

## **ALMA Reveals a Large Overdensity and Strong Clustering of Galaxies in Quasar Environments at $z \sim 4$**

García-Vergara, Cristina; Rybak, Matus; Hodge, Jacqueline; Hennawi, Joseph F.; Decarli, Roberto; González-López, Jorge; Arrigoni-Battaia, Fabrizio; Aravena, Manuel; Farina, Emanuele P.

**DOI**

[10.3847/1538-4357/ac469d](https://doi.org/10.3847/1538-4357/ac469d)

**Publication date**

2022

**Document Version**

Final published version

**Published in**

Astrophysical Journal

**Citation (APA)**

García-Vergara, C., Rybak, M., Hodge, J., Hennawi, J. F., Decarli, R., González-López, J., Arrigoni-Battaia, F., Aravena, M., & Farina, E. P. (2022). ALMA Reveals a Large Overdensity and Strong Clustering of Galaxies in Quasar Environments at  $z \sim 4$ . *Astrophysical Journal*, 927(1), Article 65.  
<https://doi.org/10.3847/1538-4357/ac469d>

**Important note**

To cite this publication, please use the final published version (if applicable).  
Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights.  
We will remove access to the work immediately and investigate your claim.



# ALMA Reveals a Large Overdensity and Strong Clustering of Galaxies in Quasar Environments at $z \sim 4$

Cristina García-Vergara<sup>1</sup> , Matus Rybak<sup>1,2</sup> , Jacqueline Hodge<sup>1</sup> , Joseph F. Hennawi<sup>3,1</sup> , Roberto Decarli<sup>4</sup> ,  
Jorge González-López<sup>5,6</sup> , Fabrizio Arrigoni-Battaia<sup>7</sup> , Manuel Aravena<sup>5</sup> , and Emanuele P. Farina<sup>7</sup>

<sup>1</sup> Leiden Observatory, Leiden University, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands

<sup>2</sup> Faculty of Electrical Engineering, Mathematics and Computer Science, TU Delft, The Netherlands

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

<sup>4</sup> INAF—Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, Via Gobetti 93/3, I-40129 Bologna, Italy

<sup>5</sup> Núcleo de Astronomía de la Facultad de Ingeniería y Ciencias, Universidad Diego Portales, Av. Ejército Libertador 441, Santiago, Chile

<sup>6</sup> Las Campanas Observatory, Carnegie Institution of Washington, Casilla 601, La Serena, Chile

<sup>7</sup> Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str 1, D-85748 Garching bei München, Germany

Received 2021 August 20; revised 2021 November 29; accepted 2021 December 23; published 2022 March 4

## Abstract

We present an Atacama Large Millimeter/submillimeter Array (ALMA) survey of CO(4–3) line emitting galaxies in 17 quasar fields at  $z \sim 4$  aimed at performing the first systematic search of dusty galaxies in high- $z$  quasar environments. Our blind search of galaxies around the quasars results in five CO emitters with  $S/N \geq 5.6$  within a projected radius of  $R \lesssim 1.5 h^{-1}$  cMpc and a velocity range of  $\delta v = \pm 1000 \text{ km s}^{-1}$  around the quasar. In blank fields, we expect to detect only 0.28 CO emitters within the same volume, implying a total overdensity of  $17.6^{+11.9}_{-7.6}$  in our fields, and indicating that quasars trace massive structures in the early universe. We quantify this overdensity by measuring the small-scale clustering of CO emitters around quasars, resulting in a cross-correlation length of  $r_{0,QG} = 8.37^{+2.42}_{-2.04} h^{-1}$  cMpc, assuming a fixed slope  $\gamma = 1.8$ . This contradicts the reported mild overdensities ( $\times 1.4$ ) of Ly $\alpha$  emitters (LAEs) in the same fields at scales of  $R \lesssim 7 h^{-1}$  cMpc, which are well described by a cross-correlation length  $3.0^{+1.5}_{-1.4}$  times lower than that measured for CO emitters. We discuss some possibilities to explain this discrepancy, including low star formation efficiency, and excess of dust in galaxies around quasars. Finally, we constrain, for the first time, the clustering of CO emitters at  $z \sim 4$ , finding an autocorrelation length of  $r_{0,CO} = 3.14 \pm 1.71 h^{-1}$  cMpc (with  $\gamma = 1.8$ ). Our work, together with the previous study of LAEs around quasars, traces simultaneously the clustering properties of both optical and dusty galaxy populations in quasars fields, stressing the importance of multiwavelength studies, and highlighting important questions about galaxy properties in high- $z$  dense environments.

*Unified Astronomy Thesaurus concepts:* Quasars (1319); Quasar–galaxy pairs (1316); High-redshift galaxies (734); CO line emission (262); Clustering (1908); Large-scale structure of the universe (902); Submillimeter astronomy (1647)

## 1. Introduction

Theoretical studies and simulations of the formation and evolution of large-scale structure in the universe have been largely developed in the last decades, but observational constraints in this field remain one of the major challenges for astrophysicists. In the standard current paradigm, the cosmological model is parameterized by the Lambda cold dark matter ( $\Lambda$ CDM) model, and the structures grow hierarchically through gravitational instability (e.g., Dodelson 2003; Padmanabhan 2006; Schneider 2015), forming a filamentary structure of dark matter and galaxies. In this framework, the most massive galaxies in the early universe are placed in the most massive dark matter halos (Springel et al. 2005; Angulo et al. 2012), tracing the most massive structures. Since the dark matter halo mass can be directly related to the clustering of a population (Cole & Kaiser 1989; Mo & White 1996; Sheth & Tormen 1999), the generic prediction is that massive galaxies should have a large autocorrelation function.

The population with the largest autocorrelation function in the early universe are quasars, and thus they represent key objects to test structure formation models at early epochs. Quasars are accreting supermassive black holes (SMBHs), and are some of the most luminous sources in the universe. They are observed to peak at  $z \sim 2\text{--}3$  (e.g., Richards et al. 2006) and have been detected up to  $z = 7.5$  (Bañados et al. 2018; Yang et al. 2020). The quasar autocorrelation has been measured as a function of redshift up to  $z \sim 4$ , and it has been found to rise steadily over the redshift range  $0 \lesssim z \lesssim 2$  (Croom et al. 2005; Myers et al. 2006; Porciani & Norberg 2006; Shen et al. 2007; da Ângela et al. 2008) with a clustering similar to that of local galaxies, and to strongly increase between  $z \sim 2$  and  $z \sim 4$  (Shen et al. 2007; White et al. 2012; Eftekharzadeh et al. 2015; Timlin et al. 2018).

Additionally, the quasar autocorrelation function has been shown to get progressively steeper on small ( $R \lesssim 40 h^{-1}$  pkpc) scales at  $z < 3$  (Hennawi et al. 2006), and the small-scale clustering has been shown to also increase from  $z \sim 3$  to  $z \sim 4$  (Shen et al. 2010). The measurements of  $z \sim 4$  quasar clustering result in an autocorrelation length  $r_0 = 24.3 \pm 2.4 h^{-1}$  cMpc (at a fixed slope  $\gamma = 2.0$ ), implying that quasars are the most clustered population at this redshift and that they inhabit very massive ( $M_{\text{halo}} > 6 \times 10^{12} M_{\odot} h^{-1}$ ) dark matter halos (Shen et al. 2007), in



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

agreement with theoretical predictions (Springel et al. 2005; Angulo et al. 2012). These results imply that  $z \sim 4$  quasars are tracers of massive structures, and thus they should be surrounded by a large overdensity of galaxies.

Several efforts have been made to detect such overdensities in quasars' environments, but the results reveal a contradictory and confusing picture. Most of the quasar environment studies at  $z \gtrsim 4$  have been performed at optical wavelengths, and they have been aimed at detecting overdensities of Lyman break galaxies (LBGs) or Ly $\alpha$  emitters (LAEs) around individual quasars. Some of the studied fields show overdensities of galaxies (Stiavelli et al. 2005; Zheng et al. 2006; Kashikawa et al. 2007; Kim et al. 2009; Utsumi et al. 2010; Capak et al. 2011; Swinbank et al. 2012; Morselli et al. 2014; Adams et al. 2015; Farina et al. 2017; Balmaverde et al. 2017; Kikuta et al. 2017; Ota et al. 2018) whereas others exhibit a similar number density of galaxies compared with blank fields (Kim et al. 2009; Willott et al. 2005; Bañados et al. 2013; Husband et al. 2013; Simpson et al. 2014; Mazzucchelli et al. 2017; Goto et al. 2017). It thus remains unclear whether  $z \gtrsim 4$  quasars reside in overdensities, as would be implied by their strong autocorrelation. The contradictory results obtained so far might be a consequence of heterogeneous methodologies, depth, and probed physical scales, together with the low-number statistics and the large cosmic variance expected to affect these studies (Buchner et al. 2019; García-Vergara et al. 2019).

One strategy that has been recently used to overcome the low-number statistics and high cosmic variance is to characterize quasar environments by targeting a larger number of fields and measuring the quasar–galaxy cross-correlation function. Using this technique, García-Vergara et al. (2017) demonstrate that LBGs at scales  $R \lesssim 9 h^{-1}$  cMpc are strongly clustered around  $z \sim 4$  quasars with a cross-correlation length  $r_0 = 8.83^{+1.39}_{-1.51} h^{-1}$  cMpc (at a fixed slope  $\gamma = 2.0$ ) and an overall overdensity of 1.5. They find a very good agreement with the expectations assuming a deterministic bias model, in which quasars and galaxies trace the same underlying dark matter distribution. García-Vergara et al. (2019) studied the environments of 17 bright quasars at  $z \sim 4$  at scales of  $R \lesssim 7 h^{-1}$  cMpc traced by LAEs (with  $\log(L_{\text{Ly}\alpha}[\text{erg s}^{-1}]) \geq 42.61$ ) selected using narrowband (NB) imaging. They find an average LAE overdensity around quasars of 1.4 for the full sample, and a cross-correlation length of  $r_0 = 2.78^{+1.16}_{-1.05} h^{-1}$  cMpc (at a fixed slope  $\gamma = 1.8$ ). Although the cross-correlation was found to be consistent with a power-law shape, indicative of a concentration of LAEs centered on quasars, they find 2.1 times fewer LAEs than the expectation computed by assuming a deterministic bias model.

This work contradicts the findings of recent deeper ( $\log(L_{\text{Ly}\alpha}[\text{erg s}^{-1}]) \geq 42.1$ ) observations of 27  $z = 3\text{--}4.5$  quasar environments performed with the Multi Unit Spectroscopic Explorer (MUSE; Bacon et al. 2010) on the Very Large Telescope (VLT) that revealed a large presence of LAEs (Fossati et al. 2021). They find an overdensity of 4.1 at scales  $R \lesssim 0.6 h^{-1}$  cMpc, and they measure a cross-correlation length  $r_0 = 3.15^{+0.36}_{-0.40} h^{-1}$  cMpc at a fixed slope  $\gamma = 1.8$ .<sup>8</sup>

The aforementioned works statistically demonstrate the existence of galaxy overdensities concentrated around quasars at  $z \sim 4$ , but they are still contradictory in the detection of the ubiquitous large galaxy concentration expected based on the theoretical deterministic bias model. Various possibilities can

be cited to explain this discrepancy. First, most of the previous studies rely on quasar redshifts determined from rest-frame UV-emission lines that are known to be offset from the quasar systemic redshifts (e.g., Richards et al. 2002; Shen et al. 2007, 2016; Coatman et al. 2017). If these offsets are large, it would imply that the search for galaxies has actually been performed at higher or lower radial comoving distances from the quasar, where the clustering signal is vanishingly small, resulting in a lower number density of galaxies, and thus the expected large overdensities would not be detectable.

A second possibility is that the complex physical processes associated with quasar radiation affect the visibility of LAEs in their vicinity. On one hand, ionizing radiation coming from the quasar could suppress star formation in nearby galaxies (e.g., Francis & Bland-Hawthorn 2004; Bruns et al. 2012), making the Ly $\alpha$  line emission and UV continuum less intense—and implying a lower number of detectable galaxies in quasar environments. On the other hand, the ionizing radiation from the quasar could induce Ly $\alpha$  fluorescent emission (Cantalupo et al. 2012), resulting in an increase of the LAE number density. These two effects work in opposite directions, and disentangling them becomes extremely difficult unless the environments are traced by a galaxy population less affected by such feedback effects. Note, however, that only small-scale environment studies would be more impacted by the mentioned effects, since the ionizing radiation is not expected to extend up to the large scales at which most previous quasar environment studies have been performed.

Finally, galaxies in quasar environments could be more dusty, and thus the optical/UV emission would be slightly obscured, making them undetectable at optical wavelengths. This would explain why only deep LAE studies reveal large overdensities (e.g., Fossati et al. 2021), while shallower observations miss a large fraction of them (e.g., García-Vergara et al. 2019). Submillimeter observations, which are less affected by dust, would be required to explore this possibility. Indeed, some serendipitous Atacama Large Millimeter/submillimeter Array (ALMA) detections of dusty galaxies have been recently reported in high- $z$  quasars environments, on scales of  $R < 0.5$  cMpc (Decarli et al. 2017; Trakhtenbrot et al. 2017; Venemans et al. 2020), suggesting a possible strong clustering of dusty galaxies around quasars. However, the quasar–galaxy cross-correlation function has not been constrained.

In summary, incomplete sampling of the galaxy population around quasars could be responsible for the contradictory results obtained so far. Systematic and multiwavelength studies are still required to statistically quantify galaxy overdensities of different galaxy populations in  $z \sim 4$  quasar environments.

