A High-Resolution Investigation of the Multi-Phase ISM in a Galaxy during the First Two Billion Years

S. Dye,[1][*] S. A. Eales,[2] H. L. Gomez,[2] G. C. Jones,[3,4] M.W.L. Smith,[2] E. Borsato,[5] A. Moss,[1]
L. Dunne,[2] J. Maresca,[1] A. Amvrosiadis,[6] M. Negrello,[2] L. Marchetti,[7,8] E. M. Corsini,[5,9] R. J. Ivison,[10]
G. J. Bendo,[11] T. Bakx,[12,13] A. Cooray,[14] P. Cox,[15] H. Dannerbauer,[16,17] S. Serjeant,[18] D. Riechers,[19]
P. Temi,[20] C. Vlahakis[21]

[1]*School of Physics and Astronomy, University of Nottingham, University Park, Nottingham, NG7 2RD, UK*
[2]*School of Physics and Astronomy, Cardiff University, The Parade, Cardiff, CF24 3AA, UK*
[3]*Cavendish Laboratory, University of Cambridge, 19 J. J. Thomson Ave., Cambridge, CB3 0HE, UK*
[4]*Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge, CB3 0HA, UK*
[5]*Dipartimento di Fisica e Astronomia "G. Galilei", Università di Padova, vicolo dell'Osservatorio 3, I-35122 Padova, Italy*
[6]*ICC, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK*
[7]*Department of Astronomy, University of Cape Town, Private Bag X3, Rondebosch 7701, Cape Town, South Africa*
[8] *INAF - Instituto di Radio astronomia, via Gobetti 101, I-40129 Bologna, Italy*
[9]*INAF - Osservatorio Astronomico di Padova, vicolo dell'Osservatorio 2, I-35122 Padova, Italy*
[10]*European Southern Observatory, Karl-Schwarzschild-Strasse 2, D-85748 Garching, Germany*
[11]*UK ALMA Regional Centre Node, Jodrell Bank Centre for Astrophysics, Department of Physics and Astronomy, The University of Manchester,
Oxford Road, Manchester M13 9PL, UK*
[12]*Division of Particle and Astrophysical Science, Graduate School of Science, Nagoya University, Aichi 464-8602, Japan.*
[13]*National Astronomical Observatory of Japan, 2-21-1, Osawa, Mitaka, Tokyo 181-8588, Japan*
[14]*Department of Physics and Astronomy, University of California, Irvine, CA92697, USA*
[15]*IRAM, 300 rue de la piscine, F-38406 Saint-Martin d'Hères, France*
[16]*Instituto de Astrofísica de Canarias (IAC), E-38205 La Laguna, Tenerife, Spain*
[17]*Universidad de La Laguna, Dpto. Astrofísica, E-38206 La Laguna, Tenerife, Spain*
[18]*School of Physical Sciences, The Open University, Milton Keynes, MK7 6AA, UK*
[19]*Cornell University, Space Sciences Building, Ithaca, NY 14853, USA*
[20]*Astrophysics Branch, NASA Ames Research Center, Moffett Field, CA 94035, USA*
[21]*National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, VA 22903-2475, USA*

ABSTRACT

We have carried out the first spatially-resolved investigation of the multi-phase interstellar medium (ISM) at high redshift, using the $z = 4.24$ strongly-lensed sub-millimetre galaxy H-ATLASJ142413.9+022303 (ID141). We present high-resolution (down to ~350 pc) ALMA observations in dust continuum emission and in the CO(7-6), $H_2O(2_{1,1} - 2_{0,2})$, [C I] (1-0) and [C I] (2-1) lines, the latter two allowing us to spatially resolve the cool phase of the ISM for the first time. Our modelling of the kinematics reveals that the system appears to be dominated by a rotationally-supported gas disk with evidence of a nearby perturber. We find that the [C I] (1-0) line has a very different distribution to the other lines, showing the existence of a reservoir of cool gas that might have been missed in studies of other galaxies. We have estimated the mass of the ISM using four different tracers, always obtaining an estimate in the range $3.2 - 3.8 \times 10^{11}$ M$_\odot$, significantly higher than our dynamical mass estimate of $0.8 - 1.3 \times 10^{11}$ M$_\odot$. We suggest that this conflict and other similar conflicts reported in the literature is because the gas-to-tracer ratios are $\simeq$4 times lower than the Galactic values used to calibrate the ISM in high-redshift galaxies. We demonstrate that this could result from a top-heavy initial mass function and strong chemical evolution. Using a variety of quantitative indicators, we show that, extreme though it is at $z = 4.24$, ID141 will likely join the population of quiescent galaxies that appears in the Universe at $z \sim 3$.

Key words: galaxies: ISM – galaxies: kinematics and dynamics – galaxies: star formation – submillimetre: galaxies – gravitational lensing: strong

[*] E-mail: simon.dye@nottingham.ac.uk

1 INTRODUCTION

Sub-millimetre galaxies (SMGs) are among the brightest objects observed at far-infrared (far-IR) and sub-millimetre (submm) wavelengths in the early universe. SMGs are typically highly optically-obscured by dust that originates from low/intermediate mass stars and supernovae (e.g., Rowlands et al. 2014) in regions undergoing extreme rates of star formation. These stellar nurseries are most abundant at redshifts around $z = 2 - 4$ when the Universe was most rapidly forming the galaxies and clusters we see around us today. SMGs have therefore become a valuable probe of the physics of the growth of these structures, particularly at the more massive end of the galaxy mass function (e.g., Casey et al. 2014; Hodge & da Cunha 2020).

The most extreme SMGs are those discovered in the wide-area surveys carried out with *Herschel* (Negrello et al. 2017; Bakx et al. 2018), the South Pole Telescope (Vieira et al. 2013; Reuter et al. 2020) and *Planck* (Cañameras et al. 2015), including ones that have star-formation rates many thousands of solar masses per year, which have been found out to a redshift of ~ 6 (e.g., Fudamoto et al. 2017). An obvious question is, what do these extreme objects evolve into at more recent epochs? In particular, do these galaxies evolve into the population of quiescent galaxies that appear in the Universe at $z \sim 3$? (e.g., Glazebrook et al. 2017; Girelli et al. 2019; Valentino et al. 2020; Forrest et al. 2020b,a; Merlin et al. 2019).

These huge star-formation rates must be fueled by large amounts of gas and many studies have found that SMGs house substantial molecular gas reservoirs of masses generally in the range $10^{10} - 10^{11}$ M$_\odot$ (e.g., Ivison et al. 2011; Bothwell et al. 2013; Huynh et al. 2017; Dye et al. 2018; Kaasinen et al. 2019). The bulk constituent of this gas, molecular hydrogen, is not rotationally excited at the typical temperatures in the molecular phase in galaxies. As such, investigations have been forced to use tracers, either dust or a variety of molecular and atomic species, thus requiring knowledge of the gas-to-tracer ratio. There is a complex hierarchy of calibrations but most of the values assumed for these ratios are ultimately based on observations in the Milky Way, the only place where it is possible to measure both the luminosity in the tracer and the mass of the underlying gas (Dunne et al. 2021).

As for galaxies in the nearby Universe, the most frequently used tracer of the molecular phase of the ISM, because it has so many bright spectral lines in atmospheric windows, is the CO molecule (Tacconi et al. 2018; Birkin et al. 2021). As a tracer of the underlying gas, CO has some well known disadvantages (Bolatto et al. 2013), in particular the fact that the J=1-0 line, the best for tracing the coldest gas, is optically thick. In addition, observations with *Fermi* (Abdo et al. 2010), *Planck* (Planck Collaboration et al. 2011) and *Herschel* (Pineda et al. 2013) all suggest that approximately one third of the molecular gas in the Galaxy does not emit CO, probably because of photo-dissociation of the CO molecule (Planck Collaboration et al. 2011). For high-resolution studies of the ISM in high-redshift galaxies there is the additional problem that the Atacama Large Millimetre Array (ALMA) is only able to observe the higher-excitation CO lines, making it possible that these studies have missed much of the gas (although lower-frequency ALMA bands will become available in the near future).

The other bright molecular lines in high-redshift galaxies are those of the water molecule, which have been used to observe both SMGs (Omont et al. 2013; Spilker et al. 2014; Yang et al. 2016; Jarugula et al. 2019) and AGN (e.g. van der Werf et al. 2011). The water lines are excited by far-IR photons (compared to collisional excitations of CO by H$_2$), explaining the correlation between line luminosity and far-infrared luminosity over three orders of magnitude (Yang et al.

2013). The water lines are therefore probably more useful as a tracer of star formation than the ISM (Jarugula et al. 2019).

Arguably the best tracers of the molecular phase of the ISM in high-redshift galaxies are the fine-structure lines of atomic carbon, despite being generally fainter. The [C I] $(^3P_1 -^3 P_0)$ and [C I] $(^3P_2 -^3 P_1)$ lines have similar excitation temperatures and critical densities to the low-J CO lines, but they are optically thin and they are emitted in all molecular regions including CO-dark regions (Papadopoulos et al. 2004; Papadopoulos & Greve 2004; Dunne et al. 2021).

The high-resolution configurations of ALMA, plus the magnification (often greater than a factor of 10) provided by gravitational lensing, have huge potential for investigating the physical conditions in the ISM in galaxies in the early Universe. This potential has been demonstrated quite dramatically using observations of the gravitationally-lensed SMG SDP81 that were carried out in the Science Verification Phase of the ALMA long baselines (ALMA Partnership et al. 2015). After being corrected for gravitational lens distortion, the reconstructed spectral-line and continuum images had a resolution of ~50 pc (Dye et al. 2015; Rybak et al. 2015a,b). The images revealed a smoothly-rotating but clumpy disk, with a Toomre Q-parameter suggesting that the disk is in a state of collapse. The resolution, approximately the size of a giant molecular cloud, made it possible to compare the Larson relations for the clumps with those in the Milky Way (Swinbank et al. 2015).

The drawback of the observations of SDP81, however, is that they were in high-J CO lines, making it uncertain whether a cool molecular phase had been missed; it is possible that these high-J CO clumps simply trace regions in the disk where the gas is particularly warm rather than genuine agglomerations of gas. In addition, at the redshift of SDP81, several spectral lines do not fall within atmospherically-transparent windows. To circumvent these shortfalls, we have therefore embarked on a second high-resolution study of the ISM using a galaxy at a redshift such that all of the key spectral lines are visible to ALMA.

One of the brightest strongly-lensed SMGs, H-ATLAS J142413.9 + 022303, hereafter referred to as 'ID141' following its label assigned by Cox et al. (2011) was discovered in the *Herschel* Astrophysical Terrahertz Large-Area Survey (Eales et al. 2010; Cox et al. 2011; Negrello et al. 2017). ID141 is not only bright because it is strongly-lensed but also because it is intrinsically very luminous. After correction for the lens magnification factor of ~ 6, the star-formation rate in ID141, estimated from the bolometric luminosity of the dust (and assuming the initial mass function of Kroupa & Weidner 2003), is $\simeq 2400$ M$_\odot$ yr^{-1}, five times higher than SDP81. With a redshift of $z = 4.24$ (Cox et al. 2011), ID141 is also one of the most distant strongly-lensed SMGs. At this redshift, both [C I] lines, the bright [C II] line at 158 μm, multiple CO lines (including the CO(1-0) line in the near future) and water lines, and many other lines are visible to ALMA.

As a result of its interest, ID141 has been observed many times in the submillimetre/millimetre wavebands (Cox et al. 2011; Bussmann et al. 2012, 2013; Omont et al. 2013; Dye et al. 2018; Cheng et al. 2020). Lens modelling by Dye et al. (2018, D18 hereafter) and Enia et al. (2018) revealed evidence of a disturbed morphology in the reconstructed continuum emission of the source (see also Bussmann et al. 2012 who showed that significant image residuals remain after subtracting the best fit lensed image of a single smooth source). A relatively high level of ionisation in the source was inferred from a high far-IR luminosity to gas mass ratio measured by Cox et al. (2011). This, along with very large line widths of ~ 800 km s^{-1} observed in CO emission led the authors to suggest the SMG is a merger, like most other SMGs (e.g., Engel et al. 2010).

In this paper, we present high-resolution (~ 0.1 arcsec) ALMA observations of ID141 in bands 3 and 4, complementing the ALMA band 7 observations presented in D18. The new observations cover the emission lines CO(7-6), $H_2O(2_{1,1} - 2_{0,2})$, $[C\,\textsc{i}]\,(^3P_2 - ^3P_1)$ and $[C\,\textsc{i}]\,(^3P_1 - ^3P_0)$, which have for the first time made it possible to carry out a high-resolution investigation of the ISM in a high-redshift galaxy that includes both the warm and the cool gas. The spectral-line observations have also allowed us to make a detailed kinematic analysis of this system, in particular addressing the question of whether ID141 is a merger.

One of the key findings of this paper is that the dynamical mass of ID141 is approximately four times lower than the gas mass inferred from Galactic gas-to-tracer ratios. SMGs with inferred gas masses that are equal to or larger than their dynamical mass have been reported in the literature for over a decade (e.g. Tacconi et al. 2008; Hodge et al. 2012; Bothwell et al. 2013). More recently, there has been an acceleration in the prevalence of such conflicting systems (e.g. Dye et al. 2015; Yang et al. 2017; Calistro Rivera et al. 2018; Neeleman et al. 2020; Mizukoshi et al. 2021). A common approach to reconcile these anomalies is to reduce the conversion factor (based on local Universe observations) that scales CO line intensity to total gas mass. In this work, we find that four different tracers yield the same excessive gas mass. It is unlikely that all four tracer conversion factors would need to be scaled by the same amount and so we have sought a more natural explanation in the form of a chemical evolution model which solves the discrepancy between the gas mass and dynamical mass in ID141.

The layout of this paper is as follows. In Section 2 we describe the acquisition and the reduction of the ALMA data. Section 3 presents our lens modelling and source reconstruction. Section 4 describes our investigation of the basic properties of the source, including the kinematic modelling and estimates of the mass of the ISM in the galaxy made five different ways. We discuss our findings in Section 5 and introduce a chemical evolution model of ID141. Section 6 is a summary of both our basic observational results and inferences from these results. Throughout this work, we have adopted the flat cosmological model with parameters $h = 0.678$, $\Omega_m = 0.308$ and $\Omega_\Lambda = 0.692$ (Planck Collaboration et al. 2016).

2 DATA ACQUISITION AND PROCESSING

The ALMA data analysed in this paper were acquired under two separate ALMA programmes. The first, 2016.1.00450.S, obtained band 4 spectral imaging on 20th August 2017 in four spectral windows (SPWs) centred on the frequencies 141.2, 143.1, 153.3 and 154.9 GHz, each with bandwidths of approximately 2 GHz. The total on-source integration time was 2360s using 44 12m antennas giving a spatial resolution of 0.15 arcsec, a maximum recoverable scale of 2.1 arcsec and a sensitivity of 0.34 mJy beam^{-1} assuming 10 km s^{-1} channels.

The second programme, 2017.1.00029.S, obtained band 3 spectral imaging on the 14th and 16th of November 2017 when 46 and 44 12m antennas were used respectively. Combining the measurement sets from both dates, the resulting images have a spatial resolution of 0.11 arcsec, a maximum recoverable scale of 2.0 arcsec and a sensitivity of 0.56 mJy beam^{-1} assuming 10 km s^{-1} channels. These observations were again carried out in four SPWs with central frequencies of 93.9, 95.8, 105.8 and 107.7 GHz with bandwidths of ~2 GHz and with a total on-source integration time of 3560s. This configuration of SPWs was chosen specifically to target the emission

Table 1. Configuration of ALMA used for spectral imaging of ID141. The table lists the central frequency, ν_{cent}, the band width, BW, and the number of channels, N_{chan}, for each spectral window.

Band	ν_{cent} (GHz)	BW (MHz)	N_{chan}
3	93.9	1875	3840
3	95.8	2000	128
3	105.8	1875	3840
3	107.7	2000	128
4	141.2	2000	128
4	143.1	2000	128
4	153.3	1875	3840
4	154.9	1875	3840

lines of CO, H_2O and $[C\,\textsc{i}]$ (see below). Table 1 gives more specific details on the spectral set up.

Calibration of both the band 3 and band 4 data was carried out using the scripts provided by the ALMA science archive using version 5.1.1 for band 3 and version 4.7.2 for band 4 of the Common Astronomy Software Applications (CASA) software package (Mc-Mullin et al. 2007). Bandpass and phase calibration was provided via the calibrators J1337-1257 and J1410+0203 respectively. For band 3, the two measurement sets (MSs) were independently calibrated and then combined into one MS using CASA. We manually verified that no erroneous visibilities remained after calibration using CASA's plotms tool. All observed images of ID141 displayed in this paper were produced by cleaning these MSs using CASA with natural weighting and a pixel scale of 0.024 arcsec. To isolate emission lines (see Section 2.1), we used the uvcontsub tool within CASA for each SPW to fit a linear spectrum to channels not containing line emission and then subtract this fit from the full visibility data. A separate set of visibilities was created for analysis of the continuum emission by excluding channels containing line emission.

In addition to the band 3 and 4 ALMA data detailed above, we have also used the ALMA band 7 data acquired under the ALMA programme 2013.1.00358.S and analysed by D18 (see Fig. 1). The observations in this programme amount to a total on-source integration time of 151s within four spectral windows spanning 336 to 352 GHz. The spectral windows contain no detectable emission lines and so we use the resulting 0.12 arcsec resolution continuum image solely for the purposes of morphological comparison of the reconstructed lensed SMG. We refer the reader to D18 for further details regarding these observations. Upon re-cleaning this dataset, we identified an error made in D18 such that the flux was over-estimated by approximately 30 per cent. The updated flux measured from the lensed image is 91 ± 5 mJy.

2.1 Emission lines

We identified four emission lines: $H_2O(2_{1,1} - 2_{0,2})$ ($\nu_{rest} = 752.033$ GHz), CO(7-6) ($\nu_{rest} = 806.652$ GHz), $[C\,\textsc{i}]\,(^3P_2 - ^3P_1)$ ($\nu_{rest} = 809.342$ GHz, referred to as $[C\,\textsc{i}]\,(2\text{-}1)$ hereafter) and $[C\,\textsc{i}]\,(^3P_1 - ^3P_0)$ ($\nu_{rest} = 492.161$ GHz, referred to as $[C\,\textsc{i}]\,(1\text{-}0)$ hereafter). Observed images of these lines are shown in Fig. 2 and the corresponding observed lensed line profiles (measured within a mask determined by selecting significant pixels) are shown in Fig. 3.

