

# The MUSE Extremely Deep Field: The cosmic web in emission at high redshift<sup>★</sup>

R. Bacon<sup>1</sup>, D. Mary<sup>2</sup>, T. Garel<sup>3,1</sup>, J. Blaizot<sup>1</sup>, M. Maseda<sup>4</sup>, J. Schaye<sup>4</sup>, L. Wisotzki<sup>5</sup>, S. Conseil<sup>6</sup>, J. Brinchmann<sup>7,4</sup>, F. Leclercq<sup>3</sup>, V. Abril-Melgarejo<sup>9</sup>, L. Boogaard<sup>4</sup>, N. F. Bouché<sup>1</sup>, T. Contini<sup>8</sup>, A. Feltre<sup>13,1</sup>, B. Guiderdoni<sup>1</sup>, C. Herenz<sup>14</sup>, W. Kollatschny<sup>10</sup>, H. Kusakabe<sup>3</sup>, J. Matthee<sup>11</sup>, L. Michel-Dansac<sup>1</sup>, T. Nanayakkara<sup>12</sup>, J. Richard<sup>1</sup>, M. Roth<sup>5</sup>, K. B. Schmidt<sup>5</sup>, M. Steinmetz<sup>5</sup>, L. Tresse<sup>1</sup>, T. Urrutia<sup>5</sup>, A. Verhamme<sup>3</sup>, P. M. Weilbacher<sup>5</sup>, J. Zabl<sup>1</sup>, and S. L. Zoutendijk<sup>4</sup>

<sup>1</sup> Univ. Lyon, Univ. Lyon1, ENS de Lyon, CNRS, Centre de Recherche Astrophysique de Lyon UMR5574, 69230 Saint-Genis-Laval, France

e-mail: roland.bacon@univ-lyon1.fr

<sup>2</sup> Laboratoire Lagrange, CNRS, Université Côte d'Azur, Observatoire de la Côte d'Azur, CS 34229, 06304 Nice, France

<sup>3</sup> Observatoire de Genève, Université de Genève, 51 Ch. des Maillettes, 1290 Versoix, Switzerland

<sup>4</sup> Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, The Netherlands

<sup>5</sup> Leibniz-Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, 14482 Potsdam, Germany

<sup>6</sup> Gemini Observatory/NSF's NOIRLab, Casilla 603, La Serena, Chile

<sup>7</sup> Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto, CAUP, Rua das Estrelas, 4150-762 Porto, Portugal

<sup>8</sup> IRAP, Institut de Recherche en Astrophysique et Planétologie, CNRS, Université de Toulouse, 14 avenue Edouard Belin, 31400 Toulouse, France

<sup>9</sup> Aix Marseille Université, CNRS, LAM (Laboratoire d'Astrophysique de Marseille) UMR 7326, 13388 Marseille, France

<sup>10</sup> Institut für Astrophysik, Universität Göttingen, Friedrich-Hund-Platz 1, 37077 Göttingen, Germany

<sup>11</sup> ETH Zurich, Institute of Astronomy, Wolfgang-Pauli-Str. 27, 8093 Zurich, Switzerland

<sup>12</sup> Centre for Astrophysics and Supercomputing, Swinburne University of Technology, Melbourne, VIC 3122, Australia

<sup>13</sup> INAF – Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, Via P. Gobetti 93/3, 40129 Bologna, Italy

<sup>14</sup> European Southern Observatory, Av. Alonso de Córdova 3107, 763 0355 Vitacura, Santiago, Chile

Received 10 November 2020 / Accepted 10 February 2021

## ABSTRACT

We report the discovery of diffuse extended Ly $\alpha$  emission from redshift 3.1 to 4.5, tracing cosmic web filaments on scales of 2.5–4 cMpc. These structures have been observed in overdensities of Ly $\alpha$  emitters in the MUSE Extremely Deep Field, a 140 h deep MUSE observation located in the *Hubble* Ultra-Deep Field. Among the 22 overdense regions identified, five are likely to harbor very extended Ly $\alpha$  emission at high significance with an average surface brightness of  $5 \times 10^{-20}$  erg s<sup>-1</sup> cm<sup>-2</sup> arcsec<sup>-2</sup>. Remarkably, 70% of the total Ly $\alpha$  luminosity from these filaments comes from beyond the circumgalactic medium of any identified Ly $\alpha$  emitter. Fluorescent Ly $\alpha$  emission powered by the cosmic UV background can only account for less than 34% of this emission at  $z \approx 3$  and for not more than 10% at higher redshift. We find that the bulk of this diffuse emission can be reproduced by the unresolved Ly $\alpha$  emission of a large population of ultra low-luminosity Ly $\alpha$  emitters ( $<10^{40}$  erg s<sup>-1</sup>), provided that the faint end of the Ly $\alpha$  luminosity function is steep ( $\alpha \lesssim -1.8$ ), it extends down to luminosities lower than  $10^{38}$ – $10^{37}$  erg s<sup>-1</sup>, and the clustering of these Ly $\alpha$  emitters is significant (filling factor  $<1/6$ ). If these Ly $\alpha$  emitters are powered by star formation, then this implies their luminosity function needs to extend down to star formation rates  $<10^{-4} M_{\odot} \text{ yr}^{-1}$ . These observations provide the first detection of the cosmic web in Ly $\alpha$  emission in typical filamentary environments and the first observational clue indicating the existence of a large population of ultra low-luminosity Ly $\alpha$  emitters at high redshift.

**Key words.** galaxies: high-redshift – galaxies: groups: general – cosmology: observations

## 1. Introduction

The current paradigm of structure formation predicts that most of the gas in the intergalactic medium (IGM) is organized in a “cosmic web” composed of nonuniform gaseous filaments connecting galaxies on scales of megaparsecs (e.g., White et al. 1987; Bond et al. 1996). These filaments feed the circumgalactic medium (CGM), the gaseous component that is responsible for regulating the gas exchange between galaxies and the surround-

ing IGM (e.g., Tumlinson et al. 2017). This complex interplay between the infall of gas from filaments through the CGM and the ejection of matter from the feedback processes is expected to play a key role in the regulation of galaxy growth through cosmic time.

The IGM has mainly been explored over the past decades using absorption line spectroscopy, which provides a powerful way to trace the neutral hydrogen observed in Lyman-alpha (Ly $\alpha$ ) absorption against bright background quasars (Gunn & Peterson 1965; Meiksin 2009). However, it has not been possible to obtain a detailed picture of these filaments as information is limited to one dimension along the line-of-sight to the

<sup>★</sup> Based on observations made with ESO telescopes at the La Silla Paranal Observatory under the large program 1101.A-0127.

background source. The low sky density of sufficiently bright background sources prevents the study of the cosmic web on scales smaller than a few megaparsecs. Only very recently have sparse two-dimensional constraints on the IGM structure at a transverse spatial sampling of a few megaparsecs started to become available (e.g., Lee et al. 2018).

Imaging the cosmic web in emission would provide the missing three-dimensional information. Filaments are predicted to emit the hydrogen Ly $\alpha$  line by fluorescence induced by the ultraviolet background (UVB) radiation. However, the low intensity of the UVB (Haardt & Madau 2012) translates into an expected surface brightness of a self-shielded filament of approximately  $10^{-20}$  erg s $^{-1}$  cm $^{-2}$  arcsec $^{-2}$  at  $z = 3$  (Gould & Weinberg 1996). This challenging level has meant that a direct detection of UVB-induced fluorescent emission from IGM filaments has remained elusive (Gallego et al. 2018).

One solution to overcome this obstacle is to observe selected regions where local ionizing sources, such as bright quasars (i.e., quasi-stellar objects, QSOs) or star forming galaxies, boost the Ly $\alpha$  emission to detectable levels (e.g., Cantalupo et al. 2005). This technique has been successfully used to map the IGM at scales of a few hundred kpc around QSOs (Cantalupo et al. 2014; Martin et al. 2015; Borisova et al. 2016; Fumagalli et al. 2016; Kikuta et al. 2019). To extend the mapping of the IGM, specific fields with multiple QSOs have recently been targeted using the MUSE instrument (Bacon et al. 2010) at ESO's Very Large Telescope (e.g., Arrigoni Battaia et al. 2019; Lusso et al. 2019).

However, the scale currently probed by all these observations is limited to a few hundred kiloparsecs, a scale larger than the CGM of the host galaxy but still too small to probe the filaments at the megaparsec scale relevant to the IGM. A notable exception are the observations of Umehata et al. (2019), who report the detection of a 1.3 Mpc long filament with a mean Ly $\alpha$  surface brightness of  $3 \times 10^{-19}$  erg s $^{-1}$  cm $^{-2}$  arcsec $^{-2}$  in the SSA22 protocluster at  $z = 3.1$  (Steidel et al. 1998).

While these observations have revealed some filamentary structures of ionized gas in very massive structures, they are biased to specific environments. For example, the SSA22 protocluster with its large overdensity of active galactic nuclei (AGNs), submillimeter galaxies and Ly $\alpha$  blobs (Lehmer et al. 2009; Umehata et al. 2018, 2019; Herenz et al. 2020) is an extreme environment. Furthermore, the environment of a QSO with anisotropic UV radiation and a possible excess of tidal debris from past interactions (Canalizo & Stockton 2001) may not be representative of the generic IGM.

Another approach was used by Daddi et al. (2020), who targeted a massive structure, corresponding to a dark matter halo of  $\approx 4 \times 10^{13} M_{\odot}$ , embedded in a giant Ly $\alpha$  nebula at  $z = 2.9$  with Keck's KCWI instrument (Martin et al. 2010). Within 300 kpc, they find three cold gas filaments connected to the central massive galaxy. Such observations give important constraints for models of the formation of galaxies in these massive structures located at the nodes of the cosmic web.

However, such very massive structures are not representative of the filamentary environment that represents 60% of the total gas mass of the Universe and where we expect most of the galaxy formation to occur (e.g., Libeskind et al. 2018; Martizzi et al. 2019). For example, the conclusions reached by Daddi et al. (2020) that gravitational energy is the main physical mechanism responsible for the Ly $\alpha$  emission may not stand in two orders of magnitude less massive environments.

It is therefore highly desirable to obtain 3D information of the IGM in emission in more representative environments. Having multiple detections of the cosmic web structure and its

evolution with redshift would also give fundamental constraints for the simulation and models of structure formation.

In recent years we have used a fraction of our MUSE guaranteed observing time to perform deep field observations in the *Hubble* Ultra-Deep Field (HUDF, Beckwith et al. 2006) at 10 and 30 h depths (Bacon et al. 2017). These observations have increased the amount of available spectroscopic information in the HUDF by more than an order of magnitude (Inami et al. 2017) and enabled several breakthroughs in our understanding of the high redshift universe, notably the discovery of ubiquitous extended Ly $\alpha$  emission from the CGMs around individual galaxies at  $z > 3$  (Wisotzki et al. 2016; Leclercq et al. 2017, 2020).

We also performed stacking experiments to explore the extended Ly $\alpha$  emission around galaxies at larger distances, extending the results for individual halos to a radius of 60 pkpc (physical kpc) or 240 ckpc (comoving kpc) at  $z = 3$  (Wisotzki et al. 2018). Although such a distance is well beyond the predicted virial radii of the host dark matter halos for the individual Ly $\alpha$  emitters of our sample, the stacking process erases all geometrical information and is not adapted to the morphological study of the filamentary cosmic web.

Given the successful experience with these MUSE spectroscopic deep fields, we recently performed a new deep field to push forward in depth. The so-called MUSE Extremely Deep Field (hereafter MXDF) is a single field located within the HUDF area (Fig. 1). It reaches a maximum depth of 140 h and benefits from improved spatial resolution thanks to the recent coupling of MUSE with the ESO Adaptive Optics Facility (AOF) as well as improved data reduction processes (Sect. 2).

In this paper we present new results regarding the cosmic web in emission at  $z > 3$  based on this new data set and complemented by existing HUDF datacubes published in Bacon et al. (2017). Galaxy formation should preferentially take place in the densest part of the cosmic web filaments. Therefore, by selecting moderately overdense regions of Ly $\alpha$  emitters, we should maximize our chances of detecting the diffuse Ly $\alpha$  emission<sup>1</sup> associated with these filaments.

We have explored the datacube in redshift space to select overdense regions of Ly $\alpha$  emitters (Sect. 3). For each overdense region, we search for and study diffuse Ly $\alpha$  emission (Sect. 4). We then discuss the implication of our discoveries in Sects. 5 and 6, and finally conclude in Sect. 7.

We use the *Planck* 2018 cosmological model (Planck Collaboration VI 2020) with  $H_0 = 67.4$  km s $^{-1}$  Mpc $^{-1}$  and  $\Omega_m = 0.315$ . Unless specified, all distances are given in comoving scale (cMpc).

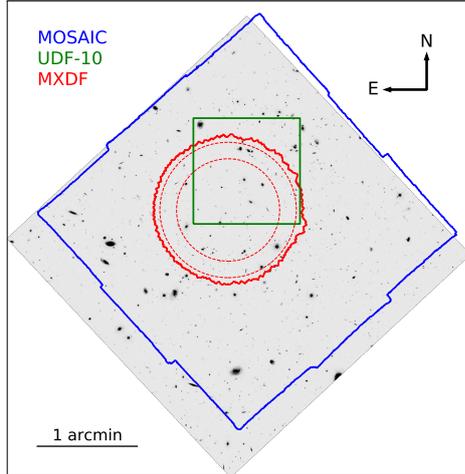
## 2. Observations and data reduction

The details of the observations, data reduction and source catalogs will be given in a forthcoming paper (Bacon et al., in prep.). We provide here a short summary of the process.

### 2.1. Observations

The observing campaign started in August 2018 and lasted until January 2019, for a total of six runs performed during the new moon periods. All observations were performed with the dedicated VLT GALACSI/AOF Ground-Layer Adaptive Optics system (Kolb et al. 2016; Madec et al. 2018). A total of 155 h

<sup>1</sup> By “diffuse emission” we mean Ly $\alpha$  emission that is spatially extended and not resolved in Ly $\alpha$  emitters at our limiting flux. A practical definition is provided in Sect. 4.2.



**Fig. 1.** Location of the three deep fields used in this paper – MXDF (140 h depth), MOSAIC (10 h depth), and UDF-10 (30 h depth) overlaid on the HST *F775W* UDF image. The two dotted red circles show the MXDF 10 and 100 h exposure time contours.

of integration were obtained. After the rejection of bad exposures, the achieved final depth is 140 h. To minimize systematics, the field of view was rotated by a few degrees between each observing block, which consist of  $4 \times 25$  min exposures. Consequently, the final combined field of view is approximately circular (Fig. 1) with a radius of  $41''$  and  $31''$  for respectively 10+ and 100+ h depths. The field center celestial coordinates are  $53^{\circ}16467, -27^{\circ}78537$  (J2000 FK5).

## 2.2. Data reduction

The data reduction is similar to the procedure developed in the first non-AO observations of the UDF (Bacon et al. 2017). It is based on the MUSE pipeline (Weilbacher et al. 2020) and includes a number of new developments. A super flatfield is performed for each exposure by combining, without any recentering and derotation, a large number of observations obtained during the same run. This “superflat” is subtracted from each individual reduced datacube, all of which are then combined into the final datacube. Remaining sky subtraction residuals are removed with ZAP (version 2.0, Soto et al. 2016). The datacube propagated variance is rescaled to take the impact of noise covariance due to the interpolation process into account (see Bacon et al. 2017). In the region with more than 100 h depth, the average  $5\sigma$  surface brightness detection limit at  $7000 \text{ \AA}$  is  $1.3 \times 10^{-19} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ arcsec}^{-2}$  for an unresolved emission line summed over  $3.75 \text{ \AA}$  (3 spectral pixels) and  $1 \text{ arcsec}^2$  ( $5 \times 5$  spaxels) and the corresponding  $5\sigma$  point source limiting flux is  $2.3 \times 10^{-19} \text{ erg s}^{-1} \text{ cm}^{-2}$  for the same emission line.

Because of the absence of bright stars in the field, we use the *muse-psfrec* software (Fusco et al. 2020) to estimate the spatial point-spread function (PSF) from the real-time Adaptive Optics telemetry information recorded by GALACSI. The PSFs of each observation are then combined to produce a measure of the final image quality. The PSF is modeled as a Moffat function (Moffat 1969) whose parameters (FWHM and  $\beta$ ) change smoothly with wavelength. Thanks to the AO performance, an excellent image quality (Moffat FWHM) of  $0.6''$  and  $\beta = 2.1$  at  $4700 \text{ \AA}$  to  $0.4''$  and  $\beta = 1.8$  at  $9300 \text{ \AA}$  is achieved for this very deep exposure.

## 2.3. Source detection and classification

We perform two types of source detection and extraction: a blind source detection with ORIGIN and source extraction and deblending using HST prior information with ODHIN. The ORIGIN software (Mary et al. 2020) has been developed to automatically detect faint line emitters in MUSE datacubes. It is optimized for the detection of compact sources with faint spatial-spectral emission signatures and provides an automated and reliable estimate of the purity (i.e., related to the proportion of false discoveries). The software was run with a purity threshold value of 0.8 and resulted in the detection of 2137 emission lines grouped into 1002 sources.

The ODHIN software (Bacher 2017) uses the higher spatial resolution provided by the HST images to perform deblending of sources in the MUSE datacube. The approach is similar to TDOSE (Schmidt et al. 2019) with the difference that it is non-parametric.

For the inspection we limit the ORIGIN sources to the inner 16+ h region (corresponding to an  $80''$  diameter) to avoid an increase in the false detection rate at the edge of the field when the S/N decreases rapidly. This reduces the number of sources to inspect to 845. Similarly, only 389 HST sources from a total of 1387 within the MXDF field have been selected for inspection, using a S/N continuum cut of 0.8 per spectral pixel.

Evaluation of the redshift solutions provided by the Marz cross-correlation software (Hinton et al. 2016) is performed by three independent experts<sup>2</sup>. After reconciliation of the disagreement between experts, a final catalog of 733 sources with redshift, matching sources and confidence is produced.

95% of the  $\text{Ly}\alpha$  emitters (LAEs) result from ORIGIN detections. In this case, an optimal extraction is performed on the raw datacube using the ORIGIN correlation pseudo narrowband image as a weighting map. While this will produce a more precise flux estimation than continuum based extraction for the case of extended  $\text{Ly}\alpha$  emission, it will nevertheless miss part of the  $\text{Ly}\alpha$  extended halo flux<sup>3</sup>.

Flux and equivalent width measurement of the  $\text{Ly}\alpha$  line is performed with *pyplatefit*, a Python enhanced version of the PLATEFIT IDL software developed for the analysis of the SDSS survey (Tremonti et al. 2004; Brinchmann et al. 2004). Among the improvements implemented in *pyplatefit* that are of interest here, one can mention the asymmetric Gaussian (simple and double) line fit and the bootstrap method of evaluating robust errors.

The  $\text{Ly}\alpha$  redshift is based on the peak of the  $\text{Ly}\alpha$  line. We note that, because of the resonant scattering properties of the  $\text{Ly}\alpha$  photons in the interstellar medium, the  $\text{Ly}\alpha$  redshift is systematically different from the systemic redshift (e.g., Shapley et al. 2003; McLinden et al. 2011; Rakic et al. 2011; Song et al. 2014). Typical velocity offsets are  $\approx 200 \text{ km s}^{-1}$  for LAEs, with larger values ( $\approx 500 \text{ km s}^{-1}$ ) for Lyman-break galaxies (Shibuya et al. 2014; Muzahid et al. 2020). As reported by Shibuya et al. (2014) and Muzahid et al. (2020), there is a strong anti-correlation between the  $\text{Ly}\alpha$  velocity offset and the  $\text{Ly}\alpha$  equivalent width. Given the high equivalent widths of our LAE sample (see Fig. 5) one can expect an average velocity offset  $\approx 200 \text{ km s}^{-1}$ .

For the vast majority of our LAE sample, there are no alternatives to  $\text{Ly}\alpha$  redshifts: nonresonant emission lines such as

<sup>2</sup> Scientists already experienced in redshift measurement with MUSE or specifically trained for this activity.

<sup>3</sup> On average, 65% of the  $\text{Ly}\alpha$  flux is in the extended halo (Leclercq et al. 2017).

CIII]<sub>1907,1909</sub> are outside the spectral window or too faint to derive a systemic redshift from and the continuum is too faint to measure any reliable absorption line or feature. A possibility is to use the empirical correlation found by Verhamme et al. (2018) between the FWHM of the Ly $\alpha$  line and the velocity offset. For double-peaked Ly $\alpha$  lines, the redshifts reported by the double asymmetric fits are measured from the averages of the two peak central wavelengths. As shown by Verhamme et al. (2018), this value is a better approximation of the systemic redshift than the Ly $\alpha$  peak location.

We note that very precise absolute redshifts are not essential for this paper given the width of the wavelength window used in the search (Sect. 3). No attempt is made to further correct the Ly $\alpha$  redshift to the systemic value. What matters, however, is the scatter of the Ly $\alpha$  redshifts around their median value at a given systemic redshift. Using the sample of 55 galaxies of Verhamme et al. (2018), we derive a scatter of  $\pm 95$  km s<sup>-1</sup> (95% probability percentile), which is smaller than the windows that we will use.

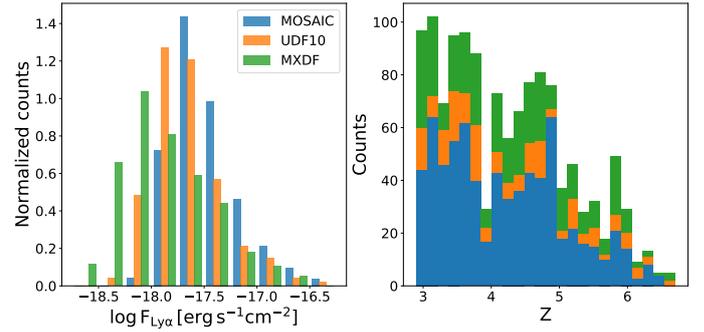
#### 2.4. Source catalog

We combine the revised UDF-10 (30 h depth,  $1 \times 1$  arcmin<sup>2</sup>) and MOSAIC (10 h depth,  $3 \times 3$  arcmin<sup>2</sup>) catalogs (Inami et al. 2017; Bacon et al., in prep.) with the new sources discovered in the MXDF (Fig. 1). Selecting all sources with  $z > 2.9$  gives a total of 1258 LAEs in the 9 arcmin<sup>2</sup> field of view of the HUDF observed with MUSE. We note that a low fraction of sources (55 that is 4%) are not strictly speaking Ly $\alpha$  emitters but are detected from their strong Ly $\alpha$  absorption, often associated with weak Ly $\alpha$  emission. These galaxies are brighter in the continuum than the overall population of Ly $\alpha$  emitters and are more typical of Lyman-break galaxies. While a few Ly $\alpha$  emitters have additional detected UV lines such as CIII]<sub>1907,1909</sub> or CIV<sub>1548,1550</sub>, there is no source in the catalog without Ly $\alpha$  detection at  $z > 2.9$ .

In the deepest MXDF region (depth  $> 100$  h), the LAE density reaches 375 galaxies per arcmin<sup>2</sup>. It is the densest collection of LAEs ever obtained in a single field. To illustrate this point, one can perform a quantitative comparison with the large Ly $\alpha$  emitter narrowband survey performed around the SSA22 field by Yamada et al. (2012). The authors identify 2161 candidate Ly $\alpha$  emitters in the redshift range [3.062–3.125], leading to a LAE average surface density of 0.20 arcmin<sup>-2</sup>. Using the same redshift range we obtain, thanks to the high sensitivity to low luminosity LAEs, a LAE surface density of 6.2 and 17.9 arcmin<sup>-2</sup> in, respectively, the MOSAIC and 100+ h depth MXDF field of view.

The catalog covers a wide redshift range from  $z = 2.90$  to 6.65 (Fig. 2), with a significant part at high redshift (346 LAEs at  $z > 4.8$ ). Redshift confidence has been attributed by experts, from one (low confidence) to three (very high confidence). There are 402, 513, and 343 Ly $\alpha$  emitters with confidence one, two and three, respectively. The lower confidence LAEs are generally reliable detections (i.e., they are obtained with ORIGIN set at a high purity level) but with lower S/N Ly $\alpha$  line profiles where the Ly $\alpha$  redshift solution is the most probable. In this paper we use all LAEs irrespective of their redshift confidence.

An important fraction (40%, 504) of LAEs has no entry in the Rafelski et al. (2015) HUDF photometric catalog. For a small part (10%, 50) it has not been possible to select a unique HST counterpart among the few possibilities. In some other cases (6%, 30) an HST counterpart can be seen in some of the HST bands that was missed by the automatic identification performed by Rafelski et al. (2015). But for the vast majority of the 504 sources without a Rafelski et al. (2015) counterpart, namely 84%



**Fig. 2.** Ly $\alpha$  flux (left panel, bin width  $0.25$  erg s<sup>-1</sup> cm<sup>-2</sup>) and redshift (right panel, bin width 0.2) distribution of the 1258 Ly $\alpha$  emitters.

