

MIT Open Access Articles

Modeling the Atomic-to-molecular Transition in Cosmological Simulations of Galaxy Formation

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

As Published: 10.3847/1538-4365/AAE387

Publisher: American Astronomical Society

Persistent URL: <https://hdl.handle.net/1721.1/136377>




Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Terms of Use: Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.





Modeling the Atomic-to-molecular Transition in Cosmological Simulations of Galaxy Formation

Benedikt Diemer¹ , Adam R. H. Stevens², John C. Forbes¹, Federico Marinacci¹, Lars Hernquist¹ , Claudia del P. Lagos^{2,3}, Amiel Sternberg^{4,5}, Annalisa Pillepich⁶ , Dylan Nelson⁷, Gergö Popping⁶, Francisco Villaescusa-Navarro⁵, Paul Torrey⁸, and Mark Vogelsberger⁹

¹ Institute for Theory and Computation, Harvard-Smithsonian Center for Astrophysics, 60 Garden St., Cambridge, MA 02138, USA
benedikt.diemer@cfa.harvard.edu

² International Centre for Radio Astronomy Research, University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia

³ ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia

⁴ Raymond and Beverly Sackler School of Physics and Astronomy, Tel Aviv University, Ramat Aviv 69978, Israel

⁵ Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, 10010, New York, NY, USA

⁶ Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany

⁷ Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str 1, D-85741 Garching, Germany

⁸ Hubble Fellow, MIT Kavli Institute for Astrophysics and Space Research, Cambridge, MA 02139, USA

⁹ Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Received 2018 June 6; revised 2018 September 17; accepted 2018 September 20; published 2018 October 23

Abstract

Large-scale cosmological simulations of galaxy formation currently do not resolve the densities at which molecular hydrogen forms, implying that the atomic-to-molecular transition must be modeled either on the fly or in postprocessing. We present an improved postprocessing framework to estimate the abundance of atomic and molecular hydrogen and apply it to the IllustrisTNG simulations. We compare five different models for the atomic-to-molecular transition, including empirical, simulation-based, and theoretical prescriptions. Most of these models rely on the surface density of neutral hydrogen and the ultraviolet (UV) flux in the Lyman–Werner band as input parameters. Computing these quantities on the kiloparsec scale resolved by the simulations emerges as the main challenge. We show that the commonly used Jeans length approximation to the column density of a system can be biased and exhibits large cell-to-cell scatter. Instead, we propose to compute all surface quantities in face-on projections and perform the modeling in two dimensions. In general, the two methods agree on average, but their predictions diverge for individual galaxies and for models based on the observed midplane pressure of galaxies. We model the UV radiation from young stars by assuming a constant escape fraction and optically thin propagation throughout the galaxy. With these improvements, we find that the five models for the atomic-to-molecular transition roughly agree on average but that the details of the modeling matter for individual galaxies and the spatial distribution of molecular hydrogen. We emphasize that the estimated molecular fractions are approximate due to the significant systematic uncertainties.

Key words: galaxies: ISM – ISM: molecules – methods: numerical

1. Introduction

Cosmological simulations that follow volumes of up to a few hundred megaparsecs on a side are integral to the study of galaxy formation (Schaye et al. 2010, 2015; Vogelsberger et al. 2013, 2014a, 2014b; Dubois et al. 2014; Somerville & Davé 2015; Davé et al. 2016; Pillepich et al. 2018a). These calculations achieve statistically representative samples of galaxies, but at the cost of relatively low resolution, typically about a million solar masses and roughly kiloparsec spatial resolution. Thus, many important physical processes cannot be followed explicitly. For example, while some higher-resolution, zoom-in simulations have been performed with chemical networks and radiative transfer (Pelupessy et al. 2006; Robertson & Kravtsov 2008; Gnedin et al. 2009; Christensen et al. 2012; Kuhlen et al. 2012; Rosdahl et al. 2015; Hu et al. 2016; Katz et al. 2017; Nickerson et al. 2018), such calculations would not necessarily be applicable at lower resolution and would also be too costly to perform in cosmological volumes and over a *Hubble* time. Thus, a number of physical quantities are not directly predicted by cosmological simulations.

One particularly desirable quantity is the phase structure of galactic gas: while most simulations now approximate the ionization balance of hydrogen due to ultraviolet (UV) radiation and self-shielding in low-density regions, they do not compute the abundances of atomic and molecular hydrogen (hereafter abbreviated H I and H₂, respectively) at high densities (see, however, Thompson et al. 2014; Davé et al. 2016). The H I and H₂ masses and distributions are critical for comparisons to observations because cold gas is often observed in a specific phase. For example, H I can be detected via its 21 cm emission or absorption in quasar spectra, while H₂ can be inferred via spectroscopic observations of CO and other molecular tracers or via the dust continuum (see Bolatto et al. 2013 for a review).

To facilitate meaningful comparisons with observations of the gas content of galaxies, we must postprocess cosmological simulations. This step is nontrivial because the transition from atomic to molecular hydrogen is physically complicated: H₂ molecules form predominantly on the surface of dust grains (which are not modeled in the simulations) and are mostly destroyed by UV radiation in the Lyman–Werner (LW) band, which is shielded by dust, H I, and H₂ (e.g., Spitzer &

Zweibel 1974; Jura 1975a, 1975b; Black & Dalgarno 1976; Shull 1978; van Dishoeck & Black 1986; Sternberg & Dalgarno 1989; Elmegreen 1993; Draine 2011). Thus, we expect molecules to form most efficiently in dense regions, at high metallicity, and to be surrounded by a layer of shielding atomic gas (Sternberg 1988; Elmegreen 1989; Browning et al. 2003; Krumholz et al. 2009a; Krumholz 2013; Sternberg et al. 2014; Bialy et al. 2017). However, the geometry of molecular clouds can be highly irregular, and the properties of dust are poorly understood (in particular, its distribution may not follow the gas; Gnedin et al. 2008; Bekki 2015; Hopkins & Lee 2016; McKinnon et al. 2016, 2017, 2018). Moreover, the efficiency of dissociation is not even across the LW band but instead is due to a series of lines, introducing complications such as line overlap (Draine & Bertoldi 1996; Gnedin & Draine 2014). In cosmological simulations, there is no way to model any of these effects in detail because the gas cells are large enough to contain a significant number of unresolved molecular clouds and sources of UV radiation.

Instead, we rely on prescriptions for the dependence of the molecular fraction, f_{mol} , on several averaged quantities, such as density, surface density, neutral gas fraction, the UV field, and metallicity. Providing physically motivated estimates of these quantities emerges as the main challenge, but once they have been computed, there are, broadly speaking, three commonly used classes of models that can be used to estimate the molecular fraction. First, it is observationally well established that, in the local universe, f_{mol} is correlated with the midplane pressure of galaxies (Blitz & Rosolowsky 2004, 2006; Leroy et al. 2008; Robertson & Kravtsov 2008). We can estimate the midplane pressure from surface densities and velocity dispersions, with no dependence on the UV field. Second, high-resolution simulations including advanced chemical networks have been used to calibrate f_{mol} as a function of surface density, metallicity, and the UV field (Gnedin & Kravtsov 2011; Gnedin & Draine 2014). Third, analytical equilibrium models of molecular clouds provide the most accurate modeling of H_2 creation and destruction, but such models need to be modified when applying them in the context of cosmological simulations, where they represent a collection of molecular clouds rather than an individual cloud (Sternberg 1988; Sternberg et al. 2014; Krumholz et al. 2009a; McKee & Krumholz 2010; Krumholz 2013; Bialy et al. 2017). We use at least one representative from each of these classes of models to estimate the molecular fraction.

Similar postprocessing exercises were first applied to semianalytical models of galaxy formation and 1D simulations (Obreschkow et al. 2009; Fu et al. 2010; Lagos et al. 2011a, 2011b; Forbes et al. 2012, 2014; Krumholz & Dekel 2012; Popping et al. 2014, 2015) and were later extended to the OWLS and EAGLE simulations (Schaye et al. 2010, 2015). Many aspects of our work are based on the effort of Lagos et al. (2015), who computed f_{mol} for EAGLE galaxies on a particle-by-particle basis. A number of follow-up studies used similar methodology and compared EAGLE to observations of the neutral and molecular gas fraction, the respective mass functions, the structure of HI disks, environmental dependencies, and the mass–size relation, among others (Rahmati et al. 2015; Bahé et al. 2016; Lagos et al. 2016; Marasco et al. 2016; Crain et al. 2017, see also Duffy et al. 2012). Marinacci et al. (2017a) applied the method of Lagos et al. (2015) to the AURIGA high-resolution zoom-in simulations

(Grand et al. 2017), and Villaescusa-Navarro et al. (2018) investigated the overall abundance and clustering of HI in the IllustrisTNG simulations. Notably, some studies found good agreement between the HI/ H_2 models tested (e.g., Marinacci et al. 2017a), while others reported significant differences between the f_{mol} predictions, especially in detailed, spatially resolved analyses (e.g., Bahé et al. 2016). Such differences are surprising because the HI/ H_2 models are supposed to agree reasonably well when they are based on the same underlying UV field, metallicity, and density (e.g., Krumholz & Gnedin 2011; Sternberg et al. 2014).

In this work, we attempt to understand and to remedy such disagreements, and to provide a comprehensive investigation of the modeling methods in general. Based on the IllustrisTNG simulations, we find that the conversion between volume and surface density can introduce systematic errors and propose a new method based on two-dimensional projections of all relevant quantities. We introduce a new way to crudely estimate the LW-band UV flux by following the propagation of radiation from star-forming populations in the optically thin limit. We investigate five different models for the HI/ H_2 transition and show that they are in good agreement on average, though their predictions diverge for individual galaxies and the spatial distribution of molecules. We will not attempt to evaluate which models are most accurate but accept their range of predictions as a systematic uncertainty.

In this paper, we focus on our methodology and reserve the quantitative comparison between IllustrisTNG and the observed gas content of galaxies for future work (B. Diemer et al. 2018, in preparation; A. Stevens et al. 2018, in preparation; G. Popping et al. 2018, in preparation). The paper is organized as follows. In Section 2, we lay out our methodology, though some of the details are to be found in Appendix A. We compare the predictions of the different HI/ H_2 models in Section 3. We discuss the most important uncertainties in Section 4 and summarize our conclusions in Section 5. We give detailed mathematical expressions and algorithmic information about the HI/ H_2 models in Appendix C.

2. Methods

In this section, we describe how we compute the molecular fraction in gas cells and galaxies in our simulations. We briefly introduce the IllustrisTNG simulation suite, focusing on those parts of the setup that are particularly relevant for our study. We discuss the two fundamental types of modeling used (volumetric and projected) and describe how we compute the neutral fraction, the UV field, and eventually the HI/ H_2 transition.

2.1. Notation

As the notation of gas densities can easily get confusing due to the large number of subspecies, we use a consistent set of subscripts. In particular, we denote the surface densities of all gas as Σ_{gas} , of all hydrogen as Σ_{H} , of neutral hydrogen as $\Sigma_{\text{HI}+\text{H}_2}$, of atomic hydrogen as Σ_{HI} , and of molecular hydrogen as Σ_{H_2} . The same subscripts are used for volumetric mass densities (ρ_{gas} , ρ_{H} , $\rho_{\text{HI}+\text{H}_2}$, ρ_{HI} , and ρ_{H_2}), as well as for number densities (e.g., n_{H}), column densities (e.g., N_{H}), and galaxy-integrated masses (e.g., M_{H}). The molecular fraction can be

Table 1
Simulations and Galaxy Sample Selection Criteria

Simulation	m_{baryon}	$M_{\text{gas,min}}$	$M_{*,\text{min}}$	N_{gas}	N_{gal}
TNG100	1.4×10^6	2.0×10^8	2.0×10^8	142	152,122
TNG300	0.9×10^7	5.0×10^9	5.0×10^{10}	454	452,664
TNG100-2	1.1×10^7	1.1×10^9	1.1×10^9	100	78,308
TNG100-3	8.9×10^7	9.0×10^9	9.0×10^9	100	20,635

Note. All masses are given in M_{\odot} . To be included in our sample, a galaxy needs to have either $M_{\text{gas,min}}$ gas mass or $M_{*,\text{min}}$ stellar mass, counting all particles bound to the galaxy. The baryonic mass units in IllustrisTNG (i.e., the stellar population particles and gas cells) can slightly evolve in mass; the mass quoted as m_{baryon} is the target mass (Pillepich et al. 2018b). On average, the minimum gas and stellar masses correspond to the listed particle numbers N_{gas} and result in sample sizes of N_{gal} galaxies.

expressed either as the ratio of atomic to molecular gas,

$$R_{\text{mol}} \equiv \frac{\Sigma_{\text{H}_2}}{\Sigma_{\text{H I}}}, \quad (1)$$

or as the fraction of neutral hydrogen that is molecular:

$$f_{\text{mol}} \equiv \frac{\Sigma_{\text{H}_2}}{\Sigma_{\text{H I}} + \Sigma_{\text{H}_2}} = \frac{R_{\text{mol}}}{R_{\text{mol}} + 1}. \quad (2)$$

When quoting the fraction of neutral hydrogen, $f_{\text{H I}+\text{H}_2}$, we note that it refers to the fraction of all gas (including helium and metals) that is in neutral hydrogen, not the fraction of hydrogen that is neutral. Furthermore, we denote the sound speed in gas as c_s , the ratio of heat capacities as $\gamma = 5/3$, the internal energy per unit mass as u , and the metallicity as Z .

2.2. The IllustrisTNG Simulations

IllustrisTNG is a suite of cosmological hydrodynamical simulations run using the moving-mesh code AREPO (Springel 2010). We predominantly use the two highest-resolution runs for each box size, hereafter referred to as TNG100 and TNG300, which model box sizes of about 100 and 300 Mpc, respectively (Marinacci et al. 2017b; Naiman et al. 2018; Nelson et al. 2018; Pillepich et al. 2018b; Springel et al. 2018). The simulations assume the Planck Collaboration et al. (2016) cosmology: $\Omega_m = 0.3089$, $\Omega_b = 0.0486$, $h = 0.6774$, and $\sigma_8 = 0.8159$. Physical models in IllustrisTNG include prescriptions for gas cooling, star formation, metal enrichment, black hole growth, stellar winds, supernovae, and active galactic nuclei (AGNs); we refer the reader to the respective papers for details (Weinberger et al. 2017; Pillepich et al. 2018a). The TNG galaxy formation model is built on the original Illustris model (Vogelsberger et al. 2013, 2014a, 2014b; Genel et al. 2014; Torrey et al. 2014; Sijacki et al. 2015).

Galaxies and halos are identified using the SUBFIND algorithm (Davis et al. 1985; Springel et al. 2001; Dolag et al. 2009). In this work, virtually all quantities are computed directly from the particle distributions, but we do use the total gravitationally bound mass in both gas and stars to select galaxies. Table 1 shows a list of the particular simulations used in this paper, as well as the lower mass limits in gas and stellar mass used to select galaxies. We emphasize that we include galaxies in our sample if they pass either the gas *or* stellar mass threshold to ensure a well-defined sample completeness regardless of which selection is applied.

2.2.1. Interstellar Medium Model and Equation of State

The results of our H I/H₂ modeling will critically depend on the hydrodynamics and star formation in dense gas in IllustrisTNG galaxies, both of which are regulated by the two-phase interstellar medium (ISM) model of Springel & Hernquist (2003, hereafter SH03). This model assumes that there is a threshold density above which star formation sets in. Below the threshold, the gas physics is determined by hydrodynamics, assuming an ideal-gas equation of state. Above the star formation threshold, the model presupposes that the ISM consists of cold, star-forming clouds with a temperature of 1000 K and hot, ionized gas. The model computes an equilibrium star formation rate (SFR) that is proportional to the cold cloud density over a free-fall time, leading to a scaling of $\text{SFR} \propto \rho^{1.5}$, which roughly matches the observed Kennicutt–Schmidt (KS) relation (Schmidt 1959; Kennicutt 1998). The SFR and the resulting heating of the ISM via supernovae balance the cooling rate. By averaging over the cold and hot gas phases, the model prescribes an effective equation of state, that is, an effective pressure and temperature as a function of density. These quantities are understood to be averages over many cold clouds and the hot ISM, and thus do not necessarily correspond to values that would be physically realized. The predictions of the SH03 model are shown as solid lines in the top row of Figure 1.

The SH03 ISM model used in IllustrisTNG is a modification of the original version that accounts for a different stellar initial mass function and provides a less-steep equation of state. In particular, the equation of state is softened such that the internal energy is $u_{\text{eff}} = 0.3u_{\text{eff,SH03}} + 0.7u_4$, where u_4 corresponds to a temperature of 10,000 K. Moreover, the parameters were changed to a star formation timescale of 3.28 Gyr, a supernova temperature of 5.73×10^7 K, and a cloud evaporation factor of 573. The mass fraction in massive stars changed to $\beta = 0.226$ due to the adoption of the Chabrier (2003) stellar initial mass function instead of the Salpeter (1955) function used in SH03. With these parameters, the threshold density for star formation is $n_{\text{H}} = 0.106 \text{ cm}^{-3}$. We refer the reader to SH03 for further details.

The histograms in Figure 1 show the distribution of all gas cells in the analyzed galaxies in TNG100. As expected, the SH03 model determines the gas properties above the star formation threshold (with a few exceptions where cells are prevented from forming stars because they are too hot). Below the star formation threshold, the cells exhibit a wider distribution of temperatures and pressures. Here, the blue lines refer to an ideal gas with a constant temperature of 10,000 K, shown to guide our understanding of how the atomic and molecular fractions evolve as a function of density.

In the following, we will often treat star-forming and quiescent gas cells differently, either because of the effects of star formation or because the pressure and temperature calculations above the threshold do not represent physical values. For this purpose, we check whether the SFR in a cell is greater than zero and treat the cell accordingly as specified in the following sections.

2.2.2. Hydrogen Fraction and Neutral Fraction

The primordial hydrogen fraction is assumed to be 0.76 in IllustrisTNG, but this fraction decreases as the gas becomes enriched with helium and metals. We use the exact hydrogen fraction in each cell for all computations in this paper. For cells within galaxies, the difference is typically a few percent.

