicity galaxies at earlier epochs (i.e., $z \approx 6$). Repeating the experiment on a much larger sample of simulated high-redshift galaxies, we aim to derive a conversion factor for the molecular gas mass of high-redshift galaxies using [C II] line luminosities, as has been done in the nearby universe ($z \sim 0.03$–0.2) for CO (Hughes et al. 2017), and as done for galaxies in the local Herschel Dwarf Galaxy Survey for [C II] (see Madden et al. 2020).

In this paper we present Simulator of Galaxy Millimeter/Submillimeter Emission (SÍGAME) simulations of [C II] line emission from 11,125 galaxies at $z \simeq 6$ in order to examine whether the line can be used to trace the gas content in normal star-forming galaxies at cosmic dawn. Section 2 of the paper describes the observation samples with which we compare our simulations in this work. Section 3 describes the simulation sample and the post-processing done by SÍGAME. Section 4 compares the integrated physical properties of the simulation sample in the context of observations. Section 5 examines [C II] as a tracer of molecular gas mass and as a tracer of SFR, derives a [C II]-to-H_2 conversion factor ($\alpha_{C\ II}$), and investigates the origin of the M_{mol}–$L_{C\ II}$ relation. Section 6 summarizes the main conclusions from our work.

2. Observed Comparison Samples

We employ two observed galaxy samples for comparison with our simulations. The first is from Z18, who presented six [C II] detections out of a sample of 10 $z \sim 2$ main-sequence

galaxies observed with ALMA and combined these with other [C II] detections from the literature. The latter spanned the redshift range $0 \leqslant z \leqslant 6$ and included diverse galaxy populations, from local dwarfs and normal star-forming galaxies to luminous starbursts. Independent estimates of the molecular gas masses for all of these galaxies were given in Z18, mostly from existing observations of CO but also via more indirect methods such as from parameterizations of the gas depletion time ($\tau_{depl} = M_{mol}/SFR$) as a function of the specific SFR ($sSFR = SFR/M_\star$). In this paper, we split the Z18 sample into $z \geqslant 4$ and $z < 4$ galaxies.

We additionally compare our simulations to results obtained from the ALPINE-ALMA [C II] Survey (henceforth referred to as ALPINE; see Dessauges-Zavadsky et al. 2020; Le Fèvre et al. 2020). ALPINE has observed 118 galaxies in the redshift range $4 < z < 6$ for [C II] and dust FIR emission. In 64% of galaxies observed, [C II] was detected. These 75 galaxies, with which we compare our simulations, consist mainly of main-sequence galaxies, with only a handful of sources lying either above or below the main sequence at these redshifts. The molecular gas masses of the ALPINE sample have been estimated from their rest-frame 850 μm continuum luminosity extrapolated from rest-frame 158 μm continuum; see Dessauges-Zavadsky et al. (2020) for details. The SFRs of the ALPINE sample are derived, via Equation (4) in Bell (2003), from their IR luminosities, which were inferred from the IRX–β relation (see Fudamoto et al. 2020). These L_{IR}-derived SFRs were added to the UV-derived SFRs (Faisst et al. 2020) in order to get the total SFRs.

We use the Z18 and ALPINE samples in combination with our simulations (see Section 3) to examine [C II] as a tracer of molecular gas.

3. Simulations

3.1. Hydrodynamic Simulations

This work builds on the analysis of snapshots taken from the SIMBA suite of cosmological galaxy formation simulations, which themselves were evolved using the meshless finite-mass hydrodynamics technique of GIZMO (Hopkins 2015, 2017; Davé et al. 2019). The SIMBA simulation set consists of three cubical volumes of 25, 50, and 100 cMpc h^{-1} on a side, all of which are used in this work to search for galaxies at $z \sim 6$. For each volume, a total of 1024^3 gas elements and 1024^3 dark matter particles are evolved from $z = 249$. Compared to its predecessor, MUFASA, new features in SIMBA include the growth and feedback of supermassive black holes, as well as a subgrid model to form and destroy dust during the simulation run; for further details we refer to Davé et al. (2019) and Li et al. (2018) for these two processes, respectively. The galaxy properties in SIMBA have been compared to various observations across cosmic time (Thomas et al. 2019; Appleby et al. 2020), including the epoch of reionization (Leung et al. 2020; Wu et al. 2020), and are in reasonable agreement.

Since version 2 of SÍGAME used here relies heavily on the molecular gas mass fraction in the simulated galaxies, we briefly describe how this fraction is derived in SIMBA, referring to Davé et al. (2020) for a more detailed description. The molecular gas mass content of each fluid element in SIMBA is calculated at each time step following the subgrid prescription of Krumholz & Gnedin (2011). This prescription relies on local metallicity and gas column density, modified to account for

variations in resolution (Davé et al. 2016). The H_2 mass function (H2MF) was found to be in good agreement with observations at redshifts $z \sim 0$–2, albeit overpredicting the H2MF at the high-mass end (Davé et al. 2020).

Importantly, when comparing the [C II] luminosity derived with SÍGAME to the molecular gas content of the simulated galaxies, we will not be using the molecular gas mass fractions from SIMBA, but rather the recalculated molecular gas masses from CLOUDY, as described in the following section. This is because the molecular gas masses as calculated in SIMBA are rough estimates based on local gas densities ($\sim 1 \times 10^2$ cm^3) and metallicities within the simulations, whereas the post-processed molecular gas masses from CLOUDY account for the much higher densities ($\sim 10^5$–10^6 cm^3) of the molecular clouds in the simulations. As a consistency check, we compared the SÍGAME versus SIMBA molecular gas masses and found, as expected, a strong correlation between the two quantities.

From the SIMBA simulations, 11,137 galaxies were selected via the yt-based (Turk et al. 2011) package CAESAR; YT is a Python package used to analyze and visualize volumetric data, and CAESAR is a six-dimensional friends-of-friends algorithm that is applied to the simulated gas and stellar particles to identify and select SIMBA galaxies. Our final sample consists of 11,125 galaxies between a [C II] luminosity range of $10^{3.82}$–$10^{8.91}$ $L_\odot$ and a molecular gas mass range of $10^{6.25}$–$10^{10.33}$ $M_\odot$. The galaxy properties derived with CAESAR include SFR, computed by dividing the stellar mass formed over a 100 Myr timescale (Leung et al. 2020); SFR surface density; stellar mass ($M_\star$); and total gas mass. In addition, we derive an SFR-weighted gas-phase metallicity, Z. All of these properties will be used for the analysis in Section 4.

3.2. SÍGAME Post-processing

The SIMBA galaxies are post-processed with version 2 of the SÍGAME module (Olsen et al. 2017).[14] At its core, this version of SÍGAME uses the spectral synthesis code CLOUDY (v17.01; Ferland et al. 2013, 2017) to model the line emission from the multiphased ISM within each simulated galaxy. Using physically motivated prescriptions, SÍGAME calculates, throughout the simulated galaxy, the local interstellar radiation field (ISRF) spectrum, the CR ionization rate, and the gas density distribution of the ionized, atomic, and molecular ISM phases. The local equilibrium gas temperatures are calculated for each ISM phase by balancing the gas heating and cooling.

SÍGAME calculates the local ISRF that impinges on a gas particle in the simulations as the sum of the radiation field from stars within one smoothing length of that gas particle. The radiation from a stellar particle is the integrated spectrum from STARBURST99 modeling of the stellar population contained in the stellar particle, based on its mass, age, and metallicity. The local CR ionization rate is assumed to scale linearly with the local ISRF far-UV strength. As for the metal abundances, gas-phase metallicities are carried over from the SIMBA simulations. Finally, the gas density requires a subgridding of the fluid elements, which are themselves too large ($\geqslant 10^5$ $M_\odot$) to follow the intricate processes taking place inside GMCs of mass 10^4–10^6 $M_\odot$. This subgridding is carried out via subdivision of the gas into three ISM phases. First, the gas mass of each fluid element is divided into a dense and a diffuse part based on the H_2 gas mass fraction of the simulation itself. The diffuse gas is

further subdivided into neutral and ionized diffuse gas, leading to three ISM phases: dense gas that is later on organized in GMCs, diffuse neutral gas, and diffuse ionized gas. Finally, the level populations and line emission are derived with CLOUDY, which also provides mass fractions of atomic, ionized, and molecular hydrogen. It is from these mass fractions that we derive the final molecular gas mass of each galaxy.

In general, the molecular gas masses recalculated in this manner do not exceed those of the original SIMBA simulation, as that sets the maximum mass of all GMCs, of which CLOUDY will deem a certain fraction (close to 100%) to be molecular. For a full description of the main methods and assumptions of SÍGAME, we refer to Olsen et al. (2017) and references therein. For this paper, we have used a slightly updated version of SÍGAME, and we refer to Leung et al. (2020) for a description of these updates and the creation of the data set analyzed here.

4. Sample Properties

4.1. Integrated Properties

In this section we perform a series of sanity checks of our simulated galaxies by examining their SFRs, masses (stellar and molecular), and metallicities in the context of known empirical correlations between these quantities. Before doing so, we stress that observed integrated properties such as stellar and molecular gas masses are likely to be uncertain by at least a factor of two (Lower et al. 2020; Gilda et al. 2021).

Figure 1(a) shows the location of our simulated galaxies in the SFR–$M_\star$ plane, where they define a sequence that extends over more than three orders of magnitude in stellar mass. The three "clouds" of simulations correspond with the three volumes of SIMBA galaxies (25, 50, 100 cMpc h^{-1}; see Section 3). The cloud with the highest-mass galaxies corresponds with the largest volume, and the cloud with the lowest-mass galaxies corresponds with the lowest volume. As explained in Leung et al. (2020), the galaxies in the larger volumes have a larger smoothing length and lower initial gas mass resolution than in the lowest volume.

