Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **478,** 1170–1184 (2018)
Advance Access publication 2018 May 3

doi:10.1093/mnras/sty1088

Kiloparsec-scale gaseous clumps and star formation at $z = 5$–7

S. Carniani,[1,2]★ R. Maiolino,[1,2] R. Amorin,[1,2] L. Pentericci,[3] A. Pallottini,[1,2,4,5] A. Ferrara,[5] C. J. Willott,[6] R. Smit,[1,2] J. Matthee,[7] D. Sobral,[7,8] P. Santini,[3] M. Castellano,[3] S. De Barros,[9,10] A. Fontana,[3] A. Grazian[3] and L. Guaita[11]

[1]*Cavendish Laboratory, University of Cambridge, 19 J. J. Thomson Ave., Cambridge CB3 0HE, UK*
[2]*Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[3]*INAF – Osservatorio Astronomico di Roma, via Frascati 33, I-00078 Monteporzio, Italy*
[4]*Centro Fermi, Museo Storico della Fisica e Centro Studi e Ricerche 'Enrico Fermi', Piazza del Viminale 1, Roma I-0184, Italy*
[5]*Scuola Normale Superiore, Piazza dei Cavalieri 7, I-56126 Pisa, Italy*
[6]*NRC Herzberg, 5071 West Saanich Rd, Victoria, BC V9E 2E7, Canada*
[7]*Leiden Observatory, Leiden University, PO Box 9513, NL-2300 RA Leiden, the Netherlands*
[8]*Department of Physics, Lancaster University, Lancaster LA1 4YB, UK*
[9]*Observatoire de Genève, Université de Genève, 51 Ch. des Maillettes, CH-1290 Versoix, Switzerland*
[10]*INAF – Osservatorio Astronomico di Bologna, via P. Gobetti 93/3, I-40129 Bologna, Italy*
[11]*Núcleo de Astronomía, Facultad de Ingeniería, Universidad Diego Portales, Av. Ejército 441, Santiago, Chile*

Accepted 2018 April 20. Received 2018 April 20; in original form 2017 December 11

ABSTRACT

We investigate the morphology of the [C II] emission in a sample of 'normal' star-forming galaxies at $5 < z < 7.2$ in relation to their UV (rest-frame) counterpart. We use new Atacama Large Millimetre/submillimetre Array (ALMA) observations of galaxies at $z \sim 6$–7, as well as a careful re-analysis of archival ALMA data. In total 29 galaxies were analysed, 21 of which are detected in [C II]. For several of the latter the [C II] emission breaks into multiple components. Only a fraction of these [C II] components, if any, is associated with the primary UV systems, while the bulk of the [C II] emission is associated either with fainter UV components, or not associated with any UV counterpart at the current limits. By taking into account the presence of all these components, we find that the $L_{[\mathrm{C\,II}]}$–SFR (star formation rate) relation at early epochs is fully consistent with the local relation, but it has a dispersion of 0.48 ± 0.07 dex, which is about two times larger than observed locally. We also find that the deviation from the local $L_{[\mathrm{C\,II}]}$–SFR relation has a weak anticorrelation with the EW(Ly α). The morphological analysis also reveals that [CII] emission is generally much more extended than the UV emission. As a consequence, these primordial galaxies are characterized by a [C II] surface brightness generally much lower than expected from the local $\Sigma_{[\mathrm{C\,II}]}-\Sigma_{\mathrm{SFR}}$ relation. These properties are likely a consequence of a combination of different effects, namely gas metallicity, [CII] emission from obscured star-forming regions, strong variations of the ionization parameter, and circumgalactic gas in accretion or ejected by these primeval galaxies.

Key words: galaxies: evolution – galaxies: high-redshift – galaxies: ISM; galaxies: formation.

1 INTRODUCTION

The morphological investigation of galaxies in the early Universe can provide important information on their formation and evolutionary processes. For instance, a disturbed and multiclump morphology, especially in the cold phase, may suggest the presence of disc instabilities, and can give indication of feedback processes, as well as minor and/or major merger events during the galaxy assembly (e.g. Tamburello et al. 2015; Ceverino et al. 2017; Fiacconi et al. 2017; Pallottini et al. 2017a,b). In this context, the advent of facilities delivering high angular resolution observations has enabled us to probe the internal structure of galaxies in the distant Universe, revealing that clumpy morphologies are more common at higher redshift than at $z = 0$ (e.g. Förster Schreiber et al. 2006; Genzel et al. 2008). For instance, the fraction of galaxies at $0.5 < z < 3$ exhibiting kpc-scale clumps with sustained star-formation activity is higher than 30 per cent (Ravindranath et al. 2006; Elmegreen et al. 2009; Guo et al. 2015). At higher redshifts, the fraction of galaxies showing disturbed morphology or multiclumps is even higher. On

★ E-mail: sc888@mrao.cam.ac.uk

a sample of 51 Ly α emitters (LAEs) and 16 Lyman break galaxies (LBGs) at $z > 5.7$ Jiang et al. (2013) found that roughly half of the brightest galaxies ($M_{1500} < -20.5$ mag) are made of multiple components that may be merging. Near-infrared (NIR) high-resolution imaging have also revealed that irregular shapes with multiclump morphology are prevalent in LAEs and LBGs at $z \sim 7$ within the epoch of reionization (Ouchi et al. 2010; Sobral et al. 2015; Bowler et al. 2017; Matthee et al. 2017).

Ly α Additional identification of clumpy systems in the early epoch and a detailed characterization of high-z clumps or satellites is fundamental to constrain galaxy assembly. The extended Atacama Large Millimetre/submillimetre Array (ALMA) configurations enable us to reach high-angular resolution and exploit far-infrared (FIR) fine structure emission lines, such as [C II] at 158 μm, as powerful diagnostics to assess the morphology of primeval galaxies. [C II] is emitted primarily in the (mostly neutral) atomic and molecular gas associated with photon-dominated regions (excited by the soft UV photons), but also in partly ionized regions and it is one of the primary coolants of the inter-stellar medium (ISM). Indeed, it is generally the strongest emission lines observed in the spectra of galaxies. Since its first detection at high redshift (Maiolino et al. 2005), this transition has then been detected in large samples of distant galaxies. However, until recently, the [CII] emission was only detected in extreme environments, such as quasar host galaxies and submillimetre galaxies (SMGs), characterized by SFRs of several hundred solar masses per year, not really representative of the bulk of the galaxy population at these epochs (e.g. Maiolino et al. 2009, 2012; De Breuck et al. 2011; Gallerani et al. 2012; Wagg et al. 2012; Carilli & Walter 2013; Carniani et al. 2013; Riechers et al. 2014; Williams et al. 2014; Yun et al. 2015; Trakhtenbrot et al. 2017; Decarli et al. 2017; Schreiber et al. 2018). Detecting [C II] in 'normal' galaxies has required the sensitivity delivered by ALMA. To date, the [C II] line has been detected in several galaxies at $z > 5$ and it is spatially resolved in most of these targets (Capak et al. 2015; Maiolino et al. 2015; Willott et al. 2015b; Knudsen et al. 2016; Pentericci et al. 2016; Bradač et al. 2017; Carniani et al. 2017, 2018; Matthee et al. 2017; Smit et al. 2018). Smit et al. (2018) recently presented [C II] observations of two galaxies at $z \sim 7$ characterized by a gradient of velocity consistent with a undisturbed rotating gas disc. However, most of the $z > 5$ galaxies show extended and clumpy [C II] emission with velocities consistent with the systematic redshift of the galaxy ($|\Delta v| < 500\,\mathrm{km\,s^{-1}}$) but spatially offset relative to the rest-frame UV counterpart (Capak et al. 2015; Maiolino et al. 2015; Willott et al. 2015b; Carniani et al. 2017, 2018; Jones et al. 2017b). In many cases, these offsets have been ignored or ascribed to astrometric uncertainties. However, based on detailed astrometric analysis, it has been shown that most of these offsets are physical (a revised analysis will be given in this paper); hence, they should be taken as an important signature of the evolutionary processes in the early phases of galaxy formation. Various scenarios have been proposed to explain the positional offsets between [CII] and star-forming regions such as stellar feedback clearing part of the ISM, gas accretion, wet mergers, dust obscuration, and variations of the ionization parameter (e.g. Vallini et al. 2015; Katz et al. 2017). Barisic et al. (2017) and Faisst et al. (2017) have recently found (rest-frame) UV faint companions whose locations is consistent with the displaced [C II] emission, suggesting that the carbon line traces star-forming regions where the UV light is absorbed by dust.

As mentioned, most previous studies have attempted to assess the nature of [C II] emission in primeval galaxies neglecting the positional offsets between the FIR line and rest-frame UV emission. The goal of this paper is to assess the connection between [C II] and SFR in the early Universe by taking into account the multiclump morphology of galaxies at $z > 5$ and by associating the components with their proper optical-UV counterparts (if detected). This is achieved by re-analysing ALMA [C II] observations of $z > 5$ star-forming galaxies and by performing a detailed kinematical analysis of the [CII] line in order to deblend the different components of the multiclump systems. In addition to previous ALMA observations, partly discussed in literature, we also make use of new ALMA data targeting five $z \sim 6$ star-forming galaxies with SFR $< 20\,\mathrm{M_\odot\,yr^{-1}}$. In Section 2, we detail the sample and the analysis of ALMA observations. The morphological analysis is presented in Section 3, while the relation between the [C II] and SFR is discussed in Section 4. In Section 5, we investigate the connection between [CII] luminosity and Ly α strength in our sample. Section 6 focuses on the spatial extension of the [C II] and UV emission and the correlation between [C II] surface brightness and SFR surface density. We discuss the findings in Section 7, while the conclusions of this work are reported in Section 8.

Throughout this paper, we assume the following cosmological parameters: $H_0 = 67.8\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_M = 0.308$, $\Omega_\Lambda = 0.685$ (Planck Collaboration et al. 2016).

2 SAMPLE, OBSERVATIONS, AND ANALYSIS

2.1 Archival data

The sample is mainly drawn from the archive and literature by selecting only spectroscopically confirmed star-forming galaxies at $z > 5$ observed with ALMA in the [C II] line. We limit our sample to those systems with SFR $\lesssim 100\,\mathrm{M_\odot\,yr^{-1}}$ since they are representative of the bulk of the galaxy population in the early Universe (e.g. Carniani et al. 2015; Robertson et al. 2015). The list of selected sources is given in Table 1. The sample does not include lensed systems (González-López et al. 2014; Schaerer et al. 2015; Knudsen et al. 2016; Bradač et al. 2017), since magnification factor uncertainties may lead to large errors on SFR and [CII] luminosity estimates, as well as on the morphology analysis.

For the purpose of our investigation, which focuses on the nature and implications of the positional offsets between [C II] and UV emission, we have retrieved and re-analysed ALMA data revealing a [C II] detection at the systemic velocity of the galaxy (see Table 1). For these objects, ALMA observations have been calibrated following the prescriptions presented in previous works.

In addition to the rest-frame FIR images, we have also used *Hubble Space Telescope* (*HST*) and Visible and Infrared Survey Telescope for Astronomy (VISTA) NIR observations (rest-frame UV at $z > 5$). ALMA, *HST*, and VISTA data have been aligned based on the location of serendipitous sources detected in both ALMA continuum and NIR images by assuming that the millimetre emission of these sources is cospatial to the NIR map. This is also supported by the fact that all foreground sources used for registering millimetre and NIR images do not exhibit any multiclump or merger-like morphologies indicating that astrometric offsets between the ALMA and NIR images are likely associated with astrometric calibrations. For those observations revealing the presence of two (or more) serendipitous sources, we have verified that the astrometric shift for each source is consistent with that estimated from the other source(s) in the same map. In all cases, we have checked that the estimated astrometric offset is consistent with those obtained by aligning NIR foreground sources and ALMA phase calibrators to their astrometric position from the GAIA Data Release 1

Table 1. Overview of the $z > 5$ star-forming galaxies observed with ALMA used in this paper, ordered by name.

