

# Confirmation of double peaked Ly $\alpha$ emission at $z = 6.593$ <sup>★</sup>

## Witnessing a galaxy directly contributing to the reionisation of the Universe

Jorryt Matthee<sup>1\*\*</sup>, David Sobral<sup>2</sup>, Max Gronke<sup>3</sup>, Ana Paulino-Afonso<sup>2,4,5</sup>, Mauro Stefanon<sup>1</sup>, and Huub Röttgering<sup>1</sup>

<sup>1</sup> Leiden Observatory, Leiden University, PO Box 9513, NL-2300 RA, Leiden, The Netherlands

<sup>2</sup> Department of Physics, Lancaster University, Lancaster, LA1 4YB, UK

<sup>3</sup> Department of Physics, University of California, Santa Barbara, CA 93106, USA

<sup>4</sup> Instituto de Astrofísica e Ciências do Espaço, Universidade de Lisboa, OAL, Tapada da Ajuda, 1349-018 Lisboa, Portugal

<sup>5</sup> Departamento de Física, Faculdade de Ciências, Universidade de Lisboa, Edifício C8, Campo Grande, 1749-016 Lisboa, Portugal

May 31, 2018

### ABSTRACT

Distant luminous Lyman- $\alpha$  emitters (LAEs) are excellent targets for spectroscopic observations of galaxies in the epoch of reionisation (EoR). We present deep high-resolution ( $R = 5000$ ) VLT/X-SHOOTER observations, along with an extensive collection of photometric data of ‘COLA1’, a proposed double peaked LAE at  $z = 6.6$  (Hu et al. 2016). We rule out that COLA1’s emission line is an [OIII] doublet at  $z = 1.475$  on the basis of i) the asymmetric red line-profile and flux ratio of the peaks (blue/red=0.31  $\pm$  0.03) and ii) an unphysical [OIII]/H $\alpha$  ratio ([OIII]/H $\alpha$  > 22). We show that COLA1’s observed  $B$ -band flux is explained by a faint extended foreground LAE, for which we detect Ly $\alpha$  and [OIII] at  $z = 2.142$ . We thus conclude that COLA1 is a real double-peaked LAE at  $z = 6.593$ , the first discovered at  $z > 6$ . COLA1 is UV luminous ( $M_{1500} = -21.6 \pm 0.3$ ), has a high equivalent width ( $EW_{0, Ly\alpha} = 85^{+70}_{-30}$  Å) and very compact Ly $\alpha$  emission ( $r_{50, Ly\alpha} = 0.33^{+0.07}_{-0.04}$  kpc). Relatively weak inferred H $\beta$ + [OIII] line-emission from *Spitzer*/IRAC indicates an extremely low metallicity of  $Z < 1/20 Z_{\odot}$  or reduced strength of nebular lines due to high escape of ionising photons. The small Ly $\alpha$  peak separation of  $220 \pm 20$  km s<sup>-1</sup> implies a low H I column density and an ionising photon escape fraction of  $\approx 15 - 30$  %, providing the first direct evidence that such galaxies contribute actively to the reionisation of the Universe at  $z > 6$ . Based on simple estimates, we find that COLA1 could have provided just enough photons to reionise its own  $\approx 0.3$  pMpc (2.3 cMpc) bubble, allowing the blue Ly $\alpha$  line to be observed. However, we also discuss alternative scenarios explaining the detected double peaked nature of COLA1. Our results show that future high-resolution observations of statistical samples of double peaked LAEs at  $z > 5$  are a promising probe of the occurrence of ionised regions around galaxies in the EoR.

**Key words.** Galaxies: high-redshift – Techniques: spectroscopic – Cosmology: dark ages, reionisation, first stars – Galaxies: formation – Galaxies: intergalactic medium

## 1. Introduction

The epoch of reionisation (EoR) is the last phase transition of the Universe. It occurred when the neutral hydrogen of the intergalactic medium (IGM) became reionised (e.g. Madau 2017). In spite of the increasingly precise measurements of the global progress of reionisation and its patchiness (e.g. Becker et al. 2015; Davies et al. 2017; Bañados et al. 2018; Bosman et al. 2018), its evolution and drivers are still largely unknown. One of the main probes of reionisation is the Lyman- $\alpha$  (Ly $\alpha$ ) line (see Dijkstra et al. 2007 for a review). Due to the low Ly $\alpha$  transmission in a partially neutral IGM, the progress of reionisation can be mapped out using the strength of detected Ly $\alpha$  emission. This is usually done through the equivalent width distribution among high-redshift galaxies (e.g. Pentericci et al. 2014; Schmidt et al. 2016), through the evolution in the luminosity function (e.g. Zheng et al. 2017) and/or its clustering signal (e.g. Jensen et al. 2014; Kakiichi et al. 2018).

Recently, wide-field narrow-band surveys have been highly efficient in identifying and confirming luminous Lyman- $\alpha$  emitters (LAEs) into the EoR ( $z \approx 7$ ; e.g. Matthee et al. 2015; Hu

et al. 2016; Santos et al. 2016; Bagley et al. 2017; Zheng et al. 2017; Shibuya et al. 2018). The number densities of extremely luminous LAEs at  $z \sim 7$  ( $L_{Ly\alpha} > 2 \times 10^{43}$  erg s<sup>-1</sup>) are higher (Matthee et al. 2015; Bagley et al. 2017) than expected based on older, smaller surveys (e.g. Ouchi et al. 2010). While the number densities of faint LAEs decrease at  $z > 6$ , the number densities of the most luminous sources are relatively constant between  $z \approx 5 - 7$  (e.g. Santos et al. 2016; Zheng et al. 2017; Konno et al. 2018). These luminous sources likely reside in early ionised bubbles (Stark et al. 2017), facilitating their Ly $\alpha$  observability during the EoR (e.g. Matthee et al. 2015; Weinberger et al. 2018).

Due to their luminosity, luminous LAEs are ideal for spectroscopic follow-up (e.g. Sobral et al. 2015; Hu et al. 2016; Matthee et al. 2017a; Shibuya et al. 2018). The most Ly $\alpha$ -luminous example among these sources at  $z \approx 6.6$  is CR7 (Sobral et al. 2015, 2017a). CR7 has spectacular properties: strong, narrow Ly $\alpha$  emission ( $EW_0 \approx 210$  Å), intense H $\beta$  and/or [OIII] emission (Bowler et al. 2017) and there are indications of strong high-ionisation rest-frame UV emission lines such as HeII (Sobral et al. 2017a). High resolution *HST* and ALMA observations reveal multiple components in the rest-frame UV separated by  $\sim 5$  kpc and spatially and spectrally resolved [CII] FIR cooling-

<sup>★</sup> Based on observations obtained with the Very Large Telescope, programs: 294.A-5039, 099.A-0254 & 100.A-0213

<sup>\*\*</sup> e-mail: matthee@strw.leidenuniv.nl

line emission (Matthee et al. 2017b), indicating the build-up of a central galaxy through accretion of satellite galaxies.

Recently, Hu et al. (2016) presented the spectroscopic confirmation of a potentially even more luminous LAE at  $z = 6.593$ , ‘COLA1’, with an unexpected double peaked Ly $\alpha$  line-profile. The presence of multiple peaks in COLA1 had been unprecedented in  $z > 6$  LAEs (e.g. Matthee et al. 2017a; Shibuya et al. 2018), but Songaila et al. (2018) show that roughly a third of a sample of extremely luminous LAEs at  $z = 6.6$  may have such complex line shapes. The transmission on the blue part of the Ly $\alpha$  line is typically very low due to inflows of H I gas (e.g. Dijkstra et al. 2007; Laursen et al. 2011), residual neutral hydrogen in the halos of galaxies (e.g. Mesinger et al. 2015) and the increasing neutral fraction of the inter galactic medium (e.g. Hu et al. 2010).

Several effects help to increase the observability of Ly $\alpha$  at high redshift: (i) the redshift of the Ly $\alpha$  line with respect to the surrounding gas (e.g. due to outflows), and (ii) the Hubble expansion that may redshift the Ly $\alpha$  line out of the resonance frequency of the neutral gas, thus lowering the optical depth in case the galaxy is surrounded by a large, highly ionised region.

However, due to the large-scale infall of gas which shifts the absorption  $\sim 100 \text{ km s}^{-1}$  towards the red, combined with the extended wing of the absorption profile of the highly neutral IGM, simulations and analytical models expect that even the red part of the Ly $\alpha$  emission is diminished significantly (e.g. Dijkstra et al. 2007; Laursen et al. 2011; Mesinger et al. 2015; Sadoun et al. 2017; Weinberger et al. 2018). Furthermore, radiative transfer processes in the interstellar medium (ISM) even shift the blue peak to lower velocities; e.g. Neufeld 1990. Therefore, a detection of intrinsically blue-shifted Ly $\alpha$  emission above  $z \gtrsim 6$  is theoretically unexpected, or even considered to be ruled out by theoretical models (e.g. Mason et al. 2018).

Since the transmission of the blue part of the Ly $\alpha$  line depends directly on the ionisation state of the environment of the emitting galaxy, COLA1 could open up a completely new probe of reionisation, if the detection of the blue Ly $\alpha$  line is confirmed. With a larger sample, the evolution of the Ly $\alpha$  line asymmetry (i.e., comparing the blue and red fluxes) can be quantified more robustly, adding an interesting additional observable to the evolution of LAE clustering and EW. Interestingly, the Ly $\alpha$  peak separation correlates strongly with the escape of Lyman-Continuum (LyC) radiation (e.g. Verhamme et al. 2015; Vanzella et al. 2018; Izotov et al. 2018). Therefore, COLA1 could also provide the first direct evidence that galaxies actively contribute to the reionisation of the Universe.

However, the double peaked line could also mean that COLA1 is an [OII] $_{3727,3730}$  emitter at  $z = 1.475$ . We therefore test whether the line may have been misinterpreted, in which case COLA1 is a low-redshift interloper, and not a LAE at  $z = 6.6$ . Because of the unique potential of the COLA1 galaxy, we investigate its nature independently using our own narrow-band data (from Sobral et al. 2013, not used in Matthee et al. 2015), new high spectral resolution VLT/X-SHOOTER observations covering a wavelength range from 0.3 to  $2.5 \mu\text{m}$  and the wealth of public available data from the COSMOS survey which were not explored in Hu et al. (2016).

In this work, we first summarise the available photometric information and present the new X-SHOOTER observations in §2. We analyse the spectrum in detail and conclude on the redshift of COLA1 in §3. §4 presents the properties of COLA1 that we measure from the spectroscopic and photometric data available. We discuss the implications in §5 and summarise the results in §6. Throughout this work we adopt a flat  $\Lambda$ CDM cosmology

with  $\Omega_{\Lambda} = 0.7$ ,  $\Omega_{\text{M}} = 0.3$  and  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ , a Salpeter (1955) initial mass function and magnitudes in the AB system (Oke 1974).

## 2. Data

### 2.1. Photometric measurements

COLA1 (Hu et al. 2016) is located in the COSMOS field (Capak et al. 2007; Scoville et al. 2007) and hence public imaging data are available in  $\approx 30$  filters from the UV to IR (e.g. Ilbert et al. 2009). This includes high resolution data in the F814W filter from the *Hubble Space Telescope* (HST)/Advanced Camera for Surveys (Koekemoer et al. 2007). COLA1 is located in a shallow region from the UltraVISTA survey (McCracken et al. 2012) and is covered by *Spitzer*/IRAC imaging from the SPLASH program (e.g. Steinhardt et al. 2014)<sup>1</sup>. We use optical data from Subaru/Suprime-cam (Taniguchi et al. 2007) and near-infrared data from UltraVISTA DR3. COLA1 is furthermore covered in Subaru/Suprime-cam NB921 images from Sobral et al. (2012, 2013), similar to the NB921 imaging from Hyper Suprime-Cam from Hu et al. (2016). We show thumbnail images in the  $u$ ,  $B$ ,  $z$ , NB921, F814W,  $Y$ ,  $J$ ,  $H$ , [3.6] and [4.5] bands in Fig. 1, where red contours illustrate the location of line-emission (measured from the NB921- $z$  image). We note that we have confirmed the astrometric alignment of the different images using the position of  $\approx 30$  stars and galaxies around COLA1. Our photometric measurements are summarised in Table 1.

#### 2.1.1. NB921 Narrow-band data

Combining the Suprime-cam NB921 narrow-band data (Sobral et al. 2013) with public  $z'$  band data, we measure a total line-flux of  $5.8^{+1.2}_{-1.1} \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$ . As the full  $z'$  band flux can be explained by line-emission, we use the  $Y$  band ( $Y = 25.2^{+0.4}_{-0.3}$ ,  $S/N \approx 3$ , measured in a  $2''$  diameter aperture and applying an aperture correction of  $-0.3 \text{ mag}$ ) to obtain a weak constraint on the observed equivalent width of  $\text{EW}_{\text{obs}} = 660^{+760}_{-350} \text{ \AA}$  assuming a UV slope  $\beta = -2.0$ . The  $Y$  band magnitude is in good agreement with the measurement in Hu et al. (2016) based on Subaru data. The line-flux we measure is a factor  $\approx 2$  lower than Hu et al. (2016), which is likely due to the use of a smaller aperture (they use apertures with a  $3''$  diameter) and therefore less contaminated by a neighbouring object (see Fig. 1).