(423 objects) no detectable HST counterpart is observed. A previous study of 103 similar sources in this field, anterior to MXDF observations, shows that these sources are the high equivalent width tail of the Ly $\alpha$  emitter population (Maseda et al. 2018) and that they have on average a high ionizing photon production efficiency (Maseda et al. 2020).

### 3. Redshift overdensities of Ly $\alpha$ emitters

A visual inspection of the datacube shows that LAEs are not distributed uniformly but exhibit strong clustering in redshift space. Here we focus on the detection and properties of LAE redshift overdensities. The method used is described in the following section.

#### 3.1. Choice of coordinate system

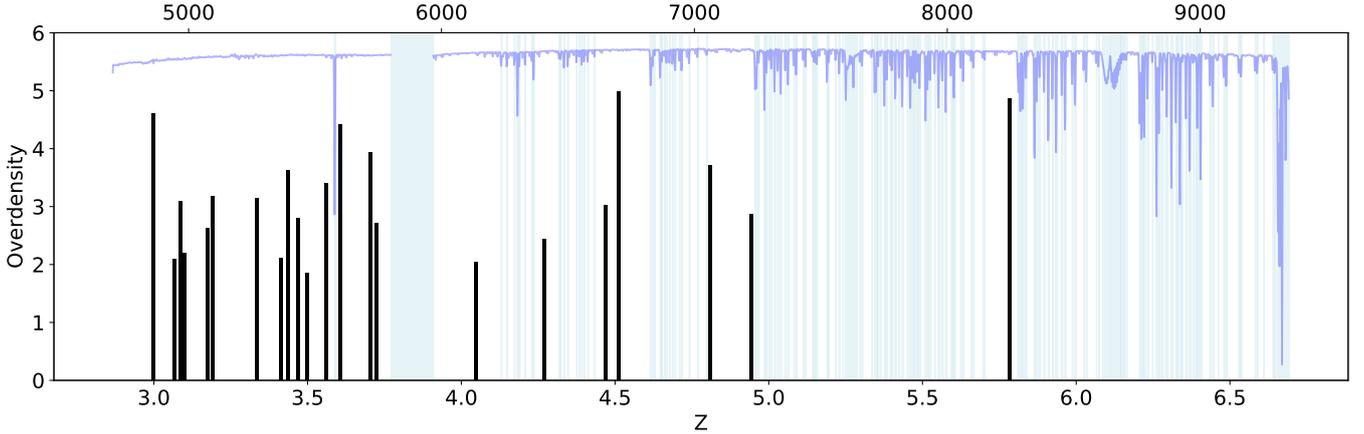
Given our small field of view, finding overdensities is nothing more than performing a binning of the redshift distribution followed by a peak detection. There are, however, some subtleties to take into account. One of them is the choice of coordinate system which is strategic given the very wide redshift range (2.9–6.7) probed by our observations.

A first possibility selected by Cohen et al. (1999) and Gilli et al. (2003) is to use bins in velocity space ( $c \ln[1 + z]$ ). A fixed bin in rest-frame velocity space has the advantage of sampling the Ly $\alpha$  line, independent of the Hubble flow. The sampling must be large enough given that the Ly $\alpha$  profile is generally broad and shifted in velocity<sup>4</sup>.

An alternative is to use a fixed physical scale which allows us to compare the sizes of physical objects independently of the redshift. In this coordinate system, a sampling of 2 Mpc corresponds to 600 km s<sup>-1</sup> at  $z = 3$  and 1400 km s<sup>-1</sup> at  $z = 6$ . Given the small evolution in redshift of the Ly $\alpha$  line profile (Hayes et al. 2021), such a large spread of the sampling with redshift will be suboptimal for the diffuse emission signal extraction.

The last possibility is to use comoving coordinates. In this coordinate system, the cosmic web filaments keep their structural properties along the expansion of the Universe. The sampling evolution with redshift is also less important than for the physical scale: an 8 cMpc sampling translates into 610 km s<sup>-1</sup> at  $z = 3$  and 800 km s<sup>-1</sup> at  $z = 6$ . We therefore selected this last

<sup>4</sup> Using the expected  $\pm 95$  km s<sup>-1</sup> velocity shift scatter around the median value computed in Sect. 2.3 and the 370 km s<sup>-1</sup> rest-frame median FWHM measured line width, we derive a value of 420 km s<sup>-1</sup> for the minimum spectral window size to capture the bulk of the Ly $\alpha$  flux.



**Fig. 3.** LAE overdensities ( $\delta$ ) found in the HUDF as a function of redshift. The top axis displays the corresponding wavelengths in  $\text{\AA}$ . The noise spectrum is shown in blue on the reverse y-axis. Regions excluded from the overdensity search are displayed in pale blue. The 5800–5966  $\text{\AA}$  masked region is due to the sodium notch filter used to filter out the light from the bright AO Laser guide star.

option, which is best-suited to the object of the study and is not far from optimal for signal detection.

### 3.2. Detection of overdensities

Although the search for diffuse emission will be done in the MXDF deep area, the search for overdensities is performed on the full LAE catalog, namely in the entire MOSAIC field of view. With an area ten times larger than in the deeper region of the MXDF, the  $3 \times 3$  arcmin<sup>2</sup> MOSAIC field of view improves the chance to detect large-scale overdensities. The search is done in two steps: we start by identifying peaks in the redshift distribution using a wide sampling. Group redshifts are then refined with a finer sampling, and their overdensity is estimated by comparing the number of LAEs in the group with the expected mean value.

We computed the histogram of the LAE population in comoving space, imposing a constant step of 8 cMpc over the whole redshift range. Given the expected faintness of the signal, we aggressively masked all redshift bins with sky lines. This removes 46% of the bins, mostly at  $z > 5$  (Fig. 3).

The background value ( $B$ ) was estimated as the mean value of the number of LAEs in a 8 cMpc bin after 3 sigma clipping. It should be noted that the possible redshift dependence of the background was ignored in this preliminary first step<sup>5</sup>. For each histogram peak, its signal to noise ratio (S/N) was estimated as the peak signal ( $S$ ), that is the excess number of LAEs over the background and divided by the noise value  $\sqrt{S + \sigma_B^2}$ , where  $\sigma_B$  is the standard deviation of the background value. After some tests we selected a S/N cut of 2.5 and a minimum number of LAEs of seven. This resulted in 24 peaks.

Naturally, the number and redshifts of overdensities is a function of the input parameters: namely the S/N cut, the minimum number of members and the window size. We checked the sensitivity of the method by playing with these parameters. Increasing the S/N cut from two to three decreases the number of overdensities from 37 to 11. The number of overdensities is stable with the imposed minimum number of LAEs up to ten,

<sup>5</sup> This assumption will introduce some bias in the overdensity selection. This could be a problem for an exhaustive study of overdensities but our aim is restricted to find the most overdense regions. As shown later in this section, the final overdensity estimate takes into account the redshift evolution of the mean LAE density.

and decreases rapidly from 24 to 15 for larger values of the minimum number of members. The selected redshift window size of 8 cMpc is optimal: The number of detections decreases from 24 to 21 and from 24 to 15 when the window size is changed to 6 and 10 cMpc, respectively.

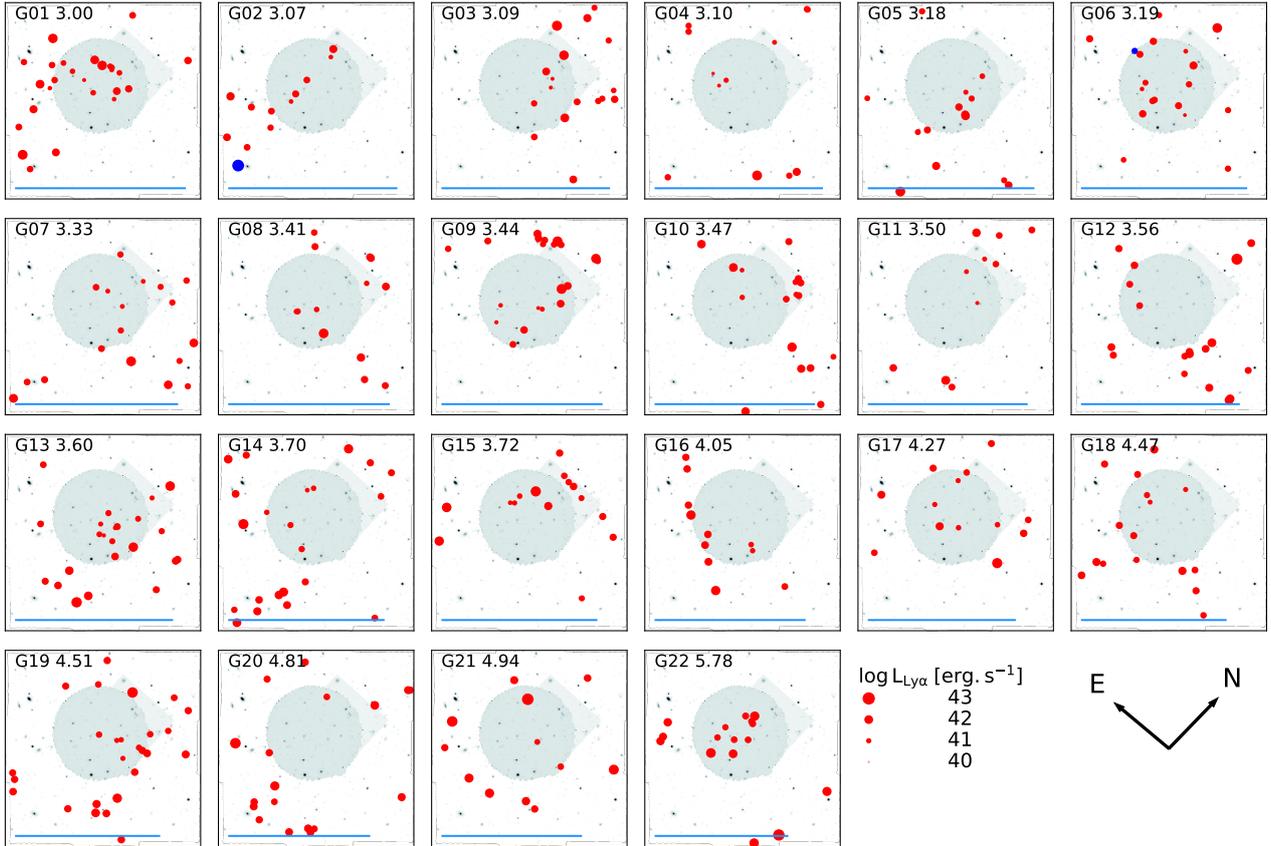
We did not try to further optimize the detection parameters given that it is not our goal to have an exhaustive list of overdensities, but to find the most overdense regions within a redshift window size suitable for the search for diffuse Ly $\alpha$  emission (i.e., large enough with respect to the Ly $\alpha$  line width but not too large to dilute the signal). We therefore proceed to use the optimal 8 cMpc window and the S/N cut of 2.5, which is a good compromise in the number of overdensities to explore.

In a second step we refined the group measurement for the selected peaks using a histogram with a finer grid of 50 kpc sampling. The 8 cMpc size of the window was kept fixed, but its center was slightly adjusted around the initial peak value to maximize the number of LAE members. Groups overlapping in redshift were merged into a single fixed 8 cMpc size window. This reduced the number of overdensities by two. The final group catalog, given in Table C.1, is composed of 22 overdensities. Groups have on average 17 members with a minimum of ten and a maximum of 26.

To estimate the overdensity, that is the ratio of the LAE number density in groups with respect to the mean of the overall population, we first computed the evolution of the mean LAE density with redshift. We partitioned the 2837 sky free redshift bins into 5 redshift bins with a similar numbers of LAEs and computed their mean LAE densities. The mean density decreases approximately linearly with redshift from 0.0201 cMpc<sup>-3</sup> in the  $z = [2.86-3.45]$  bin to 0.0063 cMpc<sup>-3</sup> in the  $z = [5.75-6.65]$  last bin.

For each group we can then estimate its overdensity ( $\delta$ ) by comparing the number of group members ( $N_{\text{lac}}$ ) to the number of expected LAEs using their mean density ( $\bar{\rho}_{\text{lac}}$ ) and the group volume ( $\Delta V = 8 \text{ cMpc} \times S$ ), with  $S$  the area of the MOSAIC field<sup>6</sup>:  $\delta = N_{\text{lac}} / (\bar{\rho}_{\text{lac}} \Delta V)$ . The group overdensities are shown in Fig. 3 and the corresponding spatial locations of LAEs in the groups in Fig. 4. The group numbering follows the redshift. There are 370 LAEs in groups, that is 29% of the total LAE sample. The mean overdensity is 3.2 and the densest group is at  $z = 4.5$  and has  $\delta = 5.0$ .

<sup>6</sup>  $S = 5.7^2 - 7.3^2 \text{ cMpc}^2$  at  $z = 3$  and 6, respectively



**Fig. 4.** Spatial distribution of LAEs in overdensities. LAEs are shown as red circles sized according to the log of their  $\text{Ly}\alpha$  luminosity. The blue circles in groups 2 and 6 identify the AGN (also sized according to the log of their  $\text{Ly}\alpha$  luminosity). The location of each group member is overlaid on the HST  $F775W$  gray scale image. The field locations are shown as a gray circle (MXDF) superimposed on a square (UDF-10). Group ID and redshift are labeled. Fields have the MOSAIC size and orientation, that is 3 arcmin and  $42^\circ$  PA. The blue line at the bottom of each image indicates the 5 cMpc scale.

### 3.3. Properties of overdensities

Figure 3 shows a clear trend with redshift: 2/3 of the groups are found at  $z < 3.8$ , in 1/3 of the free-sky comoving total accessible volume. We examine the correlation of overdensities with the AGN population. The 7 Ms *Chandra* Deep Field South (CDFS) catalog (Luo et al. 2017) has 65 identified AGN at  $z > 2.9$  in the full area ( $484 \text{ arcmin}^2$ ). We found ten overdensities with AGN at similar redshift ( $\Delta z < 0.01$ ) within the CDFs area. As might be expected, a significant fraction (45%) of overdensities harbor an AGN. However, the UDF area is a tiny fraction (2%) of the total CDFs area and there are only two overdensities with an AGN located within the UDF: group 2 at  $z = 3.07$  and group 6 at  $z = 3.19$ .

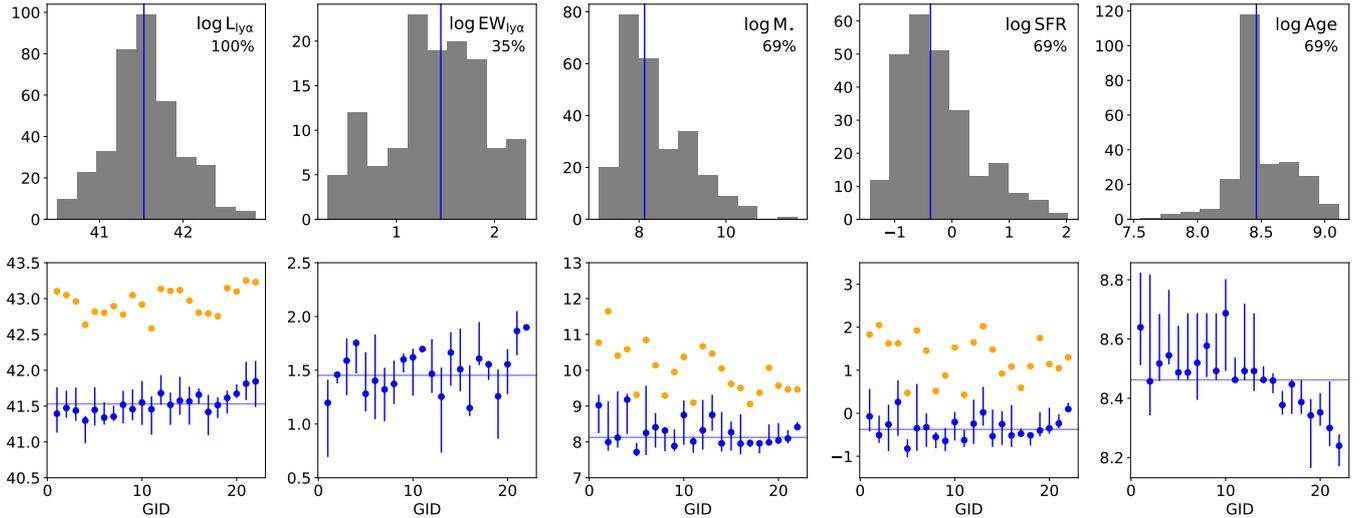
As shown in the first panel of Fig. 5, LAEs in overdensities are low-luminosity  $\text{Ly}\alpha$  emitters:  $\log(L_{\text{Ly}\alpha} \text{ erg s}^{-1}) = 41.5 \pm 0.4$ . We also display in panel 2 the  $\text{Ly}\alpha$  rest-frame equivalent width (EW) for the subset of  $\text{Ly}\alpha$  emitters that are bright enough in their continuum to derive a meaningful equivalent width (i.e., with an EW  $S/N > 3$ ). We note that for the majority of  $\text{Ly}\alpha$  emitters (65%), their continuum is too faint, and thus we can only derive lower limits for the equivalent width. As already pointed out, many of these galaxies are so faint in the continuum that they are not even detected in the deepest HST images (Maseda et al. 2018).

An important fraction (69%) of group members have an entry in the HST Rafelski et al. (2015) catalog. We use the exquisite ancillary information provided by the HST photom-

etry to derive additional physical parameters for the groups. The stellar mass, SFR and specific SFR (sSFR) are inferred via SED fitting using the high- $z$  extension of the code MAGPHYS (da Cunha et al. 2008, 2015) and enabling for a minimum stellar mass of  $10^6 M_\odot$  (see Sect. 3.2 of Maseda et al. 2017). We use HST photometry from the UVUDF catalog of Rafelski et al. (2015) which comprises WFC3/UVIS  $F225W$ ,  $F275W$  and  $F336W$ ; ACS/WFC  $F435W$ ,  $F606W$ ,  $F775W$ , and  $F850LP$  and WFC/IR  $F105W$ ,  $F125W$ ,  $F140W$  and  $F160W$ .<sup>7</sup> On average, the LAEs in overdensities are low-mass ( $1.4 \times 10^8 M_\odot$ ), young (0.3 Gyr) galaxies with high specific star formation rates ( $\text{SFR} > 0.4 M_\odot \text{ yr}^{-1}$ ;  $\text{sSFR} > 10^{-8.5} \text{ yr}^{-1}$ ), higher than the typical values at these redshifts (Schreiber et al. 2015; Salmon et al. 2015).

Regarding the remaining population (31%) of LAEs not detected in the HST images, we cannot derive quantities like SFR or stellar mass for those galaxies without robust continuum detections from HST. However, as demonstrated by Maseda et al. (2018, 2020), this selection yields faint ( $M_{\text{UV}} \approx -15$ ), star-forming ( $\beta \approx -2.5$ ) galaxies on average. While we cannot directly determine the stellar masses from the UV continuum alone, an extrapolation to the Duncan et al. (2014)  $M_{\text{UV}}-M_\odot$  relation to  $M_{\text{UV}} = -15$  implies that these galaxies should have

<sup>7</sup> We note, however, that many of these galaxies are very small and faint, and therefore do not have any reliable longer wavelength photometry. In those cases, the parameters derived from MAGPHYS are more uncertain.



**Fig. 5.** Properties of LAEs in overdensities. *From left to right columns:* log Ly $\alpha$  luminosity in  $\text{erg s}^{-1}$ , the rest-frame log Ly $\alpha$  equivalent width in  $\text{\AA}$ , the log stellar mass in  $M_{\odot}$ , the log star formation rate in  $M_{\odot} \text{ yr}^{-1}$  and the log age in years. Except for the Ly $\alpha$  luminosity, which used all LAEs, only subsets are used for the Ly $\alpha$  equivalent width (35%) and for Mass, SFR, and Age (69%). See Sect. 3.3 for the subsample definition. *Top row:* histograms of these properties. *Bottom row:* corresponding global properties for each group (GID): the blue circle symbols show the group medians and the 25% and 75% percentiles, the orange symbols are the summed property for all galaxies in the group. The median value for each property for the full sample is shown as a blue line in each panel. The GID increases with redshift.

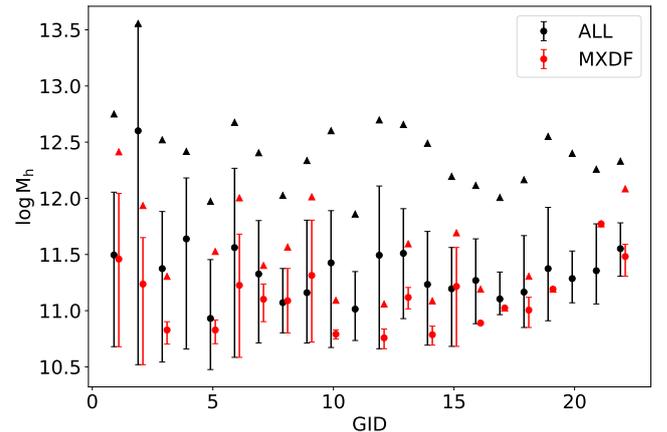
stellar masses below  $10^7 M_{\odot}$  at these redshifts. Using this average mass, we estimate the contribution of the HST-undetected LAEs to the total stellar mass of the overdensities to be on average  $\approx 1\%$ , with a maximum of 7% for group 20. Figure 5 displays the main properties derived from the SED fitting, in addition to the Ly $\alpha$  luminosity and equivalent width.

We estimate the individual dark matter halo masses associated with each group member with an HST counterpart using the galaxy stellar masses derived above and the stellar-to-halo-mass relation (SHMR) derived by Girelli et al. (2020). The overdensity mass is then the sum of the halo masses of each galaxy. We note that the observational data used by Girelli et al. (2020) to constrain their SHMR model only extend out to  $z \approx 4$  and consist of galaxies generally more massive than our sample ( $M_{\text{star}} > 10^{10} M_{\odot}$  at  $z = 4$ ). However, Girelli et al. (2020) show that their SHMR at  $z = 3$  is between the estimates of Behroozi et al. (2019) and Moster et al. (2018) who use deeper data. We thus chose the fits from Girelli et al. (2020) as a middle guess, and we note that the errors on the SHMR are probably on the order of a factor of two, which is small with respect to uncertainties on our estimates of the stellar masses. Our results are shown in Fig. 6 where we see an average halo mass  $\sim 10^{11.3} M_{\odot}$ , and almost no halos more massive than  $\sim 10^{12} M_{\odot}$ . This is typical of the LAE population we survey (e.g., Garel et al. 2015a), and our groups are likely progenitors of galaxies like our own (Garel et al. 2015b). We note the exception of group 2, which contains one very massive halo of  $10^{13.5} M_{\odot}$ , outside the MXDF region. These massive environments, typical of proto-clusters, are known to host AGNs.

In the next section we use the discovered overdensities to search for extended Ly $\alpha$  emission.

#### 4. Extended Ly $\alpha$ emission

Numerical simulations predict the Ly $\alpha$  emission to occur along filaments that are 50–100 kpc wide and a few megaparsecs long (see for example Fig. 7 of Rosdahl & Blaizot 2012). The surface brightness of the filaments is predicted to be in the range



**Fig. 6.** Estimated dark matter halo masses in the overdensities (see Sect. 3.3). The total (upper triangle), average (filled circle) and min-max (bars) values of dark matter halo masses in overdensities are displayed for all group members with an HST counterpart either within the MOSAIC field of view (in black), or restricted to the MXDF area (in red). Units are  $\log M_{\odot}$ .

$1-10 \times 10^{-20} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ arcsec}^{-2}$ , depending on the halo mass (e.g., Gould & Weinberg 1996). Such a low surface brightness will not be detectable in the 10h deep MOSAIC field, so we restrict the search to the MXDF region.

In each overdense region, the MXDF volume is a cylinder of 2.5 cMpc diameter in the transverse direction and 8 cMpc along the line-of-sight. Our goal is to detect faint Ly $\alpha$  emission in between the galaxies.

To perform this detection we developed a two-step method. In the first step, we start by performing a segmentation of the narrowband signal-to-noise (S/N) image to isolate the regions of extended emission. In the second step we compute the Ly $\alpha$  flux and its error in these segmented areas using the continuum-subtracted narrowband image. Extended Ly $\alpha$  emission is marked as detected if the computed S/N and the area of the extended

region are large enough. Here that means the estimated probability that noise can explain such a S/N value by chance is sufficiently low.

We note that the two steps are somewhat independent. Alternative methods with different signal transformation to isolate the extended emission could be used to identify diffuse emission segments. Whatever the method, the flux and error computation are always performed in the original continuum subtracted narrowband image. The details of the methodology is described in the following sections.

#### 4.1. Multiscale analysis of narrowband images

Detection and estimation of diffuse emission in the presence of sources that are locally two orders of magnitude brighter is a difficult signal processing problem. The additional complexity is that Ly $\alpha$  emitters are themselves extended and thus cannot be considered as point sources.