Target[a]	RA[b]	Dec.[c]	Ref.[d]	[C II][e]	Clumpy[f]
Literature sample					
BDF3299	337.0511	−35.1665	1,2	√	√
BDF512	336.9444	−35.1188	1		
CLM1	37.0124	−4.2717	3	√	
COSMOS13679	150.0990	2.3436	4	√	
COSMOS24108	150.1972	2.4786	4	√	√
COS-2987030247	150.1245	2.2173	5	√	
COS-3018555981	150.1245	2.2666	5	√	
CR7	150.2417	1.8042	6	√	√
Himiko	34.4898	−5.1458	7,12	√	√
HZ8	150.0168	2.6266	8	√	√
HZ7	149.8769	2.1341	8	√	
HZ6	150.0896	2.5864	8	√	√
HZ4	149.6188	2.0518	8	√	
HZ3	150.0392	2.3371	8	√	
HZ9	149.9654	2.3783	8	√	
HZ10	150.2470	1.5554	8	√	√
HZ2	150.5170	1.9289	8	√	√
HZ1	149.9718	2.1181	8	√	
IOK-1	200.9492	27.4155	9		
NTTDF6345	181.4039	−7.7561	4	√	
SDF46975	200.9292	27.3414	1		
SXDF-NB1006-2	34.7357	−5.3330	10		
UDS16291	34.3561	−5.1856	4	√	
WMH5	36.6126	−4.8773	3,11	√	√
Additional new data					
BDF2203	336.958	−35.1472		√	
GOODS3203	53.0928	−27.8826			
COSMOS20521	150.1396	2.4269			
NTTDF2313	181.3804	−7.6935			
UDS4812	34.4768	−5.2472			

[a]Name of the source. [b,c]J2000 coordinates. [d]References in which ALMA observations are presented ([1] Maiolino et al. 2015; [2] Carniani et al. 2017; [3] Willott et al. 2015b; [4] Pentericci et al. 2016; [5] Smit et al. 2018; [6] Matthee et al. 2017; [7] Ouchi et al. 2013; [8] Capak et al. 2015; [9] Ota et al. 2014; [10] Ouchi et al. 2013; [11] Jones et al. 2017b; [12] Carniani et al. 2018). [e]Check mark, √, indicates that [C II] emission has been detected at the redshift of the galaxy. [f]Check mark, √, indicates that the galaxy has a clumpy morphology.

catalogue (Gaia Collaboration et al. 2016). For those sources whose ALMA continuum map showing no serendipitous sources, we have matched the NIR foreground sources and ALMA phase calibrators to either GAIA Data Release 1 catalogue (Gaia Collaboration et al. 2016) or AllWISE catalogue (Cutri et al. 2013). We note that in all ALMA data sets the locations of the various phase calibrators are in agreement within the error with the GAIA and AllWISE catalogues, implying that, when a systemic (not physical) offset is seen, this is generally due to some small astrometric uncertainties in the optical-NIR data. These astrometric issues have been also discussed by Dunlop et al. (2017) who analysed ALMA images targeting the *Hubble Ultra Deep Field*. We therefore applied the astrometric shifts, which span a range between 0.1 and 0.25 arcsec, to the NIR images.

We note that additional systems with low SFRs have also been tentatively detected in 14 [C II] line emitting candidates at $6 < z < 8$ (Aravena et al. 2016), which are not included in this analysis as they are not spectroscopically confirmed yet and ∼60 per cent of these objects are expected to be spurious.

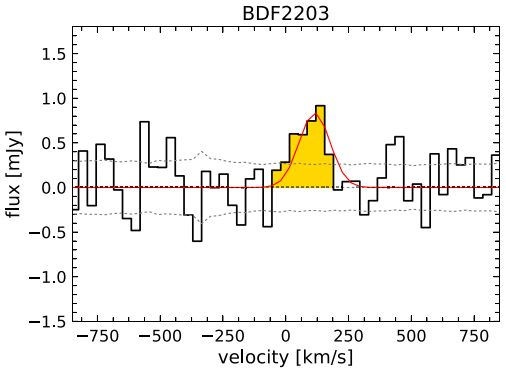

Figure 1. ALMA spectrum of BDF2203 showing a new [C II] detection. The velocity reference is set to the redshift defined by the Ly α. The dotted grey lines show the 1σ and -1σ per channel. The red line indicates the best-fitting one-dimensional Gaussian line profile.

2.2 Additional new ALMA data

In addition to the archival/literature sample, we have also included new [C II] observations of five star-forming galaxies at $z \sim 6$ with a SFR $\sim 15\ M_\odot\ yr^{-1}$ observed with ALMA in Cycles 3 and 4 (P.I. Pentericci). The five new sources, listed in Table 1, have been selected from a sample of >120 LBGs at $z \sim 6-7$. These have been spectroscopically confirmed thanks to recent ultra-deep spectroscopic observations from the ESO Large Program CANDELSz7 (P.I. L. Pentericci; De Barros et al. 2017). The proposed ALMA programmes aimed at observing [CII] emission in 10 star-forming galaxies at $z > 6$ with UV luminosities lower than −21 mag (UV SFR $<20 M_\odot\ yr^{-1}$) and spectroscopic redshift uncertainties <0.03, but only 5 galaxies have been observed. The observations and data calibrations are presented in AppendixA. We have registered NIR images to ALMA observations by matching the location of the foreground, serendipitous continuum sources, and ALMA calibrators to the position given by the GAIA Data Release 1 catalogue (Gaia Collaboration et al. 2016).

While the continuum emission is not detected at the location of any of the five galaxies, we detect two and one serendipitous sources in the COSMOS20521 and BDF2313 continuum maps, respectively. The positions of the serendipitous continuum sources is in agreement with the location of NIR foreground galaxies, thus confirming the astrometric shifts estimated from the catalogue.

The [C II] emission is undetected in all but one of these sources, that is, BDF2203. Fig. 1 shows the spectrum of the detected [C II] emission, with a spectral rebinning of $40\ km\ s^{-1}$, while the spectra of the non-detections are shown in Appendix (Fig. A1).

For the four non-detections, we assume a full width at half-maximum (FWHM) $= 100\ km\ s^{-1}$, which is consistent with [C II] line widths observed in other $z > 6$ galaxies (e.g. Pentericci et al. 2016; Carniani et al. 2017, 2018), and we infer 3σ upper limits on the [C II] luminosities (Table A1).

Ly αThe redshift of the [C II] emission detected in BDF2203 is $z_{[C\,II]} = 6.1224 \pm 0.0005$, which is in agreement with that inferred from Ly α ($z = 6.12 \pm 0.03$). By fitting a one-dimensional Gaussian profile to the [CII] line we estimate an FWHM $= 150 \pm 50\ km\ s^{-1}$ that is similar to those estimated in high-z [C II]-emitting galaxies

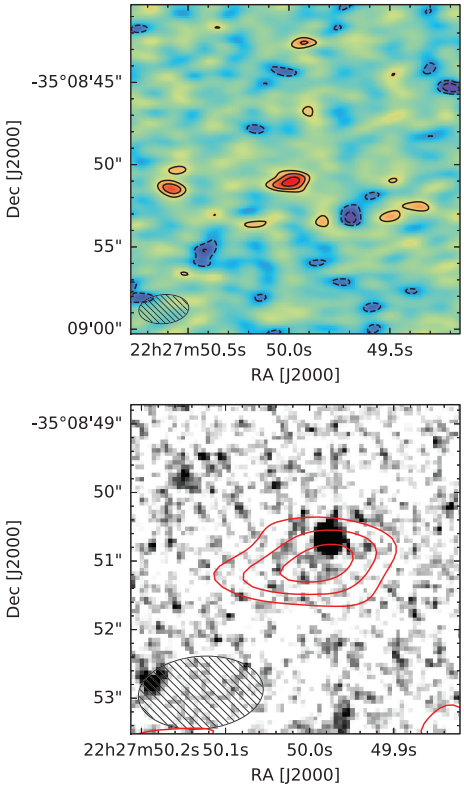

Figure 2. Top: [C II] flux map of BDF2203 obtained integrating the ALMA cube over the velocity range between −60 and 132 km s^{-1}. Red solid contours are at levels of 2σ, 3σ, and 4σ, where $\sigma = 23$ mJy km s^{-1}. Dashed black contours indicate negative values at −3 and −2 times the noise level in the same map. Bottom: zoom of the central 2.5 arcsec region around BDF2203. [CII] flux map contours are over imposed on the NIR image. Red contours are at the same levels as in the top panel. The ALMA synthesized beam in the lower left corner in both panels.

(FWHM = 50–250 km s^{-1}; Willott et al. 2015b; Pentericci et al. 2016; Matthee et al. 2017; Carniani et al. 2018). The flux map of the [C II] line, extracted with a spectral width of 200 km s^{-1}, is shown in Fig. 2. The emission is detected in the map with an S/N = 5 and has an integrated flux density of 140 ± 35 mJy km s^{-1}, which corresponds to $L_{[\rm C\,II]} = (12.5 \pm 2.5) \times 10^7$ L$_\odot$ at $z = 6.12$. By fitting a 2D Gaussian profile to the flux map, we measure a size of (2.3 ± 0.5) arcsec× (1.2 ± 0.2) arcsecwith PA = 88 deg ± 8 deg and a beam-deconvolved size of (1.4 ± 0.8) arcsec × (0.5 ± 0.4) arcsecwith a PA = 80 deg ± 20 deg. The [C II] emission is thus marginally resolved and has a diameter size of ∼5 kpc. The [C II] flux map overlaps with the rest-frame UV emission that has an extension of about 1 kpc. The centroids are separated by 0.25 arcsec. However, given the ALMA beam size (1.90 arcsec × 1.11 arcsec), such positional offset is consistent with the positional uncertainties $\Delta\theta \approx \frac{\langle \text{beam size} \rangle}{\text{S/N}} \sim 0.2{-}0.4$ arcsec. The low angular resolution and sensitivity of current observations are not sufficient to assess the morphology of the FIR line emission.

2.3 SFR estimates and morphology analysis

The continuum emission at rest-frame wavelengths around 158 μm is associated with thermal emission from dust heated by the UV emission of young stellar populations in galaxies. The continuum emission is detected only in four sources within the selected sample. When compared with the typical UV-IR spectral energy distribution (SED) of galaxies, the weak rest-frame FIR continuum indicates that these galaxies are on average characterized by low-dust masses.

The SED of the thermal dust emission can be modelled with a greybody with dust temperature $T_{\rm d}$ and spectral index emissivity β. However, one single photometric measurement is not sufficient to perform the SED fitting and constrain the two free parameters. We thus assume $T_{\rm d} = 30$ Kand $\beta = 1.5$, which are consistent with those observed in local dwarf galaxies (Ota et al. 2014), to estimate the FIR luminosity from the ALMA continuum observations. For those galaxies that are not detected in ALMA continuum images we infer a 3σ upper limit on $L_{\rm FIR}$. We note that the FIR emission strongly depends on the assumed dust temperature, yielding to a luminosity uncertainty of $\Delta\log(L_{\rm FIR}) = 0.6$ (Faisst et al. 2017).

For each galaxy, we can infer the star formation rate from both the UV (SFR$_{\rm UV}$) and FIR (SFR$_{\rm FIR}$) emissions by adopting the calibrations presented in Kennicutt & Evans (2012). Excluding HZ9 and HZ10, all sources in the sample have SFR$_{\rm FIR}$/SFR$_{\rm UV} \lessapprox 1$ with an average value of SFR$_{\rm FIR}$/SFR$_{\rm UV} \approx 0.6$. Given that most of the FIR luminosity estimates are 3σ upper limits, the average SFR$_{\rm FIR}$/SFR$_{\rm UV}$ is actually much lower than 0.6.

Given the small contribution of SFR$_{\rm FIR}$ (either in the IR-detected sources or those with upper limits), we assume total SFR ≈ SFR$_{\rm UV}$ with no dust correction for those galaxies without continuum detection. While we infer total SFR = SFR$_{\rm UV}$ + SFR$_{\rm FIR}$ for those galaxies revealing continuum emission at 158 μm.

In order to assess the multiclump morphology of the [CII] emission in our sources, we perform a kinematical analysis on the retrieved ALMA cube, extracting channel maps at different velocities relative to the redshift of the galaxies. This analysis enables us to disentangle the emission of complex systems having multiple components at different velocities (e.g. Carniani et al. 2013; Riechers et al. 2014). We also estimate the size of the individual components by fitting a two dimensional (2D) parametric profile to emission maps and by taking into account the angular resolution of the observations. For the UV regions, we calculate the the half-light radius ($R_{\rm UV}$) from the *HST* NIR observations at rest-frame λ ∼ 1600−2000 Å, which have an angular resolution of 0.2 arcsec. By using our own PYTHON script, we fit a 2D Gaussian profile to the source and correct the best-fitting size for the point spread function (PSF) smearing by subtracting the PSF size in quadrature. For those few galaxies having only VISTA NIR observations, which have angular resolution of ∼0.8 arcsec, the UV emission is not spatially resolved yielding an upper limit on $R_{\rm UV}$. We instead use the IMFIT task of CASA to estimate the half-light radius $R_{[\rm C\,II]}$ of the [C II] emission. This task uses a 2D Gaussian profile to match the [C II] flux map of the source and then estimates the deconvolved size by including the shape and inclination of the ALMA beam. In Appendix B, we show how our source-size estimates are not biased by using Monte Carlo simulations. Additionally, the results of these simulations reveal that the relative intrinsic error associated with the measurements is <30 per cent.

The properties of the [C II], FIR and UV emission, such as redshift, flux density, luminosity, SFR, and radius, are reported in Table 2.

