



September 17, 2020

Desika Narayanan^{1,2}, Eric F. Jiménez-Andrade³, Eric Murphy³, Qi Li¹, and Viviana Rosero⁴

¹ Department of Astronomy, University of Florida, 211 Bryant Space Science Center, P.O. Box 112055, Gainesville, FL 32611, USA

² Cosmic Dawn Center (DAWN), Niels Bohr Institute, University of Copenhagen, DK-2100 Copenhagen Ø; DTU-Space, Technical University of Denmark, DK-2800 Kgs. Lyngby

³ National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, VA 22903, USA

⁴ National Radio Astronomy Observatory, 1003 Lopezville Rd., Socorro, NM, 87801, USA

Abstract In this memo, we assess the ability of the Next Generation Very Large Array (ngVLA) in detecting CO across a range of transitions in high-redshift galaxies, and compare this to the current capabilities of the Atacama Large Millimeter Array (ALMA). We do this by combining thermo-radiative-chemical equilibrium models of CO line emission from giant clouds to cosmological hydrodynamic galaxy evolution simulations. We focus here on a fiducial redshift $z = 2$. We find that the ngVLA will be able to faithfully recover the majority of the CO emission from the J=1-0→3-2 states of $z = 2$ galaxies. This recovery of ground state emission is important as galaxies become increasingly compact at higher-lying transitions, and the highest-lying lines (e.g. 6-5) that are currently probed by ALMA may miss up to half of the underlying H_2 gas mass. High-resolution observations with the ngVLA will be able to detect substructure, and distinguish between physically associated clumps, and false counterparts at unrelated redshifts.

1 Introduction and Goals

Stars typically form in giant clouds comprised of relatively cold ($T \sim 10$ K) molecular hydrogen gas (H_2). While direct H_2 emission is relatively faint at these cold temperatures, tracer molecules are commonly used to probe this gas. Of these, carbon monoxide (CO) is the most popular, owing to both its bright emission lines, and relative abundance compared to other molecules. While the exact conversion factor between CO line luminosities and H_2 gas mass (or column density) remains fairly uncertain (e.g. Narayanan *et al.*, 2012; Bolatto, Wolfire & Leroy, 2013; Li *et al.*, 2018; Gong, Ostriker & Kim, 2018), there is not yet a demonstrated superior tracer of molecular gas in galaxies (Scoville *et al.*, 2014; Olsen *et al.*, 2017; Liang *et al.*, 2018; Privon, Narayanan & Davé, 2018).

CO emission lines can be used to constrain a diverse set of physical properties of galaxies. For example, their line profiles can help quantify global galaxy gas dynamics (e.g. Jiménez-Andrade *et al.*, 2018), as well as reveal molecular outflows (see the reviews by Carilli & Walter, 2013; Hodge & da Cunha, 2020; Veilleux *et al.*, 2020). At the same time, the modeling of their line spectral energy distribution can be used as a probe of the density and temperature of the emitting gas (e.g. Narayanan & Krumholz, 2014; Daddi *et al.*, 2015; Sharon *et al.*, 2016; Tunnard & Greve, 2016;

Strandet *et al.*, 2017; Kamenetzky, Privon & Narayanan, 2018; Popping *et al.*, 2019; Boogaard *et al.*, 2020).

The cosmic star formation rate density of galaxies peaks at redshifts $z \sim 1 - 3$ (Madau & Dickinson, 2014). While early studies with the SMA, PdBI and other facilities facilitated our understanding of the physical conditions in the molecular gas in these systems (e.g. Greve *et al.*, 2005; Tacconi *et al.*, 2006, 2008; Carilli *et al.*, 2010, for a few examples), these studies were typically limited to the brightest systems: submillimeter-selected galaxies and quasars (e.g. Casey, Narayanan & Cooray, 2014). In the last decade, the Atacama Large Millimeter Array (ALMA) has opened this field up, and allowed the relatively routine detection of molecular gas in main sequence galaxies (see the recent review by Hodge & da Cunha, 2020). These studies have allowed for precise redshift measurements, studies of molecular gas clumps on kpc-scales at cosmological redshifts (e.g. Swinbank *et al.*, 2015; Hodge *et al.*, 2019), and constraints on the cosmic molecular gas density evolution of the Universe (e.g. Walter *et al.*, 2014; Decarli *et al.*, 2019; Riechers *et al.*, 2019).

At the same time, most modern (sub)mm-wave facilities probe high-lying rotational CO states in high-redshift galaxies, which have effective densities 1–2 orders of magnitude greater than that of the ground state CO transition (e.g. Evans, 1999; Juneau *et al.*, 2009). Indeed, this density contrast has been evoked as a physical motivation for the nearly linear relationship seen between some high-J CO emission lines and the global star formation rate of galaxies (Krumholz & Thompson, 2007; Narayanan *et al.*, 2008; Bayet *et al.*, 2009; Greve *et al.*, 2014; Michiyama *et al.*, 2016). While this denser phase of gas is more closely related to the gas that will eventually collapse to form stars, it may trace only the tail of the metal-rich cold molecular gas mass density distribution, and therefore potentially miss significant fractions of the total gas reservoir.

The Jansky Very Large Array (JVLA) and Robert Byrd Green Bank Telescope (GBT) have demonstrated success in detecting the ground state CO line in high-redshift galaxies, enabling a measurement of the total metal-rich molecular gas mass (e.g. Carilli *et al.*, 2010; Harris *et al.*, 2010; Fu *et al.*, 2013; Sharon *et al.*, 2016). This said, the bulk of the detections have been in the most luminous systems, owing to the current sensitivity of these facilities. In this memo, we explore the efficacy of the proposed Next Generation VLA (ngVLA) in detecting CO (J=1-0) in high- z galaxies. We do this by simulating a population of galaxies in cosmological hydrodynamic simulations, and then from those, calculating the theoretical emergent CO emission line flux. In particular, we focus on the ability for the ngVLA to recover the total CO flux, the expected morphologies, and the relative spatial extent probed by ngVLA observations of the ground state CO line as compared to higher-lying submillimeter transitions as would be detected by ALMA.

1.1 Cosmological Galaxy Formation Simulations

In order to generate a population of galaxies, we employ the SIMBA cosmological galaxy formation simulation (Davé *et al.*, 2019). Briefly, SIMBA is the descendent of its ancestor, MUFASA, though with significantly updated physics modules. The simulations are run with GIZMO as the hydrodynamic solver in its meshless finite mass mode (Hopkins, 2014, 2017). In our simulations, stars form in H_2 gas, where the molecular content is derived following the Krumholz, McKee & Tumlinson (2009) dependency on the metal surface density and incident radiation field (we will discuss this in more detail in § 1.2). The star formation rate itself is given by $SFR = \epsilon_{\text{ff}} \times \rho_{H_2}/t_{\text{ff}}$, where $\epsilon_{\text{ff}} = 0.02$ is the assumed star formation efficiency, and t_{ff} is the local free-fall time. Once formed, these stars feed back on their surrounding gas following a diverse range of channels that include stellar winds, radiation pressure from massive stars, energy input from supernovae and asymptotic giant branch star feedback. The wind speeds and mass-loading factors for these winds are derived from scalings determined in high-resolution zoom-in simulations by Muratov *et al.* (2015).