In this paper, we present an ALMA survey of CO(4–3) line emitters in the environs of 17 quasars at  $z \sim 4$ , the same quasar sample recently studied for LAE overdensities by García-Vergara et al. (2019), to constrain the small-scale ( $R \lesssim 1.5 h^{-1}$  cMpc) clustering of CO emitters, and ultimately trace simultaneously the clustering properties of both optical and dusty galaxy populations around quasars. This is the first quasar sample observed with ALMA for environmental studies purposes, and thus provides an opportunity to statistically trace—for the first time—the overdensity and clustering of dusty galaxies around high- $z$  quasars.

Additionally, our observations provide a more precise measurement of the quasar redshift, previously determined from rest-frame UV-emission lines, which is important for clustering analyses. We note that, with our observations, we could also trace the number

<sup>8</sup> The  $r_0$  value published in Fossati et al. (2021) has recently been recomputed. Here, we quote the most updated  $r_0$  value, which was kindly provided by the authors via private communication.

density of continuum sources; however, this information is not useful for clustering analysis. Given the roughly constant flux density of galaxies at submillimeter wavelengths across the redshift range  $1 \lesssim z \lesssim 7$  (Blain et al. 2002), we would not distinguish between foreground and background sources, and therefore all the detected sources would be included in the clustering computation. The clustering signal would thus be integrated over large comoving volumes, which would strongly dilute the real clustering signal in the close environment of quasars.

The focus of the present paper is on the CO(4–3) emitter number counts in quasar environments, while two forthcoming papers (C. García-Vergara et al. in preparation) analyze the quasar host galaxy properties and the multiwavelength properties of galaxies in quasar fields, respectively.

The outline of the paper is as follows. In Section 2, we describe the targeted quasar sample, the available ancillary data in the fields, the ALMA observations, and data reduction. In Section 3, we describe the detection and selection of CO(4–3) emitting galaxies and present the final catalog. The galaxy number counts and clustering analysis are presented in Section 4. We discuss and interpret our results in Section 5, and we summarize in Section 6. The Appendix discusses the impact of several factors on our clustering measurement. Throughout the paper, we adopt a  $\Lambda$ CDM cosmological model  $h=0.7$ ,  $\Omega_m=0.30$ ,  $\Omega_\Lambda=0.70$ , and  $\sigma_8=0.8$ , which is consistent with Planck Collaboration et al. (2020). Comoving and proper Mpc are denoted as “cMpc” and “pMpc,” respectively. Magnitudes are given in the AB system (Oke 1974; Fukugita et al. 1995).

## 2. Observations and Data Reduction

### 2.1. Quasar Sample

The quasar sample used in this study is the same sample presented by García-Vergara et al. (2019), who observed 17 quasar fields at optical wavelengths to search for LAEs in  $z \sim 4$  quasar environments. We refer the reader to the aforementioned work for details of the quasar selection strategy. Briefly, the 17 quasars were selected from the Sloan Digital Sky Survey (SDSS; York et al. 2000) and the Baryon Oscillation Spectroscopic Survey (BOSS; Eisenstein et al. 2011; Dawson et al. 2013) quasar catalog (Pâris et al. 2014) to lie within a redshift window of  $z \sim 3.862\text{--}3.879$  (corresponding to  $\delta v = 1066 \text{ km s}^{-1}$  at  $z = 3.87$ ). This redshift window was chosen based on the coverage of the optical narrowband filter used to detect LAEs. The quasar redshifts were determined based on one or more rest-frame UV-emission lines (S IV  $\lambda$  1396, C IV  $\lambda$  1549, and C III]  $\lambda$  1908) and using the calibration of emission-line shifts from Shen et al. (2007) to estimate the systemic redshift. All the quasars have  $1\sigma$  redshift uncertainties  $\lesssim 800 \text{ km s}^{-1}$ . We tabulate the quasar positions, redshifts, and  $i$ -band magnitudes in Table 1.

### 2.2. Ancillary Data

The 17 quasar fields were previously observed (Program ID: 094.A-0900) using the FOCAL Reducer and low-dispersion Spectrograph 2 (FORS2; Appenzeller & Rupprecht 1992) instrument on the VLT in 30 hours of observing time. Optical deep observations were acquired using the narrow band HeI ( $\lambda_c = 5930 \text{ \AA}$ , FWHM =  $63 \text{ \AA}$ ) and the broad bands  $g_{\text{HIGH}}$  ( $\lambda_c = 4670 \text{ \AA}$ , hereafter  $g$ ) and  $R_{\text{SPECIAL}}$  ( $\lambda_c = 6550 \text{ \AA}$ , hereafter  $R$ ) to detect LAEs around the quasars. The total area observed per field was  $6.8 \times 6.8 \text{ arcmin}^2$ , with an image pixel scale of  $0.25''/\text{pix}$ .

**Table 1**  
Optical Properties of the Targeted Quasars

Target ID (SDSS)	R.A. (J2000)	Decl. (J2000)	Redshift	$i$ mag
J0040+1706	00:40:17.426	+17:06:19.78	$3.873 \pm 0.008$	18.91
J0042-1020	00:42:19.748	−10:20:09.53	$3.865 \pm 0.012$	18.57
J0047+0423	00:47:30.356	+04:23:04.73	$3.864 \pm 0.008$	19.94
J0119-0342	01:19:59.553	−03:42:16.51	$3.873 \pm 0.013$	20.49
J0149-0552	01:49:06.960	−05:52:18.85	$3.866 \pm 0.013$	19.80
J0202-0650	02:02:53.765	−06:50:44.54	$3.876 \pm 0.008$	20.64
J0240+0357	02:40:33.804	+03:57:01.59	$3.872 \pm 0.012$	20.03
J0850+0629	08:50:13.457	+06:29:46.91	$3.875 \pm 0.013$	20.40
J1026+0329	10:26:32.976	+03:29:50.63	$3.878 \pm 0.008$	19.74
J1044+0950	10:44:27.798	+09:50:47.98	$3.862 \pm 0.012$	20.52
J1138+1303	11:38:05.242	+13:03:32.61	$3.868 \pm 0.008$	19.10
J1205+0143	12:05:39.550	+01:43:56.52	$3.867 \pm 0.008$	19.37
J1211+1224	12:11:46.935	+12:24:19.08	$3.862 \pm 0.008$	19.97
J1224+0746	12:24:20.658	+07:46:56.33	$3.867 \pm 0.008$	19.08
J1258-0130	12:58:42.118	−01:30:22.75	$3.862 \pm 0.008$	19.58
J2250-0846	22:50:52.659	−08:46:00.22	$3.869 \pm 0.012$	19.44
J2350+0025	23:50:32.306	+00:25:17.23	$3.876 \pm 0.012$	20.61

**Note.** Quasar positions are determined from optical images (SDSS/BOSS quasar catalog; Pâris et al. 2014), redshifts are based on one or more rest-frame UV-emission lines and calibrated to estimate the systemic redshift (see Section 2.1), and magnitudes correspond to the  $i$ -band magnitudes from SDSS.

The medians of the  $5\sigma$  limiting magnitudes reached in the fields were 24.45, 25.14, and 25.81 for the HeI,  $R$ , and  $g$  images, respectively.

A catalog of 25 LAEs within  $R \lesssim 7 h^{-1} \text{ cMpc}$  in the quasar fields is available (García-Vergara et al. 2019). LAEs were selected based on a color selection criteria such that they have a (rest-frame)  $\text{EW}_{\text{Ly}\alpha} > 28 \text{ \AA}$  and a solid detection of  $\text{S/N} \geq 5.0$  in the HeI band. All the LAEs lie approximately within  $\pm 1598 \text{ km s}^{-1}$  (corresponding to  $\pm 18.7 \text{ cMpc}$  at  $z = 3.78$ ) from the quasar systemic redshift, as given by the coverage of the used NB filter. Only two out of 25 LAEs lie within a radius of  $59''$  from the central quasar, which is the area covered by the ALMA observations presented in this study.

### 2.3. ALMA Band 3 Observations

Our observations were carried out during ALMA Cycle 7, as part of the observing program #2019.1.00411.S (PI: C. García Vergara). The data were taken in Band 3 (84–116 GHz) between 2019 October 4 and 2020 March 19. Most targets were observed in a single scheduling block with 45–50 minutes on-source, except for SDSS J0040+1706, SDSSJ1138+1303, SDSSJ1205+0143, and SDSSJ1224+0746, which were observed in two scheduling blocks.

The 17 pointings were centered at the optical coordinates of each quasar (specified in Table 1). The spectral setup consists of four spectral windows (SPWs). One of the SPWs (hereafter SPW0), with a 1.875 GHz bandwidth, was centered on the expected observed frequency of the CO(4–3) emission line from the central quasar, given the quasar systemic redshifts reported in Table 1. The 1.875 GHz bandwidth correspond to  $\delta v \sim 6000 \text{ km s}^{-1}$  (or  $\delta z = 0.098$ ) at the observed frequencies in SPW0. The SPW0 was placed into the lower sideband and configured in Frequency Division Mode (FDM), which is usually used for spectral-line observations, and provides a spectral channel width of 0.488 MHz. Channels were then



**Table 2**  
ALMA Observations and Imaging Summary

Target ID	Dates observed	$N_{\text{ant}}$	Max baseline [m]	Beam size (PA) [arcsec $\times$ arcsec], [deg]	$\sigma_{\text{cont}}$ $\mu\text{Jy beam}^{-1}$	$\sigma_{\text{SPW0}}$ $\text{mJy beam}^{-1}$	$\log L'_{\text{CO}(4-3)}$ [K km s $^{-1}$ pc $^2$ ]
J0040+1706	2020 Mar 17	32	1231 m	$1.21 \times 1.00$ (36)	12.8	0.32	9.78
	2020 Mar 18	39	1231 m				
J0042-1020	2019 Oct 19	49	783 m	$1.42 \times 1.19$ (−85)	8.9	0.23	9.64
J0047+0423	2019 Oct 14	48	783 m	$1.38 \times 1.26$ (−37)	9.0	0.24	9.65
J0119-0342	2019 Oct 19	43	783 m	$1.41 \times 1.22$ (−78)	8.4	0.23	9.64
J0149-0552	2019 Oct 19	49	783 m	$1.39 \times 1.19$ (−65)	8.8	0.24	9.66
J0202-0650	2019 Oct 19	49	783 m	$1.49 \times 1.17$ (−62)	8.5	0.23	9.64
	2019 Oct 14 <sup>a</sup>	43	783 m				
J0240+0357	2019 Oct 4	48	2517 m	$1.24 \times 0.91$ (−53)	14.1	0.35	9.82
J0850+0629	2020 Mar 2	44	783 m	$1.38 \times 1.22$ (49)	10.1	0.25	9.67
J1026+0329	2019 Oct 19	46	783 m	$1.37 \times 1.24$ (−58)	9.8	0.23	9.63
J1044+0950	2020 Mar 2	43	783 m	$1.41 \times 1.30$ (23)	9.5	0.23	9.64
J1138+1303	2020 Mar 16	43	783 m	$1.30 \times 1.17$ (42)	9.3	0.24	9.66
	2020 Mar 19	45	1231 m				
J1205+0143	2019 Oct 19	46	783 m	$1.33 \times 1.23$ (−76)	8.5	0.23	9.64
	2020 Mar 3	43	784 m				
J1211+1224	2020 Mar 15	43	783 m	$1.47 \times 1.20$ (−32)	11.8	0.29	9.74
J1224+0746	2020 Mar 14	40	783 m	$1.38 \times 1.24$ (−40)	8.7	0.23	9.64
	2020 Mar 15	40	783 m				
J1258-0130	2020 Mar 3	44	783 m	$1.46 \times 1.18$ (−46)	10.7	0.30	9.75
J2250-0846	2019 Oct 18	47	783 m	$1.39 \times 1.18$ (−86)	10.4	0.25	9.67
J2350+0025	2019 Oct 14	48	783 m	$1.35 \times 1.24$ (−50)	9.0	0.25	9.67

**Notes.** For each target, we list the dates of observations, the number of antennas used, and the baseline range. The synthesized beam sizes are for naturally weighted images for the final data products. We include the rms noise measured in the dirty continuum images and the rms noise per channel (channel width of 7.8125 MHz, equivalent to  $\sim 25 \text{ km s}^{-1}$  at the observed frequency) measured in the dirty spectral cube created from SPW0 (which is targeting the expected CO(4–3) emission). We also report the limiting luminosity of the CO line at  $5.6\sigma$  (the S/N threshold used in this work), assuming a line width of  $331 \text{ km s}^{-1}$  (see details in Section 4.1). Since the rms is measured in the non-primary beam corrected image, the reported limiting luminosity corresponds to the center of the pointing for each field, but this increases with increasing radius.

<sup>a</sup> Semi-pass observation.

averaged together resulting in a final resolution of 7.8125 MHz (equivalent to  $\sim 25 \text{ km s}^{-1}$  at the observed frequency).

The remaining three SPWs were located at higher frequencies, with a 2.0 GHz bandwidth. One of them (SPW1) was adjacent to SPW0, and the other two (SPW2 and SPW3) were in the upper sideband (at  $\sim 12 \text{ GHz}$  from the lower sideband). SPWs 1, 2, and 3 were configured in Time Division Mode (TDM), which is usually used for continuum observations because it optimizes the continuum sensitivity. This provides a spectral channel width of 15.625 MHz.