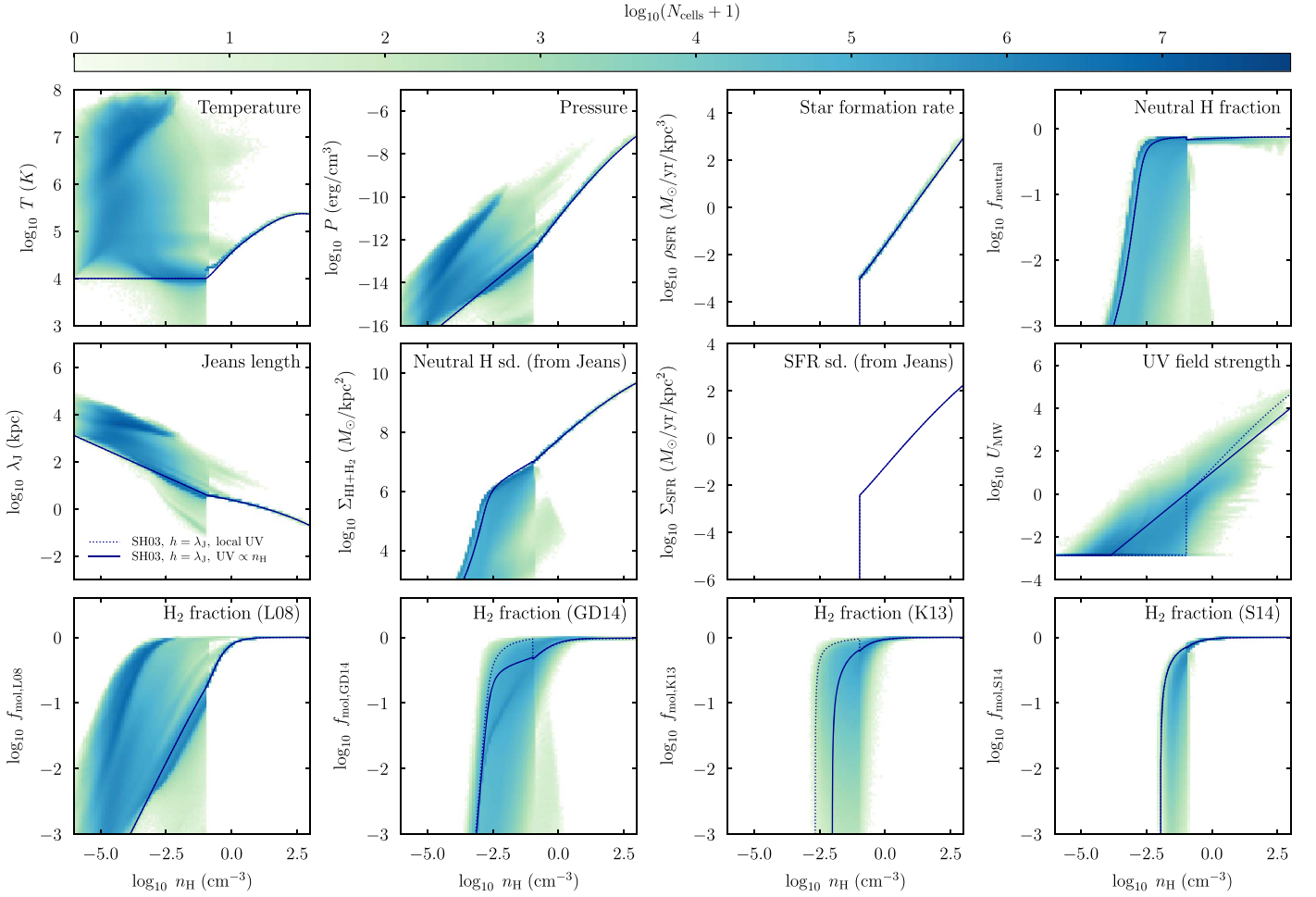


Figure 1. Physical conditions and the resulting molecular fractions as a function of density. In each panel, the histograms show the distribution of all gas cells in the analyzed galaxies in TNG100, about 1.2 billion cells. The solid blue lines represent the ISM model of **SH03** (Section 2.2.1). Top row: At densities below the star formation threshold of $n_{\text{H}} = 0.106 \text{ cm}^{-3}$, hydrodynamics governs the state of the gas, leading to a range of temperatures and pressures. To guide the eye, the blue lines correspond to an ideal-gas equation of state with a temperature of 10,000 K. Above the threshold, the **SH03** effective equation of state determines the temperature, pressure, and SFR. The top right panel shows the fraction of gas in neutral hydrogen, which is computed from the model of Rahmati et al. (2013a) below the star formation threshold (assuming $Z = Z_{\odot}$) and as the cold cloud fraction in the **SH03** model above the threshold (Section 2.2.2), introducing a slight discontinuity. Middle row: the most important ingredients used in the cell-by-cell modeling of the H I/H₂ transition (Section 2.3), namely the Jeans length, the surface densities of neutral hydrogen and star formation, and the UV field strength. The surface densities are computed by multiplying the respective volume densities by the Jeans length. The UV field is estimated from optically thin propagation (Section 2.4) and approximated as $U_{\text{MW}} \propto n_{\text{H}}$ in the solid blue lines. Bottom row: the predicted molecular fraction according to the various H I/H₂ models (the **GK11** predictions are very similar to that of **GD14**). For comparison, the dotted lines show an alternative UV model where U_{MW} is equal to the UV background below the star formation threshold and proportional to the SFR surface density above the threshold. This assumption tends to introduce an unphysical dip in f_{mol} at the star formation threshold (though it has no impact on the **L08** model, which does not depend on the UV field). However, even with the new modeling of the UV field, the star formation threshold introduces a slight discontinuity into many of the computed quantities, including the molecular fractions.

For gas below the star-formation threshold density, the neutral fraction is computed self-consistently in the IllustrisTNG simulations. The cooling rate is the sum of primordial cooling (computed from the temperature), metal-line cooling (computed from CLOUDY lookup tables, Ferland et al. 1998), and Compton cooling off cosmic microwave background photons. The photoionization rate due to the UV background is computed using the model of Faucher-Giguère et al. (2009, in the updated 2011 version), also taking into account nearby AGNs. The self-shielding of gas from the UV background at high densities is estimated based on Equation A1 in Rahmati et al. (2013a). The result is used to modify the input UVB photoionization rate to the CLOUDY computation of the cooling tables, and to modify the computation of the neutral fraction.

For star-forming cells, the situation is more complicated because the neutral fraction is not computed self-consistently in the star formation model of **SH03**. However, the model splits the gas into a hot and a cold phase, and it is safe to assume that the hot phase would be entirely ionized (Section 2.2.1). Realistically, the cold clouds would be partially ionized due to the UV radiation from young stars, but we neglect this contribution and assume that all cold gas is neutral, giving a fraction of neutral hydrogen of $f_{\text{HI}+\text{H}_2} = f_{\text{H}} \rho_{\text{cold}} / \rho$ (where ρ_{cold} is the overall volume density of cold clouds, not the density of individual clouds). The **SH03** model predicts a cold fraction between 0.9 and 1, meaning that this estimate is almost equivalent to simply assuming that all star-forming gas is neutral. We further discuss the uncertainty in the neutral fraction in Section 4.3.

2.2.3. Dust Abundance

Following numerous previous works (e.g., Gnedin & Kravtsov 2011; Krumholz 2013; Lagos et al. 2015), we assume that the dust-to-gas ratio is proportional to the metallicity. Thus, the dust-to-gas ratio observed locally is the same as the metallicity in solar units, $D_{\text{MW}} = Z_{\text{MW}} \equiv Z/Z_{\odot}$, where we use the solar metallicity assumed in IllustrisTNG, $Z_{\odot} = 0.0127$ (Asplund et al. 2009). Recently, there has been observational and theoretical evidence that this assumption may break down at low metallicities (Rémy-Ruyer et al. 2014; McKinnon et al. 2017; Popping et al. 2017).

2.3. Volumetric and Projected Modeling

In simulations, the natural unit is a particle or gas cell. Many of the models for the H I/H₂ transition, however, work in 2D quantities such as the surface densities of neutral gas or star formation. In most previous H I/H₂ modeling efforts (e.g., Lagos et al. 2015; Marinacci et al. 2017a), the molecular fraction was computed on a particle-by-particle (or cell-by-cell) basis, and the conversion to surface quantities was performed by multiplying with the Jeans length, $\Sigma_X = \lambda_J \rho_X$, where

$$\lambda_J = \sqrt{\frac{c_s^2}{G\rho}} = \sqrt{\frac{\gamma(\gamma-1)u}{G\rho}}. \quad (3)$$

Here, u and ρ are the internal energy per unit mass and total density of the gas cell, respectively. The underlying idea is that the Jeans length should approximate the size of a self-gravitating system (Schaye 2001; Schaye & Dalla Vecchia 2008). For the typical densities of IllustrisTNG gas cells, the Jeans length varies between about 0.1 and 10⁴ kpc (Figure 1).

We will refer to this approach as “volumetric” or “cell-by-cell” modeling (abbreviated to “vol” in some figure captions). In Section 4.1, we test the Jeans approximation explicitly by comparing its predictions to the projected surface densities of gas, neutral gas, and SFR in simulated galaxies. We find that the approximation leads to a significant bias in the recovered quantities and exhibits large cell-to-cell scatter. Thus, we propose an alternative way to model the H I/H₂ transition from the projected quantities directly, a method that we will refer to as “projected” or “map” modeling. Here, we compute the face-on surface quantities of a galaxy ab initio as follows. We first rotate the galaxy into a face-on projection by aligning its angular momentum vector with the z direction. There is no unique answer as to what constitutes the angular momentum of a galaxy because the rotation of dark matter, stars, and baryons is not necessarily aligned and depends on radius (e.g., Bett et al. 2010). As we are most concerned with the properties of gas, we attempt to measure the angular momentum of all gas cells within half the 3D gas half-mass radius. If there are fewer than 50 gas cells within this radius, we consider the measurement unreliable and revert to using the angular momentum of the stellar particles within two stellar 3D half-mass radii. We find, by visual inspection, that this algorithm reliably aligns disk-like galaxies in the desired face-on orientation. Some galaxies do not exhibit clear rotational symmetry, meaning that the algorithm produces a more or less random orientation.

Once a galaxy has been aligned, we project the quantities in question (density, stellar density, SFR, and so on) onto a face-on-oriented grid of 128² pixels, representing a physical

size of

$$L_{\text{map}} = 2 \times \max(2R_{1/2}^{\text{gas}}, 6R_{1/2}^*, 30\text{kpc}). \quad (4)$$

This expression accounts for both gas-rich and gas-poor galaxies, while the absolute minimum forces the map to contain a reasonable number of force resolution lengths ($\epsilon = 710$ pc in TNG100). The results shown in this paper correspond to projections of all cells or particles bound to the galaxy according to the SUBFIND halo finder. When comparing to observations with a limited integration depth, this procedure may need to be adjusted.

Given the complex shapes of gas cells in moving-mesh codes such as AREPO, it is numerically difficult to project the cells onto a pixel grid exactly. Instead, we apply a 2D Gaussian smoothing kernel with a width $\sigma = 1/2 L_{\text{cell}} = 1/2 (m_{\text{cell}}/\rho_{\text{cell}})^{1/3}$. This width is comparable to an smoothed particle hydrodynamics (SPH)-like smoothing length computed from 64 nearest neighbors but leads to less diffuse maps. We have confirmed that increasing or decreasing the smoothing scale by a factor of two changes the median H₂ mass of galaxies by less than 10%. We have also experimented with a simple particle-in-cell algorithm (i.e., without any smoothing), but we find that such maps systematically overestimate the molecular fraction because they tend to concentrate too much mass in a few pixels. Finally, we have confirmed that our results are converged with map resolution, in other words, that doubling the resolution to 256² leads to no appreciable change in the predicted average molecular masses. The median pixel size for our 128² maps is 0.8 kpc in TNG100 and 1.3 kpc in TNG300.

Figure 2 shows an example of maps for a gas-rich, Milky Way-sized disk galaxy from TNG100. The projected galaxy properties allow us to compute more faithful representations of required physical quantities such as surface densities. However, we have lost access to any volumetric quantity. For example, we cannot distinguish star-forming and quiescent cells any more. For the remainder of the paper, we will compute all models in both their volumetric and projected incarnations, implying computations on a cell-by-cell and pixel-by-pixel basis, respectively.

We note that there are characteristic length scales other than the Jeans length that could be used in cell-by-cell modeling. For example, there is the Sobolev length, $\rho/|\nabla\rho|$, the scale over which the density of a gas distribution changes by order unity (e.g., Gnedin et al. 2009). However, we do not experiment with such approximations because it is nontrivial to compute the Sobolev length from the moving-mesh grid cells in AREPO.

2.4. The UV Field

The H₂ abundance is governed by the balance between recombination on dust grains and dissociation by UV photons in the LW band. Thus, the majority of the H I/H₂ models we consider rely on an estimate of the UV radiation field. Since cosmological simulations such as IllustrisTNG currently do not include radiative transfer, we need to estimate it. This computation turns out to be one of the most uncertain and challenging aspects of modeling the H I/H₂ transition. In this section, we briefly summarize our algorithm and refer the reader to Appendix A for a more detailed description. We parameterize the UV field as U_{MW} , the field at 1000 Å in units of the observationally measured field in the local neighborhood according to Draine (1978).

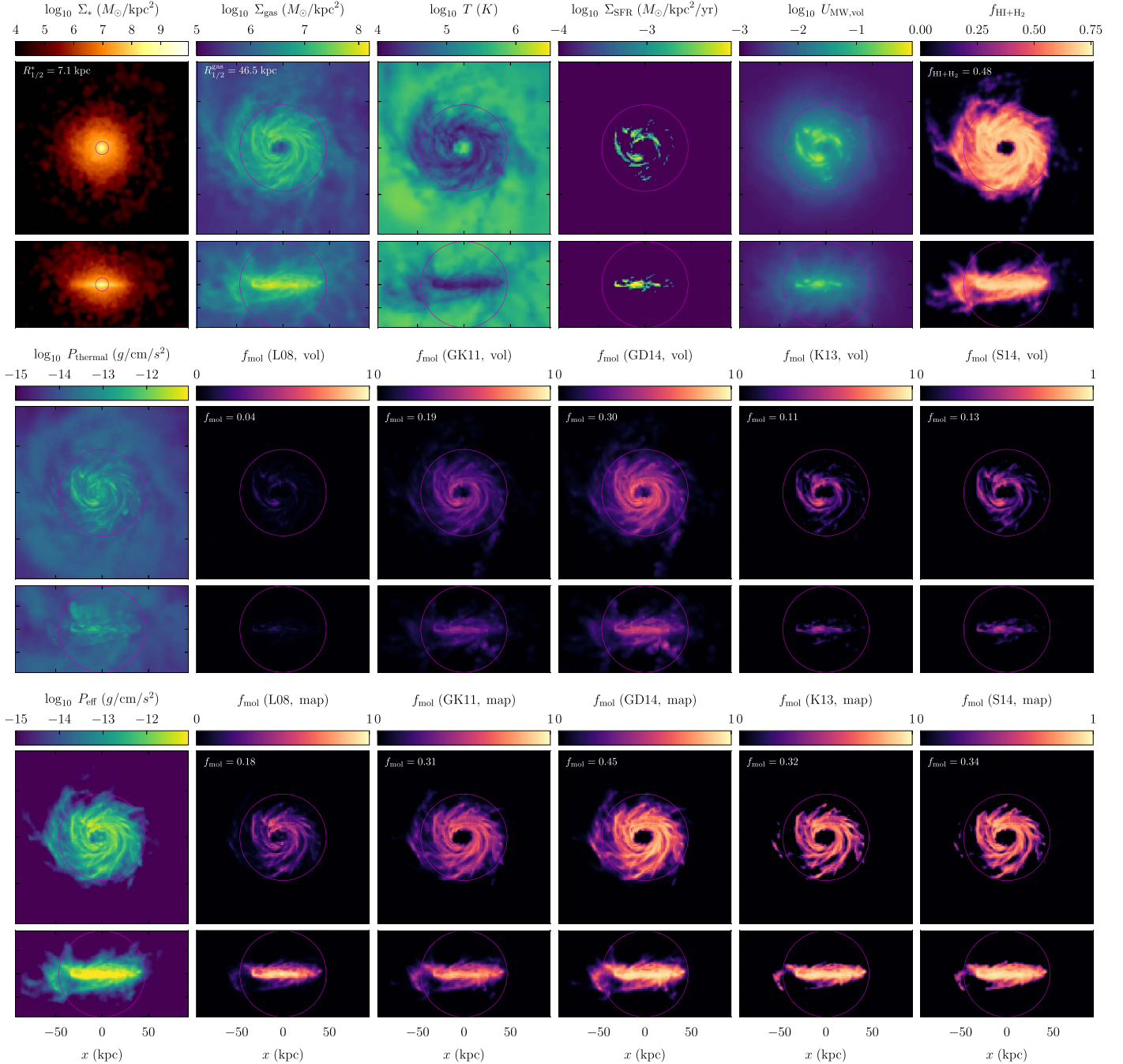


Figure 2. Projections of physical properties and the resulting molecular fractions for a gas-rich disk galaxy in TNG100. At $z = 0$, this galaxy has a halo mass of $M_{200c} = 10^{12} M_{\odot}$, a total gas mass of $10^{11} M_{\odot}$, a stellar mass of $4 \times 10^{10} M_{\odot}$, an SFR of $1.4 M_{\odot} \text{ yr}^{-1}$, and a relatively uniform metallicity of $1.2 Z_{\odot}$ on average, and it hosts a black hole of about $10^8 M_{\odot}$. The purple circles indicate the stellar and gas half-mass radii. Each set of panels shows the face-on (perpendicular to the rotation axis) and edge-on projections using all stellar population particles or gas cells that are gravitationally bound to the galaxy. The maps of temperature, thermal pressure, UV field strength, and neutral fraction are weighted by total gas mass, and the maps of molecular fractions by the neutral hydrogen mass. Top row: The UV field strength U_{MW} is sourced from star-forming regions and propagated into the surrounding regions (Section 2.4). The neutral fraction refers to the total gas density and thus cannot exceed the primordial hydrogen fraction, 0.76. Middle row: predictions for the molecular fraction according to the volumetric version of the H I/H₂ models. The white labels list the overall molecular fractions of the galaxy. Bottom row: same as the middle row, but for the projection-based version of the H I/H₂ models. With the exception of the volumetric L08 model, the H I/H₂ models vary in their predictions between 11% and 45% molecular fraction, a typical spread for individual galaxies (Section 3.2). The large difference between the volumetric and projected versions of the L08 model is clearly caused by the underlying estimates of the pressure: the thermal pressure (middle left panel) is a poor approximation to the midplane pressure (bottom left panel) that the model should be based on (Equation (6)). The volume-weighted thermal pressure would be even lower than the mass-weighted pressure shown here.