SIMBA includes a model for black hole accretion and growth. Black holes are seeded within friends-of-friends identified halos where the enclosed stellar mass is greater than $M_* \geq 10^{9.5} M_\odot$, and grow via two modes of accretion: a cold-accretion model following the Anglés-Alcázar *et al.* (2017) torque-limited accretion model, and Bondi-Hoyle-Lyttleton accretion from the hot halo (Bondi & Hoyle, 1944; Bondi, 1952). These holes feed back on the surrounding ISM via both a kinetic wind and radiative wind (see Davé *et al.*, 2019, for more details).

Finally, SIMBA includes a self-consistent on-the-fly framework to model the formation, growth, and destruction of dust (described in significantly more detail in Davé *et al.*, 2019) and Li, Narayanan & Davé (2019). Briefly, dust grains are assumed to have a single size ($a = 0.1 \mu\text{m}$), and are passively advected alongside the gas particles. This dust is formed in the ejecta of supernovae and AGB stars, grows via metal accretion from the gaseous ISM, and can be destroyed via a diverse range of thermal sputtering processes (see Li, Narayanan & Davé, 2019, for more details). These simulations have resulted in strong matches to the observed submillimetre galaxy number counts at high-redshift (Lovell *et al.*, 2020), observed dust to gas and dust to metals ratios (Li, Narayanan & Davé, 2019), [CII] luminosity functions at high- z (Leung *et al.*, 2020), CO luminosity functions at low and high- z (Davé *et al.*, 2020), galaxy radial SFR profiles (Appleby *et al.*, 2020), and served as a test bed for modern SED fitting techniques (Lower *et al.*, 2020).

We identify galaxies from our cosmological simulation using CAESAR¹, where galaxies are identified via a 6D friends-of-friends finder.

1.2 Emission Line Physics

To determine the emergent CO emission line luminosity from these model galaxies, we employ the methods of Krumholz, Leroy & McKee (2011), Narayanan & Krumholz (2017), Li *et al.* (2018), and the relatively closely related Inoue, Yoshida & Yajima (2020), and briefly overview them here. In summary, we subdivide the gas particles in our model galaxies into a series of radially stratified zones. In each zone, we simultaneously calculate the thermal, chemical, and statistical equilibrium in order to determine the CO level populations; with these in hand, we are then able to compute the emergent CO line luminosity. These calculations are done using DESPOTIC (Krumholz, 2013).

In detail, the chemical and thermal states of model clouds will depend on their volume and column densities, as well as their velocity dispersions. We assume that every gas particle in the simulations represents a potential cloud, and is spherical in geometry. For these clouds, we discard ionized gas, and consider only the neutral content. The masses, densities and metallicities are given from the hydrodynamic simulation. We assume the radius of the cloud is given by the particle smoothing length, and therefore use this radius combined with the density to compute the cloud column density. This said, to account for diffuse particles that are not resolved, we impose a floor surface density such that:

$$\Sigma_{\text{cloud}} = \max [75 M_\odot \text{pc}^{-2}, \Sigma_{\text{comp}}] \quad (1)$$

where Σ_{comp} is the computed surface density.

In order to account for the subresolution turbulent compression of gas (e.g. Ostriker, Stone & Gammie, 2001; Padoan & Nordlund, 2002; Lemaster & Stone, 2008), we scale the volumetric densities of the clouds by a factor $e^{\sigma_p^2/2}$:

$$\sigma_p^2 \approx \ln (1 + 3M_{\text{ID}}^2/4) \times f_{\text{turb}} \quad (2)$$

¹<https://github.com/dnarayanan/caesar>

as suggested by high-resolution numerical simulations of turbulent boxes. Here, M_{1D} is the one dimensional Mach number, and f_{turb} is a tunable factor of order \sim unity used to scale our sub-resolution methods. Our scalings are based on ensuring that $z = 0$ Milky Way mass models from the same simulation series presented here have a median α_{CO} comparable to that observed, and we set $f_{\text{turb}} = 0.25$. This said, calculating the Mach number of course depends on the gas phase temperature, though the temperature equilibrium calculation implicitly depends on the density. Because iterating on the gas temperature would be quite costly, we instead assume $T = 10$ K for the Mach number calculation, representing an inconsistency in our simulation methods.

With our cloud physical properties in hand, we are now prepared to compute the chemical state of the cloud. We subdivide our clouds into 16 radial zones, and compute the chemical, thermal and statistical equilibrium iteratively for each zone. In each zone, we must determine the chemical state of carbon: CO, CI or C⁺. We assume that the total abundance of carbon is $[\text{C}/\text{H}] = 2 \times 10^{-4} \times Z/Z_{\odot}$, and $[\text{O}/\text{H}] = 4 \times 10^{-4} \times Z/Z_{\odot}$. We then employ the carbon-oxygen network developed by Nelson & Langer (1999), combined with the hydrogenic chemistry network by Glover & Mac Low (2007); Glover & Clark (2012). These networks require knowledge of both the interstellar radiation field (ISRF), as well as the local cosmic ray ionization rate. We discuss these in turn.

Following Narayanan & Krumholz (2017), we parameterize the ISRF in terms of the far ultraviolet radiation field normalized to that measured in the solar neighborhood (ξ_{FUV}). We assume that the star formation rate (SFR) of the Galaxy is $1M_{\odot} \text{ yr}^{-1}$ (e.g. Robitaille & Whitney, 2010; Chomiuk & Povich, 2011), and that the FUV radiation field is proportional to the total SFR in the galaxy: $\xi_{\text{FUV}} = \text{SFR}/(1M_{\odot}\text{yr}^{-1})$. We calculate the shielding rate to dust in terms of the visual extinction $A_V = (1/2)(A_V/N_{\text{H}})/N_{\text{H}}$, where the ratio (A_V/N_{H}) is the assumed dust extinction per hydrogen nucleus in the V band, assumed to be $4 \times 10^{-22} Z' \text{ mag cm}^2$. We refer the reader to Narayanan & Krumholz (2017) for more details.

Second, we require knowing the cosmic ray ionization rate, which is unfortunately poorly constrained observationally. We assume a value for diffuse gas of $\xi_{\text{CR}} = 10^{-17} \text{ s}^{-1}$, and scale this value linearly with the star formation rate of the galaxy.

The chemical reaction rates are temperature dependent. We calculate the temperature independently for every zone in our model clouds, following Goldsmith (2001) and Krumholz, Leroy & McKee (2011) in balancing the heating from the ISRF, cosmic rays, a thermal infrared field, and energy exchange with dust, and cooling by collisions, and cooling by line cooling. Formally we consider these processes both for dust and gas respectively:

$$\Gamma_{\text{PE}} + \Gamma_{\text{CR}} - \Lambda_{\text{line}} - \Lambda_{\text{H}} + \Psi_{\text{gd}} = 0 \quad (3)$$

$$\Gamma_{\text{ISRF}} + \Gamma_{\text{thermal}} - \Lambda_{\text{thermal}} - \Psi_{\text{gd}} = 0 \quad (4)$$

where Γ terms represent heating terms, Λ terms are cooling terms, and Ψ_{gd} represent an energy exchange between gas and dust, typically relevant at densities $n_{\text{H}} \gg 10^4 \text{ cm}^{-3}$ (e.g. Narayanan *et al.*, 2011, 2012). We refer the reader to Krumholz (2013) for details on how the rates for each of these are computed.