#### 2.1.2. Ground-based optical data

Visual inspection of the thumbnail images reveals potential detections in the  $u$  and  $B$  filters. While the potential flux in the  $u$  band is offset, this is not as clear in the  $B$  band. We measure the significance of the flux in the  $B$  band by performing photometry centred on the NB921 detection in  $1.2''$  diameter apertures. We choose such small aperture to optimise the  $S/N$  ratio (the PSF-FWHM of this data is  $0.7 - 0.8''$ ) and minimise contamination from nearby objects. The noise level is measured from the standard deviation of 1000 empty aperture measurements located on random sky positions around COLA1. We measure flux at the  $2.4 \sigma$  level with  $B = 28.4^{+0.6}_{-0.4}$  (see Fig. 1, where we use a high contrast to highlight potential detections). No flux above the  $2\sigma$  level is detected in the  $u$ ,  $V$ ,  $R$  and  $I$  filters using the same aperture centered on COLA1. We measure flux at the  $2\sigma$  level in the stacked  $BVRI$  image. While the significance levels of these

<sup>1</sup> <http://splash.caltech.edu>

**Table 1.** The coordinates and photometric measurements of COLA1. Magnitudes are in the AB system. Magnitude limits are at the  $2\sigma$  level. \*As the  $H$  band photometry is contaminated by ID 593625 we only provide it as a lower limit.

Coordinates	
R.A.	10:02:35.37 (J2000)
Dec.	+02:12:13.9 (J2000)
Photometry	
$B$	$28.4^{+0.6}_{-0.4}$
$V$	$> 27.8$
$R$	$> 27.7$
$I$	$> 27.6$
F814W	$26.6^{+0.4}_{-0.3}$
NB921	$23.6^{+0.1}_{-0.1}$
$z'$	$25.5^{+0.5}_{-0.3}$
$Y$	$25.2^{+0.4}_{-0.3}$
$J$	$24.8^{+0.3}_{-0.3}$
$H$	$> 24.8^*$
$K_s$	$> 24.4$
[3.6]	$24.4^{+0.2}_{-0.1}$
[4.5]	$24.6^{+0.3}_{-0.2}$

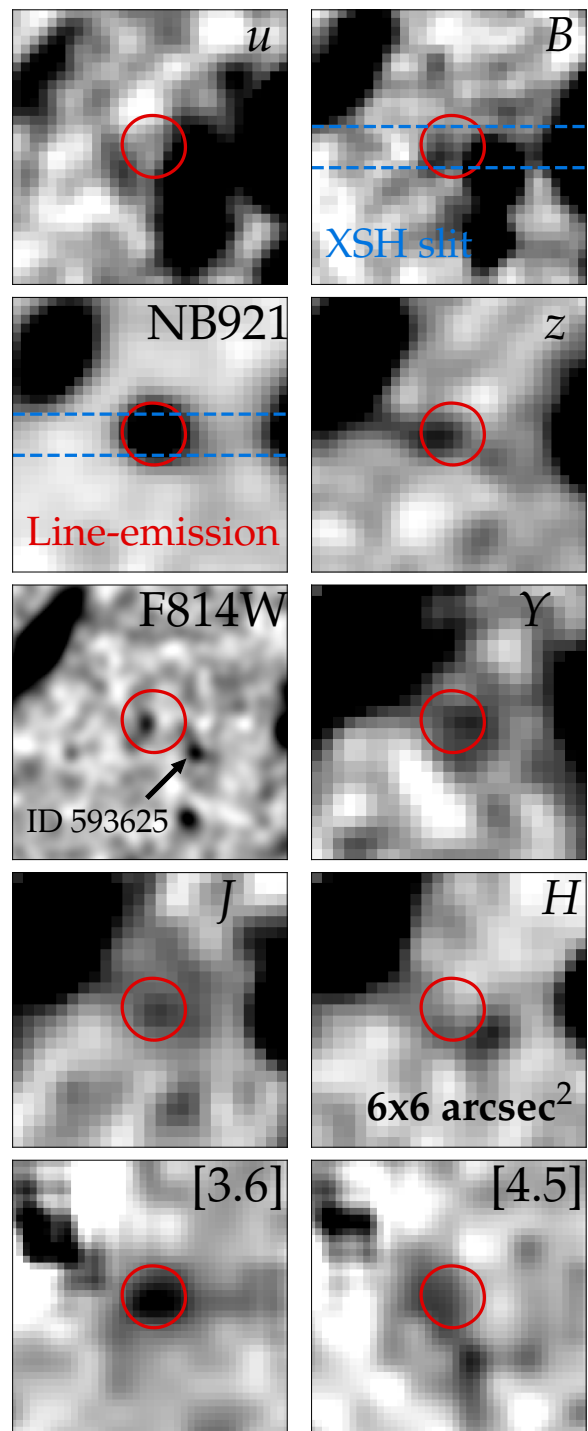
measurements is low, they may be troublesome for a galaxy at  $z > 6.5$ , motivating the need for careful spectroscopic analysis.

### 2.1.3. High resolution optical data

The *HST* imaging in the F814W filter (with  $> 20\%$  transmission between  $\lambda = 6988 - 9577 \text{ \AA}$ ,  $\lambda_{\text{eff}} = 7985 \text{ \AA}$ , PSF-FWHM =  $0.095''$ ; Koekemoer et al. 2007) reveals a faint point-like detection at the position of COLA1 (Fig. 1). Similarly to the  $u$  and  $B$  imaging, F814W imaging also clearly shows another source  $1.2''$  south-west of COLA (identified in the *HST* thumbnail in Fig. 1). This object has ID 593625 in the Laigle et al. (2016) catalogue (with photometric redshift  $p_z = 1.9^{+0.2}_{-0.1}$ ) and we will refer to it with that ID from now on. COLA1 is detected at  $3\sigma$ , with F814W =  $26.6^{+0.4}_{-0.3}$  measured in a  $0.6''$  diameter aperture and corrected for aperture losses using tabulated encircled fluxes from Bohlin (2016).

### 2.1.4. Photometric redshift and the *Spitzer*/IRAC view

COLA1 is present in the public COSMOS2015 catalogue (Laigle et al. 2016; ID 593751), where it has a photometric redshift of  $p_z = 0.99^{+0.12}_{-0.11}$ . However, the photometry that is used to estimate this redshift is measured with  $2''$  diameter apertures and may suffer contamination from ID 593625. We re-measure the photometry in the *Spitzer*/IRAC [3.6] and [4.5] filters using SPLASH data. We follow the procedure as last described in Stefanon et al. (2017), where the IRAC images are de-confused based on the *HST*/F814W images using the mophongo software (Labbé et al. 2006, 2015). We measure [3.6] =  $24.4^{+0.2}_{-0.1}$  and [4.5] =  $24.6^{+0.3}_{-0.2}$  in  $1.2''$  diameter apertures, see Fig. 1. This results in a moderately blue colour [3.6] – [4.5] =  $-0.2 \pm 0.3$ , although with significant uncertainties (see also Harikane et al. 2017, who measure [3.6] – [4.5] =  $-0.2 \pm 0.1$ ).



**Fig. 1.** High contrast thumbnail images of  $6 \times 6 \text{ arcsec}^2$  ( $32 \times 32 \text{ kpc}^2$ ) centered on the NB921 detection of COLA1. The red contour shows the  $2\sigma$  level of the NB921- $z'$  image (tracing line-emission). X-SHOOTER observations were performed with a slit alignment PA=0. A positional shift upwards along the slit corresponds to a shift to the right (west) in these thumbnails. The IRAC [3.6] and [4.5] images have been cleaned using *HST*/F814W as a prior following Stefanon et al. (2017).

### 2.2. X-SHOOTER spectroscopic observations & data reduction

We observed COLA1 with the X-SHOOTER echelle spectrograph on UT2 of the Very Large Telescope as part of ESO program 0100.A-0213 (PI Matthee). X-SHOOTER observes in

three arms simultaneously: UVB ( $\lambda = 0.30 - 0.55\mu\text{m}$ ), VIS ( $\lambda = 0.55 - 1.02\mu\text{m}$ ) and NIR ( $\lambda = 1.0 - 2.5\mu\text{m}$ ). Observations were performed on 19 January and 18-19 February 2018 with clear conditions and  $0.6 - 0.8''$  seeing. GD71 was observed as telluric standard star. We used a  $1.0''$  slit in the UVB arm and  $0.9''$  slits in the VIS and NIR arms.

We first acquired on a  $I = 13.5$  magnitude star and then applied a blind offset of  $95.86''$  to the position of COLA1 with a position angle of 0 degrees (see Fig. 1). We nodded between two positions along the slit (A and B, offset by  $3''$ ) using the AutoNodOnSlit procedure. Integration times were 700 and 780s for UVB and VIS arms respectively. Four shorter jitters with separations of  $1''$  along the direction of the slit were used for NIR exposures of 210s. This observing sequence was repeated ten times (split over two observing blocks with ABBA sequence and one with an AB sequence), resulting in total exposure times of 7.0, 7.8 and 8.4ks for the three arms respectively.

Data are reduced using standard procedures (bias and dark subtraction, flat-fielding, sky subtraction, wavelength calibration and flux calibration) incorporated in the X-SHOOTER pipeline (Modigliani et al. 2010). Individual exposures per observing block are combined using the pipeline. We then combine exposures over multiple observing blocks by computing the noise-weighted average as described in detail in Sobral et al. (2018b). Before extracting 1D spectra, we first smooth the spectrum with 2D gaussian kernels that corresponds to half the resolution ( $\sigma_x, \sigma_y$ )<sub>UVB,VIS</sub> = (0.4 Å, 0.32'') and ( $\sigma_x, \sigma_y$ )<sub>NIR</sub> = (0.6 Å, 0.21''), and bin it in the wavelength direction by 3 pixels (0.6 Å, half the resolution). The spatial extraction window in the VIS arm is  $2.2''$  which we find to optimise the S/N. Slit losses of the line at 922 nm are estimated using the NB921 image that is convolved to the PSF-FWHM of our spectroscopic observations. We measure the fraction of the total line-flux that is retrieved within the slit and extraction window. As COLA1's line-emission is compact in the NB921 data, the estimated slit losses are only 19 %. The spatial extraction windows in the UVB and NIR arms are  $1.2''$  and we do not apply slit loss corrections. Wavelengths are converted to vacuum wavelengths. We measure the effective spectral resolution using unresolved skylines on our extracted 1D spectrum. We find  $R \approx 4000$  at  $0.5\mu\text{m}$ ,  $R \approx 5000$  at  $0.9\mu\text{m}$  and  $R \approx 3800$  at  $1.6\mu\text{m}$ , corresponding to 75, 60 and 80 km s<sup>-1</sup>, respectively.

### 2.3. Spectroscopic measurements

In our X-SHOOTER spectrum, we confirm COLA1's double peaked emission line at  $\lambda_{\text{obs,vac}} = 9224, 9231 \text{ \AA}$  (S/N  $\approx 24$ ), but we also detect faint emission-lines at  $\lambda_{\text{obs,vac}} = 3821 \text{ \AA}$  (S/N  $\approx 10$ ) and  $\lambda_{\text{obs,vac}} = 15735 \text{ \AA}$  (S/N  $\approx 5$ ). The centroid of these faint emission-lines is however shifted spatially by  $1''$  to the west and are therefore not co-located with COLA1 (Fig. 1). We do not detect continuum emission. No other lines are detected in the  $0.3 - 2.5\mu\text{m}$  coverage above S/N  $> 2$ .

For the double peaked emission-line at 923nm, we measure a line-flux of  $5.90 \pm 0.24 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$  and a relative flux ratio between the blue and red component of  $0.31 \pm 0.03$ . The red line has a full width half maximum (FWHM) of  $198 \pm 14 \text{ km s}^{-1}$ , while the blue line is narrower with FWHM= $150 \pm 18 \text{ km s}^{-1}$ . While the red line is clearly asymmetric (with a weighted skewness of  $S_w = 18.0 \pm 0.9 \text{ \AA}$ ; following the definition from Kashikawa et al. 2006), the blue line is not ( $S_w = -0.2 \pm 0.3 \text{ \AA}$ ). Our measurements are summarised in Table 2.

**Table 2.** The details of the X-SHOOTER observations and spectroscopic measurements for COLA1. Flux measurements of the 923nm line are corrected for 19 % slit-losses. Upper and lower limits are at the  $2\sigma$  level. Flux limits assume a line-width FWHM=200 km s<sup>-1</sup>.

Observations	COLA1
Dates	19 Jan 2018, 18-19 Feb 2018
$R_{0.9\mu\text{m}}$	5000
$R_{1.6\mu\text{m}}$	3800
$t_{\text{exp,UVB}}$	7.0ks
$t_{\text{exp,VIS}}$	7.8ks
$t_{\text{exp,NIR}}$	8.4ks
<b>923 nm line properties</b>	
Flux	$5.90 \pm 0.24 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$
$\text{EW}_{\text{obs}}$	$660^{+760}_{-350} \text{ \AA}$
Peak separation	$220 \pm 20 \text{ km s}^{-1}$
$\text{FWHM}_{\text{red}}$	$198 \pm 14 \text{ km s}^{-1}$
$\text{FWHM}_{\text{blue}}$	$150 \pm 18 \text{ km s}^{-1}$
$\text{Flux}_{\text{blue}}/\text{Flux}_{\text{red}}$	$0.31 \pm 0.03$
$S_{w, \text{red}}$	$18.0 \pm 0.9 \text{ \AA}$
$S_{w, \text{blue}}$	$-1.5 \pm 0.4 \text{ \AA}$
$S_{w, \text{full}}$	$3.2 \pm 0.2 \text{ \AA}$
<b>Flux limits of interest</b>	
$\text{H}\alpha_{z=1.475}$	$< 5.2 \times 10^{-18} \text{ erg s}^{-1} \text{ cm}^{-2}$
$[\text{OIII}]_{4959, z=1.475}$	$< 9.8 \times 10^{-18} \text{ erg s}^{-1} \text{ cm}^{-2}$
$[\text{OIII}]_{5007, z=1.475}$	$< 9.0 \times 10^{-18} \text{ erg s}^{-1} \text{ cm}^{-2}$
$\text{CIV}_{z=6.591}$	$< 1.6 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$
$\text{HeII}_{z=6.591}$	$< 0.7 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$

The line at  $3821 \text{ \AA}$  has a line-flux of  $3.0 \pm 0.3 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$  and FWHM= 350 km s<sup>-1</sup>. The line that is observed at  $15735 \text{ \AA}$  has a line-flux of  $1.5 \pm 0.3 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$  and a width FWHM= 260 km s<sup>-1</sup>. The spatial offset of  $\approx 1''$  of these lines coincides with ID 593625.