The geometry of the problem calls for a multiscale approach. An interesting tool in this respect is the wavelet transform and, given the isotropy of astronomical sources, in particular the Isotropic Undecimated Wavelet Transform (IUWT, Bijaoui et al. 1994; Starck et al. 1998). Compared to classical Gaussian filtering, this wavelet transform allows for a separation of the sources according to its spatial scale. In the classical version of the IUWT filterbank, the original image can be directly resynthesized by the simple sum of the multiscale approximation coefficients. This feature is very interesting for detection or denoising approaches operating scale by scale (Starck et al. 1998). Such wavelet signal decompositions have been successfully used in astronomy for various applications: for example X-ray source detection (Finoguenov et al. 2020), diffuse light study in compact group of galaxies (da Rocha & Oliveira 2005) and the estimation of faint and diffuse radio components (Dabbech et al. 2015; Ammanouil et al. 2019).

For each overdensity, we started by building a narrowband S/N image by summing the continuum subtracted S/N cube over a window centered on the mean group velocity with a width corresponding to 8 cMpc (see Sect. 3). This corresponds to nine and 18 wavelength channels at  $z = 3$  and  $z = 6$ , respectively.

As the continuum subtracted datacube we use the PCA subtracted S/N cube provided by ORIGIN. The ORIGIN process (Mary et al. 2020) is efficient for removing the continuum of bright galaxies without leaving many artifacts, which can be problematic for low surface brightness detection.

We then decomposed this narrowband image with the classical perfect reconstruction filterbank of the IUWT (Bijaoui et al. 1994; Starck et al. 1998, 2007). In the considered setting, the IUWT decomposes the image into eight component images<sup>8</sup>, from high to low frequencies. The high frequency noise mostly lives in the first two scales, bright and compact sources are well captured by the next two scales and the diffuse, large scale components of the image mostly live in the four remaining scales. The angular size of the spatial structures captured in each band is determined by the impulse responses of the IUWT analysis filters, whose FWHM are respectively in the range [0.2–0.6] arcsec (high frequency band), [0.6–2.2] arcsec (medium frequency band) and [2.2–37] arcsec (low frequency band). As shown in Fig. 7 for group 2 at  $z = 3.07$ , this process filters out the noise from the signal (high freq. panel) and cleanly

separates the LAE signal (medium freq. panel) from the large-scale diffuse structure (low freq. panel).

#### 4.2. Shape and flux of Ly $\alpha$ diffuse emission

We used the IUWT S/N images obtained by the process described above to separate the signal of the extended Ly $\alpha$  emission from the Ly $\alpha$  emitter signal.

We started to identify compact sources by segmenting the medium frequency images using a threshold of  $3.5\sigma$  (right upper panel of Fig. 7). We used this segmentation map to mask all outliers and compact sources in the original S/N image, replacing the mask values by an average S/N value obtained from a first estimate of the diffuse components. This first crude estimate is the average value of coefficients above a threshold of  $1.5\sigma$  in the low frequency S/N image. We then again applied the IUWT decomposition to this S/N masked and filled image. This iteration removed most of the low frequency signal due to the compact sources (see lower left and central panels of Fig. 7).

We then performed a segmentation of the resulting low frequency image, using a threshold of  $1.5\sigma$  (lower central panel of Fig. 7). We discarded all segments that have a size smaller than 200 pixels (i.e., 8 arcsec<sup>2</sup>). This last step is motivated by the desire to keep only contiguous surfaces that are extended with respect to the PSF size ( $\approx 5 \times FWHM$ ).

At the edge of the MXDF field, the exposure time drops rapidly from 140 to 10 h. Besides having much lower S/N, this outer zone is also subject to larger systematics since it results from a smaller number of individual exposures. In this region, the probability of false detection is thus much larger. Therefore, we discarded all segments that have a mean depth (averaged over the full segment area) less than 60 h (see the corresponding location in the lower right panel of Fig. 7).

To avoid potential pollution by low- $z$  emission line interlopers, we identified all galaxies in the full source catalog with emission lines other than Ly $\alpha$  having  $S/N > 5$  and inside the group window. Any interloper which falls inside the filaments was masked out for the filament's flux estimation.

The final segmentation (lower right panel of Fig. 7) is composed of a series of extended structures, called “filaments” in the rest of the text. We also identified the area covered by the compact sources detected in the first step within each filament. The remaining area within the filaments is called the diffuse extended emission in contrast to the compact source area. We will use this terminology in the rest of the document, referring to filament, compact and diffuse area.

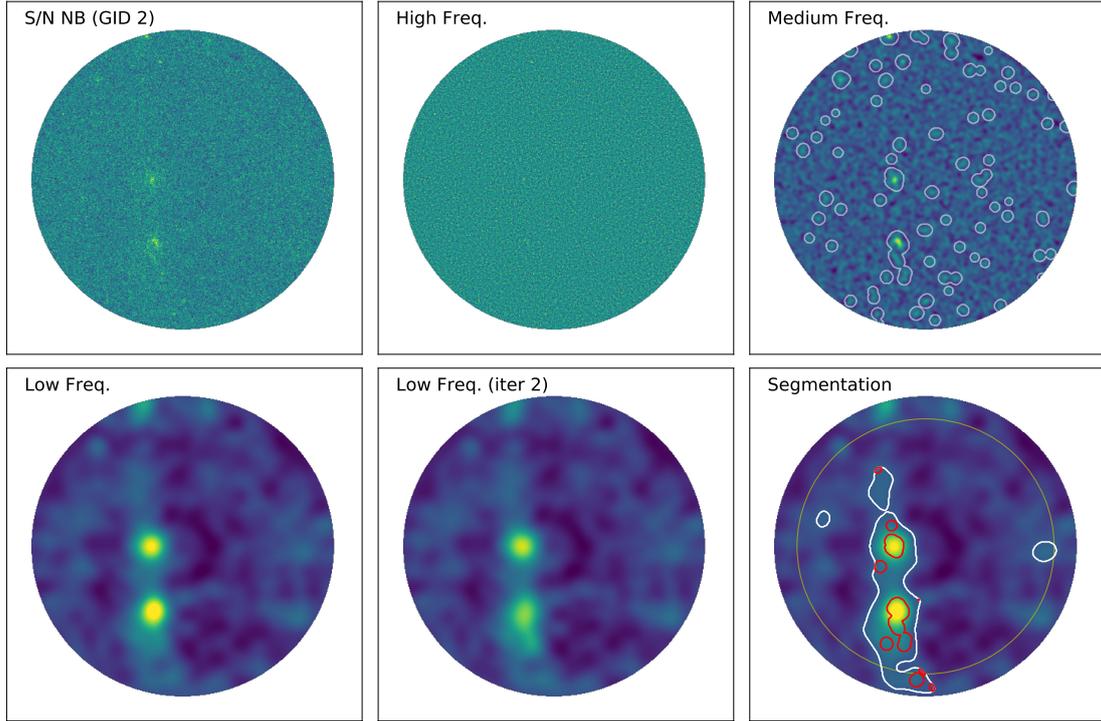
These segmentations obtained in the wavelet space are then used to compute, in the data space, the total flux of each area. The computation is performed on the original narrowband images, obtained from the ORIGIN PCA continuum-subtracted datacube.

#### 4.3. Noise and S/N estimation

Although the datacube noise properties have been carefully validated, including the impact of noise covariance, one cannot rely on formal noise propagation when working at such faint surface brightness especially after continuum subtraction processes, which may modify the noise distribution. Any remaining systematics like low background fluctuations can have a significant effect when summing the signal over a large area.

In order to estimate the standard deviation of our flux measurement, we have performed the following computation: We mask all spaxels contained within the filament segments or

<sup>8</sup> In IUWT the spatial extension of the filter grows as a power of 2 from one scale to another. Considering scales larger than 8 would lead to filters larger than the MXDF field.



**Fig. 7.** Example of IUWT wavelet multiscale decomposition of the group 2 ( $z = 3.07$ ) S/N narrowband image (*upper left panel*). The signal is split into three channels: high, medium, and low spatial frequency. *Central-lower panel*: (low freq. iter 2) shows the low frequency channel after masking the medium frequency peaks in the S/N image (see Sect. 4.2). The final segmentation is shown in the *lower right panel*. In white the “filaments”, in red the area of compact sources included in the filaments. The diffuse area is defined as the area included in the filament (white contours) after exclusion of the compact sources (red contours). The yellow ring shows the low S/N region with less than 60 h exposure time. The figure has the MXDF orientation with north up.

within the compact source segments identified in the first step (see upper and lower right panels of Fig. 7). Depending on the group this masks 10 to 30% of the total area. We start by computing the average offset and its standard deviation of the main unmasked area (1′ diameter or 120 h depth). This mean offset is subtracted from the flux image. The computed offsets are always small:  $-0.9 \pm 1.2 \times 10^{-20}$  erg s $^{-1}$  cm $^{-2}$  arcsec $^{-2}$  on average for all groups.

We then perform the following bootstrap experiment: we sum the flux of  $N$  spaxels selected by random permutation of the list of unmasked spaxels.  $N$  is the number of spaxels of the segmented area. At each iteration, we add a single offset flux value randomly drawn from a Normal distribution with zero mean and standard deviation equal to the previously computed offset error. This allows us to take into account the uncertainty of the subtracted offset value. We repeat this 1000 times and compute the standard deviation of the sum values. This gives the empirical noise estimate for the sum of  $N$  spaxels. We perform this experiment for 3 different values of  $N$ , corresponding to the number of spaxels of the segmented area covered by the entire segments, the compact sources and the diffuse area, respectively.

The standard deviation derived from the bootstrap experiment is then scaled to take into account the correlation of the noise introduced during the data reduction process (see Sect. 4.6 of Weibacher et al. 2020). This scale factor, estimated during the data reduction, is slowly evolving with wavelength, with values in the range 1.9–2.1. This results in our empirical standard deviation of the flux values for the filament, compact and extended segmented area.

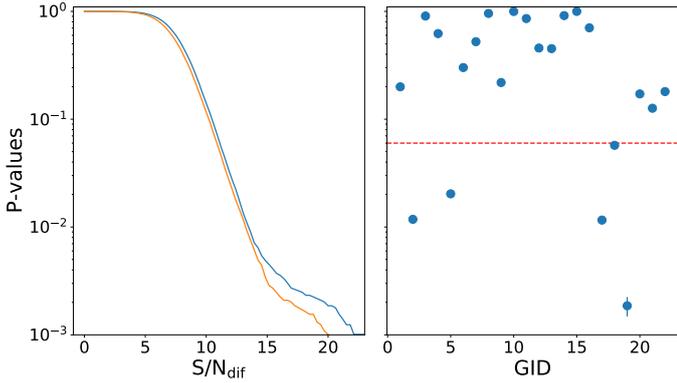
This empirical noise estimation is performed for each overdensity. We compare the computed values with the ones derived

from noise propagation, using the values given in the datacube (already corrected for noise correlation). We note that, even after taking the noise correlation into account, the propagated values underestimate the noise by a factor two ( $2 \pm 0.3$ ). This is likely due to small systematics left by the continuum subtraction. In the rest of the document the S/N is defined as the flux divided by the standard deviation of the noise estimated empirically as described above.

#### 4.4. Process validation

While the empirical error computed in the previous section gives an estimate of the expected error for the measured flux within a given diffuse emission segment identified by the algorithm, it says nothing about the probability of being fooled by noise, in other words, the probability of estimating a diffuse emission segment with the same or higher S/N when there is no diffuse emission. For each narrowband, the S/N value of the diffuse emission computed by the estimation algorithm can be used as a test statistic to discriminate between a null hypothesis (there is no diffuse emission segment) and the presence of diffuse emission. Owing to the nature of the data and to the various preprocessing leading to the computed S/N values, the distribution of these test statistics under the null hypothesis is unknown and in particular if this is not a Gaussian as we will see shortly. One must consequently resort to Monte Carlo simulations to evaluate this distribution in order to quantify the significance of our findings.

For that purpose we perform the following experiment. We select all wavelength slices free from sky pollution and outside the 22 groups’ wavelength regions. This effectively leaves 2582 wavelengths among the 3721 cube wavelengths. We then



**Fig. 8.** Estimated  $P$ -values for the diffuse component. *Left plot:*  $P$ -values as a function of the diffuse component’s S/N for the control samples (see Sect. 4.4). The samples with ten and 20 wavelengths are shown in, respectively, blue and orange colors. *Right plot:* estimated  $P$ -values for all overdensities. The dotted horizontal red line shows the 0.06  $P$ -value (see Sect. 4.5).

perform a random permutation in wavelength of this sample and split it into 258 groups of ten slices. The random permutation breaks the spectral continuity ensuring that no genuine extended structure at other redshifts than that of the overdensities can add coherently in the groups of shuffled slices. In contrast, we expect that the contribution of some diffuse emission in contiguous slices might, even if they are very faint in each slice, combine to something detectable in some of the 22 groups under examination. No spatial permutation is attempted because this would break the structure of numerous emission line sources that are present in all spectral regions of the datacube. These sources are also present in the groups under examination and even if the algorithm attempts to remove them in order to estimate diffuse emission, they do affect the algorithm’s results. The estimation algorithm (i.e., wavelet transformation, segmentation and empirical noise estimation) is then applied to each group and the resulting S/N of the diffuse emission segment are saved. We repeat this process 50 times in order to increase the size of the control sample to 12 900 groups. From the resulting statistical distribution of the S/N, one can estimate the  $P$ -value for a given overdensity ( $P_k$ ). The  $P$ -value  $P_k$ , associated with a S/N value  $S/N_k$ , is the probability of obtaining a S/N value equal or larger to  $S/N_k$  when there is no diffuse emission. We estimate the  $P$ -values by counting the empirical fraction of control groups with  $S/N > S/N_k$ , where  $S/N_k$  is the S/N of the overdensity under examination ( $k = 1, \dots, 22$ ). In practice, to take into account the increase in the wavelength range with redshift (from 9 to 18 slices), we repeat the experiment with a similar sample of 12 900 groups of 20 wavelength slices. The resulting  $P$ -values as a function of S/N are given in the left panel of Fig. 8.

Further validation of the method using independent data sets and simulations are presented in Sects. 4.6 and 5.2.2, respectively.

#### 4.5. Extended and diffuse Ly $\alpha$ emission in overdensities

The detection process described in Sect. 4.2 has been applied to the 22 groups. The resulting images are displayed in Fig. 9. As shown by this figure and the left panel of Fig. 10, extended emission<sup>9</sup> is found in most of the overdensities. If we arbitrarily

<sup>9</sup> The extended emission refers to the total flux within the filaments including the compact sources area, whereas the diffuse emission refers to the filament area after exclusion of compact sources.

set a lower limit of 300 arcsec<sup>2</sup> (or  $1.9 \times 10^4$  pkpc<sup>2</sup> at  $z = 3$ ) as the minimum area, we have 14 groups (63%) with “extended” emission. The most extended group is group 5 with an area of 1008 arcsec<sup>2</sup> or  $6 \times 10^4$  pkpc<sup>2</sup>, that is 20% of the total available MXDF area. We note that groups ten and 15 show almost no extended emission. As we will see later in Sect. 5.2.2, this is likely due to the size difference between the MOSAIC and MXDF fields of view.

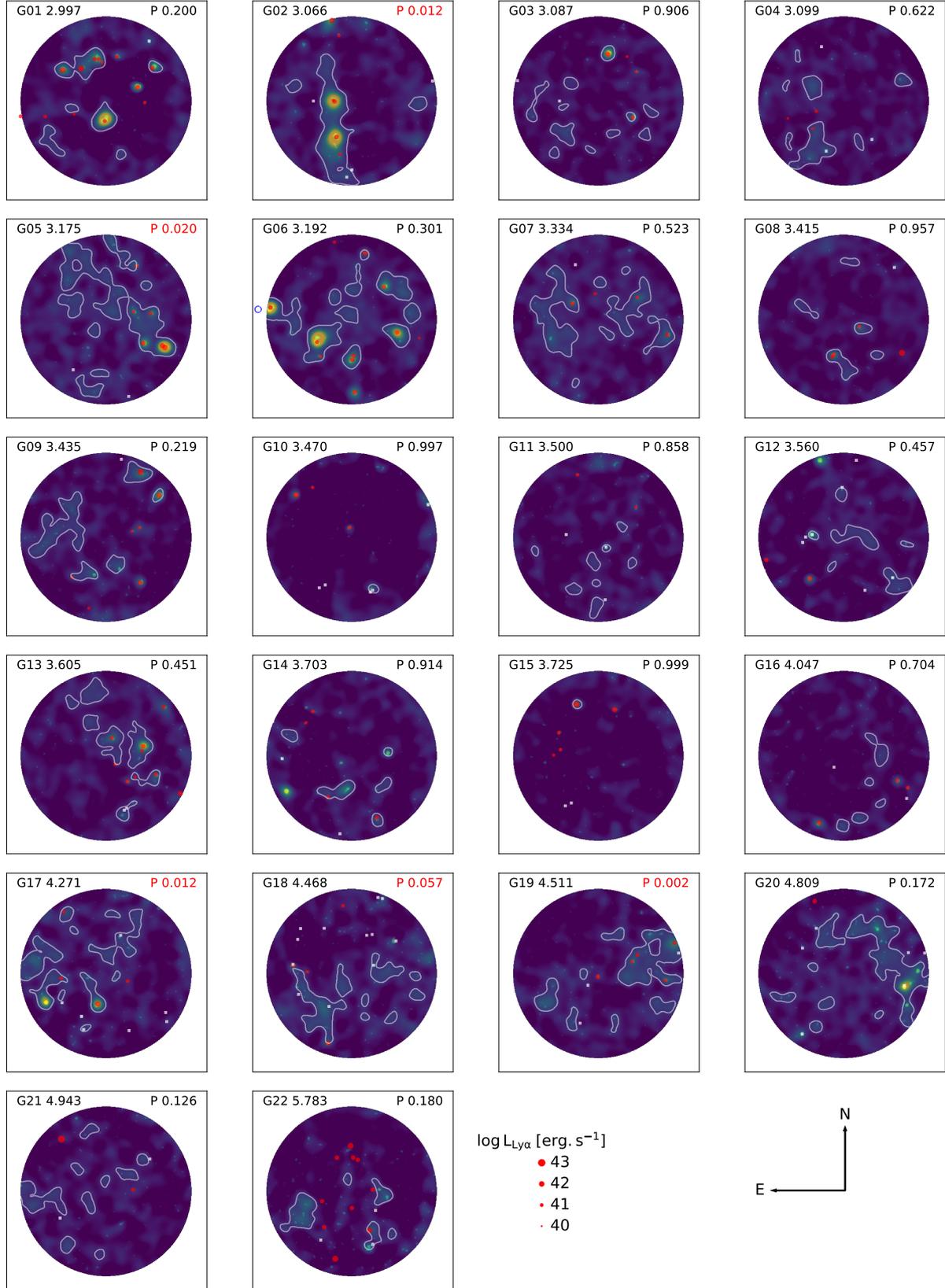
We now review the diffuse area identified within each filament (central panel of Fig. 10). To evaluate their statistical significance, we translate their S/N to a  $P$ -value by using the relation between  $P$ -value and S/N derived from the control sample (Fig. 8 left panel). For each overdensity we linearly interpolate the  $P$ -values at the corresponding number of wavelength slice of the group. The computed  $P$ -values are given in the right panel of Fig. 8. We split the groups in two subgroups: the high confidence overdensities (confidence 1) with diffuse emission signal detected with a  $P$ -value smaller than 0.06, and the rest of low confidence overdensities without statistically significant detection ( $P$ -value  $> 0.06$ , confidence 0). Among the 14 groups with extended Ly $\alpha$  emission, we have five groups with diffuse emission detected at high confidence: two at  $z \approx 3$  (groups 2 and 5) and three at higher redshifts (groups 17, 18, 19 at  $z \approx 4.5$ ). It is worth mentioning that the diffuse Ly $\alpha$  emission in the high redshift groups is detected around 7000 Å, at the peak of MUSE sensitivity where the noise is minimal (column SB of Table C.1).

It should also be pointed out that group 6, which is globally the brightest in Ly $\alpha$  flux and one of the most extended (left panel of Fig. 10), is classified as low confidence ( $P = 0.3$ ) for diffuse Ly $\alpha$  emission. The reason is that the flux and area attributed to compact LAEs prevail over its diffuse emission. The S/N for the diffuse area flux (central panel of Fig. 10) is then too low to pass the  $P$ -value threshold ( $P < 0.06$ ). For similar reasons, groups 1, 7 and 20 were ranked as low confidence despite their total flux brightness and area.

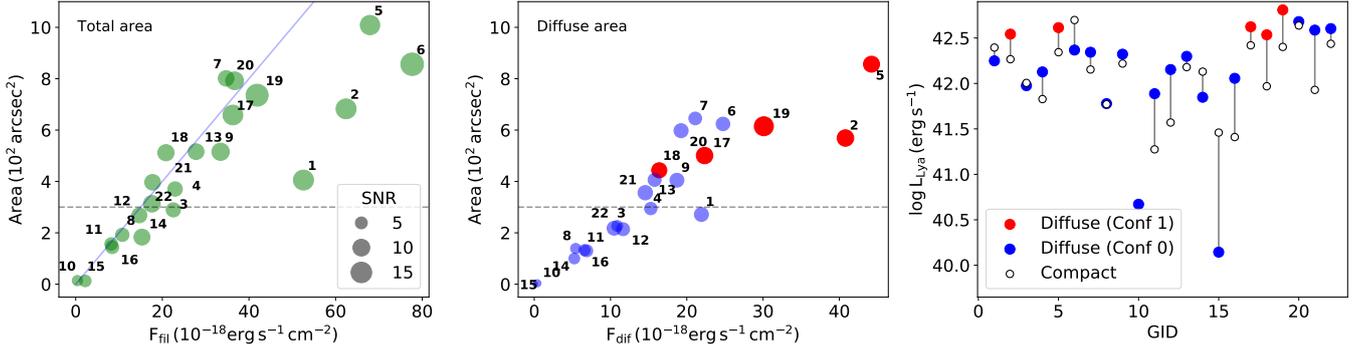
The average surface brightness of the diffuse emission in the high confidence groups is  $5.1 \pm 1.2 \times 10^{-20}$  erg s<sup>-1</sup> cm<sup>-2</sup> arcsec<sup>-2</sup>. A statistical overview of the parameters of the high confidence filaments is given in Table 1 while the details for each overdensity are given in Table C.2.

We note that, while the algorithm results in the detection of extended and diffuse emission with a given confidence, it does not tell us if this extended emission is indeed Ly $\alpha$  emission. There are other possible sources of diffuse emission in the datacube due to other emission lines such as MgII (Burchett et al. 2020; Wizotski, in prep.; Leclercq, in prep.), [OIII] (Johnson et al. 2018) or even galactic diffuse emission (e.g., H $\alpha$ , [SII]). However, the former are systematically associated with bright continuum galaxies that are not present in the identified overdensities and the latter are at known wavelengths and can thus be easily excluded. In addition, the fact that the identified diffuse emission overlaps spatially with the known Ly $\alpha$  emitters found independently during the construction of the catalog (Sect. 2.4), suggest that this diffuse emission originates from Ly $\alpha$  emission.

The boundary between compact and extended sources is based on the considered multiscale analysis. An important question is to what extent the segmentation performed on compact sources is representative of the galaxy itself only, or of the galaxy plus its surrounding CGM. In other words, we want to determine what fraction of the filament flux can be explained by the CGM of the identified galaxies. In a few cases, for example group 5, the question is not relevant given that a high fraction of the filaments have no detected LAEs. But in some cases, such as group 2, there are two bright LAEs in the main filament that might bias



**Fig. 9.** Searching for extended Ly $\alpha$  emission in groups within the MXDF field. The images are composites of low and mid-frequencies IUWT wavelets S/N components. The solid contours display the segmentation corresponding to the identified structures (named filament in the text). LAE group members and potential low- $z$  interlopers are shown, respectively, as red and white symbols. The group number, its corresponding redshift and the diffuse emission  $P$ -value ( $P$ ) (see Sect. 4.4) are labeled. The  $P$ -value is written in red for groups with confidence 1. The blue circle symbol in group 6 indicates the location of the AGN ID 788 in the *Chandra* catalog (Luo et al. 2017). The MXDF field shown here corresponds to a depth of 16+ h. It has a diameter of 80'' or 2.5 cMpc at  $z = 3$ .



**Fig. 10.** Filament properties. *Left panel:* angular area covered by the entire segments (filaments) as a function of the corresponding Ly $\alpha$  flux. The group IDs are labeled and displayed as green filled circles with radii proportional to the total flux S/N. The dotted horizontal line and the blue diagonal line show respectively the 300 arcsec<sup>2</sup> area limit and a constant surface brightness of  $5 \times 10^{-20}$  erg s<sup>-1</sup> cm<sup>-2</sup> arcsec<sup>-2</sup>. *Central panel:* same as *left panel* for the diffuse Ly $\alpha$  emission (diffuse area). The filled circles are colored according to the detection confidence in blue and red for low and high confidence, respectively (see Sect. 4.5). *Right panel:* total luminosity of diffuse Ly $\alpha$  in filaments (filled circle symbols in red for high confidence, in blue for low confidence) compared with the luminosity of the compact Ly $\alpha$  emission (black open circles) within the same filaments. We note that group 10 has no compact source detected within the filament area.