3 MULTICOMPONENTS SYSTEMS

Carniani et al. (2017) discuss the origin of the positional and spectral offsets between optical emission and FIR lines observed in high-z systems, suggesting that the observational properties can

Table 2. Star-forming galaxies at $z > 5$.

Name	ID	Redshift	$\log(L_{UV})$	SFR_{UV}	$\log(L_{[C\,II]})$	r_{UV}	$r_{[C\,II]}$	S_{mm}	$\log(L_{FIR})$	
			($L_\odot$)	($M_\odot$ yr^{-1})	($L_\odot$)	(kpc)	(kpc)	(μJy)	($L_\odot$)	
(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)
HZ8*	HZ8	5.1533	11.17	25.4	8.7	–	–	<90	<10.7	[1,2,3]
HZ8a	HZ8a	5.1533	11.04	18.8	8.4	1.2	3.2	<90	<10.7	[1,2,3]
HZ8b	HZ8b	5.1533	10.57	6.4	8.3	1.5	1.2	<90	<10.7	[1,2,3]
HZ7	HZ7	5.2532	11.05	19.3	8.7	1.0	1.9	<108	<10.8	[1,2,3]
HZ6*	HZ6	5.2928	11.47	50.7	9.2	–	–	220	11.1	[1,2,3]
HZ6a	HZ6a	5.2928	11.11	22.1	8.3	0.9	<1.9	30	10.3	[1,2,3]
HZ6b	HZ6b	5.2928	11.0	17.2	8.8	0.8	1.6	120	10.9	[1,2,3]
HZ6c	HZ6c	5.2928	10.81	11.1	8.7	0.8	1.7	100	10.8	[1,2,3]
HZ9	HZ9	5.541	10.95	15.3	9.2	0.9	1.5	516	11.5	[1,2,3]
HZ3	HZ3	5.5416	11.08	20.6	8.7	0.7	<2.3	<153	<11.0	[1,2,3]
HZ4	HZ4	5.544	11.28	32.7	9.0	0.7	1.4	202	11.1	[1,2,3]
HZ10*	HZ10	5.6566	11.19	26.6	9.4	–	–	1261	11.9	[1,2,3]
HZ10a	HZ10a	5.6566	11.14	23.7	9.3	1.1	1.6	630	11.6	[1,2,3]
HZ10b	HZ10b	5.6566	10.23	2.9	8.8	0.7	1.5	630	11.6	[1,2,3]
HZ2*	HZ2	5.6597	11.15	24.3	9.0	–	–	<87	<10.8	[1,2,3]
HZ2a	HZ2a	5.6597	11.08	20.6	8.6	0.7	1.5	<87	<10.8	This work
HZ2b	HZ2b	5.6597	10.85	12.0	8.5	1.1	<2.7	<87	<10.8	This work
HZ2c	HZ2c	5.6597	10.52	5.7	8.5	0.8	<2.7	<87	<10.8	This work
HZ1	HZ1	5.6885	11.21	28.5	8.4	1.2	0.9	<90	<10.8	[1,2,3]
WMH5*	WMH5	6.0695	11.36	59.0	8.7	–	–	91	10.8	[6,7]
WMH5a	WMH5a	6.0695	<10.5	<5.0	8.5	<2.3	0.6	42	10.5	[6,7]
WMH5b	WMH5b	6.0695	11.36	59.0	8.4	<2.3	1.4	49	10.6	[6,7]
NTTDF2313	N23	6.07	10.85	12.0	<7.7	<1.4	–	<54	<10.6	This work
BDF2203	B22	6.12	10.97	16.0	8.1	1.1	2.7	<69	<10.7	This work
CLM1	CLM1	6.1657	11.37	60.0	8.4	<2.3	<1.8	<78	<10.8	[6]
GOODS3203	GS32	6.27	11.02	18.0	<8.1	1.1	–	<123	<11.0	This work
COSMOS20521	C20	6.36	10.89	14.0	<7.7	0.7	–	<60	<10.7	This work
UDS4812	U48	6.561	10.87	13.0	<7.8	0.7	–	<72	<10.8	This work
Himiko*	H	6.595	11.07	20.4	8.1	–	–	<27	<10.4	[13,14]
Himiko-a	Ha	6.595	10.45	4.9	<7.3	0.7	–	<27	<10.4	[13,14]
Himiko-b	Hb	6.595	10.26	3.1	7.7	0.9	<2.0	<27	<10.4	[13,14]
Himiko-c	Hc	6.595	10.49	5.4	<7.3	0.7	–	<27	<10.4	[13,14]
Himiko-Lya	HL	6.595	<10.31	<3.5	7.9	–	3.4	<27	<10.4	[13,14]
CR7*	CR7	6.604	11.2	27.1	8.3	–	–	<21	<10.2	[10]
CR7a	CR7a	6.604	10.9	15.6	7.9	0.9	3.0	<21	<10.2	[10]
CR7b	CR7b	6.604	10.2	2.9	7.5	0.9	<2.2	<21	<10.2	[10]
CR7c	CR7c	6.604	10.3	4.0	7.4	0.9	3.8	<21	<10.2	[10]
COSMOS24108*	C24	6.6294	10.99	16.7	8.1	–	–	<54	<10.7	[4]
COSMOS24108a	C24a	6.6294	10.99	16.7	<7.9	1.1	–	<54	<10.7	[4]
COSMOS24108b	C24b	6.6294	<10.37	<7.3	8.1	–	2.1	<54	<10.7	[4]
UDS16291	U16	6.6381	10.71	8.8	7.9	1.0	2.4	<60	<10.7	[4]
NTTDF6345	N63	6.701	10.95	15.3	8.2	1.5	<1.7	<48	<10.6	[4]
COS-2987030247	C29	6.8076	11.11	23.0	8.6	–†	3.1	<75	<10.8	[5]
SDF46975	S46	6.844	10.94	15.4	<7.8	<2.0	–	<58	<10.7	[8]
COS-3018555981	C30	6.854	11.04	18.8	8.7	1.3	2.6	<87	<10.9	[5]
IOK-1	IOK	6.96	10.94	15.1	<7.5	0.6	–	<63	<10.8	[11]
BDF512	B51	7.008	10.54	6.0	<7.8	0.5	–	<52	<10.7	[8]
BDF3299*	B32	7.107	0.0	6.4	7.8	–	–	<23	<10.3	[8,9]
BDF3299a	B32a	7.107	10.52	5.7	<7.3	0.9	–	<23	<10.3	[8,9]
BDF3299b	B32b	7.107	<10.0	<1.7	7.8	–	1.0	<23	<10.3	[8,9]
COSMOS13679	C13	7.1453	10.91	13.9	7.9	0.8	1.4	<42	<10.6	[4]
SXDF-NB1006-2	SXDF	7.212	11.06	8.7	<7.9	–	–	<42	<10.6	[12]

Notes: **(a)** NAME OF THE SOURCE; the asterisk mark (*) indicates that the source has a multiclump morphology. **(b)** ID used to indicate the source in the figures of this paper. **(c)** Redshift of the galaxy (or system) inferred from either Ly α line or spectroscopic Lyman break. C29 and C30 do not have Ly α spectra, so the redshift is obtained from [CII]. **(d)** Rest-frame UV luminosity at 1600Å. **(e)** SFR based on the UV emission: $\log(SFR/M_\odot\ yr^{-1}) = \log(L_{UV}/erg\ s^{-1}) - 43.35$ (Hao et al. 2011; Murphy et al. 2011; Kennicutt & Evans 2012) **(f)** [C II] luminosity. **(g, h)** Half-light radius in kpc for UV and [C II] emission. **(i)** Continuum emission (or 3σ upper limit) at rest-frame 158 μm. **(j)** FIR luminosity (or 3σ upper limit) estimated from ALMA observations. **(k)** References: [1] Capak et al. (2015), [2] Barisic et al. (2017), [3] Faisst et al. (2017), [4] Willott et al. (2015b), [5] Jones et al. (2017b), [6] Ouchi et al. (2013), [7] Matthee et al. (2017), [8] Pentericci et al. (2016), [9] Smit et al. (2018), [10] Maiolino et al. (2015), [11] Ota et al. (2014), [12] Carniani et al. (2017), [13] Inoue et al. (2016), [14] Carniani et al. (2018). †The presence of a BACKGROUND source at lower redshift does not allow us to estimate the dimension of the UV region.

arise from distinct regions (or components) of galaxies. In order to understand the nature of these offsets, we have re-analysed ALMA [C II] observations for those galaxies showing [C II] detections and have performed a morphology analysis as discussed in Section 2.3. We have found a clear multiclump morphology in 9 out of 21 $z > 5$ galaxies having [C II] detections (Table 2). The UV rest-frame images and [C II] maps are shown in Fig. 3, sorted by redshift. The stamps are 3 arcsec across, or ∼17.5 kpc at the average redshift of this sample ($\langle z \rangle = 6$), and show the contours of the [C II] maps, obtained from our analysis, superimposed on the rest-frame UV emission in grey-scale.

New *HST* near-IR images of HZ6, HZ8, and HZ10 have already been presented in Barisic et al. (2017) and Faisst et al. (2017) revealing multicomponent structures. The location of the individual rest-frame UV clumps is consistent with the peak positions of the [CII] emission extracted at different velocities relative to the redshift of the brightest component (which is labelled 'a' in all stamps). The [C II] emission detected in all individual clumps has a level of significance[1] higher than 5σ and it is spatially resolved (see Table 2). The channel map analysis confirms that HZ8b and HZ10b (Hz8W and Hz10W in previous works) are at the same redshift of HZ8a and HZ10a, respectively. We note that the kinematic properties of HZ10 are consistent with the analysis reported by Jones et al. (2017a), who claim that the velocity gradient observed in this galaxy matches a merger scenario rather than a rotating gas disc model.

In the HZ2 system, the deeper *HST* observations reveal also two faint companions (HZ2b and HZ2c) close to the main galaxy (HZ2a) within a projected distance of ∼6 kpc. These sources are not discussed in previous studies (Capak et al. 2015; Barisic et al. 2017), since their redshifts were not fully spectroscopically confirmed by the [C II] line. However, a detailed kinematic investigation of the carbon line shows an extended emission with a morphology consistent with the rest-frame UV emission. Although the two faint companions are detected with a low level of significance (∼3.5σ), the match between the three UV peaks and the [C II] emission supports the reliability of the ALMA detections. The three sources are at the same redshift and form a multicomponent system similar to that observed in HZ6. The angular resolution and the low sensitivity of current ALMA observations are not sufficient to spatially resolve the [CII] emission in HZ2b and HZ2c. HZ2a is instead spatially resolved, with a diameter of 2.5 kpc.

An additional star-forming galaxy with a complex morphology is WHM5 at $z = 6.0695$, which appears to consist of multiple components in [C II]: a compact source, seen also in dust emission, and an extended component at the location of the rest-frame UV emission (Willott, Bergeron & Omont 2015a; Jones et al. 2017b). The two components are separated by a projected distance of ∼3 kpc and a velocity of ∼200 km s^{-1}. Both [C II] emission components are spatially resolved (∼1.3 and ∼2.8 kpc) in the ALMA observations with an angular resolution of 0.3 arcsec (Jones et al. 2017b).

At higher redshift, the presence of multiple [C II] components has recently been reported in CR7 at $z = 6.6$ by Sobral et al. (2015, 2017) and Matthee et al. (2015, 2017). Three out of four detected [C II] clumps coincide with the location of UV clumps. In two cases, the [C II] emission is spatially resolved with a radius of ∼3−3.8 kpc (see Matthee et al. 2017 for details).

An additional multiclump galaxy observed with ALMA is Himiko at $z \sim 6.695$ (Ouchi et al. 2013). The rest-frame UV image (see fig. 3 of Ouchi et al. 2013) reveals that the galaxy comprises three subcomponents with SFR spanning in range between 5 and 8 M$_\odot$ yr^{-1}. The projected distance between the sources is of about 3–7 kpc. One out of the three subcomponents (Himiko-a in Fig. 3) have a Ly α EW = 68 Å, while the other two have EW less than 8 Å. Although early studies had reported non-detections of [CII] in this source, recently Carniani et al. (2018) have reported a clear detection with extended/multiclump morphology. The primary [CII] emission is coincident with the Ly α peak, while the UV clumps are much weaker in [CII] or even undetected. More generally the extended [C II] emission does not resemble the UV clumpy distribution.

For the remaining systems (COS24108 and BDF3299) showing a clear positional offsets between [C II] and UV emission in Fig. 3, we cannot speculate much more than what has been already done in previous works due to the lack of deeper ALMA and *HST* observations. As discussed by Carniani et al. (2017), these offsets are certainly associated with physically distinct subcomponents, and the [CII] clumps with no UV counterpart may either be tracing star-forming regions that are heavily obscured at UV wavelengths (Katz et al. 2017), or associated with accreting/ejected gas.