Also shown in Figure 1(a) are the ALPINE and $z \geqslant 4$ Z18 samples. The galaxy main sequence is usually modeled by a loglinear relation of the form $\log \mathrm{SFR} = \alpha \log M_\star + \beta$, where the parameters α and β are functions of cosmic time, or redshift. The normalization, β, in particular, is found to increase at earlier cosmic epochs. Observational constraints on the galaxy main sequence at $z \simeq 5$–6 come from Speagle et al. (2014), Schreiber et al. (2015), Salmon et al. (2015), Pearson et al. (2018), and Khusanova et al. (2020). The Speagle et al. (2014) main sequence is based on observed galaxy samples at $z \simeq 4$–6 with $M_\star = 10^{9.5}$–$10^{11.5} M_\odot$, while Schreiber et al. (2015) derived their main sequence from a small sample of galaxies at $z \simeq 3.5$–5 with $M_\star \geqslant 10^{11} M_\odot$ and an extrapolation to $z \simeq 6$. Finally, the Salmon et al. (2015) main sequence is based on a sample of $z \simeq 6$ galaxies with stellar masses in the range $M_\star = 10^9$–$10^{10.5} M_\odot$. While the main-sequence parameterizations by Speagle et al. (2014), Salmon et al. (2015), and Khusanova et al. (2020) are in agreement with the ALPINE and Z18 data, the parameterizations by Schreiber et al. (2015) and Pearson et al. (2018) appear to have normalizations that fall below those of the data (although the slopes appear to match). This may be because the Schreiber et al. (2015) fit was made using $z < 4$ data and is thus an extrapolation toward higher redshift. Nevertheless, the main-sequence parameterizations by
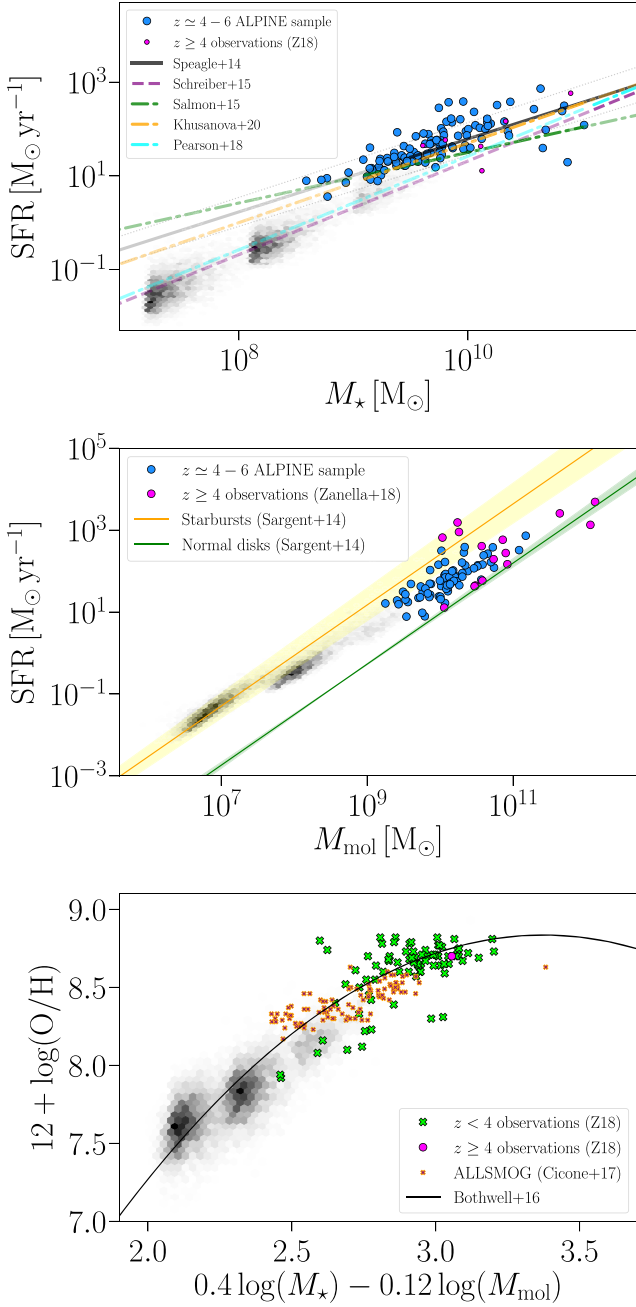
---

[14] https://kpolsen.github.io/SIGAME/index.html

Figure 1. Top: the SFR vs. $M_\star$ plane with the distribution of our simulated galaxies (gray hex bins) shown, along with the ALPINE sample (blue filled circles) and the $z > 4$ galaxies from Z18 (magenta filled circles). The lines show best-fit power laws to observations of the galaxy main sequence at $z \sim 5$–6 from Speagle et al. (2014) (solid black lines, with 1σ scatter indicated by dotted lines), Schreiber et al. (2015) (dashed purple), and Salmon et al. (2015) (dotted–dashed green). The highlighted segments of the lines indicate the stellar mass ranges of the galaxy samples used to infer the main sequence. Middle: the integrated KS relation (SFR vs. $M_{\rm mol}$) for our simulated and the ALPINE and $z > 4$ Z18 samples. Where they overlap, the simulations and observations follow the same relation, although the latter exhibit a larger scatter. Bottom: the molecular fundamental mass–metallicity relation, i.e., metallicity (12 + log(O/H)) vs. the optimum linear combination of $\log M_\star$ and $\log M_{\rm H_2}$ as determined from a PCA of the local galaxy sample ALLSMOG (see Bothwell et al. 2016 for details). The solid curve represents the quadratic fit to the ALLSMOG data, here shown using the revised data values from Cicone et al. (2017). Our simulations and the full Z18 sample are in good agreement with the ALLSMOG results.

Schreiber et al. (2015) and Pearson et al. (2018) are seen to match our simulations across the full stellar mass range spanned by the simulations, although these are based on extrapolations to lower masses than those probed by the observations.

For stellar masses $\geqslant 2 \times 10^9 \, M_\odot$ there is fair agreement between our simulations and the ALPINE and $z \geqslant 4$ Z18 data. At lower stellar masses (down to $\sim 2 \times 10^8 \, M_\odot$) there is an apparently increasing offset between the simulations and the observational data, although there are few observational data available at these low masses to warrant a fair comparison. We note that both the ALPINE and Z18 data may be somewhat biased toward high SFR values simply by virtue of having SFR estimates at these low masses. Furthermore, there may be a selection bias in the ALPINE sources due to the selected parent sample of ALPINE; this is discussed further in Khusanova et al. (2021).

Figure 1(b) shows SFR versus $M_{\rm mol}$ for our simulated galaxies, along with $z \geqslant 4$ galaxies from our comparison samples. Our simulated galaxies trace out a correlation that is consistent with the galaxy-integrated version of the KS law (Kennicutt 1998). We show parameterizations of the redshift-independent galaxy-integrated KS law for starbursts (yellow line) and normal disk galaxies (green line) from Sargent et al. (2014). The galaxies simulated within the 25 cMpc h^{-1} box are seen to agree with the KS law for starbursts, while the simulated galaxies extracted from the 50 cMpc h^{-1} box fall slightly below the starburst KS law. Galaxies from the 100 cMpc h^{-1} box—as well as the observed samples—fall in between the KS laws for starbursts and disks. They agree well with the Z18 and ALPINE samples, which are seen to also straddle the KS laws for starbursts and normal disk galaxies.

Finally, in Figure 1(c) we examine the relationship between gas-phase metallicity, $M_\star$, and $M_{\rm mol}$. Bothwell et al. (2016) showed that the SFR–metallicity relation at a given stellar mass (the so-called fundamental metallicity relation) is in fact a result of an underlying relation between metallicity and molecular gas mass. Using the ALLSMOG sample of nearby galaxies (Bothwell et al. 2016; Cicone et al. 2017), they derived a relation between metallicity (12 + log(O/H)), $M_\star$, and $M_{\rm mol}$. This relation is shown as the solid line in Figure 1(c). It represents the linear combination in $\log M_\star$ and $\log M_{\rm H_2}$, which yields the least scatter in metallicity (see Bothwell et al. 2016 for details).