Finally, we must compute the statistical/radiative equilibrium within each zone. DESPOTIC calculates this using the escape probability formalism (see the Draine, 2011, book for a nice review). Formally, we solve the linear system of equations:

$$\begin{aligned} & \sum_j f_j \left[q_{ji} + \beta_{ji}(1 + n_{\gamma,ji})A_{ji} + \beta_{ij} \frac{g_i}{g_j} n_{\gamma,ij} A_{ij} \right] \\ & = f_i \sum_k \left[q_{ik} + \beta_{ik}(1 + n_{\gamma,ik})A_{ik} + \beta_{ki} \frac{g_k}{g_i} n_{\gamma,ki} A_{ki} \right] \end{aligned} \quad (5)$$

subject to the constraint $\sum_i f_i = 1$. Here A_{ij} is the Einstein coefficient for spontaneous transitions from state i to state j , g_i and g_j are the degeneracies of the states,

$$n_{\gamma,ij} = \frac{1}{\exp(\Delta E_{ij}/k_B T_{\text{CMB}}) - 1} \quad (6)$$

is the photon occupation number of the cosmic microwave background at the frequency corresponding to the transition between the states, E_{ij} is the energy difference between the states, and β_{ij} is the escape probability for photons of this energy. For each shell, we assume a spherical geometry when calculating the statistical equilibrium. The collisional rates derive from the Leiden Atomic and Molecular Database (Schöier *et al.*, 2005).

The chemical reaction networks have temperature-dependent rates. At the same time, the temperatures depend on the emission line cooling rates, while the statistical equilibrium depends on both the number density of a given species (i.e. that determined by the chemical reaction networks) as well as the temperatures (for the collisional rates). The problem is therefore iterative. We perform a triple-iteration procedure in which we run the chemical reaction network in the outermost loop while holding the temperature fixed, followed by the temperature calculation (holding the level populations fixed), and finally the statistical equilibrium. We iterate on this procedure until the abundances, temperatures and level populations converge for each zone.

1.3 Simulating ngVLA observations with CASA

The synthetic galaxy images display velocity-integrated line intensities in units of K km s^{-1} . To simulate ngVLA observations with CASA, we convert the images pixel values to velocity-integrated flux density in Jy km s^{-1} . Because the observed line intensity decreases with redshift as $I_{\nu}^{\text{obs}} = I_{\nu}^{\text{em}}/(1+z)^3$ (Condon & Matthews, 2018), and the emitted line intensity is $I_{\nu}^{\text{em}} = 2\nu^2 kT/c^2$, the observed flux of the line is given by

$$S_{\nu}^{\text{obs}} = \Omega I_{\nu}^{\text{obs}} = \Omega I_{\nu}^{\text{em}}/(1+z)^3 = \Omega \frac{2\nu^2 k}{c^2(1+z)^3} T, \quad (7)$$

where Ω is the solid angle of the source in the sky, ν is the rest-frame frequency of the line, k the Boltzmann's constant, c the speed of light, and T is the line intensity in Kelvin. Using Eq. 7, we derive the sky models in Jy km s^{-1} for our sample of synthetic galaxies at $z = 2$.

We then use the `simobserve` task in CASA to simulate ngVLA observations of the CO J=1-0, 2-1, 3-2 line emission. We consider the Plains+Core array to simulate 24 hour of observations with a resolution of 0.2 arcsec, i.e., $\approx 1.3 \text{ kpc}$ at the redshift of our galaxy models. We estimate and add thermal noise to our synthetic measurement sets using the ngVLA performance tables reported by Rosero (2019; ngVLA memo #55²), and following the instructions presented on the ngVLA webpage³. To generate the mock ngVLA images, we employ the `tclean` task with Briggs/Robust weighting and the multiscale algorithm. We choose a robust parameter and uv tapering scheme that leads to a PSF level of 10% at a radius of one FWHM (see Table 1), which assures a reasonable synthesized beam and noise performance for a target resolution of 0.2 arcsec, cell size of 0.01 arcsec/px and image size of 2048 px. We stop the cleaning when the residual peak flux density is ≈ 2.5 times the expected noise r.m.s, or when a maximum number of iterations of 20000 is reached.

We also simulate 24 hours of ALMA observations of the CO J=4-3, 5-4, 6-5 line emission following the instructions presented in the guide to simulating ALMA data⁴. We adopt the antenna

²http://library.nrao.edu/public/memos/ngvla/NGVLA_55.pdf

³https://casaguides.nrao.edu/index.php/Simulating_ngVLA_Data-CASA5.4.1

⁴https://casaguides.nrao.edu/index.php?title=Guide_To_Simulating_ALMA_Data

Line	$\nu_{\text{obs}}(z = 5)$	Robust	TA	$\text{FWHM}_{\text{maj}} \times \text{FWHM}_{\text{min}}$
	[GHz]		[mas]	[arcsec]
CO (1 – 0)	38.42	–0.10	178.0	0.21×0.19
CO (2 – 1)	76.85	0.12	185.5	0.21×0.20
CO (3 – 2)	115.27	0.31	178.3	0.21×0.19
CO (4 – 3)	153.68	0.5	...	0.27×0.19
CO (5 – 4)	192.09	0.5	...	0.25×0.18
CO (6 – 5)	230.49	0.5	...	0.25×0.18

Table 1: Imaging parameters for the different CO emission lines

configurations models `alma.out16`, `alma.out15`, and `alma.out14`, respectively, with `robust=0.5` to reach an angular resolution of $\approx 0.25 \times 0.18$ arcsec (see Table 1), which is similar to our 0.2 arcsec resolution ngVLA imaging. We employ multi-scale cleaning with a cell size of 0.01 arcsec/px, image size of 2048 px, a cleaning threshold of 2.5 times the theoretical r.m.s. and a maximum of 20000 iterations.

In order to measure the total flux density and extent of the simulated CO line observations, we employ the `statmorph` package in Python (Rodriguez-Gomez *et al.*, 2019) to calculate non-parametric morphological properties of the simulated CO line observations. The code derives the ellipticity and centroid of the galaxy’s surface brightness distribution to create elliptical apertures and determine the half-light radius. Using those elliptical apertures, we determine the distance between the farthest pixel with $S/N > 3$ and the galaxy center. This quantity, defined here as r_{max} , provides a more robust proxy for the total extent of the observed CO line emission. To estimate the total flux density of the images we use irregular apertures, which are defined by identifying contiguous pixels around the galaxy’s center with $S/N > 1.5$.