In summary, the deeper [C II] and rest-frame UV observations have unveiled the real multicomponent nature of nine star-forming galaxies at $z > 5$ further highlighting and (partly) explaining positional offsets between UV and [C II] emission in previous studies (Capak et al. 2015; Maiolino et al. 2015; Willott et al. 2015b; Pentericci et al. 2016; Carniani et al. 2017; Carniani et al. 2018). These results suggest that future evidence of displaced FIR line and UV emission should be not ignored since it generally reveals the presence of subcomponents with different physical properties.

Such subcomponents can be ascribed to either satellites in the process of accreting (Pallottini et al. 2017a) or clumps ejected by past galactic outflows (Gallerani et al. 2018). However, in many cases both the SFR$_{\rm UV}$ and the size of the various subcomponents are comparable with each other, hence suggesting merger scenario. In addition to that, in four systems (HZ2, HZ10, CR7, and WHM5) the kinematic of the gas does not exhibit any clear evidence for ordered rotation and the narrow velocity dispersion is more consistent with a merger or inflow scenario than strong outflows (Riechers et al. 2014; Jones et al. 2017a,b; Matthee et al. 2017). This seems also to be supported by the fact that the subcomponents of CR7 have similar dynamical masses (Matthee et al. 2017). For the other systems, current observations have high neither angular resolution nor sensitivity to perform a detailed kinematic analysis, so we cannot rule out the outflow interpretation. Future ALMA observations are necessary to detect further subcomponents in these and other systems and to perform a detailed dynamical analysis, which allow us to asses the dynamical mass and nature of subcomponents.

4 THE $L_{\rm [C\,II]}$–SFR RELATION AT $z = 5$–7

A tight relation between the [C II] luminosity and the global SFR is seen in local galaxy observations, at least when excluding extreme (ULIRG-like) cases (De Looze et al. 2014; Kapala et al. 2015; Herrera-Camus et al. 2015). This finding makes the [C II] line a promising tool to investigate the properties of early galaxies and to trace their star formation. However, the behaviour of the [C II] line emission at $z > 5$ seems to be more complex than observed in the local Universe. Previous studies have shown that only a fraction of [C II] detections of early galaxies agree with the local relation, while most high-z galaxies are broadly scattered, with claims that most of them are [C II]-deficient relative to the local relation. However, most of previous high-z studies classified multicompo-

[1] σ is the rms of the channel map in which we detect the [C II] emission.

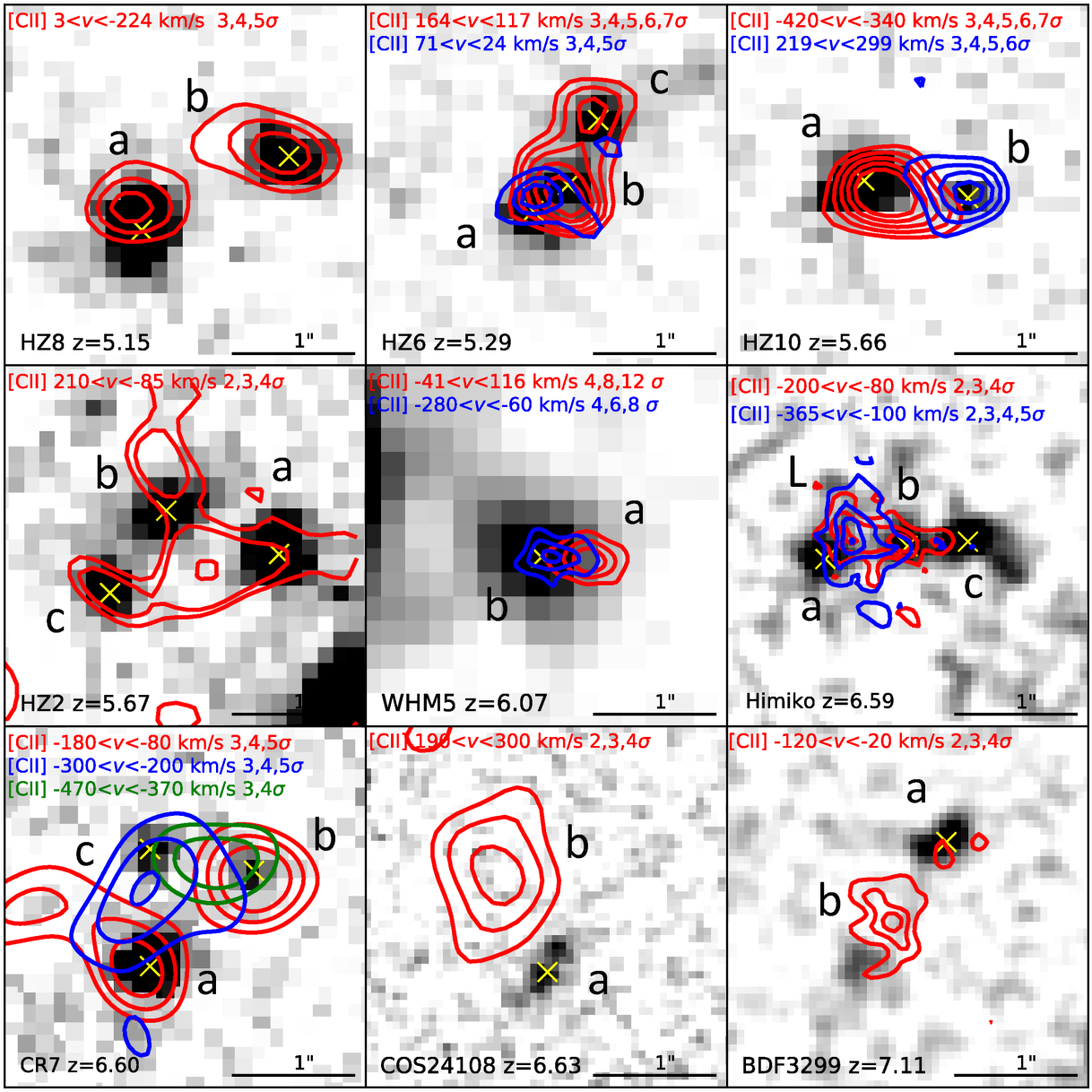

Figure 3. Rest-frame UV images of the nine $z > 5$ star-forming galaxies showing multiclump morphology in [CII] and/or rest-frame UV emission. The galaxies are ordered by redshift and each stamp is 3 arcsec on the side, with north to the top and east to the left. The red, blue, and green contours show the [CII] channel maps at different velocity intervals. The levels of the contours and channel maps velocities are indicated in the legends. Yellow crosses indicate the location of the UV regions detected in the NIR observations. The properties of each galaxy are reported in Table 2.

nent systems as single objects in the $L_{[C\,II]}$-SFR diagram. If we associate each clump and/or galaxy with its proper UV counterparts (or lack thereof), then the resulting location on the $L_{[C\,II]}$–SFR diagram changes significantly for these objects.

Fig. 4 shows $L_{[C\,II]}$ as a function of SFR. The green line illustrates the local relation obtained by De Looze et al. (2014) and its dispersion is given by the shaded area. Results for $z > 5$ galaxies, as listed in Table 2, are shown with various symbols in Fig. 4. The SFR estimation for the $z > 5$ galaxies (and their subcomponents) is discussed in Section 2.3. The multiple-component objects (HZ2, HZ6, HZ8, HZ10, WHM5, Himiko, CR7, COS24108, and BDF3299) are split into several individual components with their own SFRs and $L_{[C\,II]}$. The nine complex systems discussed in Section 3 are broken into 20 subcomponents distributed on different regions of the $L_{[C\,II]}$–SFR plane. The location of the individual subcomponents is indicated with yellow stars, while the location of these systems by integrating the whole [CII] and UV emission (i.e. ignoring that these are actually composed of different subsystems) is indicated with blue stars. The four new [C II] non-detections presented in Section 2.2 fall below the local relation, while the $L_{[C\,II]}$ for BDF2203 places this galaxy along the De Looze et al. (2014) relation.

Once the association between [C II] emission and optical-UV counterparts is properly done, we find that the resulting distribution occupies a large area of the $L_{[C\,II]}$–SFR plot with a large scatter both above and below the local relation. About 19 objects of the total sample are in agreement within 1σ with the local relation, but the remaining 24 systems have deviations, either above or below the relation, up to 3σ.

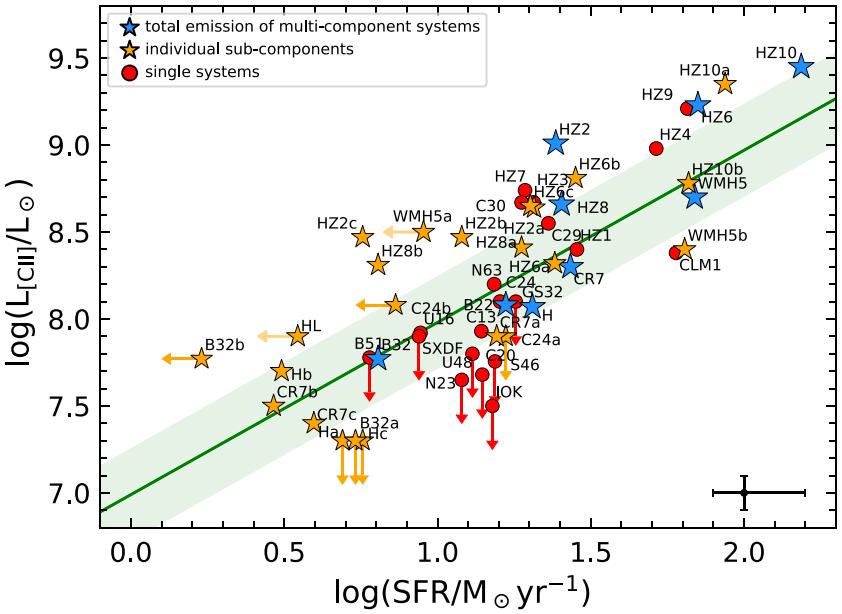

Figure 4. $L_{[\mathrm{C\,II}]}$ as a function of SFR at $z = 5-7$. Blue stars indicate the global SFR and [CII] luminosity for system having multiple subcomponents, as discussed in Section 3; in these cases, the location of the individual subcomponents are shown with yellow stars. Red circles show the location of the remaining high-z galaxies (i.e. single systems) listed in Table 2 that do not exhibit any positional offsets and/or disturbed morphology. In the bottom right of the figure, we show an error bar that is representative for the whole sample, which does not take the $\mathrm{SFR_{FIR}}$ contribution into account. The green line is the local relation for local star-forming galaxies (De Looze et al. 2014).

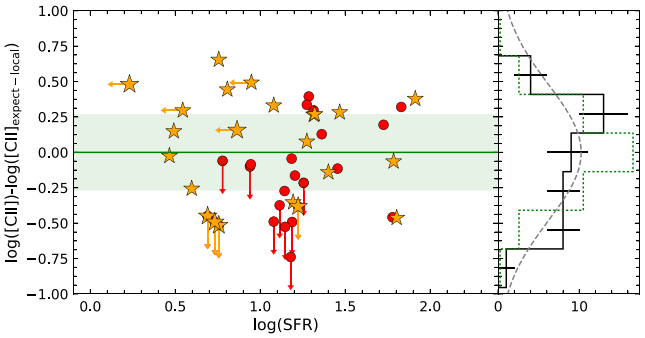

Figure 5. Deviations from the local $L_{[\mathrm{C\,II}]}$–SFR relation as a function of SFR. The 1σ dispersion of the local relation is indicated with the shaded green region. The distribution of the offsets is shown in the right-hand panel. The black histogram represents the $\log([\mathrm{C\,II}])$–$\log([\mathrm{C\,II}]_{\mathrm{expect\text{-}local}})$ distribution from our sample, while the dotted green histogram is the distribution of the local relation. The dashed grey line shows a Gaussian fit to the high-z distribution.

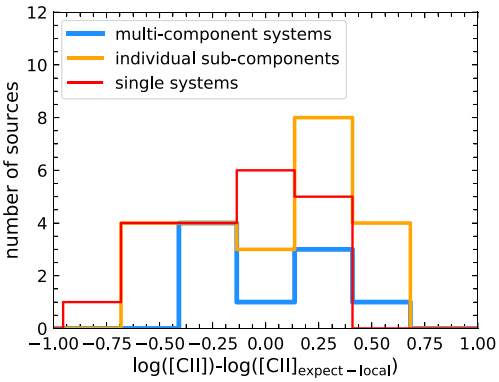

Figure 6. $\log([\mathrm{C\,II}])$–$\log([\mathrm{C\,II}]_{\mathrm{expect\text{-}local}})$ distributions for multicomponent systems (blue), individual subcomponents (orange), and single systems (red).