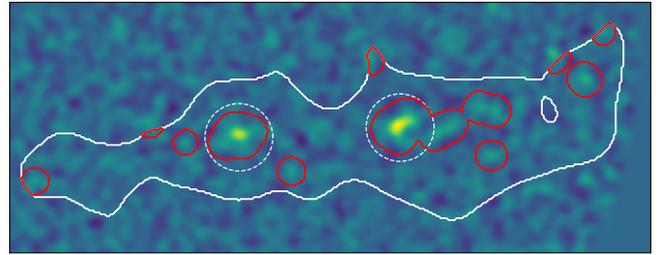
**Table 1.** Measured physical parameters for the five high confidence filaments (average values).

Name	Mean	Std	Min	Max
$F_{\text{fil}}$	45.9	17.3	20.8	67.9
$F_{\text{comp}}$	15.1	7.0	4.5	23.7
$F_{\text{dif}}$	30.8	10.6	16.4	44.2
$S/N_{\text{fil}}$	13.5	2.5	9.4	17.1
$S/N_{\text{comp}}$	10.9	2.5	5.9	12.5
$S/N_{\text{dif}}$	9.9	1.6	8.1	12.8
$\log L_{\text{fil}}$	42.8	0.1	42.6	43.0
$\log L_{\text{comp}}$	42.3	0.2	42.0	42.4
$\log L_{\text{dif}}$	42.6	0.1	42.5	42.8
$S_{\text{fil}}$	719.6	162.6	511.9	1008.8
$S_{\text{comp}}$	122.6	32.1	68.4	157.9
$S_{\text{dif}}$	597.0	142.3	443.5	856.5
$SB_{\text{dif}}$	5.1	1.2	3.7	7.2
$F_{\text{dif}}/F_{\text{fil}}$	68.5	6.1	61.5	78.6

**Notes.** Mean, standard deviation, min, and max values are given for the following three components (*c*): fil = full filament, comp = compact area within the filament, dif = diffuse area within the filament.  $F_c$ : Ly $\alpha$  flux in  $10^{-18}$  erg s<sup>-1</sup> cm<sup>-2</sup> for the component *c*,  $S/N_c$ : Flux S/N for the component *c*,  $\log L_c$ : Log of Ly $\alpha$  luminosity in erg s<sup>-1</sup> for the component *c*,  $S_c$ : Surface in arcsec<sup>2</sup> for the component *c*,  $SB_{\text{dif}}$ : Average surface brightness of the diffuse Ly $\alpha$  emission in  $10^{-20}$  erg s<sup>-1</sup> cm<sup>-2</sup> arcsec<sup>-2</sup>.  $F_{\text{dif}}/F_{\text{fil}}$ : Fraction of Ly $\alpha$  flux in the diffuse area (in %). For individual values by overdensities see Table C.2.

the filament flux measurement. In Fig. 11 we show that the area of group 2 covered by compact source segments extends up to 27 pkpc from the galaxy center. At such a distance, a very high fraction of the galaxy CGM is already included (see Fig. 15 of Leclercq et al. 2017) and the average surface brightness of the Ly $\alpha$  halos falls below  $10^{-20}$  erg s<sup>-1</sup> cm<sup>-2</sup> arcsec<sup>-2</sup> (see Fig. 2 of Wisotzki et al. 2018). We then expect the CGM of these Ly $\alpha$  emitters to have only a small contribution to the measured Ly $\alpha$  emission flux in the diffuse area.

In addition, as shown in Fig. 11, the compact source segmentation map (red contours) is not restricted to the detected LAEs, but covers all medium frequency peaks identified in the NB image. We also note that the flux of these compact sources is



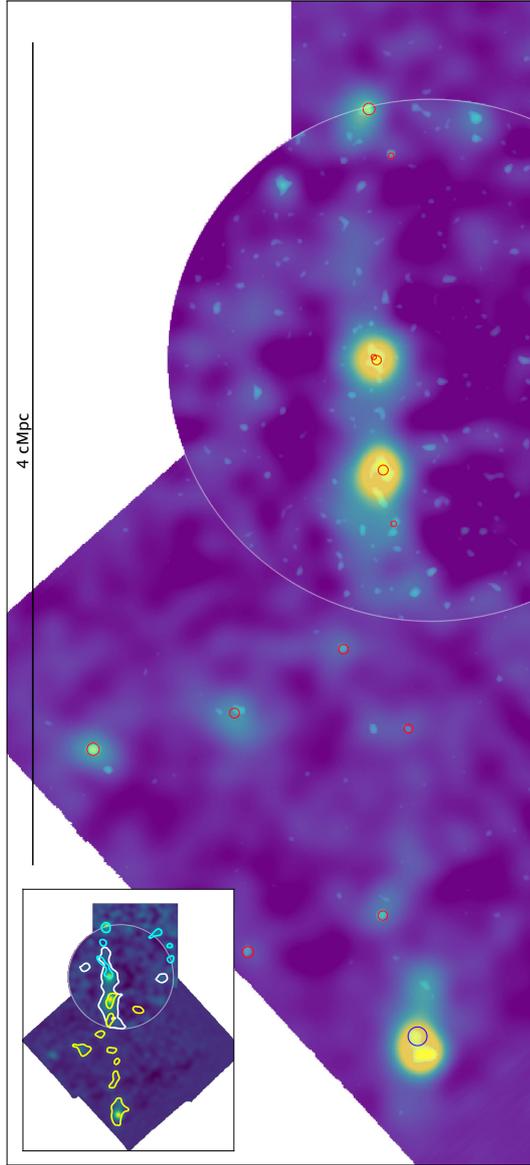
**Fig. 11.** Medium frequency narrowband image of the main group 2 filament (rotated by 90° counterclockwise). The compact source segmentation contours are shown in red. The two white dashed circles display the 3.5'' radius area (27 pkpc) where the diffuse emission from the CGM of the two LAEs (bright yellow clumps) identified in the catalog becomes insignificant.

on average three times fainter than the Ly $\alpha$  flux of the detected LAEs (Fig. 13). We conclude that our extended diffuse emission flux is a conservative lower limit on the Ly $\alpha$  emission shining outside the CGM of identified LAEs.

For each group we compute the total Ly $\alpha$  luminosity in the diffuse area and compare it to the flux measured in the compact structure area (right panel of Fig. 10, Table C.2). For most of the groups, the measured Ly $\alpha$  luminosity in the diffuse area is significantly larger than in the compact source area. This is true in particular for the five high confidence groups. As seen above, if we assume that the compact source flux is a good proxy of the total Ly $\alpha$  emission of identified galaxies, including their CGM emission, we conclude that a high fraction of the total Ly $\alpha$  emission measured in the filaments is coming from either undetected faint Ly $\alpha$  emitters and their CGM and/or intrinsic diffuse Ly $\alpha$  emission. For the high confidence groups the average log Ly $\alpha$  luminosity (erg s<sup>-1</sup>) in the diffuse area amounts to  $42.6 \pm 0.1$  and represents between 61% and 79% (average 68%) of the total filament Ly $\alpha$  luminosity.

#### 4.6. The case of the $z = 3.07$ overdensity

Although we did not expect to detect diffuse Ly $\alpha$  emission other than the Ly $\alpha$  halos in the MOSAIC and UDF-10 exposures with 10 and 30 h depth, respectively, we nevertheless check for possible exceptions in the high confidence groups. The estimation



**Fig. 12.** Extended diffuse Ly $\alpha$  emission in the  $z = 3.07$  overdensity (GID 2). Composite IUWT-S/N image formed with part of the MOSAIC 10 h depth exposure (bottom rotated square field), the UDF10 30 h depth exposure (small upper square) and the MXDF 140 h depth exposure (circle). The locations of the identified Ly $\alpha$  emitters are displayed as red circles with size proportional to the log of their luminosity. The blue circle at the bottom shows the location of the AGN and Ly $\alpha$  emitter blob. The 4 cMpc (1 pMpc) scale is shown. The inset displays the contours of the filaments in yellow, white and cyan for, respectively, the MOSAIC (only a fraction of the full field), MXDF and UDF-10 exposures.

algorithm shows no convincing emission in all groups with the notable exception of group 2, which displays a clear signal extending the main MXDF filament to the south. Figure 12 displays the complete structure which now extends to a total length of 4.6 cMpc (1.1 pMpc) with a width of 191 ckpc (47 pkpc).

As shown in the inset of Fig. 12, the filament is detected in each of the three independent exposures. This observational evidence constitutes an additional process validation and reinforces our confidence that the structure is real. The MOSAIC and UDF-10 exposures show a second filament crossing the main one at 45°.

A bright patch of Ly $\alpha$  emission is seen at the south end of the filament (bottom of Fig. 12). The source of Ly $\alpha$  emission coincides with a luminous, QSO-level type II AGN identified as ID 746 in the *Chandra* 7 Ms catalog (Luo et al. 2017), hereafter referred to as CID 746. The extended Ly $\alpha$  emission surrounding the AGN was already reported by den Brok et al. (2020) in their study of extended Ly $\alpha$  emission around type II AGN. The authors trace extended Ly $\alpha$  out to 80 pkpc from the AGN (see their Fig. 1). We note that their analysis is derived from the same data set (i.e., the MOSAIC MUSE datacube), but with a different post-processing. Our analysis confirms this extension and adds the finding of fainter diffuse Ly $\alpha$  emission at much larger distance. We consider the possible importance of this AGN in Sect. 5.1.2 below.

Figure 12 is also illustrative of the density increase of compact sources seen at the MXDF depth with respect to the MOSAIC field. Of course we are not able to prove that all these compact sources are indeed faint Ly $\alpha$  emitters, and some of them are possibly noise peaks, but the fact that a low fraction of them were independently identified as Ly $\alpha$  emitters during the construction of the catalog suggests that the filaments are populated by a high number of faint Ly $\alpha$  emitters. This property is not specific to group 2, but can be seen in the other groups as well (Fig. 9).

## 5. Analysis and discussion

Part of the Ly $\alpha$  emission that we detect overall undoubtedly comes from galaxies and their CGM. Emission from the CGM may be powered by a variety of processes, probably combined: dissipation of gravitational energy through cooling radiation, fluorescence from the UV background or from ionizing radiation emitted by local sources, scattering of galactic Ly $\alpha$  through neutral gas in the CGM, extended star formation, or undetected satellite galaxies. Regardless the origin of this emission, we have shown that our analysis separates this signal from the diffuse emission (see for example Fig. 11). It is the origin of this diffuse component, which represents the largest fraction (70%) of the flux, that we try to explain in the present section. Possible sources for this emission are: hydrogen gas heated and ionized by the external UV background, undetected faint Ly $\alpha$  emitters, or gravitational compression. We explore these options in the following sections.

### 5.1. Ly $\alpha$ fluorescence

#### 5.1.1. Ly $\alpha$ fluorescence by the cosmic UV background

Optically thick clouds illuminated by a diffuse intergalactic UV radiation field will emit Ly $\alpha$  emission by the fluorescent conversion of H-ionizing photons into Ly $\alpha$  (Hogan & Weymann 1987; Gould & Weinberg 1996). A ubiquitous source of such photons is the cosmic UV background (UVB), namely the integrated UV emission from AGN and star-forming galaxies (e.g., Haardt & Madau 1996, 2012).

Here we follow the recipes from Cantalupo et al. (2005) to estimate the expected Ly $\alpha$  surface brightness due to UVB fluorescence, in a similar way as in Gallego et al. (2018). We adopt the photoionization rate of neutral hydrogen  $\Gamma_{\text{HI}}^{\text{HM}}$  and its redshift evolution predicted by the synthesis model of Haardt & Madau (2012) as a baseline value. To assess the uncertainties we also consider the more recent model calculations by Faucher-Giguère (2020) ( $\Gamma_{\text{HI}}^{\text{FG}}$ ), as well as the empirical estimates by Becker & Bolton (2013) ( $\Gamma_{\text{HI}}^{\text{BB}}$ ). We emphasize that the

**Table 2.** Estimation of maximum fraction of Ly $\alpha$  surface brightness due to the Cosmic UV background in the five groups with high confidence diffuse emission.

ID	$z$	SB <sub>dif</sub>	$\Gamma_{\text{HI}}^{\text{HM}}$	$\Gamma_{\text{HI}}^{\text{FG}}$	$\Gamma_{\text{HI}}^{\text{BB}}$	SB <sub>uv</sub> <sup>BB</sup>	SB <sub>uv</sub> <sup>HM</sup>	$R$
2	3.07	7.17	8.08	8.96	$8.12 \pm 0.68$	$2.03 \pm 0.17$	2.02	28
5	3.18	5.16	7.75	8.64	$7.93 \pm 0.66$	$1.78 \pm 0.15$	1.74	34
17	4.27	4.46	5.19	6.17	$9.22 \pm 1.05$	$0.82 \pm 0.09$	0.46	10
18	4.47	3.69	4.89	5.80	$9.53 \pm 1.22$	$0.73 \pm 0.09$	0.37	10
19	4.51	4.90	4.84	5.73	$9.50 \pm 1.26$	$0.70 \pm 0.09$	0.36	7

**Notes.** ID: Group ID.  $z$ : redshift. SB<sub>dif</sub>: Average surface brightness of the diffuse Ly $\alpha$  emission in units of  $10^{-20}$  erg s $^{-1}$  cm $^{-2}$  arcsec $^{-2}$ .  $\Gamma_{\text{HI}}^{\text{src}}$ : Photoionization rate of neutral hydrogen from *src*, in units of  $10^{-13}$  s $^{-1}$ ; *src* takes the following values: HM (Haardt & Madau 2012), FG (Faucher-Giguère 2020), BB (Becker & Bolton 2013) SB<sub>uv</sub><sup>src</sup>: Expected Ly $\alpha$  surface brightness from the UV background in units of  $10^{-20}$  erg s $^{-1}$  cm $^{-2}$  arcsec $^{-2}$ .  $R$ : relative fraction SB<sub>uv</sub><sup>HM</sup>/SB<sub>dif</sub> in %.

Cantalupo et al. (2005) recipes consider Ly $\alpha$  photons produced in Lyman limit systems, that is, in optically thick clouds with HI column densities above  $10^{17.2}$  cm $^{-2}$  with a covering fraction of 1.

As shown in Table 2, the fluorescent Ly $\alpha$  emission powered by the UVB cannot explain the totality of the observed surface brightness, but its contribution is not negligible at  $z \approx 3$ , with  $\approx 30\%$  of the total surface brightness. At higher  $z$ , the contribution falls to  $\approx 10\%$  (or 20% if we use the Becker & Bolton 2013 photoionization rate). It is important to stress that these values are upper limits as they assume a covering fraction of 1 for the Lyman-limit systems.

We note, however, that the photoionization rates given in Table 2 are average values. Within the overdensities, given the higher number of Ly $\alpha$  emitters, one might expect the local ionising radiation field to be enhanced relative to the background value. However, because the mean free path for ionizing photons is much greater than the size of the overdense region, the typical local enhancement of the ionizing flux is expected to be much smaller than the local overdensity of sources of ionizing radiation (e.g., Schaye 2006). The cosmological radiative transfer simulations performed by Rahmati et al. (2013) confirm that local sources only dominate over the UVB on very small scales. Their Fig. 4 (left panel) displays the various contributions to the photoionization rate as a function of distance from the centers of  $10^{10.5}$ – $10^{11}$   $M_{\odot}$  dark matter halos at  $z = 3$ . One can observe that the contribution of local stellar radiation ( $\Gamma_{\text{HI}}^{\text{LSR}}$ ) is already a factor of five below  $\Gamma_{\text{HI}}^{\text{UVB}}$  at the virial radius and then drops to negligible values at larger distances. Given that the measured diffuse Ly $\alpha$  emission excludes by construction most of the local source’s CGM (Sect. 4.5), one can then safely ignore the contribution of local ionizing radiation to the Ly $\alpha$  fluorescence in the diffuse emission area.

### 5.1.2. Ly $\alpha$ fluorescence from active galactic nuclei

In the vicinity of a luminous AGN, the intense UV radiation can boost the Ly $\alpha$  emission by large factors (Haيمان & Rees 2001; Cantalupo et al. 2005; Kollmeier et al. 2010). Two high-redshift AGN are known within the MOSAIC field of view, both of which are outside the MXDF but match the redshifts of two of our groups. The spectacular case of group 2 at  $z = 3.07$  was already mentioned in Sect. 4.6; the other AGN is associated with group 6 at  $z = 3.19$ , immediately adjacent to the MXDF footprint. We now consider the importance of AGN for explaining the observed diffuse Ly $\alpha$  emission in the MXDF, beginning with group 2.

While an AGN as luminous as CID 746 is intrinsically a copious producer of UV radiation, the object is classified as a highly obscured type 2 AGN. In order to produce any fluorescent

emission in the MXDF filament the obscuration must be negligible in the transverse direction toward the MXDF filament, which (in line with den Brok et al. 2020) we assume in the following. We consider two simplified scenarios:

If the neutral hydrogen column density is high enough to make the MXDF filament fully self-shielded and if we at the same time assume negligible absorption between the QSO and the filament, its Ly $\alpha$  surface brightness will scale directly with the geometrically diluted intensity of the incident UV radiation. We can then estimate a boost factor, defined by Cantalupo et al. (2005) as the enhancement of UV illumination due to the AGN and quantified in their Eq. (14). The northern extension of the Ly $\alpha$  nebula at 80 pkpc from the AGN has a surface brightness of  $1 \times 10^{-18}$  erg s $^{-1}$  cm $^{-2}$  arcsec $^{-2}$ , which translates into a boost factor of 50 if we adopt a baseline surface brightness of  $2 \times 10^{-20}$  erg s $^{-1}$  cm $^{-2}$  arcsec $^{-2}$  for fluorescence at an optically thick HI cloud due to the UVB alone (see Table 2 in Sect. 5.1.1). This boost factor decreases with the square of the distance to the AGN, and at the MXDF location in 700 pkpc its value becomes  $1 + 50 \times (80/700)^2 \approx 1.6$ . We note that this is an upper limit given that the 3D distance is likely to be larger than the projected separation. In this scenario, the extra diffuse Ly $\alpha$  emission within the MXDF region caused by the AGN would fall short of reproducing the measured value by a factor  $\gtrsim 4$  (see Table C.2).

If on the other hand the hydrogen within the megaparsec-scale surroundings of the AGN including the MXDF region is optically thin to ionizing radiation, it is less straightforward to estimate the emergent Ly $\alpha$  surface brightness. Judging by the substantial X-ray luminosity of the AGN, the near-zone around the AGN should be much larger than the  $\sim 1$  pMpc distance relevant here (Bolton & Haehnelt 2007; Eilers et al. 2017), implying that the Ly $\alpha$  surface brightness enhancement is dominated by radiative recombinations and scales with the square of the local density. In this case it would indeed be possible that the AGN provides most or all of the additional photons needed to match the observed Ly $\alpha$  emission. Since we have no way to distinguish between these scenarios, we must leave the final judgement open for this particular group.

In case of group 6, the AGN is located so close ( $\approx 5$  arcsec) to the region of diffuse emission that, although the AGN is much fainter, the emission can be explained with both scenarios. We note, however, that this group was ranked as low confidence for diffuse emission (see Sect. 4.5). In addition, one can observe in Fig. 9 that the diffuse emission is not specifically strong at the immediate vicinity of the AGN.

In the other overdensities with extended emission we did not find any evidence for any nearby AGN, neither based on characteristic features in the MUSE spectra nor from the Chandra

catalog (Luo et al. 2017). Given the extraordinary depth of this catalog, this implies that we can confidently exclude any further strong AGN in this region. Low-luminosity AGN below the *Chandra* flux limit or very highly obscured objects – possibly even Compton-thick ones – are still possible, but these would not be significant contributors of escaping UV photons capable of powering the observed diffuse Ly $\alpha$  emission. Overall we conclude that AGN could boost extended Ly $\alpha$  emission in two of our overdensities, but that they are unlikely to be important in the majority of cases.

Finally, we note that AGN are expected to flicker on timescales of  $\sim 10^5$  yr (e.g., Schawinski et al. 2015), in which case the presence or absence of an AGN may not be correlated with AGN-induced Ly $\alpha$  emission at distance  $> 100$  pkpc.

## 5.2. Contribution of undetected Ly $\alpha$ emitters to the diffuse Ly $\alpha$ emission

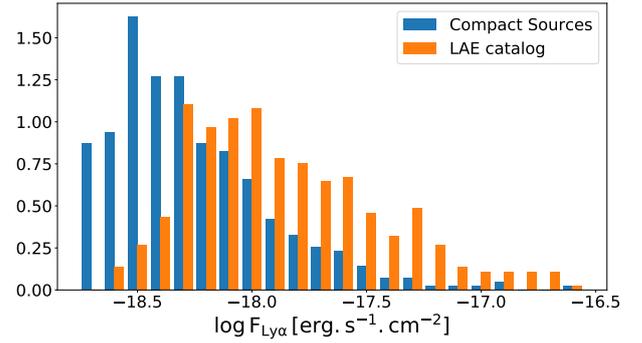
In spite of the depth of the MXDF observations, we expect a significant fraction of LAEs to fall beyond our detection limit. It is then indisputable that at least a fraction of the observed diffuse Ly $\alpha$  flux is coming from these undetected LAEs. We now investigate if these galaxies can be responsible for the totality of the measured flux. In Sect. 5.2.1, we first address this question with simple modeling based on the luminosity function of LAEs. In Sect. 5.2.2, we then use a more sophisticated approach based on the semi-analytic model GALICS (Garel et al. 2012, 2015a).

### 5.2.1. Luminosity function toy model

The luminosity function (LF) of LAEs is not observed to evolve strongly between  $z=3$  and 6 (e.g., Ouchi et al. 2008; Cassata et al. 2011; Herenz et al. 2019), and it is well described by a Schechter (1976) function  $\phi(L)dL = \phi_*(L/L_*)^\alpha \exp(-L/L_*)dL/L_*$ . The faint end behavior of the LF is given by the parameter  $\alpha$ . Thanks to MUSE observations which have provided a large sample of faint Ly $\alpha$  emitters (e.g., Bacon et al. 2017; Urrutia et al. 2019; Richard et al. 2021), we now have better constraints on  $\alpha$  down to  $\log(L_{\text{Ly}\alpha}/\text{erg s}^{-1}) = 41.5$  in deep fields and 40.5 in lensing cluster fields. Herenz et al. (2019) found  $\alpha = -1.84^{+0.42}_{-0.41}$  with no evolution with redshift, Drake et al. (2017) measured  $\alpha = -2.03^{+1.42}_{-0.07}$  at  $z=3.5$  and  $\alpha = -2.86^{+0.78}_{-\infty}$  at  $z=5.5$  and de La Vieuville et al. (2019), using MUSE lensing cluster observations, found  $\alpha = -1.63^{+0.13}_{-0.12}$  at  $z=3.5$  and a similar value at  $z=4.5$ . In the following we have selected the Herenz et al. (2019) intermediate values with  $\log(\phi_*/\text{cMpc}^{-3}) \approx -2.71$ ,  $\alpha \approx -1.84$ , and  $\log(L_*/\text{erg s}^{-1}) \approx 42.6$  as our fiducial model, but we will also explore later different values of  $\alpha$ . Assuming to first order that the amplitude of the LF changes with environment but not its shape, we may write the mean surface brightness (SB) contributed by undetected LAEs as:

$$\text{SB}_{\text{LF}}(<L_{\text{det}}) = \Delta l \times \zeta \times \int_0^{L_{\text{det}}} \delta \times L \times \phi(L)dL, \quad (1)$$

where  $\phi$  is the field LF,  $\delta$  is the over-density measured in each group (Table C.1), and the integral runs over luminosities below a luminosity  $L_{\text{det}}$  corresponding to a flux limit of  $10^{-18.5} \text{ erg s}^{-1} \text{ cm}^{-2}$  at the redshift of each group. This flux limit is derived as the peak of the Ly $\alpha$  flux distribution (Fig. 13) of compact sources identified by the detection algorithm (Sect. 4.2).  $\Delta l = 8 \text{ cMpc}$  is the depth of the narrowband slice, and  $\zeta \geq 1$  is a factor that accounts for angular clustering. We define  $\zeta$  as the



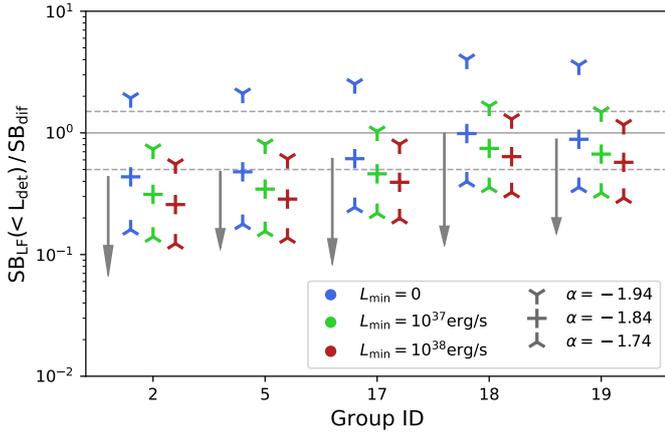
**Fig. 13.** Comparison of the Ly $\alpha$  flux probability density for the compact sources detected by the algorithm (Sect. 4.2) and the MXDF LAE catalog.

inverse of the filling factor of LAEs in a field:  $\zeta = S_{\text{MXDF}}/S_{\text{LAE}}$  where  $S_{\text{MXDF}} \approx 3632 \text{ arcsec}^2$  is the projected area of the field and  $S_{\text{LAE}}$  is the area occupied by diffuse emission from unresolved LAEs. A value  $\zeta = 1$  corresponds to sources uniformly distributed, and a higher value means a stronger angular clustering of sources. At the scales involved in the present study ( $< 1$  arcmin), the angular clustering of LAEs is not constrained. Moreover, clustering constraints are statistical by nature and would not allow us to predict a value of  $\zeta$  group by group. We thus estimate a value of  $\zeta$  for each group as  $\zeta = S_{\text{MXDF}}/S_{\text{Fil}}$ , where  $S_{\text{Fil}}$  is taken from Table 1. This assumes that all undetected galaxies are distributed within the observed diffuse emission mask. This is probably an upper estimate of  $\zeta$  as one expects fainter sources to cluster less strongly (e.g., Mo et al. 1998) and hence to occupy a larger area. The values of  $\zeta$  we obtain in this way range from 4 to 8 depending on the group, with a mean of 6.4.