We begin by recognizing that the vast majority of LW photons originate from very young stars whose UV flux decreases by orders of magnitude within 100 Myr (e.g., Leitherer et al. 1999). In IllustrisTNG, the SFR evolves on timescales of more than half a gigayear at $z \lesssim 1$ and a few

hundred megayears at higher redshift (Torrey et al. 2018). Thus, we can safely estimate the UV field from the SFR, though this approximation becomes less accurate at high redshift. In contrast, the stellar population particles in the simulation are stochastic tracers of the SFR, and only the very

most recently formed ones would contribute significant UV flux, rendering the calculation spatially inhomogeneous.

In previous works, the UV field in star-forming gas cells has been scaled to the observed local star-formation surface density, $U_{\text{MW}} = \Sigma_{\text{SFR}}/\Sigma_{\text{SFR,local}}$ (Lagos et al. 2015). With this method, however, it is not clear how to treat quiescent cells below the star formation threshold. These cells can contain a significant fraction of neutral hydrogen (Figure 1). In the past, either their H_2 contribution was ignored (Marinacci et al. 2017a) or their UV field was set to the cosmic UV background (Lagos et al. 2015), a small fraction of the local field at $z=0$ (Appendix A.1). The latter choice creates a sharp break in the UV field at the star formation threshold, which is itself an arbitrary byproduct of the ISM model (Figure 1).

Instead, we assume that some fraction of the UV emitted by star-forming cells is absorbed in situ, and that the remaining UV propagates through an optically thin medium. This approximation is likely flawed for dust-rich, high-metallicity systems, but has the advantage that it relies on only one free parameter: the escape fraction. We calibrate this parameter to 10% by comparing the SFR–UV relation to solar neighborhood values (Section 4.2). The UV background serves as a lower limit for cells far away from any star formation. An example of the resulting UV field is shown in Figure 2. In the projection-based models, we proceed equivalently for each pixel in the projected maps. We describe our procedure in detail in Appendix A. In Section A.3, we show that even drastic changes in the escape fraction have a relatively modest effect on the molecular fraction.

The center-right panel of Figure 1 shows a histogram of U_{MW} for the gas cells in TNG100 galaxies. The UV field follows a roughly linear relation with density on average, $U_{\text{MW}} \approx 10 n_{\text{H}}/(1\text{cm}^{-3})$, but with large scatter. The scatter is expected from higher-resolution simulations that resolve the ISM structure better (see, e.g., Figure 11 in Hu et al. 2017). Compared to the model based only on local star formation and the UV background (dotted line in Figure 1), our optically thin propagation transports UV radiation to the lower-density parts of galaxies. Comparing to the dotted lines in the bottom panels of Figure 1, it is clear that this radiation suppresses H_2 formation in those regions. Furthermore, our modeling avoids the large unphysical jump in f_{mol} that results from the discontinuity in the UV field at the star formation threshold. However, a smaller discontinuity remains, due to the discontinuous nature of other quantities such as the estimated neutral surface density. This discontinuity is removed in the projection-based models because a threshold in the volumetric density does not translate into a threshold in surface density. We discuss the uncertainties of our optically thin UV modeling and the escape fraction in Section 4.2.

2.5. Models for the $\text{H I}/\text{H}_2$ Transition

Of the numerous models for the molecular fraction that have been proposed, we consider those of Leroy et al. (2008, L08), Gnedin & Kravtsov (2011, GK11), Krumholz (2013, K13), Gnedin & Draine (2014, GD14), and Sternberg et al. (2014, S14). As described in Section 2.3, we compute each model on a cell-by-cell basis and in projection.

2.5.1. Observed Correlations (L08)

The ratio of molecular to atomic gas has been observed to correlate strongly with the midplane pressure (Wong & Blitz 2002; Blitz & Rosolowsky 2004, 2006; Leroy et al. 2008):

$$R_{\text{mol}} = \left(\frac{P_{\text{eff}}}{P_0} \right)^\alpha, \quad (5)$$

where we use the Leroy et al. (2008) values of $P_0/k_{\text{B}} = 1.7 \times 10^4 \text{ K cm}^{-3}$ and $\alpha = 0.8$ (their Equation (34)). This model depends only on hydrodynamical and stellar quantities but not on the UV field or metallicity.

For the cell-by-cell computation, we follow Marinacci et al. (2017a) in setting $P_{\text{eff}} \equiv P_{\text{thermal}} = u\rho(\gamma - 1)$. For star-forming cells, Marinacci et al. (2017a) used the partial pressure of the cold phase; instead, we use the total pressure. However, Equation (5) actually refers to the midplane pressure of a disk in hydrostatic equilibrium (Elmegreen 1989):

$$P_{\text{eff}} \approx \frac{\pi}{2} G \Sigma_{\text{gas}}^2 \left(1 + \frac{\sigma_{\text{gas},z}}{\sigma_{*,z}} \frac{\Sigma_{*}}{\Sigma_{\text{gas}}} \right), \quad (6)$$

where $\sigma_{\text{gas},z}$ and $\sigma_{*,z}$ are the vertical velocity dispersions of gas and stars, respectively. Thus, for the projection-based version of the L08 model, we compute P_{eff} from surface maps of the relevant quantities, as described in Appendix C.1.

We find that the mass-weighted average thermal pressure maps significantly differ from P_{eff} . Because the thermal pressure tends to be lower in high-density regions, the volumetric L08 model predicts systematically lower molecular fractions than the projected version (Figure 2 and Section 3.2). This mismatch is not surprising because the resolution of IllustrisTNG is not sufficient to resolve the disk scale height, meaning that the pressure cannot be expected to match the true midplane pressure. In high-resolution zoom-in simulations, the pressure should represent the actual hydrostatic equilibrium, which explains why Marinacci et al. (2017a) found that the L08 and GK11 models agree reasonably well when applied to the AURIGA simulations.

In the context of this work, we conclude that the $P_{\text{eff}} = p_{\text{thermal}}$ assumption does not hold, meaning that the volumetric L08 model is not physical and that its predictions should not be taken at face value. We thus omit the volumetric L08 model from most figures and conclusions.

2.5.2. Calibration from Simulations (GK11 and GD14)

The GK11 model is based on 35 high-resolution simulations of isolated disk galaxies that follow the detailed chemical evolution of the gas (see also Gnedin et al. 2009). Each simulation is initialized with a particular value of the dust-to-gas ratio and the interstellar radiation field. GK11 present a fitting function for the resulting molecular fraction averaged over a scale of 500 pc. This function takes U_{MW} , metallicity, and the surface density of neutral gas as input parameters (Appendix C.2).

Here, $\Sigma_{\text{H I}+\text{H}_2}$ is computed based on the Jeans approximation in the volumetric version of the model, and directly in the projected version. Our pixel sizes of roughly a kiloparsec are similar to the 500 pc adaptive mesh refinement cells that GK11 used to calibrate their formulae.

The **GD14** model builds on the **GK11** model by considering the effects of line overlap. The model is based on the same quantities as **GK11**, but its mathematical expressions are somewhat different and explicitly take resolution into account (Appendix C.3). We expect the **GK11** and **GD14** models to predict similar molecular fractions.

2.5.3. Analytic Models (K13 and S14)

The **K13** model represents an extension of the model of Krumholz et al. (2009a, hereafter **KMT**; see also Krumholz et al. 2009b and McKee & Krumholz 2010), who developed a formula for f_{mol} as a function of the neutral gas surface density and metallicity. Their model, however, predicts that f_{mol} rapidly approaches zero below some critical, Z -dependent surface density. This prediction is inconsistent with observations and with the requirement that the ISM maintain vertical hydrostatic equilibrium. Following Ostriker et al. (2010), **K13** added the latter condition, providing an iterative model that computes f_{mol} and the surface density of star formation as a function of $\Sigma_{\text{HI}+\text{H}_2}$ and metallicity. Instead of computing Σ_{SFR} , we impose the UV field externally and iterate to find f_{mol} as a function of $\Sigma_{\text{HI}+\text{H}_2}$, U_{MW} , and Z .

The mathematical expressions and details of our algorithm are given in Appendix C.4. In the projected version of this model, we proceed similarly to the **GK11** model, meaning that we compute all surface densities directly and apply the model equations to the 2D maps. We expect that the **K13** and **GK11** models should agree relatively well (Krumholz & Gnedin 2011).

The model of **S14** is similar in spirit to the **KMT** prescription but improves on some aspects of the physical modeling. Instead of approximating f_{mol} directly, the **S14** model computes the column density of HI as a function of $\Sigma_{\text{HI}+\text{H}_2}$, U_{MW} , and Z . This HI column can then be subtracted from the total column density to find the molecular column density. We further discuss this model in Appendix C.5.

3. Results

In this section, we evaluate the predictions of the various HI/H₂ models. We begin by checking whether the well-known relation between surface density and star formation is obeyed on a pixel-by-pixel basis. We compare the model results on a galaxy-by-galaxy basis and in summary statistics such as gas fractions and mass functions. Eventually, we consider the spatial distribution of molecular gas within galaxies and discuss potential changes at high redshift.

3.1. The KS Relation

The KS relation (Schmidt 1959; Kennicutt 1998) is a correlation between the surface densities of gas and star formation. As our HI/H₂ models do not explicitly tie f_{mol} to the SFR, this relation provides a basic test of how our modeling assigns molecular and atomic gas in relation to star formation. We perform this comparison in a spatially resolved manner, by considering the pixels in the projected maps of our galaxies. We could also average the surface densities over entire galaxies, but those quantities would be subject to the definition of the size of a galaxy.

Figure 3 shows histograms of $\Sigma_{\text{HI}+\text{H}_2}$, Σ_{HI} , and Σ_{H_2} versus Σ_{SFR} for all pixels in the maps of all galaxies in TNG100 (a large fraction of pixels lie outside the plotted range). This

stack is dominated by relatively small galaxies, but we have confirmed that the plots do not change systematically if only galaxies with Milky Way-like stellar masses are considered. We are showing Σ_{HI} and Σ_{H_2} according to the projected **GD14** model, but the other models produce similar results.

The left panel of Figure 3 shows the relation between $\Sigma_{\text{HI}+\text{H}_2}$ and Σ_{SFR} . The red line shows the Kennicutt (1998) relation with an exponent of $N = 1.4$, and orange contours roughly indicate the spatially resolved observations of Bigiel et al. (2008). Their contours refer to resolution elements of a size of about 0.75 kpc, well matched with the median pixel size for TNG100 (Section 2.3). Bigiel et al. (2008) cite lower limits on the detection of star formation surface density (roughly $10^{-4} M_{\odot} \text{ yr}^{-1} \text{ kpc}^{-2}$) and molecular surface density ($10^{0.5} M_{\odot} \text{ pc}^{-2}$), which are shown as dot-dashed gray lines in Figure 3.

Though not explicitly enforced in the simulation, we expect IllustrisTNG to obey a relation similar to that of Kennicutt (1998) because the star formation law was designed to match such observations. In particular, the SFR is proportional to $\rho^{1.5}$ in the **SH03** model (Section 2.2.1). While the projection onto surface densities could change this slope, roughly constant scale heights will naturally lead to a similar relation between the respective surface densities. Nevertheless, the simulation also matches the observed turnoff from the relation around $\Sigma_{\text{HI}+\text{H}_2} \approx 1 M_{\odot} \text{ pc}^{-2}$.

The center panel of Figure 3 shows the relation between Σ_{HI} and Σ_{SFR} . Our simulation results qualitatively match the Bigiel et al. (2008) contours. In particular, the observed ‘‘saturation’’ of HI at $\Sigma_{\text{HI}} \approx 1 M_{\odot} \text{ pc}^{-2}$ is obeyed, presumably because the HI/H₂ models were designed to match the observationally known transition to molecular gas at this surface density.

Finally, the right panel of Figure 3 shows the KS relation for molecular gas. The IllustrisTNG data match the observed contours well at intermediate densities around $\Sigma_{\text{H}_2} \approx 10 M_{\odot} \text{ pc}^{-2}$. At higher densities, however, the simulation results are well described by a steep relation, $\Sigma_{\text{SFR}} \propto \Sigma_{\text{H}_2}^{1.5}$ (dark blue line). In contrast, observations favor a shallower relation with a more or less constant depletion time, as shown by the contours and literature compilation (Bigiel et al. 2011; Schruba et al. 2011; Leroy et al. 2013, 2017; Bolatto et al. 2017). We note that the HI/H₂ modeling cannot be blamed for this difference: at surface densities above $\Sigma_{\text{H}_2} \approx 1 M_{\odot} \text{ pc}^{-2}$, virtually all hydrogen is molecular, meaning that no change in the models could shift the IllustrisTNG data to the higher surface densities required to match the observations. At lower Σ_{H_2} , the scatter increases, but it is clear that there is a population of pixels with a detectable $\Sigma_{\text{H}_2} \approx 10^{0.5} M_{\odot} \text{ pc}^2$ and very little or no star formation. High-resolution simulations with H₂ physics have also found that H₂ is not always correlated with star formation (Pelupessy et al. 2006; Hu et al. 2016, see also Krumholz et al. 2011 and Glover & Clark 2012).

Of course, the comparisons in Figure 3 are crude because we did not take observational systematics into account. For example, the observational star-formation indicators used by Bigiel et al. (2008), FUV and 24 μm fluxes, are sensitive to star formation on a 0.1–1 Gyr timescale, whereas we consider instantaneous SFRs. Moreover, the H₂ depletion time has been shown to depend on a number of variables, including galaxy stellar mass and the dynamical state of the star-forming gas (Saintonge et al. 2011; Leroy et al. 2017). Similarly, Schruba et al. (2018) find that the HI saturation density depends on

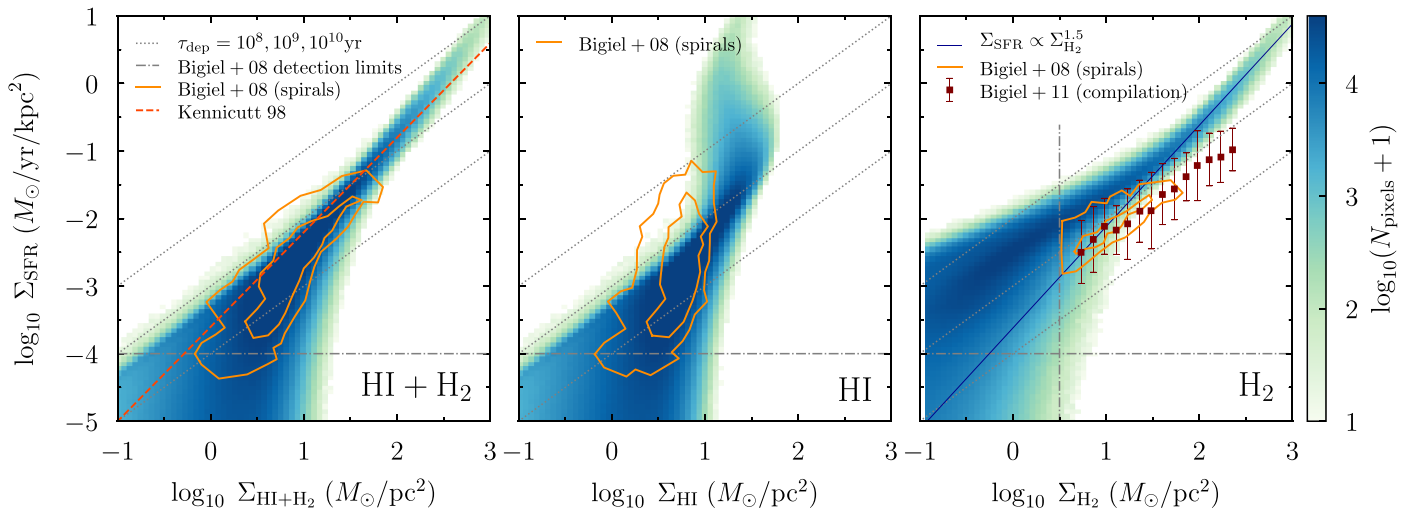


Figure 3. Kennicutt–Schmidt relation for neutral (left), atomic (center), and molecular (right) gas. The histograms show the distribution of pixels from the projected maps of TNG100 galaxies. In each panel, the dotted lines correspond to constant depletion times of $\tau_{\text{dep}} = \Sigma_{\text{gas}}/\Sigma_{\text{SFR}}$. The orange contours roughly indicate the location of the spatially resolved observations of seven spiral galaxies of Bigiel et al. (2008), and the horizontal and vertical dot-dashed lines mark their sensitivity limits for star formation and H_2 surface density. Left panel: The total gas KS relation in IllustrisTNG is independent of the $\text{H I}/\text{H}_2$ models and roughly follows the Kennicutt (1998) relation (red dashed line) at high densities, largely by construction. Around a surface density of $1 M_{\odot} \text{pc}^{-2}$, the relation steepens and falls below the Kennicutt (1998) relation, in agreement with observations. Center panel: the distribution of Σ_{HI} according to the projected GD14 model. The other models give very similar trends. The relation between Σ_{HI} and Σ_{SFR} is less pronounced than for $\Sigma_{\text{HI}+\text{H}_2}$ and Σ_{H_2} , in agreement with Bigiel et al. (2008). Right panel: The molecular KS relation is not quite obeyed by the simulation. The Bigiel et al. (2008, 2011) data exhibit a roughly constant depletion time, whereas the simulation results are better described by a superlinear relation with a slope of 1.5 (dark blue line). See Section 3.1 for a detailed discussion.

metallicity. Finally, there are hints of a steeper KS relation at high redshift (Hodge et al. 2015). These effects complicate the interpretation of the results shown in Figure 3. Nevertheless, we conclude that our $\text{H I}/\text{H}_2$ models give reasonable results for the surface densities of H I and H_2 and for their relation to the star-formation surface density.