In order to quantify the $L_{[\mathrm{C\,II}]}$-SFR offset of the high-z sample from the relation found in the local population, we investigate the distribution of offsets relative to the local relation. More specifically, for each galaxy we calculate the offset from the relation as $\Delta\log([\mathrm{C\,II}]) = \log([\mathrm{C\,II}])$–$\log([\mathrm{C\,II}]_{\mathrm{expect\text{-}local}})$, where $\log([\mathrm{C\,II}]_{\mathrm{expect\text{-}local}})$ is the [CII] luminosity expected from the local relation according to the SFR measured in the galaxy or subcomponent. The result of this distribution is shown in the left-hand panel of Fig. 5, while the right-hand panel shows the distribution of the offsets. In contrast with some previous claims based on fewer targets, the $\Delta\log([\mathrm{C\,II}])$ distribution, which includes both detections and upper limits, does not exhibit any clear shift relative to the local relation (and whose distribution is shown with the dotted green histogram. The number of objects below (20) and above (23) the $L_{[\mathrm{C\,II}]}$–SFR relation are comparable. Instead, if we look at the $\Delta\log([\mathrm{C\,II}])$ distributions obtained from the total emission of the multicomponents and single systems (Fig. 6), we notice that such distributions have a tail towards negative values that leads to the $L_{[\mathrm{C\,II}]}$–SFR offset claimed in previous studies. This offset is mitigated when we analyse the subcomponents separately. In addition, we note that previous $L_{[\mathrm{C\,II}]}$–SFR offset claims were mainly due to the various upper limits on [CII] emission that can be underestimated as shown in Carniani et al. (2018) for Himiko.

Although the new high-z $L_{[\mathrm{C\,II}]}$–SFR relation does not reveal any offset, we estimate a dispersion of 0.48 ± 0.07, which is about 1.8 times larger than the uncertainty reported by De Looze et al. (2014) for the local relation. Such larger dispersion may be associated with the presence of kpc-scale subcomponents that are not common in the local Universe. Indeed, subcomponents with either

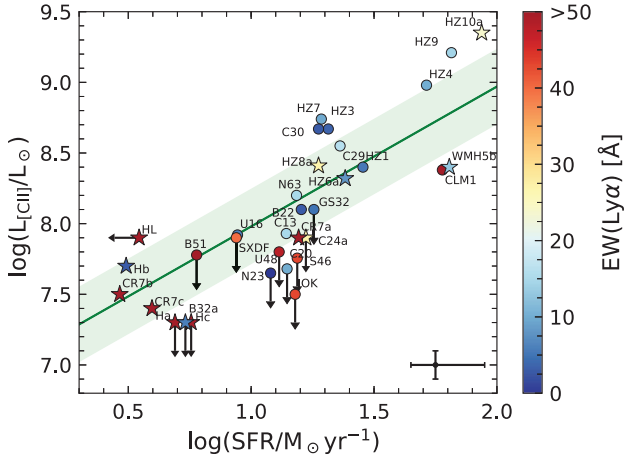

Figure 7. $L_{[\mathrm{C\,II}]}$ as a function of SFR at $z = 5-7$. Stars and circles represent subcomponents and individual galaxies, respectively. Symbols are colour-coded according to their Ly α EWs (not corrected for the inter-galactic medium absorption), as indicated on the colour bar on the right. In the bottom-right corner, we show an error bar that is representative for the whole sample. The green lines are the local relation for local star-forming galaxies (De Looze et al. 2014), while its dispersion is indicated by the shaded green region.

different properties or evolutionary stages may span a large range of [CII] luminosities for a fixed SFR value. For example, we know that the metallicity content alters the intensity of the [CII] emission up to one order of magnitude (Vallini et al. 2015; Olsen et al. 2017). We will discuss the possible origin of this dispersion in detail in the next sections.

5 THE RELATION BETWEEN LY α EW AND [CII] EMISSION

It is well known that Ly α emission depends on the level of ionizing photons produced by star formation (or AGN activity) and radiative transfer effects in the ISM. Models and observations suggest that the Ly α EW increases with decreasing metallicity and dust content (Raiter, Schaerer & Fosbury 2010; Song et al. 2014). Since the [C II] emission is sensitive on the ISM properties as well, and in particular, the ISM heating through photoelectric ejection from dust grains, we expect a relation between the [CII] luminosity and the Ly α strength (Harikane et al. 2017).

In Fig. 7, we show $L_{[\mathrm{C\,II}]}$ as function of SFR by colour-coding the different symbols according to their Ly α EWs (not corrected for the inter-galactic medium absorption). We include only those subcomponents and galaxies having a Ly α EW measurement. There is a weak tendency for galaxies with high EW(Ly α) to lie below the local $L_{[\mathrm{C\,II}]}$–SFR relation, and vice versa. This is shown better in Fig. 8 where the offset from the local $L_{[\mathrm{C\,II}]}$–SFR relation is plotted as a function of EW(Ly α). Note that, since the slope of the local relation is one, plotting the deviation from the local relation ($\Delta\log([\mathrm{C\,II}]) = \log([\mathrm{C\,II}]) - \log([\mathrm{C\,II}]_{\mathrm{expect-local}})$) is equivalent to plotting the ratio between [C II] luminosity and SFR, i.e. $L_{[\mathrm{C\,II}]}$/SFR, which is indeed given on the right-hand axis of Fig. 8. Although there is a large dispersion, there is a tentative indication that the offset from the local relation (hence, $L_{[\mathrm{C\,II}]}$/SFR) anticorrelates with Ly α EW. A linear fit gives

$$\Delta\log([\mathrm{C\,II}]) = (0.55 \pm 0.20) - (0.44 \pm 0.15)\log(\mathrm{EW(Ly\,\alpha)})$$

and the dispersion around this best fit is 0.25 dex.

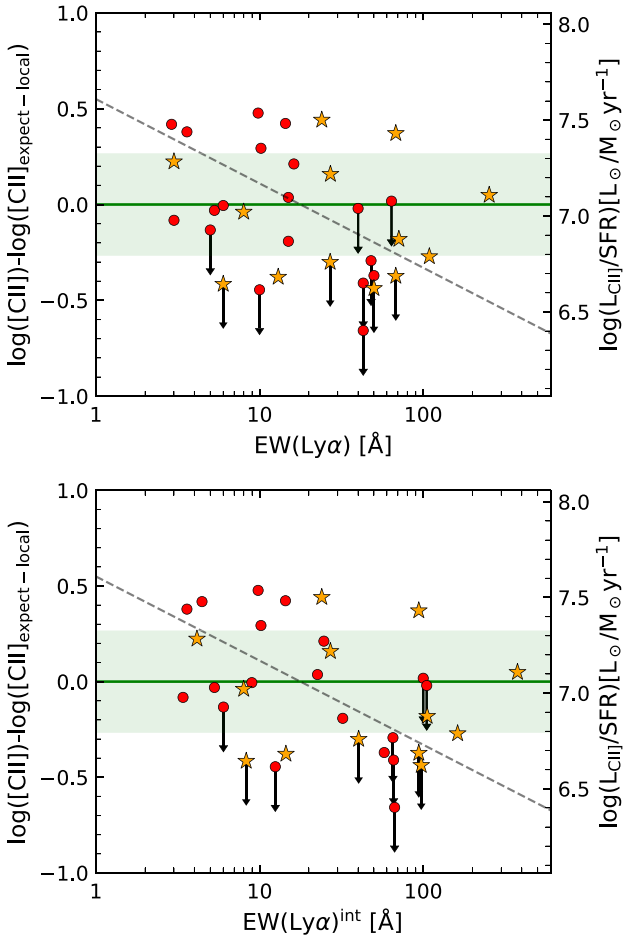

Figure 8. Top panel: Offset from the local $L_{[\mathrm{C\,II}]}$–SFR relation as a function of EW(Ly α). The right-hand axis shows the [CII]/SFR ratio corresponding to the deviations from the local $L_{[\mathrm{C\,II}]}$–SFR relation. The 1σ dispersion of the local relation is indicated by the shaded green region. Star and circles indicate subcomponents and individual galaxies, respectively. The grey dashed relationship shows our linear fit (see the text). Bottom panel: Same as the top panel but the EW(Ly α)$^{\mathrm{int}}$ has been corrected for the IGM absorption by following Harikane et al. (2017).

The result does not change significantly if we attempt to correct the EW(Ly α) for IGM absorption, by following the prescription given by Harikane et al. (2017). In this case the relation is shown in the bottom panel of Fig. 7 and the resulting best-fitting linear relation is

$$\Delta\log([\mathrm{C\,II}]) = (0.56 \pm 0.20) - (0.41 \pm 0.14)\log(\mathrm{EW(Ly\,\alpha)^{int}})$$

and the dispersion around this best fit is even larger 0.3 dex.

Harikane et al. (2017) find a relation between $L_{[\mathrm{C\,II}]}$/SFR and EW(Ly α) slightly steeper[2] than ours, though consistent within errors. The slope difference between the two relations is probably associated with measurement uncertainties due to the fact that Harikane et al. (2017) used the global properties of multicomponent galaxies, while in our analysis, we extract the $L_{[\mathrm{C\,II}]}$ and SFR from the individual subcomponents.

Some anticorrelation between deviation from the local $L_{[\mathrm{C\,II}]}$–SFR relation and EW(Ly α) (or, equivalently, between $L_{[\mathrm{C\,II}]}$–SFR and

[2]Harikane et al. (2017): $\log(L_{[\mathrm{C\,II}]}/\mathrm{SFR}) = (-0.6 \pm 0.2)\log(\mathrm{EW(Ly\,\alpha)^{int}}) + (7.6 \pm 0.3)$.

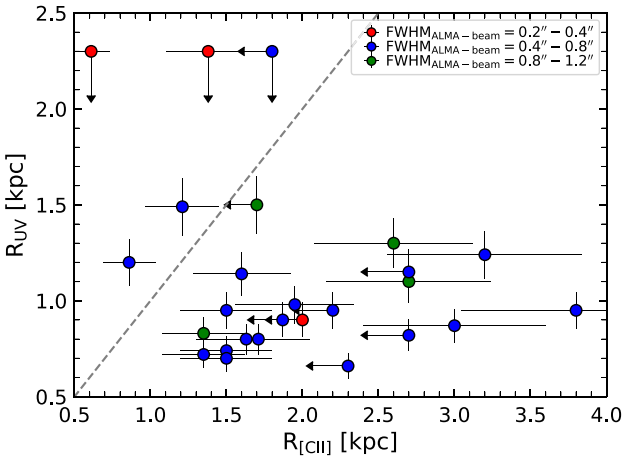

Figure 9. Half-light radii of star formation regions, as measured from the (rest-frame) UV light, compared with the half-light radii of the associated [C II] emission. Symbols are colour-coded according to the FWHM of ALMA beam. The dashed line indicates the 1:1 relation.

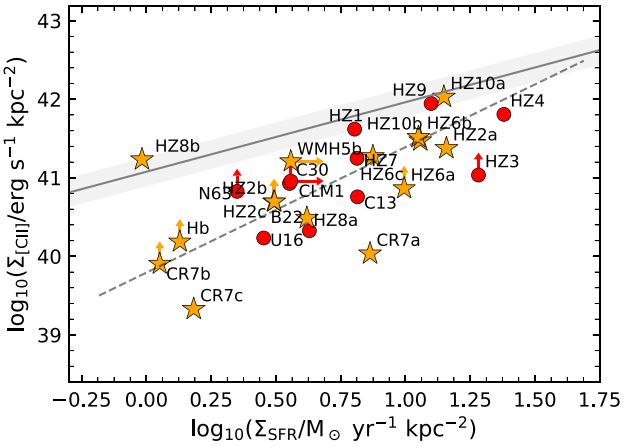

Figure 10. $\Sigma_{[C\,II]}$ versus Σ_{SFR} for $z > 5$ star-forming galaxies detected both in [C II] and rest-frame UV. Symbols are as in Fig. 4. The grey line indicates the local relation by Herrera-Camus et al. (2015). The dashed line indicates the best linear fit on the $z > 5$ sample.

EW(Ly α)^{int}) is expected from the dependence of these quantities from the metallicity, either directly or through the associated dust content, as already predicted by some models (e.g. Vallini et al. 2015; Pallottini et al. 2017b). Indeed, lower metallicity implies lower amount of carbon available for cooling, but also less dust content. Indeed, since the heating of PDRs occurs primarily through photoelectric effect on dust grains, the lower is the dust content the lower is the heating efficiency of the gas in the PDR, hence the lower is the emissions of the [CII] cooling line. On the other hand, the lower is the dust content the lower is the absorption of the Lyα resonant line, hence the higher is the EW(Ly α).

6 SPATIALLY RESOLVED $L_{[C\,II]}$–SFR RELATION

In the previous sections, we show that the $L_{[C\,II]}$–SFR relation at $z > 5$ has a intrinsic dispersion larger than observed in the local Universe. Such a large scatter suggests that the [C II] luminosity may not be good tracer of the SFR at least at early epochs.