Most observations were taken in the C43-4 array configuration with a maximum baseline length of 783 m, but several observations were taken in a more extended configuration, with a maximum baseline length of 1231 or 2517 m. Table 2 summarizes the details of individual observing blocks. The shortest baseline was always 15 m, and the maximum recoverable scale ranged between  $10''$  and  $14''$ . The number of 12 m antennas used ranged between 32 and 49.

The primary beam FWHM is  $66''$  at 94.5 GHz, and the typical size of the synthesized beam is  $1''.4 \times 1''.2$  (see Table 2).

#### 2.4. Imaging

All data processing and imaging was performed using CASA (Common Astronomy Software Applications package, McMullin et al. 2007), version 5.6.1. We process the data using the standard ALMA pipeline, however, some data sets required further manual flagging of problematic antennas. Specifically, we flagged the antennas DA60 and DA58 in J1258-0130, DV04 and DV11 in J0240+0357, DV19 in J1138+1303 (2020

March 19 observation), and DV17 for J1044+0950. The sources were too faint for a successful self-calibration.

For each quasar field, we used CASA’s task `tclean` to create both dirty-image and clean spectral cubes at native frequency resolution. We use the Högbom cleaning algorithm and recalculate the noise per visibility (setting `fastnoise=False`). All images were created using natural weighting to maximize the surface brightness sensitivity. For the cleaned images, we used a stopping threshold of  $1.5\sigma$  without manual masking. We set the pixel size to  $0''.2$ , which ensures that the synthesized beam is sufficiently subsampled. Note that the line search was performed on the dirty cubes, and the clean cubes are only used to extract the detected sources (see Section 3.1).

J0240+0357 was observed with a rather sparse UV-plane coverage, resulting in a noticeably non-Gaussian dirty beam with large sidelobes. To improve the image quality, we reimaged this source using an outer Gaussian taper of  $1''.5$ .

We use the dirty spectral cubes created from SPW0 (the cube at the expected observed frequency of the quasar emission line) without a primary beam correction to compute the rms noise per 7.813 MHz channel throughout the scanned frequency range in a rectangular region that covers most of the region within the primary beam FWHM. The rms noise varies only slightly per channel, (typically  $\sim 3\%$ ), so we compute the median rms noise measured in all the channels of the spectral cube for each field and report these values in Table 2.

Finally, we create a continuum image by combining all four SPWs. For each quasar field, we compute the rms noise in the dirty continuum images (without primary beam correction)

over the same region as for the SPW0 cube and tabulate the values in Table 2. The achieved rms is in agreement with expectations from the ALMA sensitivity calculator. Note that, in our analysis, we do not subtract the continuum from the data.

### 3. Emission Line Catalog

In this study, we focus on the measurement of the small-scale clustering of CO(4–3) emission lines around the central quasar. Therefore, we limit our emission line search to the spectral cube created from SPW0, where the central quasar is expected to be located. The procedure and catalog presented in this section only contain sources detected in this spectral cube, which cover  $\pm 3000 \text{ km s}^{-1}$  around the quasar. At the end of this section, we briefly present the detection of the quasars, and the comparison between their optical and ALMA redshifts (see Section 2.1). We present further details about the quasars in a forthcoming paper (C. García-Vergara et al. in preparation).

#### 3.1. Line Search Algorithm

We perform a blind search for emission lines in the quasar fields (i.e., on the 17 spectral cubes extracted from the SPW0). Since we do not expect to detect extremely bright sources in our cubes, we prefer to use the “dirty” cubes to perform the line search, allowing us to preserve the intrinsic properties of the noise in the cubes. We use FINDCLUMPS, an IRAF-based routine originally developed for the blind search of CO lines performed in the ALMA spectroscopic survey in the Hubble Ultra Deep Field (ASPECS; Walter et al. 2016; Decarli et al. 2019), a 3D survey of gas and dust in distant galaxies. We refer the reader to these works for further details about the algorithm implemented in the FINDCLUMPS routine. Briefly, FINDCLUMPS performs a top-hat convolution using  $N_{\text{chn}}$  consecutive frequency channels of the cube to create a combined image, with  $N_{\text{chn}}$  ranging from 3 up to 19 (corresponding to  $74\text{--}470 \text{ km s}^{-1}$  at  $z=3.87$ ) in steps of two channels. For each combined image, the rms is computed<sup>9</sup> and a search for sources is performed using SExtractor (Bertin & Arnouts 1996), resulting in one source catalog. The signal-to-noise (S/N) ratio of each source in this catalog is computed using the peak flux of the source and the rms computed for the combined image. We only keep sources with  $S/N \geq 3.5$  in the catalog. The search is performed not only within the primary beam FWHM but over the entire area covered by our observations, which is given by a circular area with a radius  $59''$  (at which the telescope sensitivity is 10% of the maximum).

Since the source search is performed repetitively using different numbers of channels, we obtain several individual source catalogs, and we combine these into one final catalog per cube; however, the resulting final catalog contains duplications. For example, if the FWHM of an emission line is equivalent to 19 channels, the algorithm will detect such a line with the maximum S/N in the combined image created by using 19 consecutive frequency channels. However, that source may also be detected in the combined image created by using fewer consecutive channels, although with lower S/N because in this case only part of the total flux is encompassed in the combined image.

<sup>9</sup> We use a rectangular region that covers most of the region within the primary beam FWHM, since the primary beam pattern should be very well-defined, and thus the image before the primary beam correction should be flat, without spurious features, and with (mostly) Gaussian noise.

Since the S/N of a line is maximized when detected combining a number of channels equivalent to the exact width of the line, we look for all the sources that fall within  $1''.5$  (which is similar to the size of the synthesized beam in our images) and  $650 \text{ km s}^{-1}$  ( $0.21 \text{ GHz}$  at  $z=3.87$ ), and we only keep in our final catalogs the source with the highest S/N. This procedure effectively removes the duplications, but we could also be missing some real close pairs of sources (for example mergers). We repeat this line search process on all our cubes, resulting in 17 source catalogs.

#### 3.2. Fidelity

The assessment of the reliability of the sources in our catalogs is crucial to distinguish real detections from spurious detections only caused by noise peaks exceeding our S/N threshold. The reliability of a line can be determined by exploring the noise properties of each cube. Specifically, we apply the same line search algorithm described in Section 3.1 to the negative cubes, and we create final catalogs containing (unphysical) negative lines detected with  $S/N \geq 3.5$ . Since negative lines are only produced by noise fluctuations, this catalog can be used to determine the probability that a positive line detection is due to noise as a function of their S/N. Note, however, that this probability also depends on the line width, such that at a fixed S/N, a source with a wider line has a higher probability to be real than a source with a narrower line (see a detailed discussion in González-López et al. 2019).

To estimate the noise distribution of each data cube, and under the assumption of a Gaussian distribution for the noise, we fit a Gaussian (centered at  $S/N=0$ ) to the S/N histogram of the negative lines for the different width lines. To improve the statistics, we have grouped detections with different widths together before performing the fitting; specifically, we grouped lines with channels width  $N_{\text{chn}}$  3 and 5, 7 and 9, etc. To better take into account the low-number statistics in the tails of the noise distribution, we perform the fitting using a Poisson maximum likelihood estimator.

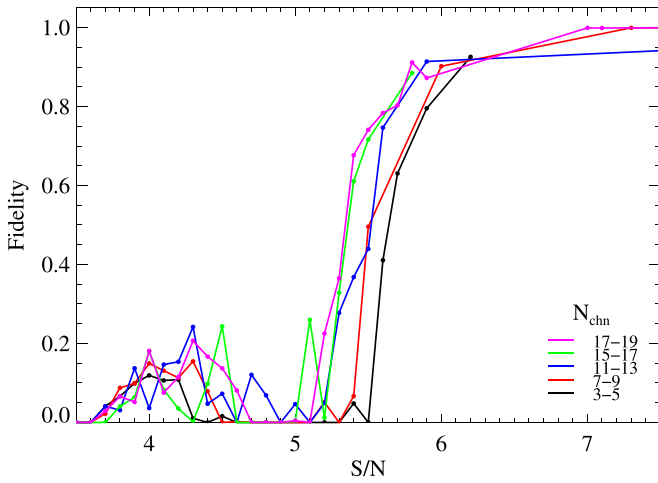
We compute the reliability (or fidelity  $F$ ) of a detection as:

$$F(S/N, N_{\text{chn}}) = 1 - \frac{N_{\text{neg}}(S/N, N_{\text{chn}})}{N_{\text{pos}}(S/N, N_{\text{chn}})}, \quad (1)$$

where  $N_{\text{pos}}(S/N, N_{\text{chn}})$  and  $N_{\text{neg}}(S/N, N_{\text{chn}})$  are the number of positive emission lines and the expected number of negative emission lines, respectively, in each S/N bin. The expected number of negative emission lines is computed using a Gaussian function with the best-fitted parameters found by the fitting process described above.

In some previous works (e.g., Walter et al. 2016; Loiacono et al. 2021), the fidelity has been computed using the cumulative number counts (i.e.,  $N_{\text{neg}}(\geq S/N)$  and  $N_{\text{pos}}(\geq S/N)$ ) instead of the differential number counts. However, if a cumulative approach is used, the computed fidelity at intermediate-S/N values (for instance at  $S/N \sim 4.5$ ) would be completely dominated by the higher-S/N detections (for instance, a detection with  $S/N \sim 9.0$ ), resulting in a boost of the fidelity for lower-S/N detections. To avoid this artificial boost of fidelity of intermediate-S/N sources, here we use  $N_{\text{neg}}$  and  $N_{\text{pos}}$  as the differential numbers of detections in bins of S/N values.

The noise properties varies among the 17 cubes, so we repeat this process for all our 17 cubes individually and determine the fidelity of each detected source in the cubes. For visualization



**Figure 1.** Median fidelity of a line candidate as a function of S/N for different line widths. In our analysis, we use a fidelity computed for each cube separately, but we show the median value in this figure for visualization. Taking into account all the fields together, we find that fidelity 0.8 is typically reached at  $S/N \sim 5.6$ ,  $S/N \sim 5.7$ ,  $S/N \sim 5.8$ , and  $S/N \sim 6.0$  for sources with line widths 15–19 ( $371\text{--}470\text{ km s}^{-1}$ ), 11–13 ( $272\text{--}322\text{ km s}^{-1}$ ), 7–9 ( $173\text{--}223\text{ km s}^{-1}$ ), and 3–5 ( $74\text{--}124\text{ km s}^{-1}$ ) channels, respectively.

purposes only, we computed the median fidelity over the cubes and show this in Figure 1. Since we compute fidelity using the differential number of detections, the fidelity cannot be constrained in S/N bins where no positive lines are detected. For computing the median, we only use the bins with constrained fidelity over the fields (data points in Figure 1) and we linearly interpolate between the S/N values.

### 3.3. Line Catalog

The choice of the fidelity threshold to keep a source in our catalogs is crucial and has to be taken carefully. This choice varies between different studies, depending on the aim of the analysis. Decarli et al. (2019) is aimed to constrain the CO luminosity function (used in our work as a reference for the computation of the background number counts as described in Section 4.1), and they choose a fidelity threshold of 0.2. However, they treat the fidelity as an upper limit and use a Monte Carlo simulation to generate several realizations of their final catalog, in which they randomly assign a fidelity between 0.0 and their upper limits, and they only keep sources if this random fidelity value exceeds their threshold of 0.2. This means that lower-fidelity sources are less likely to be kept in the catalog for each realization.

This approach is adequate for measuring the luminosity function because, even if sources with low fidelity (likely to be spurious) are included in some realizations, the overall statistics of the number counts will stay mostly unchanged (as explicitly checked in Decarli et al. 2019). Additionally, low-fidelity sources are mostly composed of faint sources, only affecting the faint end of the luminosity function, which has little impact on the overall fitting for the number counts.

In the case of clustering measurements, secure detections are required, since all the sources contribute with the same weight to the final measurement, no matter their intrinsic flux. Including spurious detections would have a strong impact on the final measurement, especially for sparse samples. For this reason, we prefer to choose a more conservative fidelity threshold of  $\geq 0.8$ , and we only keep sources fulfilling this

criterion in our final catalogs. We nevertheless explore the impact of the fidelity threshold choice on our clustering results in the Appendix.

From our median fidelity computation (see Figure 1), we find that fidelity 0.8 is typically reached at  $S/N \sim 5.6$ ,  $S/N \sim 5.7$ ,  $S/N \sim 5.8$ , and  $S/N \sim 6.0$ , for sources with line widths 15–19 ( $371\text{--}470\text{ km s}^{-1}$ ), 11–13 ( $272\text{--}322\text{ km s}^{-1}$ ), 7–9 ( $173\text{--}223\text{ km s}^{-1}$ ), and 3–5 ( $74\text{--}124\text{ km s}^{-1}$ ) channels, respectively.

Since the FINDCLUMPS algorithm determines the source S/N using a global rms computed on the primary beam FWHM region of the ALMA pointing, the S/N of the detected candidates located close to the edges of the pointing may be overestimated due to noise fluctuation affecting these regions of the pointing. However, we still could detect reliable sources in these regions, so we include sources within the entire area covered by our observations, but we increase the fidelity threshold up to  $\geq 0.9$  for sources located in the edges of the pointing (where the telescope sensitivity is lower than 20% of the maximum—or equivalently, at radii larger than  $50.14''$  from the central quasar).

Considering all the fields together, our final catalogs contain nine sources (excluding the detection of the central quasar). Note that this is the total number of sources detected in all the SPW0 cubes, but not all of them are necessarily included for the clustering computation (see details in Section 4.1).