More recently, however, Cicone et al. (2017) reanalyzed the ALLSMOG sample and derived improved metallicities, as well as stellar and molecular gas masses. Using these revised and final values, we plot the ALLSMOG sample in Figure 1(c) (orange symbols). We find that the galaxies from Z18 are consistent with the ALLSMOG sample, independent of their redshifts, and even extend the relation to the point where it turns over. To overlay the SIMBA simulations, we use the calibration 12 + log(O/H) = log($\langle Z \rangle_{\rm mass}$) + 9 derived by Ma et al. (2016) using high-resolution cosmological simulations over the redshift range $z = 0$–6. Ma et al. (2016) use the mass-weighted average gas-phase metallicity, $\langle Z \rangle_{\rm mass}$, and for that reason we have used this quantity from our simulations, and not the SFR-weighted average gas-phase metallicities (see Section 3.1), to infer 12 + log(O/H). Our simulations are seen to be in good agreement with the observations in the region
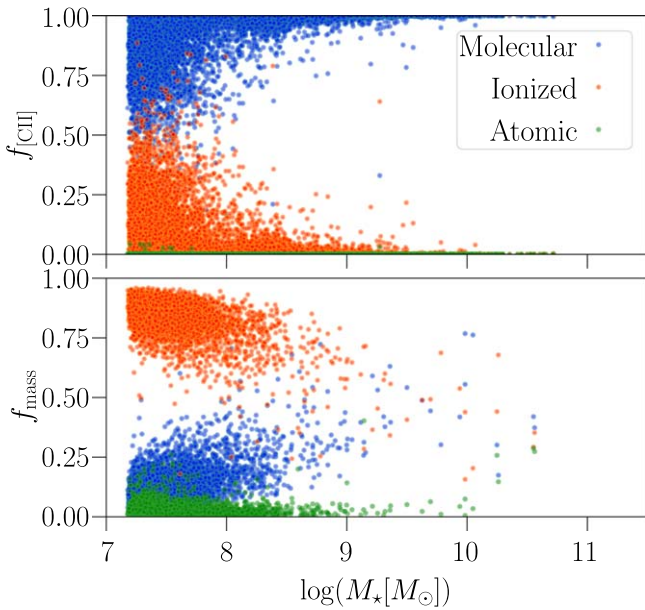
Figure 2. Top: the fraction of the total [C II] luminosity emitted by the molecular (blue), ionized (red), and atomic (green) ISM phases for each galaxy as a function of its stellar mass. Bottom: the same as in the top panel, but now showing the gas mass fraction of the different ISM phases.

where there is overlap, and they show a declining trend toward lower metallicities.

The analysis above demonstrates that within the range of masses where a comparison with observations is possible, our simulated SIMBA galaxies are in overall agreement with key scaling relations observed for $z \geqslant 4$ main-sequence galaxies. This suggests that the simulations are representative, even at the low stellar mass end that extends well below the observed range, of the bulk of the normal star-forming galaxies at these epochs.

4.2. ISM Properties

Before we explore the H_2-tracing capabilities of [C II], it is instructive to first examine the gas mass fractions of the three ISM phases in our simulations and their [C II] emission properties. We note here that the diffuse, molecular, and atomic gas masses are calculated to be the total mass of all gas particles associated with each respective phase according to CAESAR. However, for observations these phases are typically calculated within a certain radius. Therefore, the gas mass fractions are calculated to be independent of galactic region or radius for our simulations.

Figure 2 shows fractions of total [C II] luminosity ($f_{C\,II}$) and ISM gas mass (f_{mass}) as a function of stellar mass. It is seen that the molecular phase dominates the [C II] emission: it contributes ∼50% to the total [C II] emission for galaxies with $M_\star \sim 10^7\,M_\odot$ and ∼100% for $M_\star \gtrsim 10^9\,M_\odot$ galaxies. Over the same stellar mass range, the ionized gas-phase contribution to the total [C II] emission spans ∼50% at the low-mass end to only a few percent at the high-mass end. The atomic gas hardly contributes anything to the [C II] emission. Looking at the gas mass fraction, the ionized and molecular gas phases dominate, with the former accounting for ∼50%–100% and the latter with ∼10%–75%. Thus, while the molecular gas does not account for most of the ISM mass in our simulated galaxies, except for

in the most massive galaxies ($M_{star} \gtrsim 10^{9.5}\,M_\odot$), it is responsible for the bulk of the total [C II] emission.

In contrast, the atomic phase makes up $\lesssim$10% of the total ISM mass, and even less of the total [C II] emission. The low mass fraction for the atomic gas is somewhat surprising, given that the cosmic atomic gas density exceeds that of molecular gas at redshifts $z \gtrsim 3$ (Walter et al. 2020). We believe that this is due to the way that gas is classified as "ionized" in SÍGAME, which is simply when the electron-to-hydrogen fraction is $x_{e^-} > 0.5$. As a result, some of the atomic gas is lumped in with the ionized gas, and both the true atomic gas mass and [C II] emission fraction are thus likely to be higher than what is indicated in Figure 2.

The fraction of the [C II] emission coming from the molecular phase in our simulations is consistent with what is found in local star-forming galaxies (∼60%–80%; Accurso et al. 2017), as well as in high-resolution simulations of $z \simeq 6$ Lyman break galaxies (∼95%; Pallottini et al. 2017). The contribution to the [C II] emission from the ionized phase is seen to increase as the stellar mass, and thus the metallicity, decreases. Although one would expect the [C II] cooling rate to decrease at lower metallicity, this effect is negligible compared to the increase in CO photodissociation rate, and thus available C^+ ions, that comes with lower metallicities; a similar effect was seen by Accurso et al. (2017). The fact that the bulk of the [C II] emission is coming from the molecular gas suggests that we can indeed use the line as a molecular gas tracer.

5. Results and Discussion

5.1. How Well Does [C II] Trace Molecular Gas Mass?

In this section we examine whether there is a correlation between the [C II] emission and the molecular gas masses of our simulated $z \simeq 6$ galaxies, and to what extent this correlation agrees with the empirical relation found by observations.

Figure 3 shows the [C II] luminosity versus the total molecular gas mass and $M_{mol}/L_{C\,II}$ versus M_{mol} for our simulated galaxies, along with observed $z < 4$ and $z \geqslant 4$ galaxies from Z18 and $z = 4$–6 galaxies from the ALPINE survey (Dessauges-Zavadsky et al. 2020). We fit a loglinear model of the form $\log L_{C\,II} = a \log M_{mol} + b$ to the simulations only, as well as the ALPINE data only. This yields the following relations.

Simulations:

$$\log L_{C\,II} = (0.78 \pm 0.01)\log M_{mol} + (0.44 \pm 0.04), \quad (1)$$

shown as the thick solid line in Figure 3. The rms scatter of the simulations around the fitted line is 0.45 dex.

ALPINE:

$$\log L_{C\,II} = (1.00 \pm 0.01)\log M_{mol} - (1.35 \pm 0.1), \quad (2)$$

shown as the thick dashed line in Figure 3, with an rms scatter of the data around the fitted line of 0.12 dex.

For comparison, Z18 derived a relation given by $\log L_{C\,II} = (0.98 \pm 0.02)\log M_{mol} - (1.28 \pm 0.01)$, which is shown as the thin dotted line in Figure 3. They reported an rms scatter of 0.2 dex. There is good agreement between the observed data from ALPINE and Z18 (see also Dessauges-Zavadsky et al. 2020), with the fit to the ALPINE data being very similar to the Z18 relation. The Z18 data exhibit somewhat larger scatter compared to the ALPINE data. This is likely due to the ALPINE galaxies being uniformly selected
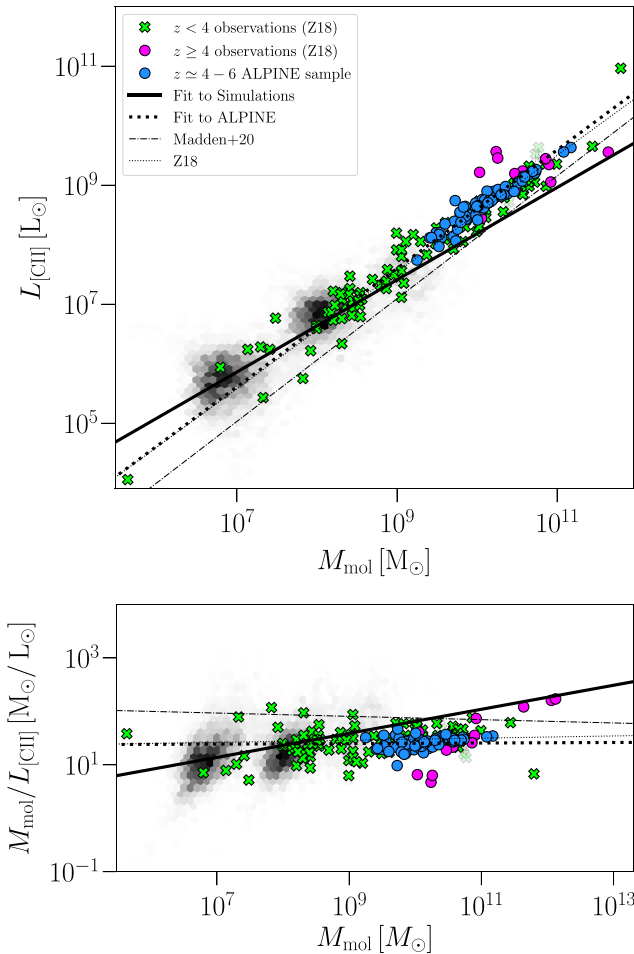
Figure 3. $L_{[C\,II]}$ vs. M_{mol} (top) and $M_{mol}/L_{[C\,II]}$ vs. M_{mol} (bottom) for our simulated $z = 6$ galaxies (gray hex bin) and observed comparison samples from Z18 and Dessauges-Zavadsky et al. (2020). Loglinear fits to our simulations (solid line), to the observed samples (dotted lined), and to simulations by Madden et al. (2020) of nearby dwarf galaxies (dotted–dashed line) are indicated.

as main-sequence galaxies at $z = 4$–6, while the Z18 sample is heterogeneous, consisting of local (U)LIRGs, as well as main-sequence galaxies and starbursts at low and high redshifts.