Recent spatially resolved studies have claimed that the [C II]–SFR relation is better behaved in terms of SFR surface density and [CII] surface brightness than in global proprieties ($L_{[C\,II]}$ and SFR), since the surface brightness calibration is more closely related to the local UV field (Herrera-Camus et al. 2015; Smith et al. 2017). It is thus worth to analyse the relation $\Sigma_{[C\,II]}$–Σ_{SFR} at $z > 5$.

In this section, we compare the spatial extension of the [CII] and UV emission and, then, we investigate the correlation between [C II] surface brightness and SFR surface density.

6.1 Spatial extension of the [C II] emission

Fig. 9 shows the extension of the [C II] emission compared with the extension of star formation traced by the UV counterpart. [C II] emission is generally much more extended than the UV emission tracing unobscured star formation. This discrepancy may be partially associated with observational effects. Indeed, while the high angular resolution of *HST* enables to resolve small clumps, it may have low sensitivity to diffuse, extended emission. However, in Appendix B we show, by using Monte Carlo simulations, that the size difference is not associated with either angular resolution or our method to estimate the source-size of the UV and [CII] emission. The simulations show that if the UV and [C II] regions had the same surface brightness profile, we would have obtained a larger number of points along the 1:1 relation. On the other hand, the resolution of the [C II] observations may, in some case, smear out clumps and result in an overall extended distribution. However, in many cases the ALMA observations achieve a resolution comparable, or even higher, than *HST* at UV rest-frame wavelengths and, despite this, we measure clearly larger [C II] sizes. Moreover, when high angular resolution observation are used, these do reveal that a significant fraction of the [C II] flux is resolved out on large scales (see e.g. discussion in Carniani et al. 2017).

In conclusion, we believe that the different sizes between [C II] and UV emission are tracing truly different distribution of the [C II] and UV emission on different scales.

There could be various explanations for these differences. If the star formation associated with the [C II] emission on large scale is heavily obscured, the UV light does not trace this component (Katz et al. 2017). In alternative, the extended component of [C II] may not be directly associated with star forming regions, but with circumgalactic gas, either in accretion or ejected by the galaxy, and which is illuminated by the strong radiation field produced by the galaxy.

6.2 Surface brightness

Once we have measured the extension of the [C II] emission and of the SFR regions, we can estimate the [C II] surface brightness and SFR surface density of each subcomponent and individual source detected in [CII] and UV. Fig. 10 shows the $\Sigma_{[C\,II]}$–Σ_{SFR} relation, where we have included only those galaxies detected in both [C II] and UV emission. Systems that are not spatially resolved in [C II] (or UV) emission are indicated with lower limits. We also show the local relation by Herrera-Camus et al. (2015) and its dispersion. In contrast to the $L_{[C\,II]}$–SFR diagram, there are no galaxies located significantly above the local relation, only a few galaxies are located on the local relation, and most galaxies spread largely below the local relation. This is primarily due to the large extension of the [CII] emission in these high redshift systems, as discussed in the previous section. By fitting the $\Sigma_{[C\,II]}$–Σ_{SFR} measurements for our

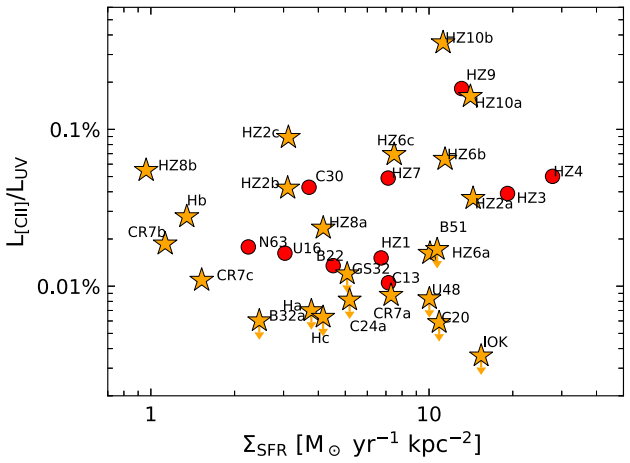

Figure 11. $L_{[C\,II]}/L_{UV}$ ratio as a function of the SFR surface density. Symbols are as in Fig. 4.

sample, we obtain the following relation:

$$\log(\Sigma_{SFR}) = (0.63 \pm 0.11) \times \log(\Sigma_{[C\,II]}) - (25 \pm 6).$$

Part of the $\Sigma_{[C\,II]}$ deficit may be ascribed to the metallicity of the gas. Indeed a similar deviations have been observed in local low-metallicity galaxies, in which the $\Sigma_{[C\,II]}$–Σ_{SFR} calibration overpredicts the $\Sigma_{[C\,II]}$ by up to a factor of 6. We note that Faisst et al. (2017) estimated a metallicity of 12 + log(O/H) > 8.5 for HZ10 and HZ9, and a metallicity of 12 + log(O/H) < 8.5 for HZ1, HZ2, and HZ4. The former agree with the local $\Sigma_{[C\,II]}$–Σ_{SFR} relation, while HZ1, HZ2, and HZ4 have low $\Sigma_{[C\,II]}$–Σ_{SFR} ratio. This interpretation is also supported by the fact that all CR7 clumps, which have metallicities 12 + log(O/H) < 8.2 (see discussion in Sobral et al. 2017), are located well below the local relation. The $\Sigma_{[C\,II]}$ deficit is also akin to the simulations by Pallottini et al. (2017a) who investigated the physical properties of a simulated galaxy at $z = 6$, with metallicity 12 + log(O/H) = 8.35. Such galaxy has a $\log(\Sigma_{[C\,II]}/\mathrm{erg\ s^{-1}\ kpc^{-2}}) = 40.797$ and $\log(\Sigma_{SFR}/\mathrm{M_\odot\ yr^{-1}}) = 1.027$, which places the mock galaxy below the $\Sigma_{[C\,II]}$–Σ_{SFR} relation found for local star-formation galaxies.

However, the offset observed in high-z galaxies is significantly larger (one order of magnitude or more) than the one observed in local low-metallicity galaxies, so other effects are likely in place, which will be discussed in the next section. Finally, still within the context of the surface brightness and SFR surface density, we mention that, for local galaxies, Smith et al. (2017) found a dependence between the $L_{[C\,II]}/L_{FIR}$ ratio and the SFR surface density, suggesting that the [C II] deficit increases strongly with increasing Σ_{SFR}. Therefore, the Σ_{SFR} dependence has a strong impact also on the $L_{[C\,II]}$-SFR relation. In our sample, we cannot verify this dependence as most galaxies have only an upper limit on the L_{FIR}. However, we can investigate the relation $L_{[C\,II]}/L_{UV}$ and Σ_{SFR}, since the contribution of SFR_{FIR} to the total SFR is negligible (see discussion in Section 2.3). Fig. 11 shows the observed $L_{[C\,II]}/L_{UV}$ ratio spanning a range between 0.002 per cent and 0.4 per cent (over two orders of magnitude), while the SFR surface density spans the range between 1 and 30 $\mathrm{M_\odot\ yr^{-1}\ kpc^{-2}}$. There is only a very weak correlation between $L_{[C\,II]}/L_{UV}$ and Σ_{SFR}, with very large dispersion (much larger than what observed locally for the $L_{[C\,II]}/L_{FIR}$ and Σ_{SFR} relation). This indicates that in these high-z systems, the $L_{[C\,II]}/L_{UV}$ line ratio (and $L_{[C\,II]}$/SFR) does not strongly depend on the areal density with which galaxies form stars at $z > 5$, and that other effects (metallicity, and other phenomena discussed in the next section) may contribute to the very large dispersion. Interestingly, we note that $L_{[C\,II]}/L_{UV}$ ratio in single systems (red circles) seems to increase with increasing Σ_{SFR}. Since the Σ_{SFR} is correlated to the molecular and metal content as natural consequence of the Schmidt–Kennicutt law, the putative $L_{[C\,II]}/L_{UV}$ trend could trace the dust content in these high-z systems. We also note that systems with high Σ_{SFR} are closer to the local $\Sigma_{[C\,II]}$–Σ_{SFR} relation supporting the fact that the [C II] deficit is lead by the metal enrichment of the system.

7 DISCUSSION ON THE [C II]–SFR SCALING RELATIONS AT $z = 5$–7

Clearly galaxies at $z > 5$ behave differently, relative to their local counterparts, for what concerns the [C II] and SFR properties. Summarizing the finding of the previous sections: (1) both the $L_{[C\,II]}$–SFR relation and the $\Sigma_{[C\,II]}$–Σ_{SFR} relation have a scatter much larger than the local relations; (2) contrary to some previous claims, the $L_{[C\,II]}$–SFR relation is not offset relative to the local relation, while the $\Sigma_{[C\,II]}$–Σ_{SFR} relation is clearly offset by showing much lower $\Sigma_{[C\,II]}$ relative to local relation; (3) the extension of the [CII] emission is larger than the extent of the star formation traced by the UV emission.

The larger scatter observed in the SFR–$L_{[C\,II]}$ relation is certainly indicative of a broader range of properties spanned by such primeval galaxies relative to the local population. Indeed, high-z cosmological simulations show that the [C II] emission strongly depends on the gas metallicity, ionization parameter, and evolutionary stage of the system and that all of these properties are expected to span a much broader range at high-z with respect to local galaxies (Vallini et al. 2015, 2017; Pallottini et al. 2015, 2017b; Olsen et al. 2017; Katz et al. 2017). Recently, Lagache, Cousin & Chatzikos (2018) investigated the expected dispersion of $L_{[C\,II]}$–SFR relation in the distant Universe by using a semi-analytical model of galaxy formation for a large sample of simulated galaxies at $z > 4$. They found an $L_{[C\,II]}$–SFR correlation with a large scatter of 0.4–0.8 dex, which is in agreement with our result. They claim that such large dispersion is associated with the combined effects of different gas contents, metallicities, and interstellar radiation fields in the simulated high-z galaxies.

As we have shown, the mild anticorrelation with EW(Ly α), as well as the analysis of some individual galaxies for which the metallicity has been estimated, does suggest that the metallicity may play a role on the [CII] emission (either simply in terms of carbon abundance and/or in terms of heating of the ISM through photoelectric effect on dust grains, whose abundance scales with the metallicity). In particular, the lower metallicity of galaxies at $z > 5$ can explain some of the scatter towards low [C II] emission in both the $L_{[C\,II]}$–SFR relation and in the $\Sigma_{[C\,II]}$–Σ_{SFR} relation. However, metallicity effects are unlikely to explain the scatter towards high [C II] emission in the $L_{[C\,II]}$–SFR relation. Moreover, for what concerns the $\Sigma_{[C\,II]}$–Σ_{SFR} relation, the offset and large spread towards low $\Sigma_{[C\,II]}$ is probably too large to be entirely ascribed to metallicity.

Large variations in ionization parameter can also contribute to the spread in [C II] emission (Graciá-Carpio et al. 2011; Katz et al. 2017). In particular, if [C II] in primeval galaxies also traces circumgalactic gas in accretion and/or expelled from the galaxy, and excited by the UV radiation of the central galaxy, this would result into the observed larger [C II] sizes, hence lower $\Sigma_{[C\,II]}$ and lower ionization parameter, and the latter would increase the total $L_{[C\,II]}$ relative to the local relation. On the other hand, young compact star forming, primeval galaxies would be characterized by higher

ionization parameter, which would reduce the [C II] emission, hence contributing to the spread towards low $L_{[C\,II]}$.

As mentioned in the previous sections, yet another possibility is that the UV emission associated with some of the [C II] clumps is heavily obscured. This would explain the scatter above the $L_{[C\,II]}$–SFR relation, in the sense that for some of these systems the underlying SFR, as traced by the UV, is heavily underestimated. This can be the case for a few galaxies. However, as we mentioned, most of these systems show no or weak continuum dust emission, indicative of low dust content.

To make further progress additional data at other wavelength will be, in the future, extremely valuable. In particular, *James Webb Space Telescope* will enable to identify obscured stellar components as well as H α emission associated with star formation. ALMA observations of other transitions, such has [OIII]88μm (though observable only in some redshift ranges), has proved extremely useful to constrain these scenarios (e.g. Inoue et al. 2016; Carniani et al. 2017).