The results of this simple model, with the parameters set to the values discussed above, are shown in Fig. 14 with the blue plusses. This figure shows the ratio of the expected Ly $\alpha$  surface brightness due to undetected galaxies to the one measured in the diffuse component ( $\text{SB}_{\text{diff}}$  from Table C.2) for each high confidence group. We find that undetected galaxies account for about 40% of the signal in groups 2 and 5, about 55% for group 17, and 80–90% for groups 18 and 19. Thus, for most groups studied here, the contribution of undetected LAEs is likely dominant. If we add the contribution from fluorescence of the UVB (Table 2), we can explain about 70% of the signal for groups 2, 5 and 17 and 90–100% for groups 18, 19.

Of course these estimates are crude at best and very uncertain due to many poorly constrained parameters. The steepness of the faint end of the Ly $\alpha$  LF, in particular, is the parameter that has the largest impact<sup>10</sup>. Assuming a steep, yet reasonable, slope  $\alpha = -1.94$ , we find that the expected SB due to undetected LAEs actually overshoots our observations by factors  $\approx 2$ –3 (top blue symbols in Fig. 14). From Eq. (1), it is clear that this tension may easily be resolved by reducing the product  $\zeta \times \delta$  by a factor of three. As mentioned above, our fiducial value of  $\zeta$  is an upper limit and the true value may be anywhere between 1 and our estimate, either due to small-scale clustering variance or to projection effects. The amplitude of the multiplicative  $\zeta$

<sup>10</sup> For simplicity, in the analysis presented here, we only vary  $\alpha$  and keep the other two parameters of the LF fixed. We have verified that our conclusions are unchanged when changing  $\phi_*$  and  $L_*$  consistently with  $\alpha$  to take into account the correlation between these three parameters shown by Herenz et al. (2019, their Fig. 12). For the steep limit, our results are barely distinguishable, and for the shallow limit, changing  $\phi_*$  and  $L_*$  would increase our model's SB by about 25%.



**Fig. 14.** Ratio of the expected SB due to undetected LAEs to the mean SB of detected diffuse emission, as a function of group ID for the high confidence groups. The blue pluses show the prediction for a fiducial LF with a faint-end slope of  $\alpha = -1.84$  (Herenz et al. 2019) extrapolated to zero luminosity. The other symbols show the predictions for steeper ( $\alpha = -1.94$ ) or shallower ( $\alpha = -1.74$ ) slopes as indicated on the plot. The red and green symbols show the result of the model when using a low-luminosity cut ( $L_{\min}$ ) of respectively  $10^{38}$  and  $10^{37}$  erg s $^{-1}$ . The arrows show the amplitude of the multiplicative  $\zeta$  correction that was included (see Sect. 5.2.1). The horizontal dotted lines display the 0.5 and 1.5 values.

correction is shown as an arrow in Fig. 14 for each group. The value of  $\delta$ , in turn, is evaluated on the full extent of the MOSAIC field, and it is not obvious that the same over-density would be measured within the smaller MXDF area. It is interesting to note that a shallower slope of the LF, say  $\alpha = -1.74$ , produces at best 30% of the observed diffuse emission (lower blue symbols on Fig. 14). This would imply that the diffuse emission we observe is for the most part not due to stellar irradiation. The precise value of the faint-end slope of the Ly $\alpha$  LF is thus crucial to understand our observations. Its uncertainties are still large, although a steep slope appears more likely (Herenz et al. 2019).

Given the potential role of faint LAEs in producing the diffuse emission we detect, it is interesting to understand down to which luminosities the contribution of these LAEs matters. This depends on the slope of the LF: the steeper the LF the more important the contribution of very faint objects. As an example, for the slopes  $\alpha = -1.74$ ,  $-1.84$ , and  $-1.94$ , 50% of the luminosity below  $L_{\text{det}}$  is contributed by sources fainter than  $0.07 \times L_{\text{det}}$ ,  $0.01 \times L_{\text{det}}$ , and  $10^{-5} \times L_{\text{det}}$ , respectively. At the redshifts of our groups,  $L_{\text{det}}$  is in the range  $\approx 2-7 \times 10^{40}$  erg s $^{-1}$ , and so half of the light from undetected LAEs is contributed by sources brighter than  $\approx 1-5 \times 10^{39}$ ,  $2-7 \times 10^{38}$ , and  $2-7 \times 10^{35}$  erg s $^{-1}$  for the three slopes, respectively. There are no observational constraints on the Ly $\alpha$  LF at such faint luminosities.

The deepest UV LFs at  $z \approx 3-6$  are consistent with a power-law extending to magnitudes as faint as  $M_{\text{AB}} \approx -13$  (Alavi et al. 2016; Bouwens et al. 2017; Livermore et al. 2017; Atek et al. 2018). Making the crude approximation that Ly $\alpha$  and UV emission are linearly related to the star formation rate and neglecting the effect of dust<sup>11</sup>, this value corresponds to a Ly $\alpha$  luminosity slightly fainter than  $10^{40}$  erg s $^{-1}$  (e.g., Garel et al. 2015a). With the same assumptions, a Ly $\alpha$  luminosity of  $L_{\text{Ly}\alpha} \approx 10^{38}$  erg s $^{-1}$

corresponds to a UV magnitude as faint as  $M_{\text{AB}} \approx -8$ , or equivalently to a star formation rate (SFR) of approximately  $10^{-4} M_{\odot} \text{ yr}^{-1}$ . We note that systems with such a low SFR are observed in the local Universe and that the UV LF at  $z = 0$  keeps rising (at least) down to this level (Bothwell et al. 2011). Extrapolating these local constraints at high redshift is difficult but it is worth pointing out that Weisz et al. (2014) have been able to reconstruct the very faint-end of the UV LF at  $z = 3-5$  down to  $M_{\text{AB}} \approx -5$  using the star formation histories of Local Group dwarf galaxies, showing no evidence for a break of the LF.

Predictions from semi-analytic models typically do not go much fainter than  $L_{\text{Ly}\alpha} \approx 10^{40}$  erg s $^{-1}$  or  $M_{\text{AB}} \approx -13$ , often because of the mass resolution of the parent  $N$ -body simulation (e.g., Garel et al. 2015a; Gurung-López et al. 2020). Alternative models, based on the Extended Press-Schechter formalism, predict that the UV LF at  $z \approx 4$  keeps rising until UV magnitudes of  $\approx -8$ , corresponding to galaxies hosted by halos at the atomic cooling limit (Yung et al. 2019). Beyond the atomic cooling limit, it is unclear whether the processes that regulate galaxy formation are strong enough to break the slope of the LF, especially at redshifts 3–4 when the fully ionized IGM easily resists the gravitational pull of low-mass dark matter halos (e.g., Okamoto et al. 2008).

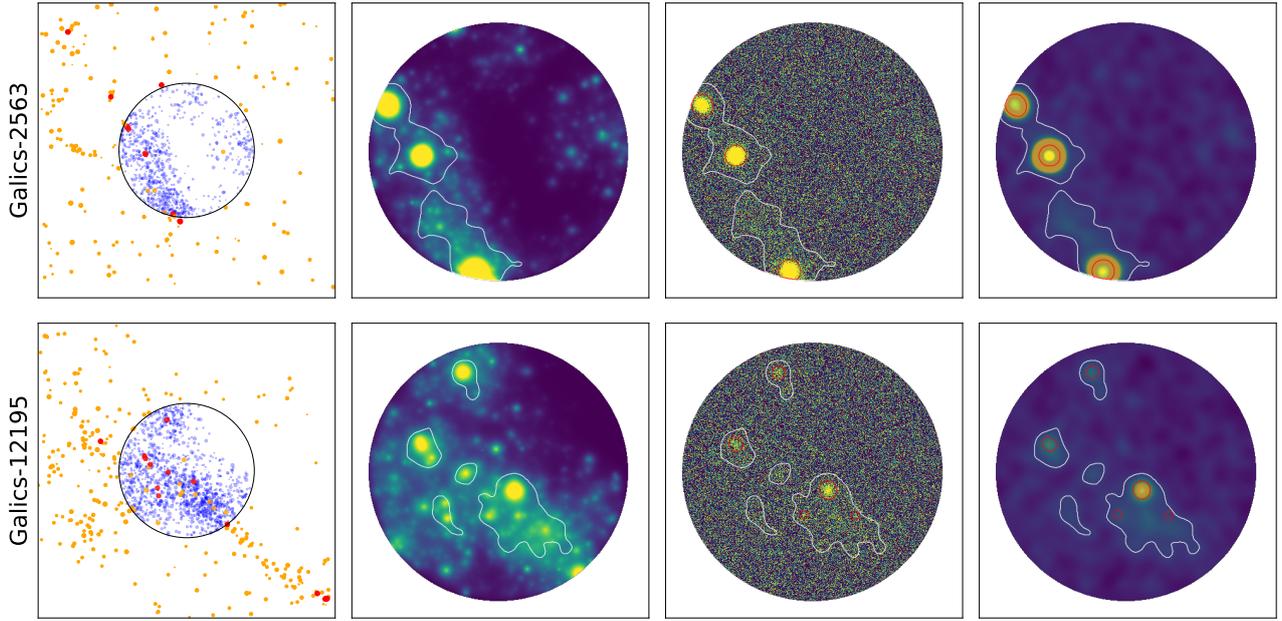
To understand how our model is affected by the uncertainty on how faint the power law behavior of the Ly $\alpha$  LF extends, we show in Fig. 14 the results of our model when we ignore the contribution of galaxies fainter than  $L_{\text{Ly}\alpha} = 10^{38}$  erg s $^{-1}$  (red symbols) and  $10^{37}$  erg s $^{-1}$  (green symbols). With these cuts and a slope  $\alpha = -1.84$ , the expected SB due to undetected LAEs is in the range 30–70% of the detected emission and thus remains an important source of luminosity. Assuming a steeper slope ( $\alpha = -1.94$ ), we see that even when the LF is truncated at  $L_{\text{Ly}\alpha} = 10^{38}$  erg s $^{-1}$ , the signal from undetected LAEs may fully explain our observations for most groups. Alternatively, we clearly see from Fig. 14 that a slope shallower than  $\alpha = -1.74$  would only reproduce 10 to 30% of the diffuse Ly $\alpha$  flux, regardless of the value of the low luminosity cut. In this case, the contribution of undetected LAEs is unable to account for the diffuse emission, unless a very strong clustering is assumed (i.e., a very high  $\zeta$ ). Our model with a sharp cut of the LF below a given luminosity is of course unrealistic as one can expect a smooth transition between the power law behavior and the LF decline. Unfortunately we have no observational clue for the Ly $\alpha$  LF shape at such low luminosity, but one can point out that using a more realistic LF shape will simply increase the required steepness of the LF.

### 5.2.2. GALICS semi-analytical model

The arguments developed in the previous section demonstrate that the faint, undetected LAEs can be sufficiently numerous to contribute most or even all of the observed diffuse Ly $\alpha$  emission. We now use GALICS, a semi-analytic model of galaxy formation coupled to numerical Ly $\alpha$  radiation transfer models (Hatton et al. 2003; Garel et al. 2012, 2015a), to test this conclusion with a more elaborated model. In the following we give a brief description of the method used to generate mock Ly $\alpha$  narrowband images. A full description of the model is given in Appendix A.

GALICS relies on a cosmological  $N$ -body simulation to follow the hierarchical growth of dark matter structures and on semi-analytic prescriptions to describe the physics of the baryonic component. We generate 100 mock lightcones that mimic the geometry and redshift range of the MUSE HUDF survey.

<sup>11</sup> Here, we make the reasonable assumption that  $L_{\text{Ly}\alpha} [\text{erg s}^{-1}] = 10^{42}$  (SFR/ $M_{\odot} \text{ yr}^{-1}$ ) and that very faint galaxies are likely to be metal-poor, such that dust attenuation is expected to be much less significant than for brighter objects (e.g., Garel et al. 2015a; Maseda et al. 2020).



**Fig. 15.** Two examples of GALICS simulations of diffuse Ly $\alpha$  emission in overdensities. *Top row:* ID 2563  $\delta = 2.8$ ,  $z = 3.10$ . *Bottom row:* ID 12195  $\delta = 6.5$ ,  $z = 4.96$ . *First column:* distribution of simulated galaxies within the MOSAIC and MXDF field of view. Bright Ly $\alpha$  emitters with  $F_{\text{Ly}\alpha} > 10^{-18}$  are marked with red symbols and those with  $10^{-19} < F_{\text{Ly}\alpha} < 10^{-18}$  with orange symbols. In the MOSAIC, the limit for bright Ly $\alpha$  emitters is set to  $10^{-17.5}$ . Flux units are  $\text{erg s}^{-1} \text{cm}^{-2}$ . Low-luminosity sources extrapolated from the LF (see text) are shown in blue. *Second column:* corresponding noiseless Ly $\alpha$  narrowband image. *Third column:* simulated MXDF narrowband image with noise. *Fourth column:* composite of low and mid frequencies IUWT wavelets S/N components. The contours of the identified filament and compact source area are shown respectively in white and red. The MXDF field of view has an 82 arcsec diameter or 2.6 cMpc at  $z = 3$ .

We then reproduce the detection method described in Sect. 3.2 to identify LAE overdensities in the mock fields. In total, this yields 13253 mock groups, of which 2475 have  $\delta \geq 2$  and  $N_{\text{LAE}} \geq 7$ .

In the previous section we have shown that a steep faint end slope ( $\alpha \lesssim -1.84$ ) of the Ly $\alpha$  LF is required to explain the observed diffuse Ly $\alpha$  emission. The model requires integrating the LF down to very low halo luminosity:  $10^{37} - 10^{38} \text{ erg s}^{-1}$ . However, the current dark matter mass resolution of the cosmological simulation used in GALICS is  $2 \times 10^9 M_{\odot}$ . This mass cut corresponds approximately to a Ly $\alpha$  luminosity of  $10^{40} - 10^{41} \text{ erg s}^{-1}$ , depending on the redshift. To overcome this limitation, we produce a large number of low-luminosity ad hoc LAEs by extrapolating each group LF down to  $10^{37} \text{ erg s}^{-1}$ . The Schechter LF model already presented in Sect. 5.2.1 with  $\alpha = -1.84$  is used. We note that from the recipes of star formation in GALICS presented in Appendix A.3, we predict that the faint-end slope of the Ly $\alpha$  LF should be about  $-1.8$ . These galaxies are then spatially placed to produce a given level of angular clustering (i.e., the factor  $\zeta$  in Eq. (1)). A Gaussian kernel with a width of 0.1 arcmin is used to randomly place the ad hoc sources around GALICS LAE in the mock fields (see Appendix A for a detailed description).

By construction, the spatial extent of Ly $\alpha$  sources is not modeled in GALICS. In order to build a simulated Ly $\alpha$  narrowband image we proceed as follows:

An empty image with the MXDF field of view is created. For each source in the selected GALICS group catalog, we derive a spatial profile using the galaxy-halo decomposition performed by Leclercq et al. (2017). Surface brightness is modeled as a sum of two circular, tow-dimensional exponential profiles. Statistical information and/or correlation with some measured properties is also available from the GALICS output, and are used to constrain

the four model parameters. Details of the model can be found in Appendix B.

The simulated Ly $\alpha$  image is then convolved with the MXDF Moffat PSF model and Gaussian noise is added using the MXDF datacube variance values summed over the corresponding spectral window. To take into account the increased noise with respect to the propagated values, as measured in Sect. 4.3, a factor of two is applied to the noise standard deviation. In addition, a Gaussian spatial filter with a FWHM of one spaxel is performed on the noisy image to simulate the noise correlation present in the datacube. This image is finally processed using the same detection parameters as the one used in Sect. 4.1.

We present two examples of simulated groups in Fig. 15. These examples have been selected to be roughly representative of our observations, with overdensity values of 3 and 6, redshifts of 3 and 5 and filament Ly $\alpha$  extended flux of  $\approx 2 \times 10^{-17} \text{ erg s}^{-1} \text{cm}^{-2}$ .

Figure 15 shows that the overall distribution of faint sources is well captured by the detection algorithm. One can see the power of the method by comparing the noisy Ly $\alpha$  narrowband image in the third column with the wavelet decomposition in the right panel. While it is hard to see more than the bright sources in the narrowband, the extended emission is clearly apparent in the low frequency IUWT images.

We note that the results obtained for these two show-case structures are illustrative of the success but also of the limitation of the detection method. In both cases, the algorithm appears sufficiently powerful to identify a diffuse emission structure when there is indeed one, even if it is hardly or not at all visible by-eye in the data. However, the second emission is so faint that the estimated S/N (see Table 3) may only be marginally higher than the S/N that the algorithm would attribute to structures identified by mistake in the noise. In effect, if similar structures and

**Table 3.** Example of filament detection in simulated GALICS overdensities.

ID	$z$	$\delta$	$F_{\text{cat}}$	$F_{\text{fil}}$	$F_{\text{dif}}$	$S/N_{\text{dif}}$	$S_{\text{dif}}$	$SB_{\text{dif}}$	$F_{\text{dif}}/F_{\text{fil}}$
2563	3.10	2.8	133.3	$113.2 \pm 0.7$	$20.0 \pm 0.6$	15.5	670	3.0	17.7
12195	4.96	6.5	80.2	$33.2 \pm 1.4$	$23.2 \pm 1.4$	8.8	611	3.8	69.9

**Notes.** ID: GALICS group ID.  $z$ : redshift.  $\delta$ : overdensity factor.  $F_{\text{cat}}$ : Total Ly $\alpha$  flux from catalog.  $F_{\text{fil}}$  and  $F_{\text{dif}}$ : Ly $\alpha$  flux measured in the filaments and the diffuse components.  $S/N_{\text{dif}}$ : S/N of diffuse Ly $\alpha$  emission.  $S_{\text{dif}}$ : Area of the diffuse component (arcsec<sup>2</sup>).  $SB_{\text{dif}}$ : Average surface brightness of the diffuse Ly $\alpha$  emission ( $10^{-20}$  erg s<sup>-1</sup> cm<sup>-2</sup> arcsec<sup>-2</sup>).  $F_{\text{dif}}/F_{\text{fil}}$ : Fraction of flux in the diffuse component in %. Flux unit are  $10^{-18}$  erg s<sup>-1</sup> cm<sup>-2</sup>.

S/N were obtained for some groups of the MXDF datacube, their associated  $P$ -values would be 0.3% and 40% respectively (see Fig. 8). Hence, in a noise-only situation, the S/N would be found to be larger than the S/N of the second show-case 40% of the time, showing that such a source is close to the detection limit. We note, however, that this reasoning assumes that the relation between  $P$ -values and S/N shown in Fig. 8 is also valid for the GALICS simulated narrowbands, which is probably pessimistic given the difference in noise properties between the real and the simulated data<sup>12</sup>. The existence of such a limit is indeed inherent to any detection method. These remarks help to interpret the results of Fig. 9, where many groups show diffuse emission that (if real) resembles the examples of the two show-cases of Fig. 15.

In Fig. 16 we show the derived filament parameters obtained when running the detection process on the 2475 GALICS groups with  $\delta \geq 2$  and  $N_{\text{LAE}} \geq 7$ . The simulations present a large scatter in the filaments properties. A large part of it is due to field-of-view effects. For statistical reasons we have used the large MOSAIC field to search for overdensities, but depending on the relative location of the filaments with respect to the smaller MXDF field, one might miss part of the structure. This is obvious in Fig. 15 where one can see that moving the MXDF location will eventually lead to no filament detection.

Compared to the GALICS simulation, the observations (black labels in Fig. 16) are located in the upper tail of the distribution of the simulated filament properties. This is particularly true for the diffuse Ly $\alpha$  flux and surface brightness (second and right panels in Fig. 16). As we can see from the figure, in most cases, the simulation is able to reproduce the observed Ly $\alpha$  surface brightness in diffuse areas ( $SB_{\text{dif}}$ ) only in very overdense regions ( $\delta > 10$ ).

Figure 15 indicates that the Ly $\alpha$  halos surrounding the numerous faint LAEs in each overdensity should overlap significantly, leading to a projected covering factor of unity over the entire area. We recall that in our previous study of the incidence rates of Ly $\alpha$  emission (Wisotzki et al. 2018) we found that without spatial clustering, a surface brightness level of  $5 \times 10^{-20}$  cgs corresponds to an incidence rate on the order of 0.5 per unit redshift, comparable to that of high column density HI absorbers. In the overdense regions considered here this incidence rate presumably increases further. It is therefore plausible (but unfortunately untestable without bright background sources) that each line-of-sight through these overdensities would also reveal strong HI absorption.

<sup>12</sup> This difference between the simulation and our observations could be explained by the idealized nature of the noise properties in the simulation. Although care was taken to properly scale the noise to reach the variance level observed in the data (see Sect. 4.3), the Gaussian noise injected in these simulations is different from the actual distribution of the noise. In particular, the imperfect continuum subtraction may leave behind observed narrowband Ly $\alpha$  image systematics, which are not present in the simulated narrowbands.

The median surface brightness estimated in the GALICS simulations is  $1.7 \times 10^{-20}$  erg s<sup>-1</sup> cm<sup>-2</sup> arcsec<sup>-2</sup> at  $\delta < 5$ , a value significantly smaller than the minimum surface brightness in the high confidence groups ( $3.7 \times 10^{-20}$  erg s<sup>-1</sup> cm<sup>-2</sup> arcsec<sup>-2</sup>). Only 8% of the simulated groups have a surface brightness brighter than this value. This is a factor of three below the 23% (5/22) success rate of the observations which implies that the GALICS model is producing a lower average Ly $\alpha$  surface brightness. The ad hoc Gaussian model we use to extrapolate GALICS results and to distribute faint sources is likely responsible for part of this disagreement. Indeed, (brighter) galaxies are typically observed to have a power-law correlation function, steepening at small scale due to the one-halo term (e.g., Hildebrandt et al. 2009; Harikane et al. 2017). The Gaussian kernel we use smooths out the one-halo term at scales below  $\approx 0.1$  arcmin which results in spreading the luminosity of faint galaxies too widely.

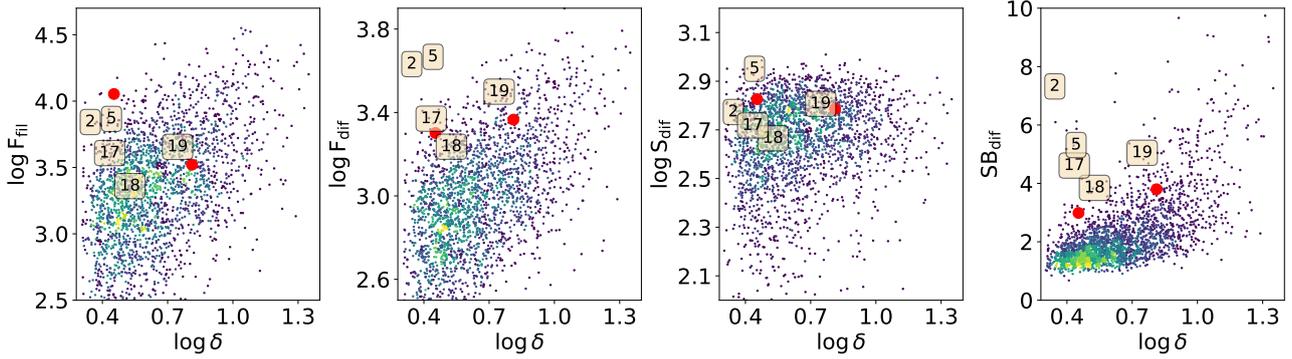
We note also that our GALICS mock catalogs are based on many other approximations, including an extrapolation of the Ly $\alpha$  LF by three orders of magnitude and an empirical model of galaxy and halo decomposition. It is beyond the scope of the present paper to present a self-consistent physical model compatible with our observations. Nevertheless, despite these approximations, the analysis of the mock data confirms that faint LAEs cannot be ignored in the photon budget and that they can produce a significant fraction or even most of the observed diffuse emission.