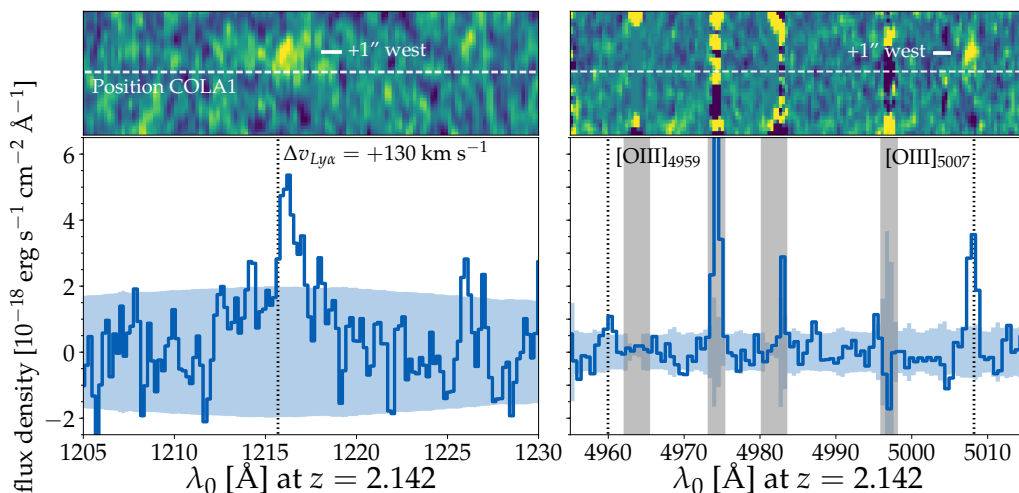
### 3. Is COLA1 a LAE at $z = 6.59$ or an [OII] emitter at $z = 1.47$ ?

The tentative detection in the *B* band (although at the  $2.4\sigma$  level) and the unlikelihood of observing Ly $\alpha$  flux blue-wards of the red asymmetric line indicate that COLA1 may not be a LAE at  $z = 6.6$ , but an [OII] emitter at  $z = 1.47$  instead. Here, we assess the implications of each observed feature to test this hypothesis.

#### 3.1. 923nm line strength and consistency with F814W

If the 923nm line from COLA1 is [OII], it would yield  $z = 1.475$  and it would have a typical luminosity ( $L_{[\text{OII}]} = 1.5 \times L^*$ ; Khostovan et al. 2015). The [OII] equivalent width would be extreme,  $\text{EW}_0 = 260^{+310}_{-140} \text{ \AA}$ , well above the typical EW for [OII] emitters ( $\text{EW}_0 = 50 \pm 20 \text{ \AA}$ ) at  $z = 1.47$  (Khostovan et al. 2016).

We test whether the F814W photometry can be explained by pure line-emission. Assuming negligible contribution from continuum emission and by spreading the line-flux homogeneously over the full transmission region of F814W results in  $\text{F814W} = 26.8 \pm 0.1$ . Hence, the F814W photometry (we measure  $\text{F814W} = 26.6^{+0.4}_{-0.3}$ ) can be perfectly explained by pure line-emission and does not indicate flux blue-wards of the emission-line. Therefore, the F814W detection does not rule out Ly $\alpha$  at  $z = 6.6$ .



**Fig. 2.** Ly $\alpha$  and [OIII] $_{5007}$  detections in the X-SHOOTER spectrum that are offset by  $\approx 1''$  to the west. This confirms that a foreground source (ID 593625 in Laigle et al. 2016) is present at  $z = 2.142$ . Ly $\alpha$  emission is redshifted by  $+130 \text{ km s}^{-1}$  compared to the systemic redshift based on [OIII] $_{5007}$ .

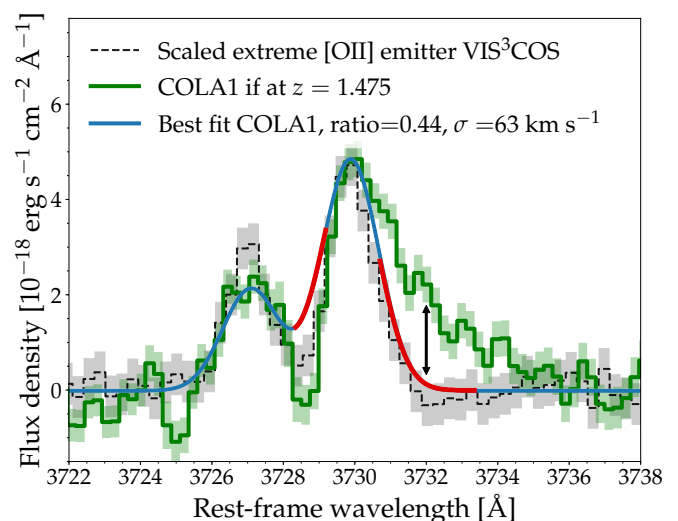
### 3.2. Line-detections at 382nm and 1573.5nm

As described above, we detect two emission-lines that are offset by  $\approx 1''$  to the west of COLA1. This corresponds roughly to the foreground source ID 593625 as can be seen in the  $B$  band thumbnail image (Fig. 1), which shows that outskirts of this galaxy fall in the X-SHOOTER slit. We identify the two lines as Ly $\alpha$  and [OIII] $_{5007}$  at  $z = 2.142$ , as illustrated in Fig. 2, where we have optimised the centroid of spatial extraction aperture for these lines. Ly $\alpha$  is redshifted by  $130 \text{ km s}^{-1}$  with respect to the systemic redshift, similar to other LAEs at  $z \sim 2$  (Trainor et al. 2015; Sobral et al. 2018b). Without correcting for slit losses, we measure a Ly $\alpha$  luminosity of  $1.0 \pm 0.1 \times 10^{42} \text{ erg s}^{-1}$ , well below the typical Ly $\alpha$  luminosity at  $z \approx 2.2$  ( $\approx 0.4 \times L^*$ ; Sobral et al. 2017b). The  $B$  band magnitude corresponding to pure line-emission at this flux-level is  $B = 28.2 \pm 0.1$ . Ly $\alpha$  emission extends to close to the peak position of COLA1 and could thus contribute significantly to the faint  $B$  band detection ( $B = 28.4^{+0.6}_{-0.4}$ ) at the COLA1 position. The [OIII] $_{5007}$  luminosity is  $0.5 \pm 0.1 \times 10^{42} \text{ erg s}^{-1}$ . We note that the centroid of the (low S/N) flux in the  $B$ - near COLA1 is shifted slightly towards a faint  $H$  band detection that could be explained by this [OIII] flux. Therefore, several detections in the images (in particular  $B$  and  $H$ ) can be attributed to LAE 593625 at  $z = 2.142$ .

### 3.3. 923nm line-profile

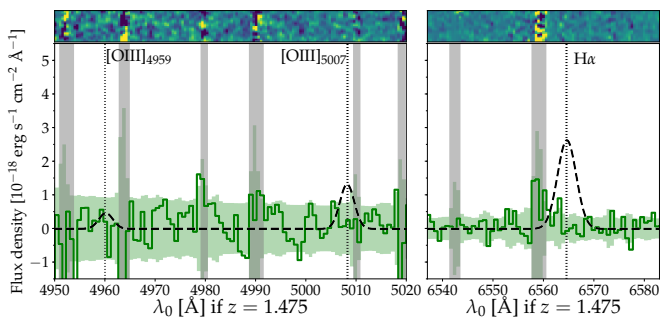
We show a detailed zoom-in figure of the line-profile of the 923nm line in Fig. 3. In this figure, we shift the spectrum to the rest-frame assuming COLA1 is at  $z = 1.475$ . We compare it to a median [OII] spectrum of galaxies with asymmetric red lines and high red/blue ratios from the VIS $^3$ COS survey (Paulino-Afonso et al. 2018a) and also show the best fitted [OII] doublet. While the peak separation ( $220 \pm 20 \text{ km s}^{-1}$ ) is fully consistent with the peak separation of the [OII] doublet, the asymmetry of the red line can not be fitted as an [OII] doublet. In particular, the blue line would have been expected to be asymmetric as well, similar to the red line. Moreover, the absence of flux between the two lines can also not easily be explained in the case of an [OII] doublet, unless lines are very narrow.

The skewness of COLA1's red line is high ( $S_{w, \text{red}} = 18.0 \pm 0.9 \text{ \AA}$ ; see Table 2), much higher than the typical maximum



**Fig. 3.** X-SHOOTER spectrum of COLA1, shifted to rest-frame wavelengths assuming it is an [OII] emitter at  $z = 1.475$  (green). Dashed black lines show the median [OII] profile of the [OII] emitters in the VIS $^3$ COS survey (Paulino-Afonso et al. 2018a) that are selected on having extreme blue/red ratios and asymmetric red lines. The blue line shows our fitted [OII] doublet to the spectrum of COLA1, and the red line indicates where the fit is  $> 3\sigma$  away from the data. The asymmetry in the red line (not present in the blue line) and the absence of flux in between the lines are at odds with COLA1 being an [OII] emitter.

skewness of a low-redshift galaxy ( $S_w = 3 \text{ \AA}$ ; Kashikawa et al. 2006), which is similar to the skewness of the full doublet. Finally, the line-ratio between [OII] $_{3726}$  and [OII] $_{3730}$  would be significantly lower than the line-ratio in our extreme [OII] emitter reference sample (with a blue-to-red fraction of  $> 0.65$ ). In fact, the line-ratio of the blue and red lines of our best-fit (0.44) is significantly lower than the theoretical minimum line-ratio for electron densities as low as  $1 \text{ cm}^{-3}$  ( $\approx 0.65$  for an electron temperature of  $10^4 \text{ K}$ ; Sanders et al. 2016, but also for temperatures between  $10^{3-5} \text{ K}$ ) and thus unphysical. Therefore, the line-profile strongly disagrees with COLA1 being an [OII] emitter at  $z = 1.475$  even though the peak separation is in perfect agreement.



**Fig. 4.** The wavelengths where [OIII] and H $\alpha$  would be observed if COLA1 is at  $z = 1.475$ . The dashed lines indicate the minimum expected [OIII] and H $\alpha$  fluxes based on known extreme line-ratios ([OIII]/[OII]=0.1 and [OII]/H $\alpha$  = 0.2, shown with FWHM=200 km s $^{-1}$ ). These line-ratios are ruled out at  $\approx 1\sigma$  and  $4.7\sigma$ , respectively.

### 3.4. No line in the NIR associated with $z = 1.475$

As we show in Fig. 4, we do not detect H $\alpha$  or [OIII] $_{5007}$  if COLA1 would be at  $z = 1.475$ . We illustrate how these lines would look in our data for the lowest empirical H $\alpha$ /[OII] = 0.2 and [OIII]/[OII] = 0.1 line-ratios (e.g. Hayashi et al. 2013). These extreme line-ratios would imply both high metallicities and high ionisation states (properties that are typically anti-correlated; Nakajima & Ouchi 2014). If we assume a FWHM of 200 km s $^{-1}$  (similar to the red line at 923nm), we measure a  $1\sigma$  limit on the H $\alpha$  flux of  $2.6 \times 10^{-18}$  erg s $^{-1}$  cm $^{-2}$  and a lower limit of  $9.4 \times 10^{-18}$  erg s $^{-1}$  cm $^{-2}$  on the combined [OIII] $_{4959,5007}$  line. While the lowest [OIII]/[OII] ratio is within the  $1\sigma$  noise level, these data rule out H $\alpha$ /[OII] = 0.2 at  $4.7\sigma$ , additional clear evidence against the interpretation that COLA1 is an [OII] line at  $z = 1.475$ .

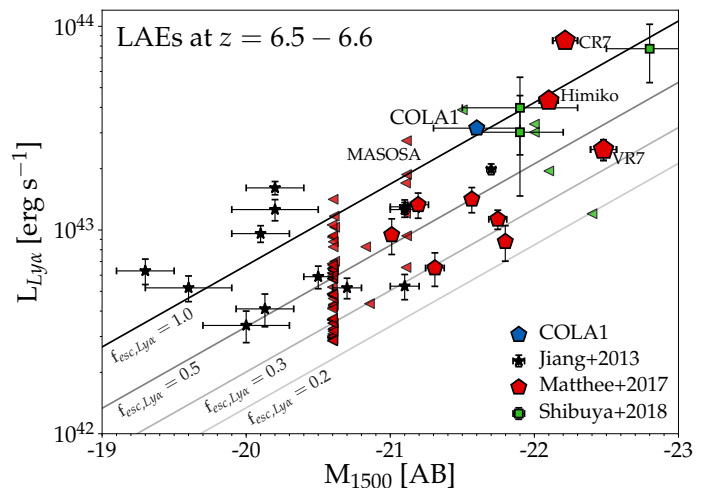
### 3.5. Concluding remarks on the redshift of COLA1

As the line-profiles of the red and blue lines differ, the double peaked emission line around 923nm can not be fitted by an [OII] doublet, even though the peak separation is similar<sup>2</sup>. Photometric indications for COLA1 being at low-redshift (in particular  $B$  and  $F814W$  detections) are explained due to line-emission from COLA1 itself ( $F814W$ ) and a foreground LAE at close separation (ID 593625 at  $z = 2.142$ ). The relatively blue *Spitzer*/IRAC colours, combined with a optical to near-infrared break of  $BVRI - Y > 3$  are also not indicative of a red dusty or old interloper at a lower redshift. Finally, if COLA1 would have been an [OII] emitter at  $z = 1.475$  the flux ratio of the lines in the [OII] doublet and the limits on [OII]/H $\alpha$  indicate unphysical conditions. Combining all the observations from above, we conclude that COLA1 is best explained as a double-peaked LAE at  $z_{Ly\alpha,red} = 6.593$ , as initially claimed by Hu et al. (2016).