Our simulations appear to be in broad agreement with the observations, despite lying in a very different part of the "galaxy parameter space" (see Section 4). Upon closer inspection, it is seen from Figure 3 that our simulations contain a number of galaxies that exhibit higher [C II] luminosities given their gas masses than would be expected from the Z18 relation. On average, the simulations have lower $M_{mol}/L_{[C\,II]}$ values than the bulk of the observed data. This is most pronounced for the lowest gas mass simulations ($M_{mol} \lesssim 10^7\,M_\odot$), while the simulations with the highest gas masses ($M_{mol} \sim 10^9\,M_\odot$) have $M_{mol}/L_{[C\,II]}$ ratios in line with the observed data. The overall effect of this is a shallower slope in the simulations-only relation and a higher rms scatter (0.45 dex). The shallower slope ($\sim$0.78) suggested by our simulations implies an $L_{C\,II}$-dependent, i.e., nonuniversal, $M_{mol}/L_{C\,II}$ ratio for $z \sim 6$ galaxies (see Z18). The small number of [C II] observations of galaxies with gas masses $\lesssim 10^8\,M_\odot$ prevents us from assessing whether a similar trend is seen in observations.
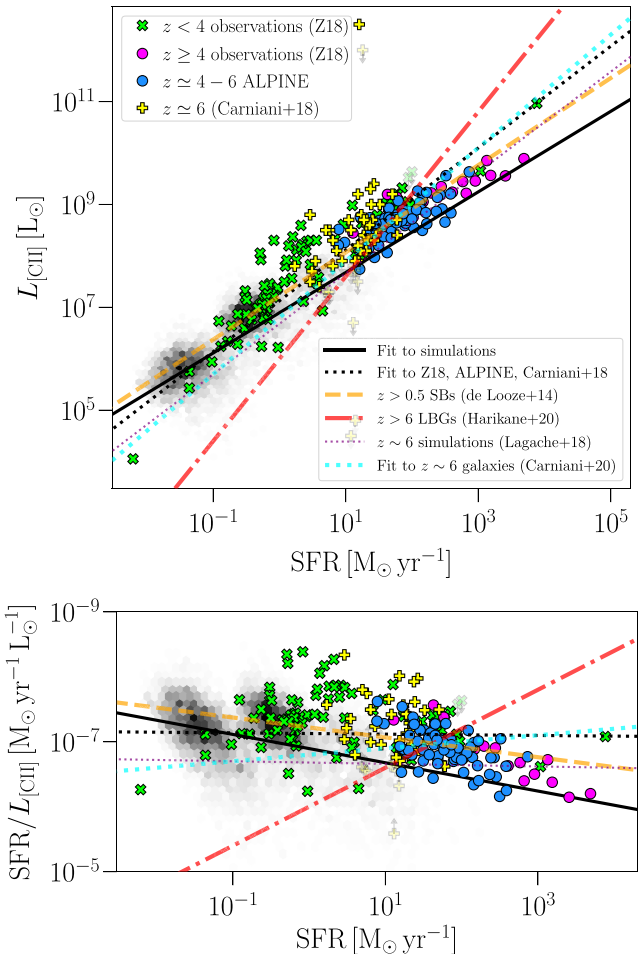

Figure 4. $L_{[C\,II]}$ vs. SFR (top) and SFR/$L_{[C\,II]}$ vs. SFR (bottom) for our simulated $z = 6$ galaxies (gray hex bin), and the observed comparison samples from Z18, Dessauges-Zavadsky et al. (2020), and Carniani et al. (2018) (circles and crosses).

From the $z \sim 0$ simulations presented by Madden et al. (2020), they derive $\log L_{C\,II} = 1.03 \log M_{mol} - 2.19$, with an rms scatter of 0.14 dex. This relation, shown as the thin dashed–dotted line in Figure 3, has a slope consistent with unity, i.e., similar to the Z18 and ALPINE relations, but the normalization is significantly lower. Using this relation would result in $\sim$3$\times$ larger molecular gas masses, which they ascribe to Z18 not accounting for CO-dark gas and thus using a CO-to-H_2 conversion factor that is too low (Madden et al. 2020). This explanation, however, is not borne out by our simulations, which examine the relationship between $L_{C\,II}$ and the full molecular gas reservoir (CO-dark or not).

5.2. How Well Does [C II] Trace Star Formation Rate?

The traditional interpretation of the [C II] line has been as a diagnostic of the physical conditions in photon-dominated regions and as a tracer of the star formation activity in galaxies (Stacey et al. 1991; De Looze et al. 2014). In Figure 4 we show $L_{C\,II}$ versus SFR for our simulations, the ALPINE sample, the Z18 sample, and the Carniani et al. (2018, hereafter C18) sample. Also shown is the loglinear fit to the combined Z18, C18, and ALPINE samples, as well as our simulations.

For comparison the global $L_{C\,II}$–SFR relation inferred for high-z starbursts by De Looze et al. (2014) is also shown, along with relations derived from observations (Harikane et al. 2020) and simulations (Lagache et al. 2018) of $z \simeq 6$ star-forming galaxies. The rms scatter of the $\log L_{[C\,II]}$–$\log$ SFR relation fitted to our simulations alone is 0.50 dex. A fit to the Z18, ALPINE, and C18 data exhibits an even larger scatter, i.e., 0.55 dex, which is higher than the scatter reported by De Looze et al. (2014) for high-z starbursts (0.40 dex). This is also higher than Schaerer et al. (2020), who found a scatter of 0.28 dex when plotting $L_{[C\,II]}$ versus SFR of the ALPINE sample after using the IRX–β relation (Fudamoto et al. 2020) to add an L_{IR}-derived SFR to the UV-derived SFRs (Faisst et al. 2020; see also Section 2).

The $\log L_{C\,II}$–$\log$ SFR relation exhibits more scatter than the $\log M_{mol}$–$\log L_{C\,II}$ relation, both for our simulations and for observed data, and we thus conclude that the [C II] luminosity is more tightly correlated with the molecular gas mass than the SFR. Our simulations are seen to agree well with the high-z relation derived by De Looze et al. (2014), who also predict a slightly decreasing SFR/$L_{C\,II}$ with increasing SFR. This trend can account for some of the scatter in the $\log L_{C\,II}$–$\log$ SFR. The scatter can also arise from other factors; we note, for instance, that SFR is prone to large observational systematics—particularly at high redshift—and this may be the reason why the scatter in the $\log L_{C\,II}$–$\log$ SFR relation is larger than the $\log M_{mol}$–$\log L_{C\,II}$ relation. Furthermore, the methods of measuring SFR, as well as uncertainties on SFR measurements, can affect the total scatter on the $\log L_{C\,II}$–$\log$ SFR relation for our observations.

5.3. The [C II] Conversion Factor

The fact that Z18 found the $L_{C\,II}$–M_{mol} relation to be loglinear with a slope close to unity allowed them to define a conversion factor of [C II] luminosity to molecular gas mass, $\alpha_{C\,II} = M_{mol}/L_{C\,II}$. They inferred a value of $\alpha_{C\,II} \sim 30\,M_\odot/L_\odot$, with an uncertainty of 0.3 dex averaged across their sample, and argued that $\alpha_{C\,II}$ was effectively invariant with redshift, metallicity, and galaxy type (e.g., starburst vs. main-sequence galaxy); see also Figure 3. For the ALPINE sample, we find a median conversion factor of $\alpha_{C\,II} \sim 27\,M_\odot/L_\odot$ and a median absolute deviation (MAD) of $4\,M_\odot/L_\odot$. From our simulated galaxies, we find a median value of $\alpha_{C\,II} \sim 18\,M_\odot/L_\odot$, with an MAD of $10\,M_\odot/L_\odot$.

In Figure 5 we plot $\alpha_{C\,II}$ versus z for our simulations and the Z18 and ALPINE samples, along with histograms of their overall $\alpha_{C\,II}$ distributions (top panel). The $\alpha_{C\,II}$-distribution for our simulated galaxies peaks at $15\,M_\odot/L_\odot$, with an extended tail toward higher values. The $\alpha_{C\,II}$-distribution for the Z18 subsample at $z < 4$ also exhibits a tail toward high $\alpha_{C\,II}$-values. In addition to high $\alpha_{C\,II}$-values, the simulated sample also contains many galaxies with $\alpha_{C\,II}$-values $< 10\,M_\odot/L_\odot$. These low values of $\alpha_{C\,II}$ derived from our simulations are in agreement with high-z [C II] observations by Rizzo et al. (2021), who inferred a median $\alpha_{C\,II} = 7^{+4}_{-1}\,M_\odot/L_\odot$ for five strongly lensed starburst galaxies between $4 \leqslant z \leqslant 5$ (see also Rizzo et al. 2020). Finally, we note that the Z18 subsample at $z \geqslant 4$ overlaps with the $\alpha_{C\,II}$-values of the ALPINE sample, and both are tightly distributed around their peak. For the ALPINE sample, in particular, this might be due to it being a relatively homogeneous sample of main-sequence galaxies. In contrast,
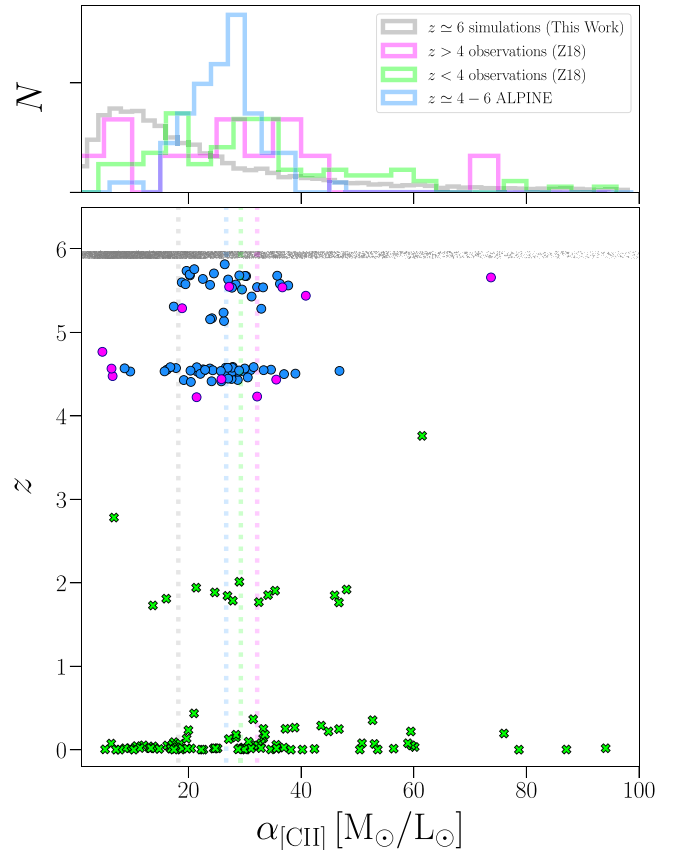