### 5.3. Gravitational heating

Gravitational compression of gas during structure formation is a net heating term that may be dissipated through the emission of Ly $\alpha$  radiation. In the low-density IGM, the thermal state of the gas is determined by a competition between cooling and heating both from gravitational compression and from the UVB (e.g., Hui & Gnedin 1997). At values of the density contrast<sup>13</sup> typical of large-scale filaments ( $\approx 10$ ), the temperature of the gas is mostly the result of the competition between photo-heating from the UVB and radiative cooling (through Ly $\alpha$  emission). From Rosdahl & Blaizot (2012, their Fig. 9), we see that the Ly $\alpha$  emissivity of gas in this regime (i.e., at densities  $< 0.01$  cm<sup>-3</sup>) is well described by a steep powerlaw  $\propto n_{\text{H}}^{2.5}$ . That the slope is steeper than  $n_{\text{H}}^2$  is due to the fact that the ionized fraction rises toward low densities. Thus, at low densities, the joint contribution of the UVB and gravitational compression to Ly $\alpha$  emission drops rapidly and is not likely to contribute significantly to the emission we observe.

Within dark matter halos, or in their immediate vicinity, part of the IGM has condensed into denser cold accretion streams (e.g., Kereš et al. 2005). These streams can produce Ly $\alpha$

<sup>13</sup> We refer here to the overdensity in the density field, not to the overdensity in the number of LAEs as used elsewhere in the paper.



**Fig. 16.** Filament properties derived from GALICS mocks. *From left to right:* as a function of the overdensity ( $\delta$ ), the filament total and diffuse Ly $\alpha$  flux in units of  $10^{-20}$  erg s $^{-1}$  cm $^{-2}$ , the area (arcsec $^2$ ) and surface brightness ( $10^{-20}$  erg s $^{-1}$  cm $^{-2}$  arcsec $^{-2}$ ) of the diffuse component. Color coding correspond to the number of points in the two-dimensional histogram bins with increasing values from blue to yellow. The group IDs for the corresponding observed filaments properties are shown as black labels. The red points display the two show cases presented in Fig. 15 and Table 3.

emission by dissipating their gravitational energy (Fardal et al. 2001; Dijkstra & Loeb 2009; Faucher-Giguère et al. 2010; Rosdahl & Blaizot 2012). In this regime, gravitational heating largely dominates over fluorescence from the UVB (Rosdahl & Blaizot 2012, their Fig. 6). This mechanism has been suggested as a possible source of the bright Ly $\alpha$  extended emission observed in Ly $\alpha$  blobs (e.g., Haiman et al. 2000; Fardal et al. 2001; Nilsson et al. 2006; Rosdahl & Blaizot 2012). For example, in their study of the R0-1001 Ly $\alpha$  nebulae, Daddi et al. (2020) conclude that the gravitational energy associated with gas infall is the most likely source of power for the observed extended Ly $\alpha$  emission ( $1.3 \times 10^{44}$  erg s $^{-1}$ ).

The same mechanism should also occur in less massive galaxies when cold gas falls into the potential wells of their dark matter halos. This may explain part of the extended Ly $\alpha$  emission observed around most LAEs with MUSE (Wisotzki et al. 2016; Leclercq et al. 2017). Again, by construction, the diffuse Ly $\alpha$  emission we measure excludes the CGM of all LAEs detected in the field, and we can thus rule out that cooling radiation is responsible for the diffuse emission we report.

## 6. Source of diffuse Ly $\alpha$ emission

In the previous section, we discussed three possible sources of energy that could power the diffuse Ly $\alpha$  radiation we observe: (1) fluorescence from the UVB (2) Ly $\alpha$  emission from an abundant population of undetected small Ly $\alpha$  emitters, and (3) dissipation of gravitational heating. We find that Ly $\alpha$  fluorescence powered by the cosmic UV background can explain at most 28–34% of the observed signal at  $z \approx 3$ , and less than 10% at  $z \approx 4.5$  (Sect. 5.1.1). Even if our estimate is relatively uncertain (for example we use the maximum value of one for the LLS covering fraction as in Cantalupo et al. 2005 and we neglect pumping from other lines that may enhance the signal by  $\approx 20\%$ , Furlanetto et al. 2005), these results suggest that this process is not responsible for the bulk of the emission we observe. Perhaps more importantly, we note that Ly $\alpha$  fluorescence from the UVB is mostly produced by relatively dense gas, with column densities typical of Lyman Limit Systems (LLS) or larger. This gas is mostly located in the CGM of galaxies (or within their dark matter halos, e.g., van der Voort et al. 2012) and its Ly $\alpha$  emission may thus be associated with galaxies in Ly $\alpha$  surveys, in the form of Ly $\alpha$  halos. Moreover, while increasing the intensity of the local UVB will increase the overall emissivity of this dense gas, it will also decrease the emissivity of lower-density gas (at  $n_{\text{H}} < 0.05$  cm $^{-1}$ , see Rosdahl & Blaizot

2012, their Fig. 16), thus pushing the emission even closer to the galaxies. It is thus not clear whether fluorescence from the UVB produces a signal that is distinguishable from that produced by galaxies and their CGM.

Similarly, we have argued that gravitational heating will lead to dissipation through Ly $\alpha$  emission only in the densest parts of the IGM, in the CGM. This emission may contribute to the extended Ly $\alpha$  halos around LAEs, in addition to fluorescence from the UVB or local star formation, to scattered Ly $\alpha$  photons from galaxies, or to extended star formation (e.g., Mas-Ribas et al. 2017). For more diffuse gas, the heating source is a combination of UVB and gravitational heating, which cannot be disentangled, but which has been measured in simulations to produce extremely faint emission, decreasing rapidly with decreasing density (Rosdahl & Blaizot 2012). Thus, it appears that fluorescence or cooling radiation, while they may play a significant role in lighting up the CGM of galaxies, are not able to explain the levels of extended emission that we detect beyond this CGM.

We therefore find that the most likely explanation for the diffuse emission we observe is an abundant population of ultra-faint undetected LAEs (Sect. 5.2). This result is in line with the recent findings of Mitchell et al. (2021) who use high-resolution radiation-hydrodynamical simulations to show that the very extended part of Ly $\alpha$  halos around galaxies is mostly due to undetected neighboring galaxies. Our conclusion relies on a number of assumptions: the faint-end slope of the Ly $\alpha$  LF must be steep enough ( $\alpha \lesssim -1.84$ ), this power-law behavior must extend down to at least  $\sim 10^{38}$ – $10^{37}$  erg s $^{-1}$  before turning over and the clustering of faint Ly $\alpha$  emitters must be favorable (filling factor  $\sim 1/6$ ). While these assumptions seem reasonable, there are today no observational constraints on the Ly $\alpha$  LF shape below  $10^{40.5}$  erg s $^{-1}$  or on the angular clustering of extremely faint Ly $\alpha$  emitters. Our observations may provide a new original and indirect constraint on this population of extremely faint LAEs.

To appreciate the nature of the extremely faint objects that may contribute to the observed signal, we remind the reader that a single massive star of  $40 M_{\odot}$  will produce of order  $10^{49}$  ionizing photons per second for  $\approx 6$  Myr (e.g., Geen et al. 2018). Assuming these photons are all processed to yield Ly $\alpha$  through case B recombination, this single star would produce  $\sim 10^{38}$  erg s $^{-1}$  in the Ly $\alpha$  line. The detailed simulations of molecular clouds by Kimm et al. (2019) show a more complete picture. These authors predict that a cloud of total mass  $10^5 M_{\odot}$  with  $10^4 M_{\odot}$  in stars has a Ly $\alpha$  luminosity that evolves from  $3 \times 10^{39}$  erg s $^{-1}$  at the onset of star formation to  $5 \times 10^{36}$  erg s $^{-1}$

when the cloud is disrupted 6 Myr later. A cloud of the same mass, but which forms ten times less mass in stars (i.e., with stellar mass  $10^3 M_{\odot}$ ) is found to have a Ly $\alpha$  luminosity evolving from  $6 \times 10^{38}$  to  $3 \times 10^{35}$  erg s $^{-1}$  over 20 Myr. In both cases, the Ly $\alpha$  luminosity of the cloud drops faster than the production of ionizing photons by the stellar population because the disruption of the cloud allows ionizing photons to escape without being reprocessed into Ly $\alpha$ . Systems with Ly $\alpha$  luminosities as faint as  $10^{37}$ – $10^{38}$  erg s $^{-1}$  thus require very few massive stars to produce such luminosities. While some of these objects may be explained by stronger star formation events seen at later stages (when the ionizing radiation has dropped and neutral gas has been blown away), or by very low Ly $\alpha$  escape fractions, it is likely that sources with  $L_{\text{Ly}\alpha} \sim 10^{37-38}$  erg s $^{-1}$  are indeed extremely small systems, with stellar masses as small as a few thousand solar masses, which are seen at the very first moments of their formation. These could be nascent galaxies or even (compact) star clusters, which have been observed at similar implied continuum magnitudes at higher redshifts (e.g., [Boylan-Kolchin 2017](#); [Vanzella et al. 2018, 2021](#)), although currently we cannot differentiate between the two possibilities.

While assessing the baryonic content of these ultra-faint LAEs is highly uncertain, we do not expect that these objects make a significant contribution to the global stellar mass budget. By extrapolating the SFR– $M_{\star}$  relation at  $z \approx 3$ –4 (e.g., [Salmon et al. 2015](#)) and assuming our fiducial LF slope  $\alpha = -1.84$ , we estimate that LAEs with  $10^{38} < L_{\text{Ly}\alpha} < 10^{42}$  erg s $^{-1}$  should only contribute  $\lesssim 10$ – $20\%$  to the cosmic stellar mass density.

The existence of a large number of faint galaxies also has some implications for synthesis models of the UVB. One of the key ingredients of these models is the total emissivity of galaxies, which scales with the integral of the UV LF down to faint luminosities:  $0.01 L_{\star}$  ([Haardt & Madau 2012](#)) or  $M_{\text{AB}} = -13$  ([Faucher-Giguère 2020](#)). Extending these integrals down to  $M_{\text{AB}} \sim -8$  as suggested above would increase the global UV emissivities of these models by  $\sim 50\%$  ([Haardt & Madau 2012](#)) and  $\sim 20\%$  ([Faucher-Giguère 2020](#)). While these models can probably absorb this difference by readjusting their free parameters to reproduce the same observed values of  $\Gamma_{\text{HI}}$ , it is interesting to note that FG20’s model requires values of the escape fractions of ionizing radiation as low as 1% (at  $z = 3$ ), which is already in tension with observational results ([Steidel et al. 2018](#)) and numerical works (e.g., [Rosdahl et al. 2018](#)). Increasing the total emissivity of galaxies by extending the LF to the faint-end makes this situation yet more uncomfortable. If confirmed, the very faint population of LAEs that we discussed above may thus require more important adaptations of these models.

We note that our analysis assumes that the Ly $\alpha$  emission from the CGM is counted as part of each galaxy’s luminosity. Indeed, while previous measurements of the Ly $\alpha$  LF were restricted to the Ly $\alpha$  flux from galaxies, and were thus missing a large part of the Ly $\alpha$  flux (70% on average, [Leclercq et al. 2017](#)), the latest measurements with MUSE carefully include the halo flux (e.g., [Drake et al. 2017](#); [Herenz et al. 2019](#)). Given the difficulty of measuring the extended Ly $\alpha$  flux at large galactocentric distance, the Ly $\alpha$  flux in the CGM may still be underestimated. This uncertainty affects the requirement stated above on the shape the LF should have to reproduce our observations.

Including the CGM Ly $\alpha$  emission in each galaxy’s luminosity makes sense in a scenario where this emission is due to scattered light from the central galaxy. In that case, the total Ly $\alpha$  luminosity better relates to the star formation rate. This is a strong assumption, however, which is still heavily debated in the theoretical work (e.g., [Mitchell et al. 2021](#); [Byrohl et al. 2020](#)). It

is interesting that the other sources of energy (UVB or gravitational compression) are also expected to produce most of their Ly $\alpha$  emission in the CGM. In that respect, these different processes are not strictly additive. Indeed, it is likely that the signal we detect does include a contribution from these terms, and that this contribution is present mostly in the CGM of faint galaxies, which we account for in the LF regardless of its physical origin.

Finally, we note that our results do not confirm the conclusion of [Elias et al. \(2020\)](#) who have predicted that diffuse Ly $\alpha$  emission from cosmic web filaments at  $z \approx 3$  and restricted to the IGM, should be easily detected by MUSE in a 30 h deep exposure. Our measurements within the MXDF, a 5 times deeper exposure than the one foreseen in their analysis, confirm that their prediction, as already pointed out by the authors, was overly optimistic. Moreover, we show that the majority of the diffuse flux is due to the Ly $\alpha$  emission within the CGM of undetected galaxies and that only a small fraction can be coming from the IGM proper.

## 7. Summary and conclusions

We introduced and analyzed observations carried out with MUSE in the MXDF, a single, 140 h deep field located in the HUDF area, complemented by MUSE observations of the entire HUDF area at 10 h depth. This resulted in datacubes of exquisite sensitivity and a catalog of 1258 Ly $\alpha$  emitters covering redshifts 2.9 to 6.7. We analyzed this unique data set to (i) detect and quantify Ly $\alpha$  emitter overdensities and (ii) search and characterize diffuse Ly $\alpha$  emission within each detected overdensity. Our major findings are summarized below:

Ly $\alpha$  emitters are strongly clustered in redshift space, with 30% of the 1258 Ly $\alpha$  emitters residing in 22 overdensities spanning redshifts 3.0 to 5.8. The overdensity, measured within volumes of  $260 \text{ cMpc}^3$  ( $z = 3.0$ ) to  $415 \text{ cMpc}^3$  ( $z = 5.8$ ), is on average 3.2 and reaches 5.0 for the densest groups at  $z = 4.5$  and  $z = 4.7$ . LAEs in overdensities have a low average Ly $\alpha$  luminosity ( $10^{41.5}$  erg s $^{-1}$ ) and their mean number density is  $0.055 \pm 0.017 \text{ cMpc}^{-3}$ .

SED photometric analysis of the 67% of the LAEs with an HST counterpart shows that the galaxy population in overdensities is mainly composed of low-mass ( $1.4 \times 10^8 M_{\odot}$ ), young (0.3 Gyr) galaxies with high specific star formation rates ( $10^{-8.5} \text{ yr}^{-1}$ ). The majority of the remaining LAEs without an entry in the [Rafelski et al. \(2015\)](#) photometric catalog display very high Ly $\alpha$  equivalent widths and no visible HST counterpart. On average, these very faint ( $M_{\text{UV}} \sim -15$ ) star forming galaxies must have stellar mass below  $10^7 M_{\odot}$ . Overdensities are expected to be populated by dark matter halos with masses around  $10^{11.3} M_{\odot}$ , with the notable exception of the  $z = 3.07$  group 2 for which we estimate a halo mass of  $10^{13.5} M_{\odot}$ .

A search for diffuse Ly $\alpha$  emission within the MXDF area with an original method based on multiscale undecimated isotropic wavelet transforms, resulted in the identification of 14 overdensities with extended Ly $\alpha$  emission. This extended Ly $\alpha$  emission arises from filamentary structures of cMpc size and covers an area of 0.4–1.1 cMpc $^2$ , corresponding to a significant fraction (10–20%) of the total MXDF area. Group 2 at  $z = 3.07$  is the only group displaying a filament outside the MXDF field, in an extended area covered with MUSE at 10 h depth. This additional detection extends the MXDF filament length to 4.6 cMpc, and reveals a second filament crossing the field.

We use the mid and large spatial scale segmentation images resulting from the wavelet decomposition process to split filaments into areas corresponding to compact sources and diffuse

emission. We show that the compact source area can account for the total Ly $\alpha$  emission of identified galaxies, including the emission from their CGM. Using Monte Carlo simulations we show that five overdensities among the 14 with extended Ly $\alpha$  emission display an extended diffuse Ly $\alpha$  emission signal with a mean surface brightness of  $5.1 \pm 1.2 \times 10^{-20} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ arcsec}^{-2}$  and a high confidence level ( $P$ -value  $< 0.06$ ). The Ly $\alpha$  luminosity from this diffuse area represents a large fraction of the total filament flux:  $68 \pm 6\%$ .

We have investigated the potential impact of AGN and concluded that, except for groups 2 and 6 where we cannot rule out a boost in Ly $\alpha$  surface brightness, AGN are unlikely to be a significant contributor to the observed extended Ly $\alpha$  emission in the 14 groups. This is especially true for 4 of the 5 overdensities with high confidence diffuse Ly $\alpha$  emission.

At  $z \approx 3$ , a maximum of 28% and 34% of the observed surface brightness of the diffuse emission can be explained by Ly $\alpha$  fluorescence powered by the cosmic UV background. At higher  $z$ , this fraction is reduced to below 10%. The measured diffuse Ly $\alpha$  surface brightness can be reproduced by a population of undetected ultra low-luminosity Ly $\alpha$  emitters, provided that the faint end of the Ly $\alpha$  LF is steep enough ( $\alpha \lesssim -1.84$ ), that it extends to luminosities lower than  $10^{38} - 10^{37} \text{ erg s}^{-1}$  and that the clustering of faint Ly $\alpha$  emitters is significant (filling factor  $< 1/6$ ).

Narrowband Ly $\alpha$  emitter wide field surveys, followed-up by spectroscopic multi-object observations, have produced a wealth of information on LAE clustering and the large-scale structure of galaxies at high  $z$  (e.g., Francis et al. 2004; Matsuda et al. 2005; Zheng et al. 2016; Ouchi et al. 2020). They cover a large volume: for example  $200 \times 200 \times 80 \text{ Mpc}^3$  for the SILVER-RUSH survey (Ouchi et al. 2018; Shibuya et al. 2017), but at the expense of a bright limiting luminosity (typically  $\sim 10^{43} \text{ erg s}^{-1}$ ) and sparse sampling (e.g., 179 LAEs for the Harikane et al. 2019 study at  $z = 5.7$  and  $6.6$ , that is  $5.6 \times 10^{-5} \text{ Mpc}^{-3}$ ). With its two orders of magnitude higher average number density of LAEs (e.g.,  $6.3 \times 10^{-3} \text{ Mpc}^{-3}$  at  $z = 6$ ), our study for the first time probes the large-scale structure at high  $z$  with a much finer sampling. Thanks to this zoomed view of the cosmic web, we can resolve the filamentary structure in unprecedented detail. This is very complementary to the classical wide field approach.

Using both a simple analytical LF model and the GALICS semi-analytical model, we show that our measurements imply that a large population of ultra low-luminosity Ly $\alpha$  emitters ( $< 10^{40} \text{ erg s}^{-1}$ ) powers the diffuse Ly $\alpha$  emission within the filaments. The Ly $\alpha$  luminosity of this ultra faint population corresponds to star formation rates smaller than  $< 10^{-4} M_{\odot} \text{ yr}^{-1}$ . An important consequence of these results is that it is unclear whether intergalactic (as opposed to circumgalactic) gas is even observable through its Ly $\alpha$  emission because the denser regions where this emission is most likely to be detected will be crowded with a high surface density of small Ly $\alpha$  emitting galaxies. Even if the study of IGM proper in emission could be problematic, understanding how star formation proceeds in these dwarf galaxies, down to what Ly $\alpha$  luminosity the LF extends, what their clustering properties are and what the relative importance is of the various processes that produce the Ly $\alpha$  emission within the CGM and IGM, will be very valuable to understand galaxy formation. Last – but not least – the population of faint Ly $\alpha$  emitters appears to be a good tracer of the cosmic web.

This first detection of the cosmic web structure in Ly $\alpha$  emission in typical filamentary environments, namely outside massive structures typical of web nodes, is a milestone in the long search for the cosmic web signature at high  $z$ . This has been possible because of the unprecedented faint surface brightness of  $5 \times 10^{-20} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ arcsec}^{-2}$  achieved by 140 h MUSE observations on the VLT. However, because of the limited size of the

MXDF, we cannot trace the full length of the filaments, which extend at least to the megaparsec (physical) scale as demonstrated in the case of the  $z = 3.07$  filament (Fig. 11). Repeating this study on a larger field and different environments would be very valuable to increase the detection statistics and to provide better constraints on the filaments' physical parameters. This remains, nevertheless, a costly investment in telescope time. In the longer term, the BlueMUSE project (Richard et al. 2019) will allow us to probe the Ly $\alpha$  redshift range 2–4 with a larger field of view. The lower Ly $\alpha$  redshift of BlueMUSE will be very beneficial for such a study. With a wavelength range mostly free of bright sky lines, a reduced impact of redshift dimming, and access to the peak of the cosmic star formation history, BlueMUSE should be able to perform similar studies in less telescope time and over larger fields of view.

*Acknowledgements.* We warmly thank ESO Paranal staff for their great professional support during all MXDF GTO observing runs. We thank the anonymous referee for a careful reading of the manuscript and helpful comments. We also thank Matthew Lehnert for fruitful discussions. RB, AF, SC acknowledge support from the ERC advanced grant 339659-MUSICOS. JB acknowledges support by Fundação para a Ciência e a Tecnologia (FCT) through the research grants UID/FIS/04434/2019, UIDB/04434/2020, UIDP/04434/2020 and through the Investigador FCT Contract No. IF/01654/2014/CP1215/CT0003. TG, AV acknowledges support from the European Research Council under grant agreement ERC-stg-757258 (TRIPLE). DM acknowledges A. Dabbech for useful interactions about IUWT and support from the GDR ISIS through the Projets exploratoires program (project TASTY). AF acknowledges the support from grant PRIN MIUR2017-20173ML3WW\_001. SLZ acknowledges support by The Netherlands Organisation for Scientific Research (NWO) through a TOP Grant Module 1 under project number 614.001.652. This research made use of the following open-source software and we are thankful to the developers of these: GNU Octave (Eaton et al. 2018) and its statistics, signal and image packages, the Python packages Matplotlib (Hunter 2007), Numpy (van der Walt et al. 2010), MPDAF (Piqueras et al. 2017), Astropy (Astropy Collaboration 2018), PyWavelets (Lee et al. 2019).