3.2. Galaxy-by-galaxy Comparison of Model Results

We now assess how well the $\text{H I}/\text{H}_2$ models agree with each other by considering the integrated molecular masses of individual galaxies. Figure 4 shows histograms of the ratio of H_2 masses predicted by each pair of models. In each panel, the numbers on the bottom right give the median ratio and scatter in dex. These values are shown separately for TNG100 and TNG300, and the two simulations are weighted equally in the histograms to account for the larger number of galaxies in TNG300. The lower left sector compares the volumetric versions of the models, the upper right sector the projected versions, and the panels along the diagonal compare the volumetric and projected version of each model.

As expected from the previous discussion, the volume-based L08 model (left column) predicts molecular masses 0.2–0.6 dex lower than other models, and we will ignore it for the rest of this section. The volumetric and projected versions of each model (red histograms) agree to better than 0.02 dex in TNG100, though much less well (0.3 dex) in TNG300. In both simulations, the scatter around this average is large, between 0.2 and 0.4 dex. We conclude that while the volumetric and projected versions agree reasonably well on average, their predictions diverge for many individual galaxies. This scatter is likely a consequence of the fact that the Jeans approximation to the scale height works on average but exhibits significant scatter (Section 4.1).

Comparing the different volumetric models, we find that they agree to 0.13 dex in TNG100 and 0.3 dex in TNG300.

Given that the GK11 and GD14 models are based on the same underlying physics, it is not surprising that they exhibit a smaller scatter than the other pairs of models. Comparing the projected models, we find a large range of offsets between zero and 0.25 dex. Interestingly, the agreement is roughly equally good in TNG100 and TNG300, indicating that the projected modeling is less sensitive to resolution effects.

We note that the scatter visibly increases toward lower H_2 masses, meaning that the values quoted in Figure 4 are sensitive to the lower mass cutoff. For this reason, we will consider only galaxies with $M_{\text{H}_2} > 10^8 M_{\odot}$ when comparing mass functions in Section 3.4. We have confirmed that the trends shown in Figure 4 hold even when only galaxies with high stellar mass or high SFR are selected. Such cuts do not necessarily reduce the scatter between the models, indicating that the differences are not unique to low-mass galaxies.

3.3. Gas Fractions

Having assessed the differences between $\text{H I}/\text{H}_2$ models for individual galaxies, we now turn to averaged quantities such as the gas fractions shown in Figure 5. We refrain from observational comparisons at this point because all masses correspond to the total gas or stellar mass bound to a galaxy, which is generally not what is measured observationally. Moreover, we have not made any attempt to mimic observational sample selections or account for nondetections. Our sample is complete down to the stellar masses shown because we selected galaxies by either their stellar or gas mass (Table 1).

Before analyzing the trends in H I and H_2 , it is instructive to consider the total neutral gas fraction because it is independent of the $\text{H I}/\text{H}_2$ modeling and sets a baseline for the abundance of the atomic and molecular phases. At fixed stellar mass, TNG300 galaxies contain about three times less neutral gas than TNG100 galaxies (left panel of Figure 5), a resolution effect that also affects the H I and H_2 fractions because they

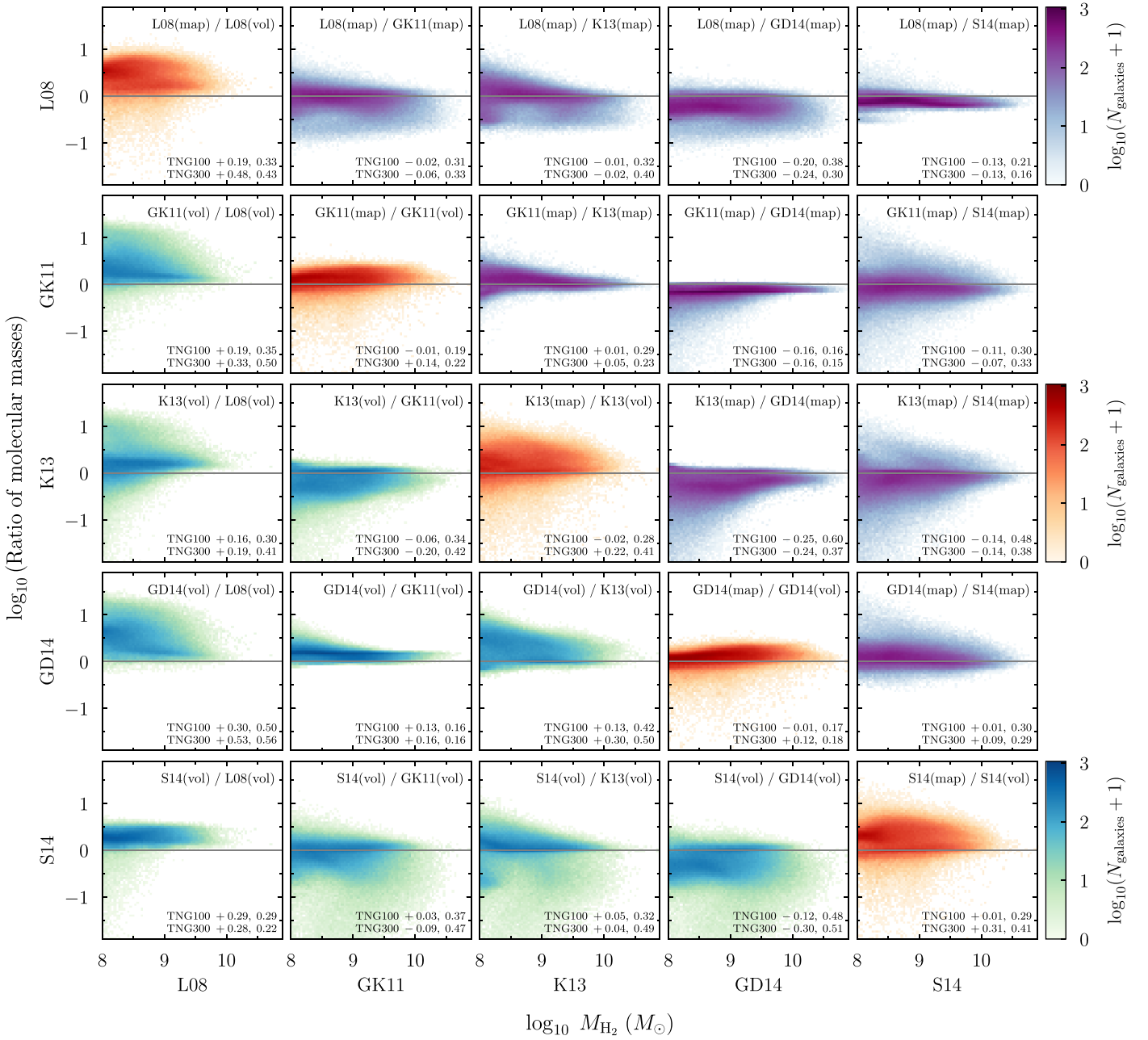


Figure 4. Galaxy-by-galaxy comparison of the H_2 masses predicted by the models studied in this paper. Each panel shows a histogram of TNG100 and TNG300 galaxies, with equal weight on each of the two simulations. The bottom left (blue) panels compare the H_2 masses predicted by the volumetric versions of the models, the top right (purple) panels the projection-based versions, and the diagonal (red) panels compare the volumetric and projection results for each model. The numbers at the bottom right of each panel give the median offset in dex as well as the difference between the 84th and 16th percentiles (i.e., the 68% scatter in dex). Those numbers are given for TNG100 and TNG300 separately and refer only to the galaxies above the threshold of $M_{H_2} > 10^8 M_\odot$ according to the model plotted on the x axis. The volumetric and projection-based models tend to give similar predictions on average (especially in TNG100) but exhibit significant galaxy-to-galaxy scatter, about 0.2–0.4 dex. The exception is the L08 model (left column), where the projection-based model predicts systematically more H_2 because the volumetric version is not physical (Section 2.5.1). See Section 3.2 for a detailed discussion.

depend on the surface density of neutral gas. The stellar mass in TNG100 is about 1.4 times greater than in TNG300 at fixed halo mass (Pillepich et al. 2018b), but we find the same disagreement in neutral gas mass when comparing at fixed halo mass.

The center and right panels of Figure 5 shows the HI and H_2 fractions, respectively. As expected from the average agreement between HI/ H_2 models in Figure 4, we find that the median gas fractions according to the different models agree reasonably well. In particular, we cannot discern significant

overall offsets between the volumetric or projected models. The H_2 fraction in TNG100 exhibits a typical (median) range of 0.4 dex within stellar mass bins, with a maximum of 1 dex. These differences lead to smaller disagreements in the HI fraction, typically 0.1 dex and up to 0.3 dex. The agreement is slightly worse in TNG300.

We conclude that the spread in the HI and H_2 fractions due to differences between the HI/ H_2 models is relatively small compared to the systematic uncertainties in our modeling, especially for the better resolved TNG100. The HI gas fraction

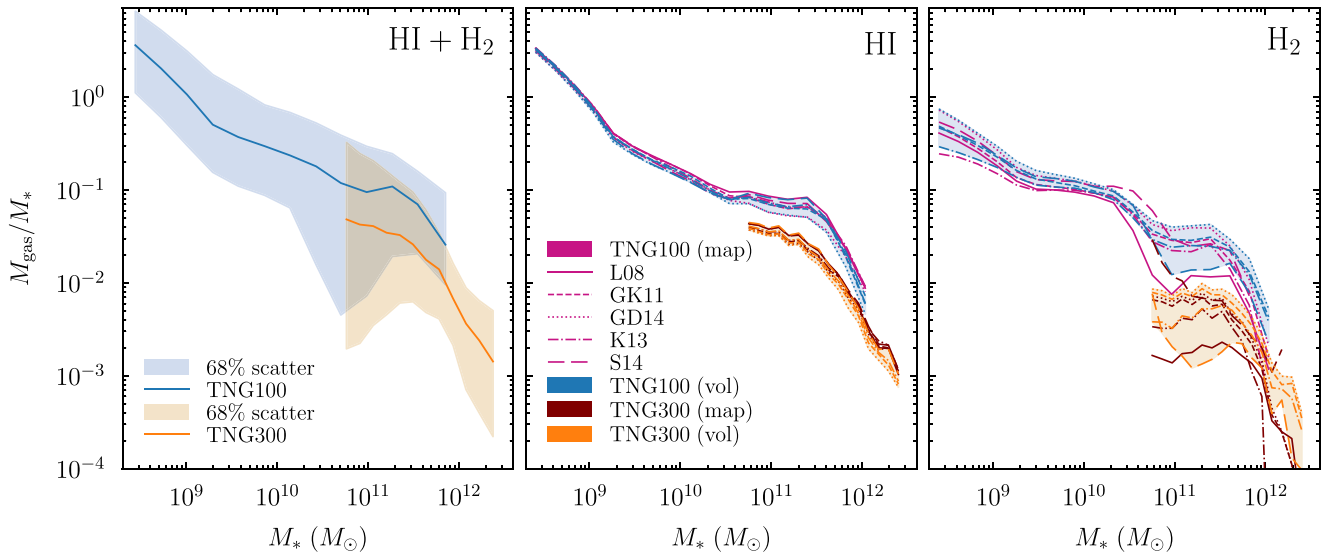


Figure 5. Gas-to-stellar mass ratio for neutral (left), atomic (center), and molecular gas (right). The total neutral gas fraction is shown as a median (solid lines) and 68% scatter (shaded areas). The gas fraction is a factor of two to three lower in TNG300 at fixed stellar mass, indicating that this quantity is not converged with mass resolution. In the center and right panels, the blue and orange lines show the median predictions of the volumetric versions of the H I/H₂ models, and the pink and brown lines show the projected versions. The area between the volumetric models is shaded to indicate the range of their predictions (the shading does not indicate scatter). For TNG100, the molecular masses agree to about 0.4 dex in most mass bins, with differences up to 1 dex. The agreement is worse in TNG300, with a typical spread of 0.8 dex. Because molecular gas is subdominant at $z = 0$, these differences translate into smaller spreads in the H I mass, typically 0.1 dex in TNG100 and 0.15 dex in TNG300, with a maximum of 0.3 dex. The gas fractions include all gas and stellar mass bound to the galaxy and are thus not suitable for comparisons to observations that correspond to a particular aperture or sensitivity limit.

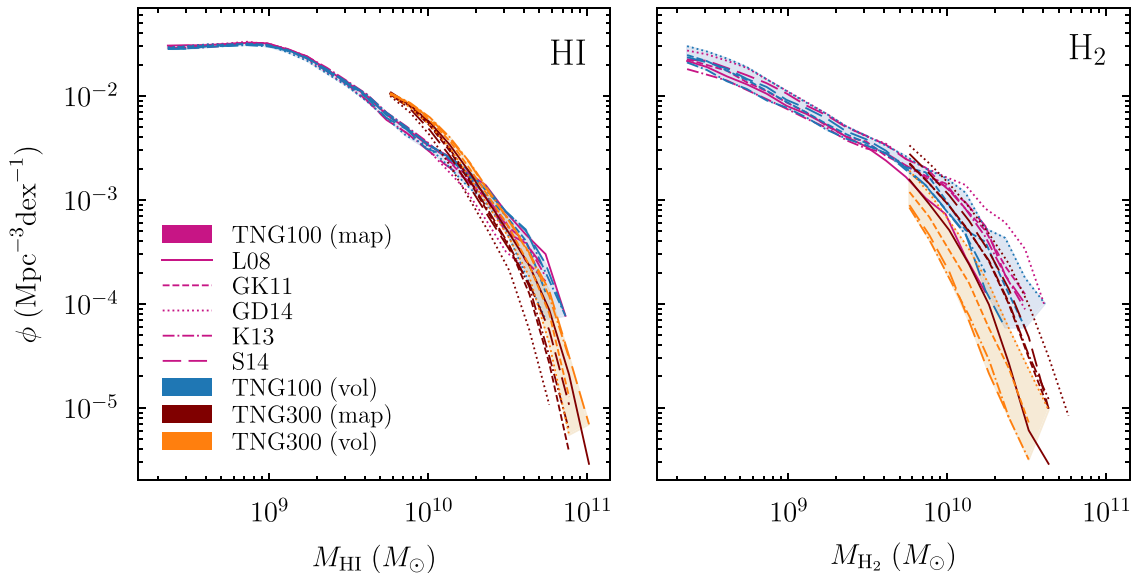


Figure 6. Mass functions of H I (left) and H₂ (right). The lines and shaded regions have the same meaning as in Figure 5. The different H I/H₂ models agree very well with each other, with the exception that the volumetric versions predict a lower H₂ mass function in TNG300. Generally, the agreement between TNG100 and TNG300 appears to be better than in the gas mass fractions. As in Figure 5, the gas masses contain all cells bound to a galaxy and should thus not be compared to observations.

is particularly well constrained, allowing detailed comparisons to observations in future work.

3.4. Mass Functions

Given that the gas fractions at fixed stellar mass agree reasonably well between the H I/H₂ models, we might expect that the mass functions would match as well. However, we investigated only the median fractions in Figure 5. Large scatter in any one model could change the predictions for the mass

functions significantly because they fall off steeply toward high masses.

However, Figure 6 shows that the H I and H₂ mass functions agree very well between the different H I/H₂ models, particularly for TNG100. In TNG300, we witness a surprising reversal of the trend seen in Figure 5: while the volumetric models predict a slightly higher median f_{mol} , they predict a slightly lower H₂ mass function at the high-mass end. This seeming contradiction highlights that the mass function depends on the full distribution of gas masses and thus represents a separate check on the modeling. Overall, the mass

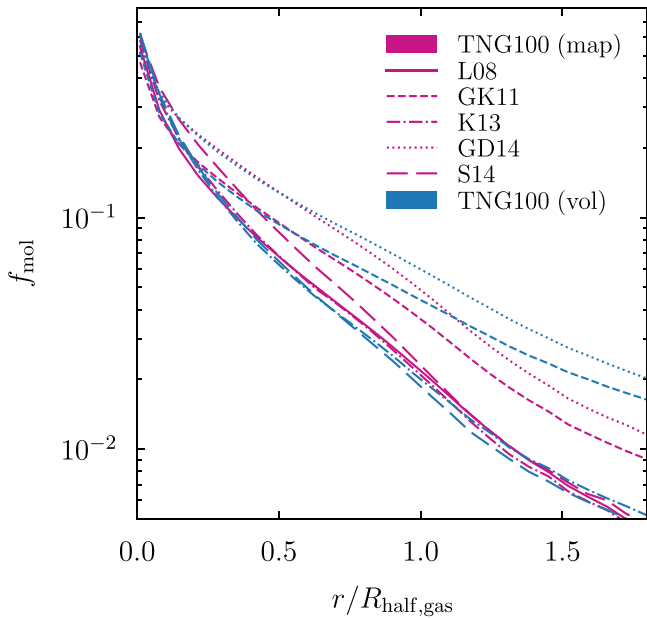


Figure 7. Projected mean radial profiles of the molecular fraction for TNG100 galaxies with $M_{\text{H}_2} > 10^8 M_\odot$. The line styles and colors have the same meaning as in Figures 5 and 6. The GK11 and GD14 models (short-dashed and dotted lines) predict higher molecular fractions at large radii than the other models because they allow for more molecular gas at low densities (Appendix C). While the mean profiles are dominated by low-mass galaxies, the profiles for higher-mass samples lead to the same qualitative conclusions. The corresponding median profiles fall to zero between 0.3 and 1.5 half-mass radii.

functions agree better between TNG100 and TNG300 than the gas fractions, particularly for the projected models.