8 CONCLUSIONS

In this work, we have investigated the nature of the [CII] emission in star-forming galaxies at $z = 5-7$. In particular, we have explored the positional offsets between UV and FIR line emissions, the presence of multiple components, and the implications on the [C II]–SFR scaling relations in the distant Universe, once the correct association between [CII] and UV emission is properly taken into account. We have performed our investigation in a sample of 29 $z > 5$ 'normal' star-forming galaxies (SFR $< 100\,M_\odot\,yr^{-1}$) observed with ALMA in the [C II] line. In addition to the re-analysis of archival objects, we have also included new ALMA observations targeting five star-forming galaxies at $z \sim 6$ with SFR $\sim 15\,M_\odot\,yr^{-1}$, resulting into a new detection. Our main results are as follows:

(i) The continuum emission around 158 μm is not detected in most of the $z > 5$ galaxies observed with ALMA, indicating a low dust content. By modelling the dust emission with a greybody spectrum with dust temperature $T_d = 30$ K and emissivity index $\beta = 1.5$, we have found that the SFR based on the FIR emission is, on average, lower than the SFR measured from the UV emission by at least a factor of 0.6, but probably much more (due to several upper limits).

(ii) By accurately registering ALMA and NIR images, and by kinematically discriminating multiple [C II] components, our analysis has revealed that the [C II] emission breaks into multiple subcomponents in 9 out of the 21 galaxies having [CII] detections. In these nine targets, we have observed the presence of 19 FIR-line emitting clumps. Only very few of these, if any, are associated with the primary (brightest) UV counterpart, while the bulk of the [CII] is associated with fainter UV components. In only three cases (COSMOS24108, Himiko, and BDF3299), the shallow NIR images have not enabled us to detect the UV counterparts associated with some of the [CII] clumps.

(iii) We have studied the relation between [C II] and SFR on the high-z sample by taking into account the presence of these subcomponents and the proper associations between [CII] and UV components. The distribution of $z > 5$ galaxies on the $L_{[C\,II]}$–SFR diagram follows the local relation, but the dispersion is 1.8 times larger than that observed in nearby galaxies.

(iv) The deviation from the local $L_{[C\,II]}$–SFR relation shows a weak anticorrelation with EW(Ly α) though shallower and with larger dispersion than what found in other studies that did not account for the multicomponent nature of these systems.

(v) Most of the objects in the high-z sample are spatially resolved in [C II] and UV emission. The extension of the [C II] emission is generally much larger than the extension of star forming regions traced by the UV emission.

(vi) In the $\Sigma_{[C\,II]}$–Σ_{SFR} diagram, $z > 5$ galaxies are characterized by a large scatter with respect to local galaxies and are mostly distributed below the local relation (i.e. fainter $\Sigma_{[C\,II]}$ at a given Σ_{SFR}).

We have suggested that a combination of different effects may be responsible for the different properties of high-z galaxies in terms of [C II]–SFR properties relative to local galaxies. More specifically: (1) the low metallicity of high-z galaxies may be responsible (also indirectly through the lower dust photoelectric heating) for part of the scatter towards lower [C II] emission relative to the local relations; (2) the presence of circumnuclear gas in accretion and/or expelled from the galaxy may be responsible for the larger size in [CII] relative to the SFR distribution and may also be responsible for the scatter of the $L_{[C\,II]}$–SFR distribution above the local relation as a consequence of lower ionization parameter; (3) in compact young star-forming regions the increased ionization parameter and higher gas density may be responsible for the suppression of [CII] for galaxies which are below the local relation; (4) dust obscuration may be responsible for both the different morphology between [CII] and UV emission and also for the scatter of sources above the local $L_{[C\,II]}$–SFR relation.

ACKNOWLEDGEMENTS

This paper makes use of the following ALMA data: ADS/JAO.ALMA#2012.1.00719.S, ADS/JAO.ALMA#2012.A.00040.S, ADS/JAO.ALMA#2013.A.00433.S, ADS/JAO.ALMA#2011.0.00115.S, ADS/JAO.ALMA#2012.1.00033.S, ADS/JAO.ALMA#2012.1.00523.S, ADS/JAO.ALMA#2013.1.00815.S, ADS/JAO.ALMA#2015.1.00834.S., ADS/JAO.ALMA#2015.1.01105.S, AND ADS/JAO.ALMA#2016.1.01240.S which can be retrieved from the ALMA data archive: https://almascience.eso.org/ alma-data/archive. ALMA is a partnership of ESO (representing its member states), NSF (USA) and NINS (Japan), together with NRC (Canada) and NSC and ASIAA (Taiwan), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO, and NAOJ. We are grateful to G. Jones to for providing his [C II] flux maps. RM and SC acknowledge support by the Science and Technology Facilities Council (STFC). RM acknowledges ERC Advanced Grant 695671 'QUENCH'. AF acknowledges support from the ERC Advanced Grant INTERSTELLAR H2020/740120.

REFERENCES

Aravena M. et al., 2016, ApJ, 833, 71
Barisic I. et al., 2017, ApJ, 845, 41
Bowler R. A. A., Dunlop J. S., McLure R. J., McLeod D. J., 2017, MNRAS, 466, 3612
Bradač M. et al., 2017, ApJ, 836, L2
Capak P. L. et al., 2015, Nature, 522, 455
Carilli C. L., Walter F., 2013, ARA&A, 51, 105
Carniani S. et al., 2013, A&A, 559, A29
Carniani S. et al., 2015, A&A, 584, A78
Carniani S. et al., 2017, A&A, 605, A42
Carniani S., Maiolino R., Smit R., Amorín R., 2018, ApJ, 854, L7

Ceverino D., Primack J., Dekel A., Kassin S. A., 2017, MNRAS, 467, 2664
Cutri R. M. et al., 2013, VizieR Online Data Catalog, 2328
De Barros S. et al., 2017, A&A, 608, A123
De Breuck C., Maiolino R., Caselli P., Coppin K., Hailey-Dunsheath S., Nagao T., 2011, A&A, 530, L8
De Looze I. et al., 2014, A&A, 568, A62
Decarli R. et al., 2017, Nature, 545, 457
Dunlop J. S. et al., 2017, MNRAS, 466, 861
Elmegreen B. G., Elmegreen D. M., Fernandez M. X., Lemonias J. J., 2009, ApJ, 692, 12
Faisst A. L. et al., 2017, ApJ, 847, 21
Fiacconi D., Mayer L., Madau P., Lupi A., Dotti M., Haardt F., 2017, MNRAS, 467, 4080
Förster Schreiber N. M. et al., 2006, ApJ, 645, 1062
Gaia Collaboration et al., 2016, A&A, 595, A2
Gallerani S. et al., 2012, A&A, 543, A114
Gallerani S., Pallottini A., Feruglio C., Ferrara A., Maiolino R., Vallini L., Riechers D. A., Pavesi R., 2018, MNRAS, 473, 1909
Genzel R. et al., 2008, ApJ, 687, 59
González-López J. et al., 2014, ApJ, 784, 99
Graciá-Carpio J. et al., 2011, ApJ, 728, L7
Guo Y. et al., 2015, ApJ, 800, 39
Hao C.-N., Kennicutt R. C., Johnson B. D., Calzetti D., Dale D. A., Moustakas J., 2011, ApJ, 741, 124
Harikane Y. et al., 2018, ApJ, 859, 84
Herrera-Camus R. et al., 2015, ApJ, 800, 1
Huang K.-H., Ferguson H. C., Ravindranath S., Su J., 2013, ApJ, 765, 68
Inoue A. K. et al., 2016, Science, 352, 1559
Jiang L. et al., 2013, ApJ, 773, 153
Jones G. C. et al., 2017a, ApJ, 850, 180
Jones G. C., Willott C. J., Carilli C. L., Ferrara A., Wang R., Wagg J., 2017b, ApJ, 845, 175
Kapala M. J. et al., 2015, ApJ, 798, 24
Katz H., Kimm T., Sijacki D., Haehnelt M. G., 2017, MNRAS, 468, 4831
Kennicutt R. C., Evans N. J., 2012, ARA&A, 50, 531
Knudsen K. K., Richard J., Kneib J.-P., Jauzac M., Clément B., Drouart G., Egami E., Lindroos L., 2016, MNRAS, 462, L6
Lagache G., Cousin M., Chatzikos M., 2018, A&A, 609, A130
Maiolino R. et al., 2005, A&A, 440, L51
Maiolino R., Caselli P., Nagao T., Walmsley M., De Breuck C., Meneghetti M., 2009, A&A, 500, L1
Maiolino R. et al., 2012, MNRAS, 425, L66
Maiolino R. et al., 2015, MNRAS, 452, 54
Matthee J., Sobral D., Santos S., Röttgering H., Darvish B., Mobasher B., 2015, MNRAS, 451, 400
Matthee J. et al., 2017, ApJ, 851, 145
McMullin J. P., Waters B., Schiebel D., Young W., Golap K., 2007, in Shaw R. A., Hill F., Bell D. J., eds, ASP Conf. Ser. Vol. 376, Astronomical Data Analysis Software and Systems XVI. Astron. Soc. Pac., San Francisco, p. 127
Murphy E. J. et al., 2011, ApJ, 737, 67
Olsen K., Greve T. R., Narayanan D., Thompson R., Davé R., Niebla Rios L., Stawinski S., 2017, ApJ, 846, 105
Ota K. et al., 2014, ApJ, 792, 34
Ouchi M. et al., 2010, ApJ, 723, 869
Ouchi M. et al., 2013, ApJ, 778, 102
Pallottini A., Gallerani S., Ferrara A., Yue B., Vallini L., Maiolino R., Feruglio C., 2015, MNRAS, 453, 1898
Pallottini A., Ferrara A., Bovino S., Vallini L., Gallerani S., Maiolino R., Salvadori S., 2017a, MNRAS, 471, 4128
Pallottini A., Ferrara A., Gallerani S., Vallini L., Maiolino R., Salvadori S., 2017b, MNRAS, 465, 2540
Paulino-Afonso A. et al., 2018, MNRAS, 476, 5479
Pentericci L. et al., 2016, ApJ, 829, L11
Planck Collaboration et al., 2016, A&A, 594, A13
Raiter A., Schaerer D., Fosbury R. A. E., 2010, A&A, 523, A64
Ravindranath S. et al., 2006, ApJ, 652, 963
Riechers D. A. et al., 2014, ApJ, 796, 84
Robertson B. E., Ellis R. S., Furlanetto S. R., Dunlop J. S., 2015, ApJ, 802, L19
Schaerer D. et al., 2015, A&A, 576, L2
Schreiber C. et al., 2018, A&A, 611, A22
Smit R. et al., 2018, Nature, 553, 178
Smith J. D. T. et al., 2017, ApJ, 834, 5
Sobral D., Matthee J., Darvish B., Schaerer D., Mobasher B., Röttgering H. J. A., Santos S., Hemmati S., 2015, ApJ, 808, 139
Sobral D. et al., 2017, preprint (arXiv:1710.08422)
tSong M. et al., 2014, ApJ, 791, 3
Tamburello V., Mayer L., Shen S., Wadsley J., 2015, MNRAS, 453, 2490
Trakhtenbrot B., Lira P., Netzer H., Cicone C., Maiolino R., Shemmer O., 2017, ApJ, 836, 8
Vallini L., Gallerani S., Ferrara A., Pallottini A., Yue B., 2015, ApJ, 813, 36
Vallini L., Ferrara A., Pallottini A., Gallerani S., 2017, MNRAS, 467, 1300
Wagg J. et al., 2012, ApJ, 752, L30
Williams R. J. et al., 2014, MNRAS, 439, 2096
Willott C. J., Bergeron J., Omont A., 2015a, ApJ, 801, 123
Willott C. J., Carilli C. L., Wagg J., Wang R., 2015b, ApJ, 807, 180
Yun M. S. et al., 2015, MNRAS, 454, 3485

APPENDIX A: ADDITIONAL ALMA DATA – OBSERVATION AND DATA REDUCTION

The five $z\sim6$ star-forming galaxies listed in Table A1 were observed with ALMA in band 6 during Cycle 3 and Cycle 4 (programme ID #2015.1.01105.S and #2015.1.01240.S). ALMA observations were carried out with a semicompact array configuration with angular resolutions ranging from 0.3 to 0.8 arcsec. The sources were observed for a total on source integration time of 0.1–0.8 h with a precipitable water vapour of 0.4–1.4mm, depending on the specific observation. For each target we used four spectral windows (SPWs) set up in frequency division mode with a spectral resolution of $\sim$30 MHz ($\sim$35 km s^{-1}) and bandwidth of 1.875 GHz. One of the four SPWs was tuned to the expected frequency of the [C II] line. The phase of each observation was centred at the NIR position of the source.

J2248-3235, J0948-002, J1147-0724, J0239-0234, and J0552-3627 were observed as phase calibrator, respectively, for the four sources. The flux calibrators were J2056-4714, Ganymede, J1229+0203, J1229+0623, and J0334-4008, while bandpass calibrations were carried out through the observations of J2258-2758, J1058+0133, J1229+0203, J0238+1636, and J0522-3627.

ALMA observations were calibrated by using CASA software version v4.5.2 (McMullin et al. 2007). Continuum and data cube images were obtained by using the CASA task CLEAN and natural weighting. The final angular resolution and sensitivity reached in each set of data are listed in Table A1.