2 Results

2.1 Model Galaxies

For the purposes of this study, we employ the SIMBA m25n512 simulation (with an initial baryon resolution of $\sim 10^6 M_{\odot}$). We use the 25 most massive galaxies for our study here selected at $z = 2$. The ngVLA will be able to image CO J=1-0, 2-1 and 3-2 at this redshift. In Figure 1, we plot these galaxies on the $z = 2$ star formation rate (SFR)- M_{\star} relation. The bulk of our galaxies fall on the main sequence (though at the upper SFR/ M_{\star} end), with star formation rates $\sim 10^2 M_{\odot} \text{ yr}^{-1}$, though a few of our model galaxies are quenched/quiescent. The galaxies modeled here resemble galaxies ranging from disks at high-redshift that are actively forming stars through CO-detected quiescent galaxies (e.g. Bezanson *et al.*, 2019; Spilker *et al.*, 2019; Neeleman *et al.*, 2020).

2.2 Recovery of Total CO flux and Extended CO Emission

In Figure 2, we show an example of the results from this modeling procedure for three model galaxies (chosen arbitrarily). We show the (from left to right) theoretical CO emission map, the ngVLA recovered image, the residuals, and the point spread function (PSF). The ngVLA beam size is in the bottom left of the second column. While the galaxies shown here are arbitrarily chosen, they are representative of our model sample.

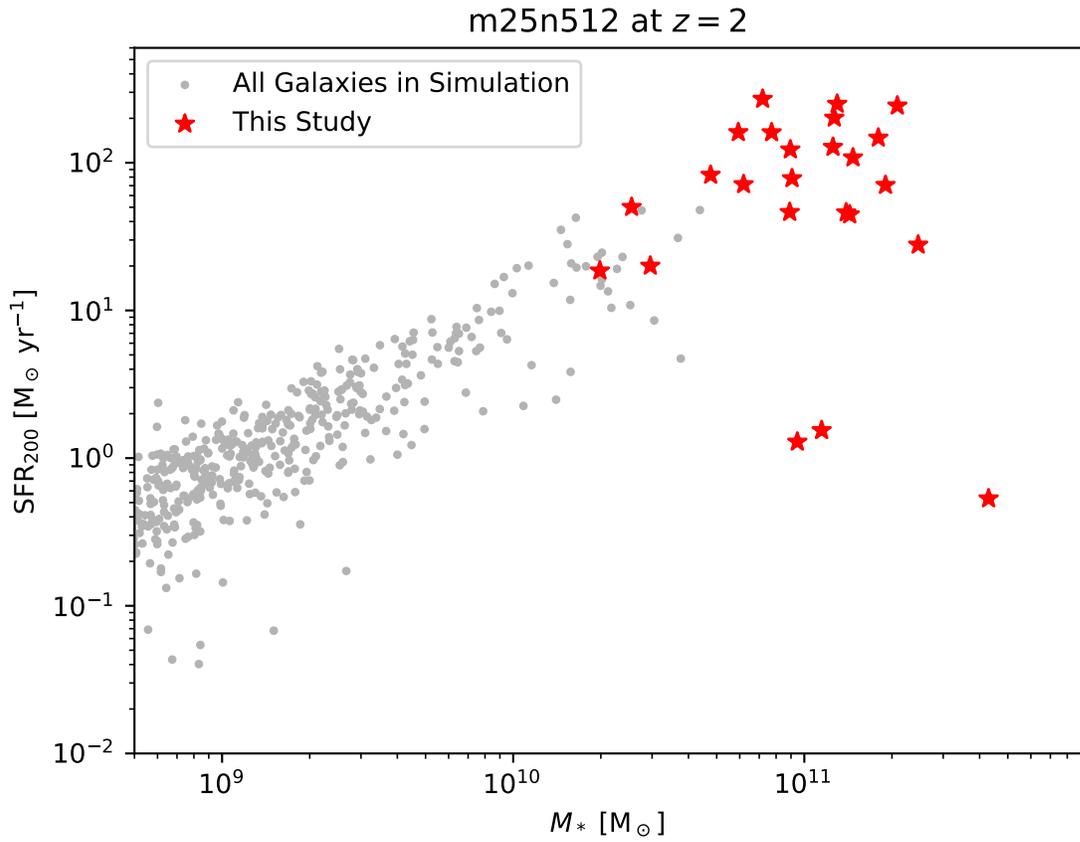


Figure 1: Star Formation Rate-Stellar Mass relation for our model galaxies. Grey points show all galaxies in the model snapshot, while red points show the galaxies employed for the study here.

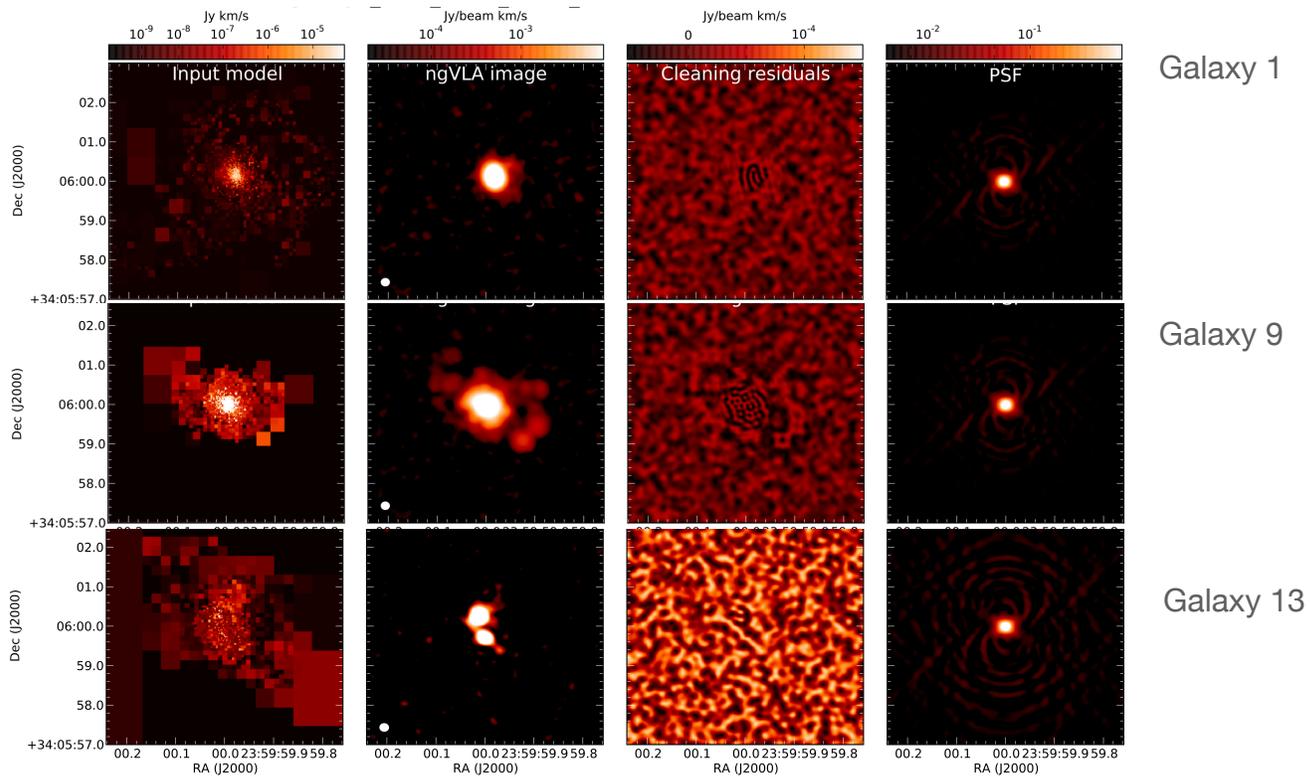


Figure 2: Simulated ngVLA images of three arbitrarily chosen model galaxies in CO ($J=1-0$). The columns from left to right show: (i) the theoretical input model that was input into CASA; (ii) the recovered ngVLA image; (iii) the cleaning residuals; (iv) the simulated point spread function.