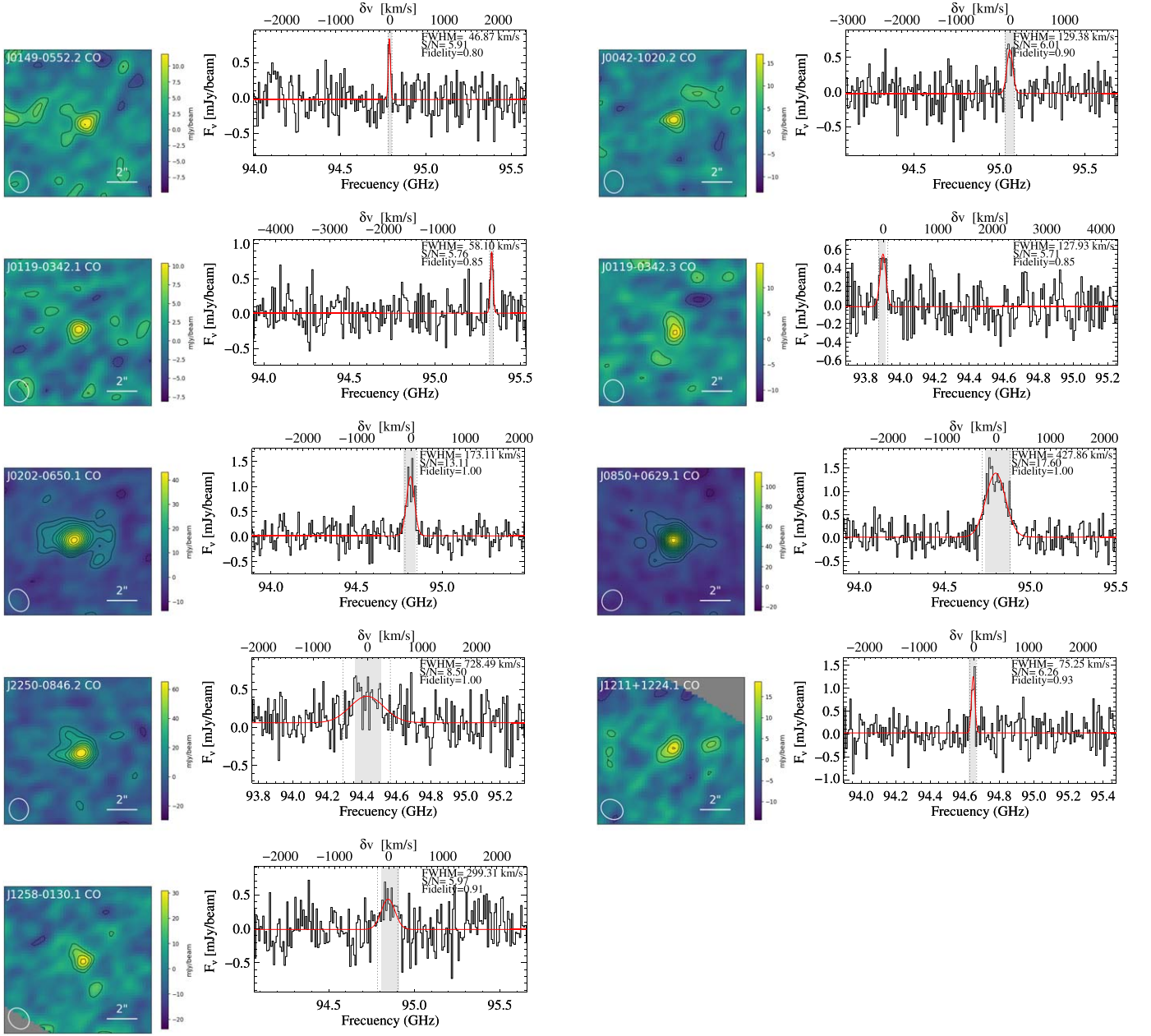
We extract the 1D spectra for all the selected candidates from the clean cubes. At the resolution of our observations, the sources are not significantly resolved. Therefore, we perform the spectral extraction on the brightest pixel of the source, and we perform a Gaussian fit with a flat continuum to constrain the line center ( $\nu_{\text{line}}$ ) and the line standard deviation ( $\sigma_{\text{line}}$ ). We use the `immoments` task from the CASA package to create a 0th moment map for each source. For that, the data cube is integrated over a frequency width given by  $2.8\sigma_{\text{line}}$  centered on  $\nu_{\text{line}}$ , which recovers  $\sim 84\%$  of the line flux.

We measured the total integrated line flux from the 1D spectra by adding the flux from all the channels within  $\pm 1.4\sigma_{\text{line}}$  from the emission line. We note that this may result in a slightly different S/N ratio than the reported from the line search algorithm, since the computation is made over a different number of channels in some cases (see the gray area versus the dotted lines in the 1D spectra shown in Figure 2). The fluxes are not corrected by the primary beam response of the telescope. Note that the integrated line flux of the individual objects is not relevant for the clustering analysis, since clustering is based only on the position of the sources. The only requirement is that the total integrated line flux of all the sources should be always greater than or equal to  $5.6\sigma$  (the S/N threshold used in this work), with  $\sigma$  being the limiting flux of the survey. The limiting flux of the survey is the only relevant flux measurement used in the clustering analysis (see Section 4.1), and it is computed based on the rms of the cubes (see Table 2).

In Figure 2, we show the moment 0 maps and the 1D extracted spectra for all the sources in our catalogs, and the best-fitted parameters for the lines are listed in Table 3. In Table 5, we summarize the number of emission lines in each individual field.

One of the sources (J2250-0846.2) is identified by our algorithm as two close ( $\delta\theta = 0''.4$ ,  $\delta v \sim 625\text{ km s}^{-1}$ ) sources. This could be either a galaxy merger or a galaxy with a two-peaked emission line, with total width FWHM  $> 700\text{ km s}^{-1}$ .





**Figure 2.** *Left:* Line maps for all the sources detected in our survey with fidelity  $\geq 0.8$  (or  $\geq 0.9$ , if they are located at radius  $> 50''$  14 from the pointing center), integrated over a frequency width given by  $2.8\sigma_{\text{line}}$  around the center of the line. Contours start at  $\pm 2\sigma$  and increase in steps of  $1\sigma$ . The white ellipse in the lower left corner shows the FWHM beam size. *Right:* 1D extracted spectra on the brightest pixel of each source. We show the Gaussian plus flat continuum fit as a red curve. The shaded gray area shows the line width as detected by the line search algorithm. The vertical lines show the channels where the integrated line flux was measured, corresponding to a width of  $2.8\sigma_{\text{line}}$  centered on  $\nu_{\text{line}}$ . In each panel, we report the best-fitted FWHM value, the S/N of the line detected by the line search algorithm (which is computed over the gray area shown in the spectra), and the fidelity. Line maps and spectra are both extracted from the cleaned cubes, and they are not corrected by the primary beam response.

In this study, we consider this as a single object, since even in the merger scenario this would likely trace the same dark matter halo. We study this object further in a forthcoming paper (C. García-Vergara et al. in preparation).

Our candidates span a large range of line widths, ranging from  $47 \text{ km s}^{-1}$  to  $728 \text{ km s}^{-1}$ , with three of the lines showing small FWHM values ( $< 100 \text{ km s}^{-1}$ ).<sup>10</sup> To explore whether the existence of such narrow lines is expected and reasonable, we checked the line widths of the 16 CO-emitting galaxies blindly found in the ASPECS survey with fidelity  $> 0.9$  (González-López

et al. 2019). We find that their FWHMs range from  $40 \text{ km s}^{-1}$  to  $617 \text{ km s}^{-1}$ , including two very narrow lines with  $40 \text{ km s}^{-1}$  (with  $S/N = 7.9$ ) and  $50 \text{ km s}^{-1}$  (with  $S/N = 9.5$ ), respectively. Using HST counterparts for the CO emitters detected in ASPECS, Aravena et al. (2019) suggest that these galaxies are likely very face-on, which would explain the detection of such narrow line widths in CO-emitting galaxy surveys.

### 3.4. Line Identification and Contamination

As we targeted quasars with known redshifts, which are expected to be surrounded by large overdensities of galaxies, and given that we are tracing relatively small areas in the sky,

<sup>10</sup> We note that one of these narrow lines is the object J1211+1224.1, which has a LAE counterpart at the same redshift.



**Table 3**  
Properties of the Emission Lines Detected in All the 17 Fields Properties of the Emission Lines Detected in All the 17 Fields

ID	R.A. (J2000) [deg]	Decl. (J2000) [deg]	$\nu_{\text{line}}$ [GHz]	$\text{FWHM}_{\text{line}}$ [km s <sup>-1</sup> ]	S/N	Fidelity	$\delta\theta$ ["]	$\delta v$ [km s <sup>-1</sup> ]	$F$ [mJy km s <sup>-1</sup> ]	Optical Counterpart
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
J0042-1020.2	00:42:18.48	-10:20:36.48	95.06	129	6.01	0.90	33.0	-1767.19	78.08 ± 13.75	$R > 25.6, g > 26.3$
J0119-0342.1	01:20:00.37	-03:42:59.90	95.33	58	5.76	0.85	45.1	-2266.36	48.65 ± 9.14	$R > 25.7, g > 26.4$
J0119-0342.3	01:20:02.15	-03:42:16.96	93.90	128	5.71	0.85	39.0	2270.94	70.32 ± 14.13	$R > 25.7, g > 26.4$
J0149-0552.2 <sup>a</sup>	01:49:08.41	-05:52:41.34	94.79	47	5.91	0.80	31.3	-851.97	42.31 ± 9.73	$R > 25.7, g > 26.5$
J0202-0650.1 <sup>a</sup>	02:02:54.72	-06:50:51.88	94.82	173	13.11	1.00	16.1	-829.85	193.75 ± 15.79	$R = 24.4, g = 25.7$ Lyman break detection with $g - R = 1.3$ located at 1''2
J0850+0629.1 <sup>a</sup>	08:50:14.22	06:29:49.93	94.80	428	17.60	1.00	11.9	-710.84	556.37 ± 27.58	$R > 25.7, g > 26.3$
J1211+1224.1 <sup>a</sup>	12:11:45.06	12:25:06.25	94.65	75	6.26	0.93	54.9	559.31	90.54 ± 16.91	$R > 25.3, g > 26.3$ LAE located at 1''6
J1258-0130.1	12:58:43.81	-01:31:09.35	94.85	299	5.97	0.91	53.1	-1220.26	120.69 ± 26.47	$R > 25.7, g > 26.4$
J2250-0846.2 <sup>a</sup>	22:50:50.92	-08:45:40.39	94.43	728	8.50	1.00	32.7	398.63	277.69 ± 34.97	$R > 25.6, g > 26.4$

**Notes.** This table contains all the emission lines detected with Fidelity  $\geq 0.8$  (or  $\geq 0.9$  if they are Located at Radius  $> 50''$  14 from the Pointing Center). Columns (1), (2), and (3) indicate the name of the candidate and the R.A., decl. sexagesimal coordinates. Columns (4) and (5) indicate the central frequency and the FWHM of the detected line as determined by the best Gaussian+continuum fitting. Column (6) shows the S/N of the detection as determined by the line search algorithm. Column (7) shows the fidelity computed using Equation (1). Columns (8) and (9) show the angular distance and the velocity distance, respectively, between the emission line and the central quasar, assuming that the observed line corresponds to the CO(4–3) transition. For the computation of the velocity distance, we use the redshift based on the ALMA detection of the quasar, when it is detected with fidelity  $\geq 0.8$ ; otherwise, we use the redshift as determined from the optical quasar spectrum. Column (10) indicates the (non-primary beam corrected) integrated line flux measured from the 1D spectra presented in Figure 2. Column (11) shows the optical color–magnitude  $g$  and  $R$  measured in a 2'' aperture in the position of the source (or at the position of a counterpart, if one exists within 2''). The lower limits for the magnitude correspond to  $3\sigma$  limits. We provide comments about optical counterparts for the source.

<sup>a</sup> Included for the clustering measurement.

we assume that all the detected lines correspond to the CO(4–3) transition and that the serendipitous detection of emission lines at other redshifts is negligible. To support this assumption, here we quantify the probability that each of our detected lines is actually a galaxy at a different redshift, and we also use our optical images to look for a counterpart of the detected emission lines.

#### 3.4.1. Probability of Low- $z$ Contaminants

First, we compute the probability that each source in our catalog is a low- $z$  galaxy. For this computation, we first explore what transitions at low- $z$  could be detectable given the frequency setup of our survey, and we compute the comoving volume traced for them in each quasar field. We consider a circular area given by  $R \leq 59''$  (the same used to look for emission lines) and a redshift coverage given by (i)  $\delta v = \pm 3000 \text{ km s}^{-1}$ , corresponding to the width of the whole cube, the same as where we detect our nine candidates, and (ii)  $\delta v = \pm 1000 \text{ km s}^{-1}$ , corresponding to the width used for the clustering analysis, where we detect five candidates (see details in Section 4.1). We summarize this information in Table 4. The main possible contaminants in our sample are the CO transitions CO(1–0) at  $z \sim 0.2$ , CO(2–1) at  $z \sim 1.4$ , and CO(3–2) at  $z \sim 2.7$ .

To compute the probability, we use the CO luminosity function of the different CO transitions (Decarli et al. 2019) at the corresponding redshift, and integrate it down to the  $5.6\sigma$  limiting luminosity of our survey (this is the same S/N threshold that is used in this work for the detections of CO emitters; see Section 3.3) to compute the number density of sources for each transition. We multiply this quantity by the traced comoving volume to obtain the expected number of contaminants in one field, and we show the result in Table 4.