Table 2 lists the four line fluxes measured in the observed (i.e. lensed) images. The CO(7-6), $[C\,\textsc{i}]\,(1\text{-}0)$ and $[C\,\textsc{i}]\,(2\text{-}1)$ lines were also detected by Cox et al. (2011). The line flux measured for the $[C\,\textsc{i}]\,(1\text{-}0)$ line measured by Cox et al. (2011) is consistent with our measurement, but the Cox et al. fluxes of the other two lines are nearly
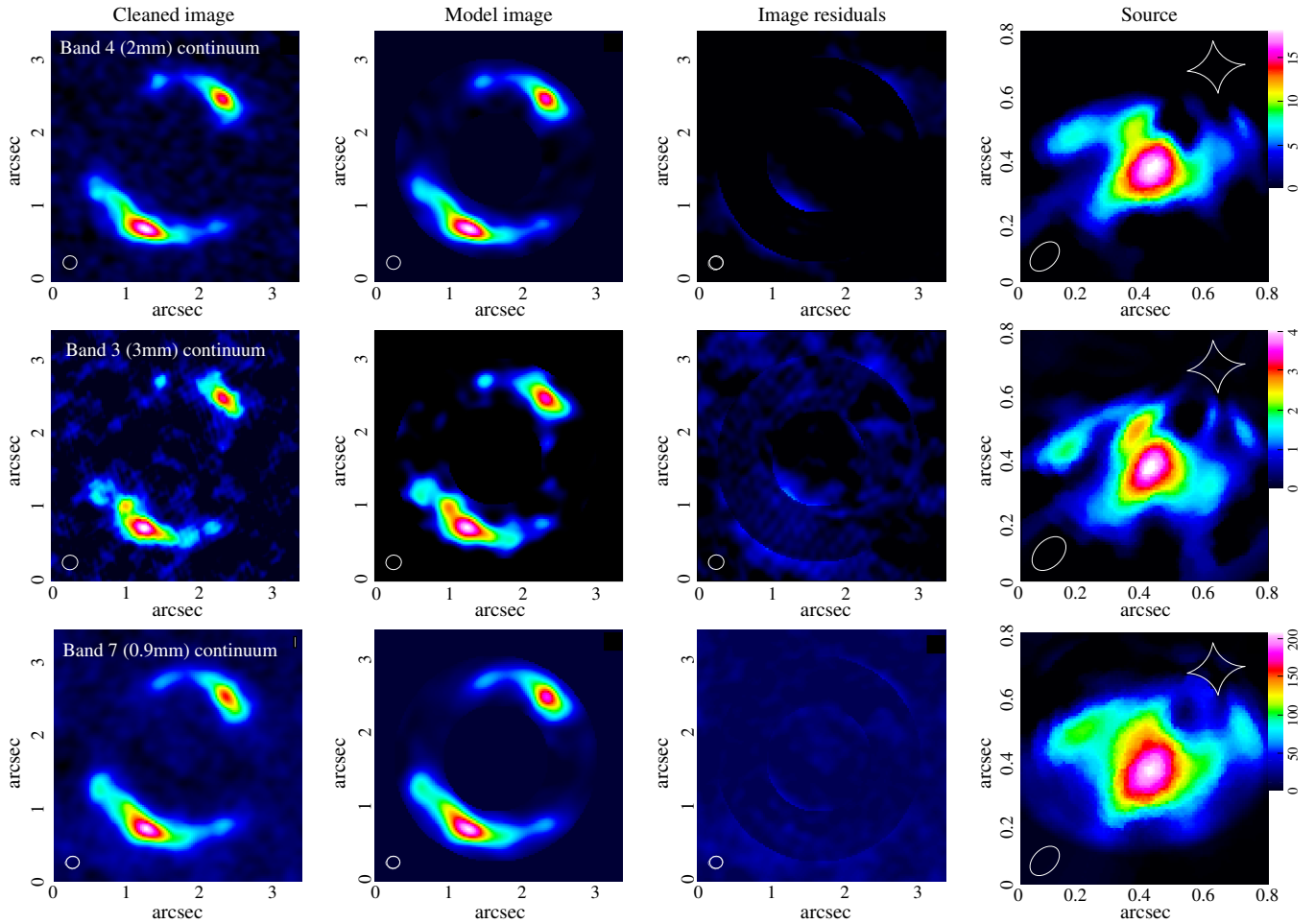
Figure 1. From left to right, columns show observed cleaned images, model lensed images of the reconstructed source emission, residuals (observed minus model) and reconstructed source (from the visibilities – the caustic is shown in white) of the continuum emission in band 3, 4 and 7. Observed images have been cleaned with a natural weighting with a pixel scale of 0.024 arcsec. The white ellipse shown in the lower left corners of the plots indicates the FWHM of the beam. (For the source plane, the effective central beam is plotted; see Section 3.2.) The colour wedge at the far right indicates the surface brightness in units of mJy arcsec^{-2}.

a factor of two lower than those measured by ALMA. Although the discrepancies are larger than expected, the errors quoted in Cox et al. (2011) do not include any uncertainty due to continuum subtraction which are large enough to eliminate any significant disagreement (R. Neri, private communication).

Fig. 3 also shows the emission lines determined from the reconstructed source plane emission (see Section 3.2). In this case, we extracted the line emission from the source plane region that, according to the lens model, maps to the image plane mask used to determine the image plane emission. It is clear from the three lines with the highest signal-to-noise ratio (SNR), i.e. H_2O, CO(7-6) and [C I] (2-1), that emission lines extracted from the image have slightly stronger flux at negative velocities whereas the reverse is true in the lines extracted from the source plane. Our lensing model confirms that this is due to differential magnification; the region of the source responsible for the emission of the negative side of the lines lies closer to the lens caustic where the magnification is higher.

Taking the median observed frequency of each of the H_2O, CO(7-6) and [C I] (2-1) lines measured from the reconstructed source and calculating the mean of these weighted by their SNR results in an inferred source redshift of $z = 4.24417 \pm 0.00003$ (see Table 2 - note

Table 2. Emission lines detected in the band 3 and band 4 ALMA data. The rest-frame frequencies, ν_{rest}, are measured from the reconstructed source cubes and therefore eliminate differential magnification of the emission lines. The SNR-weighted mean of the redshifts implied by the H_2O, CO(7-6) and [C I] (2-1) lines is $z = 4.24417 \pm 0.00003$. The final column gives the line flux in the observed (i.e. lensed) image.

Line	ν_{rest} (GHz)	ν_{obs} (GHz)	$z_{inferred}$	flux (Jy km s^{-1})
H_2O	752.033	143.403	4.24419	4.1 ± 0.4
CO(7-6)	806.652	153.820	4.24413	11.4 ± 1.1
[C I] (2-1)	809.342	154.332	4.24416	6.2 ± 0.8
[C I] (1-0)	492.161	93.8479	(4.24424)	3.1 ± 0.4

that we did not include the [C I] (1-0) line in this measurement due to its significantly lower SNR). This measurement is consistent with the redshift of $z = 4.243 \pm 0.001$ determined by Cox et al. (2011).
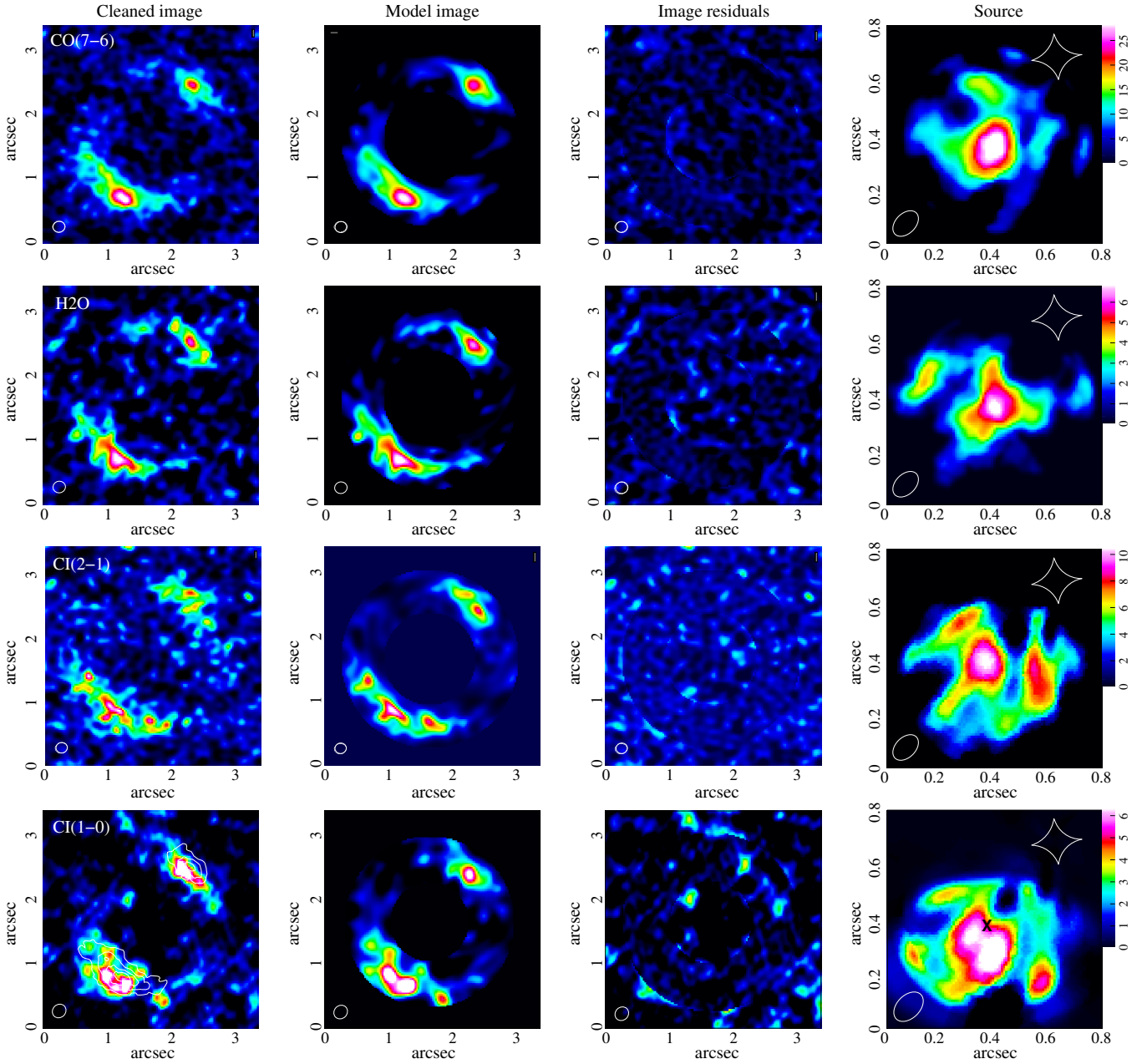
Figure 2. From left to right, columns show observed cleaned images, model lensed images of the reconstructed source emission, residuals (observed minus model) and reconstructed source (from the visibilities – the caustic is shown in white) of the four emission lines, CO(7-6), H_2O, [C I](2-1) and [C I](1-0) identified in our observations. Observed images have been cleaned with a natural weighting with a pixel scale of 0.024 arcsec. The white ellipse shown in the lower left corners of the plots indicates the FWHM of the beam. (For the source plane, the effective central beam is plotted; see Section 3.2.) The black cross in the panel showing the reconstructed [C I](1-0) line emission (bottom-right) shows the location of the peak reconstructed [C I](2-1) emission. The white contours in the observed [C I](1-0) line map (bottom-left) which follow the band 3 the continuum flux illustrate that the [C I](1-0) is spatially offset. The colour wedge at the far right indicates the line surface brightness in units of Jy km s^{-1} arcsec^{-2}.

3 LENS MODELLING

To recover an undistorted image of the lensed SMG, we applied the modelling process of D18. Using a given model for the distribution of mass in the foreground lensing galaxy, this process reconstructs a pixelised map of the source surface brightness (see Warren & Dye 2003). In this paper, we modelled solely the interferometric visibilities. The process works by computing a set of model visibilities from the lensed image of each source pixel. We computed these visibilities using `uvmodel` contained within the `MIRIAD` software suite (Sault et al. 2011). The linear combination of the sets of model visibilities that best fits the observed visibilities then recovers the source surface brightness distribution for a given lens model. The lens model is optimised iteratively, repeating the linear inversion with every iteration of the lens model. In this work, lens model optimisation was performed with `NeuralNest` (Moss 2020) which uses a neural network
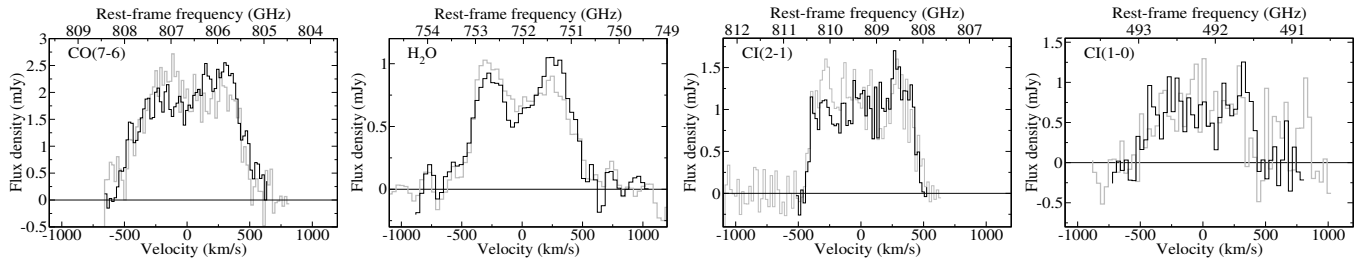
Figure 3. Emission lines, $H_2O(2_{1,1} - 2_{0,2})$, CO(7-6), [C I] (2-1) and [C I] (1-0), measured within a masked area determined by selecting significant pixels in their cleaned observed images (grey) and the corresponding line profile in their reconstructed source plane cubes (black). Image plane line profiles have been corrected for total lens magnification. Differential magnification causes the lensed image profiles to contain more flux at negative velocities even though intrinsically the source emits more flux at positive velocities.

to generate highly efficient proposals for Markov Chain Monte Carlo sampling.

3.1 Lens Model

Following the lens modelling in D18, we used an elliptical power-law density profile for the foreground lens, using the form given by Kassiola & Kovner (1993) which has a surface mass density

$$\kappa = \kappa_0 \, (\tilde{r}/1\text{kpc})^{1-\alpha} . \tag{1}$$

where κ_0 is the normalisation surface mass density and α is the power-law index of the volume mass density profile. The elliptical radial co-ordinate, $\tilde{r}$, is defined in terms of the Euclidean co-ordinates, $(x.y)$ relative to the ellipse centre, and the elongation, ϵ (i.e., the ratio of semi-major to semi-minor axis length), by $\tilde{r}^2 = x^2 + y^2/\epsilon^2$. The orientation of the semi-major axis, θ is measured east of north and the co-ordinates of the centre of the lens in the image plane are (x_c, y_c). Initial lens modeling tests with a model including external shear were attempted but it was found that negligible shear was required in the best fit model. This agrees with the findings of D18 and hence a parameterisation of external shear was not included in the final lens model which comprised six parameters in total. We also tried a lens model comprising two separate mass components as used by Bussmann et al. (2012) and as in D18, we found no significant improvement in the fit compared to a single ellipsoidal power-law. Although the two models give rise to a difference in the total source plane magnification of approximately 5 per cent, this is negligible in comparison to the errors on the intrinsic source properties we derive (see Section 5.2 for further discussion). Additionally, the two models yield almost identical source morphology and kinematical properties.

To determine our optimal lens model, we simultaneously fitted the visibilities from all four SPWs in the band 4 continuum data from this paper and the band 7 visibilities from D18. Combining bands and including the continuum in this way maximises the signal-to-noise ratio (SNR) to give the best possible lens model. We used the band 7 visibilities as presented in D18 but we time-averaged the more numerous band 4 visibilities into bins of 20 s to improve computational efficiency. We found that the size of this temporal bin can be changed by a factor of several without making any noticeable difference to the reconstructed lens model or source. We obtained a model that improves upon the model presented in D18 when applied solely to the band 7 visibilities. We attribute this to a combination of the new optimisation strategy (i.e., NeuralNest) used in this work and the complicated parameter space that results from randomising the source plane pixelisation (see Section 3.2).

The parameters of our best-fit lens model are: $\kappa_0 = (0.60 \pm 0.01) \times 10^{10} M_\odot \, \text{kpc}^{-2}$, $\epsilon = 1.14 \pm 0.02$, $\theta = (96 \pm 2)^\circ$, $\alpha = 2.01 \pm 0.03$. Comparing against the model in D18, there are two significant changes; the elongation has increased from 1.07 to 1.14 and the lens orientation has increased from 85° to 96° east of north. The source structure qualitatively looks very similar (see Fig. 1) but the total magnification of the source continuum has fallen from 6.4 to 6.1 in band 7. The Einstein radius of 0.97 ± 0.04 arcsec remains the same.

3.2 Source Reconstruction

We reconstructed the lensed source on a 0.8×0.8 arcsec source plane consisting of 400 Voronoi pixels formed using the method described in D18. The Voronoi tessellation is formed within the source plane region traced by mapping the region enclosed within an annular mask applied to the ring images as can be seen in Figs. 1 and 2. When optimising the lens model, we computed seven randomisations of the source plane pixelisation for each trial set of lens model parameters and returned the median likelihood of these seven randomisations to NeuralNest. This alleviates the effect reported in Nightingale et al. (2018) whereby lens model optimisation can occasionally stagnate when exceptionally good pixelisations occur by chance.

When reconstructing source cubes, we extracted the set of visibilities for each channel and applied the best-fit lens model determined in Section 3.1. We reconstructed the visibilities corresponding to each slice of each cube 50 times, randomising over source plane pixelisations every time. All 50 random Voronoi pixelisations were then individually binned onto a fine grid of 100×100 square pixels across the source plane and their mean computed to arrive at the final source reconstruction for a given cube slice.

The reconstruction scheme applied to cube slices results in an effective point spread function (PSF) that varies with lens magnification and therefore with position in the source plane. To quantify the resolution of our reconstructions, we computed the spatial variation of this PSF in the following way. First, we generated a set of model visibilities of a lensed image of a point source created using the best fit lens model. These visibilities were obtained using MIRIAD's uvmodel task which evaluates them at the same uv co-ordinates as the observed visibilities. We then applied our reconstruction method, including averaging the 50 random Voronoi source planes as described above to arrive at a reconstructed point source. This procedure was repeated for a grid of 3×3 point source positions in the source plane (arranged to cover the actual observed source light) and the resulting superposition of these recovered point sources is shown in Fig. 4. The dimensions of the PSFs obtained in this way for band 3 and 4 are very similar and these are 0.17×0.10 arcsec, 0.08×0.05 arcsec and
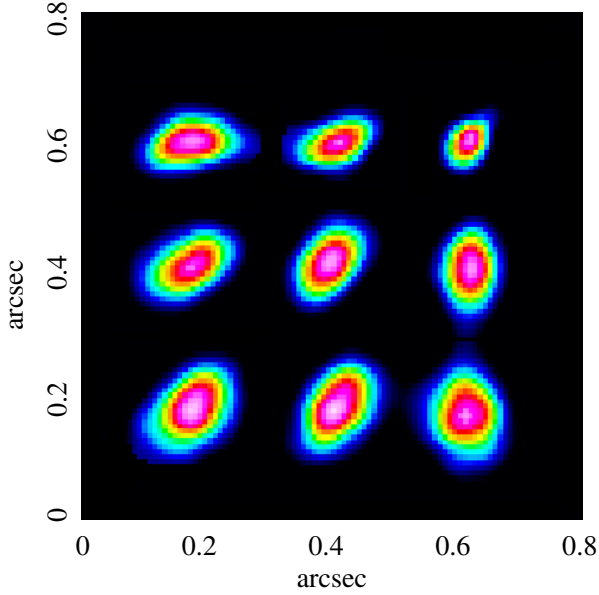
Figure 4. Variation of PSF across reconstructed source plane for band 4 data. The figure is a superposition of the reconstruction of nine individual point sources placed at one of nine positions in a 3×3 grid in the source plane. For each point source, we made a model set of visibilities and reconstructed the set with the same source pixelisation scheme used for reconstructing the lensed SMG (see main text for more details). The beam at the centre of the grid has a FWHM of 0.12 and 0.08 arcsec along its major and minor axis respectively. For the band 3 data, the PSF shows the same qualitative variation across the source plane but has a linear scale that is approximately 15 per larger than the band 4 PSF.