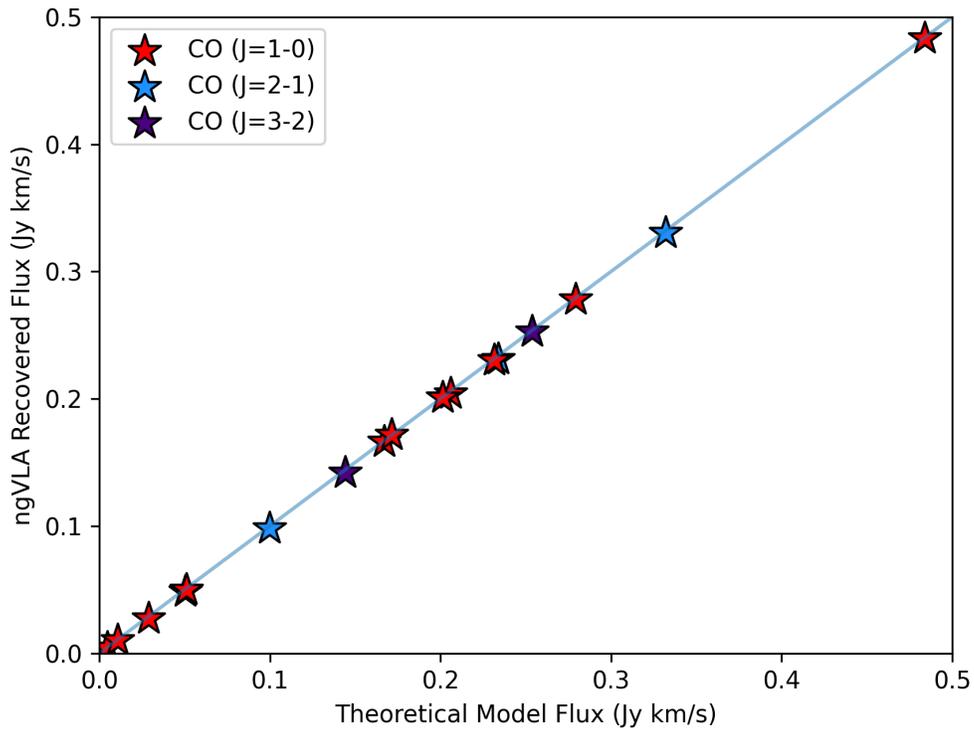


Figure 3: ngVLA recovered CO flux compared to the true theoretical CO flux for all galaxies in our sample from CO ($J=1-0$) through CO $J=3-2$. To guide the eye, the 1:1 line is additionally shown. The model galaxies are sufficiently compact that the facility faithfully recovers the entirety of the emission.

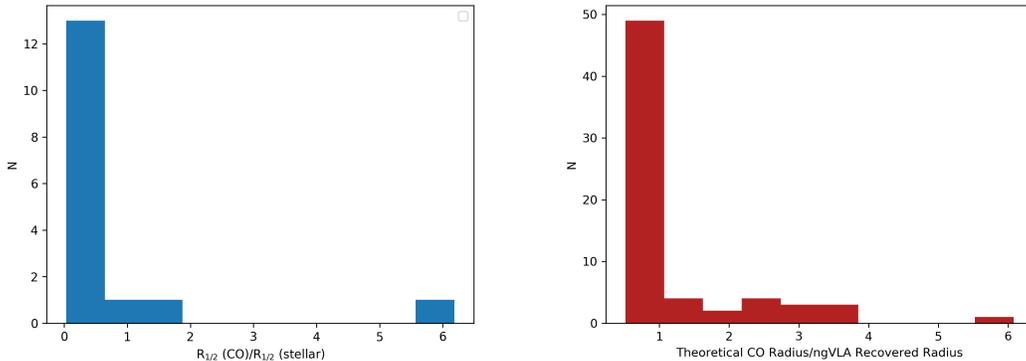


Figure 4: **Left:** Ratio of the CO (J=1-0) half light radius to the stellar half mass radius for our sample of galaxies. The CO radius is typically half that of the stellar radius. **Right:** Ratio of theoretical half-light CO (J=1-0) radius to the recovered ngVLA half-light radius. The ngVLA is able to recover the bulk of the extended emission

The ngVLA in the modeled configuration presented here is able to faithfully recover the bulk of the CO flux emitted. In Figure 3, we present the recovered CO flux for CO transitions J=1-0→3-2 for our modeled galaxies vs their true theoretical flux as returned from the models. To guide the eye, we additionally show the 1:1 line. As is evident, the ngVLA is able to recover the total true flux for all modeled CO transitions.

At the same time, much of the CO emission can be relatively extended (depending on the transition). It is therefore worth asking how much of this extended emission from $z = 2$ galaxies the ngVLA will be able to recover. The spatial extent of CO emission varies amongst our simulated galaxy sample, though typically extends to approximately half the stellar radius. In the left panel of Figure 4, we show a histogram of the ratio of the half CO (J=1-0) radius to the stellar half mass radius. The latter is computed with CAESAR⁵, where galaxies are identified via a 6D friends-of-friends algorithm, and the stellar half mass radius is the radius that encompasses half of the total galaxy stellar mass. The half-brightness radii for CO are typically $\sim 0.5 - 1''$, or < 10 kpc at cosmological redshifts. This is much less than the spatial extent of CO emission observed in, e.g. the Spiderweb protocluster by Emonts *et al.* (2016), though the halo mass of this system is likely much larger than any we are considering here.

In Figure 4, we compare the ngVLA derived CO 1/2 light radius to the theoretical 1/2 light CO radius. We show this as a histogram of the ratio between the two, and plot this for all modeled CO transitions at once. We manually remove galaxies that have clear multiple components, since our method for deriving the CO radii with STATMORPH only considers the brightest, central component.

Broadly, the ngVLA recovers the bulk of the extended CO emission. There are some outliers when the morphology is especially extended, though these instances are relatively rare, and contribute insubstantially to the total CO flux (e.g. Figure 3). We conclude that the ngVLA will be able to accurately derive both the total CO flux from high- z galaxies, as well as on average detect the bulk of the extended low surface brightness emission.