## 4. Properties of COLA1 – a unique LAE at $z = 6.6$

Now we have established that COLA1 is a real LAE at  $z = 6.593$ , we can have a better look at its properties based on the X-SHOOTER spectrum and available imaging data. The properties are summarised in Table 3.

<sup>2</sup> As we show in Section 4, the Ly $\alpha$  line profile is well-fitted by a Ly $\alpha$  shell model and shows properties similar to normal LAEs at  $z < 6$ .



**Fig. 5.** UV luminosity versus Ly $\alpha$  luminosity of known LAEs at  $z = 6.5 - 6.6$  identified through narrow-band surveys (Jiang et al. 2013; Matthee et al. 2017a; Shibuya et al. 2018). Leftward pointing triangles indicate  $2\sigma$  upper limits. Black lines show lines of constant escape fraction ( $f_{esc, Ly\alpha}$ ); assuming a Salpeter IMF and a constant SFR for 100 Myr; e.g. Kennicutt 1998). COLA1 has a high Ly $\alpha$  escape fraction, similar to other LAEs known at  $z = 6.6$ .

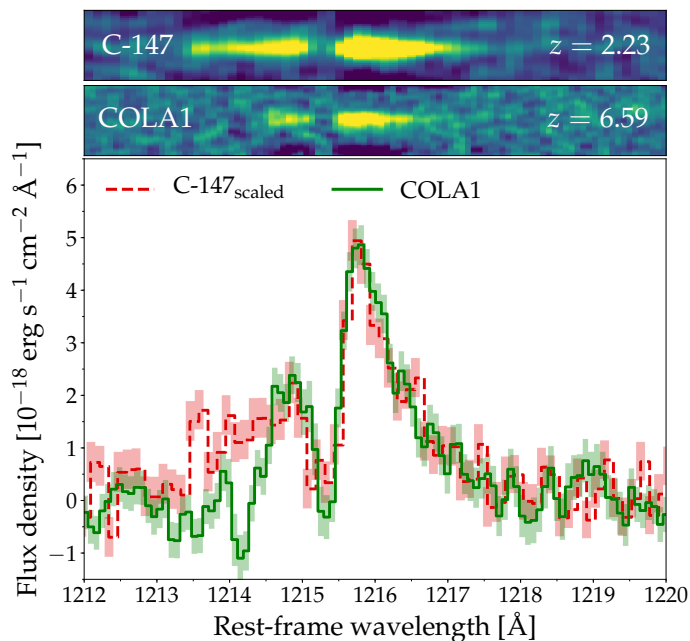
### 4.1. Ly $\alpha$ luminosity and spectral energy distribution

The slit corrected line-flux measured in the X-SHOOTER spectrum implies a Ly $\alpha$  luminosity of  $L_{Ly\alpha} = 2.9 \times 10^{43}$  erg s $^{-1}$ , which is among the most luminous LAEs known at  $z \approx 5 - 7$  (see a compilation in Matthee et al. 2017a). The Ly $\alpha$  EW is high ( $EW_0 = 85^{+70}_{-40}$  Å, based on the continuum estimated from the  $Y$  band flux), but this is a rather common property of LAEs at high-redshift (e.g. Hashimoto et al. 2017; Sobral et al. 2018a) and we note that it is poorly constrained due to uncertainties on the continuum magnitude. The Ly $\alpha$  luminosity implies a comoving number density  $\approx 1 \times 10^{-5}$  Mpc $^{-3}$  (Matthee et al. 2015).

COLA1 is weakly detected in the  $Y$  and  $J$  bands ( $S/N \approx 3$  and  $2.5$ , respectively), see Fig. 1. An off-centred  $H$  band detection is likely contaminated by strong [OIII] $_{4959,5007}$  emission from the foreground galaxy at  $z = 2.14$ , and we can only provide a  $2\sigma$  limit on the magnitude of  $H > 24.8$ . COLA1 is undetected in the  $K_s$  band ( $K_s > 24.4$ ). Due to the large uncertainties in the photometry no meaningful constraints can be obtained on the UV slope, and we assume a flat UV slope  $\beta = -2$  in the rest of the paper (this is similar to other LAEs at  $z = 6.6$ ; Ono et al. 2010; Matthee et al. 2017a).

The  $Y$  band magnitude implies an absolute magnitude  $M_{1500} = -21.6 \pm 0.3$ , slightly above  $M_{1500}^*$  at  $z \approx 7$  (Bouwens et al. 2015), but fainter than other luminous LAEs at  $z = 6.6$  (e.g. CR7 and VR7; Matthee et al. 2017a), see Fig. 5. Following Shibuya et al. (2015), such a UV luminosity indicates  $M_{star} \approx 10^{10} M_{\odot}$  at  $z \sim 7$ , although this could be significantly underestimated in case the source is strongly dust-obscured (which however is at odds with its strong Ly $\alpha$  emission; e.g. Matthee et al. 2016). The narrowness of the Ly $\alpha$  line indicates that COLA1 is likely a star-forming galaxy, as AGN typically have broader Ly $\alpha$  lines (e.g. Matsuoka et al. 2016). COLA1 is unlikely to be significantly magnified: no massive foreground structures are known and no distortions are seen on the images.

COLA1's colours in the shortest *Spitzer*/IRAC channels are blue,  $[3.6] - [4.5] = -0.2 \pm 0.3$ . This is likely a consequence of  $H\beta$  and/or [OIII] slightly boosting the  $[3.6]$  flux more than



**Fig. 6.** X-SHOOTER spectrum of COLA1, shifted to rest-frame Ly $\alpha$  at  $z = 6.593$  (green), compared to spectra of C-147 (red dashed line, a double peaked LAE at  $z = 2.23$  confirmed through H $\alpha$  emission, scaled to the peak flux in the red line; Sobral et al. 2018b). The profile of the red line in COLA1 resembles the Ly $\alpha$  profile of C-147 remarkably well. The blue component of COLA1 is cut at  $\Delta v \approx -250$  km s $^{-1}$ , and thus narrower than C-147, potentially due to a higher optical depth at large distance from COLA1.

H $\alpha$  is boosting the [4.5] flux (which is in a low transmission wavelength of the [4.5] band). However, COLA1 is not nearly as blue as other confirmed galaxies at  $z \sim 6.6$ . For example, CR7 and Himiko have colours  $[3.6] - [4.5] = -1.3 \pm 0.3$  and  $[3.6] - [4.5] = -0.7 \pm 0.4$ , respectively (Harikane et al. 2017). This could indicate that COLA1’s [OIII] line is relatively weaker due to a lower metallicity. When we compare this colour to the predicted colours at  $z = 6.6$  using photoionisation analysis and BPASS SED models presented in Bowler et al. (2017), we find that the implied metallicity lies between  $Z \approx 10^{-4} - 10^{-3}$ , but always consistent with  $Z < 10^{-3} (< 1/20 Z_{\odot})$  within the uncertainties. Alternatively, the IRAC colour could also indicate that an older stellar population is present in COLA1, or that the strength of nebular lines is lower due to a high LyC escape fraction (e.g. Zackrisson et al. 2017).

#### 4.2. Ly $\alpha$ size and star formation rate density

As discussed in §3, the detection of COLA1 in the F814W band can be explained by pure Ly $\alpha$  emission. As the HST/ACS imaging has a small PSF, we can therefore use these data to constrain the size of the Ly $\alpha$  emitting region in COLA1 accurately (e.g. Paulino-Afonso et al. 2018b). We use a non-parametric method to measure the median half-light radius ( $r_{50, \text{obs}}$ ) for 5000 random realisations of the image, where each pixel count is drawn from a Poissonian distribution. Then, we correct for PSF-broadening using  $r_{50} = \sqrt{r_{50, \text{obs}}^2 - r_{\text{PSF}}^2}$ , where  $r_{\text{PSF}} = 0.38$  kpc. This results in  $r_{50} = 0.33^{+0.07}_{-0.04}$  kpc. This size is smaller than typical UV-selected galaxies with  $M_{1500} \approx -21$ , similar to galaxies with  $M_{1500} \approx -18.4$  in the UV luminosity - size relation from Shibuya et al. (2015) and similar to the largest star cluster complexes (Bouwens et al. 2017). As the UV sizes of galaxies are typically

**Table 3.** Derived properties of COLA1 as described in §4. Limits are at the  $2\sigma$  level.

Property	Best estimate
Spectral analysis	
$L_{\text{Ly}\alpha}$	$2.9 \pm 0.1 \times 10^{43}$ erg s $^{-1}$
$\text{EW}_{0, \text{Ly}\alpha}$	$85^{+70}_{-40}$ Å
$M_{1500}$	$-21.6 \pm 0.3$
$M_{\text{star}}$	$\approx 10^{10} M_{\odot}$
$r_{50}$	$0.33^{+0.07}_{-0.04}$ kpc
$f_{\text{esc, LyC}}$	$\approx 15 - 30$ %
$\text{SFR}_{\text{Ly}\alpha}$	$75^{+60}_{-35} M_{\odot} \text{ yr}^{-1}$
$\text{SFR}_{\text{UV, nodust}}$	$27^{+8}_{-7} M_{\odot} \text{ yr}^{-1}$
$\Sigma_{\text{SFR, UV, nodust}}$	$95^{+50}_{-35} M_{\odot} \text{ yr}^{-1} \text{ kpc}^{-2}$
Gas-phase metallicity $Z$	$\lesssim 10^{-3} (1/20 Z_{\odot})$
Shell model fitting	
$v_{\text{exp}}$	$78^{+5}_{-6}$ km s $^{-1}$
$\log_{10}(N_{\text{HI}}/\text{cm}^{-2})$	$17.0^{+0.3}_{-0.3}$
$\log_{10}(T/\text{K})$	$4.2^{+0.1}_{-0.2}$
$\sigma_{\text{intrinsic}}$	$159^{+4}_{-4}$ km s $^{-1}$
$\tau_d$	$4.2^{+0.5}_{-0.8}$
$\tau_{\text{sys}}$	$6.5930^{+0.0001}_{-0.0002}$

much smaller than the Ly $\alpha$  sizes (e.g. Wisotzki et al. 2016), it could be possible that the galaxy is even more compact in the UV. On the other hand, it is likely that more diffuse Ly $\alpha$  emission is resolved out in the broad, high resolution F814W image (similar to e.g. in CR7; more luminous, extended Ly $\alpha$  emission that is undetected in similar F814W data), and that the core Ly $\alpha$  profile resembles the UV profile. Ly $\alpha$  emission is expected to be compact in galaxies with a low H $\alpha$  column density and high  $f_{\text{esc, LyC}}$  (Mas-Ribas et al. 2017) – conditions that are likely present in COLA1 as discussed next.

We can estimate the SFR of COLA1 from the UV continuum or from the Ly $\alpha$  luminosity. Assuming a UV slope  $\beta = -2.0$  and a Meurer et al. (1999) attenuation law, we find a dust-corrected  $\text{SFR}_{\text{UV}} = 40^{+12}_{-10} M_{\odot} \text{ yr}^{-1}$  ( $27^{+8}_{-7} M_{\odot} \text{ yr}^{-1}$  without correcting for dust). Alternatively, the Ly $\alpha$  luminosity indicates  $\text{SFR}_{\text{Ly}\alpha} = 75^{+60}_{-35} M_{\odot} \text{ yr}^{-1}$  (assuming a 40 % Ly $\alpha$  escape fraction and 15 % LyC escape fraction; following Sobral & Matthee 2018).

If we assume that the UV size is similar to the Ly $\alpha$  size, these measurements imply a minimum average SFR surface density  $\Sigma_{\text{SFR, UV, nodust}} = 95^{+50}_{-35} M_{\odot} \text{ yr}^{-1} \text{ kpc}^{-2}$ . This number would be higher if the UV size is more compact or if significant dust attenuation is present. This  $\Sigma_{\text{SFR}}$  is significantly higher than other galaxies known at  $z > 6$  (e.g. Carniani et al. 2017). The SFR surface density is well above the threshold required to drive galactic outflows (e.g. Heckman et al. 2001), hence such outflows may well be present.

#### 4.3. Ly $\alpha$ line properties and modelling

In Fig. 6 we show COLA1’s rest-frame spectrum assuming  $z = 6.593$  (the redshift of the red peak) and compare it to the luminous LAE C-147 (a Ly $\alpha$ -selected star-forming galaxy at  $z = 2.23$  for which a blue peak is also detected in the Ly $\alpha$  spectrum and the systemic redshift is confirmed through H $\alpha$  and [OIII] emission; Sobral et al. 2017b, 2018b). The red part of the Ly $\alpha$  line of C-147 resembles the red asymmetric line in COLA1 very well. C-147 has a broader blue line, but the peak separation and the

fact that there is no significant flux at line-centre is similar to COLA1.