We registered NIR images to ALMA observations by matching the location of the foreground sources and ALMA calibrators to the position given by the GAIA Data Release 1 catalogue (Gaia Collaboration et al. 2016).

As discussed in Section 2.2, [C II] emission was detected only in one target, BDF3203. Figs 1 and A1 show the ALMA [C II] spectra for the five star-forming galaxies extracted from a region as large as the ALMA beam and centred at the location of the rest-frame UV regions.

Table A1. UV and FIR properties of the new five $z \sim 6-7$ sources observed with ALMA.

Name	RA (J2000) (deg)	Dec. (J200) (deg)	$z_{\mathrm{Ly}\,\alpha}$ [a]	$\mathrm{SFR}_{\mathrm{UV}}^{b}$ ($\mathrm{M}_{\odot}\ \mathrm{yr}^{-1}$)	$\nu_{\mathrm{obs}}([\mathrm{C\,II}])^{c}$ (GHz)	Beam[d] (min″ × maj″)	t_{exp}[e] (h)	$\sigma_{\mathrm{cont}}^{f}$ (μJy)	$\sigma_{\mathrm{line}}^{g}$ (mJy)	$L_{[\mathrm{C\,II}]}{}^{h}$ ($10^7\ \mathrm{L}_{\odot}$)
NTTDF2313	181.3804	− 7.6935	6.07	12	268.817	0.98 × 0.71	0.7	18	0.15	<4.5
BDF2203	336.958	− 35.1472	6.12	16	266.93	1.90 × 1.11	0.4	23	0.2	12.5 ± 2.5
GOODS3203	53.0928	− 27.8826	6.27	18	250.236	1.26 × 1.03	0.1	41	0.4	<12.0
COSMOS20521	150.1396	2.4269	6.36	14	258.225	1.46 × 1.20	0.8	20	0.15	<4.8
UDS4821	34.4768	− 5.24728	6.561	13	251.361	0.24 × 0.22	0.3	24	0.2	<6.7

$_{\mathrm{Ly}\,\alpha}$[a]Redshift from either Ly α line or spectroscopic Lyman break. The uncertainty is <0.04. [b]SFR inferred from the rest-frame UV continuum adopting the calibration discussed in Kennicutt & Evans (2012). [c]Expected [C II] frequency according to z. [d]ALMA synthesized beam. [e]On-source integration time. [f]Sensitivity in ALMA continuum map. [g]Sensitivity in spectral channels of 100 km s^{-1}. [h][C II] luminosity. The upper limits on the $L_{[\mathrm{C\,II}]}$ are at 3σ and are calculated on a channel width of 100 km s^{-1}.

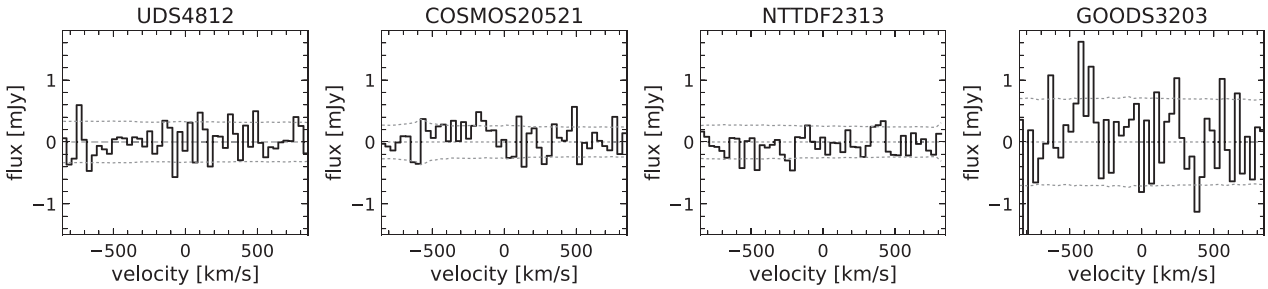

Figure A1. ALMA spectra of UDS4812, COSMOS20521, NTTDF2313, and GOODS3203. The spectra have been extracted at the location of the UV emission within a ALMA beam region. The velocity reference is set to the redshift inferred from spectroscopic rest-frame optical observations. The dotted grey lines show the 1σ and -1σ.

APPENDIX B: SOURCE-SIZE MEASUREMENTS

The extension of the UV and [C II] regions has been estimated by fitting a 2D Gaussian profile to the emission and taking into account the effect of the PSF or ALMA beam. We have used own custom-built PYTHON routine to perform the fitting to the UV emission, which returns the deconvoled half-light radius by subtracting in quadrature the PSF size. Instead we use the IMFIT task of CASA to obtain the extension of the [C II] emission.

Although recent studies have already shown the source-size measurements based on parametric method are less biased than non-parametric measurements (Huang et al. 2013; Paulino-Afonso et al. 2018), we have performed Monte Carlo simulations to evaluate the robustness of our source-size measurements. We have also assessed if the discrepancy between the extension of the UV and [C II] emission shown Fig. 9 might be due to the different angular resolutions between *HST* and ALMA observations.

Initially, we have verified the robustness of our R_{UV} estimations. We have generated 1000 mock UV sources at $z = 6$, whose surface brightnesses are described by Sérsic profiles with indices randomly selected between $n = 0.5$ and $n = 2.5$. The half-light radius of all mock sources has been fixed to $R_{\mathrm{UV}}^{\mathrm{model}} = 0.95$ kpc that is the average UV radius estimated from our high-z systems. We have also assigned to each source a random axis ratio ($0.1 < b/a < 1.0$) and position angle. After convolving the surface brightnesses with an empirical PSF derived from *HST* images using a number of isolated bright stars, we have positioned each convolved mock source in a blank region of the *HST* images in order to include the background noise. The intrinsic flux of each source has been tuned to have all synthetic observations with the same S/N. Finally, we have measured the radii by performing a 2D Gaussian fitting and subtracting the PSF size in quadrature. The top panel of Fig. B1 shows the distribution of the source-size measurements for two simulations that differ in the S/N associated with the mock sources. The source-size distribution

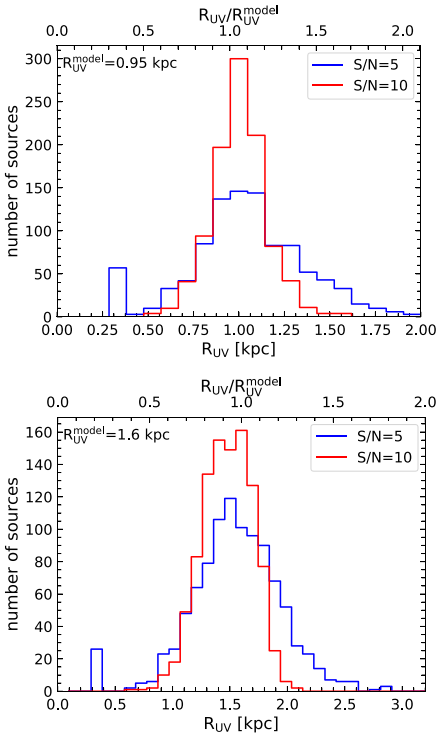

Figure B1. Distribution of the half-light radius R_{UV} estimated from 1000 simulated UV sources with S/N = 5 (blue histogram) and S/N = 10 (red histogram). Top and bottom panels show the results of the simulation for the two effective radii used in the model: $R_{\mathrm{UV}}^{\mathrm{model}} = 0.95$kpc and $R_{\mathrm{UV}}^{\mathrm{model}} = 1.6$ kpc.

for sources with S/N = 5 and S/N = 10 peak at $R_{\rm UV} = 1.0 \pm 0.3$ kpc and $R_{\rm UV} = 1.00 \pm 0.15$ kpc, respectively. We stress that the two selected values of S/N used for the mock galaxies mirror the level of confidence of our UV $z > 5$ sources and the results of the simulations evidence that our method for estimating $R_{\rm UV}$ is not biased and has uncertainty of ∼15−30 per cent depending on the S/N of the detection.

In Section 6.1, we have mentioned that the high-angular resolution of *HST* observations may miss out the diffuse and faint UV emission, so altering the size measurement of the UV regions and leading to the discrepancy between $R_{\rm UV}$ and $R_{\rm [C\,II]}$. We have thus tested the scenario where the UV regions have the same extension of the [C II] emission. We have performed the same simulations, but have set $R_{\rm UV}^{\rm model} = 1.6$ kpc that is the average $R_{\rm [C\,II]}$ of our real observations. The results of these simulations are shown in the bottom panel of Fig. B1. At S/N > 5, the radius of the emission is well estimated with a uncertainty of ∼10−20 per cent indicating that the discrepancy between $R_{\rm UV}$ and $R_{\rm [C\,II]}$ is not associated with the UV source-size measurement method.

After verifying the quality of our UV measurements, we have assessed the robustness the [C II] source-size measurement method. In this case, we have used the SIMOBSERVE task of CASA to produce synthetic interferometric observations with sensitivities similar to our real observations and angular resolution of 0.6 arcsec (∼3.5 kpc at $z = 6$) that is as large as the average ALMA beam of our data. As we have done for the UV simulations, in each synthetic observations we have included a mock source with a Sérsic profile and a half-light radius $R_{\rm [C\,II]}^{\rm model}$. We have randomly assigned a axis ratio and position angle to each mock source. We have also added a thermal noise component by setting the parameter THERMAL NOISE of SIMOBSERVE to *tsys-atm* with a precipitable water vapour of 1.0 mm and ambient temperature of 269 K, which are typical values of our observations. In each synthetic map, we have changed each time the parameter SEED with a random value, which allows us to generate a random thermal noise for each mock observation. Following this prescription, we have generated 1000 mock observations for sources having the same S/N and $R_{\rm [C\,II]}^{\rm model}$. The size of mock source has been then estimated by using IMFIT. The results of these simulations are shown in Fig. B2.

At $R_{\rm [C\,II]}^{\rm model} = 0.95$ kpc, IMFIT has returned point-source solutions for ∼50 per cent mock sources at both S/N = 5 and S/N = 10. For remaining half, the distribution of the source-size measurements have a median consistent with the $R_{\rm [C\,II]}^{\rm model}$ ($R_{\rm [C\,II]} = 0.9$ kpc and $R_{\rm [C\,II]} = 1.0$ kpc) and a dispersion of $\sigma(R_{\rm [C\,II]}) = 0.3$ kpc at S/N = 5 and $\sigma(R_{\rm [C\,II]}) = 0.2$ kpc and at S/N = 10. This result suggests that if the [C II] regions were as compact as the UV emission we would have a large number of upper limits on $R_{\rm [C\,II]}$ and a larger number of point consistent with the 1:1 relation of Fig. 9. In the case $R_{\rm [C\,II]}^{\rm model} = 1.6$ kpc, most of results are consistent with a spatially resolved emission with a $R_{\rm [C\,II]} = 1.5 \pm 0.4$ kpc and $R_{\rm [C\,II]} = 1.6 \pm 0.3$ kpc at S/N = 5 and S/N = 10, respectively.

The distribution of the source-size measurements from our simulations indicates the source-size measurements are not biased and their relative errors are not sufficient to explain the $R_{\rm [C\,II]}/R_{\rm UV} \gtrsim 2$ observed in the high-z systems.

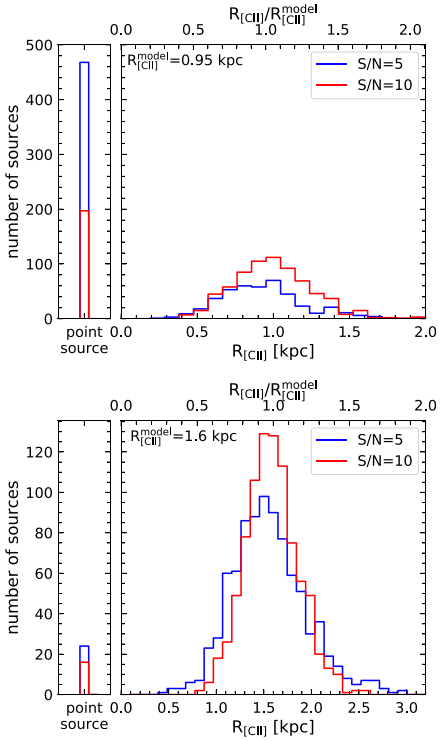

Figure B2. Distribution of the half-light radius $R_{\rm [C\,II]}$ estimated from 1000 simulated [C II] sources with S/N = 5 (blue histogram) and S/N = 10 (red histogram). The synthetic ALMA observations have been generated with the task SIMOBSERVE and with an angular resolution of 0.6 arcsec (∼3.5 kpc at z= 6). Top and bottom panels show the results of the simulation for the two effective radii used in the model: $R_{\rm [C\,II]}^{\rm model} = 0.95$ kpc and $R_{\rm [C\,II]}^{\rm model} = 1.6$ kpc.

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