Figure 5. Top panel: the distributions of $\alpha_{C\,II}$ values for our simulations (gray horizontal band) and comparison samples. Bottom panel: $\alpha_{C\,II}$ vs. redshift for our simulations and comparison samples. The vertical dashed lines indicate the median $\alpha_{[C\,II]}$ values of the different samples. For clarity, our $z = 6$ simulations have been spread out in the z-direction slightly.
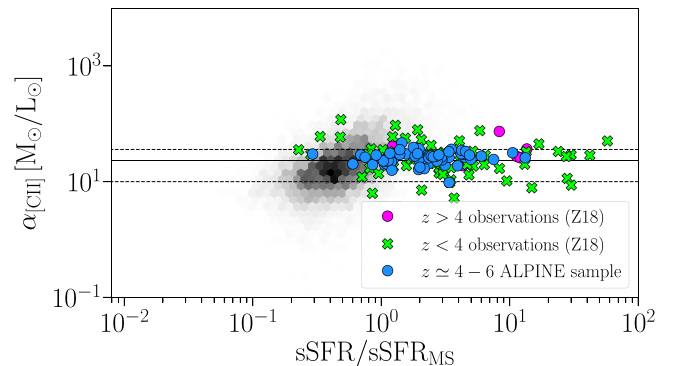

Figure 6. $\alpha_{[C\,II]}$ vs. offset from the main sequence for our simulations (gray hex bins) and observed galaxies from Z18 (green and magenta symbols) and the ALPINE survey (blue filled circles; Dessauges-Zavadsky et al. 2020).

the full Z18 sample and our simulations span a wider range in SFR and masses.

In Figure 6 we show $\alpha_{C\,II}$ versus offset from the main sequence for our simulated galaxies, which for a given galaxy is defined as the ratio of its specific SFR (sSFR = SFR/$M_\star$) and the specific star formation of a galaxy with the same stellar mass on the main-sequence relation (at the given redshift). We adopt the prescription for the main sequence at $z = 6$ as given by Schreiber et al. (2015), since it agrees well with our simulations (Section 4.1). The Z18 and ALPINE comparison samples have also been normalized with the Schreiber et al. (2015)
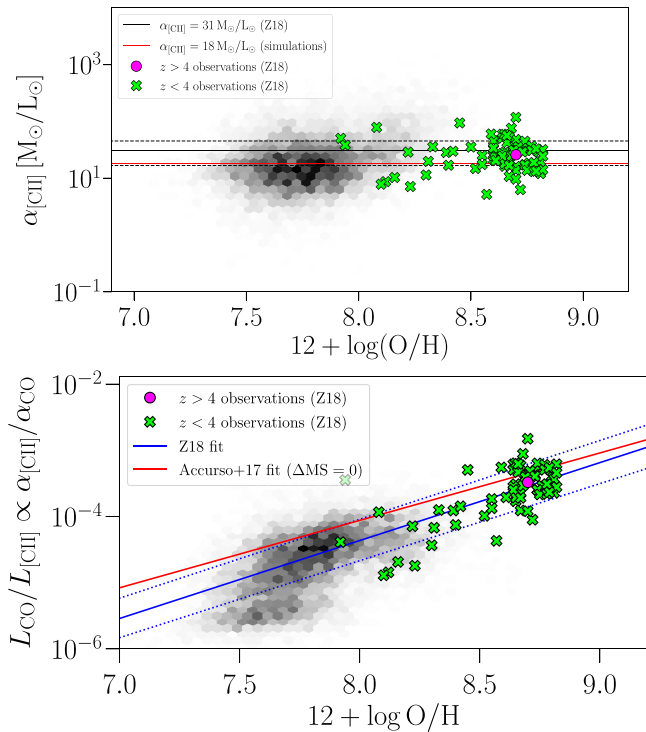
Figure 7. Top: $\alpha_{[\mathrm{C~II}]}$ vs. gas-phase metallicity for our simulated galaxies (gray hex bins), along with observed samples from Z18 (green and magenta symbols). The black solid line indicates the median $\alpha_{[\mathrm{C~II}]}$ value (31 $M_\odot/L_\odot$) found by Z18 and its 0.2 dex scatter (dotted lines), while the red line indicates the median $\alpha_{[\mathrm{C~II}]}$ value (18 $M_\odot/L_\odot$) of our $z = 6$ simulations. Bottom: the CO (1–0)-to-[C II] luminosity ratio (both in units of solar luminosities) vs. the gas-phase metallicity. The luminosity ratio, which scales with the ratio of the [C II]-to-CO (1–0) conversion factors ($\alpha_{\mathrm{C~II}}/\alpha_{\mathrm{CO}}$), is seen to strongly decrease with decreasing metallicity. This is a result of the increasing photodissociation of CO in unshielded environments in metal-poor galaxies. The blue and red lines are fits from Zanella et al. (2018) and Accurso et al. (2017), respectively.

main-sequence parameterizations at the relevant redshifts of each individual galaxy. Our simulations show a weak trend of increasing $\alpha_{[\mathrm{C~II}]}$ with main-sequence offset. In contrast, both the Z18 and ALPINE samples show constant $\alpha_{[\mathrm{C~II}]}$. This is true even for the sources with large offsets from the main sequence.

5.4. Is [C II] a Metallicity-invariant Tracer of Molecular Gas?

In Figure 7 (top) we show $\alpha_{\mathrm{C~II}}$ versus gas-phase metallicity (i.e., $12 + \log(\mathrm{O/H})$) for our simulations, along with the Z18 sample (note that we do not have metallicity estimates for the ALPINE sample). Across the metallicity range probed by our simulations ($12 + \log \mathrm{O/H} \sim 7.0$ to ~ 9.0), the $\alpha_{\mathrm{C~II}}$ values show no strong dependence on metallicity. As already noted, the simulated $\alpha_{\mathrm{C~II}}$ values are, on average, lower than observed values, although with significant overlap between the two distributions. Only a few observations, however, are of galaxies with as low metallicities as the simulations, and where there is overlap in terms of metallicity, the agreement is good between the observed and simulated $\alpha_{\mathrm{C~II}}$ values (although the data points are sufficiently sparse that a quantitative comparison is difficult). Figure 7 suggests that the on average lower $\alpha_{[\mathrm{C~II}]}$-value for the simulations is, at least in part, due to their much lower metallicities. Z18 found $\alpha_{[\mathrm{C~II}]}$ to be invariant with the metallicity. It should be noted, however, that the bulk of their sample falls within a relatively small range of

metallicities, making it hard to tease out any trend. Our simulations suggest that there is in fact a weak dependency on metallicity, which becomes apparent when one probes extremely low metallicities ($12 + \log(\mathrm{O/H}) \lesssim 8.0$), i.e., much lower metallicities than probed by Z18, and as low as those of local metal-poor dwarfs (De Looze et al. 2014).

Naively, one would expect $L_{\mathrm{C~II}}$ to increase with metallicity, due to a higher abundance of carbon. This in turn would imply a decreasing trend in $\alpha_{\mathrm{C~II}}$ with metallicity, which is not what is seen in Figure 7 (top). However, an increase in metallicity also implies a higher dust content; thus, fewer UV photons are available to ionize neutral carbon (photoionization potential of 11.3 eV). It also implies fewer UV photons capable of photodissociating CO (requires photon energies >11.1 eV) in neutral PDRs, which ultimately means less carbon available for ionization. A higher dust content also implies more UV-shielded regions, which would give rise to an increase in $\alpha_{[\mathrm{C~II}]}$ with gas-phase metallicity.

In SÍGAME, the ISRF that impinges on the diffuse and molecular clouds is attenuated by the dust inside the clouds. The fraction of ionized carbon thus depends on the radiation field hitting the clouds, as well as the dust fraction (which scales with the metallicity) of the clouds. CO as a tracer of molecular gas is known to deteriorate in low-metallicity environments, where the photodissociation of CO leads to a significant fraction of the molecular ISM being CO-dark (e.g., Narayanan & Krumholz 2017; Li et al. 2018; Madden et al. 2020).