**Table 4**  
Probability of the Presence of Low- $z$  Contaminants in Our Survey

Transition	$z_{\text{min}}$	$z_{\text{max}}$	Volume <sub>3000</sub> [ $h^{-1} \text{cMpc}$ ]	$N_{3000}$	$N_{1000}$
(1)	(2)	(3)	(4)	(5)	(6)
CO(1–0)	0.206	0.231	6.486	0.023	0.008
CO(2–1)	1.413	1.461	147.276	0.250	0.083
CO(3–2)	2.619	2.691	253.083	0.103	0.034

**Note.** For the different transitions, we report the redshift and the comoving volume traced in one quasar field at the redshift of the transition. The comoving volume is given by a circular area with radius  $R \leq 59''$  and a redshift coverage given by  $\delta v = \pm 3000 \text{ km s}^{-1}$ . Columns (5) and (6) show the expected number of lines per quasar field in the comoving volume indicated in column (4), and in 1/3 of that volume (i.e., assuming  $\delta v = \pm 1000 \text{ km s}^{-1}$ ), respectively. Note that, in the absence of clustering, the number of expected lines only depends on the volume; therefore, the number of lines simply decrease by a factor of three in column (6) compared with column (5).

The limiting luminosity at the center of the pointing is reported in Table 2. Note that the limiting luminosity varies over the ALMA pointing, due to the decrease of sensitivity with the radius, thus we include this in our computation, following the procedure described in Section 4.1.

The probability that each of the lines presented in our final catalog is CO(1–0), CO(2–1), and CO(3–2) is 2%, 25%, and 10%, respectively. If we only focus on the sample used for the clustering analysis composed of five sources, we find that the probability that each of these sources is CO(1–0), CO(2–1), and CO(3–2) is 0.8%, 8.3%, and 3.4%, respectively.

#### 3.4.2. Optical Counterparts

The second approach that we use to explore possible contamination from low- $z$  galaxies is to cross-match the spatial

position of the detected lines with the position of LAEs detected in the optical images available for our 17 fields. We limit the search for the optical counterparts of our emission lines to a radius of  $2''$ . Using the available LAE catalog (see Section 2.2), we find that one of the detected candidates (J1211+1224.1) coincides with the position of an LAE detected in our previous survey. We also explored the optical images at the position of all the line detections, but we do not find any LAE candidate at these positions, even when relaxing the LAE detection threshold down to  $S/N > 3$ .

We also check the possibility of including additional low-fidelity line emission sources at the position of the only other  $S/N > 5$  LAEs that were located within the ALMA Field-of-view. However, even by relaxing the fidelity criterion down to 0.3 in the ALMA catalog, no line detections were found within  $3''$  from the position of the LAE.

Since the UV-continuum emission of  $z \sim 4$  galaxies is characterized by a strong flux break at  $\lambda_{\text{RF}} = 1216 \text{ \AA}$  (the so-called Lyman break), owing to the absorption of photons with  $\lambda_{\text{RF}} \leq 1216 \text{ \AA}$  by neutral hydrogen in the intergalactic medium, we can use the  $R$ - and  $g$ -band optical images to detect the expected flux break. Although all the galaxies should exhibit this break, not all of them will exhibit the LAE emission line, and thus they may not be included in the LAE catalog. Note that the presence of the flux break by itself cannot confirm the redshift of the source, since we would need an additional observation on the UV continuum at longer wavelengths to exclude possible interlopers, but the absence of the break can be used to confirm that the source is located at lower redshift.

For all the detected lines, we search for a  $\geq 3\sigma$  detection in the  $R$  and  $g$  images within  $2''$ . If a source is found in either of these bands, we perform a  $2''$  aperture photometry on that position in the two bands following the same criteria as in García-Vergara et al. (2019). If no source is found, we perform the aperture photometry at the position of the detected line and compute  $3\sigma$  upper limits for the fluxes in the  $g$  and  $R$  bands. In all the cases, we compute the color-magnitude  $g - R$  and report these values in Table 3. As in García-Vergara et al. (2019), the Lyman break is defined by the color criteria  $g - R \geq 0.7$ .

In addition to the LAE mentioned above (J1211+1224.1), we find that one other CO-emitting galaxy (J0202-0650.1) exhibits the Lyman break, suggesting that it is possibly an LBG at  $z \sim 4$ . All the other galaxies are not detected in either  $g$  and  $R$  bands, so we do not have information to trace the Lyman break, and we can not confirm that any of these are low- $z$  interlopers. However, the lack of an optical counterpart at the VLT survey (with  $3\sigma$  limiting magnitude 25.61 and 26.37 for the  $R$  and  $g$  images respectively) makes these objects unlikely to be lower- $z$  ( $z \sim 0.2$  for CO(1–0),  $z \sim 1.4$  for CO(2–1), and  $z \sim 2.6$  for CO(3–2); see Table 4) galaxies.

Specifically, we check the typical  $R$  magnitude for CO(2–1) emitters at  $z \sim 1.4$ , which are our main potential contaminants (see Table 4). We used the photometric catalog available from Skelton et al. (2014) to check the median  $R$  magnitudes<sup>11</sup> for the 11 CO(2–1) sources detected with high fidelity in the ASPECS survey (see Table 6 in González-López et al. 2019). We find that these sources have a median magnitude  $R = 24.83$ , which is much lower than the  $3\sigma$  limiting magnitude in the  $R$  band for our candidates (see Table 3), making them unlikely to be CO(2–1) emitters at  $z \sim 1.4$ .

<sup>11</sup>  $R$  magnitudes correspond to that measured in images taken using ESO/WFI with the filter  $R_c$  (see Skelton et al. 2014 for details).

### 3.5. Quasar Detection

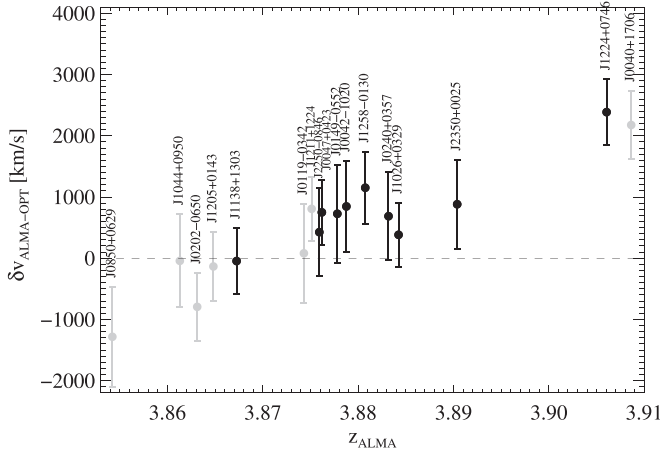
We detect an emission line in 10/17 quasars, which we assume to be the CO(4–3) transition, as they all have an optical counterpart and spectroscopic redshift from SDSS spectra. All 10 lines have fidelity  $\geq 0.8$  and are detected with  $S/N$ s ranging from 5.9 up to 24.5. We relax our  $S/N$  criterion down to  $S/N > 4$ , and we look for possible line candidates for the other seven quasars, restricting the search to detections with line widths  $\geq 124 \text{ km s}^{-1}$ . When more than one candidate was found at the expected sky projected position of their optical counterpart (within  $3''$ ), we selected the one with higher  $S/N$ . We do not impose a requirement in the velocity space position of the quasar counterpart, since the redshift of optical quasars could be associated with larger uncertainties than reported in previous studies. An uncertainty in the optical computed redshift of  $2970 \text{ km s}^{-1}$  (one half of the bandwidth) could be still possible. This search results in the detection of all the other seven quasar counterparts, detected with  $S/N$ s ranging from 4.1 up to 5.3, but with low fidelity ( $< 0.4$ ). Therefore, some of these detections could be noise fluctuations.

We compute the offset between the CO(4–3) redshift versus the redshifted based on rest-frame UV-emission lines for all the quasars, and we show our results in Figure 3 and Table 5. For the secure quasar detections, we find a median velocity offset of  $|\delta v| = 738 \pm 651 \text{ km s}^{-1}$ . This is consistent with the typical reported uncertainties of the optical-based redshifts (see Table 1). The only exception is the quasar located in the field J1224+0746, which exhibits a large offset of  $|\delta v| = 2386 \text{ km s}^{-1}$  and therefore is located at  $1004 \text{ km s}^{-1}$  from the edge of the SPW0. For completeness, we also include the velocity offset computed for the quasar detected with low fidelity in our ALMA observations, but in this study, we only use the redshifts from the secure detections. We present a detailed analysis of the quasar properties in a forthcoming study (C. García-Vergara et al. in preparation).

In Figure 4, we show the sky distribution of the detected companion lines around the central quasar, and in Figure 5 we show the distribution of the velocity offsets between the lines and the central quasar for all our fields together. For the computation of the velocity offsets, we use the redshift based on the ALMA detection of the quasar, when it is detected with fidelity  $\geq 0.8$ ; otherwise, we use the redshift as determined from the optical quasar spectrum (reported in Table 1).

### 4. Clustering Analysis

In this section, we measure the clustering properties of the CO(4–3) emitters around quasars, following a procedure analogous to the one in our previous studies about the clustering of LAE and LBG around  $z \sim 4$  quasars (García-Vergara et al. 2017, 2019). We refer the reader to these works for detailed descriptions, but in Section 4.1 we provide a brief overview. We also use the measured cross-correlation function to infer the clustering of CO emitters in blank fields in Section 4.3. We note that, in this study, we do not analyze the clustering of continuum sources, since the clustering signal would be strongly diluted when projected over the large radial comoving distance traced by the almost flat selection function of these sources over the redshift range  $1 \lesssim z \lesssim 7$ .



**Figure 3.** Velocity offset between the quasar redshifts determined from the CO(4–3) emission lines and the rest-frame UV-emission lines. We include the secure quasar detections (black) and the low-fidelity detections (gray). Based only on the secure detections, we find a median offset of  $|\delta v| = 738 \pm 651 \text{ km s}^{-1}$ .

#### 4.1. Quasar–CO Cross-correlation Function

We measure a volume-averaged projected cross-correlation function between quasars and CO(4–3) emitters defined by

$$\chi(R) = \frac{\int_{V_{\text{eff}}} \xi(R, Z) dV}{V_{\text{eff}}}, \quad (2)$$

where  $\xi(R, Z)$  is the real-space correlation function, assumed to have a power-law shape  $\xi(r) = (r/r_{0,\text{QG}})^{-\gamma}$ , where  $r_{0,\text{QG}}$  and  $\gamma$  are the correlation length and the slope of the correlation function, respectively. The real-space separation between objects,  $r$ , has been written in Equation (2) as a function of their two components: the transverse comoving separation  $R$ , and the radial comoving separation  $Z$ .  $V_{\text{eff}}$  is the effective volume of the survey, which is a cylindrical shell centered on each quasar, with inner and outer radius given by  $R_{\text{min}}$  and  $R_{\text{max}}$ , respectively, and with a height defined by the radial comoving distance  $Z_{\text{max}} - Z_{\text{min}}$ .

We compute  $\chi(R)$  in logarithmically spaced radial bins centered on the quasar by using the estimator

$$\chi(R) = \frac{\langle QG(R) \rangle}{\langle QR(R) \rangle} - 1, \quad (3)$$

where  $\langle QG(R) \rangle$  is the number of quasar–galaxy pairs observed in our survey, within the cylindrical shell volume, and  $\langle QR(R) \rangle$  is the number of quasar–galaxy pairs that we expect to observe within the same volume but in the absence of clustering (i.e., assuming the background number density of CO-emitting galaxies at  $z \sim 4$ ).

If we had ALMA observations covering large areas of the sky at random locations (i.e., regions not containing quasars), using the same setup as for our data, we could directly measure  $\langle QR(R) \rangle$  from the data. However, this is not the case. We cannot use the three quasar-free SPWs (SPWs 1, 2, and 3) for this purpose, due to the following reasons. First, the central frequency (at least for SPW1) is relatively close

( $< 6000 \text{ km s}^{-1}$ ) to the quasar location, and thus the background number counts could not necessarily be reached.<sup>12</sup> This leaves us with only the other two SPWs, which are far enough away, reaching the background number density, although they would be tracing the number density of CO(4–3) at slightly lower redshift ( $z \sim 3.2$ ). Second, the volume covered by SPW2 and SPW3 is small, resulting in extremely low statistics for the number counts.<sup>13</sup> This would result in large uncertainty in the computed  $\langle QR(R) \rangle$ , and therefore in the  $\chi(R)$  value.

Third, the SPW1, SPW2, and SPW3 were observed in TDM mode, whereas the SPW0 was observed in FDM mode (see Section 2.3). This results in different spectral noise properties, as well as different spectral resolutions (the spectral resolution in SPW1, SPW2, and SPW3 is two times lower than the resolution in SPW0), which complicates the comparison of the line number density in the different cubes (we would be sensitive to a different type of lines in SPW1, SPW2, and SPW3 compared to the lines detected in SPW0). Considering all the mentioned reasons, we thus decided to estimate  $\langle QR(R) \rangle$  using the CO(4–3) luminosity function measured at  $z = 3.8$  from the ASPECS survey (Decarli et al. 2019). We detail this in what follows.