⁵<https://github.com/dnarayanan/caesar>

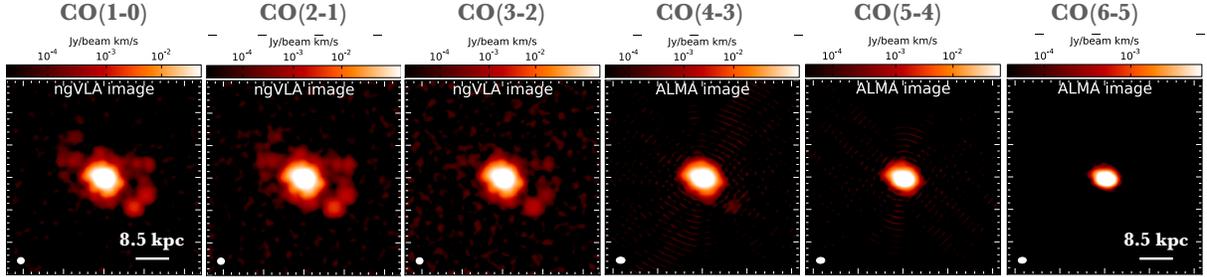


Figure 5: Mock CO emission map over CO transitions $1-0 \rightarrow 6-5$ as modeled in our mock ngVLA and ALMA observations. While the ground state transition can capture diffuse emission associated with spiral arms in this $z = 2$ galaxy, the ALMA-traced higher lying transitions primarily capture the nuclear regions.

2.3 CO Spatial Extent as a Function of Rotational State

Having established the ngVLA’s efficacy at detecting low-J CO emission at high- z , we now ask the question: how important is it to detect this emission? In other words, the bulk of existing submillimeter-wave interferometers detect high-J CO emission in high- z galaxies. Does this emission trace the bulk of the molecular mass in these galaxies?

We find that higher-lying CO transitions will generally emit from a more compact region, which will allow the ngVLA to, on average, more faithfully recover the entire spatial extent. In Figure 5, we show a CO map of an arbitrarily chosen galaxy in CO $J=1-0$ through $6-5$ as recovered by our mock ngVLA and ALMA observations. The emission is naturally more compact in the higher lying lines, owing to the higher effective density gas that they trace.

We quantify the effect seen visually in Figure 5 in Figure 6, where we show the full emitting size of our model galaxies as a function of CO transition. The solid line shows the median size from our sample of model galaxies, while the error bars denote the error of the median. Higher lying CO transitions probe increasingly compact morphologies owing to their higher effective densities. The median difference from the ground state as compared to CO ($J=6-5$) is $\sim 50\%$, with the extrema in the distributions probing even greater differences in spatial extent. The low frequency bands of the ngVLA will allow detection of cold reservoirs of gas that would otherwise be missed by the higher-lying transitions probed by ALMA in the submillimeter.

In Figure 7, we quantify this further by showing a histogram of the ratio of H_2 mass traced by CO ($J=6-5$) (observational half light radius) to the total H_2 mass in the galaxy as traced by CO $J=1-0$ (and assuming a constant X_{CO}). In making this figure, we assume that the CO ($J=1-0$) serves as a reliable and invariant tracer of the molecular gas mass (though see Bolatto, Wolfire & Leroy, 2013; Narayanan & Davé, 2012; Li *et al.*, 2018), and in reality are comparing the fraction of CO ($J=1-0$) emission that is not traced by the $6-5$ emitting area. While the most compact galaxies see the bulk of their molecular mass traced by CO $J=6-5$, the radial stratification as a function of CO energy level (e.g. Figure 6) for most galaxies causes up to 80% of the total molecular gas mass to be missed when observing CO ($J=6-5$) as a proxy for the molecular gas.

2.4 Morphologies and False Counterparts

Galaxies at high- z are often clumpy, with multiple concentrations of star formation (for a review, see Shapley, 2011). Bright knots of gas and dust can sometimes be connected via diffuse and low surface brightness bridges. At the same time, these clumps may be crucial to our understanding

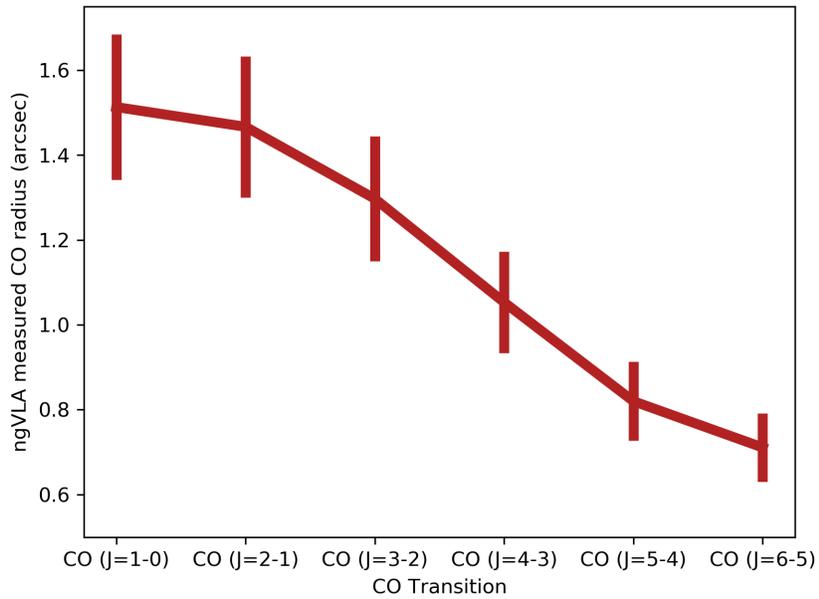


Figure 6: Full emitting region size recovered by the ngVLA for model galaxies as a function of CO transition. The horizontal line shows the median in the distribution at each transition, and the vertical line shows the error of the median. The CO $J=1-0 \rightarrow 3-2$ transitions are modeled with the ngVLA, while the higher lying lines are modeled as observed with ALMA. Generally, higher order rotational CO states probe increasingly high density gas, and therefore more compact regions in galaxies. The full emitting radius of the highest lying transition ($J=6-5$) can be up to $\sim 50\%$ smaller than the ground state emitting region.