This effect is illustrated in Figure 7 (bottom), where the CO-to-[C II] luminosity ratio ($L_{\mathrm{CO}}/L_{\mathrm{[C~II]}} \propto \alpha_{[\mathrm{C~II}]}/\alpha_{\mathrm{CO}}$) is seen to sharply decrease with decreasing metallicity. This is a direct result of α_{CO} increasing with decreasing metallicity (e.g., CO-dark gas), while $\alpha_{[\mathrm{C~II}]}$ remains largely invariant as is seen in Figure 7 (top). Across the metallicity range $7.0 \lesssim 12 + \log(\mathrm{O/H}) \lesssim 9.0$, $\alpha_{[\mathrm{C~II}]}/\alpha_{\mathrm{CO}}$ increases by more than two orders of magnitude, which is in line with a similar trend found by Z18 and Accurso et al. (2017) (red and blue lines in Figure 7, bottom panel).

5.5. What Is the Origin of the $\log \mathrm{M}_{mol}$–$\log \mathrm{L}_{[C II]}$ Relation?

In this section we examine whether combining the loglinear relations $\log L_{[\mathrm{C~II}]} - \log M_{\mathrm{mol}}$ and $\log L_{[\mathrm{C~II}]} - \log \mathrm{SFR}$, derived for our simulated galaxies in the previous section, leads to a better determination of M_{mol}. Both $L_{[\mathrm{C~II}]}$ and SFR are observables that can be measured with some confidence toward $z \sim 6$ galaxies. The unobscured SFR can be measured in a number of ways, from the luminosity of the rest-frame continuum at 1500 Å, for example, or from the Lyα emission. Obscured SFRs are typically inferred from (sub)millimeter or FIR continuum measurements. We consider here the total SFRs of our simulated galaxies and do not distinguish between obscured or unobscured SFRs. We will also examine whether including the stellar masses might lead to more accurate gas mass estimates. To this end, we perform a principal component analysis (PCA) on the parameter space spanned by $\log L_{[\mathrm{C~II}]}$, $\log M_{\mathrm{mol}}$, $\log \mathrm{SFR}$, and $\log M_\star$. This gives us the following principal components:

$$\mathrm{PC1} = 0.47 \log L_{[\mathrm{C~II}]} + 0.51 \log \mathrm{SFR} + 0.51 \log M_{\mathrm{mol}} + 0.51 \log M_\star \qquad (3)$$

$$\mathrm{PC2} = -0.87 \log L_{[\mathrm{C~II}]} + 0.40 \log \mathrm{SFR} + 0.26 \log M_{\mathrm{mol}} + 0.14 \log M_\star \qquad (4)$$

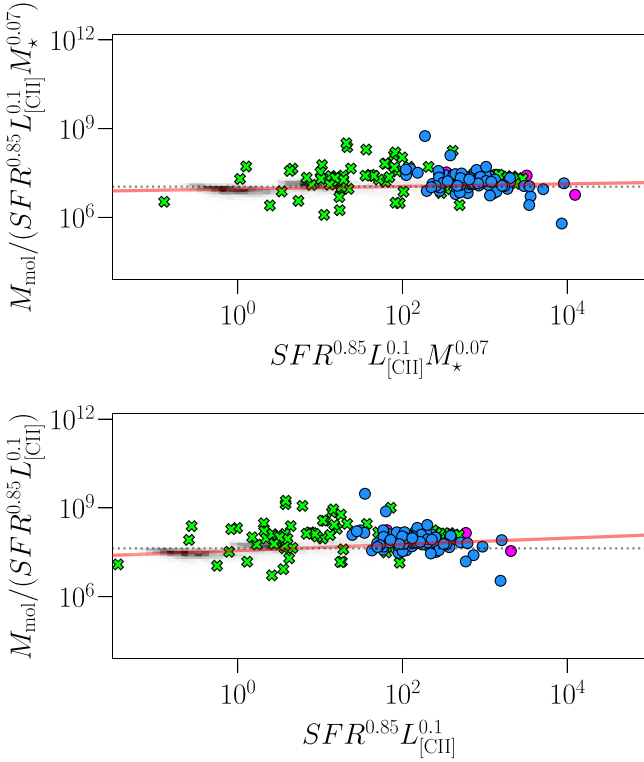**Figure 8.** Equations (7) (top) and (8) (bottom) for our simulated galaxies, ALPINE galaxies, and Z18 galaxies.

$$PC3 = -0.14 \log L_{[C\,II]} - 0.41 \log SFR$$
$$- 0.31 \log M_{mol} + 0.85 \log M_\star \qquad (5)$$

$$PC4 = 0.08 \log L_{[C\,II]} + 0.64 \log SFR$$
$$- 0.76 \log M_{mol} + 0.05 \log M_\star. \qquad (6)$$

Of the four principal components, PC1 is responsible for 91% of the variance, and PC2 and PC3 are responsible for 7% and 2%, respectively. Given its negligible contribution to the variance, we can set PC4 to zero, which in turn allows us to derive the following expression for the molecular gas:

$$\log M_{mol} = 0.10 \log L_{[C\,II]} + 0.85 \log SFR + 0.07 \log M_\star.$$
$$\qquad (7)$$

The dominant principal component in this prescription is the SFR, as expected given that star formation is fueled by molecular gas. The [C II] luminosity and stellar mass also account for some of the correlation with molecular gas mass, each with about 10% of the contribution. We thus interpret the $\log L_{C\,II} - \log M_{mol}$ relation at high z as a second-order result of the KS law, where [C II] luminosity is a coincidental tracer of the molecular gas mass as a result of the [C II]–SFR relation.

In Figure 8 we have plotted Equation (7) for our simulations and comparison data, with (top) and without (bottom) the $M_\star$-term. The residual scatter in the two cases is virtually identical, i.e., 0.18 and 0.19 dex, respectively, and we can therefore drop $M_\star$, which is the most difficult observable from Equation (7). Thus, a useful relation to obtain the molecular gas mass to within a fractional scatter of $0.19 \ln(10) \simeq 44\%$ is

$$\log M_{mol} = 0.10 \log L_{[C\,II]} + 0.85 \log SFR. \qquad (8)$$

The conclusion from our PCA is that the $\log L_{C\,II} - \log M_{mol}$ relation comes about from a combination of the

$\log L_{[C\,II]} - \log$ SFR relationship and the KS relation, which relates SFR to M_{mol}.

Sommovigo et al. (2021) came to a similar conclusion in that they provided a plausible physical explanation for the invariant $\alpha_{[C\,II]}$ by combing the $\Sigma_{[C\,II]} - \Sigma_{SFR}$ relation (De Looze et al. 2014) and the $\Sigma_{SFR} - \Sigma_{mol}$ relation (Kennicutt 1998; Kennicutt & Evans 2012). They found that $\alpha_{[C\,II]} = 30.3\tau_{depl}\Sigma_{SFR}^{-0.075}$, where $\tau_{depl} = M_{mol}/SFR$ is the molecular gas depletion timescale in Gyr and Σ_{SFR} is the SFR surface density in $M_\odot\, yr^{-1}\, kpc^{-2}$. Given the weak dependence on Σ_{SFR} and observational indications that for main-sequence galaxies τ_{depl} is approximately constant ($\sim$0.4–0.7 Gyr) across the redshift range $z \sim 0$–2, they find a nearly constant value of $\alpha_{[C\,II]} = (12$–$21)\Sigma_{SFR}^{-0.075}$, which is in good agreement with the Z18 values ($\simeq 30\, M_\odot/L_\odot$), as well as the $\alpha_{[C\,II]}$ values for our $z = 6$ simulations ($\simeq 18\, M_\odot/L_\odot$).

Following Sommovigo et al. (2021), we derive an integrated version of the above relation. To this end, we adopt $SFR = 10^{-8.52}L_{C\,II}^{1.18}$, which De Looze et al. (2014) found applied to $z > 0.5$ starburst galaxies. For the integrated KS relation we adopt $M_{mol} = 10^{8.00}SFR^{0.83}$, which Sargent et al. (2014) found applied to starbursts. Combining the two, we derive

$$\alpha_{C\,II} = 60.3\, \tau_{depl}\, SFR^{0.16}, \qquad (9)$$

where τ_{depl} is in Gyr. This relation is similar to that of Sommovigo et al. (2021), in terms of both its normalization and the weak SFR dependence, albeit with a positive exponent rather than a negative one. In Figure 9 we plot $\alpha_{C\,II}$ for our simulations derived from the Sommovigo et al. (2021) relation (panel (a)) and Equation (9) (panel (b)) versus their $\alpha_{C\,II} = M_{mol}/L_{C\,II}$ values.

It is seen that neither the $\alpha_{C\,II}$ estimator provided by Sommovigo et al. (2021) nor its volumetric version derived here correlate well with the $\alpha_{C\,II} = M_{mol}/L_{C\,II}$ values of the simulations, and significant scatter around the 1:1 relation is seen along the horizontal direction. However, both estimators agree in the mean with the simulation values. The apparent bimodality seen in Figure 9(b) is due to the fact that our simulations, which come from three simulation boxes (see Section 3), span a wide range in both SFR and τ_{depl} (Figure 1, top and bottom panels). The bottom cloud of points in Figure 9(b) is predominantly from the low-SFR and low-τ_{depl} galaxies extracted from the 25 cMpc h^{-1} box, while the top cloud consists of galaxies from the 50 and 100 cMpc h^{-1} box. This bimodality is not seen in Figure 9(a) since (i) the dependency on Σ_{SFR} in the relation by Sommovigo et al. (2021) is significantly weaker and (ii) the simulations span a relatively narrow range in Σ_{SFR}.