0.12×0.08 arcsec, in the south-eastern, north-western and middle of the source respectively where the dimensions quoted are full widths at half maximum (FWHM) along the major and minor axes. At the source redshift (where 1 arcsec corresponds to a physical scale of 6.90 kpc), the smallest physical scale we are able to probe is therefore approximately 350 pc. The median resolution across the source, computed as the geometric mean of the PSF FWHMs, is approximately 550 pc.

To determine the unlensed continuum flux of ID141, for each band, we applied the source reconstruction method to the emission line-subtracted visibilities obtained as outlined in section 2. The reconstructed source continuum in bands 3, 4 and 7 is shown in Fig. 1. Note that the same morphological features seen in the band 7 continuum source shown in this figure are also seen in D18, despite the different lens model and the different source pixelisation scheme. The quantitative difference is a 5 per cent reduction in magnification compared to D18 which results in a visually imperceptible increase in the linear scale of the source shown in Fig. 1 of 2–3 per cent. To determine the unlensed fluxes of the four identified emission lines, we reconstructed the source using the visibilities containing line emission, for each line. These are shown in Fig. 2.

With the source reconstructed in this way, the flux was then measured in the source plane aperture that maps to the image plane annulus for both the continua and emission lines. Line fluxes are given in Table 4 and the continuum fluxes measured in this way are 0.26 ± 0.04 mJy, 1.16 ± 0.14 mJy and 14.9 ± 0.8 mJy in band 3, 4 and 7 respectively. The continuum and line magnifications were computed

Table 3. Total lens magnifications, μ_{tot} of the continuum in bands 3, 4 and 7 and the emission lines identified in the ALMA data.

Band:	3	4	7	
μ_{tot} :	5.5	5.9	6.1	

Line:	CO(7-6)	H_2O	[C I] (2-1)	[C I] (1-0)
μ_{tot} :	6.2	5.8	6.3	5.6

in more restricted apertures. We formed an image-plane mask by selecting all pixels with a significance of at least 3σ and then mapped this to the source plane. Magnifications were then determined as the ratio of the image flux to source flux in these apertures. Table 3 lists the magnifications obtained in this way for the emission lines and band 3, 4 and 7 continua.

4 THE PROPERTIES OF THE LENSED SOURCE

In this section we give consideration to the reconstructed morphology of ID141, its spectral energy distribution (SED) including the inferred quantities of dust mass, dust temperature and star formation rate and the gas mass of the ISM. Section 4.6 presents our kinematical analysis. We discuss the implications of these results in Section 5.

4.1 Source Morphology

The reconstructed emission line images of ID141 in Fig. 2 show clear morphological differences to the reconstructed continuum emission in Fig. 1. Qualitatively, the line emission is more irregular than the continuum emission. This is partly an effect of the lower SNR of the reconstructed line flux but not entirely; the lensed emission line images in the leftmost column of Fig. 2 clearly have higher noise than the lensed continuum images in Fig. 1 but many of the clumpy features seen in the reconstructed line images in the rightmost column of Fig. 2 are multiply-imaged in the observed ring and can be seen at different wavelengths.

One particularly striking feature is an offset in the peak emission seen in the [C I] (1-0) line relative to the other lines and relative to the continuum emission. This is apparent in both the reconstructed source and the lensed image. The bottom-left panel of Fig. 2 illustrates this offset by over-plotting contours of the band 3 continuum flux on the line emission map. Since [C I] (1-0) has a lower excitation temperature than the other lines, this suggests the presence of cooler gas in the southern parts of the galaxy. We return to this point in Section 4.3.

4.2 Far-IR Spectral Energy Distribution

Fig. 5 plots measurements of the continuum flux of ID141 at different rest-frame wavelengths. The filled points are taken from D18 but reduced by 5 per cent to account for the updated lens magnification and from Cheng et al. (2020). We also updated the $850 \, \mu m$ flux originally taken from Bakx et al. (2018) with the flux of 90 mJy following Bakx et al. (2020). The new band 3 and 4 continuum fluxes are shown in this figure with empty circles.

Following D18, we fitted the rest-frame photometry with a modified black-body SED (referred to hereafter as the 'optically thick SED') comprising a single dust temperature component as well as an optically thin SED with two dust temperature components. This is an over-simplistic treatment of the expected continuous range of dust
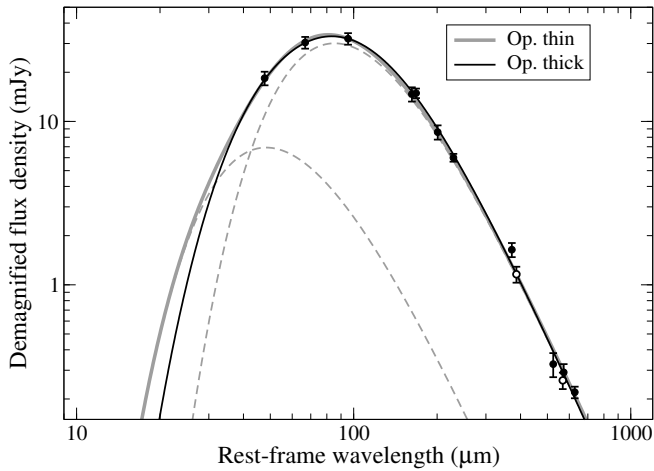
Figure 5. Spectral energy distribution of ID141. Data points with filled circles are from D18 (but adjusted for the new lens model) and Cheng et al. (2020). The open circles are the band 3 and band 4 continuum fluxes from the present work. The solid grey and black lines show an optically thin SED with two temperature components and an optically thick single-temperature SED fit respectively (see text for more details). The dashed lines are the two separate components that comprise the optically thin SED.

Figure 6. Gas temperature as a function of distance moving northwards across ID141. The temperature has been estimated from the ratio of the luminosity of the [C I] (2-1) and [C I] (1-0) lines in east-west slices of length 0.2 arcsec and width 0.08 arcsec, on the assumption of local thermodynamic equilibrium. The vertical dashed line on the left shows where the [C I] (1-0) emission peaks (see also Fig. 2) and the vertical dashed line on the right shows where the [C I] (2-1) emission peaks.

temperatures which we incorporate in later SED fitting (see Section 4.5). Nevertheless, these two simple SED models provide a good estimate of the range of possible dust masses. We measured the dust mass predicted by each SED using the method described in Dunne et al. (2011). For this, we took the observed ALMA 880 μm flux density of 91 ± 5 mJy from the band 7 data and a dust mass absorption coefficient at the corresponding rest-frame wavelength, 168 μm, of 1.7 m^{2}kg^{-1} from James et al. (2002). Both SEDs were fitted with a fixed emissivity index of 2.0 (Smith et al. 2013).

We allowed the temperature and normalisation of both components to vary in the fit for the optically thin SED giving four degrees of freedom. For the optically thick SED, temperature, normalisation and the 100 μm-opacity, τ_{100}, were varied in the fit, giving three degrees of freedom. The best-fit parameters are $T_{\rm warm} = 60 \pm 6$ K and $T_{\rm cold} = 34 \pm 3$ K with a ratio of cold-to-warm dust mass of 74 ± 13 for the optically thin SED. For the optically thick SED, we found $T = 57 \pm 1$ K and $\tau_{100} = 4.8$. The predicted dust masses are $(1.4 \pm 0.3) \times 10^9$ M$_\odot$ and $(7.4 \pm 0.8) \times 10^8$ M$_\odot$ for the optically thin and optically thick SEDs respectively.

Integrating these SEDs from 8-1000 μm gives a far-IR luminosity, $L_{\rm FIR}$, of 1.6×10^{13} L$_\odot$ and 1.5×10^{13} L$_\odot$ for the optically thin and thick SEDs respectively. These values are consistent with Cox et al. (2011) once corrected for lens magnification and places ID141 in the category of hyper-luminous infra-red galaxies (Rowan-Robinson 2000). Taking the mean of these luminosities and converting it to a star formation rate (SFR) according to the prescription given by Kennicutt & Evans (2012), which uses the initial mass function given by Kroupa & Weidner (2003), gives a value of 2400 ± 500 M$_\odot$yr^{-1}.

This is consistent with the SFR calculated in D18 but is 5 per cent higher due to the 5 per cent reduction in magnification of the new lens model. As discussed in D18, given its dust mass, this places ID141 significantly above the range of SFRs for $z < 0.5$ H-ATLAS galaxies and in the upper envelop of high redshift SMGs (Rowlands et al. 2014).
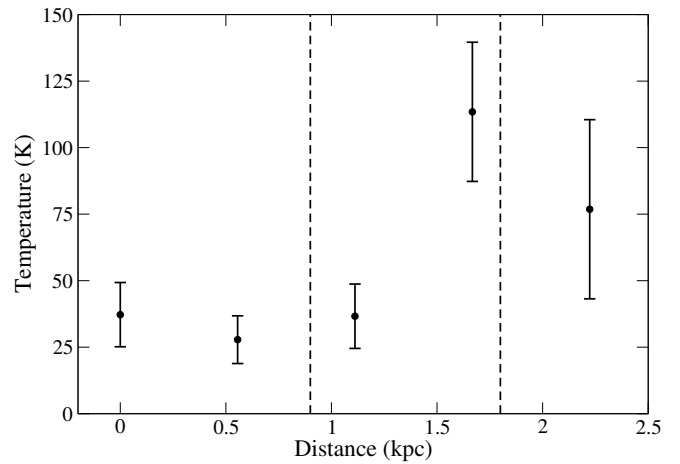
4.3 The Temperature Distribution in the ISM

There is a significant offset of the peak emission seen in the [C I] (1-0) line compared to the peak of emission seen in the other three lines. Initially concerned that this may simply be the result of an offset in the world co-ordinate system between the band 3 data (where the [C I] (1-0) line is located) and band 4 data, we confirmed that the continuum images between both bands align perfectly. Indeed, the [C I] (1-0) emission is offset significantly from the band 3 continuum (see Fig. 2) and this is clearly apparent in both the reconstructed source and the lensed image. The offset between the [C I] (1-0) emission and the [C I] (2-1) immediately suggests the temperature of the ISM is varying across the galaxy.

The advantages of the [C I] lines for tracing the cool ISM over the traditional CO method has become increasingly recognised in the last few years (Papadopoulos et al. 2004; Dunne et al. 2021, Dunne et al. *in preparation*). The [C I] lines have similar excitation temperatures and critical densities to the low-J CO lines, but their advantages over the CO lines are that the [C I] lines are mostly optically thin and atomic carbon is present everywhere in the molecular phase (Papadopoulos et al. 2004). In contrast, a significant and variable fraction of the molecular phase in galaxies appears to contain no CO, presumably because of photo-dissociation (Abdo et al. 2010; Pineda et al. 2013; Planck Collaboration et al. 2011; Bigiel et al. 2020; Dunne et al. 2021).

Since the [C I] lines are optically thin, it is reasonably straight-forward to use the ratio of the [C I] (2-1) and [C I] (1-0) lines to investigate the variation of the temperature in the ISM. Therefore, we divided the [C I] (2-1) and [C I] (1-0) source images in Fig. 2 into a stack of five slices running east-west, each 0.2 arcsec long and 0.08 arcsec wide, with the stack positioned to cover the main [C I] emission. We calculated the ratio of the luminosity (measured in K km s^{-1} pc^2, Solomon & Vanden Bout 2005) of the two lines, $R = L'(2-1)/L'(1-0)$, in each of these slices. Assuming local thermodynamic equilibrium, we then converted the line ratio into an excitation temperature using $T_{\rm ex} = 38.8$K/ln$(2.11/R_{\rm CI})$ (Stutzki et al. 1997).
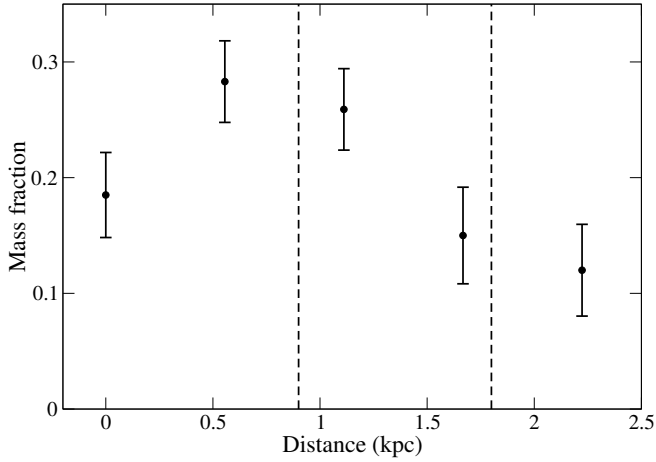
Figure 7. Fraction of the total ISM mass in the molecular phase as a function of distance moving northwards across ID141. Each data point has been determined from the [C I] (1-0) flux in a series of east-west slices of length 0.2 arcsec and width 0.08 arcsec arranged across ID141 and the excitation temperature estimated from the ratio of the luminosity in the [C I] (1-0) and [C I] (2-1) lines (Fig. 6). The vertical dashed line on the left shows where the [C I] (1-0) emission peaks (see also Fig. 2) and the vertical dashed line on the right shows where the [C I] (2-1) emission peaks.

Fig. 6 shows the variation of temperature in the stacked slices moving northwards across ID141. A $\simeq 2.5\sigma$ increase in temperature is seen rising from $\simeq$40 K in the three southern bins to $\gtrsim$ 100 K in the two northern bins. This trend is in accord with the other images in Fig. 2 because the water line and the CO(7-6) line also peak in the north, and both lines are likely to be from high-temperature gas (Papadopoulos et al. 2012a; Liu et al. 2017). The excitation temperature of the galaxy as a whole, estimated from the global fluxes for the [C I] lines (Table 4) and the equation above, is 33 K. To put this into context, in the sample of 49 galaxies of Dunne et al. (*in preparation*), ID141 would rank as the fifth highest [C I] excitation temperature.

The mass of the molecular phase of the ISM is proportional to $L'([C I] (1 - 0))/Q_{10}$, where Q_{10} is the partition function for the [C I] (1-0) line (Solomon & Vanden Bout 2005), which depends on the excitation temperature. Fig. 7 shows the fraction of the molecular mass in each of our five slices calculated from the [C I] (1-0) flux and our estimate of the excitation temperature in each slice, based on the assumption of local thermodynamic equilibrium. This demonstrates that although the gas is much hotter in the north, most of the mass of the molecular phase is in the colder gas to the south.

4.4 First Estimates of the Mass of the ISM

Any estimate of the mass of the molecular ISM in a galaxy other than our own requires the use of an ISM tracer such as CO molecules or dust. This in turn requires an estimate of the constant of proportionality (the calibration factor) between the luminosity of the tracer and the mass of the ISM. In this section, we estimate the mass of the ISM in ID141 using four different tracers and five different methods. The uncertainties quoted on the ISM mass estimates include the flux uncertainty and the uncertainty of the conversion factors used.

4.4.1 The gas mass from CO

We have two measurements of CO lines, a CO(7-6) line luminosity which has been corrected by the lensing model derived in this paper, and a CO(1-0) line luminosity which has not been corrected for lensing because the observations did not have high enough resolution (Dunne et al. *in preparation*). Since most of the work on calibrating the CO method has used the CO(1-0) line and since the ratio of high-J to low-J line flux has a large dispersion (Cañameras et al. 2018), we have estimated the gas mass from the CO(1-0) measurement.

The total CO(1-0) line flux is 0.61 ± 0.09 Jy km s^{-1} (Dunne et al. *in preparation*). We have corrected this using the magnification factor of 5.5 derived for band 3. Dunne et al. (2021) have derived the CO(1-0) calibration factor from a sample of H-ATLAS galaxies with $z \sim 0.35$ with ALMA measurements of CO(1-0), [C I] (1-0) and the continuum dust emission. By combining the measurements, they derive calibration factors for three methods of estimating the mass of molecular hydrogen. All three of these methods scale linearly with the mean gas-to-dust ratio (GDR) in the Universe, which they assume is 135, similar to the estimates of the value in the Milky Way. They derive a calibration factor $\alpha_{CO} = 3.0 \pm 0.5$ M$_\odot$ (K km s^{-1}pc^2)$^{-1}$. Using this value and including a factor of 1.36 to allow for the elements other than hydrogen gives an estimate of the mass of the ISM of $(3.3 \pm 0.7) \times 10^{11}$ M$_\odot$ (Table 5). This mass is also consistent with Cox et al. (2011) after correcting for lens magnification and using the value of α_{CO} adopted above.

4.4.2 The gas mass from [CI]

A major problem with using the CO molecule as a tracer of the ISM is the mounting evidence that a significant fraction of the molecular phase of the ISM does not contain any CO molecules, probably as the result of photo-dissociation. For example, there is evidence from *Fermi* (Abdo et al. 2010), *Planck* (Planck Collaboration et al. 2011) and *Herschel* (Pineda et al. 2013) that roughly one third of the molecular gas in the Milky Way does not contain any CO. There is also increasing evidence that the fraction of CO-dark molecular gas varies from galaxy to galaxy (Bigiel et al. 2020; Dunne et al. 2021). The [C I] (1-0) line is an attractive alternative to CO because it does not suffer from the problem of photo-dissociation, is optically thin, and has a similar critical density and excitation temperature to CO (Papadopoulos et al. 2004; Geach & Papadopoulos 2012; Tomassetti et al. 2014; Bourne et al. 2019). We have estimated the gas mass from the line flux of the [C I] (^{3}P$_1$ −^{3} P$_0$) transition, corrected with our lensing model (Table 4). Using their calibration sample (see above), Dunne et al. (2021) have derived a value for the [C I] calibration factor of $\alpha_{[CI]} = 18.8$ M$_\odot$(K km s^{-1}pc^2)$^{-1}$. The excitation function for the line (Q_{10}) varies only weakly with temperature (Dunne et al. 2021) and so we have used this calibration factor rather than trying to make any correction to it using our estimate of the temperature from the ratio of the two [CI] lines (Section 4.3). We estimate that the mass of the ISM in ID141, including all elements, is $(3.8 \pm 0.8) \times 10^{11}$ M$_\odot$ (Table 5).

4.4.3 The gas mass from [CII]

There is a strong correlation between the luminosity of the fine-structure line of [C II] at 158 μm and the mass of molecular gas in the ISM (Zanella et al. 2018). Zanella et al. derive a [C II] to H_2 conversion ratio of $\alpha_{CII} \simeq 30$ M$_\odot$/L$_\odot$. We do not have a high-resolution map of the [C II] line and therefore no specifically-derived lensing model for it. Instead, we simply used the global line flux given

by Cox et al. (2011), corrected it for the magnification factor of 5.9 we measured for the dust continuum emission in band 4, the closest in wavelength to the [C II] line, and then applied the calibration factor of Zanella et al. We obtained a total gas mass of $(3.2 \pm 0.6) \times 10^{11}$ M$_\odot$ (Table 5).