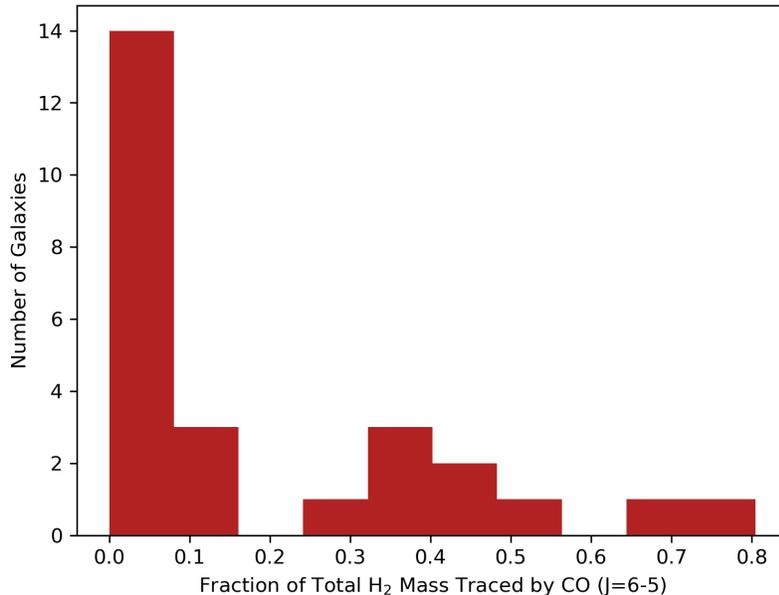


Figure 7: Distribution of fraction of total H₂ gas mass detected within the CO (J=6-5) half light radius. Telescopes that probe submillimeter-wave emission only at high-*z* such as ALMA will miss up to 90% of the underlying molecular gas mass.

of high-*z* galaxy formation. Clumps of gas formed from gas accretion and disk instabilities (e.g. Agertz, Teyssier & Moore, 2009; Dekel, Sari & Ceverino, 2009) may contribute to central bulge formation (Ceverino, Dekel & Bournaud, 2010; Krumholz & Dekel, 2010). At the same time, ongoing major and minor mergers, which can often have similar morphological signatures (e.g. Robertson & Bullock, 2008; Narayanan *et al.*, 2010, 2015; Hayward *et al.*, 2012; Abruzzo *et al.*, 2018) can be critical in shaping galaxy morphologies, star formation rates, and supermassive black hole growth.

Some simulation results suggest that kinematically unassociated blends of galaxies can mimic the signature of disk instabilities and galaxy mergers (Hayward *et al.*, 2013, 2018, though see Lovell *et al.* (2020)). While dust continuum imaging has traditionally served as a common vehicle for imaging the gaseous morphology of high-*z* galaxies (e.g. Gullberg *et al.*, 2019; Hodge *et al.*, 2019), spectroscopic information is necessary to determine kinematic association with the central luminosity source/galaxy (and avoid confusion by, e.g. chance projection of clumps/galaxies at different redshifts). In the bottom row of Figure 8, we show that the ngVLA is able to detect the bright star-forming clumps high-*z* galaxies in CO. These clumps are star-forming knots in the galaxy that appear as separate clumps owing to undetected diffuse gas (compare, for example, the first and second columns showing the theoretical vs observed CO morphologies for the three example galaxies shown in Figure 8).

Acknowledgements

This work was supported by the ngVLA Community Studies program, coordinated by the National Radio Astronomy Observatory (NRAO), which is a facility of the National Science Foundation

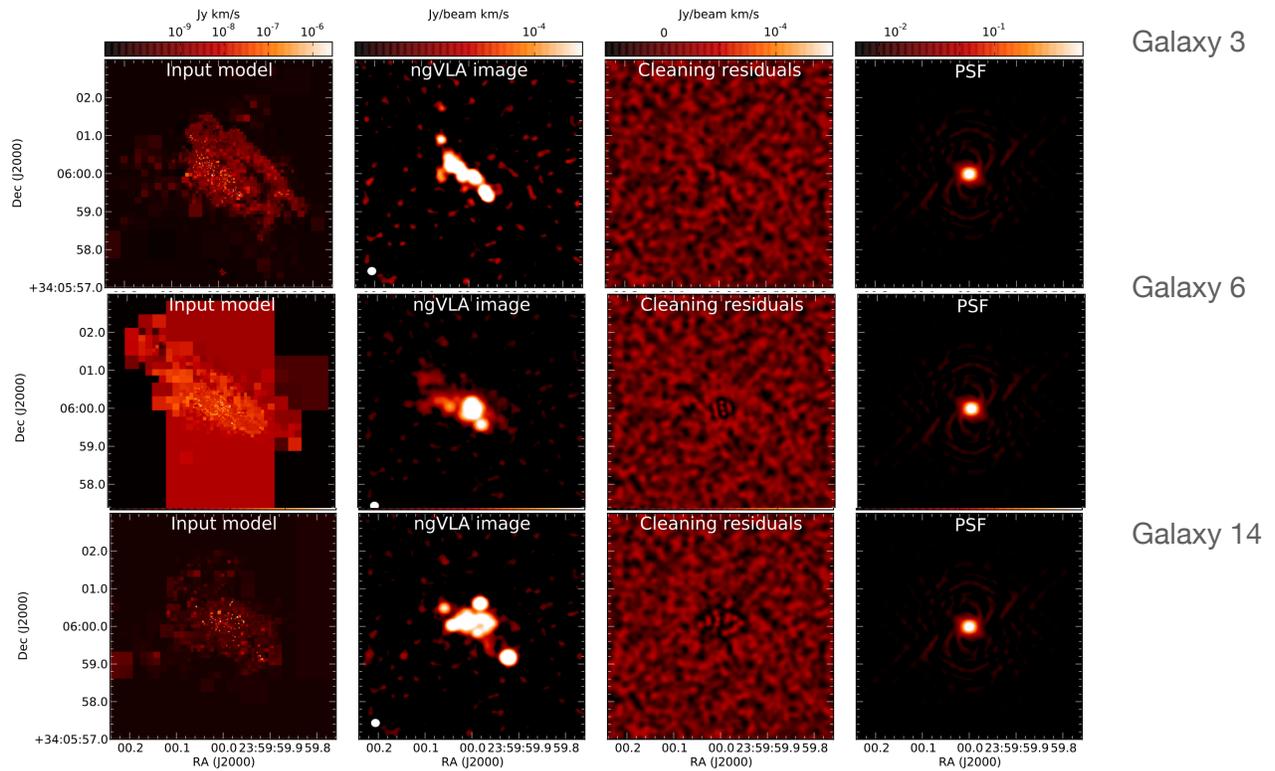


Figure 8: Example CO morphologies of three arbitrarily selected galaxies that appear clumpy in the ngVLA recovered image. The ngVLA will be instrumental in determining if detected gas clumps in galaxies are kinematically associated with the central emission source, or if they are unassociated blends (i.e. clumps of gas at different redshifts) as often confuses the interpretation of continuum imaging.

(NSF) operated under cooperative agreement by Associated Universities, Inc. The authors thank Mark Krumholz for making DESPOTIC public, as well as the developers of YT, and the broader YT community. Both codes facilitated this study. D.N. Alberto Bolatto, Chris Carilli, Caitlin Casey, Jackie Champagne, Aaron Evans, Jackie Hodge, Chao-Ling Hung, and George Privon for helpful conversations throughout this study.