The number of quasar–galaxy pairs that we expect to observe in the absence of clustering is given by  $\langle QR(R) \rangle = \int_{V_{\text{eff}}} n_{\text{CO}}(L' \geq L'_{\text{Lim}}) dV$ , and for the cylindrical shell volume defined above, we can write

$$\langle QR(R) \rangle = \int_{Z_{\text{min}}}^{Z_{\text{max}}} \int_{R_{\text{min}}}^{R_{\text{max}}} n_{\text{CO}}(L' \geq L'_{\text{Lim}}) 2\pi R dR dZ, \quad (4)$$

where  $n_{\text{CO}}(L' \geq L'_{\text{Lim}})$  is the number density of CO(4–3) emitters with line luminosity above the limiting luminosity of our survey. To compute the limiting luminosity, we follow Solomon et al. (1997),

$$\frac{L'_{\text{Lim}}}{\text{K km s}^{-1} \text{pc}^2} = \frac{3.257 \times 10^7}{1+z} \frac{F_{\text{Lim}}}{\text{Jy km s}^{-1}} \left( \frac{\nu_0}{\text{GHz}} \right)^{-2} \left( \frac{D_L}{\text{Mpc}} \right)^2, \quad (5)$$

where  $\nu_0$  is the rest-frame frequency of the line ( $\nu_0 = 461.04$  for CO(4–3)),  $D_L$  is the luminosity distance at the redshift of the source,  $z$ , and  $F_{\text{Lim}}$  is the limiting integrated line flux.

The limiting integrated line flux for each quasar field is computed using the rms noise of the cubes from Table 2, as well as the S/N threshold defined by our fidelity threshold. Specifically, we assume a typical FWHM line width of  $331 \text{ km s}^{-1}$ , which is the median width measured in the high-fidelity CO line sample of the ASPECS survey.<sup>14</sup>

<sup>12</sup> At this distance, we expect that the number density of CO(4–3) lines is 2.3 times the background number density. We have computed this quantity assuming a power-law shape for the cross-correlation function with cross-correlation length  $r_0 = 8.37 h^{-1} \text{ cMpc}$  (the cross-correlation length measured in this work) for a fixed  $\gamma = 1.8$ . We take into account the volume of the SPW1 cube, and its distance from the quasar.

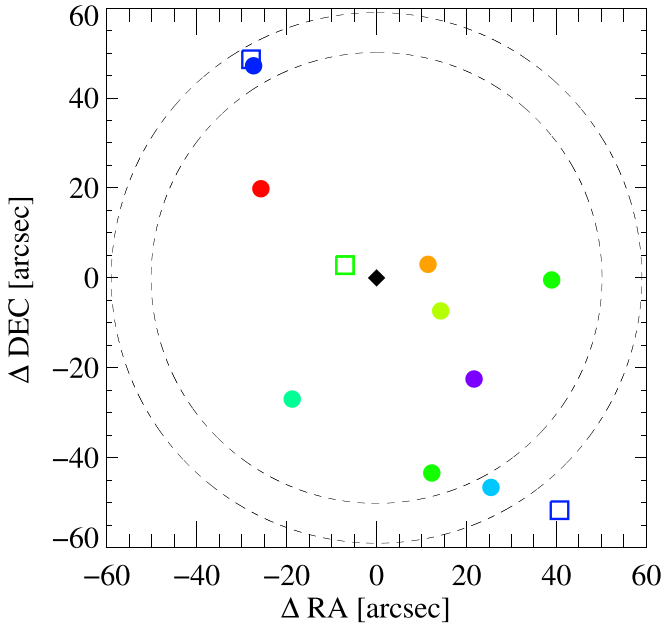
<sup>13</sup> We computed the probability to detect a CO(4–3) line in each SPW, assuming the background number density from the CO(4–3) luminosity function measured in Decarli et al. (2019; more details of this computation are provided in the main text). This probability is 7% and 7% for SPW2 and SPW3, respectively, which would result in a total of  $\sim 2.4$  CO(4–3) lines for the 17 fields detected in the volume covered by the two SPWs.

<sup>14</sup> We note that this is larger than the median FWHM of the five sources used for our clustering analysis (median FWHM is  $173 \pm 161$  for the five sources with fidelity  $> 80\%$  and  $301 \pm 183$  for the four sources with fidelity  $> 90\%$ ), but still consistent within error bars. We nevertheless checked that the assumed line width has only a small impact on our final clustering results.

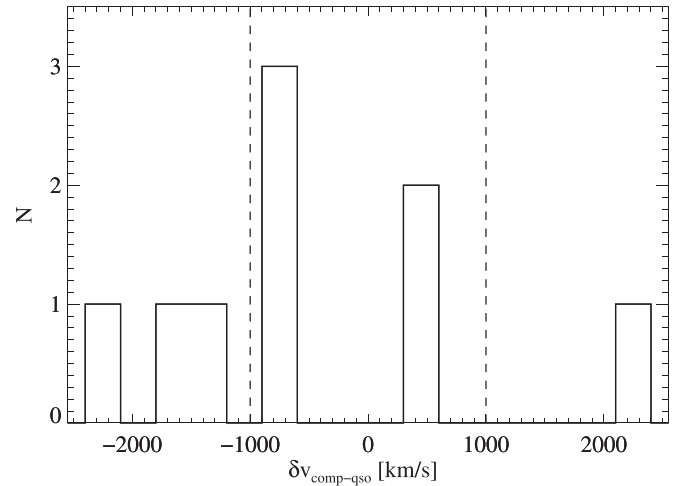
**Table 5**  
Quasar Redshift Determined from ALMA Observations, and Number of Emission Lines in Each Field

Field (1)	$z_{\text{opt}}$ (2)	$z_{\text{ALMA}}$ (3)	$\delta v$ [km s <sup>-1</sup> ] (4)	$N_{3000}(\delta_{\text{CO}})$ (5)	$N_{1000}(\delta_{\text{CO}})$ (6)	$\delta_{\text{LAE}}$ (7)
J0040+1706	3.873	3.910 <sup>a</sup>	2266 <sup>a</sup>	0 (00.00)	0 (00.00)	2.66
J0042-1020	3.865	3.878	805	1 (17.77)	0 (00.00)	0.00
J0047+0423	3.864	3.877	780	0 (00.00)	0 (00.00)	0.00
J0119-0342	3.873	3.874 <sup>a</sup>	46 <sup>a</sup>	2 (34.90)	0 (00.00)	3.41
J0149-0552	3.866	3.879	781	1 (18.61)	1 (55.87)	0.00
J0202-0650	3.876	3.864 <sup>a</sup>	-741 <sup>a</sup>	1 (17.63)	1 (52.91)	0.00
J0240+0357	3.872	3.883	661	0 (00.00)	0 (00.00)	0.92
J0850+0629	3.875	3.854 <sup>a</sup>	-1310 <sup>a</sup>	1 (19.40)	1 (58.22)	2.41
J1026+0329	3.878	3.884	380	0 (00.00)	0 (00.00)	2.16
J1044+0950	3.862	3.861 <sup>a</sup>	-37 <sup>a</sup>	0 (00.00)	0 (00.00)	0.78
J1138+1303	3.868	3.867	-62	0 (00.00)	0 (00.00)	0.00
J1205+0143	3.867	3.864 <sup>a</sup>	-161 <sup>a</sup>	0 (00.00)	0 (00.00)	1.34
J1211+1224	3.862	3.875 <sup>a</sup>	805 <sup>a</sup>	1 (25.97)	1 (77.95)	3.75
J1224+0746	3.867	3.905	2355	0 (00.00)	0 (00.00)	0.00
J1258-0130	3.862	3.882	1203	1 (27.34)	0 (00.00)	1.08
J2250-0846	3.869	3.876	433	1 (19.61)	1 (58.86)	0.77
J2350+0025	3.876	3.891	920	0 (00.00)	0 (00.00)	0.59
ALL				9 (10.56)	5 (17.60)	1.36

**Notes.** For each field, column (2) shows the quasar redshift determined from UV rest-frame emission lines, column (3) shows the redshift determined from the CO(4–3) emission line, column (4) shows the velocity offset between these (with typical uncertainties of 651 km s<sup>-1</sup>), and columns (5) and (6) show the number of CO emitters and the corresponding overdensity within  $\pm 3000$  km s<sup>-1</sup> and  $\pm 1000$  km s<sup>-1</sup> from the quasar redshift, respectively (using the redshift determined from the CO(4–3) line, if detected with high fidelity ( $\geq 0.8$ ), and the UV rest-frame emission lines otherwise). Note that, due to the quasar redshift offsets, we are not tracing a symmetric volume of  $\pm 3000$  km s<sup>-1</sup> in all the fields; however, we are symmetrically tracing 1000 km s<sup>-1</sup> in all the cases, since all the quasars are located at least at 1000 km s<sup>-1</sup> from the edge of the cube. The overdensity is computed as the number of detected galaxies over the number of expected galaxies in blank fields over the same traced volume (see Section 4.1). Column (7) show the overdensity of LAEs at  $R \lesssim 7 h^{-1}$  cMpc within  $\pm 1600$  km s<sup>-1</sup> in each field (García-Vergara et al. 2019).  
<sup>a</sup> Indicates the cases where the CO(4–3) line is detected with low ( $< 0.8$ ) fidelity.



**Figure 4.** Sky distribution of the emission lines around the central quasar for our 17 fields combined. The central black diamond indicates the quasar position. The lines are indicated as filled circles, with different colors indicating different fields in which they were detected. We include the position of the LAEs detected with  $S/N \geq 5$  in the field as open squares. The largest dashed line circle shows the region where we search for line candidates, which corresponds to the whole ALMA pointing (a radius of 59'' from the central quasar). The smaller dashed line circle shows the radius at which the telescope sensitivity is  $\geq 20\%$  of the maximum. Sources located outside of this limit radius were included in our catalog only if they have a fidelity  $\geq 0.9$ .



**Figure 5.** Velocity offset distribution of the CO(4–3) lines around the central quasar for our 17 fields together. The vertical dashed lines indicate the velocity window that was chosen to measure the clustering of sources around quasars (see Section 4.1).

(González-López et al. 2019), and we use the corresponding  $S/N$  threshold for such a width ( $S/N = 5.6$  for a detection with fidelity 0.8 as shown by the magenta curve in Figure 1). The limiting luminosity of the survey at the center of the pointing is reported in Table 2.

If the rms sensitivity of the survey were roughly flat, the term  $n_{\text{CO}}(L' \geq L'_{\text{Lim}})$  in Equation (4) could be taken out from the integral and it could be simply computed by integrating the CO(4–3) luminosity function down to the limiting luminosity



of the survey computed from Equation (5). However, since the ALMA sensitivity decreases when the radius within the pointing increases, the limiting luminosity actually depends on  $R$ , and therefore we have to model this dependency and numerically compute the integral over  $R$  in Equation (4).

We model the sensitivity variation over the pointing as a 2D Gaussian with FWHM given by the primary beam size ( $\sim 66''$ ), and compute the primary-beam-corrected limiting integrated line flux as a function of  $R$ , which gives us the limiting luminosity as a function of  $R$  from Equation (5).

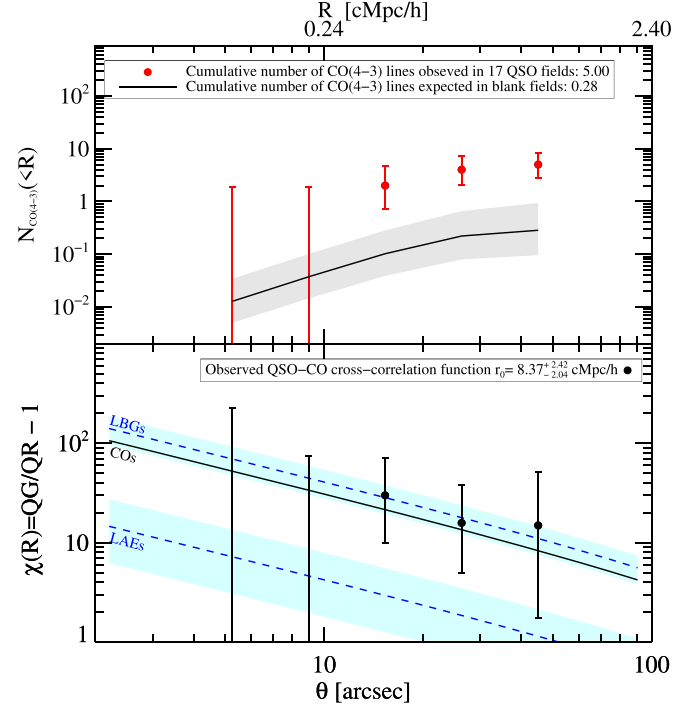
Finally, the limiting luminosity as a function of  $R$  is used to compute  $n_{\text{CO}}(L' \geq L'_{\text{Lim}})$  at every  $dR$  integration step in Equation (4). Here, we use the CO(4–3) luminosity function at  $z = 3.8$  measured by Decarli et al. (2019), which is given by the Schechter parameters  $\log \Phi_* [\text{Mpc}^{-3} \text{dex}^{-1}] = -3.43$ ,  $\log L'_* [\text{K km s}^{-1} \text{pc}^2] = 9.98$ , for a fixed  $\alpha = -0.2$ .

In principle, in Equation (4) we should include a multiplicative term encapsulating the completeness of the sample. This completeness refers to the fraction of sources detectable by the search algorithm, and it is dependent on the line flux, line width, fidelity, and S/N of the sources.<sup>15</sup> Based on previous estimations of the performance of the line search algorithm FINDCLUMPS using simulated mock sources (e.g., González-López et al. 2019; Decarli et al. 2020; Loiacono et al. 2021), we expect that at the high fidelity and S/N threshold used in our study, the completeness is near 100%, so we have neglected this completeness correction. This could result in a slight underestimation of the clustering of sources around quasars, but our main conclusions will not significantly change.