3.5. Spatial Distribution of H_2

While the $\text{H I}/\text{H}_2$ models generally agree on the average gas fractions and mass functions, they differ in the spatial distribution of H_2 . Figure 7 shows mean radial profiles of f_{mol} (without any mass weighting). We include only galaxies with $M_{\text{H}_2} > 10^8 M_\odot$ according to the respective $\text{H I}/\text{H}_2$ model because the profiles of molecule-poor galaxies tend to be noisy. While the stacked sample is dominated by low-mass galaxies, we have confirmed that the profiles look qualitatively the same for galaxies with high stellar mass. We scale the radii by the gas half-mass radius, excluding galaxies with a half-mass radius below 5 kpc. Both the volumetric and projected model profiles refer to projected radii in the face-on orientation.

Figure 7 demonstrates that the mean radial profiles of the molecular fraction are remarkably similar when the L08, K13, or S14 models are used. In contrast, the GK11 and GD14 profiles fall off less steeply with radius. We show in Appendix C that the latter models allow for slightly more molecular gas at lower densities, which explains the offset in the profiles.

3.6. Results at Higher Redshifts

All of the results shown to this point have referred to $z=0$. In this section, we check whether our conclusions are changed at higher redshifts, in particular $z=2$ and $z=4$ in TNG100. We expect some systematic trends because high-redshift galaxies tend to be richer in gas, denser, and subject to a higher UV background field.

Comparing to Figure 4, we find that the galaxy-to-galaxy scatter between the volumetric and projected versions of the models decreases significantly at high z , from about 0.2–0.3 dex at $z=0$ to roughly 0.14 dex at $z=4$, while the good average agreement remains intact. The scatter also tends to decrease between pairs of models, although there are some exceptions. The agreement between different pairs of models evolves differently. For example, the agreement between the projected GK11 and K13 models gets better (0.25 dex at $z=0$ and 0.06 at $z=4$), whereas the volumetric GK11 and S14 models agree much less (0.03 dex at $z=0$ and 0.45 at $z=4$). We defer a detailed investigation of the reasons for this behavior to future work.

Compared to Figure 5, both the average neutral gas fraction ($M_{\text{H I}+\text{H}_2}/M_*$) and molecular fraction (M_{H_2}/M_*) increase by about an order of magnitude by $z=4$, with unchanged dependencies on stellar mass. The differences between the $\text{H I}/\text{H}_2$ models remain roughly the same as at $z=0$, which translates into larger differences in the H I fractions because molecular gas makes up for a larger fraction of the total neutral masses. These conclusions are independent of whether the projected or volumetric model versions are used.

The H_2 mass function at $z=2$ is essentially unchanged from $z=0$ and increases by a factor of two to five by $z=4$. Again, the spread between the $\text{H I}/\text{H}_2$ models and the agreement between volumetric and projected models remain roughly constant. The H I mass function falls slightly with redshift, and the spread of the model increases for the same reasons as explained above.

We conclude that the scatter between the $\text{H I}/\text{H}_2$ models decreases at high z , though some models disagree more systematically. Overall, these details have little effect on the averaged gas fractions and mass functions.

4. Discussion

We have presented two avenues for modeling the molecular fraction in simulated galaxies, based on volume elements and on projections. We have applied these methods to five models of the $\text{H I}/\text{H}_2$ transition and shown that the models agree well in their average predictions but disagree for individual galaxies and in the spatial dependence of the molecular fraction. In this section, we further discuss the reasons for the differences between the volumetric and projected modeling and highlight some of the most important systematic uncertainties that affect our analysis.

4.1. How Accurate Is the Jeans Approximation?

The $\text{H I}/\text{H}_2$ models proposed in the literature generally depend on column densities or surface densities of various quantities. Whenever such calculations are to be performed for each gas cell, the cell’s volume density must be converted to a column (or surface) density, necessitating a characteristic length scale. A common solution has been to multiply the volume density by the Jeans length, that is, to compute the surface density of a quantity X as $\Sigma_X = \lambda_J \rho_X$ (e.g., Lagos et al. 2015; Marinacci et al. 2017a). In this section, we investigate the accuracy of this approximation specifically in the context of galaxies in cosmological simulations.

By definition, the Jeans length quantifies the size of a system in hydrostatic equilibrium (Equation (3)). Schaye (2001) thus argued that it should be a good proxy for the sizes of the gas

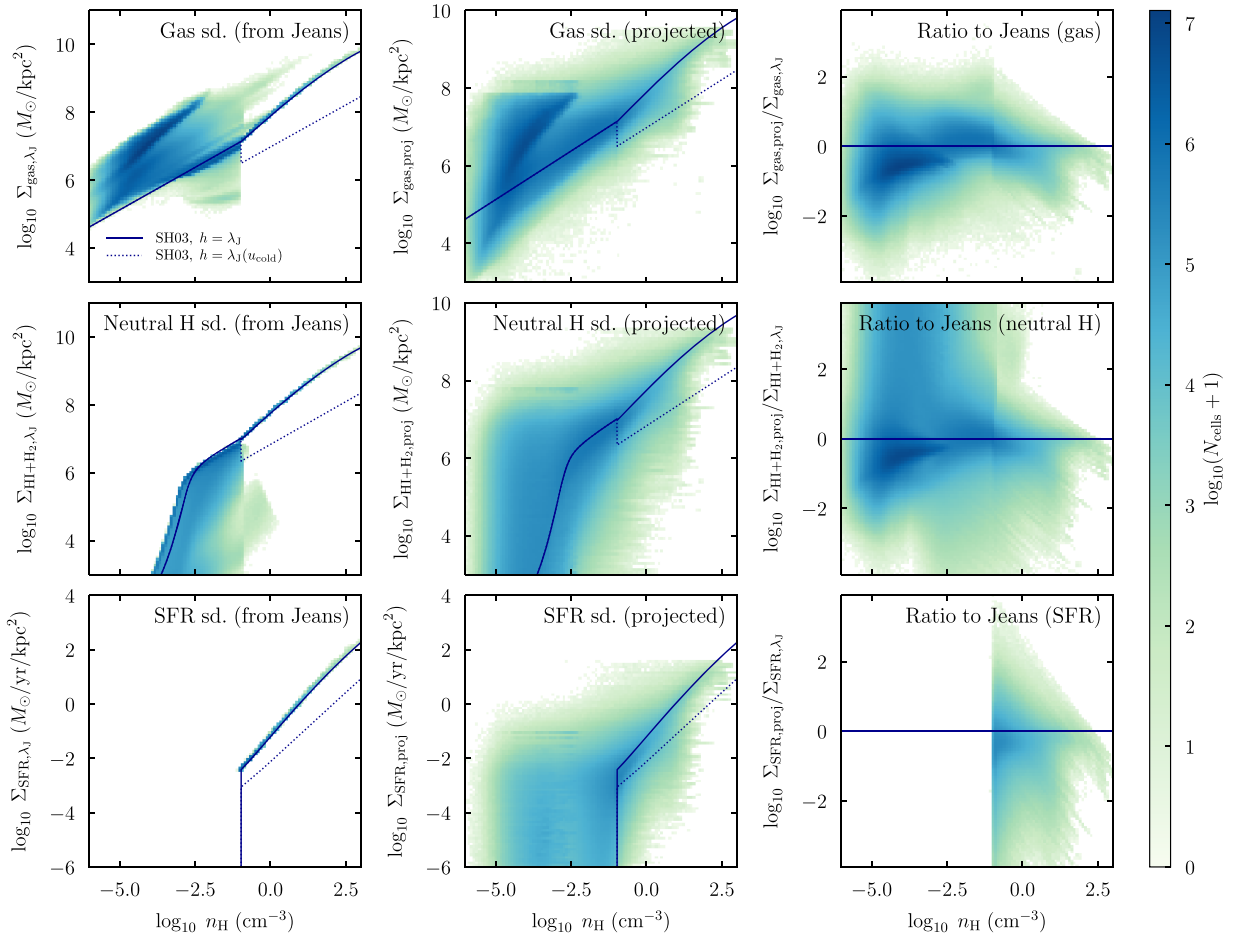


Figure 8. Test of the Jeans approximation by comparing it to projected surface densities. The histograms in each panel show the distribution of particles in TNG100 (see Figure 1). The left column shows the surface densities estimated from the Jeans approximation ($\Sigma_X = \lambda_J \rho_X$) for each gas cell, the center column shows the corresponding face-on projected surface density at the 2D position of each gas cell, and the right column shows the ratio between the two surface densities. The three rows refer to total gas density, neutral gas density, and the density of star formation. As in Figure 1, the blue lines show a calculation based on the SH03 equation of state, and the dotted lines show a version that uses the Jeans length of cold gas. The Jeans approximation leads to both a bias in the average surface density of gas as well as large cell-to-cell scatter. The approximation gets worse for the neutral and SFR surface densities.

clouds responsible for the Ly α forest. This assumption was further tested when applied to the HI column density distribution (Tepper-García et al. 2012; Rahmati et al. 2013a), but those results refer to random lines of sight through the universe. While Schaye & Dalla Vecchia (2008) argued that the Jeans length should also be a good proxy for the disk scale height, our simulated galaxies contain large amounts of gas outside their disks, and many objects are not disk-dominated in the first place. Moreover, the internal energy is governed by the effective equation of state of the SH03 ISM model at high densities (Section 2.2.1), rendering the resulting Jeans length somewhat artificial.

To clarify the impact of these caveats, we directly test the Jeans approximation by comparing its predictions to the actual projected surface densities. We caution that we specifically choose the face-on orientation of the galaxies for the projection and include all gas bound to the subhalo (according to SUBFIND). In the context of our HI/H₂ modeling, this comparison is appropriate because we are interested in the surface densities to which a gas cell contributes. These surface densities are, in turn, used to compute the molecular fraction of the gas cell. We emphasize that this comparison represents a special case that may not apply in other contexts.

The top row of Figure 8 shows histograms of the distribution of the same gas cells in TNG100 as in Figure 1, about 1.2 billion cells. Mass-weighted histograms would appear very similar because the range of gas cell masses is, by design, narrow in AREPO simulations. The panels show the surface density of gas derived using the Jeans approximation, the projected face-on surface density at the 2D position of the center of the gas cell, and the ratio of the two surface densities. The solid blue lines represent the Jeans length according to the effective equation of state used in IllustrisTNG (as in Figure 1). At high densities, the Jeans length (and thus the surface density) is a function of only density, whereas the projected surface densities exhibit scatter. On average, the Jeans approximation predicts surface densities 0.4 dex (a factor of 2.6) lower than the projection, with a scatter of 0.4 dex. The scatter increases toward low densities.

The second and third rows show the same comparison for the surface densities of neutral hydrogen and star formation, respectively. These quantities are more relevant for this work because they are used in computing the molecular fraction. The Jeans approximation performs worse than for the total gas density, with scatters of 1 and 0.6 dex, respectively. This change is not surprising: the Jeans length does not depend on the neutral fraction or SFR of a cell and is thus insensitive to

how these quantities evolve with density. The scatter shown in the right column of Figure 8 largely explains why the volumetric and projection-based versions of the H I/H₂ models tend to perform similarly on average but disagree for a given galaxy.

A particular problem arises for star-forming cells. Here, the internal energy (and temperature) are not physical because they correspond to an average of the hot ISM and star-forming gas (SH03). The Jeans length corresponding to this temperature does not correspond to a scale on which gas would actually collapse. One way to attempt to remedy this issue is to use the properties of the cold component, $u_{\text{cold}} = 1000$ K and ρ_{cold} (e.g., Marinacci et al. 2017a). The change in density is not significant because $\rho_{\text{cold}}/\rho \approx 1$ in the SH03 model, but the lower temperature reduces the Jeans length and thus the estimated surface densities. This assumption appears to provide a slightly worse match to the projected surface densities (dotted lines in the center column of Figure 8). Thus, we use the Jeans length based on the effective temperature.

We emphasize that our evaluation of the Jeans approximation corresponds to a particular scenario where we are interested in projected surface densities. Even for a perfectly spherical ball of gas in hydrostatic equilibrium, the geometry would lead to some scatter around the projected surface density. However, our test does demonstrate that the surface densities predicted using the Jeans approximation are off by many orders of magnitude for a significant fraction of gas cells. This issue is exacerbated at low densities and when estimating quantities other than total surface density.

4.2. Uncertainties in the UV Flux

We estimate the LW-band UV flux by assuming a constant escape fraction f_{esc} from the local star-forming region and optically thin propagation through the galaxy (Section 2.4 and Appendix A). Thus, f_{esc} is the only free parameter in our modeling and corresponds to an overall normalization of the UV field that impacts our results significantly (Appendix A.3). In this section, we discuss our assumptions and attempt to calculate rough estimates for f_{esc} .

Unfortunately, the escape fraction is hard to simulate because of the unknown small-scale density structure, the uncertain behavior of dust, and absorption below the resolution level. Let us illustrate these difficulties with a simple calculation. For example, we could assume that the radiation from young stars is attenuated by the dust contained in a certain column density of hydrogen N_{H} such that the optical depth becomes

$$\tau = \frac{1}{2} N_{\text{H}} \sigma_{\text{H},1000} \frac{Z}{Z_{\odot}}, \quad (7)$$

where $\sigma_{\text{H},1000}$ is the cross section for absorption. At solar metallicity, $\sigma_{\text{H},1000} = 1.4 \times 10^{-21} \text{cm}^2$ per atom (Draine 2003a, 2003b). We could estimate the column density in several ways: for example, we could assume that all stars are born in molecular clouds with a roughly constant column density of $N_{\text{H}} \approx 2 \times 10^{22} \text{cm}^{-2}$ (Larson 1981). However, the corresponding optical depth would be $\tau \approx 14 Z/Z_{\odot}$, resulting in a negligible escape fraction wherever $Z/Z_{\odot} \gtrsim 0.1$. Clearly, this scenario is not realized in nature; for example, we would not observe any H_{α} flux either. In reality, star-forming clouds

are optically thick initially but can be quickly dispersed by photoionization from young stars (e.g., Dale et al. 2012).

Alternatively, we could assume that the bulk attenuation happens as photons traverse the lower-density regions of the ISM and derive the optical depth from the disk scale height. This approximation leads to much higher escape fractions but is still almost entirely dictated by the metallicity. While the effect of metallicity is, of course, real, it does not take into account complex geometries where certain lines of sight might be entirely unobscured while clouds block other sight lines (e.g., Narayanan et al. 2018).

Instead, we estimate the escape fraction based on the solar neighborhood values for the UV field and the star-formation surface density. This idea follows the Lagos et al. (2015) method for estimating the UV field as $U_{\text{MW}} = \Sigma_{\text{SFR}}/\Sigma_{\text{SFR,local}}$, where $\Sigma_{\text{SFR,local}}$ is the observed star-formation surface density in the solar neighborhood. This number is rather uncertain. Lagos et al. (2015) suggest $10^{-3} M_{\odot} \text{yr}^{-1} \text{kpc}^{-2}$ following Bonatto & Bica (2011), whose study is based on star clusters (see also Lada & Lada 2003). This value is low compared to studies based on field stars, for example the rate of $3\text{--}7 \times 10^{-3} M_{\odot} \text{yr}^{-1} \text{kpc}^{-2}$ found by Miller & Scalo (1979). Moreover, the local UV field does not necessarily represent a galactic average because it is patchy and depends strongly on the distance to the nearest O star (e.g., Parravano et al. 2003; Hamden et al. 2013).

Despite these uncertainties, we undertake a rough calibration of f_{esc} by comparing the Σ_{SFR} and U_{MW} as predicted by our optically thin calculation. Although not shown in any figure, we find $U_{\text{MW}} \propto \Sigma_{\text{SFR}}$ at high Σ_{SFR} . Such a linear relation is expected, given that star-forming cells are the source of UV in our calculation. We assume an intermediate local SFR surface density of $\Sigma_{\text{SFR,local}} = 4 \times 10^{-3} M_{\odot} \text{yr}^{-1} \text{kpc}^{-2}$ (Robertson & Kravtsov 2008) and find that an escape fraction of $f_{\text{esc}} = 10\%$ normalizes the relation to the local UV field, as observed by Draine (1978).

Based on the above estimates, we can revisit the assumption that the galaxy is optically thin. It is clear that dense, high-metallicity regions will block LW radiation effectively. Our propagation method and escape fraction represent crude averages over a complex, clumpy ISM with irregular geometry. Thus, it is not clear that the escape fraction should be constant for different types of galaxies and as a function of redshift. For example, Whitaker et al. (2017) estimate the fraction of obscured UV light from star formation by comparing the UV and IR fluxes. They find that the obscured fraction strongly increases with galaxy stellar mass, indicating that the escape fraction might be lower for high-mass galaxies.

Encouragingly, we show in Appendix A.3 that even extreme escape fractions of 0.01 and 1 change the molecular fraction by a factor of only about three. Nevertheless, a better understanding of the UV field would provide the most important systematic improvement in our modeling.

4.3. Uncertainties in the Neutral Fraction

So far, we have discussed uncertainties that affect how we split neutral hydrogen into H I and H₂, but the underlying neutral fraction carries systematic uncertainties as well. Errors in the neutral fraction would not affect the atomic and molecular phases equally because f_{mol} nonlinearly depends on the neutral surface density.

One potential source of error could be ionizing radiation from young stars or AGNs. Rahmati et al. (2013b) found stellar radiation to significantly reduce the neutral fraction, particularly at high redshift. In their simulations, this reduction primarily affects gas with densities between $n_{\text{H}} \approx 0.01$ and 1 cm^{-3} (Figure 3 in Rahmati et al. 2013b). Much of this range lies above the star formation threshold of $n_{\text{H}} = 0.106 \text{ cm}^{-3}$, meaning that its neutral fraction is assumed to be about unity, a number that might be reduced in realistic radiative transfer simulations.

However, the effects of local ionizing radiation would be difficult to include either in the ionization state calculations in IllustrisTNG or in the SH03 ISM model because the escape fraction for ionizing radiation is highly uncertain (Gnedin et al. 2008). As ionizing photons are absorbed by HI, their mean free path is much shorter than that of LW photons, meaning that an optically thin propagation would be a very poor approximation. One could imagine a cell-by-cell correction based on the local SFR, but such a calculation is beyond the scope of this paper.