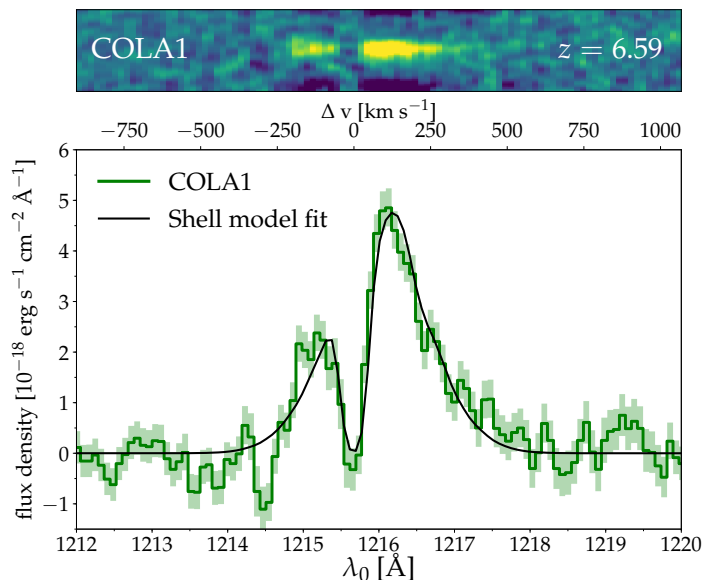
If we assume that the transmitted Ly $\alpha$  flux is zero at line-centre (in between the two peaks; e.g. Yang et al. 2017), we measure a systemic redshift  $z = 6.591$  for COLA1. The peak separation of the blue and red Ly $\alpha$  lines is  $220 \pm 20 \text{ km s}^{-1}$ . In LyC leaking galaxies in the local Universe, the peak separation is anti-correlated with the amount of leakage of LyC photons (Verhamme et al. 2015, 2017). If this correlation would hold up to  $z = 6.6$ , we would infer  $f_{\text{esc,LyC}} \approx 15 \%$  for COLA1, but potentially up to  $f_{\text{esc,LyC}} \approx 30 \%$  if we compare to the most recent results (Izotov et al. 2018). A comparison of the blue-to-red flux-ratio of local LyC leakers also indicates  $f_{\text{esc,LyC}} \approx 15 \%$  (although we note that the IGM could absorb part of the blue line-flux, see §5). For the rest of the paper we thus conservatively assume  $f_{\text{esc,LyC}} \approx 15 \%$ . Therefore, the likely non-zero LyC escape fraction of COLA1 presents the best direct evidence to date that star-forming galaxies contributed to the reionisation of the Universe at  $z \sim 7$ . In fact, as we discuss in §5, such a contribution is required to ionise a large enough region to be able to detect the blue peak at its redshift. The prospects for directly detecting LyC photons at  $z = 6.6$  are very low because of the typical opacity in the IGM is high at  $z > 6$  (Inoue et al. 2014). Moreover, while foreground contamination is a major issue (Siana et al. 2015), measuring the Ly $\alpha$  peak separation may be a promising alternative.

#### 4.3.1. Ly $\alpha$ spectral modelling

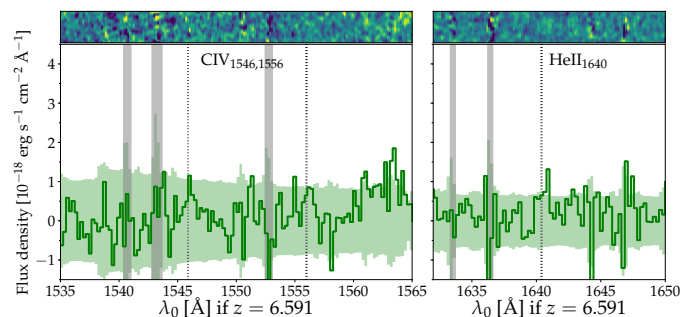
In order to get a more quantitative view of the ISM properties implied by the observed Ly $\alpha$  line, we fit the Ly $\alpha$  profile of COLA1 using a five-parameter shell model as in Dijkstra et al. (2016). This model (e.g. Ahn et al. 2001; Verhamme et al. 2006) consists of a shell of neutral gas and dust around a central ionising source and parameters include the H I column density, expansion velocity, temperature and dust optical depth of the shell and the intrinsic width of the Ly $\alpha$  line (Gronke et al. 2015). This fit does not include transmission through the IGM.

The best fitted parameters are listed in Table 3. Compared to  $z \approx 3 - 5$  LAEs (Gronke 2017), these parameters indicate a relatively low H I column density  $N_{\text{HI}} = 10^{17.0 \pm 0.3} \text{ cm}^{-2}$  and a high dust optical depth  $\tau_d = 4.2^{+0.5}_{-0.8}$ , while the other parameters are quite common for LAEs. The low H I column density is inferred from the low peak separation and suggests the possible escape of LyC photons. The high dust optical depth could alternatively be interpreted due to IGM opacity that could lower and narrow the blue peak. The H I column density and expansion velocity are also significantly lower than the inferred column density around CR7 (e.g. Dijkstra et al. 2016). We note that the shell-model prediction of the systemic velocity of CR7 presented in Dijkstra et al. (2016) agrees perfectly with recent [C II] measurements in Matthee et al. (2017b), suggesting that shell-model fitting is a good tool to recover the systemic redshift of a LAE at high redshift. The column density is also significantly lower than the column density inferred from absorption line measurements in local LyC leakers (Gazagnes et al. 2018), which lead these authors to conclude that the H I covering fraction in these galaxies is non-uniform. This is consistent with their larger Ly $\alpha$  peak separation compared to COLA1, and their typical escape fraction of  $\approx 5 \%$ .

As detailed in Verhamme et al. (2015), the Ly $\alpha$  line profile can be used as a tracer of  $f_{\text{esc,LyC}}$  as both are sensitive to H I column density as  $N_{\text{HI}} = -\ln(f_{\text{esc,LyC}})/\sigma_0$  where  $\sigma_0 = 6.3 \times 10^{-18} \text{ cm}^2$  is the ionisation cross section. According to this equation,



**Fig. 7.** The observed Ly $\alpha$  profile of COLA1 (green) is well modelled by an expanding shell of neutral gas (black). The shell-model parameters indicate a low neutral hydrogen column density  $N_{\text{HI}} = 10^{17.0 \pm 0.3} \text{ cm}^{-2}$  and low expansion velocity  $v_{\text{exp}} = 78^{+5}_{-6} \text{ km s}^{-1}$ . We note that we have assumed here that the systemic redshift lies at line-centre (i.e. between the red and blue peak, where the flux is consistent with zero), or  $z_{\text{Ly}\alpha} = 6.591$ .



**Fig. 8.** The wavelength coverage of the rest-UV CIV and HeII lines, redshifted to  $z = 6.591$ . We note that the background in a larger region around  $\lambda_{0,z=6.591} = 1560 - 1565 \text{ \AA}$  is boosted due to low atmospheric transmission and faint skylines. CIV and HeII lines are not significantly detected in COLA1, implying  $EW_{0,\text{CIV}} < 25 \text{ \AA}$  and  $EW_{0,\text{HeII}} < 12 \text{ \AA}$ .

the column density derived from the shell-model fit (see Table 3) implies  $f_{\text{esc,LyC}} \approx 50 \%$ . This is consistent with the result one would infer by using the correlation between Ly $\alpha$  EW $_0$  and  $f_{\text{esc,LyC}}$  at  $z \sim 3$  (Steidel et al. 2018), although we note this correlation has not been tested beyond  $EW_0 \gtrsim 50 \text{ \AA}$  and likely breaks down due to a reduced Ly $\alpha$  strength in the high  $f_{\text{esc,LyC}}$  regime. It also points more towards a 30 % escape fraction as implied by comparison with the most recent measurements in local LyC leakers (Izotov et al. 2018). Such high escape fractions would also affect the strength of nebular emission lines such as H $\alpha$  and H $\beta$  and could therefore be tested with future spectroscopic observations with the *James Webb Space Telescope* (e.g. Jensen et al. 2016; Zackrisson et al. 2017). Indeed, the fact that COLA1 shows a [3.6] – [4.5] colour that is much closer to zero than other luminous LAEs at the same redshift could potentially indicate a reduced strength of nebular emission lines, see §4.1.



#### 4.4. High ionisation rest-UV emission-lines

Thanks to the wavelength coverage of the X-SHOOTER spectrum, we can inspect the spectrum for the presence of strong rest-UV emission line features (e.g. Stark et al. 2015); see Fig. 8. We do not detect a significant feature besides Ly $\alpha$ . We measure  $2\sigma$  limiting line-fluxes of 1.6 and  $0.7 \times 10^{17}$  erg s $^{-1}$  cm $^{-2}$  for the C IV $_{1546,1556}$  and He II $_{1640}$  emission lines at  $z = 6.591$ , respectively (over 250 km s $^{-1}$  extraction boxes), corresponding to  $EW_{0,CIV} < 25$  Å and  $EW_{0,HeII} < 12$  Å. These non-detections are not surprising given the fact that the limits are shallower than other unsuccessful spectroscopic follow-up of LAEs at  $z = 6.6$  (e.g. Shibuya et al. 2018), and EW limits are higher than detections in e.g. Stark et al. (2017).

## 5. Discussion: witnessing a galaxy reionising its surroundings

### 5.1. The unlikelihood of observing a blue Ly $\alpha$ line at $z > 6$

Ly $\alpha$  photons resonantly scatter in the presence of neutral hydrogen. Once Ly $\alpha$  photons are absorbed, they are re-emitted in a random direction in the rest-frame of the absorbing hydrogen atom, resulting in a diffusion process in real and frequency space. Analytical models show that this process results in a double peaked Ly $\alpha$  spectrum in a static medium (e.g. Neufeld 1990). In the presence of outflows, Ly $\alpha$  photons see a larger optical depth towards the blue, resulting in a redshifted asymmetric spectrum. Hence, the Ly $\alpha$  profile is sensitive to the neutral hydrogen content and the velocity field of the gas in a galaxy (e.g. Loeb & Rybicki 1999; Santos 2004; Dijkstra et al. 2007).

After escaping from the ISM, Ly $\alpha$  photons are also affected by H I in the circum-galactic medium (CGM). The CGM predominantly transmits red Ly $\alpha$  photons, enhancing the asymmetry between the red and the blue peak (e.g. Laursen et al. 2011). This is consistent with observations at high-redshift ( $z > 2$ ), where high H I column densities in the CGM result in a majority of Ly $\alpha$  profiles that consist of a single, red asymmetric line (e.g. Erb et al. 2014). Due to the increasing neutral fraction of the IGM, the asymmetry between the transmission of the red and blue line increases and the chances of observing double peaked emission decrease.

This is also found in semi-analytical and hydrodynamical models of the EoR, which typically predict negligible transmission bluewards of the systemic velocity at  $z > 6$  (Dijkstra et al. 2007; Laursen et al. 2011; Weinberger et al. 2018). In the radiative transfer simulations from Laursen et al. (2011) the median IGM transmission at the blue peak is  $> 30\%$  only at  $z < 4.5$ , while it is  $< 0.1\%$  at  $z > 5$  (similar to more recent simulations by Weinberger et al. 2018). Given that these models predict such low transmission bluewards of the systemic redshift, how is it possible that we observe a strong blue peak in the Ly $\alpha$  profile of COLA1 at  $z = 6.59$ , a redshift where this is highly unexpected? Here, we propose three scenarios that facilitate the transmission of blue Ly $\alpha$  photons and that may explain COLA1's Ly $\alpha$  profile.

### 5.2. Three scenarios to explain double peaked Ly $\alpha$ emission at $z = 6.6$

In Fig. 9, we sketch three different physical scenarios that may explain the Ly $\alpha$  profile of COLA1. We first describe these models and then we propose observations that may test and differentiate between them. We note that the transmission, line-widths and offsets in Fig. 9 are chosen for illustrative purposes.

In Model 1, COLA1 is surrounded by a highly ionised bubble that is large enough for Ly $\alpha$  photons to redshift out of resonance due to the expansion of the Universe before encountering a relatively neutral IGM. The Ly $\alpha$  profile is therefore only determined by the ISM conditions of the galaxy. In Model 2 the bubble itself is smaller, but is embedded in a larger relatively ionised region. The Ly $\alpha$  profile here is mostly affected by the gas in the CGM and IGM. Model 3 attempts to explain COLA1's Ly $\alpha$  profile without invoking IGM attenuation, but rather by an infalling self-shielded cloud of neutral hydrogen. This neutral hydrogen cloud absorbs part of the emitted Ly $\alpha$  line, resulting in a double peaked spectrum.

#### 5.2.1. Model 1: Large, highly ionised bubble in a neutral IGM

If COLA1 resides in a large, highly ionised region, it is possible that blue Ly $\alpha$  photons redshift out of the resonance wavelength due to the Hubble expansion prior to encountering a neutral IGM (e.g. Malhotra & Rhoads 2006). Depending on the size of this highly ionised region, it is possible that only (part of) the blue peak is attenuated. In the left panel of Fig. 9, we illustrate this scenario where a double peaked Ly $\alpha$  line escapes from the ISM of the galaxy (resulting from resonant scattering effects, e.g. Neufeld 1990; Gronke et al. 2017), before the blue part of the line is attenuated by a sigmoid transmission function of the IGM.