## References

- Alavi, A., Siana, B., Richard, J., et al. 2016, *ApJ*, **832**, 56
- Ammanouil, R., Ferrari, A., Mary, D., Ferrari, C., & Loi, F. 2019, *MNRAS*, **490**, 37
- Arrigoni Battaia, F., Obreja, A., Prochaska, J. X., et al. 2019, *A&A*, **631**, A18
- Ashour, S. K., & Abdel-hameed, M. A. 2010, *J. Adv. Res.*, **1**, 341
- Astropy Collaboration (Price-Whelan, A. M., et al.) 2018, *AJ*, **156**, 123
- Atek, H., Richard, J., Kneib, J.-P., & Schaerer, D. 2018, *MNRAS*, **479**, 5184
- Bacher, R. 2017, PhD Thesis, Université Grenoble Alpes
- Bacon, R., Accardo, M., Adjali, L., et al. 2010, in *Ground-based and Airborne Instrumentation for Astronomy III*, eds. I. S. McLean, S. K. Ramsay, & H. Takami, *SPIE Conf. Ser.*, **7735**, 773508
- Bacon, R., Conseil, S., Mary, D., et al. 2017, *A&A*, **608**, A1
- Becker, G. D., & Bolton, J. S. 2013, *MNRAS*, **436**, 1023
- Beckwith, S. V. W., Stiavelli, M., Koekemoer, A. M., et al. 2006, *AJ*, **132**, 1729
- Behroozi, P., Wechsler, R. H., Hearin, A. P., & Conroy, C. 2019, *MNRAS*, **488**, 3143
- Bijaoui, A., Starck, J.-L., & Murthag, F. 1994, *Traitement du Signal*, **11**, 229
- Bolton, J. S., & Haehnelt, M. G. 2007, *MNRAS*, **381**, L35
- Bond, J. R., Kofman, L., & Pogosyan, D. 1996, *Nature*, **380**, 603
- Borisova, E., Cantalupo, S., Lilly, S. J., et al. 2016, *ApJ*, **831**, 39
- Bothwell, M. S., Kenicutt, R. C., Johnson, B. D., et al. 2011, *MNRAS*, **415**, 1815
- Bouwens, R. J., Oesch, P. A., Illingworth, G. D., Ellis, R. S., & Stefanon, M. 2017, *ApJ*, **843**, 129
- Boylan-Kolchin, M. 2017, *MNRAS*, **472**, 3120
- Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, *MNRAS*, **351**, 1151
- Burchett, J. N., Rubin, K. H. R., Prochaska, J. X., et al. 2020, *ApJ*, submitted [arXiv:2005.03017]
- Byrohl, C., Nelson, D., Behrens, C., et al. 2020, *ArXiv e-prints* [arXiv:2009.07283]
- Canalizo, G., & Stockton, A. 2001, *ApJ*, **555**, 719
- Cantalupo, S., Porciani, C., Lilly, S. J., & Miniati, F. 2005, *ApJ*, **628**, 61
- Cantalupo, S., Arrigoni-Battaia, F., Prochaska, J. X., Hennawi, J. F., & Madau, P. 2014, *Nature*, **506**, 63
- Cassata, P., Fèvre, O. L., Garilli, B., et al. 2011, *A&A*, **525**, A143

- Cohen, J. G., Blandford, R., Hogg, D. W., Pahre, M. A., & Shopbell, P. L. 1999, *ApJ*, **512**, 30
- da Cunha, E., Charlot, S., & Elbaz, D. 2008, *MNRAS*, **388**, 1595
- da Cunha, E., Walter, F., Smail, I. R., et al. 2015, *ApJ*, **806**, 110
- da Rocha, C., & Oliveira, C. M. D. 2005, *MNRAS*, **364**, 1069
- Dabbech, A., Ferrari, C., Mary, D., et al. 2015, *A&A*, **576**, A7
- Daddi, E., Valentino, F., Rich, R. M., et al. 2020, *A&A*, submitted [arXiv:2006.11089]
- de La Vieuville, G., Bina, D., Pello, R., et al. 2019, *A&A*, **628**, A3
- den Brok, J. S., Cantalupo, S., Mackenzie, R., et al. 2020, *MNRAS*, **495**, 1874
- Dijkstra, M., & Loeb, A. 2009, *MNRAS*, **400**, 1109
- Drake, A. B., Garel, T., Wisotzki, L., et al. 2017, *A&A*, **608**, A6
- Duncan, K., Conselice, C. J., Mortlock, A., et al. 2014, *MNRAS*, **444**, 2960
- Eaton, J. W., Bateman, D., Hauberg, S., & Wehbring, R. 2018, *GNU Octave Version 4.4.0 Manual: A High-Level Interactive Language for Numerical Computations*
- Eilers, A. C., Davies, F. B., Hennawi, J. F., et al. 2017, *ApJ*, **840**, 24
- Elias, L. M., Genel, S., Sternberg, A., et al. 2020, *MNRAS*, **494**, 5439
- Fardal, M. A., Katz, N., Gardner, J. P., et al. 2001, *ApJ*, **562**, 605
- Faucher-Giguère, C.-A. 2020, *MNRAS*, **493**, 1614
- Faucher-Giguère, C.-A., Kereš, D., Dijkstra, M., Hernquist, L., & Zaldarriaga, M. 2010, *ApJ*, **725**, 633
- Finoguenov, A., Rykoff, E., Clerc, N., et al. 2020, *A&A*, **638**, A114
- Francis, P. J., Palunas, P., Teplitz, H. I., Williger, G. M., & Woodgate, B. E. 2004, *ApJ*, **614**, 75
- Fumagalli, M., Cantalupo, S., Dekel, A., et al. 2016, *MNRAS*, **462**, 1978
- Furlanetto, S. R., Schaye, J., Springel, V., & Hernquist, L. 2005, *ApJ*, **622**, 7
- Fusco, T., Bacon, R., Kamann, S., et al. 2020, *A&A*, **635**, A208
- Gallego, S. G., Cantalupo, S., Lilly, S., et al. 2018, *MNRAS*, **475**, 3854
- Garel, T., Blaizot, J., Guiderdoni, B., et al. 2012, *MNRAS*, **422**, 310
- Garel, T., Blaizot, J., Guiderdoni, B., et al. 2015a, *MNRAS*, **450**, 1279
- Garel, T., Guiderdoni, B., & Blaizot, J. 2015b, *MNRAS*, **455**, 3436
- Geen, S., Watson, S. K., Rosdahl, J., et al. 2018, *MNRAS*, **481**, 2548
- Gilli, R., Cimatti, A., Daddi, E., et al. 2003, *ApJ*, **592**, 721
- Girelli, G., Pozzetti, L., Bolzonella, M., et al. 2020, *A&A*, **634**, A135
- Gould, A., & Weinberg, D. H. 1996, *ApJ*, **468**, 462
- Gunn, J. E., & Peterson, B. A. 1965, *ApJ*, **142**, 1633
- Gurung-López, S., Orsi, Á. A., Bonoli, S., et al. 2020, *MNRAS*, **491**, 3266
- Haardt, F., & Madau, P. 1996, *ApJ*, **461**, 20
- Haardt, F., & Madau, P. 2012, *ApJ*, **746**, 125
- Haiman, Z., & Rees, M. J. 2001, *ApJ*, **556**, 87
- Haiman, Z., Spaans, M., & Quataert, E. 2000, *ApJ*, **537**, L5
- Harikane, Y., Ouchi, M., Ono, Y., et al. 2017, *PASJ*, **70**, S11
- Harikane, Y., Ouchi, M., Ono, Y., et al. 2019, *ApJ*, **883**, 142
- Hatton, S., Devriendt, J. E. G., Ninin, S., et al. 2003, *MNRAS*, **343**, 75
- Hayes, M. J., Runnholm, A., Gronke, M., & Scarlata, C. 2021, *ApJ*, **908**, 36
- Herenz, E. C., Wisotzki, L., Saust, R., et al. 2019, *A&A*, **621**, A107
- Herenz, E. C., Hayes, M., & Scarlata, C. 2020, *A&A*, **642**, A55
- Hildebrandt, H., Pielorz, J., Erben, T., et al. 2009, *A&A*, **498**, 725
- Hinton, S. R., Davis, T. M., Lidman, C., Glazebrook, K., & Lewis, G. F. 2016, *Astron. Comput.*, **15**, 61
- Hogan, C. J., & Weymann, R. J. 1987, *MNRAS*, **225**, IP
- Hui, L., & Gnedin, N. Y. 1997, *MNRAS*, **292**, 27
- Hunter, J. D. 2007, *Comput. Sci. Eng.*, **9**, 90
- Inami, H., Bacon, R., Brinchmann, J., et al. 2017, *A&A*, **608**, A2
- Johnson, S. D., Chen, H.-W., Straka, L. A., et al. 2018, *ApJ*, **869**, L1
- Kereš, D., Katz, N., Weinberg, D. H., & Davé, R. 2005, *MNRAS*, **363**, 2
- Kikuta, S., Matsuda, Y., Cen, R., et al. 2019, *PASJ*, **71**, L2
- Kimm, T., Blaizot, J., Garel, T., et al. 2019, *MNRAS*, **486**, 2215
- Kolb, J., Madec, P. Y., Arsenault, R., et al. 2016, in *Adaptive Optics Systems V*, eds. E. Marchetti, L. M. Close, & J. P. Véran, *SPIE Conf. Ser.*, **9909**, 99092S
- Kollmeier, J. A., Zheng, Z., Davé, R., et al. 2010, *ApJ*, **708**, 1048
- Laursen, P., Sommer-Larsen, J., Milvang-Jensen, B., Fynbo, J. P. U., & Razoumov, A. O. 2019, *A&A*, **627**, A84
- Leclercq, F., Bacon, R., Wisotzki, L., et al. 2017, *A&A*, **608**, A8
- Leclercq, F., Bacon, R., Verhamme, A., et al. 2020, *A&A*, **635**, A82
- Lee, K.-G., Krolewski, A., White, M., et al. 2018, *ApJS*, **237**, 31
- Lee, G., Gommers, R., Waselewski, F., Wohlfahrt, K., & O'Leary, A. 2019, *J. Open Sour. Softw.*, **4**, 1237
- Lehmer, B. D., Alexander, D. M., Geach, J. E., et al. 2009, *ApJ*, **691**, 687
- Libeskind, N. I., van de Weygaert, R., Cautun, M., et al. 2018, *MNRAS*, **473**, 1195
- Livermore, R. C., Finkelstein, S. L., & Lotz, J. M. 2017, *ApJ*, **835**, 113
- Luo, B., Brandt, W. N., Xue, Y. Q., et al. 2017, *ApJS*, **228**, 2
- Lusso, E., Fumagalli, M., Fossati, M., et al. 2019, *MNRAS*, **485**, L62
- Madec, P. Y., Arsenault, R., Kuntschner, H., et al. 2018, in *Adaptive Optics Systems VI*, eds. L. M. Close, L. Schreiber, & D. Schmidt, *SPIE Conf. Ser.*, **10703**, 1070302
- Martin, C., Moore, A., Morrissey, P., et al. 2010, in *Ground-based and Airborne Instrumentation for Astronomy III*, eds. I. S. McLean, S. K. Ramsay, & H. Takami, *SPIE Conf. Ser.*, **7735**, 77350M
- Martin, D. C., Matuszewski, M., Morrissey, P., et al. 2015, *Nature*, **524**, 192
- Martizzi, D., Vogelsberger, M., Artale, M. C., et al. 2019, *MNRAS*, **486**, 3766
- Mary, D., Bacon, R., Conseil, S., Piqueras, L., & Schutz, A. 2020, *A&A*, **635**, A194
- Mas-Ribas, L., Dijkstra, M., Hennawi, J. F., et al. 2017, *ApJ*, **841**, 19
- Maseda, M. V., Brinchmann, J., Franx, M., et al. 2017, *A&A*, **608**, A4
- Maseda, M. V., Bacon, R., Franx, M., et al. 2018, *ApJ*, **865**, L1
- Maseda, M. V., Bacon, R., Lam, D., et al. 2020, *MNRAS*, **493**, 5120
- Matsuda, Y., Yamada, T., Hayashino, T., et al. 2005, *ApJ*, **634**, L125
- McLinden, E. M., Finkelstein, S. L., Rhoads, J. E., et al. 2011, *ApJ*, **730**, 136
- Meiksin, A. A. 2009, *Rev. Mod. Phys.*, **81**, 1405
- Mitchell, P., Blaizot, J., Cadiou, C., & Dubois, Y. 2021, *MNRAS*, **501**, 5757
- Mo, H. J., Mao, S., & White, S. D. M. 1998, *MNRAS*, **295**, 319
- Moffat, A. F. J. 1969, *A&A*, **3**, 455
- Moster, B. P., Naab, T., & White, S. D. M. 2018, *MNRAS*, **477**, 1822
- Muzahid, S., Schaye, J., Marino, R. A., et al. 2020, *MNRAS*, **496**, 1013
- Nilsson, K. K., Fynbo, J. P. U., Møller, P., Sommer-Larsen, J., & Ledoux, C. 2006, *A&A*, **452**, L23
- Okamoto, T., Gao, L., & Theuns, T. 2008, *MNRAS*, **390**, 920
- Ouchi, M., Shimasaku, K., Akiyama, M., et al. 2008, *ApJS*, **176**, 301
- Ouchi, M., Harikane, Y., Shibuya, T., et al. 2018, *PASJ*, **70**, S13
- Ouchi, M., Ono, Y., & Shibuya, T. 2020, *ARA&A*, **58**, 617
- Piqueras, L., Conseil, S., Shepherd, M., et al. 2017, *ArXiv e-prints* [arXiv:1710.03554]
- Planck Collaboration VI. 2020, *A&A*, **641**, A6
- Rafelski, M., Teplitz, H. I., Gardner, J. P., et al. 2015, *AJ*, **150**, 31
- Rahmati, A., Schaye, J., Pawlik, A. H., & Raičević, M. 2013, *MNRAS*, **431**, 2261
- Rakic, O., Schaye, J., Steidel, C. C., & Rudie, G. C. 2011, *MNRAS*, **414**, 3265
- Richard, J., Bacon, R., Blaizot, J., et al. 2019, *ArXiv e-prints* [arXiv:1906.01657]
- Richard, J., Claeysens, A., Lagattuta, D. J., et al. 2021, *A&A*, **646**, A83
- Rosdahl, J., & Blaizot, J. 2012, *MNRAS*, **423**, 344
- Rosdahl, J., Katz, H., Blaizot, J., et al. 2018, *MNRAS*, **479**, 994
- Salmon, B., Papovich, C., Finkelstein, S. L., et al. 2015, *ApJ*, **799**, 183
- Schaerer, D., Hayes, M., Verhamme, A., & Teyssier, R. 2011, *A&A*, **531**, A12
- Schawinski, K., Koss, M., Berney, S., & Sartori, L. F. 2015, *MNRAS*, **451**, 2517
- Schaye, J. 2006, *ApJ*, **643**, 59
- Schechter, P. 1976, *ApJ*, **203**, 297
- Schmidt, K. B., Wisotzki, L., Urrutia, T., et al. 2019, *A&A*, **628**, A91
- Schreiber, C., Pannella, M., Elbaz, D., et al. 2015, *A&A*, **575**, A74
- Shapley, A. E., Steidel, C. C., Pettini, M., & Adelberger, K. L. 2003, *ApJ*, **588**, 65
- Shibuya, T., Ouchi, M., Nakajima, K., et al. 2014, *ApJ*, **788**, 74
- Shibuya, T., Ouchi, M., Konno, A., et al. 2017, *PASJ*, **70**, S14
- Song, M., Finkelstein, S. L., Gebhardt, K., et al. 2014, *ApJ*, **791**, 3
- Soto, K. T., Lilly, S. J., Bacon, R., Richard, J., & Conseil, S. 2016, *MNRAS*, **458**, 3210
- Starck, J.-L., Murtagh, F. D., & Bijaoui, A. 1998, *Image Processing and Data Analysis* (Cambridge: Cambridge University Press)
- Starck, J.-L., Fadili, J., & Murtagh, F. 2007, *IEEE Trans. Image Process.*, **16**, 297
- Steidel, C. C., Adelberger, K. L., Dickinson, M., et al. 1998, *ApJ*, **492**, 428
- Steidel, C. C., Bogosavljević, M., Shapley, A. E., et al. 2018, *ApJ*, **869**, 123
- Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, *ApJ*, **613**, 898
- Tumlinson, J., Peeples, M. S., & Werk, J. K. 2017, *ARA&A*, **55**, 389
- Umehata, H., Hatsukade, B., Smail, I., et al. 2018, *PASJ*, **70**, 65
- Umehata, H., Fumagalli, M., Smail, I., et al. 2019, *Science*, **366**, 97
- Urrutia, T., Wisotzki, L., Kerutt, J., et al. 2019, *A&A*, **624**, A141
- van der Voort, F., Schaye, J., Altay, G., & Theuns, T. 2012, *MNRAS*, **421**, 2809
- van der Walt, S., Colbert, S. C., & Varoquaux, G. 2010, *Comput. Sci. Eng.*, **13**, 22
- Vanzella, E., Calura, F., Meneghetti, M., et al. 2018, *MNRAS*, **483**, 3618
- Vanzella, E., Caminha, G. B., Rosati, P., et al. 2021, *A&A*, **646**, A57
- Verhamme, A., Schaefer, D., Atek, H., & Tapken, C. 2008, *A&A*, **491**, 89
- Verhamme, A., Garel, T., Ventou, E., et al. 2018, *MNRAS*, **478**, L60
- Weilbacher, P. M., Palsa, R., Streicher, O., et al. 2020, *A&A*, **641**, A28
- Weisz, D. R., Johnson, B. D., & Conroy, C. 2014, *ApJ*, **794**, L3
- White, S. D. M., Frenk, C. S., Davis, M., & Efstathiou, G. 1987, *ApJ*, **313**, 505
- Wisotzki, L., Bacon, R., Blaizot, J., et al. 2016, *A&A*, **587**, A98
- Wisotzki, L., Bacon, R., Brinchmann, J., et al. 2018, *Nature*, **562**, 229
- Yamada, T., Nakamura, Y., Matsuda, Y., et al. 2012, *AJ*, **143**, 79
- Yung, L. Y. A., Somerville, R. S., Popping, G., et al. 2019, *MNRAS*, **490**, 2855
- Zheng, Z.-Y., Malhotra, S., Rhoads, J. E., et al. 2016, *ApJS*, **226**, 23

## Appendix A: GALICS

### A.1. Model description

The GALICS model (Hatton et al. 2003) presented in Garel et al. (2015a) is specifically designed to study the formation and evolution of galaxies in the high redshift Universe. It relies on a cosmological  $N$ -body simulations to follow the hierarchical growth of dark matter structures and on semi-analytic prescriptions to describe the physics of the baryonic component. The simulation was run in a box of  $100 h^{-1}$  cMpc on a side and it assumes a standard cosmology that is consistent with the WMAP-5 results. The simulation contains  $1024^3$  dark matter particles with an individual mass of  $\approx 8.5 \times 10^7 M_\odot$ . Halos are identified if they contain at least 20 dark matter particles which corresponds to a minimum halo mass of about  $10^9 M_\odot$ .

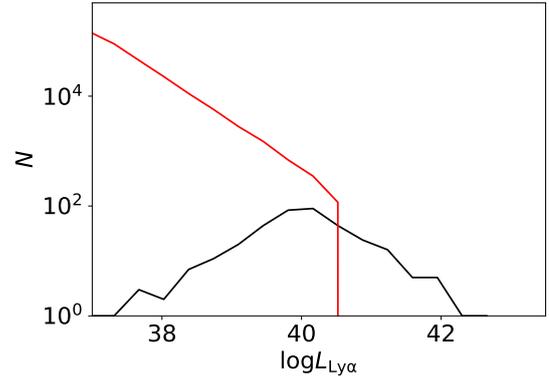
The intrinsic Ly $\alpha$  emission from galaxies is estimated from the production rate of hydrogen-ionizing photons (estimated from the stellar spectral energy distributions) assuming case B recombination. To compute the observed Ly $\alpha$  properties of galaxies, GALICS is combined with the library of radiative transfer simulations of Schaerer et al. (2011) which predicts the escape fraction of Ly $\alpha$  photons through dusty galactic outflows (see Verhamme et al. 2008; Garel et al. 2012). The GALICS Ly $\alpha$  luminosities used in the analysis presented in Sect. 5.2.2 therefore correspond to the dust-attenuated luminosities that emerge from the CGM of galaxies. As shown in Garel et al. (2015a,b), the model was tuned to reproduce various fundamental observational constraints at  $z \approx 3$ –7, including the UV and Ly $\alpha$  luminosity functions, the stellar mass functions and the SFR to stellar mass relation.

For the sake of the present study, we follow the procedure of Garel et al. (2015b) to generate 100 mock lightcones that mimic the geometry and redshift range of the MUSE HUDF survey (that is, a square field of  $\approx 9$  arcmin<sup>2</sup> and  $3 \leq z \leq 6.7$ ). We then reproduce the detection method described in Sect. 3.2 to identify LAE overdensities in the mock fields. In practice, we bin the redshift distribution in slices of 8 cMpc, count the number of detected LAEs using a LAE flux detection limit of  $10^{-18}$  erg s<sup>-1</sup> cm<sup>-2</sup> inside the MXDF and  $10^{-17.5}$  elsewhere, and compute the overdensity  $\delta$  corresponding to each group. In total, this yields 13 253 mock groups, among which 2475 have  $\delta \geq 2$  and  $N_{\text{LAE}} \geq 7$ .

### A.2. Faint-end extrapolation of GALICS mocks with ad hoc LAEs

Due to the limited mass resolution of the cosmological simulation, dark matter halos less massive than  $2 \times 10^9 M_\odot$  are not detected in GALICS. As a consequence, the number density of LAEs can be under-predicted at the faint-end. Our LAE sample is estimated to be incomplete below a Ly $\alpha$  flux limit of  $F_{\text{lim}} \approx 2 \times 10^{-19}$  erg s<sup>-1</sup> cm<sup>-2</sup> (see Garel et al. 2015b).

In an attempt to correct for this effect, we add ad hoc LAEs in our mock fields (for simplicity, the MXDF is assumed to lie exactly at the center of the HUDF field). We do so by extrapolating the Ly $\alpha$  LF between  $L_{\text{lim}}$ , the Ly $\alpha$  luminosity limit corresponding to  $F_{\text{lim}}$  at the redshift of the group, and a minimum Ly $\alpha$  luminosity,  $L_{\text{min}}$ . In practice, the Ly $\alpha$  luminosity limit varies from  $1.6 \times 10^{40}$  erg s<sup>-1</sup> at  $z = 3$  to  $1.1 \times 10^{41}$  erg s<sup>-1</sup> at  $z = 6.7$ . The choice of  $L_{\text{min}}$  is somewhat arbitrary because the Ly $\alpha$  LF is unconstrained in this luminosity range. These considerations are beyond the scope of the present study, therefore we assume a fixed value of  $L_{\text{min}} = 10^{37}$  erg s<sup>-1</sup>, which is more than three orders of magnitude smaller than the faintest observed LAEs.



**Fig. A.1.** Example of the distribution of Ly $\alpha$  luminosities in a GALICS group. The black and red histograms represent the number of true sources and ad hoc sources respectively. The luminosities of the ad hoc sources have been drawn from a power-law distribution with a slope  $\alpha = -1.84$ , in agreement with the Ly $\alpha$  LF from Herenz et al. (2019).

Based on the canonical relation between Ly $\alpha$ , star formation rate and UV luminosity (e.g., Garel et al. 2015a), this value would translate into SFR  $\approx 10^{-5} M_\odot \text{ yr}^{-1}$  and  $M_{\text{UV}} \approx -6$ .

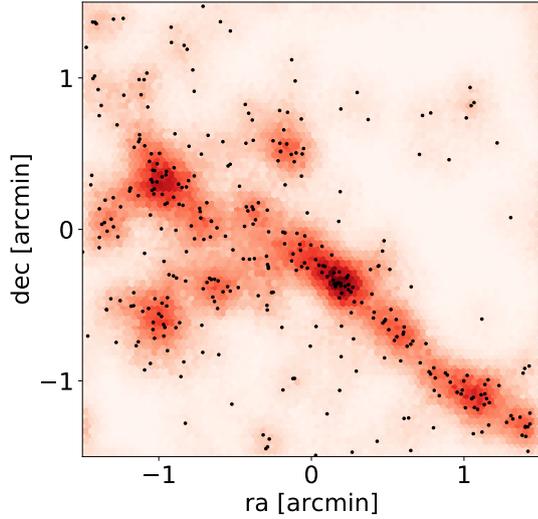
Using the best-fit parameters for the Ly $\alpha$  LF at  $3 < z < 6$  of Herenz et al. (2019),  $\log \Phi_\star = -2.71$ ,  $\log L_\star = 42.66$ ,  $\alpha = -1.84$ , the number density of LAEs is given by the Schechter function,  $\Phi(L)$ , which can be expressed as  $\Phi(L) \approx \frac{\Phi_\star}{L_\star} \left(\frac{L}{L_\star}\right)^\alpha$  for  $L \ll L_\star$ . Then, the extrapolated number density between  $L_{\text{min}}$  and  $L_{\text{lim}}$  is given by  $n_{\text{ext}} = \int_{L_{\text{min}}}^{L_{\text{lim}}} \Phi(L) dL = \frac{\Phi_\star}{(\alpha+1)L_\star^{\alpha+1}} (L_{\text{lim}}^{\alpha+1} - L_{\text{min}}^{\alpha+1})$ . Then, we compute the total number of ad hoc LAEs per group,  $N_{\text{ext}}$ , as the product of  $n_{\text{ext}}$  by the comoving volume of the group,  $V_{\text{gp}}$ , and the group overdensity,  $\delta$ . From there, we randomly draw  $N_{\text{ext}}$  luminosities between  $L_{\text{min}}$  and  $L_{\text{lim}}$  according to a power-law distribution with a slope  $\alpha$ .

As an example, we present the result of this procedure in Fig. A.1 which shows the luminosity distribution of LAEs in the mock GALICS group 12195. This group is located at  $z = 4.96$  and has an overdensity  $\delta = 6.5$ . The black histogram corresponds to the number of GALICS LAEs while the red curve shows the extrapolated distribution of ad hoc LAEs between  $L_{\text{min}}$  and  $L_{\text{lim}}$ . In this case, there are 361 true sources in the initial GALICS catalogue. As a result of the extrapolation between  $L_{\text{min}}$  and  $L_{\text{lim}}$ , 328 049 ad hoc sources are added to the catalog. We note that GALICS sources that are fainter than  $L_{\text{lim}}$  are discarded from the final catalog such that we keep only true LAEs at  $L > L_{\text{lim}}$  and only ad hoc LAEs at  $L \leq L_{\text{lim}}$ .

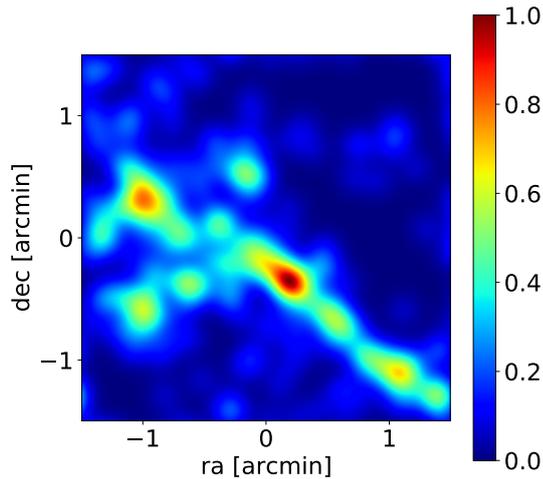
### A.3. Expected faint-end slope index

While GALICS cannot reproduce the full population of very faint LAEs because of the mass resolution limit, galaxies with Ly $\alpha$  luminosities  $L_{\text{Ly}\alpha} = 10^{37}$ – $10^{39}$  erg s<sup>-1</sup> do exist in the current simulation (see Fig. A.1). These are mostly located in young halos at the mass resolution threshold that recently appeared in the simulation, while a smaller fraction ( $\approx 25\%$ ) correspond to satellites sitting in more massive halos. From scaling arguments, we expect that the faint-end slope index is on the order  $-1.8$ , which is consistent with the value assumed above.