4.4.4 The gas mass from dust emission

Many studies have made the case that dust grains are a better way of tracing the ISM than the CO molecule (e.g. Eales et al. 2012; Scoville et al. 2014, 2016, 2017). Some of the advantages are that the dust grains do not suffer photo-dissociation like CO molecules, the dust emission is optically thin, and the relationship between the calibration factor and metallicity appears to be simpler (Eales et al. 2012, 2018). We used the calibration factor between dust luminosity at 850 μm and ISM mass derived by (Scoville et al. 2016, their Eq. A11):

$$\alpha_{850\mu m} = < L_\nu / M_{ISM} > 6.7 \pm 1.7 \times 10^{19} \text{ergs s}^{-1} \text{ Hz}^{-1} \text{M}_\odot^{-1} . \quad (2)$$

We used the global flux density measured by Cox et al. (2011) at 3290 μm of 1.2 ± 0.1 mJy, which makes it the closest flux measurement in rest-frame wavelength (628 μm) to the wavelength used in the calibration factor above, minimising the uncertainty in extrapolating over wavelength. We estimated the luminosity at rest-frame 850 μm, using a dust temperature of 25K and an emissivity index of ($\beta = 1.8$), the same values used by Scoville et al. (2016). We then calculated the ISM mass using the magnification factor of 5.5 derived from our band 3 lensing model and the calibration factor above, obtaining a total ISM mass of $(3.6 \pm 0.7) \times 10^{11}$ M$_\odot$ (Table 5).

4.4.5 Combining Three Methods

Dunne et al. (2021) have used several samples of calibration galaxies to derive an optimal technique for estimating the calibration factors (α_{CO}, $\alpha_{[CI]}$, α_{dust}) for an individual galaxy and for producing a best estimate of the ISM mass in the galaxy that uses all three. After correcting for lens magnification, using this technique on ID141 gives an estimate of the ISM mass of $(3.6 \pm 0.7) \times 10^{11}$ M$_\odot$ (Table 5).

4.4.6 Results

Table 5 lists the ISM mass estimates from the five different methods. We note a few important points. One caveat is that we could have used the lower value of α_{CO} that has been claimed for ultra-luminous infrared galaxies. This would lower the ISM mass estimated from CO by a factor of a few, although there are now some compelling arguments that the claim that ultra-luminous infrared galaxies have a different α_{CO} than other galaxies is incorrect (Scoville et al. 2016, Dunne et al. *in preparation*). A second caveat is that the only estimate in which the observations of the tracer have been corrected for gravitational lensing from the observations themselves is the estimate from the [C I] (1-0) line. For every other estimate, we have corrected the tracer using a total magnification value derived at some other wavelength. There is therefore the possibility of differential magnification (see, for example Serjeant 2012; Omont et al. 2013). However, the similarity between the line profiles measured in the image plane and those in the source plane, together with the similarity of the seven magnification values presented in Table 3 (from four line images and three continuum images), indicate that this is only a minor effect. We have quantified a typical size of the effect

by computing the magnification factor of a source that has twice the diameter of the band 3 source shown in Fig. 1, finding this to be only 3 per cent lower than the original magnification. Another important point is that all ISM estimates are only for the molecular phase of the ISM because of the way the methods are calibrated. Finally, da Cunha et al. (2013) and Zhang et al. (2016) have shown that the effect of the cosmic microwave background radiation will make it more difficult to detect the full extent of the ISM in high-redshift galaxies, with this effect depending on the frequency of the observations.

The mass estimates lie between 3.2×10^{11} M$_\odot$ and 3.8×10^{11} M$_\odot$, which are very large values and suggest we are unlikely to be missing much of the ISM. The agreement between the values is also quite remarkable considering that four of the methods use different tracers of the ISM and we are observing the galaxy only 1.5 billion years after the Big Bang. However, none of the methods are truly independent of each other because they are all calibrated from observations of the Milky Way. We discuss this further in Section 5.2.

4.5 The stellar mass

An additional means of establishing the evolutionary stage of a galaxy is by measuring its mass in stars. We return to the theme of stellar mass in Section 5.3 where we consider a chemical evolution model for ID141, but in this section, we outline our estimation of stellar mass of ID141 from photometry following the procedure outlined in Hopwood et al. (2011). The key ingredient is the mid-IR source flux density which we have determined using Spitzer Space Telescope observations (proposal ID. 80156) acquired with the Infra Red Array Camera (IRAC) as described below.

To begin, we measured the surface brightness profiles of the lensing mass from the K_s-band data presented in Bussmann et al. (2012) using GALFIT (Peng et al. 2002). As Bussmann et al. (2012) found, subtracting the GALFIT model light profiles from the observed K_s-band image showed negligible residual lens light and no evidence of the lensed source image. We then used these profiles to construct model IRAC 3.6 μm and 4.5 μm surface brightness profiles and subtracted these from the IRAC images to leave residuals resembling the ALMA image (see Fig. A1). We repeated this removal procedure for three other galaxies selected to have similar morphology and flux in the IRAC data and found only negligible residuals. This demonstrates that an accurate IRAC PSF was obtained and that the near-IR traces the mid-IR well in these three systems, although we are unable to rule out the possibility of there being a mis-match in the specific case of the lensing mass in ID141. A final measure of the lensed source IRAC flux was then extracted from an annulus surrounding the observed residual flux (see Fig. A1) resulting in fluxes of 8.7 ± 2.2 μJy and 8.1 ± 1.8 μJy at 3.6 μm and 4.5 μm respectively. The errors here do not include any uncertainty arising from lens light removal.

In the final part of the process, following Hopwood et al. (2011), we used version 2 of the SED fitting code MAGPHYS (Battisti et al. 2020; da Cunha et al. 2008) and supplied this with the two IRAC fluxes as described above, a limiting K_s-band flux, the far-IR fluxes as described in Section 4.2 and Wide-field Infrared Survey Explorer fluxes at 12 μm and 22 μm (WISE; Lang et al. 2016). We refer the reader to Section A for further details.

To de-magnify the quantities output by MAGPHYS, we assumed a magnification of 5.8, equal to the mean of the continuum magnifications given in Table 3. In addition, since MAGPHYS assumes a Chabrier initial mass function (Chabrier 2003), we scaled the stellar mass, dust mass and SFR output by MAGPHYS by a factor of 1.41 for comparison with the values determined in Section 4.2. This factor of 1.41 was determined from the scalings derived in Madau & Dickinson (2014).

Table 4. Emission line fluxes used in the estimation of the gas mass of ID141. The CO(7-6), H_2O and two [C I] lines are from the present work and have been determined in the source plane. Errors quoted for these lines include uncertainty due to the lens model and continuum subtraction. The CO(1-0) and [C II] lines were taken from Dunne et al. *in preparation* and Cox et al. (2011) respectively and have been corrected for lens magnification using the band 3 and band 4 magnification factors respectively.

Line	ν_{obs} (GHz)	Line Flux (Jy km s^{-1})	L ($10^9 L_\odot$)	L' (10^{10} K km s^{-1} pc^2)	Reference
CO(1-0)	21.991	0.111 ± 0.016	0.0040 ± 0.0007	8.0 ± 1.6	Dunne et al. *in preparation*
CO(7-6)	153.876	1.84 ± 0.31	0.45 ± 0.07	2.7 ± 0.5	this paper
[C I] ($^3P_2 - ^3P_1$)	154.339	0.98 ± 0.19	0.224 ± 0.045	1.3 ± 0.2	this paper
[C I] ($^3P_1 - ^3P_0$)	93.870	0.55 ± 0.10	0.076 ± 0.014	2.0 ± 0.4	this paper
[C II] ($^3P_{3/2} - ^2P_{1/2}$)	362.45	18.1 ± 2.9	10.4 ± 1.7	4.8 ± 0.8	Cox et al. (2011)
H_2O	143.436	0.70 ± 0.14	0.16 ± 0.03	1.14 ± 0.23	this paper

Table 5. Gas mass estimates determined from five different methods (see main text for details).

Method	ISM mass ($10^{11} M_\odot$)	Calibration reference
CO(1-0)	3.3 ± 0.7	Dunne et al. (2021)
[C I] (1-0)	3.8 ± 0.8	Dunne et al. (2021)
[C II] 158 μm	3.2 ± 0.7	Zanella et al. (2018)
dust	3.6 ± 0.7	Scoville et al. (2016)
CO/[C I]/dust	3.6 ± 0.7	Dunne et al. (2021)

The resulting stellar mass is $M_\star = 4.8^{+2.5}_{-2.3} \times 10^{11} M_\odot$. In addition to stellar mass, MAGPHYS also gives a dust mass of $1.8 \pm 0.3 \times 10^9 M_\odot$ and a SFR of $2100 \pm 500 M_\odot$ yr^{-1}. The 1σ errors quoted are taken from the distributions output by MAGPHYS (see Appendix A for more details). These results are consistent with those obtained in Section 4.2 based on our own SED derived from solely the far-IR photometry.

4.6 The kinematics of the source

The double-peaked nature of our highest SNR emission line profile, the H_2O line, shown in Fig. 3 suggests that either the source is a rotating disk-like system with a prominent outer ring (see, for example Stewart et al. 2014), or that it is composed of two or more individual merging components.

To better understand the dynamical characteristics of the source, we investigated the spatial extent of the kinematics using the kinematic modelling code ^{3D}Barolo (Di Teodoro & Fraternali 2015). Instead of fitting a kinematical model to the 2D velocity field, this code fits directly to the full datacube. This avoids the effect of artificial flattening of velocity fields due to inadequate spatial resolution and allows for the fitting of additional galaxy properties (i.e., velocity dispersion, disk scale height, brightness profile).

4.6.1 The kinematical model

First, we identified the signal in the data cube by using the 'SEARCH' routine of ^{3D}Barolo, which is based on the DUCHAMP source finder (Whiting 2012). In short, this routine automatically identifies the noise level of the data cube, searches each channel for spaxels that are above a user-provided SNR (here 3.5σ), creates a mask by extrapolating outwards to a second user-provided S/N threshold (here 2.0σ), and combines overlapping masks in different channels into discrete 'sources'. All sources that are too spectrally thin (i.e., possessing only one channel) or spatially compact (i.e., smaller than the synthesized beam of the data cube) are rejected. This final three-dimensional mask contains all significant emission, while excluding noise peaks that would be included using a flat S/N threshold.

With the signal identified, we next performed a preliminary analysis of the masked data cube in order to produce physically motivated estimates for further dynamical fitting. The CASA toolkit task image.moments was applied to the masked cubes to create maps of integrated intensity (zeroth-moment), line-of-sight velocity (first-moment), and velocity dispersion (second-moment). The zeroth-moment map was fitted with a two-dimensional Gaussian using the CASA toolkit task image.fitcomponents, resulting in a central position, morphological position angle, and FWHM of the minor and major axes. The initial estimate of the rotational velocity at this central position was then extracted from the first-moment map.

Before performing the fitting, it is crucial to determine the size of each ring, as well as the maximum radius of the model. In order to avoid over-resolving the data, we adopted a minimum ring width of a factor of 0.4 times the FWHM of the minor axis of the restoring beam. The factor of 0.4 here is comparable to factors used in other similar works (e.g., Talia et al. 2018; Fan et al. 2019). Similarly, we adopted a maximum model radius of a factor of 0.8 times the FWHM of the major axis of the 2D Gaussian fitted to the zeroth-moment map. The factor of 0.8 is the largest value that does not return non-physical results at large radii.

With these initial parameters and estimates in hand, we engaged the 3DFIT routine of ^{3D}Barolo to fit the masked data cube with a tilted ring model comprising a series of concentric rings which each rotate with an independent velocity. The number of rings and ring width were determined as detailed in the previous paragraph, while the model center was fixed to the central position based on the 2D Gaussian fitted to the zeroth-moment map. We assumed a thin disk (i.e., $z_o = 0.01$ arcsec). This routine begins by populating a model physical space (i.e., X, Y, Z) with gas clumps, so that these initial estimates of galaxy properties are replicated. This physical space is then translated into an observed-space data cube (i.e., RA, Dec, velocity), and the variable galaxy properties (i.e., rotational velocity, velocity dispersion, inclination, position angle, and systemic velocity) are perturbed until a minimum residual is met. This process is repeated for each ring, after which most variables are averaged, and the fitting process is repeated with a smaller number of variables (i.e., rotational velocity and velocity dispersion). This process results in best-fit values of each parameter and a model data cube that may be compared to the original data cube on a pixel-by-pixel basis.

The normalisation of the flux density within each ring can be either 'local', meaning that the integral along the spectral axis of a given model pixel in the ring must match that of the corresponding datacube pixel or, 'azimuthal', whereby the model is normalised

to the azimuthally-averaged flux in each ring. Local normalisation allows for non-axially symmetric distributions of emission across the disk that can bias fitting of azimuthally-averaged fits. In this work, we have used both normalisation types for comparison.

We applied ^{3D}Barolo to the two emission lines with the highest SNR, namely the CO(7-6) and H_2O lines.

4.6.2 *The results of the kinematic model*

By applying this tilted ring fitting code to the CO(7-6) and H_2O data cubes, we found good fits (as seen in the comparisons of Fig. 8). This implies that ID141 is a strongly rotation-dominated disk, a fact that is also apparent from the rotation curves shown in Fig. 9. The resulting dynamical masses and galaxy properties are listed in Table 6 and show close agreement with each other, regardless of the line or normalization employed (local or azimuthal). Comparing these dynamical mass estimates with our ISM mass estimates in the preceding section, we find a very significant discrepancy to the extent that the ISM masses are a factor of approximately four times those of the dynamical mass estimates. We discuss possible remedies for this disparity in Section 5.

One of the strengths of three-dimensional dynamical characterization is the ability to compare the data and best-fit model on a pixel-by-pixel basis. Inspecting Fig. 8, it is clear that both CO and H_2O are poorly fit by models using azimuthal averaging (note the strong residuals in zeroth-moment maps and spectra). However, the velocity fields, velocity dispersion maps, and general shape of each position-velocity diagram are well-reproduced. This raises an interesting issue: the kinematics of the galaxy are well-fit by a rotating disk, but the distribution of emission (i.e., morphology) is not. This feature will be discussed in Section 5.1 where we consider the possibility of ID141 being a merger of two or more galaxies.

5 DISCUSSION

5.1 One galaxy or several?

While the strong emission lines in this paper (i.e., CO and H_2O) are well-fit by ^{3D}Barolo (see Section 4.6.2) and return similar dynamical masses, the source morphologies are complex, resulting in imperfect fits. One key piece of information in interpreting this issue is the lack of symmetry between the velocity field structure to the north and south. While the velocity field to the north is a continuation of the central gradient, it extends to the east, and lacks a strong southern counterpart. When combined with the strong line emission and the H_2O dispersion peak in the northern region, this raises the possibility that the northern emission is not an extension of ID141, but a separate galaxy.

To investigate this possibility, we masked the northern section of the CO and H_2O data cubes, and fitted this "main" galaxy using ^{3D}Barolo. In each case, we assumed azimuthal normalization and the same model central position as in the unmasked data cubes. The fit to these masked data significantly improve upon the azimuthal fit to the unmasked data with a disk-like model, as shown in Fig 10.

Next, we took the best-fit models of these masked CO and H_2O data cubes and subtracted them from the unmasked data cube (CASA task immath) to isolate the signal to the north (hereafter referred to as the 'secondary'). As seen in the lower row of Fig. 10, the secondary signal shows a continuous velocity gradient. The corresponding dynamical mass of this secondary is $(3.9 \pm 0.9) \times 10^{10}\,M_\odot$ and when added to the dynamical mass of the main galaxy, increases the total dynamical

mass by approximately 50 per cent compared to the single-galaxy model. The secondary exhibits an offset velocity dispersion peak and two separate components in each zeroth-moment map. Because of these properties, ^{3D}Barolo returns a poor fit using azimuthal normalization, and a better but still unsatisfactory fit using local normalization (which we also show in Fig. 10). While ^{3D}Barolo is able to reproduce the brightness distribution and overall velocity gradient of each data cube, the significant residuals in each moment map and position-velocity diagram suggest that this secondary source is not a single, well-ordered rotating galaxy.

Taking these findings into account, there are three obvious possibilities for the nature of ID141:

(i) The system may simply be a single rotating disk. In this case, the asymmetry to the north would be a minor disturbance.

(ii) The source to the north may be a separate, rotating galaxy that is closely associated to ID141. This is supported by the continuous velocity gradients shown in the unmasked data cube (see Fig. 8) and reasonable fits to the H_2O emission from the 'main' and 'secondary' sources (see right column of Fig. 10).

(iii) Because the 'secondary' source shows two kinematically-separated clumps in both CO and H_2O emission, it is conceivable that the ID141 system is composed of three significant clumps: the main galaxy, a source to the north that is strong in CO emission, and a third source to the northeast that is strong in H_2O emission.

Due to the diversity in observed continuum (Fig, 1) and line (Fig. 2) morphologies, we are unable to ascertain which of these scenarios, if any[1], is the most likely without further observations although most bright SMGs are known to comprise mergers (e.g. Engel et al. 2010).

5.2 Why do the ISM mass estimates agree so well and why are they greater than the total mass?

The five gas masses derived in Section 4.4 (see Table 5) have a remarkably small range of $(3.2 - 3.8) \times 10^{11}\,M_\odot$, especially since four of the methods use different tracers. They are also remarkably larger than the total mass estimated from the kinematics of the galaxy; both application of ^{3D}Barolo and a simple estimate from $v^2 r/G$ with velocity $v = 430\,\mathrm{km\,s^{-1}}$ and radius $r = 2\,\mathrm{kpc}$ give a dynamical mass of $\sim 0.8 \times 10^{11}\,M_\odot$. We first consider possible observational explanations of the mass discrepancy and then suggest an astrophysical explanation of both points.