Extreme starburst galaxies have molecular gas depletion timescales of $\sim$100 Myr (e.g., Aravena et al. 2016), i.e., of the order 4–7$\times$ shorter than main-sequence galaxies, which would imply correspondingly lower $\alpha_{C\,II}$-values for starbursts (see Equation (9)). The $\sim$5–10$\times$ higher SFRs of starburst galaxies are not sufficient to fully offset the effect of the shorter gas depletion timescales on $\alpha_{C\,II}$. Assuming that the gas depletion timescales and SFRs for the main-sequence galaxies (MS) and starburst galaxies (SB) roughly scale as $\tau_{SB} \sim (0.4$–$0.7) \times \tau_{MS}$ and $SFR_{SB} \sim (5$–$10) \times SFR_{MS}$, we find from Equation (9) that $\alpha_{[C\,II],SB} \sim (0.2$–$0.4) \times \alpha_{[C\,II],MS}$. Inserting our median value of $\alpha_{C\,II} = 18\, M_\odot/L_\odot$ from our simulations then yields an expected conversion factor for starbursts of $\alpha_{[C\,II],SB} \sim 4$–$7\, M_\odot/L_\odot$. This
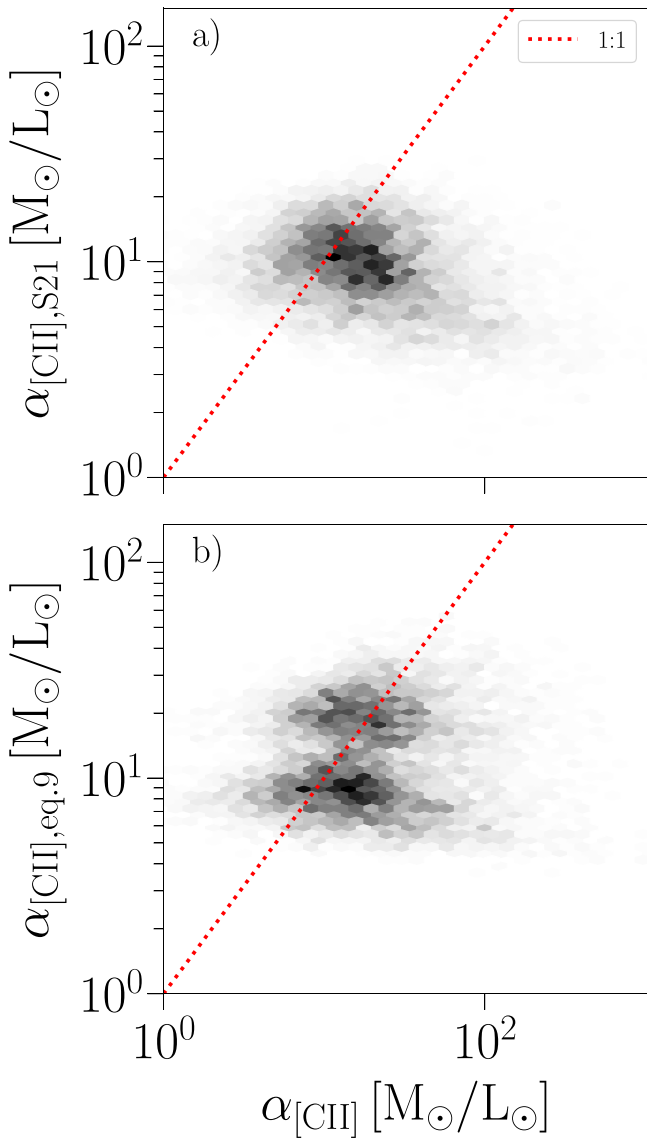
Figure 9. $\alpha_{\rm [C\,II]}$-values calculated for our simulations using the relation from Sommovigo et al. (2021) (panel (a)) and Equation (9) (panel (b)) vs. the true $\alpha_{\rm [C\,II]} = M_{\rm mol}/L_{\rm [C\,II]}$ values for our simulations. The red dotted lines denotes the 1:1 relationship.

is consistent with the low values found in high-z dusty starburst galaxies by Rizzo et al. (2021) (see Section 5.3).

6. Conclusion

Using SÍGAME post-processing of galaxies from SIMBA cosmological simulations, we have simulated and analyzed the [C II] emission from 11,125 $z = 6$ galaxies with the goal of examining the viability of [C II] as a tracer of the ISM gas mass, and in particular the molecular gas mass, in normal star-forming galaxies at this epoch. The SIMBA simulated galaxies span the stellar mass range from $1.5 \times 10^7\, M_\odot$ to $5 \times 10^{10}\, M_\odot$ and define a star-forming main sequence at $z = 6$ that at the high-mass end agrees with observations and at the low-mass end agrees with extrapolations of the observed main sequence (Schreiber et al. 2015). The simulated galaxies also follow the empirical KS law and the mass–metallicity relation. Our main findings are presented below:

1. For the most massive ($M_\star \gtrsim 10^9\, M_\odot$) galaxies in our simulation, the [C II] emission is almost entirely coming from the molecular gas phase. For galaxies with lower stellar masses ($M_{\rm star} = 10^7$–$10^9\, M_\odot$) there is a larger spread in the fraction ($\sim$50%–100%) of the total [C II] emission coming from the molecular gas, but the general trend is that the average fraction decreases toward lower masses. The atomic gas phase contributes a negligible amount ($\lesssim$5%) to the total [C II] emission. As a result, the contribution from the ionized ISM to the [C II] emission is the mirror opposite to that of the molecular gas.

2. We find and parameterize a loglinear correlation between the [C II] emission and molecular gas masses of our galaxies: $\log L_{\rm C\,II} = (0.78 \pm 0.01)\log M_{\rm mol} + (0.44 \pm 0.04)$. The slope is shallower than the near-unity slope derived by Z18, and the scatter in the correlation is 0.45 dex, which is larger than the 0.2 dex scatter in the Z18 relation. The shallower slope is due to the simulated galaxies at the low-mass end ($\lesssim 10^8\, M_\odot$) generally having higher [C II] luminosities compared to the extrapolation of the linear Z18 relation to these low masses. However, the observations that are available at the low-mass end (with the exception of Madden et al. 2020) are consistent with the simulation predictions, albeit with a larger scatter. Thus, the difference in slope might be attributable to a large number of dwarf galaxies in our sample that lack comparable observational data at high redshifts. At the high-mass end, our simulations are in good agreement with the Z18 relation and observational data from Z18 and ALPINE.

3. We derive a [C II]-to-$M_{\rm mol}$ conversion factor, $\alpha_{\rm C\,II}$, based on our simulations, finding a median value of $\alpha_{\rm C\,II} = 18\, M_\odot/L_\odot$ with an MAD of $10\, M_\odot/L_\odot$. This is lower than the average conversion factor derived by Z18, although there is a significant overlap in the distributions between our simulated and the observed $\alpha_{\rm C\,II}$-values. We attribute the lower average $\alpha_{\rm C\,II}$-values in our simulations to their much lower gas-phase metallicities, which allow for a higher CO photodissociation rate and for neutral carbon atoms in the ISM to be more readily ionized.

4. Using PCA, we determine a relation between molecular gas mass, SFR, and [C II] emission, $\log M_{\rm mol} = 0.10 \log L_{\rm [C\,II]} + 0.85 \log {\rm SFR}$, which has a scatter of only 0.18 dex. Our analysis suggests that the tight relationship between the SFR and the molecular gas mass is ultimately responsible for [C II] as a tracer of molecular gas mass.

ORCID iDs

T. R. Greve ● https://orcid.org/0000-0002-2554-1837
K. P. Olsen ● https://orcid.org/0000-0003-1250-5287
A. Zanella ● https://orcid.org/0000-0001-8600-7008
D. Narayanan ● https://orcid.org/0000-0002-7064-4309
R. Davè ● https://orcid.org/0000-0003-2842-9434
G. E. Magdis ● https://orcid.org/0000-0002-4872-2294
G. Popping ● https://orcid.org/0000-0003-1151-4659
F. Valentino ● https://orcid.org/0000-0001-6477-4011
K. E. Heintz ● https://orcid.org/0000-0002-9389-7413