5. Conclusions

We have presented a significantly improved methodological framework to estimate the molecular fraction in galaxies in cosmological simulations. We have tested five models for the HI/H₂ transition, most of which rely on estimates of the neutral hydrogen surface density, the LW-band UV flux, and metallicity. We have applied our framework to the IllustrisTNG simulation suite, but we emphasize that it can be applied to any cosmological hydrodynamical simulation. Our main conclusions are as follows:

1. We have proposed a new method to model the molecular fraction based on projected maps rather than for individual volume elements. We find that the two methods agree reasonably well in their average predictions but that they diverge for individual galaxies and in the spatial distribution of molecular hydrogen. The projected modeling is slightly less sensitive to resolution effects than the volumetric modeling.
2. Projected modeling must be used when predicting the molecular fraction based on its observed correlation with the midplane pressure because the local thermal pressure in cosmological simulations is a poor approximation to the midplane pressure.
3. We have provided a rough estimate of the LW-band UV flux based on the optically thin propagation of radiation from young stars. Our calculation reproduces a scaling with star-formation surface density that matches the observed local values for an escape fraction of about 10%. The inferred molecular masses vary by a factor of three between extreme escape fractions of 1% and 100%.
4. The average molecular fraction differs by a factor of up to three between different mass resolution levels (Appendix B). The same difference is observed between the H₂ fractions in TNG100 and TNG300.
5. We critically investigate the Jeans length approximation to the column density of a gas cell. We find that, on average, it recovers projected surface densities to 0.4 dex (or a factor of 2.5), with equally large scatter.

Modeling the molecular fraction remains a difficult enterprise. This paper represents an improvement but certainly not a conclusive answer to this challenge. We emphasize that all postprocessing calculations similar to our modeling are subject to numerous important, systematic uncertainties, namely the fraction of neutral gas, the modeling of the LW flux, and simplifications due to geometry.

We are grateful to Yannick Bahé Shmuel Bialy, Blakesley Burkhardt, Claude-André Faucher-Giguère, Robert Feldmann, Jonathan Freundlich, Nick Gnedin, Ben Johnson, Rahul Kannan, Andrey Kravtsov, Mark Krumholz, Joel Leja, Philip Mansfield, Rüdiger Pakmor, Joop Schaye, and Sandro Tacchella for useful discussions. All cosmological calculations were performed using the Python toolkit COLOSSUS (Diemer 2017). B.D. and J.F. gratefully acknowledge the financial support of an Institute for Theory and Computation Fellowship. C.L. is funded by an Australian Research Council Discovery Early Career Researcher Award (DE150100618) and by the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), through project number CE170100013. P.T. is supported through Hubble Fellowship grant HST-HF2-51384.001-A awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS5-26555.

Appendix

In this appendix, we discuss technical aspects of our methodology, namely our method to compute the UV field, the convergence of our results with mass resolution, and mathematical details of the HI/H₂ models.

Appendix A The UV Field

We set the UV field to the maximum of two components, the cosmic UV background (A.1) and radiation from young stars (A.2). Our model relies on one free parameter, the escape fraction, whose value we have calibrated to 10% (Section 4.2). Most of the HI/H₂ models discussed in Section 2.5 work not with a physical value of the UV field but with U_{MW} , the intensity of the UV field in units of the field at the solar neighborhood. That number was constrained by Habing (1968) and Draine (1978), the latter finding a field 1.7 times stronger than the former (see, e.g., Sternberg et al. 2014). Most of the HI/H₂ models assume the Draine (1978) field, and we follow this convention. To convert a UV field in physical units to a value of U_{MW} , a number of ways have been suggested, including

1. comparing the flux at 1000 Å, a very good proxy for the destruction of H₂ because the absorption of photons is efficient only in a relatively narrow frequency range (e.g., Draine 2011; Gnedin & Kravtsov 2011);
2. comparing the total flux in the LW band (912–1108 Å or 11.2–13.6 eV), that is, the spectral region relevant to photodissociation; and
3. comparing the photoionization or photoheating rates.

We choose the first option: normalize U_{MW} at 1000 Å. At this wavelength, the Draine (1978) flux is $3.43 \times 10^{-8} \text{ photons s}^{-1} \text{ cm}^{-2} \text{ Hz}^{-1}$.

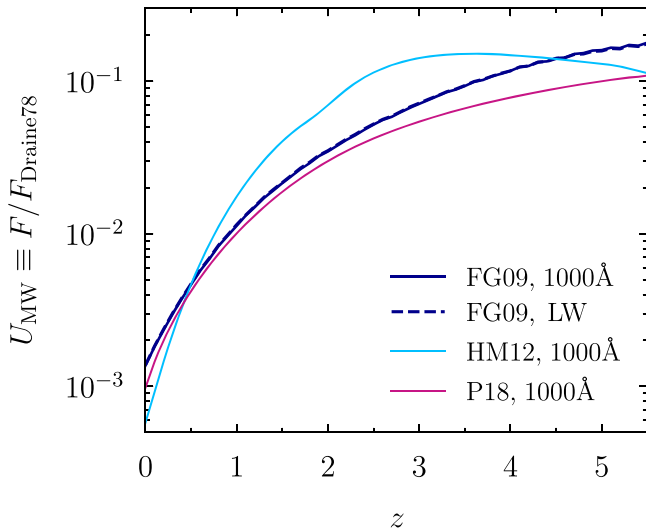


Figure 9. Comparison of methods to compute U_{MW} , the UV background field in units of the locally observed field according to Draine (1978). Based on the Faucher-Giguère et al. (2009) model, we compare the flux at 1000 Å (dark blue solid line) and the flux in the LW band (dark blue dashed line). The results are almost indistinguishable because the Faucher-Giguère et al. (2009) and Draine (1978) spectra resemble each other in shape in the relevant energy range. Both methods predict a low $U_{\text{MW}} \approx 10^{-3}$ at $z = 0$. For comparison, we also show U_{MW} assuming the Haardt & Madau (2012) and Puchwein et al. (2018) UV background models (light blue and pink lines).

A.1. UV Background

We adopt the time-dependent UV background of Faucher-Giguère et al. (2009, in the updated 2011 version), the same model used in the computation of the ionization state of gas in IllustrisTNG. Figure 9 demonstrates that the different ways to compute U_{MW} from the UV background spectrum give almost identical results because the Faucher-Giguère et al. (2009) and Draine (1978) spectra have similar shapes in the relevant region. Most importantly, U_{MW} is small at low redshift, about 10^{-3} at $z = 0$. We note that our computations give much lower values for U_{MW} than the computation used in Lagos et al. (2015), where the photoheating rate was used, $U_{\text{MW}} = \dot{q}_{\text{HI}}/2.2 \times 10^{-12} \text{ eV s}^{-1}$, where \dot{q}_{HI} is the HI heating rate from the Haardt & Madau (2001) tables. At $z = 0$, this calculation gives $U_{\text{MW}} = 0.072$, about 50 times higher than our value.

A.2. UV from Young Stars

In most of the ISM, the UV flux is dominated by the radiation from nearby young, massive stars (e.g., Parravano et al. 2003). However, if we apply this UV flux only to star-forming gas cells, we create a sharp break in U_{MW} at the (somewhat arbitrary) star formation threshold (Figure 1). Moreover, cells that do not form stars will receive only the UV background, allowing significant (and presumably unrealistic) H_2 production even at very low densities (dotted lines in bottom row of Figure 1). In real galaxies, even gas that is not star-forming itself experiences UV flux due to nearby star formation. To properly estimate the corresponding flux, we would need radiative transfer calculations that are far beyond the scope of this paper.

Instead, we assume that star-forming clouds emit UV at a rate proportional to their star formation and that a fraction of this radiation is absorbed in the cloud where it originated. We assume that the resulting radiation propagates through an

optically thin medium, in other words, that it is not attenuated any further and decreases with an inverse square law. We sum the contributions from all star-forming cells at the locations of the non-star-forming cells.

In detail, we simulate a continuously forming population of stars using STARBURST99 (Leitherer et al. 1999). We assume a Kroupa (2001) IMF because, out of the available IMF models, it is the most similar to the Chabrier (2003) IMF used in IllustrisTNG. After a few hundred megayears, the spectrum of a continuously forming population reaches equilibrium. As with the UV background, we compute U_{MW} based on the flux at 1000 Å. At this wavelength, the spectrum depends on metallicity only weakly, with less than 50% variation between the most extreme metallicities considered by STARBURST99. According to the STARBURST99 calculation, the flux due to an SFR of $1 M_{\odot} \text{ yr}^{-1}$ at a distance of 1 kpc is $3.3 \times 10^{-6} \text{ photons s}^{-1} \text{ cm}^{-2} \text{ Hz}^{-1}$. For comparison, this value corresponds to $U_{\text{MW}} \approx 100$ and is about 500 times stronger than the Faucher-Giguère et al. (2009) UV background at $z = 0$. Clearly, the contribution from young stars will dominate over the UV background as long as even a small fraction of the UV photons leave the star-forming regions and the galaxy.

We propagate the radiation using a fast Fourier transform (FFT) technique to convolve the SFR distribution with the $1/r^2$ Green’s function. We perform this calculation twice on two 128^3 grids, one grid that spans all gas cells bound to the galaxy and one higher-resolution grid that spans two gas half-mass radii. Without the high-resolution grid, the UV at the center of the galaxy can be underestimated significantly. As we cannot resolve the propagation of flux inside the grid cells, we assume a distance of $1/\sqrt{3}(r_{\text{cell}}/2)^3$ for the contributions from star formation inside a given grid cell. We linearly interpolate the resulting flux grid to the positions of the gas cells. In the case of projected modeling, we simply apply the FFT to the 2D map of star formation surface density.

Figures 1 and 2 show the results of this computation. For the cell-by-cell calculation, U_{MW} follows a more or less linear scaling with density, though with large scatter. As expected, our method spreads the UV radiation to lower-density cells, reducing their molecular fractions compared to a purely local UV field. This difference manifests itself in a spatial distribution that falls off more sharply with radius compared to a purely local UV field.

A.3. Convergence with Escape Fraction

The UV model described above has only one free parameter, the escape fraction, which determines the overall scaling of the field. In Section 4.2, we calibrated this number to 10% based on the solar neighborhood values of the UV field and star-formation surface density. However, these measurements are uncertain, and our UV propagation model is extremely simplistic as it does not take dust attenuation into account (Section 4.2). Thus, we need to investigate the effect that a much lower or higher UV flux would have on our results.

Figure 10 shows the effect on the H_2 mass fraction of varying the escape fraction by a factor of 100. We emphasize that an escape fraction of 100% is certainly unrealistic. Nevertheless, even with these extreme values, the average H_2 masses vary by a factor of about three or less. This conclusion holds regardless of the HI/H_2 model and whether the volumetric or projection version is used. Considering the H_2 mass functions rather than fractions leads to the same

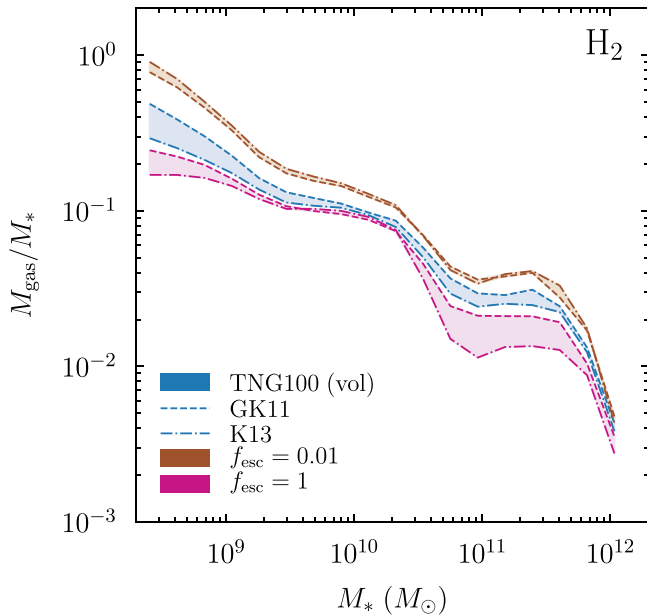


Figure 10. Impact of the UV escape fraction on the H_2 mass fraction for the fiducial escape fraction of 10% (blue), an extremely low escape fraction (1%, brown), and no absorption (100%, pink). The lines show the predictions of the volumetric GK11 and K13 models, but the same figures for other H I/ H_2 models or the projection-based versions look almost identical. Given that the probed escape fractions (and thus overall UV strengths) vary by a factor of 100, the variation of the H_2 mass by a factor of about three is rather moderate.

conclusions. Using a purely local UV field as in Lagos et al. (2015) corresponds to gas fractions similar to the $f_{\text{esc}} = 0.01$ case.

We conclude that the treatment of the UV field is likely to be our largest systematic uncertainty, but that the global properties of the galaxy population such as the H_2 mass fraction and mass function are reasonably robust to our modeling choices.

Appendix B Convergence with Mass Resolution

To assess how well converged the various gas fractions are with mass and force resolution, we have run our models on the lower resolutions of TNG100, TNG100-2, and TNG100-3 (see Table 1). Figure 11 shows the median ratios of neutral gas to stellar mass and H_2 mass to neutral gas mass. One of the main issues is readily apparent from the left panel: the median $M_{\text{H I}+\text{H}_2}/M_*$ ratio (which is independent of the H I/ H_2 models) differs by a factor of about three between resolution levels. This finding is not surprising given that TNG300 (which has a mass resolution comparable to TNG100-2) shows similar deviations in the neutral gas fraction (Figure 5). This difference will directly affect the $M_{\text{H I}}/M_*$ and M_{H_2}/M_* ratios, making them less valuable as a convergence check. Instead, we compare the $M_{\text{H I}}/M_{\text{H I}+\text{H}_2}$ and $M_{\text{H}_2}/M_{\text{H I}+\text{H}_2}$ ratios in the center and right panels of Figure 11.

The model predictions for the lowest resolution level, TNG100-3, differ by orders of magnitude and tend toward negligible H_2 fractions. Clearly, our modeling breaks down at such low resolutions, and we thus focus on the comparison between TNG100 and TNG100-2. Here, the convergence is strongly dependent on mass: at stellar masses between 10^9 and $10^{10} M_\odot$, TNG100-2 and TNG100 are almost perfectly converged, while they diverge by a factor of up to four at

higher stellar masses. We have repeated the same experiment while binning in halo mass instead of stellar mass and find the same results. The convergence of the projected models is slightly better, particularly when ignoring the L08 model, which is not plotted for the volumetric models.

Figure 12 shows the convergence of the H_2 mass function. The volumetric models again converge well at masses around $M_{\text{H}_2} \approx 10^9 M_\odot$ and differ by a factor of up to four in the mass function at the highest masses. The projected models perform better in this metric, with differences within a factor of about two at all masses.

In Section 3.2, we noted that the projected versions of the models exhibited slightly smaller differences between TNG100 and TNG300. Combined with the convergence tests in Figures 11 and 12, we conclude that the projected versions are slightly less sensitive to resolution effects. However, this statement depends on the H I/ H_2 model used and is difficult to generalize.

Appendix C Details of H I/ H_2 Models

In this appendix, we list the mathematical expressions for the H I/ H_2 models used in this paper and describe how their predictions are computed in detail. If symbols appear in multiple models but with different meanings, they are understood to be defined only within the respective section. Figure 13 compares the f_{mol} predictions of the models as a function of density.

C.1. Leroy et al. 2008 (L08)

As discussed in Section 2.5.1, we ignore the volumetric version of the L08 model because the assumption that the midplane pressure is approximated by the thermal pressure does not hold in our simulations. To compute the projected midplane pressure in projection as given in Equation (6), we need maps of stellar and gas surface density and velocity dispersion in the z direction. For the surface density, we follow L08 in setting Σ_{gas} to all neutral gas, that is, $\Sigma_{\text{H I}+\text{H}_2} + \Sigma_{\text{He}}$. Including ionized gas results in molecular masses about 0.1 dex higher.

In computing the velocity dispersion maps for gas and stars, we need to make certain choices because their exact meaning depends on the observations we wish to compare to. Here, we assume that the velocity dispersion is measured from a line width in each map pixel (e.g., using an integral field unit spectrograph). We compute a map of the mean z -velocity, find the pixel at the x - y position of each stellar particle or gas cell i , and subtract its value from the z -velocity of the particles, giving their relative velocity $v_{z,i,\text{rel}}$. For the gas velocity dispersion map, we weight particles by their neutral hydrogen mass. We also include the velocity dispersion due to the thermal energy of the gas cells,

$$\sigma_{\text{gas,pixel}} = \sqrt{\frac{\sum_i m_{\text{H I}+\text{H}_2,i} (v_{z,i,\text{rel}}^2 + \sigma_{p,i}^2)}{\sum_i m_{\text{H I}+\text{H}_2,i}}}, \quad (8)$$

where $\sigma_p^2 = P/\rho = u(\gamma - 1)$. We set a floor of 1 km s^{-1} on our velocity dispersion maps to avoid spurious results in regions with few particles or gas cells. We note that L08 assume a fixed $\sigma_{\text{gas}} = 11 \text{ km s}^{-1}$; we have verified that using this constant has a negligible impact on our results.