The Ly $\alpha$  photons in the blue peak of COLA1 need to redshift out of the resonance wavelength by at least  $\geq 250$  km s $^{-1}$  (the maximum velocity at which we observe blue Ly $\alpha$  flux compared to line-centre, see Fig. 7) before encountering significant amounts of neutral hydrogen. This requires a large ionised region. Ignoring peculiar velocities of inflowing gas, we can calculate the required size as follows:

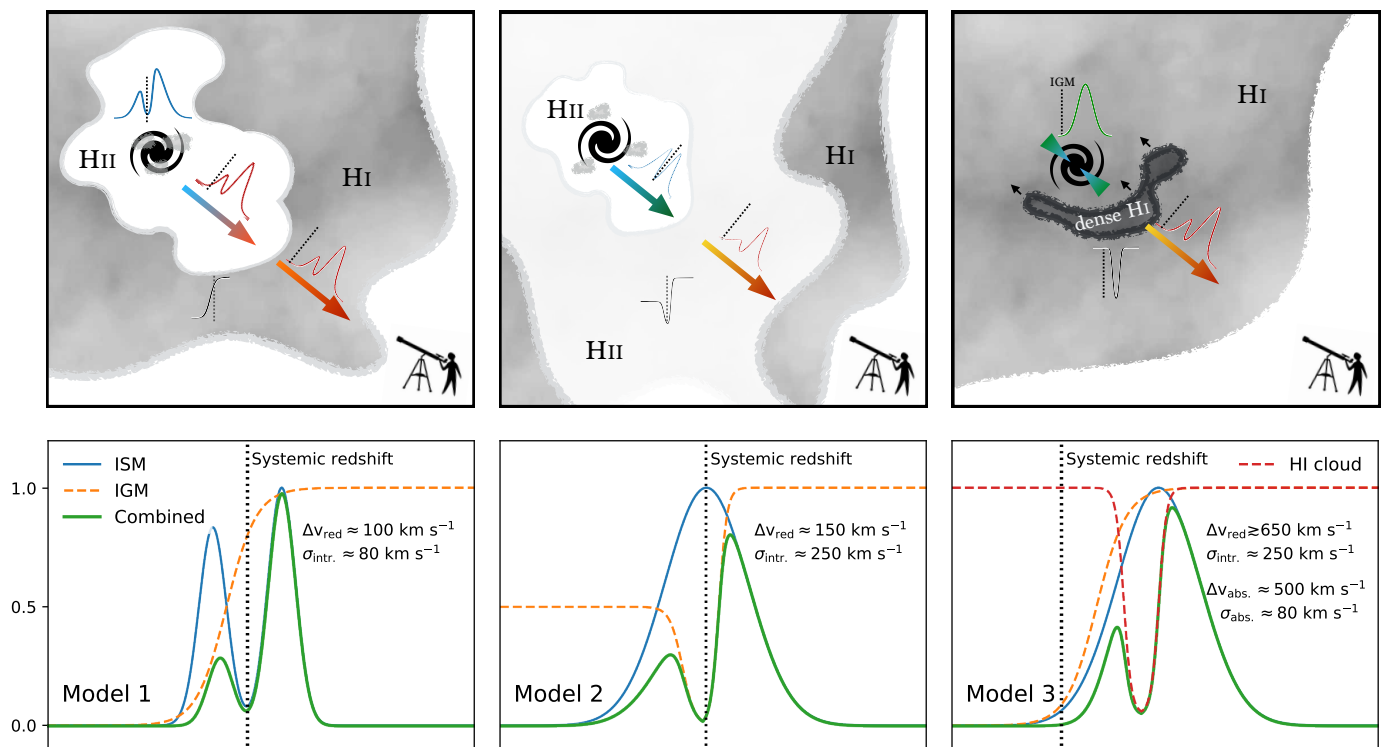
$$d_{\text{prop}} = \frac{\Delta v}{H_{z=6.59}} \text{Mpc}. \quad (1)$$

Here,  $d_{\text{prop}}$  is the proper distance photons are required to travel,  $\Delta v$  the required velocity offset and  $H_{z=6.59}$  is the Hubble parameter at  $z = 6.59$ , which is  $H_{z=6.59} = 803.85$  km s $^{-1}$  Mpc $^{-1}$  in our assumed cosmological model. Therefore, in order for Ly $\alpha$  photons to redshift by  $\geq 250$  km s $^{-1}$  would require an ionised sightline/region of at least 0.3 pMpc; or  $\geq 2.3$  cMpc. Simulations indicate that this bubble-size is similar to the characteristic bubble size for a global ionised fraction of  $x_{\text{HI}} \approx 50\%$  (e.g. Furlanetto et al. 2006; Lin et al. 2016). As the ionised region needs to be at least 0.3 pMpc, and could be larger, this means the bubble size of COLA1 would correspond to a mean IGM  $x_{\text{HI}} < 0.5$  at  $z = 6.6$ . On the other hand, the bubble around COLA1 may also be an outlier as COLA1 is likely a relatively massive galaxy. This could imply a higher neutral fraction.

Can such an ionised region be explained using the observed properties of COLA1? Following Haïman (2002), we estimate the maximum proper radius of the ionised region around a galaxy (in the absence of neighbouring ionising sources and ignoring recombinations):

$$R_s = 2.5(f_{\text{esc,LyC}} Q_{\text{ion}}/10^{54} \text{s}^{-1})^{1/3} (t_{\text{burst}}/10^7 \text{yr})^{1/3} (1+z)^{-1} \text{Mpc}. \quad (2)$$

Here,  $R_s$  is the radius of the Strömgren sphere,  $f_{\text{esc,LyC}}$  the escape fraction of ionising photons,  $Q_{\text{ion}}$  the produced number of ionising photons per second and  $t$  the age of the burst of star formation. As listed in Table 4, we estimate  $f_{\text{esc,LyC}} = 0.15$  (based



**Fig. 9.** Sketches of different scenarios that may explain the observed Ly $\alpha$  line-profile of COLA1. The top row shows sketches of the physical scenarios and the bottom row shows the emerging Ly $\alpha$  spectrum. On the left, Model 1 shows a double peaked emission line emerging from scattering in the ISM that is redshifted due to the expansion of the Universe before encountering the IGM, that preferentially attenuates the blue component. In the middle, Model 2 shows that a double peak may originate due to a low transmission at line centre (due to neutral hydrogen in the CGM), while there is a relatively high transmission at further distances from the galaxy due to a large relatively ionised region (similar to the IGM transmission curve at  $z \approx 4$  in Laursen et al. 2011). On the right, Model 3 shows that a double peaked profile can also arise without IGM attenuation in case there is an HI absorber slightly blue shifted with respect to the systemic Ly $\alpha$  velocity (which in this model is significantly redshifted with respect to the IGM).

on the Ly $\alpha$  peak offset),  $Q_{\text{ion}} \approx 5 - 7 \times 10^{54} \text{ s}^{-1}$  (based on either the Ly $\alpha$  luminosity, or based on the UV luminosity, combined with a ionising photon production efficiency of  $\xi_{\text{ion}} = 10^{25.4} \text{ Hz erg}^{-1}$ ; Bouwens et al. 2016) and  $t_{\text{burst}} = 10^7 \text{ yr}$  (based on the high Ly $\alpha$  EW; e.g. Charlot & Fall 1993) at  $z = 6.591$ . This results in  $R_{s,\text{max,Ly}\alpha} = 0.33^{+0.07}_{-0.06} \text{ pMpc}^3$  and  $R_{s,\text{max,UV}} = 0.29^{+0.03}_{-0.03} \text{ pMpc}$ , corresponding to  $\approx 2.5 \text{ cMpc}$ . The maximum radius would marginally increase by 0.03 pMpc when correcting the UV luminosity for a (high) attenuation  $A_{\text{UV}} = 0.45$ .

Therefore, under basic assumptions, the star-formation in COLA1 may provide enough photons that can ionise a large enough region for allowing the blue peak to be observed up to  $\approx -250 \text{ km s}^{-1}$  from line-centre. However, there are important caveats that require attention. For example, our estimate conservatively assumes that the IGM does not affect Ly $\alpha$  photons red-wards of line-centre (as illustrated in the left panel of Fig. 9) and the real required ionised region may have to be significantly larger.

Moreover, the calculation so far also ignores peculiar velocities that typically blueshift Ly $\alpha$  photons with respect to neutral gas in the IGM (Laursen et al. 2011). Our calculation also ignores self-shielded neutral regions in the CGM around galaxies, that could be challenging to ionise by the galaxy itself and therefore may form a major source of opacity (e.g. Mesinger et al. 2015; Sadoun et al. 2017). The recent simulations from

Weinberger et al. (2018) suggest that these self-shielded regions may be more common for galaxies that reside in halos with  $M_{\text{halo}} \sim 10^{11} M_{\odot}$ , for which the fraction of sight-lines that has a significant transmission on the blue side of the systemic redshift is consequently extremely low. Sight-lines with high blue-transmission do exist in a significant fraction of low mass halos ( $M_{\text{halo}} \sim 10^9 M_{\odot}$ ). These simulations therefore suggest that COLA1 resides in a low mass halo, unless the ionisation from galaxies themselves have been under-estimated.

Finally, as noted in Haiman (2002), the radius of the ionised sphere is over-estimated in case recombinations are important, for example when the clumping factor in the IGM is high ( $C > 10$ ) or the age of the star formation burst is  $> 10^8 \text{ yr}$ . It is thus unclear whether the ionised bubble can also be sustained, in particular when the star formation rate of COLA1 would decline or if the escape fraction would decrease. A solution would be if a quasar or (faint) neighbouring galaxies contribute to the local ionising budget (e.g. Kakiichi et al. 2018). COLA1 is not located nearby a known quasar (Bañados et al. 2016). However, COLA1 is at a relatively close separation to CR7, the most luminous LAE known at  $z = 6.6$ . The projected distance on the sky of  $34.17'$ , which corresponds to a comoving distance of 77 Mpc (proper distance 10.1 Mpc) and the velocity difference is  $\approx 350 \text{ km s}^{-1}$ . No (faint) neighbouring galaxies are known around COLA1 (Bowler et al. 2014; Matthee et al. 2015), although the sensitivity to neighbouring galaxies is limited (corresponding to  $\text{SFR} \gtrsim 25 M_{\odot} \text{ yr}^{-1}$ ) and deeper observations are required.

<sup>3</sup> We note that these uncertainties are the propagated errors corresponding to the UV and Ly $\alpha$  luminosities and ignore uncertainties in the age of the burst of star formation and the escape fraction.

**Table 4.** Parameters used in calculations of the ionised bubble around COLA1 at  $z = 6.591$  (§5.2.1).

Property	Value	Motivation
$f_{\text{esc,LyC}}$	15 %	(Conservative) Ly $\alpha$ peak separation and blue-to-red flux ratio in local LyC leakers
$f_{\text{esc,Ly}\alpha}$	$40^{+30}_{-20}$ %	Ly $\alpha$ EW calibration, tested at $z = 0 - 2.6$ (Sobral & Matthee 2018)
$Q_{\text{ion,Ly}\alpha}$	$7^{+5}_{-3} \times 10^{54} \text{ s}^{-1}$	Ly $\alpha$ luminosity, $f_{\text{esc,LyC}}$ and $f_{\text{esc,Ly}\alpha}$ (Sobral & Matthee 2018)
$Q_{\text{ion,UV}}$	$5^{+1}_{-1} \times 10^{54} \text{ s}^{-1}$	UV luminosity (no dust correction) and $\xi_{\text{ion}} = 10^{25.4} \text{ Hz erg}^{-1}$
$t_{\text{burst}}$	$10^7 \text{ yr}$	High Ly $\alpha$ EW (Charlot & Fall 1993)
$R_{s,\text{required}}$	$> 0.3 \text{ pMpc}$	Velocity offset $\gtrsim 250 \text{ km s}^{-1}$ due to expansion of the Universe
$R_{s,\text{max,Ly}\alpha}$	$0.33^{+0.07}_{-0.06} \text{ pMpc}$	Eq. 2, assuming no recombinations and ionisation by the Ly $\alpha$ source of COLA1
$R_{s,\text{max,UV}}$	$0.29^{+0.03}_{-0.03} \text{ pMpc}$	Eq. 2, assuming no recombinations and ionisation by the UV source of COLA1

### 5.2.2. Model 2: Large mildly ionised region

In our second scenario, COLA1 resides in a relatively large ionised region with a high transmission away from line-centre, while the galaxy itself is not able to ionise gas clouds in its direct vicinity, causing the low transmission at line-centre. This scenario is illustrated in the centre panel of Fig. 9. Here, the double peaked emission does not emerge from the galaxy itself, but is caused by the environment of COLA1 that transmits flux at different velocities differentially.

Here, the gas in the CGM of the galaxy still contains significant amounts of neutral hydrogen (e.g. due to self-shielding), while the gas at larger distances from the galaxy is mildly ionised and has a transmission of  $\approx 50 \%$  at  $< -1000 \text{ km s}^{-1}$  from the systemic redshift (similar to the simulated IGM properties at  $z \approx 4$  in Laursen et al. 2011). The double peaked emission originates from residual hydrogen in the vicinity of COLA1, absorbing flux around the systemic velocity. In this scenario, the intrinsic Ly $\alpha$  profile (that escapes the ISM) is broader than the observed red Ly $\alpha$  line, but the resulting blue and red lines have similar widths. While this scenario still requires a highly ionised region (a transmission of 50 % corresponds to  $\tau = 0.7$ , or a mean neutral fraction of  $\sim 2 \times 10^{-6}$  at  $z = 6.6$ ), less ionising photons originating from COLA1 are required compared to Model 1.

We point out that double peaked Ly $\alpha$  profiles with low flux at line centre are the norm among Green Pea galaxies in the local Universe ( $z \approx 0.3$ ; e.g. Henry et al. 2015; Yang et al. 2017), emerging from scattering in the ISM. Therefore, for an emerging spectrum that has a similar profile at  $z = 6.6$  (see also Fig. 6), but due to the effect of the IGM (and without scattering in the ISM) would be a remarkable coincidence.

### 5.2.3. Model 3: Large velocity offset and intervening absorber

The last scenario follows an alternative explanation which does not require a strong imprint from the IGM (either highly ionised, neutral, or partly ionised) in the vicinity of the emitting galaxy. Here, the emergent Ly $\alpha$  spectrum consists of a significantly redshifted broad peak (e.g. due to an outflow), such that line is not affected by H $\alpha$  in the CGM or IGM. The right panel of Fig. 9 shows that the double peaked profile originates from an absorbing system on top of this emission profile. This scenario is similar to the ‘cold flow’ scenario that is used to explain the Ly $\alpha$  profiles of high-redshift radio galaxies (e.g. van Ojik et al. 1997), but with the difference that the column density is much lower.

The absorbing system could be interpreted as an inflow (e.g. Matsuda et al. 2006), or potentially a wall between two ionised regions (although we note that the IGM in model 3 can also be fully neutral), such as a filament that has not yet reionised (e.g. Finlator et al. 2009). We explore the H $\alpha$  column density that such

an absorber would imply by fitting the line-profile with a voigt profile absorber on top of a gaussian emission line. The upper H $\alpha$  column density limit is mostly determined by the width of the absorber, which implies a marginally self-shielded column density  $N_{\text{HI}} \lesssim 10^{17.7} \text{ cm}^{-2}$ , blue shifted by  $\approx 80 \text{ km s}^{-1}$  with respect to the peak of the emission. The intrinsic line-width FWHM is  $380 \pm 20 \text{ km s}^{-1}$ .