The origin of this relation comes from the Kennicutt law for star formation used in GALICS (see Garel et al. 2012, 2015a) which can be expressed as SFR  $\propto M_{\text{gas}}^{1.4} / r_{\text{gal}}^{0.8}$  where  $M_{\text{gas}}$  is the gas mass reservoir in the galaxy and  $r_{\text{gal}}$  is the galaxy disk



**Fig. A.2.** Projected map of mock LAEs in the MOSAIC. The black dots correspond to the true GALICS sources. The red bins show the positions of the ad hoc sources that are used to extrapolate the number density of LAEs at the faint end. These ad hoc LAEs are distributed according to the spatial clustering of the true sources (see text).



**Fig. A.3.** Two-dimensional image of the spatial distribution of GALICS sources in the MOSAIC convolved with a Gaussian ( $\sigma = 0.1'$ ). This map is used as a two-dimensional probability distribution function to draw the positions of ad hoc LAEs in our mock GALICS groups. The color-code shown in the colorbar gives the probability per pixel.

radius. Cooling operates on short timescales in low-mass halos at high redshift, making the baryonic gas available for star formation before supernova feedback, and the  $\lambda$  parameter that drives the conservation of angular momentum in DM halos has a narrow distribution that peaks around 0.05 (Mo et al. 1998). Thus, assuming  $M_{\text{gas}} = f_b M_{\text{vir}}$  (where  $f_b$  is the universal baryonic fraction),  $r_{\text{gal}} = \lambda R_{\text{vir}}$  (where  $R_{\text{vir}}$  is the halo virial radius) and  $R_{\text{vir}} \propto M_{\text{vir}}^{1/3}$ , the star formation rate is expected to scale with the halo virial mass as  $\text{SFR} \propto M_{\text{vir}}^{1.13}$ .

Linking the SFR with the  $\text{Ly}\alpha$  luminosity and neglecting the effect of dust in such faint galaxies (see Sect. 5.2.1), we can write the  $\text{Ly}\alpha$  luminosity as  $L_{\text{Ly}\alpha} \propto \text{SFR} \propto M_{\text{vir}}^{1.13}$ . With a low-mass function of halos  $\phi(M)dM \propto M^{-2}dM$ , (for  $\Lambda\text{CDM}$ ), this relation gives a low luminosity LF:

$$\phi(L_{\text{Ly}\alpha})dL_{\text{Ly}\alpha} \propto L_{\text{Ly}\alpha}^{-1.77}. \quad (\text{A.1})$$

So, the choice of connecting an analytical low luminosity LF with a typical slope  $-1.8$  to the GALICS  $\text{Ly}\alpha$  LF at higher luminosity is consistent with the assumption on the SFR in the simulation and with expectations from first principles.

#### A.4. Spatial distribution of ad hoc LAEs

Once the fake  $\text{Ly}\alpha$  luminosities have been computed, we then have to generate the spatial positions of our ad hoc sources. On the one hand, the redshifts are simply drawn randomly between the minimum and maximum redshifts of the group. On the other hand, the celestial coordinates of ad hoc sources are determined from the spatial clustering of GALICS sources within the MOSAIC. Here, we make the assumption that faint sources follow a clustering pattern similar to that of bright galaxies, rather than being purely randomly distributed. In practice, we start by making a projected map of GALICS objects within the MOSAIC and we apply a two-dimensional Gaussian smoothing to the image, assuming a standard deviation  $\sigma$  of  $0.1'$ . Although this value is arbitrary, it is justified by the fact that  $0.1'$  corresponds roughly to the virial radius of the typical dark matter halos in our MOSAIC groups in the redshift range of our survey ( $M_{\text{vir}} \approx 10^{11} M_{\odot}$ ; see Sect. 3.3). Then, the smoothed image is normalized such that it can be used as a two-dimensional probability distribution function for the spatial sampling of ad hoc LAEs. For each object, we draw random sky coordinates and implement a rejection method to sample our two-dimensional distribution. We repeat this process until all ad hoc sources in the field,  $N_{\text{ext}}$ , have been successfully sampled.

Figures A.2 and A.3 illustrate the successive steps of our spatial sampling procedure applied to the same mock group as in the previous section. The initial projected map of GALICS LAEs within the MOSAIC is represented by the black dots in Fig. A.2. Figure A.3 depicts the corresponding two-dimensional probability distribution function obtained from the Gaussian smoothing. Finally, the red dots in Fig. A.2 show the resulting spatial distribution of ad hoc LAEs.

## Appendix B: Galaxy halo model

The surface brightness (SB) of the galaxy plus halo is modeled as a sum of two circular, two-dimensional exponential distributions, following Wisotzki et al. (2016) and Leclercq et al. (2017) prescription:

$$\text{SB}(r) = F_{\text{Ly}\alpha} \left[ (1 - X_h) e^{-r/r_g} + X_h e^{-r/r_h} \right],$$

with  $F_{\text{Ly}\alpha}$  the total  $\text{Ly}\alpha$  flux given by GALICS,  $X_h$  the fraction of flux in the halo, and the two scale lengths  $r_g$ ,  $r_h$  of, respectively, the galaxy and the halo.

$X_h$  is randomly drawn from a skewed normal distribution (Ashour & Abdel-hameed 2010) fitted to the Leclercq et al. (2017) measurements of 145  $\text{Ly}\alpha$  extended halos from the MUSE observations. The parameters of the distribution are  $\text{loc} = 0.914$ ,  $\text{scale} = 0.298$ , and  $\alpha = -4.049$ , where  $\text{loc}$  and  $\text{scale}$  are the mean and standard deviation of the Normal distribution and  $\alpha$  the skewness parameter.

The galaxy scale length  $r_g$  (in kpc) is derived from the  $M_{\text{uv}}$  GALICS value, using the linear regression measured from the Leclercq et al. (2017) data (Fig. 12 top).

$$r_g = -a_g M_{\text{uv}} + b_g,$$

with  $a_g = 0.080 \pm 0.017$  and  $b_g = -1.012 \pm 0.315$ . The measured regression errors are used to produce  $r_g$  random values compatible with the measured dispersion.

Similarly the  $r_h$  halo scale lengths (in kpc) is derived using the computed linear regression of Halo flux (in  $\text{erg s}^{-1} \text{cm}^{-2}$ ) versus scale length (Leclercq et al. 2017, Fig. 8).

$$r_h = a_h \log F_{\text{Ly}\alpha} + b_h,$$

with  $a_h = 0.066 \pm 0.017$  and  $b_h = 3.477 \pm 0.264$ .

The surface brightness is truncated at the virial radius derived from the GALICS halo mass.

We note that for the low-luminosity sources below the mass cut of GALICS and extrapolated from the  $\text{Ly}\alpha$  luminosity func-

tion (see Sect. 5.2.2), we do not have UV magnitudes or halo masses. In those cases we use  $M_{\text{UV}} = -15$  and  $M_h = 10^8 M_\odot$  to infer a value of the galaxy scale length and the virial radius. In practice these ultra faint galaxies are smaller than the MXDF PSF and can be considered as point sources.

## Appendix C: Additional tables

**Table C.1.** Overdensities of  $\text{Ly}\alpha$  emitters in the HUDF.

ID	$z$	$\delta$	$\Delta$	$N_s$	SB	$V$	$N_l$	$N_x$	$N_r$	$\log L_{\text{Ly}\alpha}$	$\log L_{\text{Ly}\alpha}^T$	$\log E_{\text{Ly}\alpha}$	$\log M_\star$	$\log M_\star^T$	$\log \text{SFR}$	$\log \text{SFR}^T$	$\log \text{Age}$
1	2.9966	4.6	0.096	9	7.2	259.7	26	13	18	$41.69 \pm 0.04$	$43.10 \pm 0.01$	$1.41 \pm 0.16$	$9.51^{+0.18}_{-0.29}$	$10.77^{+0.05}_{-0.05}$	$0.58^{+0.15}_{-0.31}$	$1.83^{+0.04}_{-0.06}$	$8.71^{+0.18}_{-0.32}$
2	3.0657	2.1	0.072	9	6.7	290.8	13	6	9	$41.93 \pm 0.03$	$43.05 \pm 0.01$	$1.41 \pm 0.07$	$10.69^{+0.31}_{-0.63}$	$11.65^{+0.13}_{-0.13}$	$1.10^{+0.48}_{-0.45}$	$2.05^{+0.22}_{-0.11}$	$8.69^{+0.21}_{-0.26}$
3	3.0868	3.1	0.066	10	7.1	258.1	17	4	14	$41.73 \pm 0.04$	$42.96 \pm 0.01$	$1.78 \pm 0.18$	$9.26^{+0.27}_{-1.10}$	$10.41^{+0.09}_{-0.12}$	$0.48^{+0.35}_{-0.85}$	$1.62^{+0.12}_{-0.11}$	$8.69^{+0.17}_{-0.29}$
4	3.0989	2.2	0.047	9	6.4	257.9	12	3	6	$41.56 \pm 0.06$	$42.64 \pm 0.02$	$1.62 \pm 0.10$	$9.81^{+0.15}_{-0.27}$	$10.59^{+0.07}_{-0.09}$	$0.85^{+0.25}_{-0.25}$	$1.63^{+0.12}_{-0.08}$	$8.71^{+0.17}_{-0.24}$
5	3.1754	2.6	0.051	10	6.2	256.5	14	6	11	$41.67 \pm 0.04$	$42.82 \pm 0.01$	$1.65 \pm 0.18$	$8.27^{+0.29}_{-0.16}$	$9.31^{+0.11}_{-0.14}$	$-0.57^{+0.15}_{-0.14}$	$0.47^{+0.05}_{-0.04}$	$8.60^{+0.20}_{-0.31}$
6	3.1924	3.2	0.066	9	5.9	288.2	19	11	13	$41.52 \pm 0.06$	$42.80 \pm 0.01$	$1.59 \pm 0.10$	$9.73^{+0.16}_{-0.11}$	$10.84^{+0.05}_{-0.03}$	$0.81^{+0.29}_{-0.45}$	$1.92^{+0.10}_{-0.09}$	$8.66^{+0.20}_{-0.35}$
7	3.3344	3.2	0.063	10	5.5	285.1	18	4	12	$41.64 \pm 0.05$	$42.89 \pm 0.01$	$1.48 \pm 0.14$	$9.06^{+0.18}_{-0.78}$	$10.14^{+0.06}_{-0.12}$	$0.38^{+0.24}_{-\infty}$	$1.46^{+0.08}_{-0.16}$	$8.64^{+0.17}_{-0.27}$
8	3.4149	2.1	0.045	10	5.7	314.6	13	4	9	$41.66 \pm 0.05$	$42.78 \pm 0.01$	$1.54 \pm 0.08$	$8.34^{+0.12}_{-0.25}$	$9.29^{+0.04}_{-0.07}$	$-0.44^{+0.12}_{-0.13}$	$0.52^{+0.04}_{-0.04}$	$8.63^{+0.18}_{-0.33}$
9	3.4351	3.6	0.071	11	5.9	282.6	20	7	15	$41.75 \pm 0.03$	$43.05 \pm 0.01$	$1.55 \pm 0.08$	$8.78^{+0.04}_{-0.52}$	$9.95^{+0.01}_{-0.09}$	$-0.30^{+0.16}_{-0.11}$	$0.88^{+0.05}_{-0.03}$	$8.64^{+0.17}_{-0.26}$
10	3.4703	2.8	0.054	10	5.6	313.1	17	3	15	$41.69 \pm 0.04$	$42.92 \pm 0.01$	$1.63 \pm 0.11$	$9.20^{+0.19}_{-0.64}$	$10.37^{+0.06}_{-0.10}$	$0.35^{+0.22}_{-0.32}$	$1.53^{+0.07}_{-0.06}$	$8.68^{+0.15}_{-0.26}$
11	3.4998	1.9	0.036	11	5.9	281.1	10	2	7	$41.58 \pm 0.06$	$42.58 \pm 0.02$	$1.70 \pm 0.11$	$8.25^{+0.44}_{-0.43}$	$9.10^{+0.22}_{-0.37}$	$-0.42^{+0.14}_{-0.37}$	$0.43^{+0.06}_{-0.11}$	$8.54^{+0.20}_{-0.32}$
12	3.5595	3.4	0.057	11	6.2	279.5	18	2	16	$41.88 \pm 0.38$	$43.13 \pm 0.09$	$1.61 \pm 0.12$	$9.46^{+0.16}_{-0.30}$	$10.67^{+0.05}_{-0.06}$	$0.44^{+0.30}_{-0.45}$	$1.65^{+0.10}_{-0.08}$	$8.62^{+0.18}_{-0.34}$
13	3.6048	4.4	0.061	11	5.9	278.4	23	9	15	$41.74 \pm 0.04$	$43.11 \pm 0.01$	$1.53 \pm 0.15$	$9.32^{+0.44}_{-\infty}$	$10.47^{+0.17}_{-0.17}$	$0.87^{+0.24}_{-\infty}$	$2.02^{+0.08}_{-0.22}$	$8.57^{+0.26}_{-0.73}$
14	3.7032	3.9	0.069	11	6.0	306.5	22	4	18	$41.77 \pm 0.04$	$43.12 \pm 0.01$	$1.74 \pm 0.13$	$8.79^{+0.33}_{-1.13}$	$10.05^{+0.10}_{-0.11}$	$0.23^{+0.49}_{-1.06}$	$1.48^{+0.18}_{-0.11}$	$8.53^{+0.21}_{-0.38}$
15	3.7245	2.7	0.052	12	6.2	305.8	15	5	11	$41.79 \pm 0.03$	$42.97 \pm 0.01$	$1.72 \pm 0.14$	$8.62^{+0.35}_{-\infty}$	$9.62^{+0.14}_{-0.19}$	$-0.08^{+0.30}_{-0.30}$	$0.92^{+0.12}_{-0.07}$	$8.54^{+0.26}_{-0.87}$
16	4.0469	2.0	0.034	13	5.7	325.7	11	3	8	$41.76 \pm 0.05$	$42.80 \pm 0.02$	$1.43 \pm 0.08$	$8.66^{+0.37}_{-0.86}$	$9.51^{+0.18}_{-0.17}$	$0.24^{+0.33}_{-0.29}$	$1.09^{+0.16}_{-0.09}$	$8.44^{+0.22}_{-0.39}$
17	4.2714	2.4	0.035	13	5.0	347.0	13	4	8	$41.68 \pm 0.05$	$42.79 \pm 0.01$	$1.84 \pm 0.12$	$8.15^{+0.39}_{-0.37}$	$9.05^{+0.18}_{-0.10}$	$-0.31^{+0.22}_{-0.28}$	$0.59^{+0.09}_{-0.08}$	$8.41^{+0.22}_{-0.42}$
18	4.4684	3.0	0.043	14	4.6	368.0	16	4	10	$41.55 \pm 0.07$	$42.75 \pm 0.02$	$1.49 \pm 0.12$	$8.37^{+0.39}_{-\infty}$	$9.37^{+0.17}_{-0.20}$	$0.09^{+0.38}_{-\infty}$	$1.09^{+0.16}_{-0.17}$	$8.38^{+0.23}_{-0.52}$
19	4.5109	5.0	0.074	13	4.5	366.4	26	5	15	$41.73 \pm 0.05$	$43.15 \pm 0.01$	$1.65 \pm 0.13$	$8.89^{+0.45}_{-\infty}$	$10.07^{+0.17}_{-0.18}$	$0.57^{+0.46}_{-\infty}$	$1.75^{+0.17}_{-0.13}$	$8.34^{+0.29}_{-0.84}$
20	4.8093	3.7	0.047	15	4.9	381.7	18	1	13	$41.84 \pm 0.05$	$43.10 \pm 0.01$	$1.53 \pm 0.10$	$8.46^{+0.38}_{-0.88}$	$9.57^{+0.14}_{-0.12}$	$0.03^{+0.34}_{-0.70}$	$1.14^{+0.12}_{-0.11}$	$8.38^{+0.22}_{-0.45}$
21	4.9433	2.9	0.035	16	4.9	376.1	13	2	8	$42.14 \pm 0.02$	$43.25 \pm 0.01$	$1.92 \pm 0.10$	$8.56^{+0.19}_{-\infty}$	$9.46^{+0.08}_{-0.23}$	$0.14^{+0.13}_{-0.18}$	$1.05^{+0.05}_{-0.06}$	$8.30^{+0.27}_{-0.86}$
22	5.7828	4.9	0.041	18	6.3	415.5	16	10	6	$42.03 \pm 0.04$	$43.23 \pm 0.01$	$1.90 \pm 0.09$	$8.69^{+0.33}_{-0.35}$	$9.46^{+0.16}_{-0.11}$	$0.52^{+0.19}_{-\infty}$	$1.30^{+0.09}_{-0.23}$	$8.22^{+0.27}_{-0.65}$

**Notes.** ID: Group ID.  $z$ : redshift.  $\delta$ : overdensity factor.  $\Delta$ : group density in  $\text{cMpc}^{-3}$ .  $N_s$ : number of wavelength slices. SB limiting surface brightness ( $1\sigma$ , 1 arcsec<sup>2</sup>) in  $10^{-20} \text{erg s}^{-1} \text{cm}^{-2} \text{arcsec}^{-2}$  unit.  $V$ : volume in  $\text{cMpc}^3$ .  $N_l$ : number of LAE.  $N_x$ : number of LAE within MXDF field.  $N_r$ : number of LAE with HST counterpart.  $L_{\text{Ly}\alpha}$ :  $\text{Ly}\alpha$  luminosity (group average value) in  $\text{erg s}^{-1}$  unit.  $L_{\text{Ly}\alpha}^T$ :  $\text{Ly}\alpha$  luminosity (group total value).  $E_{\text{Ly}\alpha}$ :  $\text{Ly}\alpha$  Equivalent width in  $\text{\AA}$ .  $M_\star$ : stellar mass in solar mass (group average value) in  $M_\odot$  unit.  $M_\star^T$ : stellar mass in solar mass (group total value). SFR: SFR (group average value) in  $M_\odot \text{yr}^{-1}$  unit.  $\text{SFR}^T$ : SFR (group total value). Age: age in year. Asymmetric error propagation was performed with the public code developed by Laursen et al. (2019).

**Table C.2.** Ly $\alpha$  diffuse emission in overdensities.

ID	Conf	$P_{\text{dif}}$	$F_{\text{dif}}$	$S/N_{\text{dif}}$	$\log L_{\text{dif}}$	$SB_{\text{dif}}$	$S_{\text{dif}}$	$F_{\text{comp}}$	$S/N_{\text{comp}}$	$\log L_{\text{comp}}$	$S_{\text{comp}}$	$F_{\text{fil}}$	$S/N_{\text{fil}}$	$\log L_{\text{fil}}$	$S_{\text{fil}}$
1	0	0.200	21.9 ± 3.1	7.0	42.25 ± 0.06	8.1 ± 1.3	271.6	30.7 ± 2.1	14.7	42.39 ± 0.03	133.9	52.6 ± 3.7	14.2	42.63 ± 0.03	405.5
2	1	0.012	40.8 ± 4.2	9.7	42.54 ± 0.04	7.2 ± 0.8	569.0	21.6 ± 1.9	12.4	42.27 ± 0.04	113.6	62.4 ± 4.6	14.1	42.73 ± 0.03	682.7
3	0	0.906	10.9 ± 2.6	4.2	41.98 ± 0.10	4.8 ± 1.2	226.9	11.7 ± 1.4	8.3	42.01 ± 0.05	62.4	22.5 ± 3.2	7.1	42.29 ± 0.06	289.3
4	0	0.622	15.3 ± 2.8	5.4	42.13 ± 0.08	5.2 ± 1.0	294.6	7.7 ± 1.4	6.2	41.83 ± 0.08	75.9	22.9 ± 3.3	7.4	42.30 ± 0.06	370.5
5	1	0.020	44.2 ± 4.9	9.1	42.61 ± 0.05	5.2 ± 0.6	856.5	23.7 ± 1.9	12.5	42.34 ± 0.03	152.3	67.9 ± 5.1	13.3	42.80 ± 0.03	1008.8
6	0	0.301	24.7 ± 3.8	6.5	42.37 ± 0.07	4.0 ± 0.6	624.0	52.9 ± 2.3	23.4	42.70 ± 0.02	232.3	77.7 ± 4.3	18.0	42.86 ± 0.02	856.2
7	0	0.523	21.1 ± 3.7	5.8	42.34 ± 0.08	3.3 ± 0.6	645.3	13.6 ± 1.7	7.9	42.15 ± 0.06	156.0	34.7 ± 4.0	8.6	42.56 ± 0.05	801.4
8	0	0.957	5.4 ± 1.5	3.6	41.78 ± 0.12	3.9 ± 1.1	139.2	5.3 ± 0.9	5.8	41.77 ± 0.07	53.1	10.8 ± 1.7	6.3	42.08 ± 0.07	192.3
9	0	0.219	18.7 ± 2.7	6.8	42.32 ± 0.06	4.6 ± 0.7	404.9	14.7 ± 1.4	10.6	42.22 ± 0.04	110.7	33.4 ± 3.1	10.7	42.57 ± 0.04	515.6
10	0	0.997	0.4 ± 0.2	1.8	40.67 ± 0.24	11.2 ± 7.4	3.6	-0.0 ± 0.4	3.2		10.4	0.4 ± 0.4	3.7	40.67 ± 0.46	14.0
11	0	0.858	6.6 ± 1.5	4.4	41.89 ± 0.10	5.0 ± 1.2	132.5	1.6 ± 0.7	3.7	41.27 ± 0.18	24.2	8.2 ± 1.6	5.5	41.98 ± 0.09	156.7
12	0	0.457	11.6 ± 2.0	6.0	42.15 ± 0.07	5.4 ± 1.0	214.1	3.0 ± 1.0	6.8	41.57 ± 0.14	54.1	14.7 ± 2.3	8.0	42.25 ± 0.07	268.2
13	0	0.451	15.8 ± 2.6	6.0	42.30 ± 0.07	3.9 ± 0.7	406.1	12.0 ± 1.4	8.4	42.18 ± 0.05	109.8	27.8 ± 3.2	8.8	42.54 ± 0.05	516.0
14	0	0.914	5.3 ± 1.3	4.1	41.85 ± 0.11	5.2 ± 1.4	100.0	10.1 ± 1.1	8.8	42.13 ± 0.05	83.3	15.3 ± 1.7	9.0	42.31 ± 0.05	183.4
15	0	0.999	0.1 ± 0.1	0.8	40.14 ± 0.53	10.7 ± 15.1	1.0	2.1 ± 0.4	5.0	41.46 ± 0.09	11.7	2.2 ± 0.5	4.8	41.48 ± 0.09	12.6
16	0	0.704	6.9 ± 1.4	5.1	42.06 ± 0.09	5.3 ± 1.1	130.1	1.6 ± 0.4	3.6	41.41 ± 0.12	14.3	8.4 ± 1.4	6.0	42.14 ± 0.07	144.4
17	1	0.012	22.3 ± 2.3	9.7	42.62 ± 0.04	4.5 ± 0.5	500.8	14.0 ± 1.2	12.2	42.42 ± 0.04	157.9	36.3 ± 2.7	13.8	42.83 ± 0.03	658.7
18	1	0.057	16.4 ± 2.0	8.1	42.54 ± 0.05	3.7 ± 0.5	443.5	4.5 ± 0.8	5.9	41.97 ± 0.08	68.4	20.8 ± 2.3	9.4	42.64 ± 0.05	511.9
19	1	0.002	30.1 ± 2.3	12.8	42.81 ± 0.03	4.9 ± 0.4	614.9	11.8 ± 1.0	11.7	42.40 ± 0.04	120.9	41.9 ± 2.4	17.1	42.95 ± 0.03	735.8
20	0	0.172	19.2 ± 2.7	7.0	42.68 ± 0.06	3.2 ± 0.5	597.7	17.4 ± 1.5	11.7	42.64 ± 0.04	194.2	36.7 ± 3.3	11.2	42.96 ± 0.04	791.9
21	0	0.126	14.6 ± 2.0	7.3	42.59 ± 0.06	4.1 ± 0.6	356.2	3.2 ± 0.6	5.0	41.93 ± 0.09	40.6	17.8 ± 2.0	8.7	42.67 ± 0.05	396.8
22	0	0.180	10.5 ± 1.5	6.9	42.60 ± 0.06	4.8 ± 0.8	217.9	7.1 ± 1.0	7.6	42.44 ± 0.06	95.2	17.6 ± 1.8	9.8	42.83 ± 0.04	313.1

**Notes.** ID: Group ID. Conf: confidence.  $P_{\text{dif}}$ :  $P$ -value for the diffuse component,  $F_c$ : Ly $\alpha$  flux in  $10^{-18}$  erg s $^{-1}$  cm $^{-2}$  for the component  $c$ ,  $S/N_c$ : Flux S/N for the component  $i$ ,  $\log L_c$ : Log of Ly $\alpha$  luminosity in erg s $^{-1}$  for the component  $i$ .  $SB_{\text{dif}}$ : Average surface brightness of the diffuse Ly $\alpha$  emission in  $10^{-20}$  erg s $^{-1}$  cm $^{-2}$  arcsec $^{-2}$ .  $S_c$ : Surface in arcsec $^2$  for the component  $c$ . Components ( $c$ ), dif: diffuse emission, comp: compact source emission, fil: full filament.