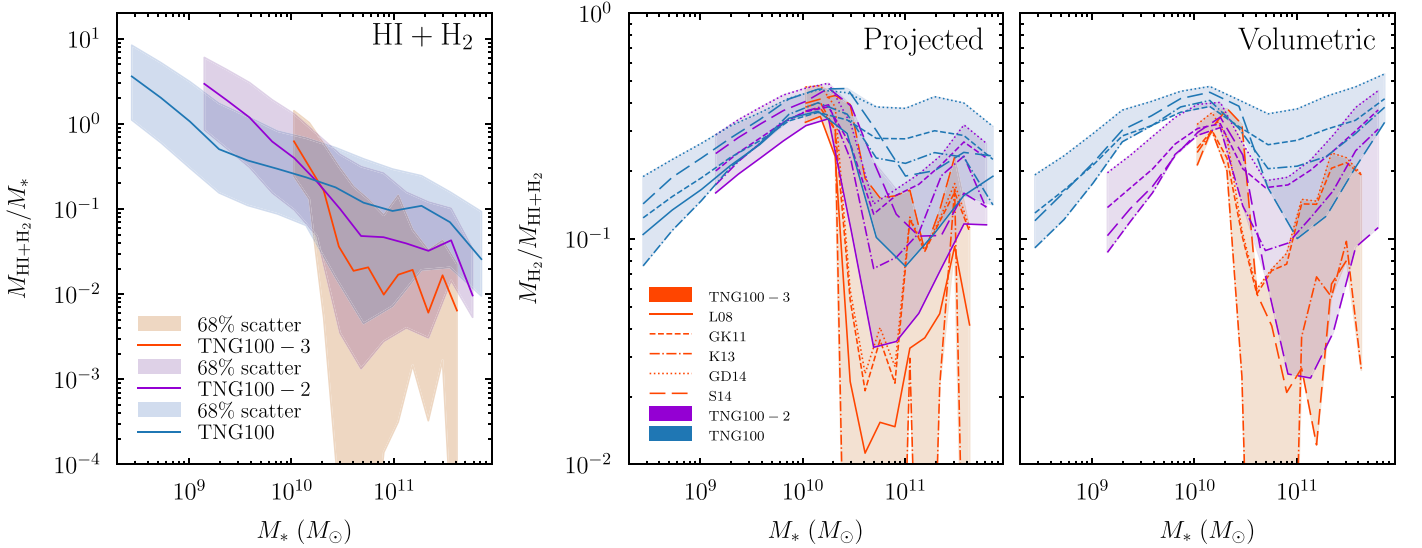


Figure 11. Convergence with mass resolution of the neutral gas fraction (left) and the $M_{\text{H}_2}/M_{\text{HI}+\text{H}_2}$ ratio (center and right) as a function of stellar mass. The blue lines and shaded areas show the TNG100 simulations, and the purple and red lines show the lower-resolution counterparts TNG100-2 and TNG100-3, respectively. The neutral gas fraction is converged only to a factor of about three at fixed mass between the resolution levels, meaning that the atomic and molecular fractions of stellar mass will be equally affected. Thus, instead of showing the fraction with respect to stellar mass, we consider the ratio of H_2 and neutral gas, both for the projection-based (center) and volumetric (right) versions of our HI/H_2 models. Comparing the TNG100 and TNG100-2 lines for each model, they are also converged to about a factor of two to three for most stellar masses. In TNG100-3, the convergence gets significantly worse.

Finally, we note that the stellar and gas velocity dispersions are likely subject to resolution effects, meaning that the second term in Equation (6) may be inaccurate. For the majority of map pixels, the first term strongly dominates, meaning that P_{eff} depends only on the gas surface density. Moreover, Figures 11 and 12 show that the projected L08 model converges at least as well as the other models with resolution. Nevertheless, a detailed study of the reliability of stellar and gas velocity dispersions would be desirable.

C.2. Gnedin & Kravtsov 2011 (GK11)

GK11 find that their simulation results for the molecular fraction are well described by

$$f_{\text{mol}} = \left(1 + \frac{\Sigma_{\text{c}}}{\Sigma_{\text{HI}+\text{H}_2}} \right)^{-2}, \quad (9)$$

where Σ_{c} is a critical threshold surface density that depends only on U_{MW} and D_{MW} ,

$$\Sigma_{\text{c}} = 2 \times 10^7 \frac{M_{\odot}}{\text{kpc}^2} \left(\frac{[\ln(1 + g D_{\text{MW}}^{3/7} (U_{\text{MW}}/15)^{4/7})]^{4/7}}{D_{\text{MW}} \sqrt{1 + U_{\text{MW}} D_{\text{MW}}^2}} \right), \quad (10)$$

where

$$g = \frac{1 + \alpha s + s^2}{1 + s}, \quad \alpha = 5 \frac{U_{\text{MW}}/2}{1 + (U_{\text{MW}}/2)^2}, \quad (11)$$

and

$$s = \frac{0.04}{D_{\text{MW}} + 1.5 \times 10^{-3} \times \ln(1 + (3U_{\text{MW}})^{1.7})}. \quad (12)$$

GK11 point out that this formula becomes inaccurate at very low $D_{\text{MW}} \lesssim 0.01$.

C.3. Gnedin & Draine 2014 (GD14)

The model is mathematically somewhat different from the GK11 model:

$$R_{\text{mol}} = \left(\frac{\Sigma_{\text{HI}+\text{H}_2}}{\Sigma_{\text{c}}} \right)^{\alpha}, \quad (13)$$

with

$$\alpha = 0.5 + \frac{1}{1 + \sqrt{U_{\text{MW}} D_{\text{MW}}^2 / 600}}. \quad (14)$$

As in the GK11 model, Σ_{c} is a critical threshold surface density:

$$\Sigma_{\text{c}} = 5 \times 10^7 \frac{M_{\odot}}{\text{kpc}^2} \left(\frac{\sqrt{0.001 + 0.1 U_{\text{MW}}}}{g(1 + 1.69 \sqrt{0.001 + 0.1 U_{\text{MW}}})} \right), \quad (15)$$

where the factor g depends on the dust-to-gas ratio,

$$g = \sqrt{D_{\text{MW}}^2 + D_*^2}, \quad (16)$$

and

$$D_* = 0.17 \frac{2 + S^5}{1 + S^5}. \quad (17)$$

The last equation includes a factor that depends on the resolution, $S \equiv L_{\text{cell}}/100 \text{ pc}$, where L_{cell} is the size of the computational cells. Cells in a moving-mesh code have no well-defined side length; we estimate it as $L_{\text{cell}} = V_{\text{cell}}^{1/3} = (m_{\text{cell}}/\rho_{\text{cell}})^{1/3}$. For the

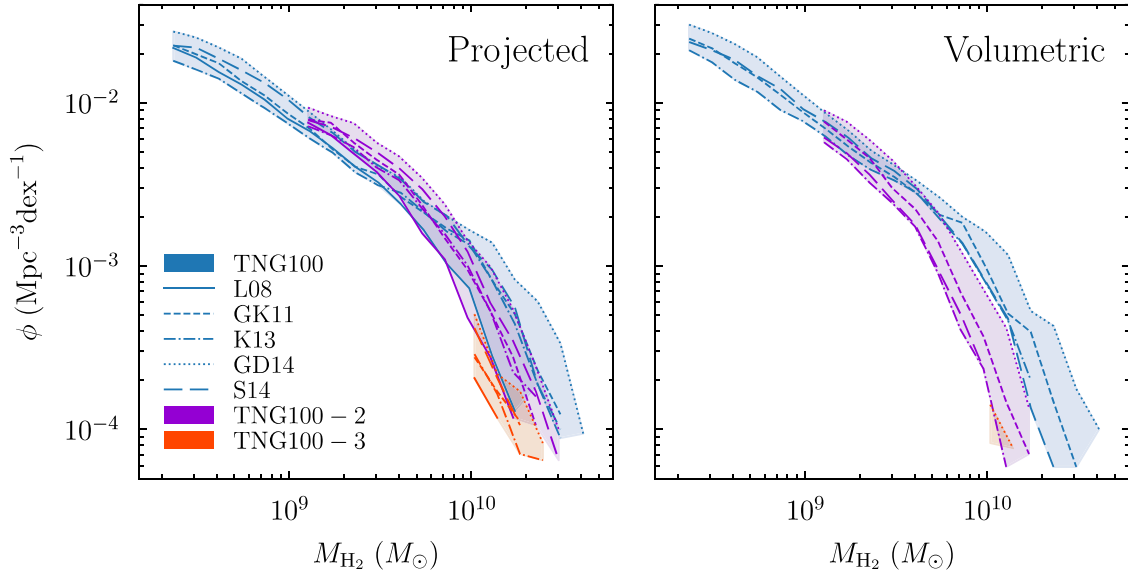


Figure 12. Same as Figure 11 but for the H_2 mass function predicted by the projected (left) and volumetric (right) versions of our $H\text{ I}/H_2$ models. The projection-based models are slightly better converged in this metric than the volumetric models.

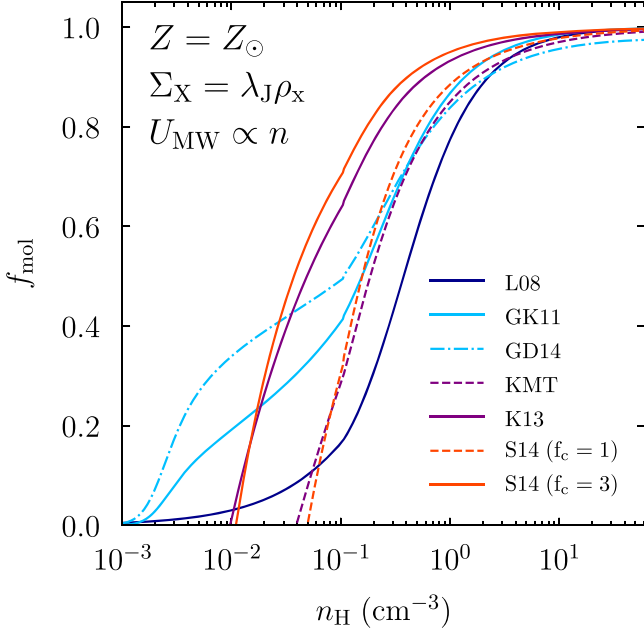


Figure 13. Molecular fraction as a function of density as predicted by the $H\text{ I}/H_2$ models used in this paper. This calculation depends on a number of variables that are computed similarly to the blue lines in Figure 1: we choose solar metallicity, $T = 10,000$ K below the star formation threshold, the **SH03** model’s equation of state above the star formation threshold, and the $U_{\text{MW}} \approx 10 n_{\text{H}}/(1\text{cm}^{-3})$ relation shown in Figure 1. In reality, these conditions vary across our galaxy sample. This figure corresponds to the volumetric implementation of the models in that we estimate surface densities from the density on the x axis and the Jeans length. The dark blue line shows the prediction of the volumetric **L08** model, which we have shown to be unphysical (Section 2.5.1). The light blue lines show that the **GK11** and **GD14** models agree well and that they predict nonzero molecular fractions out to lower densities than other models. The dashed lines show the original **KMT** and **S14** models, which agree well. The **K13** model predicts higher molecular fractions at low density due to the addition of the hydrostatic pressure floor (Appendix C.4). Applying a clumping factor of $f_c = 3$ to the **S14** model brings it into close agreement with **K13** (Appendix C.5).

projection-based version of this model, we use the side length of the map pixels to compute S . As expected, we find that the **GK11** and **GD14** models agree well. Figure 13 shows that they predict

slightly higher molecular fractions at low density compared to the **K13** and **S14** models.

C.4. Krumholz 2013 (K13)

The **K13** model is based on considerations of the interplay between H_2 recombination and UV dissociation in molecular clouds. McKee & Krumholz (2010) approximate the molecular fraction as

$$f_{\text{mol}} = \begin{cases} 1 - 3s/(4 + s) & \text{if } s < 2 \\ 0 & \text{if } s \geq 2 \end{cases}, \quad (18)$$

where

$$s \equiv \frac{\ln(1 + 0.6\chi + 0.01\chi^2)}{0.6\tau_c}. \quad (19)$$

The two critical variables in this model are χ , which quantifies the balance between UV dissociation and recombination, and τ_c , the optical depth of a cloud. In the **KMT** and **K13** models, the UV field can be determined self-consistently from the predicted SFR. In this context, the simulation has already computed an SFR, which underlies our estimate of the UV field. Given these input variables, **K13** provides an approximation for χ :

$$\chi \equiv 7.2U_{\text{MW}} \left(\frac{n_{\text{CNM}}}{10 \text{ cm}^{-3}} \right)^{-1}, \quad (20)$$

where the density appears in the denominator because a higher density leads to a higher recombination rate. However, this density is not given by the density in our simulated gas cells because their density represents an average over a large region, whereas n_{CNM} above refers to the typical density of the cold neutral medium (CNM), about 100 K. Thus, we use the **K13** estimate of n_{CNM} , which is based on the key insight that there is a relatively narrow range of pressures where a two-phase ISM can exist in equilibrium (Wolfire et al. 2003). The pressure corresponding to this density range is modified by the UV field,

which sets the overall cooling rate. Wolfire et al. (2003) give an analytical expression for the minimum density in two-phase equilibrium. KMT postulate that the equilibrium density of the CNM, $n_{\text{CNM},2\text{p}}$, should be three times that minimum density and, after certain simplifying assumptions, write

$$n_{\text{CNM},2\text{p}} = 3n_{\text{CNM},\text{min}} = 23U_{\text{MW}} \frac{4.1}{1 + 3.1D_{\text{MW}}^{0.365}} \text{cm}^{-3}. \quad (21)$$

The problem with this expression is that it predicts a vanishing density and pressure in regions with a vanishing UV field, which is unrealistic because hydrostatic equilibrium demands a finite pressure. Thus, a minimum pressure is set at low densities according to the hydrostatic equilibrium,

$$n_{\text{CNM}} = \max(n_{\text{CNM},2\text{p}}, n_{\text{CNM},\text{hydro}}), \quad (22)$$

with

$$n_{\text{CNM},\text{hydro}} = \frac{P_{\text{th}}}{1.1k_{\text{B}}T_{\text{CNM},\text{max}}}, \quad (23)$$

where the factor of 1.1 accounts for helium. The temperature of $T_{\text{CNM},\text{max}} = 243$ K represents the maximum temperature of the CNM according to Wolfire et al. (2003), meaning that $n_{\text{CNM},\text{hydro}}$ represents a lower limit. As with the L08 model, we might be tempted to interpret P_{th} as the thermal pressure in the simulated gas cells, but the high-pressure region at the midplane of the disk is not resolved, meaning that the cells' thermal pressure would underestimate the hydrostatic pressure. Instead, we use K13's prescription, which is based on the hydrostatic equilibrium model of Ostriker et al. (2010):

$$P_{\text{th}} = \frac{\pi G \Sigma_{\text{H I}}^2}{4\alpha} (1 + 2R_{\text{mol}} + \sqrt{(1 + 2R_{\text{mol}})^2 + \frac{32\zeta_{\text{d}}\alpha\tilde{f}_{\text{w}}c_{\text{w}}^2\rho_{\text{sd}}}{\pi G \Sigma_{\text{H I}}^2}}), \quad (24)$$

where $\alpha = 5$ reduces the thermal pressure to account for turbulent, magnetic, and cosmic-ray pressure, $\zeta_{\text{d}} = 0.33$ is a geometric factor, $c_{\text{w}} = 8 \text{ km s}^{-1}$ is the sound speed in the warm neutral medium, and $\tilde{f}_{\text{w}} = 0.5$ is the ratio of the thermal velocity dispersion to c_{w} .

Up to this point, we have not used the density or surface density in our simulation. These quantities enter via the ρ_{sd} parameter, which represents the midplane density of stars and dark matter. We could set ρ_{sd} to a constant (e.g., Lagos et al. 2015), but that would mean neglecting the spatial dependence of the hydrostatic floor: we expect this term to be larger at the dense centers of galaxies and to become insignificant in the outskirts. Moreover, the values of ρ_{sd} quoted in the observational literature vary between 10^{-5} and $0.1 M_{\odot} \text{ pc}^{-3}$, presumably due to the different types of galaxies observed and the radial variation of the density (Holmberg & Flynn 2000; Bruzzone et al. 2015; Watts et al. 2018). The value of ρ_{sd} significantly influences the predicted molecular fractions, with order-of-magnitude changes over the possible range of ρ_{sd} . Thus, we attempt to approximate ρ_{sd} from the simulation data. For the cell-by-cell modeling, we subtract the density of each gas cell from the total density of matter surrounding the cell (computed using an SPH kernel over the 64 nearest neighbors). In some cells, this number becomes negative and

we set it to zero. In the case of projection-based modeling, we compute the stellar and dark matter densities separately: $\rho_{\text{sd}} = \rho_{*} + \rho_{\text{dm}}$. We base the stellar density on a simple disk model, $\rho_{*} \approx \Sigma_{*}/(2h_{*})$, with a scale height of

$$h_{*} = \frac{\sigma_{*}^2}{\pi G (\Sigma_{*} + \Sigma_{\text{H I}+\text{H}_2})}, \quad (25)$$

where σ_{*} is computed as discussed in Appendix C.1 (though without a contribution from thermal pressure). We approximate the dark matter density in the disk as a Navarro et al. (1997) profile with a concentration given by the relation of Diemer & Kravtsov (2015). For host halos (as defined by the halo finder), we take $M_{200\text{c}}$ from the halo catalog. For subhalos, we estimate it as the mass in dark matter that is bound to the galaxy. This approximation is far from exact, but varying it makes no appreciable difference to the predicted f_{mol} because the dark matter tends to be a subdominant contribution to ρ_{sd} in regions of appreciable molecular density.

Based on this modeling, we find values of ρ_{sd} that vary between zero and $\sim 1 M_{\odot} \text{ pc}^{-3}$. The exact range depends strongly on the properties of a given galaxy. While setting ρ_{sd} to a constant leads to similar average molecular masses, we find that modeling a spatially dependent ρ_{sd} changes the spatial distribution of H_2 significantly.

The second important factor in Equation (19) is the optical depth of dust:

$$\tau_{\text{c}} \equiv 0.066f_{\text{c}}D_{\text{MW}} \frac{\Sigma_{\text{H I}+\text{H}_2}}{M_{\odot} \text{ pc}^{-2}}. \quad (26)$$

Here, f_{c} is a clumping factor that accounts for the scale on which the surface density is measured; the actual surface density of star-forming clouds is much higher than the surface density that occurs in the large simulation cells. On scales of 100 pc or less, we expect $f_{\text{c}} \approx 1$. For kiloparsec scales, K13 suggests $f_{\text{c}} \approx 5$, which is the value we adopt. We note that KMT give an alternative formula for χ , which does not explicitly depend on the UV field. This expression contains another clumping factor of 30 that accounts for higher H_2 formation rates in dense clouds that are not resolved in the simulation. Conversely, the surface density measured in simulations may represent the column density from multiple clouds and would thus overestimate the column density of individual clouds. Thus, the clumping factor accounts for multiple mismatches between theory and simulation and should be seen as a free parameter.

The expression for P_{th} (Equation (24)) depends on R_{mol} , which is the quantity we wish to compute, meaning that we need to iterate to find the solution. If implemented naively, the R_{mol} can take more than 50 iterations to converge to an accuracy of 10^{-3} because the solution tends to oscillate for certain values. We find a great speed-up to fewer than 10 iterations when initializing the equation with $f_{\text{mol}} = 0.5$ and averaging the new and old solution with a weight of 0.7 toward the new solution. Compared to the solution where the R_{mol} terms are dropped from Equation (24), iterating makes a difference of more than 20% for some values of the input parameters.