How likely is such a scenario? In this model, both observed lines are redshifted with respect to the systemic redshift. This is not seen in  $z = 0 - 3$  galaxies with double peaked Ly $\alpha$  emission, where the systemic redshift typically lies between peaks (e.g. Kulas et al. 2012; Henry et al. 2015; Sobral et al. 2018b). Vanzella et al. (2018) identify a luminous system at  $z = 4$  ( $M_{1500} = -22.2$ ,  $L_{\text{Ly}\alpha} \approx 1.4 \times 10^{43} \text{ erg s}^{-1}$ ) that has four Ly $\alpha$  components, of which only one is redshifted by  $\approx 350 \text{ km s}^{-1}$ , while the others are either systemic or blueshifted. Multiple redshifted lines are seen in high-redshift radio galaxies (van Ojik et al. 1997), but these lines are typically much broader (FWHM  $\gtrsim 1000 \text{ km s}^{-1}$ ) and likely a consequence of large amounts of hydrogen present around the massive galaxies that harbour radio AGN. Therefore, unless particular conditions in the EoR are important, this scenario may be quite unlikely.

### 5.2.4. Summary and predictions

The sketched scenarios in Fig. 9 may be over-simplified and reality may be a combination of different aspects of the models. Even though Models 2 and 3 may be unlikely, some aspects of these models may be combined with the general scenario in Model 1. For example, the absorber from Model 3 could reside in the highly ionised region from Model 1. Other complications would arise if the galaxy is undergoing a major merger, with multiple emission line regions at different relative velocities (although there are currently no indications for this scenario). How can observations discriminate between the different explanations and make progress?

As illustrated in Fig. 9, it is clear that the key quantities that would allow to differentiate between the models are the intrinsic line-width and the systemic redshift (marked as dashed, vertical lines in Fig. 9). These properties can be measured from the H $\alpha$  or the H $\beta$  lines, accessible with the *James Webb Space Telescope*. The systemic redshift can alternatively also be measured using [C $\text{II}$ ] FIR emission with ALMA (e.g. Matthee et al. 2017b), which can also test whether a merger of multiple components is ongoing. Furthermore, if COLA1 is surrounded by a large ionised region, additional sources of ionising photons are likely present. It is therefore expected that COLA1 resides in a larger scale over-density of faint galaxies, potentially observable using deep Ly $\alpha$  observations with VLT/MUSE and/or with ALMA observations that can identify [C $\text{II}$ ].

### 5.3. Implications & considerations for surveys

Why are there no other LAEs known at  $z = 5-7$  that have double peaked Ly $\alpha$  emission (c.f. [Songaila et al. 2018](#))? Why do we find a double peaked LAE at a redshift where the IGM transmission towards the blue is likely very low, instead of around  $z \approx 5-6$ ? Could it be that more double-peaked LAEs have been found, but they have been discarded as [OII] interlopers?

#### 5.3.1. Observational biases against identifying double peaked LAEs?

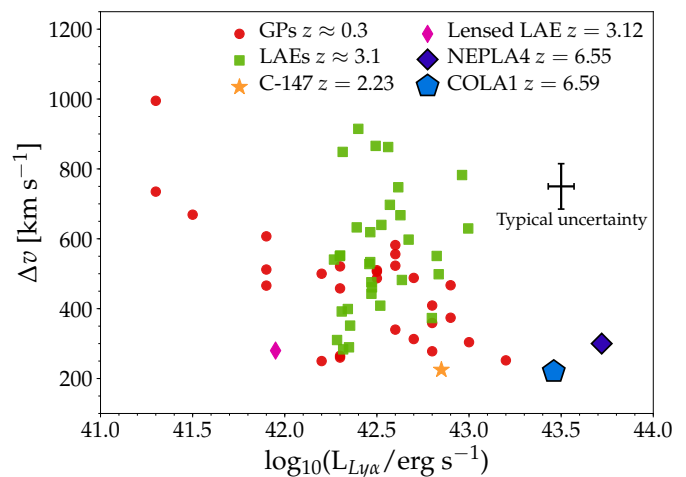
As already noted in [Dijkstra et al. \(2007\)](#), LAEs with a strong blue peak show a relatively low asymmetry (quantified by skewness) and could thus be classed as low-redshift interlopers in the case of low resolution spectroscopy. Moreover, the range of observed peak separations in double peaked LAEs ( $\Delta v \approx 200-400$  km s $^{-1}$ ; [Verhamme et al. 2017](#)) corresponds to observed wavelength differences  $\Delta\lambda_{\text{obs}} = (1 + z_{\text{Ly}\alpha}) \times [0.8 - 1.6]$  Å, similar to the observed wavelength difference of the [OII] doublet  $\Delta\lambda_{\text{obs,[OII]}} = (1 + z_{\text{[OII]}}) \times 2.79$  Å at all observed wavelengths in the optical. Therefore, ‘single’ line-identifications may be biased against confirming double peaked LAEs. This bias may particularly be important when spectral wavelength coverage is limited to the optical, but less important for  $z_{\text{Ly}\alpha} \lesssim 4.8$ , as potential interlopers of LAEs at these wavelengths will likely be detected in multiple lines (such as H $\beta$  and [OIII], besides [OII]).

There are also other explanations of why the number of known double-peaked LAEs at  $z = 5-6$  may be low. One explanation is that the fraction of double-peaked LAEs increases strongly with Ly $\alpha$  luminosity (as discussed in [Hu et al. 2016](#)) and current samples at  $z = 5-6$  do not yet include enough luminous sources (as the majority of very wide-field surveys prioritised  $z > 6$ ; e.g. [Songaila et al. 2018](#)). As the typical Ly $\alpha$  luminosity increases with redshift, this explanation would indirectly imply that the LyC escape fraction increases with Ly $\alpha$  luminosity and with redshift (as  $L_{\text{Ly}\alpha}^*$  increases; [Sobral et al. 2018a](#)).

Alternatively, it may be possible that spectroscopic follow-up observations and/or publication efforts may be biased towards  $z > 6$  and double-peaked LAEs at  $z = 3-6$  simply have not yet been observed spectroscopically or published; or that the typical S/N of spectroscopic observations at  $z > 4$  is too low to identify similar blue peaks. The flux in the blue peak is a factor  $\approx 0.3$  lower than the brighter red component and can therefore only be detected at  $3\sigma$  significance if the red peak is detected at  $\approx 10\sigma$  significance, which is not always the case.

#### 5.3.2. A preference for observing narrow peak separations only in luminous LAEs at $z > 5$ ?

Multiple peaked Ly $\alpha$  profiles have regularly been observed in LAEs at  $z = 2-3$  (e.g. [Kulas et al. 2012](#); [Yamada et al. 2012](#); [Trainor et al. 2015](#); [Vanzella et al. 2016](#); [Rivera-Thorsen et al. 2017](#)), but also in quasars and in high-redshift radio galaxies (e.g. [van Ojik et al. 1997](#); [Miley & De Breuck 2008](#)). As COLA1 does not show evidence for AGN activity, we focus our comparison to LAEs at  $z = 2-3$  (combining data from [Yamada et al. 2012](#), [Saez et al. 2015](#) and [Sobral et al. 2018b](#)) and Ly $\alpha$  analogues (i.e. Green Pea galaxies; GPs; [Yang et al. 2017](#)). The fraction of double peaked LAEs at  $z = 2-3$  is  $\approx 30-50\%$  ([Kulas et al. 2012](#); [Yamada et al. 2012](#); [Trainor et al. 2015](#)), who find typical peak separations of  $\sim 500-750$  km s $^{-1}$ , although we note that the spectral resolution of these observations ( $\approx 150-200$  km



**Fig. 10.** Ly $\alpha$  peak separation versus the observed Ly $\alpha$  luminosity of a compilation of star-forming LAEs. COLA1 (blue pentagon) has the lowest peak separation known, while NEPLA4 (dark blue diamond; [Songaila et al. 2018](#)) is the most luminous LAE known with a double peak. While there is no clear relation between the peak separation and Ly $\alpha$  luminosity at  $z \approx 2-3$  (orange star: luminous LAE C-147 at  $z = 2.23$ , that is shown in Fig. 6; green squares: points based on published results from [Yamada et al. 2012](#) and [Saez et al. 2015](#); magenta diamond: a low mass  $M_{\text{star}} < 10^7 M_{\odot}$  compact LAE presented in [Vanzella et al. 2016](#)), a clear anti-correlation is seen between peak separation and luminosity for Ly $\alpha$  emitting Green Pea galaxies (GPs; red points, [Yang et al. 2017](#)). This anti-correlation may be a selection effect, as GPs are selected on compactness and strong [OIII]/[OII] ratios that may trace density bounded HII regions, and not a pure Ly $\alpha$  selection.

s $^{-1}$ ) may not identify peaks with low velocity separations. The peak separation in COLA1 is smaller than the peak separation of the ultra-faint ( $M_{1500} = -17$ ), lensed LAE at  $z = 3.12$  identified in [Vanzella et al. \(2016\)](#), which is an extremely compact, highly ionising star-forming galaxy.

As shown in Fig. 10, there is no clear relation between peak separation and Ly $\alpha$  luminosity for LAEs at  $z \approx 3$  (although the dynamic range is limited). On the other hand, a clear anti-correlation is seen between peak separation and luminosity for Ly $\alpha$  emitting GPs. COLA1, which is the most luminous LAE with the lowest known peak separation, seems to follow this relation. However, we note that this anti-correlation may be a selection effect, as GPs are selected on compactness and strong [OIII]/[OII] ratios ([Cardamone et al. 2009](#)), that may trace density bounded HII regions. For example, GPs may all have a similar intrinsic Ly $\alpha$  luminosity, but because of a correlation between the Ly $\alpha$  escape fraction and the peak separation ([Yang et al. 2017](#)) a strong anti-correlation is observed (Fig. 10).

If Ly $\alpha$  luminosity is anti-correlated with the peak separation at  $z > 6$ , this may explain why double peaked Ly $\alpha$  profiles may only be observed in the most luminous LAEs. As the IGM transmission strongly decreases towards the blue of the systemic redshift (e.g. [Laursen et al. 2011](#); [Weinberger et al. 2018](#)), the blue lines of faint LAEs are more likely to be reduced. Luminous LAEs may however reside in larger ionised bubbles (because of a higher amount of ionising photons), facilitating the observability of a blue peak. This would for example imply that NEPLA4 ([Songaila et al. 2018](#)) resides in a larger ionised bubble than COLA1. On the other hand, simulations argue that the blue peak transmission may be higher for fainter galaxies, as there is less self-shielded HI present in their CGM (e.g. [Weinberger et al.](#)

2018). The low *observed* peak separation in COLA1 could also indicate that the CGM/IGM attenuated a significant amount of the flux at even bluer wavelengths. Hence, it may simply be expected that observed peak separations at  $z > 6$  are lower than at  $z < 5$ . A crucial test would therefore be to map out the evolution of the relation between peak separation and Ly $\alpha$  luminosity from  $z \approx 5 - 7$  with high resolution spectroscopy, and explore how the blue-to-red flux ratio evolves with redshift.

## 6. Summary

We have presented a detailed analysis of the COLA1 galaxy, a double peaked LAE at  $z = 6.59$  (Hu et al. 2016), using deep multi-wavelength photometry and new deep, high resolution optical and near infrared spectroscopy with VLT/X-SHOOTER. Due to the opacity of the CGM and IGM for blue Ly $\alpha$  photons, detecting a blue peak at  $z > 6$  is highly unlikely. We therefore test whether the line profile is interpreted correctly as Ly $\alpha$  at  $z = 6.6$  and study the properties of COLA1 and their implications. Our results are summarised as follows:

- We rule out that the double peaked line is [OII] emission at  $z = 1.475$  based on 1) the shape of the line-profile: the asymmetric wing of the red line is not observed in the blue line and no flux is detected in the middle of the blue and red lines; 2) the extremely low blue-to-red flux ratio that is inconsistent with [OII] and 3) the non-detection of H $\alpha$  at  $z = 1.475$  ruling out even the lowest observed ratios with respect to [OII].
- A tentative detection in the *B* band is explained by a contribution from Ly $\alpha$  emission in a foreground LAE at  $z = 2.142$ , confirmed through [OIII]<sub>5007</sub> emission in addition to Ly $\alpha$ .
- We confirm that COLA1 is a double peaked LAE at  $z_{\text{Ly}\alpha, \text{red}} = 6.593$  and summarise its properties in Table 3. The Ly $\alpha$  luminosity and EW are high ( $L_{\text{Ly}\alpha} = 2.9 \pm 0.1 \times 10^{43}$  erg s<sup>-1</sup> and  $\text{EW}_0 = 85^{+70}_{-40}$  Å) and the Ly $\alpha$  emission is compact ( $r_{50} = 0.33^{+0.07}_{-0.04}$  kpc based on *HST* imaging).
- The Ly $\alpha$  peak separation is  $220 \pm 20$  km s<sup>-1</sup> and the flux of the blue peak is  $0.31 \pm 0.03$  times the flux in the red peak. The Ly $\alpha$  lines are narrow (FWHM =  $198 \pm 14$  km s<sup>-1</sup> and  $150 \pm 18$  km s<sup>-1</sup> for the red and blue peak, respectively) and the full Ly $\alpha$  line-profile resembles the profile of a luminous LAE at  $z = 2.2$  with a prominent blue peak (Fig. 6). Due to the narrowness of the Ly $\alpha$  line, COLA1 is unlikely to be powered by an AGN.
- By modelling the Ly $\alpha$  profile with a five-parameter shell model, we find that the line-profile is characterised by a very low H I column density ( $N_{\text{HI}} = 10^{17.0 \pm 0.3}$  cm<sup>-2</sup>), indicating a non-zero escape of ionising photons. Based on correlations between the escape fraction and the peak separation in LyC leaking galaxies, we infer an escape fraction of  $\approx 15\%$ , but potentially up to  $\approx 30\%$ . Other inferences imply escape fractions up to  $f_{\text{esc, LyC}} \sim 50\%$ . This means that we are witnessing a star-forming galaxy that is actively contributing to the reionisation of the Universe.
- COLA1 has a high UV luminosity,  $M_{1500} = -21.6 \pm 0.3$ , implying a SFR  $\approx 30 M_{\odot} \text{ yr}^{-1}$ . The *Spitzer*/IRAC colours  $[3.6] - [4.5] = -0.2 \pm 0.3$  imply relatively weak contribution from [OIII]/H $\beta$  contributing to the [3.6] band, implying a metallicity as low as  $Z < 10^{-3} (1/20 Z_{\odot})$  or reduced strength of nebular emission lines due to escaping ionising photons.
- The detectability of blue Ly $\alpha$  emission from COLA1 implies a low Ly $\alpha$  optical depth out to  $\approx -250$  km s<sup>-1</sup> of the galaxy, potentially due to a surrounding  $\gtrsim 0.3$  pMpc (2.3 cMpc)

ionised region. Based on the observed properties of COLA1 and simple assumptions (detailed in Table 4), we find that enough ionising photons have escaped COLA1 to ionise a  $\approx 0.3$  pMpc (2.3 cMpc) bubble. However, keeping this region ionised likely requires a contribution from faint neighbouring sources. No faint neighbouring galaxies are known, but COLA1 is likely in the same large scale over-density as CR7 with a separation of  $\approx 30'$  ( $\approx 10$  pMpc) and  $\approx 350$  km s<sup>-1</sup>.