The computation of  $\langle QG(R) \rangle$  and  $\langle QR(R) \rangle$  is performed individually for each of the 17 fields, and then they are stacked to obtain the final  $\chi(R)$  value using Equation (3). We measure the clustering using  $Z_{\text{max}} = -Z_{\text{min}} = 8.19 h^{-1} \text{ cMpc}$  (corresponding to  $\delta v = \pm 1000 \text{ km s}^{-1}$  at  $z = 3.87$ ). We note that the quasar with the largest redshift offset is located at  $1004 \text{ km s}^{-1}$  from the edge of SPW0 (see Section 2.1), and therefore the choice of  $\delta v = \pm 1000 \text{ km s}^{-1}$  allows us to trace a symmetric and complete volume around the quasar for all the fields. This choice results in a total of five sources included in the clustering analysis (see Table 3). Given the small size of our sample, we assume that Poisson error dominates our measurement, and we compute the one-sided Poisson confidence interval for small number statistics from Gehrels (1986). We show the measured cross-correlation in Figure 6 and tabulate it in Table 6.

Note that choosing a larger radial comoving range ( $\delta v$ ) would result in more sources for the measurement, increasing the statistics, but the clustering signal would be diluted because we are integrating the signal in a larger volume and up to large distances from the quasar, where the background is close to being reached. We also note that keeping the volume small also decreases the probability of having a contaminant in the sample (see Section 4). We explore the effect of the chosen  $\delta v$  in the Appendix.

For the seven quasars not (securely) detected in our observations, we use the optical-based redshifts, which could have an offset from the ALMA-based redshift of  $\sim 700\text{--}800 \text{ km s}^{-1}$  (see Figure 3). This would mean that, for these quasars, we may be including or excluding companions at larger and smaller redshift distances, respectively. The confirmation of the precise redshift for



**Figure 6.** *Top:* The cumulative number counts of CO(4–3) lines observed in our 17 quasar fields ( $\langle QG(\leq R) \rangle$ ) within  $\delta v = 1000 \text{ km s}^{-1}$  from the central quasar (red points) with Poisson error bars, compared to the expectation for CO(4–3) lines in blank fields in our 17 quasar fields ( $\langle QR(\leq R) \rangle$ ) computed using Equation (4) (black line). The gray area shows the uncertainty in this expectation. Our observations yield five CO(4–3) lines from the ensemble of the 17 fields, while we expect only 0.28 CO(4–3) lines from the background alone, resulting in a total overdensity of  $17.6^{+11.9}_{-7.6}$ . *Bottom:* Quasar–CO cross-correlation function  $\chi(R)$  with  $1\sigma$  Poisson error bars for the 17 fields (black data points) computed using Equation (3), and the best maximum likelihood estimator for both  $r_{0,\text{QG}}$ , assuming a fixed  $\gamma = 1.8$  (black line). For comparison, we show the quasar–LBG and quasar–LAE cross-correlation function (blue dashed lines) observed at  $z \sim 4$ , and their respective uncertainties showed as blue shaded areas (García-Vergara et al. 2017, 2019).

**Table 6**  
Quasar–CO Cross-correlation Function

$R_{\text{min}}$ $h^{-1} \text{ cMpc}$	$R_{\text{max}}$ $h^{-1} \text{ cMpc}$	$\langle QG(R) \rangle$	$\langle QR(R) \rangle$	$\chi(R)$
0.096	0.165	0.000	0.008	$-1.00^{+226.13}_{-0.00}$
0.165	0.282	0.000	0.025	$-1.00^{+74.91}_{-0.00}$
0.282	0.483	2.000	0.065	$29.80^{+40.75}_{-19.96}$
0.483	0.827	2.000	0.119	$15.79^{+22.15}_{-10.85}$
0.827	1.417	1.000	0.063	$14.88^{+36.53}_{-13.14}$

these quasars is the only way to correct for these uncertainties in our clustering measurement, but we explore how much the clustering signal changes if we use ALMA-based redshifts for all the quasars, even if they are detected at low fidelity. We find that the same five CO emitters are kept in this case, confirming the stability of the obtained correlation function.

To determine the real-space cross-correlation parameter  $r_{0,\text{QG}}$  that best fits our data, we use a Poisson maximum likelihood estimator. Given the noise of the data and the small physical scales proved in our study, we follow common practice and fit our data assuming a fixed slope  $\gamma = 1.8$  (e.g., Ouchi et al. 2004; García-Vergara et al. 2017, 2019; Fossati et al. 2021). We find that the maximum likelihood and  $1\sigma$

<sup>15</sup> Note that this is not the completeness due to the sensitivity variation over the pointing, which is indeed corrected in our computation as explained above.

confidence interval for the correlation length is  $r_{0,QG} = 8.37^{+2.42}_{-2.04} h^{-1} \text{cMpc}$ . We use the best-fitted parameter in Equation (2) to compute the corresponding  $\chi(R)$  value, which is shown as a black line in Figure 6.

Finally, we use our  $\langle QG(R) \rangle$  and  $\langle QR(R) \rangle$  binned values to compute the observed and expected cumulative number counts of CO(4–3) lines in our 17 quasar fields ( $\langle QG(\leq R) \rangle$ ), and show this in Figure 6. In the whole volume survey ( $1751.7 h^{-3} \text{cMpc}^3$  for the 17 fields over  $\delta v = 1000 \text{ km s}^{-1}$ ), we find a total of five CO(4–3) lines, while we expect only 0.28 CO(4–3) lines from the background alone, resulting in a total CO(4–3) line overdensity of  $17.6^{+11.9}_{-7.6}$  in quasar fields. We note that, due to the low-number statistics, the computed overdensity is significant at the  $2.2\sigma$  level, and either deeper observations or observations of a larger sample of quasar fields would be required to confirm the overdensity with higher significance. We also compute the total overdensity per field as the ratio of  $\langle QG(R) \rangle$  per field over  $\langle QR(R) \rangle$  integrated over the radial bins and provide these values in Table 5. Although the individual overdensity is dominated by low-number statistics and affected by cosmic variance, we include this to study possible correlations between the overdensities and the quasar properties (C. García-Vergara et al. in preparation).

#### 4.2. Comparison with the Clustering of Other Populations around Quasars

We compare our results with the clustering of LBG and LAE around  $z \sim 4$  quasars. Since the previous LBG and LAE clustering studies extend up to larger scales ( $R \lesssim 9 h^{-1} \text{cMpc}$ ) than those traced in our study ( $R \lesssim 1.5 h^{-1} \text{cMpc}$ ), for this comparison we assume that the small-scale quasar–CO cross-correlation function can be extrapolated toward larger scales following a single power-law shape.

We find that the QSO–CO cross-correlation length is slightly lower, but consistent within error bars with the QSO–LBG cross-correlation length, which is given by  $r_{0,Q-LBG} = 9.78^{+1.68}_{-1.86} h^{-1} \text{cMpc}$  for a fixed  $\gamma = 1.8$ <sup>16</sup> (García-Vergara et al. 2017). This suggests that CO emitters and LBGs would inhabit dark matter halos of similar masses at  $z \sim 4$  (we further discuss this point in Section 4.3).

We can also compare our measurements with the clustering of LAEs around this same quasar sample, thus providing a direct comparison of optical and dusty galaxy populations in these fields. We find that the cross-correlation length for CO-emitting galaxies is  $3.0^{+1.5}_{-1.4}$  times higher than the cross-correlation length for LAEs ( $r_{0,Q-LAE} = 2.78^{+1.16}_{-1.05} h^{-1} \text{cMpc}$  with  $\gamma = 1.8$ ; García-Vergara et al. 2019) around quasars, although we caution that the uncertainties are still large, and therefore the cross-correlation lengths are consistent within  $2\sigma$  error bars. We note that the redshift window traced by both studies is similar ( $\delta v = \pm 1000 \text{ km s}^{-1}$  for CO emitters, and  $\delta v = \pm 1600 \text{ km s}^{-1}$  for LAEs), and thus we do not expect that this discrepancy is the result of a dilution in the signal due to differences in the traced volume. We have explicitly checked this in the Appendix, where we find that the quasar–CO cross-correlation length is only 1.1 times smaller if we assume a  $\delta v = \pm 1600 \text{ km s}^{-1}$  (see Figure 8).

Tracing a different quasar sample at  $z = 3\text{--}4.5$ , and using deeper optical observations, (Fossati et al. 2021) measure a slightly higher quasar–LAE cross-correlation length of  $r_0 = 3.15^{+0.36}_{-0.40} h^{-1} \text{cMpc}$  at a fixed slope  $\gamma = 1.8$ <sup>17</sup>. This is still  $2.7^{+0.8}_{-0.8}$  times lower than the cross-correlation length for CO-emitting galaxies. We note that the physical scale traced in this study ( $R \lesssim 0.6 h^{-1} \text{cMpc}$ ) is slightly smaller than that traced in our analysis.

The difference in the clustering of CO emitters and LAEs around quasars is unlikely to be caused by differences in the halo mass hosting both populations (see Section 4.3), therefore we suggest that this discrepancy is related to physical processes affecting the visibility of the LAEs around quasars. We further discuss this interpretation in section Section 5.

#### 4.3. Autocorrelation of CO Emitters at $z \sim 4$

The autocorrelation of a galaxy population is a powerful tool, since it can be directly related to the dark matter halo mass in which that population reside (e.g., Cole & Kaiser 1989; Mo & White 1996; Sheth & Tormen 1999). The autocorrelation of CO-emitting galaxies at  $z \sim 4$  has never been measured before, mainly due to the lack of large and deep surveys of these galaxies at high- $z$ . The largest samples of CO emitters at  $z \sim 4$  currently available are composed of only a few tens of sources (e.g., Decarli et al. 2016, 2019), which do not provide enough statistics for an autocorrelation measurement. Therefore, up to now it has been challenging to compute precise masses of CO-emitting galaxies and understand how they fit into different evolutionary scenarios for high- $z$  galaxies.

However, under certain assumptions, the cross-correlation between CO emitters and quasars can be used to infer the clustering of CO-emitting galaxies in blank fields, providing initial constraints on the clustering of this population at  $z \sim 4$ .

First, we assume that our small-scale cross-correlation can be extrapolated toward larger scales following a power-law shape given by  $\xi(r) = (r/r_{0,QG})^{-\gamma}$ . Although the autocorrelation function of quasars and galaxies has been found to slightly deviate from a power law toward smaller ( $\lesssim 0.2 h^{-1} \text{cMpc}$ ) scales (e.g., Ouchi et al. 2005; Hennawi et al. 2006, but see also Shen et al. 2010), likely due to the transition between the one-halo to two-halo terms, the autocorrelation is typically reasonably well-approximated as a power-law, and thus we assume that the one-halo term does not strongly boost the signal in our measurement.

Second, we assume a deterministic bias model, in which the QSO–galaxy cross-correlation function can be written as  $\xi_{QG} = \sqrt{\xi_{QQ}\xi_{GG}}$ , where  $\xi_{QQ}$  and  $\xi_{GG}$  are the autocorrelation of quasar and galaxies, respectively. We also assume that  $\xi_{QQ}$  and  $\xi_{GG}$  have a power-law shape with the same slope  $\gamma = 1.8$ . Under these assumptions, the correlation lengths can be related by

$$r_{0,QG} = \sqrt{r_{0,QQ}r_{0,GG}}. \quad (6)$$

Using the quasar autocorrelation length previously reported at  $z \sim 4$ , given by  $r_{0,QQ} = 22.3 \pm 2.5 h^{-1} \text{cMpc}$  (recomputed from Shen et al. 2007 with a fixed  $\gamma = 1.8$ ), and our measured QSO–galaxy cross-correlation length, we obtain that the

<sup>16</sup> García-Vergara et al. (2017) fitted the quasar–LBG cross-correlation function using a fixed  $\gamma = 2.0$ , therefore we refitted their measurements using a fixed  $\gamma = 1.8$ , which results in  $r_{0,Q-LBG} = 9.78^{+1.68}_{-1.86} h^{-1} \text{cMpc}$ .

<sup>17</sup> The  $r_0$  value published in Fossati et al. (2021) has recently been recomputed. Here, we quote the most updated  $r_0$  value, which was kindly provided by the authors via private communication.

autocorrelation length of CO emitters at  $z \sim 4$  is given by  $r_{0,GG} = 3.14 \pm 1.71 h^{-1} \text{ cMpc}$ .

This measurement is slightly lower than the LBG autocorrelation length at  $z \sim 4$  ( $r_{0,LBG} = 4.1^{+0.2}_{-0.2} h^{-1} \text{ cMpc}$ ; Ouchi et al. 2004), and is slightly higher than the LAE autocorrelation length at  $z \sim 4$  ( $r_{0,LAE} = 2.74^{+0.58}_{-0.72} h^{-1} \text{ cMpc}$ , Ouchi et al. 2010). We caution that the uncertainties of our estimation are still large, which makes the CO autocorrelation length consistent within uncertainties with the autocorrelation length of both LAE and LBG.