For a comparison between the KMT and K13 models, Figure 13 shows their predictions for a specific set of physical parameters. The addition of the hydrostatic pressure floor and the different clumping factors lead to higher molecular fractions at low density in the K13 model.

C.5. Sternberg et al. 2014 (S14)

The S14 model is based on a physical picture similar to that in the KMT model. It predicts the total column density of atomic hydrogen in a slab irradiated by UV on both sides:

$$N_{\text{HI}} = 5.3 \times 10^{20} \text{cm}^{-2} \left(\frac{1}{Z/Z_{\odot}} \ln \left[\frac{\alpha G / f_c}{2} + 1 \right] \right). \quad (27)$$

Like χ in the KMT model, the αG parameter quantifies the ratio of H_2 photodissociation and recombination (Sternberg 1988; Bialy & Sternberg 2016; Bialy et al. 2017):

$$\alpha G = \frac{D_0 G}{R n_{\text{HI}+\text{H}_2}}, \quad (28)$$

where $D_0 = 5.8 \times 10^{-11} U_{\text{MWS}}^{-1}$ is the photodissociation rate, $R = 3 \times 10^{-17} \text{cm}^{-3} \text{s}^{-1}$ is the recombination rate, and the self-shielding factor is

$$G = 3 \times 10^{-5} Z / Z_{\odot} \left(\frac{9.9}{1 + 8.9 Z / Z_{\odot}} \right). \quad (29)$$

The difference between αG and χ lies in the treatment of the LW absorption by H_2 molecules as opposed to dust (Section 4.1 in S14). For $Z = Z_{\odot}$, we reproduce Equation (3) in Bialy et al. (2017),

$$\alpha G \approx 60 U_{\text{MW}} \left(\frac{n_{\text{HI}+\text{H}_2}}{\text{cm}^{-3}} \right)^{-1}, \quad (30)$$

highlighting that αG depends only on the volume density of neutral hydrogen, the radiation field, and metallicity. We translate the atomic column density from Equation (27) into a molecular fraction as

$$f_{\text{mol}} = \begin{cases} 0 & \text{if } N_{\text{HI}} > N_{\text{HI}+\text{H}_2} \\ 1 - N_{\text{HI}} / N_{\text{HI}+\text{H}_2} & \text{if } N_{\text{HI}} \leq N_{\text{HI}+\text{H}_2} \end{cases}. \quad (31)$$

For the volumetric version of the model, we once again use the column density of neutral gas derived from the Jeans approximation. However, as discussed in Appendix C.4, we cannot set $n_{\text{HI}+\text{H}_2}$ to the volume density in the simulation cell because it refers to the density of molecular clouds, orders of magnitude higher than typical densities in the simulation. We revert to the same strategy as the K13 model and set $n_{\text{HI}+\text{H}_2} = n_{\text{CNM},2\text{p}}$ (Equation (21)). With this modification, the S14 model agrees almost exactly with the original KMT model (Figure 13). Without the clumping factor f_c , however, this model predicts molecular fractions an order of magnitude lower than all other models due to the mismatch between the simulated surface densities and those of actual molecular clouds. Thus, we have introduced a clumping factor of $f_c = 3$ into the S14 model in Equation (27). This factor brings the model into good overall agreement with K13, as shown in Section 3.

ORCID iDs

Benedikt Diemer  <https://orcid.org/0000-0001-9568-7287>
Lars Hernquist  <https://orcid.org/0000-0001-6950-1629>
Annalisa Pillepich  <https://orcid.org/0000-0003-1065-9274>

References

- Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, *ARA&A*, 47, 481
Bahé, Y. M., Crain, R. A., Kauffmann, G., et al. 2016, *MNRAS*, 456, 1115
Bekki, K. 2015, *MNRAS*, 449, 1625
Bett, P., Eke, V., Frenk, C. S., Jenkins, A., & Okamoto, T. 2010, *MNRAS*, 404, 1137
Bialy, S., Burkhardt, B., & Sternberg, A. 2017, *ApJ*, 843, 92
Bialy, S., & Sternberg, A. 2016, *ApJ*, 822, 83
Bigiel, F., Leroy, A., Walter, F., et al. 2008, *AJ*, 136, 2846
Bigiel, F., Leroy, A. K., Walter, F., et al. 2011, *ApJL*, 730, L13
Black, J. H., & Dalgarno, A. 1976, *ApJ*, 203, 132
Blitz, L., & Rosolowsky, E. 2004, *ApJL*, 612, L29
Blitz, L., & Rosolowsky, E. 2006, *ApJ*, 650, 933
Bolatto, A. D., Wolfire, M., & Leroy, A. K. 2013, *ARA&A*, 51, 207
Bolatto, A. D., Wong, T., Utomo, D., et al. 2017, *ApJ*, 846, 159
Bonatto, C., & Bica, E. 2011, *MNRAS*, 415, 2827
Browning, M. K., Tumlinson, J., & Shull, J. M. 2003, *ApJ*, 582, 810
Bruzzeze, S. M., Meurer, G. R., Lagos, C. D. P., et al. 2015, *MNRAS*, 447, 618
Chabrier, G. 2003, *PASP*, 115, 763
Christensen, C., Quinn, T., Governato, F., et al. 2012, *MNRAS*, 425, 3058
Crain, R. A., Bahé, Y. M., Lagos, C. D. P., et al. 2017, *MNRAS*, 464, 4204
Dale, J. E., Ercolano, B., & Bonnell, I. A. 2012, *MNRAS*, 424, 377
Davé, R., Thompson, R., & Hopkins, P. F. 2016, *MNRAS*, 462, 3265
Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, *ApJ*, 292, 371
Diemer, B. 2017, arXiv:1712.04512
Diemer, B., & Kravtsov, A. V. 2015, *ApJ*, 799, 108
Dolag, K., Borgani, S., Murante, G., & Springel, V. 2009, *MNRAS*, 399, 497
Draine, B. T. 1978, *ApJS*, 36, 595
Draine, B. T. 2003a, *ARA&A*, 41, 241
Draine, B. T. 2003b, *ApJ*, 598, 1017
Draine, B. T. 2011, *Physics of the Interstellar and Intergalactic Medium* (Princeton: Princeton Univ. Press)
Draine, B. T., & Bertoldi, F. 1996, *ApJ*, 468, 269
Dubois, Y., Pichon, C., Welker, C., et al. 2014, *MNRAS*, 444, 1453
Duffy, A. R., Kay, S. T., Battye, R. A., et al. 2012, *MNRAS*, 420, 2799
Elmegreen, B. G. 1989, *ApJ*, 338, 178
Elmegreen, B. G. 1993, *ApJ*, 411, 170
Faucher-Giguère, C.-A., Lidz, A., Zaldarriaga, M., & Hernquist, L. 2009, *ApJ*, 703, 1416
Ferland, G. J., Korista, K. T., Verner, D. A., et al. 1998, *PASP*, 110, 761
Forbes, J., Krumholz, M., & Burkert, A. 2012, *ApJ*, 754, 48
Forbes, J. C., Krumholz, M. R., Burkert, A., & Dekel, A. 2014, *MNRAS*, 438, 1552
Fu, J., Guo, Q., Kauffmann, G., & Krumholz, M. R. 2010, *MNRAS*, 409, 515
Genel, S., Vogelsberger, M., Springel, V., et al. 2014, *MNRAS*, 445, 175
Glover, S. C. O., & Clark, P. C. 2012, *MNRAS*, 421, 9
Gnedin, N. Y., & Draine, B. T. 2014, *ApJ*, 795, 37
Gnedin, N. Y., & Kravtsov, A. V. 2011, *ApJ*, 728, 88
Gnedin, N. Y., Kravtsov, A. V., & Chen, H.-W. 2008, *ApJ*, 672, 765
Gnedin, N. Y., Tassis, K., & Kravtsov, A. V. 2009, *ApJ*, 697, 55
Grand, R. J. J., Gómez, F. A., Marinacci, F., et al. 2017, *MNRAS*, 467, 179
Haardt, F., & Madau, P. 2001, in *Clusters of Galaxies and the High Redshift Universe Observed in X-rays*, ed. D. M. Neumann & J. T. V. Tran, 64
Haardt, F., & Madau, P. 2012, *ApJ*, 746, 125
Habing, H. J. 1968, *BAN*, 19, 421
Hamden, E. T., Schiminovich, D., & Seibert, M. 2013, *ApJ*, 779, 180
Hodge, J. A., Riechers, D., Decarli, R., et al. 2015, *ApJL*, 798, L18
Holmberg, J., & Flynn, C. 2000, *MNRAS*, 313, 209
Hopkins, P. F., & Lee, H. 2016, *MNRAS*, 456, 4174
Hu, C.-Y., Naab, T., Glover, S. C. O., Walch, S., & Clark, P. C. 2017, *MNRAS*, 471, 2151
Hu, C.-Y., Naab, T., Walch, S., Glover, S. C. O., & Clark, P. C. 2016, *MNRAS*, 458, 3528
Jura, M. 1975a, *ApJ*, 197, 581
Jura, M. 1975b, *ApJ*, 197, 575
Katz, H., Kimm, T., Sijacki, D., & Haehnelt, M. G. 2017, *MNRAS*, 468, 4831
Kennicutt, R. C., Jr. 1998, *ApJ*, 498, 541
Kroupa, P. 2001, *MNRAS*, 322, 231
Krumholz, M. R. 2013, *MNRAS*, 436, 2747
Krumholz, M. R., & Dekel, A. 2012, *ApJ*, 753, 16
Krumholz, M. R., & Gnedin, N. Y. 2011, *ApJ*, 729, 36
Krumholz, M. R., Leroy, A. K., & McKee, C. F. 2011, *ApJ*, 731, 25
Krumholz, M. R., McKee, C. F., & Tumlinson, J. 2009a, *ApJ*, 699, 850
Krumholz, M. R., McKee, C. F., & Tumlinson, J. 2009b, *ApJ*, 693, 216

- Kuhlen, M., Krumholz, M. R., Madau, P., Smith, B. D., & Wise, J. 2012, *ApJ*, **749**, 36
- Lada, C. J., & Lada, E. A. 2003, *ARA&A*, **41**, 57
- Lagos, C. D. P., Baugh, C. M., Lacey, C. G., et al. 2011a, *MNRAS*, **418**, 1649
- Lagos, C. D. P., Lacey, C. G., Baugh, C. M., Bower, R. G., & Benson, A. J. 2011b, *MNRAS*, **416**, 1566
- Lagos, C. d. P., Crain, R. A., Schaye, J., et al. 2015, *MNRAS*, **452**, 3815
- Lagos, C. d. P., Theuns, T., Schaye, J., et al. 2016, *MNRAS*, **459**, 2632
- Larson, R. B. 1981, *MNRAS*, **194**, 809
- Leitherer, C., Schaerer, D., Goldader, J. D., et al. 1999, *ApJS*, **123**, 3
- Leroy, A. K., Walter, F., Brinks, E., et al. 2008, *AJ*, **136**, 2782
- Leroy, A. K., Walter, F., Sandstrom, K., et al. 2013, *AJ*, **146**, 19
- Leroy, A. K., Schinnerer, E., Hughes, A., et al. 2017, *ApJ*, **846**, 71
- Marasco, A., Crain, R. A., Schaye, J., et al. 2016, *MNRAS*, **461**, 2630
- Marinacci, F., Grand, R. J. J., Pakmor, R., et al. 2017a, *MNRAS*, **466**, 3859
- Marinacci, F., Vogelsberger, M., Pakmor, R., et al. 2017b, arXiv:1707.03396
- McKee, C. F., & Krumholz, M. R. 2010, *ApJ*, **709**, 308
- McKinnon, R., Torrey, P., & Vogelsberger, M. 2016, *MNRAS*, **457**, 3775
- McKinnon, R., Torrey, P., Vogelsberger, M., Hayward, C. C., & Marinacci, F. 2017, *MNRAS*, **468**, 1505
- McKinnon, R., Vogelsberger, M., Torrey, P., Marinacci, F., & Kannan, R. 2018, *MNRAS*, **478**, 2851
- Miller, G. E., & Scalo, J. M. 1979, *ApJS*, **41**, 513
- Naiman, J. P., Pillepich, A., Springel, V., et al. 2018, *MNRAS*, arXiv:1707.03401
- Narayanan, D., Conroy, C., Dave, R., Johnson, B., & Popping, G. 2018, arXiv:1805.06905
- Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, *ApJ*, **490**, 493
- Nelson, D., Pillepich, A., Springel, V., et al. 2018, *MNRAS*, **475**, 624
- Nickerson, S., Teyssier, R., & Rosdahl, J. 2018, *MNRAS*, **477**, 1206
- Obreschkow, D., Klöckner, H.-R., Heywood, I., Levrier, F., & Rawlings, S. 2009, *ApJ*, **703**, 1890
- Ostriker, E. C., McKee, C. F., & Leroy, A. K. 2010, *ApJ*, **721**, 975
- Parravano, A., Hollenbach, D. J., & McKee, C. F. 2003, *ApJ*, **584**, 797
- Pelupessy, F. I., Papadopoulos, P. P., & van der Werf, P. 2006, *ApJ*, **645**, 1024
- Pillepich, A., Springel, V., Nelson, D., et al. 2018a, *MNRAS*, **473**, 4077
- Pillepich, A., Nelson, D., Hernquist, L., et al. 2018b, *MNRAS*, **475**, 648
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, *A&A*, **594**, A13
- Popping, G., Behroozi, P. S., & Peebles, M. S. 2015, *MNRAS*, **449**, 477
- Popping, G., Somerville, R. S., & Galametz, M. 2017, *MNRAS*, **471**, 3152
- Popping, G., Somerville, R. S., & Trager, S. C. 2014, *MNRAS*, **442**, 2398
- Puchwein, E., Haardt, F., Haehnelt, M. G., & Madau, P. 2018, arXiv:1801.04931
- Rahmati, A., Pawlik, A. H., Raičević, M., & Schaye, J. 2013a, *MNRAS*, **430**, 2427
- Rahmati, A., Schaye, J., Bower, R. G., et al. 2015, *MNRAS*, **452**, 2034
- Rahmati, A., Schaye, J., Pawlik, A. H., & Raičević, M. 2013b, *MNRAS*, **431**, 2261
- Rémy-Ruyer, A., Madden, S. C., Galliano, F., et al. 2014, *A&A*, **563**, A31
- Robertson, B. E., & Kravtsov, A. V. 2008, *ApJ*, **680**, 1083
- Rosdahl, J., Schaye, J., Teyssier, R., & Agertz, O. 2015, *MNRAS*, **451**, 34
- Saintonge, A., Kauffmann, G., Wang, J., et al. 2011, *MNRAS*, **415**, 61
- Salpeter, E. E. 1955, *ApJ*, **121**, 161
- Schaye, J. 2001, *ApJ*, **559**, 507
- Schaye, J., & Dalla Vecchia, C. 2008, *MNRAS*, **383**, 1210
- Schaye, J., Dalla Vecchia, C., Booth, C. M., et al. 2010, *MNRAS*, **402**, 1536
- Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, *MNRAS*, **446**, 521
- Schmidt, M. 1959, *ApJ*, **129**, 243
- Schruba, A., Bialy, S., & Sternberg, A. 2018, *ApJ*, **862**, 110
- Schruba, A., Leroy, A. K., Walter, F., et al. 2011, *AJ*, **142**, 37
- Shull, J. M. 1978, *ApJ*, **219**, 877
- Sijacki, D., Vogelsberger, M., Genel, S., et al. 2015, *MNRAS*, **452**, 575
- Somerville, R. S., & Davé, R. 2015, *ARA&A*, **53**, 51
- Spitzer, L. J., & Zweibel, E. G. 1974, *ApJL*, **191**, L127
- Springel, V. 2010, *MNRAS*, **401**, 791
- Springel, V., & Hernquist, L. 2003, *MNRAS*, **339**, 289
- Springel, V., White, S. D. M., Tormen, G., & Kauffmann, G. 2001, *MNRAS*, **328**, 726
- Springel, V., Pakmor, R., Pillepich, A., et al. 2018, *MNRAS*, **475**, 676
- Sternberg, A. 1988, *ApJ*, **332**, 400
- Sternberg, A., & Dalgarno, A. 1989, *ApJ*, **338**, 197
- Sternberg, A., Le Petit, F., Roueff, E., & Le Bourlot, J. 2014, *ApJ*, **790**, 10
- Tepper-García, T., Richter, P., Schaye, J., et al. 2012, *MNRAS*, **425**, 1640
- Thompson, R., Nagamine, K., Jaacks, J., & Choi, J.-H. 2014, *ApJ*, **780**, 145
- Torrey, P., Vogelsberger, M., Genel, S., et al. 2014, *MNRAS*, **438**, 1985
- Torrey, P., Vogelsberger, M., Hernquist, L., et al. 2018, *MNRAS*, **477**, L16
- van Dishoeck, E. F., & Black, J. H. 1986, *ApJS*, **62**, 109
- Villaescusa-Navarro, F., Genel, S., Castorina, E., et al. 2018, arXiv:1804.09180
- Vogelsberger, M., Genel, S., Sijacki, D., et al. 2013, *MNRAS*, **436**, 3031
- Vogelsberger, M., Genel, S., Springel, V., et al. 2014a, *MNRAS*, **444**, 1518
- Vogelsberger, M., Genel, S., Springel, V., et al. 2014b, *Natur*, **509**, 177
- Watts, A. B., Meurer, G. R., Lagos, C. D. P., et al. 2018, *MNRAS*, **477**, 5554
- Weinberger, R., Springel, V., Hernquist, L., et al. 2017, *MNRAS*, **465**, 3291
- Whitaker, K. E., Pope, A., Cybulski, R., et al. 2017, *ApJ*, **850**, 208
- Wolfire, M. G., McKee, C. F., Hollenbach, D., & Tielens, A. G. G. M. 2003, *ApJ*, **587**, 278
- Wong, T., & Blitz, L. 2002, *ApJ*, **569**, 157