As we have shown, the confirmation of COLA1 as a double peaked LAE at  $z = 6.59$  allows us to witness directly that galaxies contribute to the reionisation of the Universe. Several assumptions made here can be overcome with future observations of the intrinsic line-profile and systemic redshift of COLA1. In particular, future high resolution spectroscopic observations of samples of LAEs may reveal the dependence on double peaked Ly $\alpha$  profiles on luminosity and redshift, providing a new probe of the reionisation process.

## Acknowledgements

JM acknowledges the award of a Huygens PhD fellowship from Leiden University. MG acknowledges support from NASA grant NNX17AK58G. APA, PhD::SPACE fellow, acknowledges support from the FCT through the fellowship PD/BD/52706/2014. Based on observations made with ESO Telescopes at the La Silla Paranal Observatory under programme IDs 294.A-5018, 098.A-0819, 099.A-0254 and 0100.A-0213. We are grateful for the excellent data-sets from the COSMOS and UltraVISTA survey teams. This research was supported by the Munich Institute for Astro- and Particle Physics (MIAPP) of the DFG cluster of excellence "Origin and Structure of the Universe". We thank Len Cowie, Koki Kakiichi, Peter Laursen, Eros Vanzella, Lewis Weinberger and Johannes Zabl for discussions. We have benefited from the public available programming language Python, including the *numpy*, *matplotlib*, *scipy* and *astropy* packages (Hunter 2007; Astropy Collaboration et al. 2013), the astronomical imaging tools *Swarp* (Bertin 2010) and *ds9* and the *Topcat* analysis tool (Taylor 2013).

## References

- Ahn, S.-H., Lee, H.-W., & Lee, H. M. 2001, *ApJ*, 554, 604  
 Astropy Collaboration et al. 2013, *AAP*, 558, A33  
 Bañados, E. et al. 2016, *ApJS*, 227, 11  
 Bañados, E. et al. 2018, *Nature*, 553, 473  
 Bagley, M. B. et al. 2017, *ApJ*, 837, 11  
 Becker, G. D., Bolton, J. S., Madau, P., et al. 2015, *MNRAS*, 447, 3402  
 Bertin, E. 2010, *SWarp: Resampling and Co-adding FITS Images Together*, Astrophysics Source Code Library  
 Bohlin, R. C. 2016, *AJ*, 152, 60  
 Bosman, S. E. I., Fan, X., Jiang, L., et al. 2018, *ArXiv e-prints* [arXiv:1802.08177]  
 Bouwens, R. J., van Dokkum, P. G., Illingworth, G. D., et al. 2017, *ArXiv e-prints* [arXiv:1711.02090]  
 Bouwens, R. J. et al. 2015, *ApJ*, 803, 34  
 Bouwens, R. J. et al. 2016, *ApJ*, 831, 176  
 Bowler, R. A. A., Dunlop, J. S., McLure, R. J., et al. 2014, *MNRAS*, 440, 2810  
 Bowler, R. A. A. et al. 2017, *MNRAS*, 466, 3612  
 Capak, P. et al. 2007, *ApJS*, 172, 99  
 Cardamone, C. et al. 2009, *MNRAS*, 399, 1191  
 Carniani, S. et al. 2017, *ArXiv e-prints* [arXiv:1712.03985]  
 Charlot, S. & Fall, S. M. 1993, *ApJ*, 415, 580  
 Davies, F. B., Becker, G. D., & Furlanetto, S. R. 2017, *ArXiv e-prints* [arXiv:1708.08927]  
 Dijkstra, M., Gronke, M., & Sobral, D. 2016, *ApJ*, 823, 74  
 Dijkstra, M., Lidz, A., & Wyithe, J. S. B. 2007, *MNRAS*, 377, 1175  
 Erb, D. K. et al. 2014, *ApJ*, 795, 33

- Finlator, K., Özel, F., Davé, R., & Oppenheimer, B. D. 2009, MNRAS, 400, 1049
- Furlanetto, S. R., McQuinn, M., & Hernquist, L. 2006, MNRAS, 365, 115
- Gazagnes, S., Chisholm, J., Schaerer, D., et al. 2018, ArXiv e-prints [arXiv:1802.06378]
- Gronke, M. 2017, AAP, 608, A139
- Gronke, M., Bull, P., & Dijkstra, M. 2015, ApJ, 812, 123
- Gronke, M., Dijkstra, M., McCourt, M., & Peng Oh, S. 2017, AAP, 607, A71
- Haiman, Z. 2002, ApJL, 576, L1
- Harikane, Y. et al. 2017, ArXiv e-prints [arXiv:1711.03735]
- Hashimoto, T. et al. 2017, AAP, 608, A10
- Hayashi, M., Sobral, D., Best, P. N., Smail, I., & Kodama, T. 2013, MNRAS, 430, 1042
- Heckman, T. M., Sembach, K. R., Meurer, G. R., et al. 2001, ApJ, 558, 56
- Henry, A., Scarlata, C., Martin, C. L., & Erb, D. 2015, ApJ, 809, 19
- Hu, E. M., Cowie, L. L., Barger, A. J., et al. 2010, ApJ, 725, 394
- Hu, E. M., Cowie, L. L., Songaila, A., et al. 2016, ApJL, 825, L7
- Hunter, J. D. 2007, Computing in Science and Engineering, 9, 90
- Ilbert, O. et al. 2009, ApJ, 690, 1236
- Inoue, A. K., Shimizu, I., Iwata, I., & Tanaka, M. 2014, MNRAS, 442, 1805
- Izotov, Y. I., Worseck, G., Schaerer, D., et al. 2018, ArXiv e-prints [arXiv:1805.09865]
- Jensen, H., Hayes, M., Iliiev, I. T., et al. 2014, MNRAS, 444, 2114
- Jensen, H., Zackrisson, E., Pelckmans, K., et al. 2016, ApJ, 827, 5
- Jiang, L. et al. 2013, ApJ, 772, 99
- Kakiuchi, K., Ellis, R. S., Laporte, N., et al. 2018, ArXiv e-prints [arXiv:1803.02981]
- Kashikawa, N. et al. 2006, ApJ, 648, 7
- Kennicutt, Jr., R. C. 1998, ARAA, 36, 189
- Khostovan, A. A., Sobral, D., Mobasher, B., et al. 2015, MNRAS, 452, 3948
- Khostovan, A. A., Sobral, D., Mobasher, B., et al. 2016, MNRAS, 463, 2363
- Koekemoer, A. M. et al. 2007, ApJS, 172, 196
- Konno, A. et al. 2018, PASJ, 70, S16
- Kulas, K. R., Shapley, A. E., Kollmeier, J. A., et al. 2012, ApJ, 745, 33
- Labbé, I., Bouwens, R., Illingworth, G. D., & Franx, M. 2006, ApJL, 649, L67
- Labbé, I. et al. 2015, ApJS, 221, 23
- Laigle, C. et al. 2016, ApJS, 224, 24
- Laursen, P., Sommer-Larsen, J., & Razoumov, A. O. 2011, ApJ, 728, 52
- Lin, Y., Oh, S. P., Furlanetto, S. R., & Sutter, P. M. 2016, MNRAS, 461, 3361
- Loeb, A. & Rybicki, G. B. 1999, ApJ, 524, 527
- Madau, P. 2017, ApJ, 851, 50
- Malhotra, S. & Rhoads, J. E. 2006, ApJL, 647, L95
- Mas-Ribas, L., Hennawi, J. F., Dijkstra, M., et al. 2017, ApJ, 846, 11
- Mason, C. A., Treu, T., Dijkstra, M., et al. 2018, ApJ, 856, 2
- Matsuda, Y., Yamada, T., Hayashino, T., Yamauchi, R., & Nakamura, Y. 2006, ApJL, 640, L123
- Matsuoka, Y. et al. 2016, ApJ, 828, 26
- Matthee, J., Sobral, D., Darvish, B., et al. 2017a, MNRAS, 472, 772
- Matthee, J., Sobral, D., Oteo, I., et al. 2016, MNRAS, 458, 449
- Matthee, J., Sobral, D., Santos, S., et al. 2015, MNRAS, 451, 400
- Matthee, J. et al. 2017b, ApJ, 851, 145
- McCracken, H. J. et al. 2012, AAP, 544, A156
- Mesinger, A., Aykotalp, A., Vanzella, E., et al. 2015, MNRAS, 446, 566
- Meurer, G. R., Heckman, T. M., & Calzetti, D. 1999, ApJ, 521, 64
- Miley, G. & De Breuck, C. 2008, AAPR, 15, 67
- Modigliani, A. et al. 2010, in SPIE, Vol. 7737, Observatory Operations: Strategies, Processes, and Systems III, 773728
- Nakajima, K. & Ouchi, M. 2014, MNRAS, 442, 900
- Neufeld, D. A. 1990, ApJ, 350, 216
- Oke, J. B. 1974, ApJS, 27, 21
- Ono, Y., Ouchi, M., Shimasaku, K., et al. 2010, ApJ, 724, 1524
- Ouchi, M. et al. 2010, ApJ, 723, 869
- Paulino-Afonso, A., Sobral, D., Darvish, B., et al. 2018a, ArXiv e-prints [arXiv:1805.07371]
- Paulino-Afonso, A. et al. 2018b, MNRAS, 476, 5479
- Pentericci, L. et al. 2014, ApJ, 793, 113
- Rivera-Thorsen, T. E. et al. 2017, AAP, 608, L4
- Sadoun, R., Zheng, Z., & Miralda-Escudé, J. 2017, ApJ, 839, 44
- Saez, C. et al. 2015, MNRAS, 450, 2615
- Salpeter, E. E. 1955, ApJ, 121, 161
- Sanders, R. L. et al. 2016, ApJ, 816, 23
- Santos, M. R. 2004, MNRAS, 349, 1137
- Santos, S., Sobral, D., & Matthee, J. 2016, MNRAS, 463, 1678
- Schmidt, K. B., Treu, T., Bradač, M., et al. 2016, ApJ, 818, 38
- Scoville, N. et al. 2007, ApJS, 172, 1
- Shibuya, T., Ouchi, M., & Harikane, Y. 2015, ApJS, 219, 15
- Shibuya, T. et al. 2018, PASJ, 70, S15
- Siana, B. et al. 2015, ApJ, 804, 17
- Sobral, D., Best, P. N., Matsuda, Y., et al. 2012, MNRAS, 420, 1926
- Sobral, D. & Matthee, J. 2018, ArXiv e-prints [arXiv:1803.08923]
- Sobral, D., Matthee, J., Darvish, B., et al. 2015, ApJ, 808, 139
- Sobral, D., Santos, S., Matthee, J., et al. 2018a, MNRAS, 476, 4725
- Sobral, D., Smail, I., Best, P. N., et al. 2013, MNRAS, 428, 1128
- Sobral, D. et al. 2017a, ArXiv e-prints [arXiv:1710.08422]
- Sobral, D. et al. 2017b, MNRAS, 466, 1242
- Sobral, D. et al. 2018b, ArXiv e-prints [arXiv:1802.10102]
- Songaila, A., Hu, E. M., Barger, A. J., et al. 2018, ArXiv e-prints [arXiv:1805.00490]
- Stark, D. P. et al. 2015, MNRAS, 450, 1846
- Stark, D. P. et al. 2017, MNRAS, 464, 469
- Stefanon, M. et al. 2017, ApJ, 851, 43
- Steidel, C. C., Bogosavlevic, M., Shapley, A. E., et al. 2018, ArXiv e-prints [arXiv:1805.06071]
- Steinhardt, C. L. et al. 2014, ApJL, 791, L25
- Taniguchi, Y. et al. 2007, ApJS, 172, 9
- Taylor, M. 2013, Starlink User Note, 253
- Trainor, R. F., Steidel, C. C., Strom, A. L., & Rudie, G. C. 2015, ApJ, 809, 89
- van Ojik, R., Roettgering, H. J. A., Miley, G. K., & Hunstead, R. W. 1997, AAP, 317, 358
- Vanzella, E. et al. 2016, ApJL, 821, L27
- Vanzella, E. et al. 2018, MNRAS, 476, L15
- Verhamme, A., Orlitová, I., Schaerer, D., & Hayes, M. 2015, AAP, 578, A7
- Verhamme, A., Orlitová, I., Schaerer, D., et al. 2017, AAP, 597, A13
- Verhamme, A., Schaerer, D., & Maselli, A. 2006, AAP, 460, 397
- Weinberger, L. H., Kulkarni, G., Haehnelt, M. G., Choudhury, T. R., & Puchwein, E. 2018, ArXiv e-prints [arXiv:1803.03789]
- Wisotzki, L. et al. 2016, AAP, 587, A98
- Yamada, T., Matsuda, Y., Kousai, K., et al. 2012, ApJ, 751, 29
- Yang, H. et al. 2017, ApJ, 844, 171
- Zackrisson, E. et al. 2017, ApJ, 836, 78
- Zheng, Z.-Y. et al. 2017, ApJL, 842, L22