References

- Abruzzo, M. W. *et al.* 2018, arXiv/1803.02374.
- Agertz, O., Teyssier, R., and Moore, B. 2009, MNRAS, 397(1), L64.
- Anglés-Alcázar, D. *et al.* 2017, MNRAS, 464(3), 2840.
- Appleby, S. *et al.* 2020, MNRAS, 494(4), 6053.
- Bayet, E. *et al.* 2009, MNRAS, 399, 264.
- Bezanson, R. *et al.* 2019, ApJL, 873(2), L19.
- Bolatto, A. D., Wolfire, M., and Leroy, A. K. 2013, ARA&A, 51, 207.
- Bondi, H. 1952, MNRAS, 112, 195.
- Bondi, H. and Hoyle, F. 1944, MNRAS, 104, 273.
- Boogaard, L. A. *et al.* 2020, arXiv e-prints, arXiv:2009.04348.
- Carilli, C. L. and Walter, F. 2013, ARA&A, 51, 105.
- Carilli, C. L. *et al.* 2010, ApJ, 714, 1407.
- Casey, C. M., Narayanan, D., and Cooray, A. 2014, Physics Reports, 541, 45.
- Ceverino, D., Dekel, A., and Bournaud, F. 2010, MNRAS, 404(4), 2151.
- Chomiuk, L. and Povich, M. S. 2011, AJ, 142, 197.
- Condon, J. J. and Matthews, A. M. 2018, , 130(989), 073001.
- Daddi, E. *et al.* 2015, A&A, 577, A46.
- Davé, R. *et al.* 2019, arXiv1901.10203.
- Davé, R. *et al.* 2020, MNRAS, 497(1), 146.
- Decarli, R. *et al.* 2019, ApJ, 882(2), 138.
- Dekel, A., Sari, R., and Ceverino, D. 2009, ApJ, 703(1), 785.
- Draine, B. T. 2011, Physics of the Interstellar and Intergalactic Medium, : Princeton University Press: Princeton, NJ).
- Emonts, B. H. C. *et al.* 2016, Science, 354(6316), 1128.
- Evans, II, N. J. 1999, ARA&A, 37, 311.

Fu, H. *et al.* 2013, *Nature*, 498(7454), 338.

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

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

Goldsmith, P. F. 2001, *ApJ*, 557, 736.

Gong, M., Ostriker, E. C., and Kim, C.-G. 2018, *ApJ*, 858(1), 16.

Greve, T. R. *et al.* 2014, *ApJ*, 794(2), 142.

Greve, T. R. *et al.* 2005, *MNRAS*, 359, 1165.

Gullberg, B. *et al.* 2019, *MNRAS*, 490(4), 4956.

Harris, A. I. *et al.* 2010, arXiv/1006.3691.

Hayward, C. C. *et al.* 2013, *MNRAS*, 434, 2572.

Hayward, C. C. *et al.* 2018, *MNRAS*, 476(2), 2278.

Hayward, C. C. *et al.* 2012, *MNRAS*, 424, 951.

Hodge, J. A. and da Cunha, E. 2020, arXiv e-prints, arXiv:2004.00934.

Hodge, J. A. *et al.* 2019, *ApJ*, 876(2), 130.

Hopkins, P. F. 2014, arXiv/1409.7395.

Hopkins, P. F. 2017, arXiv/1712.01294.

Inoue, S., Yoshida, N., and Yajima, H. 2020, arXiv e-prints, arXiv:2008.12484.

Jiménez-Andrade, E. F. *et al.* 2018, *A&A*, 615, A25.

Juneau, S. *et al.* 2009, *ApJ*, 707, 1217.

Kamenetzky, J., Privon, G. C., and Narayanan, D. 2018, *ApJ*, 859, 9.

Krumholz, M. R. 2013, *MNRAS*.

Krumholz, M. R. and Dekel, A. 2010, *MNRAS*, 406(1), 112.

Krumholz, M. R., Leroy, A. K., and McKee, C. F. 2011, *ApJ*, 731, 25.

Krumholz, M. R., McKee, C. F., and Tumlinson, J. 2009, *ApJ*, 693, 216.

Krumholz, M. R. and Thompson, T. A. 2007, *ApJ*, 669, 289.

Lemaster, M. N. and Stone, J. M. 2008, *ApJL*, 682, L97.

Leung, T. K. D. *et al.* 2020, arXiv/2004.11912, arXiv:2004.11912.

Li, Q., Narayanan, D., and Davé, R. 2019, *MNRAS*, 490(1), 1425.

Li, Q. *et al.* 2018, *ApJ*, 869, 73.

Liang, L. *et al.* 2018, MNRAS, 478, L83.

Lovell, C. C. *et al.* 2020, arXiv/2006.15156, arXiv:2006.15156.

Lower, S. *et al.* 2020, arXiv/2006.03599, arXiv:2006.03599.

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

Michiyama, T. *et al.* 2016, , 68(6), 96.

Muratov, A. L. *et al.* 2015, MNRAS, 454, 2691.

Narayanan, D. *et al.* 2008, ApJ, 684, 996.

Narayanan, D. and Davé, R. 2012, MNRAS, 423, 3601.

Narayanan, D. *et al.* 2010, MNRAS, 401, 1613.

Narayanan, D. *et al.* 2011, MNRAS, 418, 664.

Narayanan, D. and Krumholz, M. R. 2014, MNRAS, 442, 1411.

Narayanan, D. and Krumholz, M. R. 2017, MNRAS, 467, 50.

Narayanan, D. *et al.* 2012, MNRAS, 421, 3127.

Narayanan, D. *et al.* 2015, *Nature*, 525, 496.

Neeleman, M. *et al.* 2020, *Nature*, 581(7808), 269.

Nelson, R. P. and Langer, W. D. 1999, ApJ, 524, 923.

Olsen, K. *et al.* 2017, ApJ, 846, 105.

Ostriker, E. C., Stone, J. M., and Gammie, C. F. 2001, ApJ, 546, 980.

Padoan, P. and Nordlund, Å. 2002, ApJ, 576, 870.

Popping, G. *et al.* 2019, MNRAS, 482(4), 4906.

Privon, G. C., Narayanan, D., and Davé, R. 2018, arXiv/1805.03469.

Riechers, D. A. *et al.* 2019, ApJ, 872(1), 7.

Robertson, B. E. and Bullock, J. S. 2008, ApJL, 685(1), L27.

Robitaille, T. P. and Whitney, B. A. 2010, ApJL, 710, L11.

Rodriguez-Gomez, V. *et al.* 2019, MNRAS, 483(3), 4140.

Schöier, F. L. *et al.* 2005, A&A, 432, 369.

Scoville, N. *et al.* 2014, ApJ, 783(2), 84.

Shapley, A. E. 2011, ARA&A, 49, 525.

Sharon, C. E. *et al.* 2016, ApJ, 827(1), 18.

- Spilker, J. S. *et al.* 2019, ApJ, 883(1), 81.
- Strandet, M. L. *et al.* 2017, ApJL, 842(2), L15.
- Swinbank, A. M. *et al.* 2015, ApJL, 806(1), L17.
- Tacconi, L. J. *et al.* 2008, ApJ, 680, 246.
- Tacconi, L. J. *et al.* 2006, ApJ, 640, 228.
- Tunnard, R. and Greve, T. R. 2016, ApJ, 819(2), 161.
- Veilleux, S. *et al.* 2020, , 28(1), 2.
- Walter, F. *et al.* 2014, ApJ, 782(2), 79.