References

Accurso, G., Saintonge, A., Bisbas, T. G., & Viti, S. 2017, MNRAS, 464, 3315
Appleby, S., Davé, R., Kraljic, K., Anglés-Alcázar, D., & Narayanan, D. 2020, MNRAS, 494, 6053
Aravena, M., Spilker, J. S., Bethermin, M., et al. 2016, MNRAS, 457, 4406
Arimoto, N., Sofue, Y., & Tsujimoto, T. 1996, PASJ, 48, 275
Bell, E. F. 2003, ApJ, 586, 794
Bisbas, T. G., Papadopoulos, P. P., & Viti, S. 2015, ApJ, 803, 37
Bisbas, T. G., van Dishoeck, E. F., Papadopoulos, P. P., et al. 2017, ApJ, 839, 90
Bolatto, A. D., Wolfire, M., & Leroy, A. K. 2013, ARA&A, 51, 207
Bothwell, M. S., Maiolino, R., Cicone, C., Peng, Y., & Wagg, J. 2016, A&A, 595, A48
Carilli, C. L., & Walter, F. 2013, ARA&A, 51, 105
Carniani, S., Maiolino, R., Amorin, R., et al. 2018, MNRAS, 478, 1170
Cicone, C., Bothwell, M., Wagg, J., et al. 2017, A&A, 604, A53
Cox, D. P. 2005, ARA&A, 43, 337
da Cunha, E., Groves, B., Walter, F., et al. 2013, ApJ, 766, 13
Davé, R., Anglés-Alcázar, D., Narayanan, D., et al. 2019, MNRAS, 486, 2827
Davé, R., Crain, R. A., Stevens, A. R. H., et al. 2020, MNRAS, 497, 146
Davé, R., Thompson, R., & Hopkins, P. F. 2016, MNRAS, 462, 3265
De Looze, I., Cormier, D., Lebouteiller, V., et al. 2014, A&A, 568, A62
Dessauges-Zavadsky, M., Ginolfi, M., Pozzi, F., et al. 2020, A&A, 643, A5
Díaz-Santos, T., Armus, L., Charmandaris, V., et al. 2013, ApJ, 774, 68
Dickman, R. L., Snell, R. L., & Schloerb, F. P. 1986, ApJ, 309, 326
Downes, D., & Solomon, P. M. 1998, ApJ, 507, 615
Draine, B. T. 2011, Physics of the Interstellar and Intergalactic Medium (Princeton, NJ: Princeton Univ. Press)
Eales, S., Smith, M. W. L., Auld, R., et al. 2012, ApJ, 761, 168
Faisst, A. L., Schaerer, D., Lemaux, B. C., et al. 2020, ApJS, 247, 61
Ferland, G. J., Chatzikos, M., Guzmán, F., et al. 2017, RMxAA, 53, 385
Ferland, G. J., Porter, R. L., van Hoof, P. A. M., et al. 2013, RMxAA, 49, 137
Fudamoto, Y., Oesch, P. A., Faisst, A., et al. 2020, A&A, 643, A4
Genzel, R., Tacconi, L. J., Combes, F., et al. 2012, ApJ, 746, 69
Gerin, M., & Phillips, T. G. 2000, ApJ, 537, 644
Gilda, S., Lower, S., & Narayanan, D. 2021, ApJ, 916, 43
Glover, S. C. O., Clark, P. C., Micic, M., & Molina, F. 2015, MNRAS, 448, 1607
Guelin, M., Zylka, R., Mezger, P. G., Haslam, C. G. T., & Kreysa, E. 1995, A&A, 298, L29
Guelin, M., Zylka, R., Mezger, P. G., et al. 1993, A&A, 279, L37

Harikane, Y., Ouchi, M., Inoue, A. K., et al. 2020, ApJ, 896, 93
Heintz, K. E., & Watson, D. 2020, ApJL, 889, L7
Heintz, K. E., Watson, D., Oesch, P., Narayanan, D., & Madden, S. C. 2021, ApJ, 922, 147
Herrera-Camus, R., Bolatto, A. D., Wolfire, M. G., et al. 2015, ApJ, 800, 1
Hildebrand, R. H. 1983, QJRAS, 24, 267
Hopkins, P. F. 2015, MNRAS, 450, 53
Hopkins, P. F. 2017, arXiv:1712.01294
Hughes, T. M., Ibar, E., Villanueva, V., et al. 2017, A&A, 602, A49
Ikeda, M., Oka, T., Tatematsu, K., Sekimoto, Y., & Yamamoto, S. 2002, ApJS, 139, 467
Ivison, R. J., Papadopoulos, P. P., Smail, I., et al. 2011, MNRAS, 412, 1913
Kaasinen, M., Scoville, N., Walter, F., et al. 2019, ApJ, 880, 15
Kaufman, M. J., Wolfire, M. G., Hollenbach, D. J., & Luhman, M. L. 1999, ApJ, 527, 795
Kennicutt, R. C. 1998, ARA&A, 36, 189
Kennicutt, R. C., & Evans, N. J. 2012, ARA&A, 50, 531
Khusanova, Y., Bethermin, M., Le Fèvre, O., et al. 2021, A&A, 649, A152
Khusanova, Y., Le Fèvre, O., Cassata, P., et al. 2020, A&A, 634, A97
Krumholz, M. R. 2012, ApJ, 759, 9
Krumholz, M. R., & Gnedin, N. Y. 2011, ApJ, 729, 36
Lagache, G., Cousin, M., & Chatzikos, M. 2018, A&A, 609, A130
Le Fèvre, O., Béthermin, M., Faisst, A., et al. 2020, A&A, 643, A1
Leung, T. K. D., Olsen, K. P., Somerville, R. S., et al. 2020, ApJ, 905, 102
Li, Q., Narayanan, D., & Davé, R. 2019, MNRAS, 490, 1425
Li, Q., Narayanan, D., Davè, R., & Krumholz, M. R. 2018, ApJ, 869, 73
Liang, L., Feldmann, R., Faucher-Giguère, C.-A., et al. 2018, MNRAS, 478, L83
Lower, S., Narayanan, D., Leja, J., et al. 2020, ApJ, 904, 33
Ma, X., Hopkins, P. F., Faucher-Giguère, C.-A., et al. 2016, MNRAS, 456, 2140
Madden, S. C., Cormier, D., Hony, S., et al. 2020, A&A, 643, A141
Magdis, G. E., Daddi, E., Béthermin, M., et al. 2012, ApJ, 760, 6
Magdis, G. E., Rigopoulou, D., Hopwood, R., et al. 2014, ApJ, 796, 63
Malhotra, S., Helou, G., Stacey, G., et al. 1997, ApJL, 491, L27
Narayanan, D., & Krumholz, M. R. 2017, MNRAS, 467, 50
Narayanan, D., Krumholz, M. R., Ostriker, E. C., & Hernquist, L. 2012, MNRAS, 421, 3127
Offner, S. S. R., Bisbas, T. G., Bell, T. A., & Viti, S. 2014, MNRAS, 440, L81
Olsen, K., Greve, T. R., Narayanan, D., et al. 2017, ApJ, 846, 105
Olsen, K., Greve, T. R., Narayanan, D., et al. 2018, ApJ, 857, 148
Pallottini, A., Ferrara, A., Gallerani, S., et al. 2017, MNRAS, 465, 2540
Papadopoulos, P. P., & Greve, T. R. 2004, ApJL, 615, L29
Papadopoulos, P. P., van der Werf, P., Xilouris, E., Isaak, K. G., & Gao, Y. 2012, ApJ, 751, 10
Pearson, W. J., Wang, L., Hurley, P. D., et al. 2018, A&A, 615, A146
Pineda, J. L., Langer, W. D., & Goldsmith, P. F. 2014, A&A, 570, A121
Popping, G., Narayanan, D., Somerville, R. S., Faisst, A. L., & Krumholz, M. R. 2019, MNRAS, 482, 4906
Privon, G. C., Narayanan, D., & Davé, R. 2018, ApJ, 867, 102
Rémy-Ruyer, A., Madden, S. C., Galliano, F., et al. 2014, A&A, 563, A31
Rizzo, F., Vegetti, S., Fraternali, F., Stacey, H. R., & Powell, D. 2021, MNRAS, 507, 3952
Rizzo, F., Vegetti, S., Powell, D., et al. 2020, Natur, 584, 201
Salmon, B., Papovich, C., Finkelstein, S. L., et al. 2015, ApJ, 799, 183
Sargent, M. T., Daddi, E., Béthermin, M., et al. 2014, ApJ, 793, 19
Schaerer, D., Ginolfi, M., Béthermin, M., et al. 2020, A&A, 643, A3
Schreiber, C., Pannella, M., Elbaz, D., et al. 2015, A&A, 575, A74
Scoville, N., Aussel, H., Sheth, K., et al. 2014, ApJ, 783, 84
Scoville, N., Sheth, K., Aussel, H., et al. 2016, ApJ, 820, 83
Solomon, P. M., & Barrett, J. W. 1991, in Dynamics of Galaxies and their Molecular Cloud Distributions, ed. F. Combes & F. Casoli, Vol. 146 (Cambridge: Cambridge Univ. Press), 235
Sommovigo, L., Ferrara, A., Carniani, S., et al. 2021, MNRAS, 503, 4878
Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. 2014, ApJS, 214, 15
Stacey, G. J., Geis, N., Genzel, R., et al. 1991, ApJ, 373, 423
Tacconi, L. J., Genzel, R., Saintonge, A., et al. 2018, ApJ, 853, 179
Tacconi, L. J., Genzel, R., Smail, I., et al. 2008, ApJ, 680, 246
Tacconi, L. J., Neri, R., Genzel, R., et al. 2013, ApJ, 768, 74
Thomas, N., Davé, R., Anglés-Alcázar, D., & Jarvis, M. 2019, MNRAS, 487, 5764
Tunnard, R., & Greve, T. R. 2016, ApJ, 819, 161
Turk, M. J., Smith, B. D., Oishi, J. S., et al. 2011, ApJS, 192, 9
Valentino, F., Daddi, E., Puglisi, A., et al. 2020a, A&A, 641, A155

Valentino, F., Magdis, G. E., Daddi, E., et al. 2018, ApJ, 869, 27
Valentino, F., Magdis, G. E., Daddi, E., et al. 2020b, ApJ, 890, 24
Vallini, L., Gallerani, S., Ferrara, A., Pallottini, A., & Yue, B. 2015, ApJ, 813, 36
Walter, F., Carilli, C., Neeleman, M., et al. 2020, ApJ, 902, 111

Wilson, C. D. 1995, ApJL, 448, L97
Wolfire, M. G., Hollenbach, D., & McKee, C. F. 2010, ApJ, 716, 1191
Wu, X., Davé, R., Tacchella, S., & Lotz, J. 2020, MNRAS, 494, 5636
Zanella, A., Daddi, E., Magdis, G., et al. 2018, MNRAS, 481, 1976
Zhang, Z.-Y., Papadopoulos, P. P., Ivison, R. J., et al. 2016, RSOS, 3, 160025