One possible explanation for the mass discrepancy is that the lens model and therefore the lens magnification is incorrect. In this case, the derived intrinsic source properties will be systematically biased. We can estimate the extent to which such an error will propagate into the ratio of gas mass to dynamical mass. Regarding the gas mass, since $M_g \propto L_{int} \propto L_{obs}/\mu$, where L_{int} and L_{obs} are the intrinsic and observed (i.e., lensed) line luminosities, the gas mass scales inversely proportionally with magnification. Under the assumption of circular orbits used by ^{3D}Barolo, the dynamical mass scales as $M_{dyn} \propto v_c^2 r_{int}$, where v_c is the circular velocity and r_{int} is some intrinsic physical length scale. Since the observed v_c is invariant under lensing and the intrinsic physical length scales as $r_{int} \propto r_{obs}/\sqrt{\mu}$, the dynamical mass scales with magnification as $M_{dyn} \propto \mu^{-1/2}$ and so $M_g/M_{dyn} \propto \mu^{-1/2}$. Therefore, the magnification we have obtained

---

[1] The strong water emission may also be indicative of an Eddington-limited starburst in which case the complex morphology may, in part, be an outflow. See van der Werf et al. (2011) for example
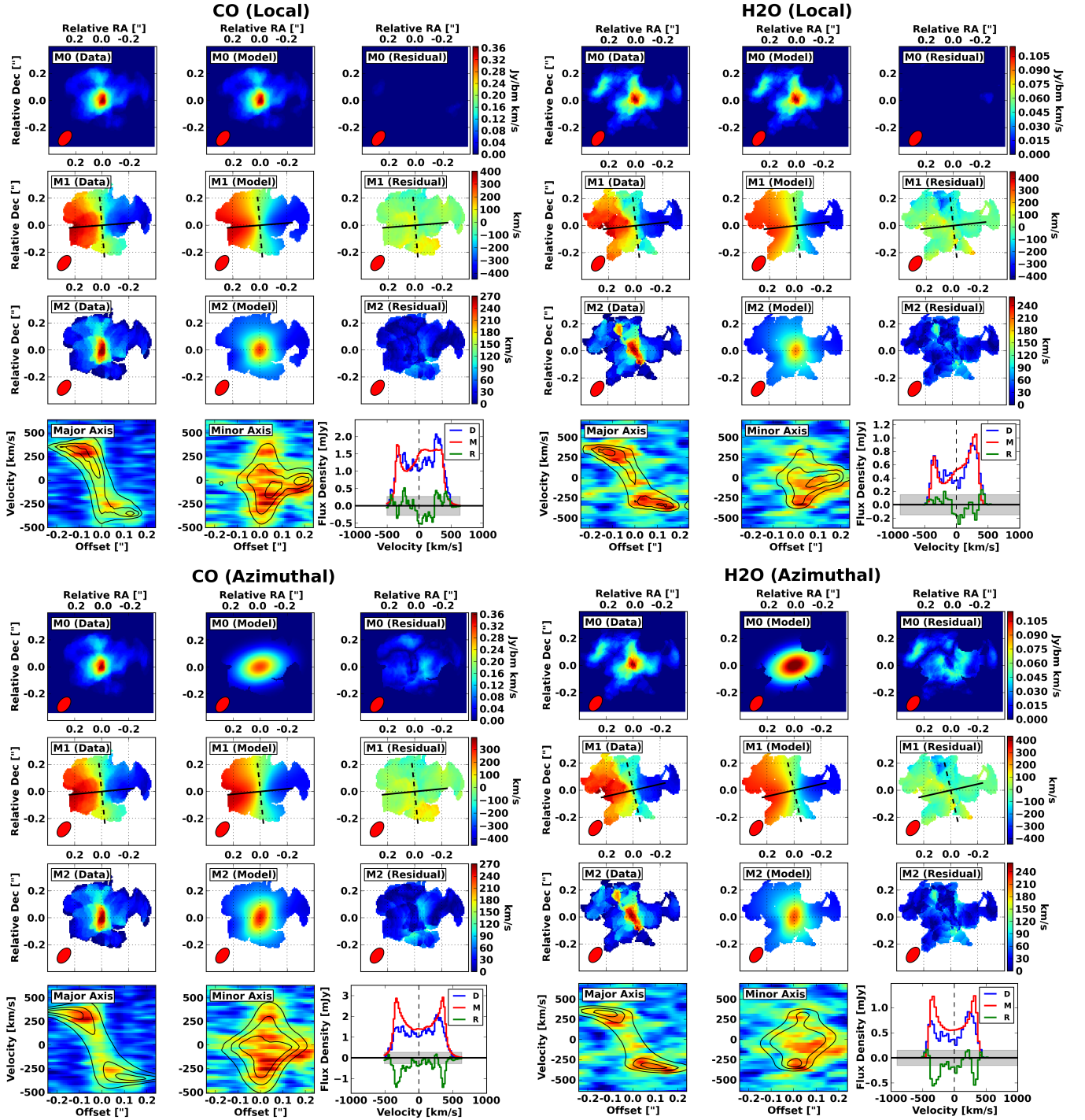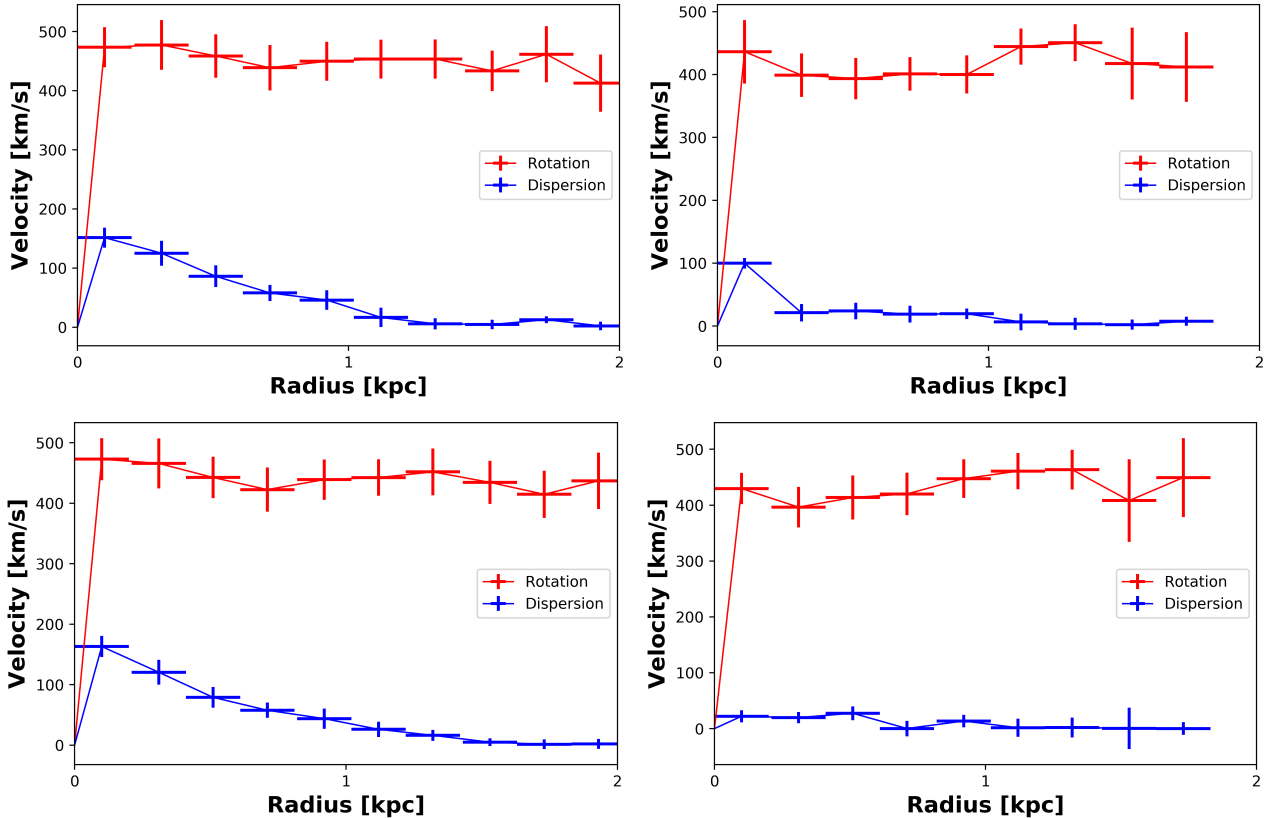
Figure 8. Comparison of observed emission lines and best-fit ^{3D}`Barolo` models for local normalization (top row) and azimuthal normalization (bottom row), as well as both CO (left column) and H$_2$O (right column). In each set of panels, we show the zeroth-moment (i.e., integrated intensity) maps (top row), first-moment (i.e., line-of-sight velocity) maps (second row; solid and dashed line indicates major and minor kinematical axis respectively), and second-moment (i.e., velocity dispersion) maps (third row) for the data (left column), model (middle column), and the residual between the two (right row). We also show the position-velocity diagrams taken along the major axis (bottom row, left column) and minor axis (bottom row, middle column) of the data cube (colour) and model cube (contours). Finally, we compare the spectra of the data (blue), model (red), and residual (green) in the lower right panel, showing the 1σ noise level by the shaded grey region.

Table 6. Best-fit morphological (from a 2-D Gaussian fit to each zeroth-moment map) and kinematic (from ^{3D}`Barolo` fitting) parameters.

Line	Normalization	Size (arcsec × arcsec)	i_M (°)	i_K (°)	PA_M (°)	PA_K (°)	$M_{dyn}(10^{10}M_\odot)$
H_2O	Local	$(0.36 \pm 0.06) \times (0.23 \pm 0.05)$	50 ± 12	61 ± 11	66 ± 17	96 ± 10	6.9 ± 1.8
H_2O	Azimuthal	–	–	57 ± 12	–	103 ± 5	8.1 ± 2.4
CO	Local	$(0.39 \pm 0.05) \times (0.33 \pm 0.04)$	33 ± 16	54 ± 10	80 ± 31	95 ± 6	7.7 ± 1.7
CO	Azimuthal	–	–	65 ± 9	0	96 ± 7	8.6 ± 2.0

Figure 9. Comparison of best-fit rotation curves (red) and velocity dispersion profiles (blue) for local normalization (top row) and azimuthal normalization (bottom row), as well as both CO (left column) and H_2O (right column).

from our lens modelling would have to be underestimated by a factor of 16 to obtain a gas mass that is equal to the dynamical mass.

Although the statistical uncertainties we quote are orders of magnitude beneath this, larger errors can arise if the lens model we have adopted is not a faithful representation of the true lensing mass. As previously stated, changing our lens model by adding an additional mass profile to match a possible second lensing galaxy only modifies the magnification by approximately 5 per cent. An alternative is to use a mass profile that does not follow a power law. Schneider & Sluse (2013) present an example of a composite lens model comprising two significantly different density profiles with a resulting magnification that differs from the power-law model by 40 per cent but still broadly reproduces the observed lensed images. Even this rather extreme example is over an order of magnitude beneath the size of shift required to reconcile our dynamical and gas mass estimates.

Another possible explanation for the discrepancy is that the total (dynamical) mass has been estimated from the CO (7-6) and water lines while the ISM estimates have been obtained from observations

of the continuum and different spectral lines. The [C I] (1-0) line, which is probably the best tracer of the cool ISM but which is too weak for us to perform a full kinematic analysis, does have a different spatial distribution from the lines used to investigate the kinematics (Fig. 2). However, the velocity extent of all three lines is very similar (Fig. 3) and since the kinematic analysis extends out to a similar radius to the [C I] (1-0) emission (Fig. 3), it is unlikely that our estimate of the total mass is too low. A related issue that has been argued in the literature (e.g., da Cunha et al. 2013; Zhang et al. 2016) is that the mass estimates of high-redshift galaxies may be too low because the increasing energy density in the cosmic microwave background with redshift makes it increasingly hard to detect extended gas. This is unlikely to remedy such a large discrepancy because, to first order, this effect simultaneously reduces our measurements of the total mass (by causing an under-estimation of dynamical radius) and the ISM mass.

An additional concern to consider is in the kinematic analysis from Section 5.1. We were ultimately unable to confirm that ID141
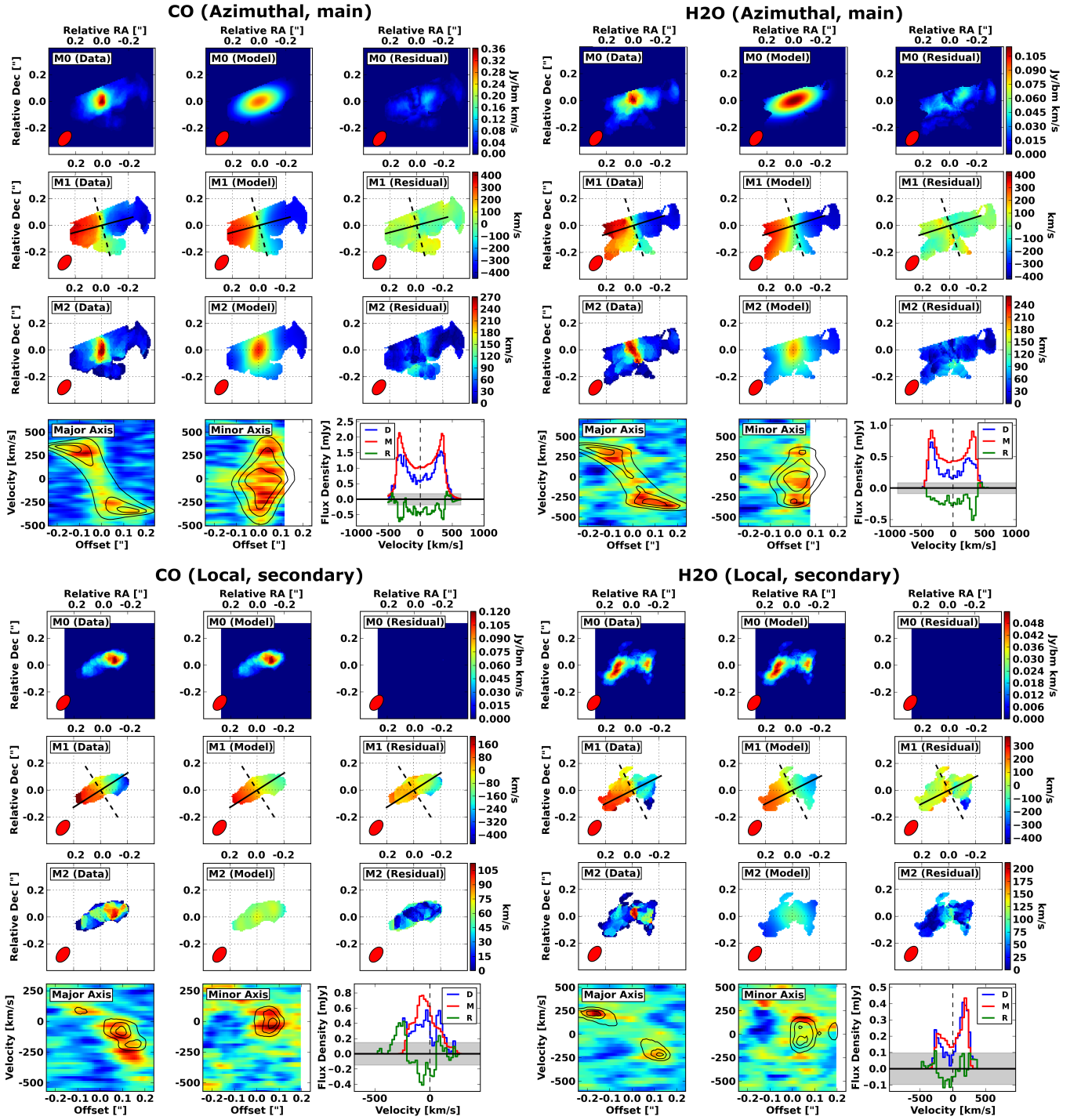
Figure 10. Comparison of observed emission lines and best-fit ^{3D}Barolo models for masked CO (left column) and H$_2$O (right column) data cubes. In each set of panels, we present the zeroth-moment (i.e., integrated intensity) maps (top row), first-moment (i.e., line-of-sight velocity) maps (second row), and second-moment (i.e., velocity dispersion) maps (third row) for the data (left column), model (middle column), and the residual between the two (right column). We also present the position-velocity diagrams taken along the major axis (bottom row, left column) and minor axis (bottom row, middle column) of the data cube (colour) and model cube (contours). Finally, we compare the spectra of the data (blue), model (red), and residual (green) in the lower right panel, showing the 1σ noise level by the shaded grey region.

is a single rotating disk rather than several galaxies. The strongly-lensed SMG SMM J02399-0136, initially identified as a large disk galaxy (Genzel et al. 2003) and subsequently resolved into a complex system of several components (Ivison et al. 2010) is an example of why caution must be taken. Nevertheless, in the case of ID141, this is unlikely to explain the large discrepancy. If we consider option (ii) in Section 5.1, that ID141 consists of two rotating disks, this gives the largest of our dynamical mass estimates of $1.3 \times 10^{11}\,M_\odot$ which is still significantly short of explaining the discrepancy.

Finally, we note that the discrepancy becomes even worse if we include stellar mass. Our estimate is that the stellar mass in this system is $M_\star = 4.8^{+2.5}_{-2.3} \times 10^{11}\,M_\odot$, which in itself is greater than the dynamical mass. However, the uncertainty on this measurement is high and furthermore, does not include any uncertainty arising from removal of the lens light.

We now consider a possible astrophysical solution, which might also partly explain the surprising agreement between the mass estimates made with four different tracers, a solution which arises because none of the methods is really independent. The dust method of Scoville et al. (2016) is calibrated using CO(1-0) measurements of galaxies and is based on the CO(1-0) calibration factor derived for giant molecular clouds in the Milky Way; the calibration factor for [C II] (Zanella et al. 2018) is based on dust measurements, which are again ultimately based on the CO(1-0) calibration factor from galactic giant molecular clouds; and all of the calibration factors discussed in Dunne et al. (2021) are based on the assumption that the GDR is the same as in the Milky Way. All the methods are ultimately benchmarked to the same astrophysical source, for the obvious reason that it is only in the Milky Way (and to a much lesser extent in galaxies in the Local Group) where it is possible to make the direct measurements of the ISM masses that are needed to calibrate the tracer methods.

A possible solution is if the calibration factors for ID141 are very different to the Galactic ones. To explain the agreement observed between the different mass estimates, all of the gas-to-tracer ratios would have to be $\simeq 4$ times lower than in the Milky Way. Since all of the tracers are composed of heavy elements, the most direct way to bring some insight to the problem is with a chemical evolution model.

5.3 Chemical Evolution in the Epoch of Galaxy Formation

In this section, we describe the results of a chemical evolution model we have constructed to investigate the likely evolution of the interstellar dust, gas and stellar mass of ID141. The details of this model can be found in Appendix B. We have used this model to investigate the evolution of the GDR. Although we have not modelled the evolution of the other tracers (gas-to-CO etc.), it is likely that the results will be very similar since all of the tracers must depend on the abundances of the metals in the galaxy. The similarity of the ISM estimates for four different tracers provides strong circumstantial evidence in support of this.

In the model, which is an extension of the model in De Vis et al. (2017), the galaxy starts as a cloud of gas with no heavy elements that is gradually converted into stars. The model incorporates gas flowing in to the galaxy via accretion from the cosmic web and out from it via stellar feedback and active galactic nuclei. Dust is produced by stars, both asymptotic giant branch stars and supernovae, and grain growth and it is destroyed by astration, supernova shocks and outflows.

We considered two star formation histories, one with continuous star formation starting at some formation redshift, z_f, and a second where we are currently observing 40 per cent of the way through a

burst lasting 200 Myr[2]. Hereafter, we refer to these as the 'conSF' and the 'SB' model.

To account for possible variations in the IMF (Appendix B), we used three functions that span most of the range of IMFs considered in the literature, namely a Salpeter IMF (Salpeter 1955), the Chabrier IMF (Chabrier 2003) and the more top-heavy IMF proposed by Zhang et al. (2018), required to explain isotope ratios at high redshifts[3]. In Section 4.2 we estimated a SFR for ID141 from its far-infrared luminosity of $2400 \pm 500\,M_\odot\,yr^{-1}$, based on the assumption of an IMF from Kroupa & Weidner (2003). If we instead use the IMFs considered here, we obtain an SFR of 2800, 1800 and 1300 $M_\odot\,yr^{-1}$ for the Salpeter, Chabrier and top-heavy IMF. In our model we set $z_f = 4.8$, 5.5 and 7 respectively (the higher observed current-day SFRs for the Salpeter and Chabrier IMFs leads to a more rapid conversion of gas into stars, and so their star formation needs to begin later). The parameters of each model have been chosen so that at the redshift of ID141 ($z = 4.24$) the SFR predicted by the model is similar to the SFR estimated for ID141 from its far-infrared luminosity.

For the conSF star formation model, we assume a SFR that follows the Kennicutt-Schmidt law (Kennicutt 1998), i.e., $\psi(t) = \epsilon M_g^k(t)$ where ϵ is the star formation efficiency, M_g is the total gas mass and k is set to unity. ϵ is estimated using $\psi = \epsilon M_g/t_{dyn}$ where t_{dyn} is the dynamical (free-fall) timescale $t_{dyn} = \sqrt{r^3/2GM}$ (Walter et al. 2009; Swinbank et al. 2010) and ψ is the observed SFR (see also Appendix B). Given an effective galaxy radius of 1.4 kpc (measured from the reconstructed source continuum) and an observed ISM mass of $3.5 \times 10^{11}\,M_\odot$, the maximum SFE is 1 per cent for the Salpeter IMF. If we instead assume that the true ISM mass must be less than this (because of the limit from the dynamical mass) and set the ISM mass equal to the dynamical mass of the one-disk model ($M_{dyn} \simeq 0.8 \times 10^{11}\,M_\odot$), the range of maximum SFE obtained is 5 and 2 per cent for the Salpeter and top-heavy IMFs respectively. Although ID141 is one of the most extreme objects in the universe in terms of its SFR, both estimates of SFE suggest that it is not forming stars at a significantly more efficient rate than normal systems. The SFE is similar to other high redshift SMGs (Swinbank et al. 2010; Riechers et al. 2013) but an order of magnitude lower than the $z = 6.4$ hyper-starburst from Walter et al. (2009). Using the current observed level of SFR and assuming that $M_g \approx M_{dyn}$ when the galaxy started forming stars, the gas reservoir would be exhausted within 40 Myr for a Salpeter IMF if there is no additional replenishment of gas via mergers or inflows.

In Table 7, we compare the model outputs predicted at the redshift of ID141 of $z = 4.24$. The Salpeter IMF models produce lower dust masses (since there are fewer high mass stars producing metals and dust and there is more gas being locked up in stellar remnants) and higher stellar masses than the Chabrier and top heavy IMF models. The resulting gas fractions are also lower, indicative of a much later evolutionary stage. Davé et al. (2011) predict gas fractions of ~ 0.4 at $z = 4$ for a current day halo mass of $M_h(z = 0) = 10^{14}\,M_\odot$. This is larger than the Salpeter IMF results but more in line with the results from the starburst, top heavy IMF model. The total baryonic mass

---

[2] The duration of high-z starbursts are thought to range from 50-250 Myr, e.g. Danielson et al. (2017); typical duty cycles of high-redshift SMGs are thought to range from 40-100 Myr (Coppin et al. 2008; Toft et al. 2014; Narayanan et al. 2015).

[3] Note that the so-called top heavy IMF in this work is different to the top heavy IMF evoked by Baugh et al. (2005) to explain number counts of bursty submillimetre galaxies (Appendix B).

$(M_g + M_*)$ predicted by the conSF models is ~ 5 times larger than the observed dynamical mass of ID141, though the starburst models predict a more comparable mass ($\sim 1 - 3\times$).

5.3.1 Evolution of the Gas to Dust Ratio

In Fig. 11 we show the evolution of the model GDR as a ratio of the Milky Way GDR, $(M_g/M_d)_{\text{model}}/(M_g/M_d)_{\text{MW}}$, for each model plotted against $f_g = M_g/(M_g + M_\star)$, the fraction of baryonic mass that is gas. For clarity, we only show and discuss the results for the Salpeter and top heavy IMFs as the results assuming a Chabrier IMF sits mid-way between the two. The shaded area around each line shows the effect of changing the parameter governing the rate of dust grain growth in the dense clouds (Appendix B) from 700 to 15,000, where the former value is appropriate for the Milky Way today and the latter indicates extremely rapid grain growth (Mattsson & Andersen 2012; Asano et al. 2013; De Vis et al. 2017). The GDR is high at high gas fractions because the abundance of metals (and therefore dust) is still low. The GDR decreases as the gas is converted into stars and as dust is formed in stellar winds, supernovae and, after reaching some critical metallicity, via grain growth.

We have assumed a value for the GDR ratio of the Milky Way that is appropriate for the Scoville et al. (2016) calibration used in Section 4.4.4 to convert dust mass and submm continuum luminosity to gas mass. We have estimated this by assuming $\beta = 1.8$, $T_d = 25$ K, $\kappa_{850} = 0.077\,\text{m}^2\,\text{kg}^{-1}$ (James et al. 2002), and $\kappa_{\text{ISM}} = 4.8 \times 10^{-4}\,\text{m}^2\,\text{kg}^{-1}$ (Scoville et al. 2014). Since the models account for the total gas, whereas the calibration of dust-to-gas (α_{850}) provided by Scoville et al. is for the molecular component only, we follow Scoville et al. (2014) and assume that there are roughly equal amounts of HI and H_2. We also account for Helium. This gives a GDR for the Milky Way of 160 given the assumed κ_{850} (a factor of ~ 1.2 higher than that assumed in Dunne et al. 2021). Table 7 shows that, for all models, the GDR reached by redshift 4.24 is < 160 (i.e., the ratio of the model GDR to the Milky Way GDR is less than unity), suggesting *it is not appropriate to assume a Milky Way calibration when estimating gas masses from the dust emission in SMGs like ID141.* Such low GDRs have been measured in other high redshift SMGs (see, for example, Berta et al. 2021; Yang et al. 2017, and references therein).

Fig. 11 also shows that the GDR can vary by over an order of magnitude during the SMG phase of a galaxy. Therefore it is also dangerous to assume a constant value for the GDR when estimating the ISM mass from the submm continuum emission. However, we can use the model to our advantage: the results in Fig. 11 provide a way to use the model itself to estimate the true GDR, and thus the mass of the ISM. We can also use it to probe how far through ID141 is in its conversion of gas into stars i.e., its evolutionary phase. Defining $M_g(\text{obs})$ to be the ISM mass we have estimated on the assumption of the Galactic GDR, the true ISM mass, including the contribution of the atomic gas, is given by:

$$M_g(\text{true}) = M_g(\text{obs})\frac{(M_g/M_d)_{\text{model}}}{(M_g/M_d)_{\text{MW}}}. \tag{3}$$

The total mass of the model galaxy is then given by:

$$M_{\text{true,tot}} = \frac{M_g(\text{true})}{f_g}. \tag{4}$$

We have made the assumption that there is no contribution from dark matter (see Genzel et al. 2017), but if there is a contribution from dark matter it will strengthen the conclusions below.

We show the evolution of $M_{\text{true,tot}}$ in Fig. 11 as dashed lines. The shaded horizontal region indicates where the total baryonic mass (gas plus stars) is consistent with the dynamical mass ($M_g + M_\star \leq M_{\text{dyn}}$) within the uncertainty, defined as $3\,\sigma$.

We can immediately draw the following conclusions:

(i) it is only possible to make the masses consistent if the IMF is top-heavy, adding to the evidence from isotopic abundances (Zhang et al. 2018) that the IMF is different in high-redshift star-forming galaxies.

(ii) It is possible to make the masses consistent when the galaxy is roughly halfway through its evolution ($f_g \sim 0.5$). Although it is often assumed that galaxies like ID141 must have very high gas fractions, the discrepancy between the mass estimates gets rapidly worse above $f_g \sim 0.6$.

We also notice that the true total mass calculated using Eq. 4 overestimates M_{dyn} in comparison to the model M_{tot} when assuming a Salpeter IMF. This is due to a combination of the faster consumption of gas due to the higher SFR and the lower dust masses formed per stellar population with a Salpeter IMF. For example, to reach a total mass $\lesssim 1 \times 10^{11}\,M_\odot$ using Eq. 4, the conSF Salpeter model would require a GDR/$f_g \lesssim 50$, or alternatively, for a galaxy with $f_g < 0.5$, a GDR of $\lesssim 25$ is needed. Such low GDRs are not reached in the Salpeter IMF models.

Although the conSF top heavy IMF appears to result in Eqn. 4 producing a total mass within the shaded region, this is misleading. The variation of the GDR ratio with f_g from this model is low enough that scaling the observed gas mass produces a true total mass from Eqn. 4 that lies below the dynamical mass limit (in the shaded region). However, with the continuous SF model, such low GDRs can only be produced in a model that has already consumed a lot of its gas and built up a large reservoir of stars, hence violating the mass criteria we are trying to satisfy (M_{tot} (conSF, top heavy) $> 4 \times 10^{11}\,M_\odot$ in Table 7).

Finally, we note that if we increase the amount of gas accreted in the model, although this would increase the gas fraction, it would also increase both the GDR and the stellar mass, hence would not solve this issue.

5.3.2 Metallicity Evolution

The low GDRs we predict for ID141 imply these galaxies will reach supersolar metallicities by $z = 4.24$ (where $Z = M_{\text{metals}}/M_g = 1.5 - 4.4\,Z_\odot$[4], Table 7) depending on the assumed IMF and SFH. To check if such high metallicities are expected in these galaxies, we next compare the evolution of the gas-to-dust ratios (Fig. 11) with metallicity in Fig. 12. We convert the model oxygen abundance using $12 + \log(\text{O/H}) = 12 + \log(X_O/X_g)$ where $X_O = M_O/16$ and $X_g = M_g/1.32$. In these units, the metallicity reached by the model assuming a top-heavy IMF is $12 + \log(\text{O/H}) = 9.33$ and 9.16 for the constant SF and starburst respectively.

Local GDR-Z scaling relations from *Herschel* surveys, e.g. Rémy-Ruyer et al. (2014); De Vis et al. (2019), are shown in Fig. 12. These relationships imply values of $12 + \log(\text{O/H})$ that are consistent with the predicted model GDRs of 25-50 for ID141. At the same metallicity as the SB model with a top heavy IMF, the Rémy-Ruyer et al. (2014) relationships predict GDRs ranging from 28-55. The properties of local galaxies from the DustPedia sample of De Vis et al. (2019) also reach low gas-to-dust ratios, but at lower metallicities than those predicted for ID141 (Fig. 12). This sample is made up mostly

---

[4] Solar metallicity taken from Asplund et al. (2009).

Model	f_g	SFR ($M_\odot\,yr^{-1}$)	M_g ($10^{11}M_\odot$)	$M_\star$ ($10^{11}M_\odot$)	M_{tot} ($10^{11}M_\odot$)	Z	M_d ($10^9 M_\odot$)	GDR
Continuous (conSF)								
Top Heavy	0.24	1336	1.16	3.66	4.83	0.06	4.10	28
Chabrier	0.20	1754	1.14	4.61	5.91	0.03	1.65	69
Salpeter	0.23	2790	1.30	4.39	5.69	0.02	1.30	100
Starburst (SB)								
Top Heavy	0.50	1300	0.79	0.80	1.60	0.04	1.84	43
Chabrier	0.25	1800	0.48	1.41	2.71	0.03	0.71	68
Salpeter	0.15	2800	0.41	2.39	2.80	0.02	0.51	80

Table 7. The values predicted by our chemical evolution models for the properties of our ID141-like galaxy at the redshift at which we observe ID141: $z = 4.24$. We give results for two different star-formation histories - starburst ('SB') and continuous ('conSF') - and for the Salpeter, Chabrier and top heavy initial mass functions. f_g is the gas fraction. GDR denotes the gas-to-dust ratio and M_{tot} is the total (gas mass + stellar mass) returned by the model. The metallicity, Z, is defined as the ratio of the metal mass and gas mass. By design, the SFR predicted by each model is similar to the SFR estimated for ID141 from its far-infrared luminosity.
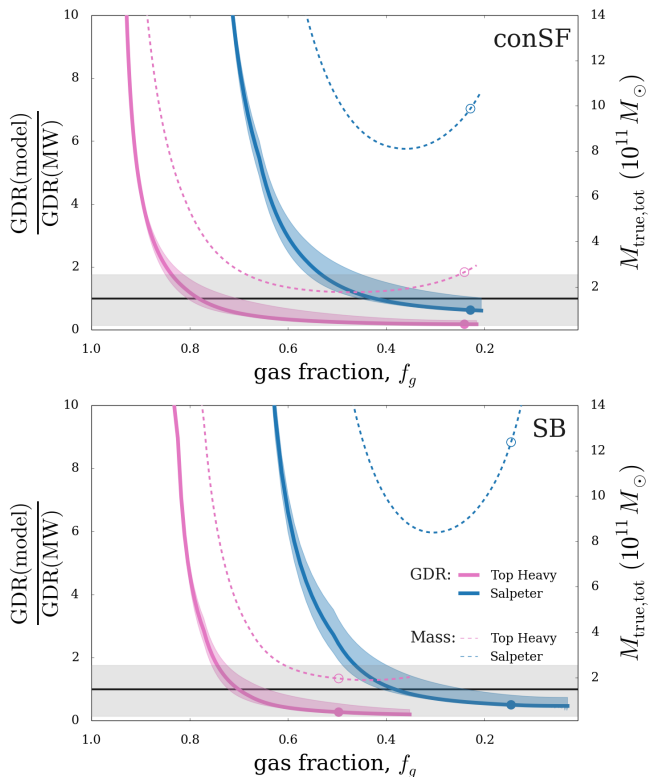
Figure 11. The gas-to-dust ratio (solid lines), as a fraction of the Milky Way value, predicted by standard dust-evolution models plotted against f_g, the fraction of baryon mass in the form of gas for *top:* the continuous star forming (conSF) model and *bottom:* the starburst (SB) model. The pink and blue colours denote the top-heavy and Salpeter IMFs respectively and the circles (filled and solid) show the values at $z = 4.24$. The dashed lines show the 'true' total mass ($M_{true,tot}$) derived from the GDRs from the models and the observed gas mass assuming a MW-like GDR (Eqn. 4) as the galaxy evolves. The black horizontal line indicates a GDR equal to the Milky Way. The shaded grey region indicates $M_g + M_\star \lesssim M_{dyn}$, where we use the maximal value ($M_{dyn} = 1.3 \times 10^{11}\,M_\odot$) with a $3\,\sigma$ uncertainty.

of late type galaxies, with stellar masses $> 10^8\,M_\odot$ making up only 4 per cent.

We also show the relationship from Santini et al. (2014) derived by combining observed dust masses of galaxies out to $z = 2.5$ with metallicities derived from the stellar mass-SFR-metallicity relationship of Mannucci et al. (2010). Although a number of mass-metallicity relations exist in the literature (eg Bothwell et al. 2013; Peng et al. 2015, Maiolino & Mannucci 2019 and references therein) we caution here that there are very few studies carried out at high enough stellar mass, at redshifts > 3, and with SFRs $> 100\,M_\odot\,yr^{-1}$, making it difficult to apply these relationships to ID141.

What we can say, is that if we assume that ID141 will shortly turn into a massive quiescent early-type galaxy with very little future evolution, then our models predict high mass galaxies with super solar metallicities $\sim 1.3 - 3\,Z_\odot$ at present day. This is in agreement with observations of massive ellipticals in the local universe (Gallazzi et al. 2005; Peng et al. 2015, see also Maiolino & Mannucci 2019, their Fig. 12).

5.4 Broader Implications of Low Gas-to-Dust Ratios

The conflict between dynamical mass and ISM mass when estimating the latter using CO has been previously reported in the literature (e.g., Tacconi et al. 2008; Hodge et al. 2012; Bothwell et al. 2013; Calistro Rivera et al. 2018). A common conclusion is to reduce the CO conversion factor, $\alpha_{CO} \simeq 1$ (but, in contrast, see Boogaard et al. 2021). However, even with this low value, the tension between the dynamical mass and the ISM mass is sometimes still seen, requiring an additional explanation (e.g. in changing the structure of the galaxy Bothwell et al. 2013). Low values of α_{CO} seemed plausible given the evidence for a low value in low-redshift Ultraluminous Infrared Galalaxies (ULIRGs) (Downes & Solomon 1998). More recently, this evidence has weakened (Papadopoulos et al. 2012b,a, also see Dunne et al. *in preparation*), which makes the low α_{CO} explanation of the mass tension given in these previous studies less likely. The advantage of this work is that we have used four different tracers of the ISM: [C I], [C II], dust and CO. Therefore, although we might resolve the tension in our ISM estimate from the CO tracer by reducing α_{CO} by a factor of ~ 4, we would also have to reduce the three other calibration factors by the same amount.

Given the quite startling agreement between the ISM mass estimates for ID141, which has been noticed for other galaxies (e.g., Dye
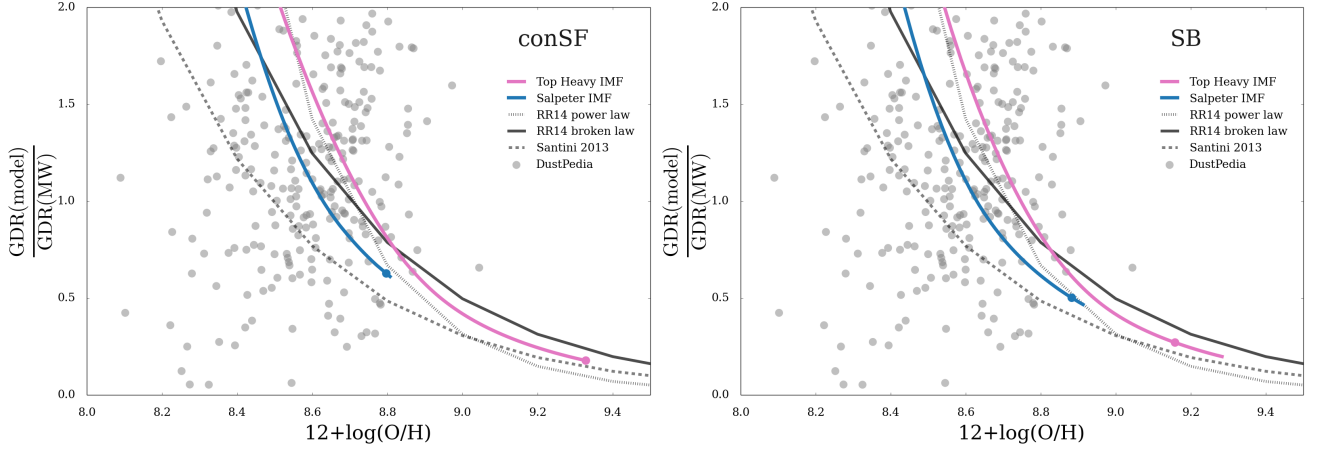
Figure 12. The gas-to-dust ratio (solid lines), as a fraction of the Milky Way value, predicted by the dust-evolution models as a function of metallicity. *Left:* the continuous star forming (conSF) model and *right*: the starburst (SB) model. The pink and blue colours denote the top-heavy and Salpeter IMFs respectively and the coloured circles show the values at $z = 4.24$. The black lines show the predictions from (i) Rémy-Ruyer et al. (2014) (their empirical fits to GDR and metallicity for local galaxies, note that their broken power laws calibrated to MW $X_{CO,MW}$ or assuming $X_{CO,Z}$ are the same in this regime), (ii) Santini et al. (2014). The grey scatter points show the observed values for 454 local galaxies with metallicity measurements from DustPedia (De Vis et al. 2019)(assuming metallicities calibrated using the O3N2 method, and gas masses derived using their Eqs. 7-9).

et al. 2015), it seems more natural to assume that the true explanation of the mass discrepancy is due to chemical evolution rather than the assumption that the physical conditions in the ISM are producing a low value of α_{CO}. The former naturally explains why the calibration factors for all four gas tracers are lower in ID141.

This may seem counter-intuitive at first since galaxies in the first few billion years might be expected to have higher GDRs than in low-redshift galaxies because there has been less time for heavy elements to form (e.g., see Péroux & Howk 2020). There is one strong argument why, in this situation, common sense is probably wrong. All the high-redshift galaxies observed have been found because they contain a lot of dust or because they already contain a large mass of stars (Scoville et al. 2014, 2017; Tacconi et al. 2018; Millard et al. 2020), and so there is a strong selection effect that will mean that samples of high-redshift galaxies are likely to contain galaxies in which the stellar mass or dust mass is already high. The simplest closed-box models for the formation of dust (Eales & Edmunds 1996, Dunne et al. 2003), for example, predict that the dust mass in a galaxy is at a peak when $f_g \sim 0.5$, and thus the galaxies in a sample compiled from a submm continuum survey are likely to have high dust masses and to have already turned much of their gas into stars.

If our conclusion that the gas-to-tracer ratio is $\simeq 4$ times lower than in the Milky Way is true for high redshift galaxies in general, it would have a dramatic effect on our assessment of their evolutionary states, reducing all ISM mass estimates by a factor of 4. However, the variation in the GDR seen in our chemical-evolution model (Fig. 11) shows that calibrating all the tracers using ID141 would be just as risky as using the Galactic calibrations. The indication from Fig. 11 is that there is no perfect calibration, and it is better to think of a gas-to-tracer ratio as being a new useful tool for assessing the evolutionary state of a galaxy.

5.5 How quickly will ID141 become a quiescent galaxy?

In this section, we consider the evolution of ID141 itself. In particular, we examine whether the evolution might be fast enough for it to join the population of quiescent galaxies that is seen at $z \sim 3$ (Glazebrook et al. 2017; Girelli et al. 2019; Valentino et al. 2020; Forrest et al.

2020b,a; Merlin et al. 2019). For reference, the time interval between the two redshifts is 7.2×10^8 yr.

To begin addressing this question, we turn to our kinematic analysis (Section 4.6) which shows evidence for a large rotating disk, with possibly one or two additional merging systems. Since high-redshift galaxies are often separated into two classes - disks or mergers - the kinematics of ID141 suggest that both are often going on at the same time. The escape velocity of the system is given by:

$$v_{esc} = \sqrt{\frac{2GM}{R}}. \tag{5}$$

Given a dynamical mass of 1.3×10^{11} $M_\odot$ and a radius of 1.4 kpc, the escape velocity is ~ 910 km s^{-1}. This is similar to the velocity range of the system (Fig. 3), which suggests that most of the material we see is destined to end up in a single galaxy.

There are two other observational results that are relevant to the dynamical evolution of the system. First, the [C I] (1-0) image (Fig. 1), which likely gives us the most unbiased view of the gas, shows that the gas is divided into a number of clumps. The existence of clumps in cleaned interferometric data needs to be treated with care (e.g., see Ivison et al. 2020) since there are observational effects that can lead to them being falsely identified. However, when clumps are multiply-imaged in a strong lensing system such as ID141, these observational effects are mitigated. A particularly striking example of this is in the system SDP81 which unequivocally shows multiply-imaged clumps in high resolution, high signal-to-noise ALMA data (Dye et al. 2015).

Clumps are expected to migrate into the centre of the galaxy through dynamical friction on a timescale given by (Genzel et al. 2011):

$$t_{inspiral} \simeq \left(\frac{v_c}{\sigma_0}\right)^2 t_{dyn}, \tag{6}$$

where σ_0 is the velocity dispersion and v_c is the circular velocity. The ratio of v_c and σ_0 is $\simeq 10$, giving a migration timescale of $t_{inspiral} \simeq 100 t_{dyn} \simeq 2.3 \times 10^8$ yr.

Alternatively, we can consider the dynamical evolution of the system by calculating the Toomre Q-parameter (Toomre 1964), which quantifies the stability of rotating, thin symmetric disks against grav-

itational collapse (Toomre 1964). For a flat rotation curve, the Q-parameter is given by (Genzel et al. 2011):

$$Q_{\text{gas}} = \sqrt{2} \, \frac{\sigma_0}{v_c} \, \frac{M_{\text{tot}}}{M_{\text{gas}}} \, . \tag{7}$$

Using the results of our chemical evolution model, we assume $M_{\text{gas}}/M_{\text{tot}} \simeq 0.5$. With this value of the mass ratio, the results from the kinematical modelling of the CO line (Fig. 9 - left panels) imply that $Q \simeq 0.6$ in the central few hundred parsecs of the galaxy, falling to a lower value beyond a radius of 1 kpc. The results from the kinematical modelling of the H_2O line (Fig. 9 - right panels) imply much lower values of Q_{gas}. Our results therefore show that the gas disk is unstable. We reached the same conclusion for the disk of the other gravitationally-lensed *Herschel* source that has been observed at high resolution with ALMA, SDP81.

All of these results show that the dynamic evolution of ID141 is very fast and suggest that it could easily turn into a single fairly undisturbed galaxy by $z \sim 3$.

Turning now to the evolution of ID141's gas, the only chemical evolution models that produce galaxies consistent with the tight mass limits are those with a top-heavy IMF and a starburst (Table 7). The conSF model has a gas mass and star-formation rate at $z = 4.24$ of 1.16×10^{11} $M_\odot$ and 1336 $M_\odot$ yr^{-1} respectively, but the model predicts a stellar mass greater than the dynamical mass limit. However, if we assume that the galaxy continues to form stars at the same rate, and that the galaxy accretes no fresh gas, the gas will be consumed in $\simeq 9 \times 10^7$ yr. If we perform the same calculation with the SB model, we obtain a timescale of $\simeq 6 \times 10^7$ yr.

In conclusion, both the dynamics and the other properties of ID141 are consistent with the idea that, although it is an extreme system at $z = 4.24$, by a redshift of $z \sim 3$ it will have evolved into a single, fairly undisturbed galaxy, containing little gas.

6 CONCLUSIONS

ID141 is a gravitationally-lensed hyper-luminous SMG at a redshift of 4.24 with a star-formation rate of 2400 $M_\odot$ yr^{-1}, making it an extreme example of an already extreme class. The system was chosen as a target for a comprehensive high-resolution investigation of the phases of the ISM in high-redshift galaxies because at its redshift, the key ISM spectral lines fall in the ALMA bands. In this paper we have presented high-resolution observations in the [C I](1-0) line, the [C I](2-1) line, the CO(7-6) line and the $H_2O(2_{1,1} - 2_{0,2})$ lines. After the lensing reconstruction, the images in the source plane reach down to a resolution of $\simeq 350$ pc. In this section, we outline the observational results and our inferences from these results.

Our observational results are:

(i) The gas is distributed over a region of $\simeq 3$ kpc. The [C I](1-0) emission is offset by $\simeq 1$ kpc from the other lines, including the [C I](2-1) line. This shows that the galaxy contains a large reservoir of cool gas that is missed by the other spectral lines that we have observed with ALMA, although the estimates of the ISM mass from the [C I](1-0) and CO(1-0) lines are remarkably similar (see below).

(ii) We have used the ratio of the [C I](2-1) and [C I](1-0) lines to measure the variation in excitation temperature over the galaxy. We find that the temperature increases from $T_{\text{ex}} \sim 40$K where the [C I](1-0) emission peaks to $T_{\text{ex}} \gtrsim 100$K where the CO and water lines are at a peak.

(iii) The gas disk appears to be broken into clumps.

(iv) We have used the CO and water lines to carry out a kinematic analysis of the galaxy. We find strong evidence of a rotating system

but cannot distinguish between there being a single rotating galaxy or there also being one or two additional galaxies. Our estimate of the dynamical mass depends on which of these scenarios we choose, but our estimates lie in the range $0.8 - 1.3 \times 10^{11}$ $M_\odot$.

(v) We have estimated the ISM mass in the galaxy five ways: (1) from the [C I](1-0); (2) from the dust continuum emission: (3) from a measurement of the global CO(1-0) flux; (4) from a measurement of the global flux in the [C II]158-μm line; (5) from a method that combines (1), (2) and (3). Our mass estimates lie in the range $3.2 - 3.8 \times 10^{11}$ $M_\odot$.

(vi) After subtracting the lens from a *Spitzer* image, we estimate that the stellar mass of ID141 is $M_\star = 4.8^{+2.5}_{-2.3} \times 10^{11}$ $M_\odot$.

From these basic observational results we make the following inferences:

(i) All five of our ISM estimates are very similar and are substantially larger than the dynamical mass of the system (without even considering the stellar mass and any dark matter). Inconsistencies like this have been observed before and have always been attributed to a low value of α_{CO}, which is then attributed to physical conditions in the ISM of an SMG probably being different from the Milky Way. This can not be the explanation in ID141 because we are using four different tracers, each of which will be affected by different physical conditions in the ISM in different ways. The similarity of the ISM mass estimates using the four different tracers suggests that chemical evolution could explain the mass discrepancy because the abundances of all the tracers will depend on the abundances of heavy elements in the galaxy.

(ii) We have used a chemical evolution model to show that the GDR does change substantially during the evolution of an SMG. It can reach values much lower than the Milky Way value as long as the IMF is top-heavy in SMGs for which there is now good evidence from spectroscopy of molecules containing rare isotopes (Zhang et al. 2018).

(iii) Although there is a complex web of calibrations, all estimates of the ISM masses in high-redshift galaxies are ultimately calibrated in the Milky Way. Our conclusion that the gas-to-tracer ratio for four different tracers is four times lower than in the Milky Way puts all these estimates into doubt. On a positive note, we suggest that the gas-to-tracer ratio may be a useful new tool for assessing the evolutionary state of a galaxy.

(iv) We have used ID141 to test whether extreme SMGs like this might evolve into the population of quiescent galaxies that is seen at $z \sim 3$. Several indicators (the Toomre Q-parameter, the inspiral time for the clumps, and the escape velocity system of the system) suggest that ID141 should be a single, fairly undisturbed galaxy by $z \sim 3$. The depletion time for the gas also implies that, on the assumption of no gas infall, the galaxy will be devoid of gas by this redshift. Therefore, extreme though it is at $z = 4.2$, it seems plausible that ID141 will evolve into an unexceptional galaxy by $z \sim 3$.

DATA AVAILABILITY

All data presented in this paper can be obtained from the ALMA science archive under the project IDs 2013.1.00358.S, 2016.1.00450.S and 2017.1.00029.S.

REFERENCES

ALMA Partnership et al., 2015, ApJ, 808, L4
Abdo A. A., et al., 2010, ApJ, 710, 133
Asano R. S., Takeuchi T. T., Hirashita H., Inoue A. K., 2013, Earth, Planets, and Space, 65, 213
Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, ARA&A, 47, 481
Bakx T. J. L. C., et al., 2018, MNRAS, 473, 1751
Bakx T. J. L. C., et al., 2020, MNRAS, 494, 10
Battisti A. J., da Cunha E., Grasha K., Cosmos Collaboration 2020, in American Astronomical Society Meeting Abstracts #236. p. 110.05
Baugh C. M., Lacey C. G., Frenk C. S., Granato G. L., Silva L., Bressan A., Benson A. J., Cole S., 2005, MNRAS, 356, 1191
Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57
Belfiore F., Vincenzo F., Maiolino R., Matteucci F., 2019, MNRAS, 487, 456
Berta S., et al., 2021, A&A, 646, A122
Bigiel F., et al., 2020, ApJ, 903, 30
Birkin J. E., et al., 2021, MNRAS, 501, 3926
Bolatto A. D., Wolfire M., Leroy A. K., 2013, ARA&A, 51, 207
Boogaard L. A., et al., 2021, arXiv e-prints, p. arXiv:2105.12489
Bothwell M. S., Maiolino R., Kennicutt R., Cresci G., Mannucci F., Marconi A., Cicone C., 2013, MNRAS, 433, 1425
Bourne N., Dunlop J. S., Simpson J. M., Rowland s K. E., Geach J. E., McLeod D. J., 2019, MNRAS, 482, 3135
Bussmann R. S., et al., 2012, ApJ, 756, 134
Bussmann R. S., et al., 2013, ApJ, 779, 25
Cañameras R., et al., 2015, A&A, 581, A105
Cañameras R., et al., 2018, A&A, 620, A61
Cai Z.-Y., De Zotti G., Bonato M., 2020, ApJ, 891, 74
Calistro Rivera G., et al., 2018, ApJ, 863, 56
Casey C. M., Narayanan D., Cooray A., 2014, Phys. Rep., 541, 45
Chabrier G., 2003, PASP, 115, 763
Cheng C., et al., 2020, ApJ, 898, 33
Coppin K. E. K., et al., 2008, MNRAS, 389, 45
Cox P., et al., 2011, ApJ, 740, 63
Danielson A. L. R., et al., 2017, ApJ, 840, 78
Davé R., Finlator K., Oppenheimer B. D., 2011, MNRAS, 416, 1354
De Vis P., et al., 2017, MNRAS, 471, 1743
De Vis P., et al., 2019, A&A, 623, A5
Di Teodoro E. M., Fraternali F., 2015, MNRAS, 451, 3021
Downes D., Solomon P. M., 1998, ApJ, 507, 615
Dunne L., Eales S. A., Edmunds M. G., 2003, MNRAS, 341, 589
Dunne L., et al., 2011, MNRAS, 417, 1510
Dunne L., Maddox S. J., Vlahakis C., Gomez H. L., 2021, MNRAS, 501, 2573
Dye S., et al., 2015, MNRAS, 452, 2258
Dye S., et al., 2018, MNRAS, 476, 4383
Eales S. A., Edmunds M. G., 1996, MNRAS, 280, 1167
Eales S., et al., 2010, PASP, 122, 499
Eales S., et al., 2012, ApJ, 761, 168
Eales S., et al., 2018, MNRAS, 473, 3507
Engel H., et al., 2010, ApJ, 724, 233
Enia A., et al., 2018, MNRAS, 475, 3467
Fakhouri O., Ma C.-P., Boylan-Kolchin M., 2010, MNRAS, 406, 2267
Fan L., Knudsen K. K., Han Y., Tan Q.-h., 2019, ApJ, 887, 74
Forbes J. C., Krumholz M. R., Burkert A., Dekel A., 2014, MNRAS, 438, 1552
Forrest B., et al., 2020a, ApJ, 890, L1
Forrest B., et al., 2020b, ApJ, 903, 47
Fudamoto Y., et al., 2017, MNRAS, 472, 2028
Gallazzi A., Charlot S., Brinchmann J., White S. D. M., Tremonti C. A., 2005, MNRAS, 362, 41
Geach J. E., Papadopoulos P. P., 2012, ApJ, 757, 156
Genzel R., Baker A. J., Tacconi L. J., Lutz D., Cox P., Guilloteau S., Omont A., 2003, ApJ, 584, 633
Genzel R., et al., 2011, ApJ, 733, 101
Genzel R., et al., 2017, Nature, 543, 397
Girelli G., Bolzonella M., Cimatti A., 2019, A&A, 632, A80
Glazebrook K., et al., 2017, Nature, 544, 71
Hodge J. A., da Cunha E., 2020, Royal Society Open Science, 7, 200556
Hodge J. A., Carilli C. L., Walter F., de Blok W. J. G., Riechers D., Daddi E., Lentati L., 2012, ApJ, 760, 11
Hopwood R., et al., 2011, ApJ, 728, L4
Huynh M. T., et al., 2017, MNRAS, 467, 1222
Inoue A. K., 2003, PASJ, 55, 901
Ivison R. J., Smail I., Papadopoulos P. P., Wold I., Richard J., Swinbank A. M., Kneib J. P., Owen F. N., 2010, MNRAS, 404, 198
Ivison R. J., Papadopoulos P. P., Smail I., Greve T. R., Thomson A. P., Xilouris E. M., Chapman S. C., 2011, MNRAS, 412, 1913
Ivison R. J., Richard J., Biggs A. D., Zwaan M. A., Falgarone E., Arumugam V., van der Werf P. P., Rujopakarn W., 2020, MNRAS, 495, L1
James A., Dunne L., Eales S., Edmunds M. G., 2002, MNRAS, 335, 753
Jarugula S., et al., 2019, ApJ, 880, 92
Kaasinen M., et al., 2019, ApJ, 880, 15
Kassiola A., Kovner I., 1993, ApJ, 417, 450
Kennicutt Robert C. J., 1998, ApJ, 498, 541
Kennicutt R. C., Evans N. J., 2012, ARA&A, 50, 531
Kroupa P., Weidner C., 2003, ApJ, 598, 1076
Lang D., Hogg D. W., Schlegel D. J., 2016, AJ, 151, 36
Liu L., et al., 2017, ApJ, 846, 5
Madau P., Dickinson M., 2014, ARA&A, 52, 415
Maiolino R., Mannucci F., 2019, A&ARv, 27, 3
Mannucci F., Cresci G., Maiolino R., Marconi A., Gnerucci A., 2010, MNRAS, 408, 2115
Mattsson L., Andersen A. C., 2012, MNRAS, 423, 38
McBride J., Fakhouri O., Ma C.-P., 2009, MNRAS, 398, 1858
McMullin J. P., Waters B., Schiebel D., Young W., Golap K., 2007, in Shaw R. A., Hill F., Bell D. J., eds, Astronomical Society of the Pacific Conference Series Vol. 376, Astronomical Data Analysis Software and Systems XVI. p. 127
Merlin E., et al., 2019, MNRAS, 490, 3309
Millard J. S., et al., 2020, MNRAS, 494, 293
Mizukoshi S., et al., 2021, arXiv e-prints, p. arXiv:2105.08894
Moss A., 2020, MNRAS, 496, 328
Narayanan D., et al., 2015, Nature, 525, 496

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

Negrello M., et al., 2017, MNRAS, 465, 3558

Neistein E., Dekel A., 2008, MNRAS, 383, 615

Nightingale J. W., Dye S., Massey R. J., 2018, MNRAS, 478, 4738

Omont A., et al., 2013, A&A, 551, A115

Papadopoulos P. P., Greve T. R., 2004, ApJ, 615, L29

Papadopoulos P. P., Thi W. F., Viti S., 2004, MNRAS, 351, 147

Papadopoulos P. P., van der Werf P. P., Xilouris E. M., Isaak K. G., Gao Y., Mühle S., 2012a, MNRAS, 426, 2601

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

Peng C. Y., Ho L. C., Impey C. D., Rix H.-W., 2002, AJ, 124, 266

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

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

Pineda J. L., Langer W. D., Velusamy T., Goldsmith P. F., 2013, A&A, 554, A103

Planck Collaboration et al., 2011, A&A, 536, A19

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

Rémy-Ruyer A., et al., 2014, A&A, 563, A31

Reuter C., et al., 2020, ApJ, 902, 78

Riechers D. A., et al., 2013, Nature, 496, 329

Romano D., Matteucci F., Zhang Z. Y., Papadopoulos P. P., Ivison R. J., 2017, MNRAS, 470, 401

Rowan-Robinson M., 2000, MNRAS, 316, 885

Rowlands K., Gomez H. L., Dunne L., Aragón-Salamanca A., Dye S., Maddox S., da Cunha E., van der Werf P., 2014, MNRAS, 441, 1040

Rybak M., McKean J. P., Vegetti S., Andreani P., White S. D. M., 2015a, MNRAS, 451, L40

Rybak M., Vegetti S., McKean J. P., Andreani P., White S. D. M., 2015b, MNRAS, 453, L26

Salpeter E. E., 1955, ApJ, 121, 161

Santini P., et al., 2014, A&A, 562, A30

Sault R. J., Teuben P., Wright M. C. H., 2011, MIRIAD: Multi-channel Image Reconstruction, Image Analysis, and Display (ascl:1106.007)

Schneider P., Sluse D., 2013, A&A, 559, A37

Scoville N., et al., 2014, ApJ, 783, 84

Scoville N., et al., 2016, ApJ, 820, 83

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

Serjeant S., 2012, MNRAS, 424, 2429

Smith D. J. B., et al., 2013, MNRAS, 436, 2435

Solomon P. M., Vanden Bout P. A., 2005, ARA&A, 43, 677

Spilker J. S., et al., 2014, ApJ, 785, 149

Stewart I. M., Blyth S. L., de Blok W. J. G., 2014, A&A, 567, A61

Stutzki J., et al., 1997, ApJ, 477, L33

Swinbank A. M., et al., 2010, Nature, 464, 733

Swinbank A. M., et al., 2015, ApJ, 806, L17

Tacconi L. J., et al., 2008, ApJ, 680, 246

Tacconi L. J., et al., 2018, ApJ, 853, 179

Talia M., et al., 2018, MNRAS, 476, 3956

Toft S., et al., 2014, ApJ, 782, 68

Tomassetti M., Porciani C., Romano-Diaz E., Ludlow A. D., Papadopoulos P. P., 2014, MNRAS, 445, L124

Toomre A., 1964, ApJ, 139, 1217

Valentino F., et al., 2020, ApJ, 889, 93

Vieira J. D., et al., 2013, Nature, 495, 344

Walter F., Riechers D., Cox P., Neri R., Carilli C., Bertoldi F., Weiss A., Maiolino R., 2009, Nature, 457, 699

Warren S. J., Dye S., 2003, ApJ, 590, 673

Wechsler R. H., Tinker J. L., 2018, ARA&A, 56, 435

Whiting M. T., 2012, MNRAS, 421, 3242

Yang C., Gao Y., Omont A., Liu D., Isaak K. G., Downes D., van der Werf P. P., Lu N., 2013, ApJ, 771, L24

Yang C., et al., 2016, A&A, 595, A80

Yang C., et al., 2017, A&A, 608, A144

Yoshida N., Stoehr F., Springel V., White S. D. M., 2002, MNRAS, 335, 762

Zanella A., et al., 2018, MNRAS, 481, 1976

Zhang Z.-Y., Papadopoulos P. P., Ivison R. J., Galametz M., Smith M. W. L., Xilouris E. M., 2016, Royal Society Open Science, 3, 160025

Zhang Z.-Y., Romano D., Ivison R. J., Papadopoulos P. P., Matteucci F., 2018, Nature, 558, 260

Zhukovska S., 2014, A&A, 562, A76

da Cunha E., Charlot S., Elbaz D., 2008, MNRAS, 388, 1595

da Cunha E., et al., 2013, ApJ, 766, 13

van der Werf P. P., et al., 2011, ApJ, 741, L38

APPENDIX A: STELLAR MASS ESTIMATION

Our estimate of the stellar mass of ID141 follows the procedure used by Hopwood et al. (2011) who used MAGPHYS to fit an SED to observed optical, mid-IR and far-IR photometry. In this paper, we have used version 2 of MAGPHYS (see Battisti et al. 2020).

The dominant constraints on the SED in the mid-IR region are provided by Spitzer Space Telescope data which were obtained under proposal ID 80156 using the Infra Red Array Camera (IRAC) operating at $3.6\,\mu m$ and $4.5\,\mu m$. To obtain a measure of the source emission, it was first necessary to remove the lens emission since the two are blended due to the relatively large IRAC PSF. This was done by first determining the surface brightness profiles of the two foreground galaxies using the K_s-band data presented in Bussmann et al. (2012) using GALFIT (Peng et al. 2002). Subtracting the model profiles of these two well-resolved foreground galaxies which lie well within the lensed source light observed by ALMA left only negligible residuals and no evidence of any source emission. We then used these profiles to construct model surface brightness profiles of the lens galaxies as they would be observed by IRAC at $3.6\,\mu m$ and $4.5\,\mu m$. To achieve this, we extracted the PSF at both wavelengths using five stars in the field of view of the IRAC observations and convolved these with the K_s-band profiles. Subtracting these model light profiles from the observed IRAC profiles left residuals which resemble the lensed source image as seen by ALMA (see Fig. A1).

To test whether these residuals were the result of a poorly-estimated PSF or because the mid-IR does not closely follow the near-IR light, we repeated the procedure for three galaxies of similar flux and morphology selected at random within the field of view common to both the K_s-band and IRAC observations. Removing the model IRAC profiles left no significant residuals in any of the three galaxies we tested. We therefore assumed that the residuals left from removing the lens profiles were indeed the source image although we can not rule out the possibility that the light profile of the lens in the near-IR is different to that in the mid-IR in the ID141 system. To obtain a final measure of the lensed source IRAC flux, we summed the residuals within an annulus surrounding the observed residual flux as shown in Fig. A1. This resulted in fluxes of $8.7 \pm 2.2\,\mu Jy$ and $8.1 \pm 1.8\,\mu Jy$ at $3.6\,\mu m$ and $4.5\,\mu m$ respectively. The errors quoted here are estimated from the image noise and do not include any uncertainty arising from lens light removal.

To constrain the far-IR part of the SED fitted by MAGPHYS, we used the photometry as discussed in Section 4.2. The optical/near-IR part of the SED was constrained using the Keck K_s-band data. Since the lensed source was not detected in the lens-subtracted K_s-band image, we set the flux in this band to be three times the standard deviation of the background flux and assumed a generous 1σ error equal to this value. In this way, the K_s flux acts like an upper limit in the SED fitting. Finally, we also included Wide-field Infrared Survey Explorer fluxes at $12\,\mu m$ and $22\,\mu m$ (WISE; Lang et al. 2016). Since the WISE fluxes are an unknown combination of the lens and source light, we attributed 50 per cent of the flux in each passband to the source and took a 1σ error equal to this value so that again, the WISE data
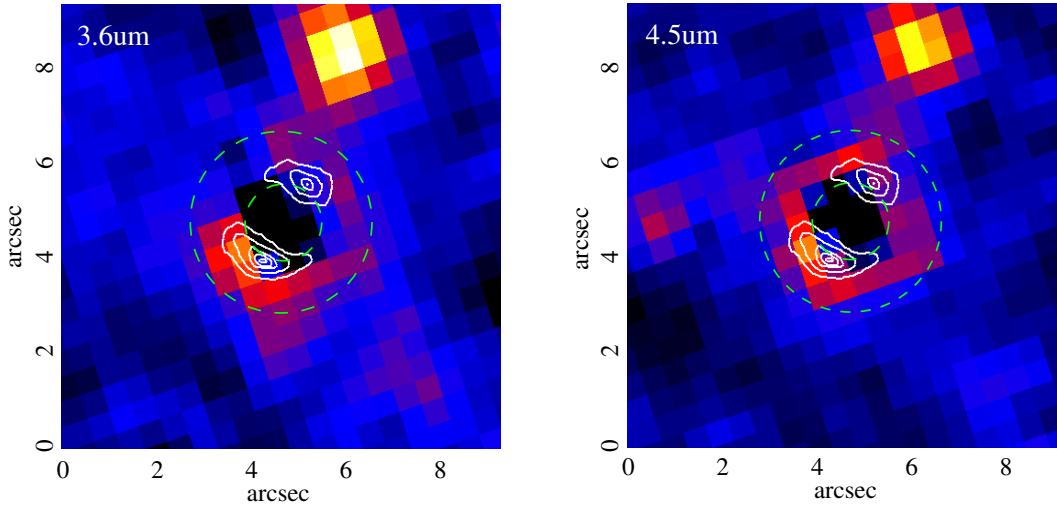
Figure A1. Residual flux remaining at 3.6 μm and 4.5 μm after subtraction of the model lens light profiles generated using GALFIT (see main text). The white contours show the ALMA band 4 continuum flux and the green dashed lines denote the annulus within which the mid-IR flux was extracted.

act like upper limits. To convert the quantities derived by MAGPHYS from observed (i.e., lensed) to intrinsic quantities of the unlensed source, we assumed a magnification of 5.8, equal to the mean of the continuum magnifications given in Table 3. The resulting stellar mass we obtained is $3.4^{+1.8}_{-1.7} \times 10^{11}$ M$_\odot$ where the quoted 1σ errors are taken from the distribution of masses output by MAGPHYS. Fig. A2 shows the corresponding MAGPHYS SED.

APPENDIX B: CHEMICAL EVOLUTION MODELLING

In this Section, we describe a one-zone chemical evolution model that we developed to compare the evolution of the interstellar dust, gas and stellar mass of a ID141-like system. The model is based on De Vis et al. (2017): the galaxy starts as a cloud of gas, which is converted into stars via the initial mass function and ongoing star formation. It includes a realistic description of the amount of gas flowing in (accretion from cosmic web) and out (supernovae and AGN) from the galaxy.

The dust prescription accounts for dust formation by stars (both low-intermediate mass stars in their asymptotic giant branch phase and massive-star supernovae). It also includes dust grain growth in the diffuse and dense environments of the galaxy, thought to be a dominant production route of dust in galaxies above a metallicity threshold (Asano et al. 2013; Dunne et al. 2011; Zhukovska 2014; De Vis et al. 2017). Dust is removed from the ISM via astration (locked up in stellar remnants), dust destruction by SN shocks and outflows.

As detailed in the main text, we used two star formation histories, one with a fixed continuous star formation rate and one with a starburst. For the latter, we assumed that we are currently observing a starburst at 40 per cent of the way through a burst lasting 200 Myr. This is motivated by the fact that the duration of high-z starbursts are thought to range from 50-250 Myr (e.g. Danielson et al. 2017); typical duty cycles of high-redshift SMGs range from 40-100 Myr (Coppin et al. 2008; Toft et al. 2014; Narayanan et al. 2015).

The three chosen IMFs adhere to the form $dN/d\log m \propto m^{-\alpha}$ where m is stellar mass. The Salpeter IMF is described by $\alpha = 1.7$. Both the Chabrier and top-heavy IMF of Zhang et al. (2018) are broken power laws with exponents $\alpha_{\mathrm{Chab}} = [0.4, 1.35]$ and $\alpha_{\mathrm{TH}} =$

IMF	α_0	α_1	m_0	m_1	m_2
Salpeter	1.7	1.7	0.1	0.5	100
Chabrier	0.4	1.35	0.4	1.0	100
Top heavy (this work)	0.3	1.1	0.1	0.5	100

Table B1. The IMFs used in the model to show extremes of the gas and dust evolution. The IMF slopes α are in terms of mass, and m_0 and m_2 are the upper and lower limits of stellar masses for which the IMF is normalised to unity. m_1 indicates the mass at which the slope may change. The top-heavy IMF used here is from Zhang et al. (2018), see also Romano et al. (2017); Cai et al. (2020). This is a less extreme form than the Baugh et al. (2005) top-heavy IMF, where $\alpha = 0$ across the mass range.

[0.3, 1.1] (see Table B1). Both IMFs are normalised to unity between stellar masses of 0.1 M$_\odot$ and 100 M$_\odot$.

Following Zhukovska (2014), we set the gas inflow rate to $I(t) \propto e^{-t/\tau}$ where τ is the infall timescale. We normalise this such that the galaxy accretes a total mass of gas M_{acc} over its lifetime (see below). To check that this is a realistic level of inflow, we compared with the halo baryonic mass accretion rate expected for a massive star-forming galaxy such as ID141 in the following way. Extreme star-forming galaxies with SFRs > 1000 M$_\odot$ yr^{-1} at $z \sim 4$ exist in dark matter halo masses of a few $\times 10^{12}$ M$_\odot$ (Behroozi et al. 2013). The mass of the halo with redshift $M_h(z)$ can be estimated using the prescriptions in McBride et al. (2009) and Fakhouri et al. (2010) where $M_h(z) = M_h(z = 0) \times (1 + z)^\beta e^{-\gamma z}$. $M_h(z = 0)$ is the halo mass the galaxy resides in at $z = 0$, β and γ are set to 0.1 and 0.69 following McBride et al. (2009). The accretion rate for a given halo mass and redshift can be estimated using the relationships from Forbes et al. (2014) (their Eqns. 15 and 22) and Neistein & Dekel (2008) where $\dot{M}_{\mathrm{ext}}(\mathrm{M}_\odot \, yr^{-1}) = \epsilon_{\mathrm{in}} f_b \dot{M}_h(z)$. f_b is the baryonic fraction (0.17) and ϵ_{in} denotes an efficiency. As an example, a galaxy with $M_h(z = 0) = 1 \times 10^{14}$ M$_\odot$ (Narayanan et al. 2015) would have a mass accretion rate of $\dot{M}_{\mathrm{ext}} \sim 700$ M$_\odot$ yr^{-1} at $z = 4.2$, and ~ 400 M$_\odot$ yr^{-1} at $z = 7$. A current day halo mass of 10^{13} M$_\odot$ would
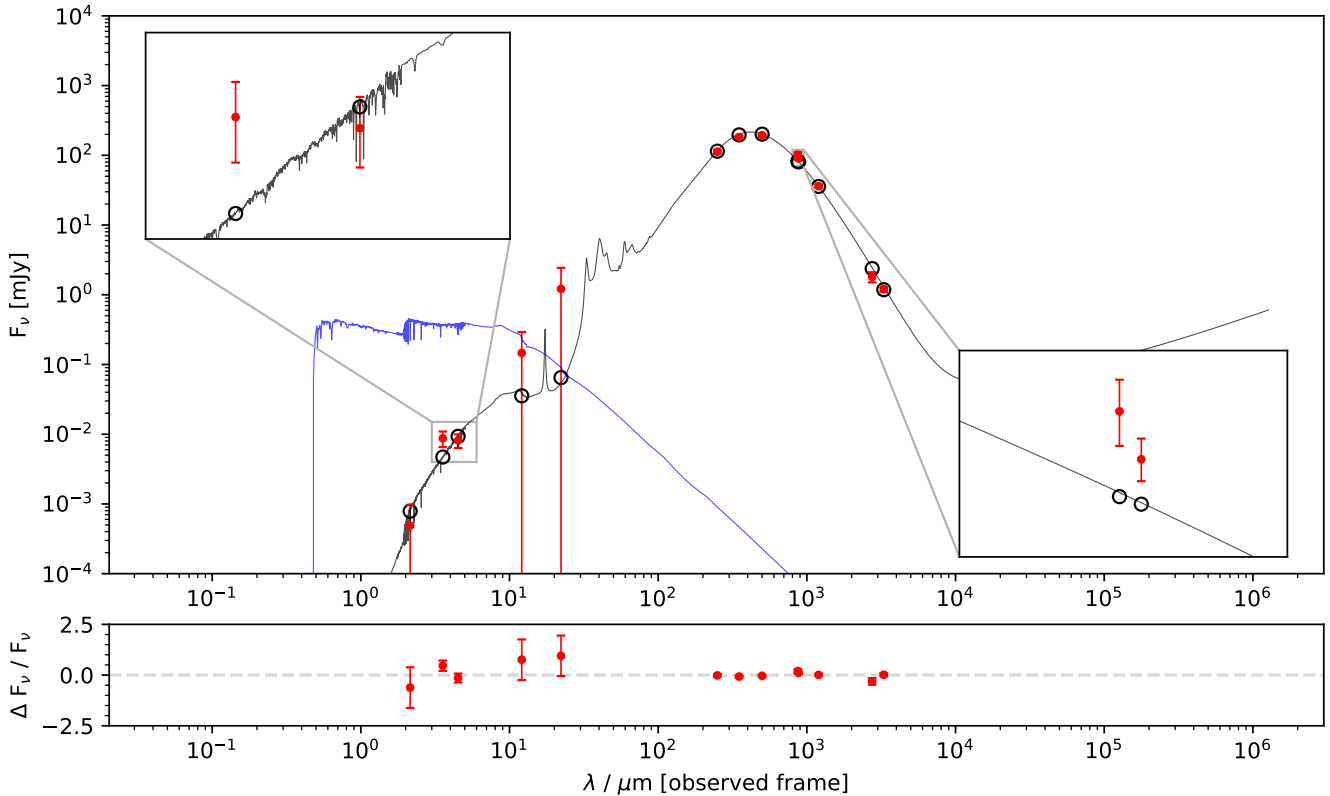
Figure A2. MAGPHYS SED (black line) fitted to Keck K_s band, Spitzer-IRAC (3.6 μm and 4.5 μm), WISE and far-IR photometry as described in main text. The unattenuated spectrum is shown as the blue line. Red points and black circles show observed and predicted fluxes respectively. The bottom plot shows the residuals calculated as observed minus predicted flux. Note that all photometry in this plot is as observed, i.e., not lens de-magnified.

have a mass accretion rate of roughly 8 times lower at these redshifts[5]. For the top heavy IMF with conSF SFR prescription, accretion of a total mass of 5×10^{11} M$_\odot$ of gas over the lifetime of the galaxy in our model, produces very similar inflow rates to those predicted from the halo mass growth described above. For the Salpeter IMF, the mass accretion rates need to be approximately twice as high to sustain the higher SFRs.

For the star formation histories, we also account for the different redshifts at which different mass halos undergo their peak SFE. Generally, this is at higher redshifts for higher mass halos. We scale the star formation efficiency by redshift such that the peak efficiency occurs at $z = 3 - 5$ but decreases at higher and lower redshifts. For the halo described above, Behroozi et al. (2013) (their Fig. 8) shows that this varies from $(0.4 - 1) \times \epsilon(z = 4)$. We fit a polynomial to their function.

We set the dimensionless parameter controlling the rate of dust grain growth in the dense interstellar medium to 7000. Recent work by De Vis et al. (2017) and De Vis et al. *in review* suggests an appropriate range of values for the rate of grain growth is 4000-7000. As discussed in the main text, this was allowed to vary from 700-15,000 in our analysis of ID141. In addition, we allow 60 per cent of the metals in the cloud to be 'available' for dust formation in order to produce a higher dust mass (lower gas to dust ratio). This factor is predicted to be ~0.5 in nearby spirals (De Vis et al. *in press*). The

grain growth parameter depends on the galaxy metal mass fraction, the current dust-to-metal ratio and the SFE, as well as the fraction of the ISM in dense clouds (we set this to 0.6, Inoue 2003; Yoshida et al. 2002). Finally, we let the galaxy evolve until (i) the stellar mass reaches 5×10^{11} M$_\odot$ where we assume an outflow clears the galaxy of the ISM (Romano et al. 2017), (ii) in the case of the starburst model, we stop the model when the burst has ended or (iii) when the gas fraction f_g falls to less than 3 per cent.

This paper has been typeset from a TEX/LATEX file prepared by the author.

---

[5] Similar results are obtained using the halo mass accretion prescriptions in Behroozi et al. (2013); Wechsler & Tinker (2018); Belfiore et al. (2019).