Interestingly, the reported autocorrelation length of CO emitters is lower than the autocorrelation length of  $S_{870\mu\text{m}} > 4.0 \text{ mJy}$  submillimeter galaxies (SMGs) detected with ALMA at  $1.5 < z < 3$  ( $r_{0,SMGs} = 7.7^{+2.8}_{-2.6} h^{-1} \text{ cMpc}$ ; Stach et al. 2021, but see García-Vergara et al. 2020). If the SMG clustering does not strongly evolve with redshift as suggested by Stach et al. (2021), then this would mean that CO emitters are lower-mass galaxies compared to the population of SMGs typically detected in continuum surveys.

While our measurements are still noisy, they provide a first rough constraint of the clustering of CO-emitting galaxies, allowing us for the first time to locate them within the context of evolutionary galaxy models. Larger and deeper surveys of emitting lines around quasars are still needed in order to constrain the clustering of CO emitters more precisely. Alternatively, large and deep surveys of CO-emitting galaxies in blank fields would offer an independent and more direct constraint of the CO clustering at these redshifts. However, such large surveys are very expensive and extremely challenging because of the required sensitivity and the low number density of CO emitters in blank fields.

## 5. Discussion

Our study reveals a large overdensity of CO-emitting galaxies ( $17.6^{+11.9}_{-7.6}$ ) and a strong clustering of them around quasars ( $r_{0,QG} = 8.37^{+2.42}_{-2.04} h^{-1} \text{ cMpc}$ ) at scales  $R \lesssim 1.5 h^{-1} \text{ cMpc}$ . This result helps clarify the current confused picture of quasar environments at high  $z$  and provides strong observational evidence in favor of  $z \sim 4$  quasars as tracers of massive structures. By comparing with the previous measurement of clustering of LAEs around the same quasar sample at scales  $R \lesssim 7 h^{-1} \text{ cMpc}$ , we find that CO-emitting galaxies are more clustered than LAEs around quasars (with a cross-correlation length  $3.0^{+1.5}_{-1.4}$  times higher), resulting in large overdensities of these galaxies, whereas only a mild overdensity of LAEs ( $1.4^{+0.4}_{-0.4}$ ) was found in these fields (García-Vergara et al. 2019). A comparison with the other available LAE study in quasar environments (Fossati et al. 2021), which is performed at smaller physical scales ( $R \lesssim 0.6 h^{-1} \text{ cMpc}$ ), leads to similar conclusions, with the quasar–CO cross-correlation length being  $2.7^{+0.8}_{-0.8}$  times higher than the quasar–LAE cross-correlation length (see Section 4.2). In the following, we discuss the possible reasons that could explain this discrepancy.

First, we explore the possibility that CO-emitting galaxies inhabit more massive dark matter halos compared to LAEs, which would result in a significant difference in the quasar–galaxy cross-correlation for both populations. Although the dark matter halos for CO emitters at  $z \sim 4$  have not been constrained yet, the LAE autocorrelation length at  $z \sim 4$  is well-constrained and is found to be  $r_{0,LAE} = 2.74^{+0.58}_{-0.72} h^{-1} \text{ cMpc}$ , (Ouchi et al. 2010). If we focus on clustering hierarchy arguments only and assume a deterministic bias model, the difference of a factor of three in the measured quasar–CO and

quasar–LAE cross-correlation length implies that the autocorrelation length of CO emitters would be nine times larger than the autocorrelation length of the LAEs (see Equation (6)), resulting in  $r_{0,CO} \sim 25 h^{-1} \text{ cMpc}$ .

This is even higher than the quasar autocorrelation length at  $z \sim 4$  ( $r_{0,QQ} = 22.3 \pm 2.5 h^{-1} \text{ cMpc}$  (Shen et al. 2007), which implies halo masses of  $M_{\text{halo}} > 6 \times 10^{12} M_{\odot} h^{-1}$ ), and it would imply that CO emitters inhabit halos more massive than quasars. This seems to be an unlikely scenario and inconsistent with the CO dark matter halo mass inferred from our results, assuming a deterministic bias model ( $M_{\text{halo}} = 8.31^{+49.29}_{-8.27} \times 10^{10} M_{\odot}$ ; see Section 4.3).

Another possibility to explain the discrepancy in the overdensity of CO emitters and LAEs around quasars is related to the uncertainties on the rest-frame UV quasar redshift of our targets. If these quasar redshifts were associated with large uncertainties, the LAE search would have been offset from the real quasar position, which would result in a lower number density of detected galaxies, and thus a lower quasar–LAE cross-correlation. However, with our ALMA observations, we could confirm the redshift of 10 quasars (and tentatively do so for the other seven quasars detected with lower fidelity), and for them we find relatively low redshift offset compared with the ones determined from rest-frame UV-emission lines ( $|\delta v| = 738 \text{ km s}^{-1}$ ; see Section 3.3). These uncertainties in the quasar redshifts are much smaller than the FWHM of the narrow band used in the LAEs study ( $3197 \text{ km s}^{-1}$ ), implying that the LAE search was performed at the correct redshift.

If we assume that the remaining seven quasars have similar redshift uncertainties (as indeed suggested by their lower-fidelity detections; see Table 5), then this is not a convincing explanation for the observed discrepancy. Additionally, we note that five out of these seven quasars exhibit an overdensity of LAEs (see Table 5), which is an indication that the LAE search was performed at the correct redshift in these cases. We note that one of the quasars exhibits a larger offset ( $|\delta v| = 2386 \text{ km s}^{-1}$  for J1224+0746), in which case the LAE search was performed at larger distances from the quasar, where the background number density would be expected. However, if we exclude that field in the LAE study, we find that the total overdensity increases from 1.36 to 1.43 and the autocorrelation length increases by a factor of 1.09, which would not change our main conclusions.

Ruling out the two mentioned explanations as the main reasons to explain the large overdensity of CO emitters compared to the mild overdensity of LAEs in the same fields, we explore the possibility that particular physical properties in galaxies around quasars could be impacting their visibility and detection. While the CO(4–3) emission line traces the molecular gas in a galaxy,<sup>18</sup> the Ly $\alpha$  emission line is a tracer of instantaneous star formation, and thus, a relatively small star formation efficiency in galaxies around quasars could explain the lack of LAEs in these fields.

Although the molecular gas reservoirs in galaxies have been found to correlate well with their star formation rate, and stellar mass (e.g., Kennicutt 1998; Bigiel et al. 2008; Leroy et al. 2008), these scaling relations have been mostly measured in field galaxies, but the validation of such relations in high- $z$  dense environments is still poorly constrained. The few

<sup>18</sup> The luminosity of the CO(4–3) emission line  $L'_{\text{CO}(4-3)}$  can be converted to a molecular hydrogen gas mass  $M_{\text{mol}}$  by assuming an excitation correction  $L'_{\text{CO}(4-3)}/L'_{\text{CO}(1-0)}$  and a conversion factor  $\alpha_{\text{CO}}$  (e.g., Carilli & Walter 2013).



available studies have been aimed to trace the properties of galaxies in  $z \sim 1.5$  galaxy clusters, and they show evidence of systematic deviations from the scaling relations of the field toward larger molecular gas masses (e.g., Noble et al. 2017; Hayashi et al. 2018, but see also Rudnick et al. 2017). Although such systems have a much more dense and evolved intercluster medium than that in protoclusters at  $z \sim 4$ , the results of these studies suggest that complex physical processes may be involved in the galaxy evolution in dense environments.

We note, however, that the  $\text{Ly}\alpha$  emission does not depend only on the instantaneous star formation, but it also depends on other factors such as the interstellar and circumgalactic medium conditions, which impact the complex radiative transfer process of the  $\text{Ly}\alpha$  photons. Exploring the aforementioned scenario in the context of our galaxy samples would require the acquisition of more observations. Specifically, multiwavelength photometry would be useful to constrain the properties of the galaxies, including SFRs. Additionally, James Webb Space Telescope (JWST; Gardner et al. 2006) observations can add key information of kinematics, stellar masses, and SFRs, which would help to build a physical scenario for these systems. And at mm wavelengths, ALMA [C II] 158  $\mu\text{m}$  emission observations are sensitive tracers of the interstellar medium conditions, enabling the study of scaling relations in high- $z$  dense environments.

Finally, as previously suggested by García-Vergara et al. (2019), galaxies around quasars could be more dusty, affecting the visibility of the  $\text{Ly}\alpha$  line. Since  $\text{Ly}\alpha$  photons are easily absorbed by dust, the  $\text{Ly}\alpha$  emission line could be partially suppressed, becoming hardly detectable. This would result in a decreased number of detected LAEs around quasars. The CO emission line is instead unaffected by dust absorption, making it detectable even in galaxies with large fractions of dust. This could explain why the deeper LAE study (Fossati et al. 2021) reveals a stronger clustering of LAEs around high- $z$  quasars compared to that computed from shallower LAEs observations (García-Vergara et al. 2019).

We note that, in this scenario, the detection of LBGs would also be affected by the dust, implying a lower LBG overdensity around quasars than that expected. However, the escape of the  $\text{Ly}\alpha$  photons is particularly sensitive to the presence of dust, resulting in a strongly attenuated  $\text{Ly}\alpha$  emission line, whereas the UV continuum may be less impacted, making LBGs still detectable. ALMA continuum observations of our quasar fields would allow us to directly trace the dust in these galaxies, and together with additional multiwavelength data, would provide invaluable information to completely characterize the galaxy properties through SED modeling, testing their dependence on environment.

We caution that our clustering constraints are still dominated by the low-number statistics provided by our relatively small quasar sample. Deeper and/or larger surveys of quasar fields would constrain the galaxy clustering around quasars with higher S/N than that achieved in this work. Alternatively, targeting a brighter emission line, such as the emission line [C II] at 158- $\mu\text{m}$  (which is the strongest line from star-forming galaxies at radio wavelengths; Carilli & Walter 2013) would offer better statistics with the same exposure time despite the smaller area ( $R \lesssim 0.3 h^{-1} \text{cMpc}$ ) that we would trace using ALMA/band 8, at which this line is detectable at  $z = 3.87$ . Specifically, we computed that, for the same exposure time used in our CO emitters study, we expect to detect 17 [C II]

companions for all the quasar fields over the area covered with band 8.<sup>19</sup> Finally, we stress the importance of performing quasar environment studies using multiwavelength information, since the visibility of the galaxies around quasars seems to be strongly dependent on the studied wavelengths.

## 6. Summary

We use ALMA band 3 observations to perform a blind search for CO(4–3) emitting galaxies in the environment of 17  $z \sim 4$  quasars. The quasars were selected from the SDSS and BOSS quasar catalogs to lie within a redshift window of  $z \sim 3.862 - 3.879$  set by an optical narrow-band filter used in a previous work to detect LAEs around the quasars. The spectroscopic redshifts of the quasars are determined from the UV-emission lines (with typical uncertainties of  $< 800 \text{ km s}^{-1}$ ).

We explore a cylindrical volume around the quasars defined by a projected radius of  $R \lesssim 1.5 h^{-1} \text{cMpc}$  and a velocity coverage of  $\delta v \sim \pm 3000 \text{ km s}^{-1}$ . We find nine CO-emitting galaxies with  $\text{S/N} \geq 5.6$ , and line widths ranging from  $47 \text{ km s}^{-1}$  to  $728 \text{ km s}^{-1}$ . The S/N threshold was chosen to only select sources with fidelity  $\geq 0.8$  in our survey.

We also detect the CO(4–3) emission line from 10 quasars with fidelity  $\geq 0.8$ , and they typically show relatively low redshift offsets (the median offset is  $|\delta v| = 738 \text{ km s}^{-1}$ ) with respect to the redshifts determined from the UV-emission lines, except for one quasar that exhibit an offset of  $|\delta v| = 2386 \text{ km s}^{-1}$ . The other seven quasars were tentatively detected with lower fidelity ( $< 0.4$ ), and S/Ns ranging from 4.1 to 5.3.

We quantify the clustering properties of the galaxies around the quasars. For this, we only focus on a small velocity range of  $\delta v \sim \pm 1000 \text{ km s}^{-1}$  around the quasar, to avoid dilution of the small-scale clustering signal. Five of the CO-emitting galaxies lie within this volume, and we use this sample to measure the volume-averaged cross-correlation function. For this measurement, we estimate the background number density from the CO luminosity function previously measured at  $z = 3.8$  from blank fields (Decarli et al. 2019).

We find that the expected number density of CO(4–3) emitting galaxies in blank fields is 0.28 for the whole sample of quasars, while five were detected in our survey, which results in an overdensity of  $17.6^{+11.9}_{-7.6}$ . We also fit the quasar–CO cross-correlation in our fields and find a cross-correlation length of  $r_{0,\text{QG}} = 8.37^{+2.42}_{-2.04} h^{-1} \text{cMpc}$ , assuming a fixed slope  $\gamma = 1.8$ .

In a previous study, we performed a search for LAEs in the same 17 quasar fields, allowing us to simultaneously trace the clustering properties of optical and dusty galaxies around quasars for the first time. That study revealed only a mild overdensity ( $\times 1.4$ ) of LAEs in these fields, and a quasar–galaxy cross-correlation length of  $r_{0,\text{Q-LAE}} = 2.78^{+1.16}_{-1.05} h^{-1} \text{cMpc}$  (García-Vergara et al. 2019), which is  $3.0^{+1.5}_{-1.4}$  times lower than the cross-correlation length found for CO emitters.

We argue that differences in the halo mass hosting the two galaxy populations, as well as uncertainties associated with the optically based quasar redshifts, are unlikely reasons to explain the observed discrepancy. We suggest instead that the properties of galaxies in quasar environments could impact their visibility and thus detectability in  $\text{Ly}\alpha$  emission. Specifically, galaxies in quasar environments could have low star formation

<sup>19</sup> For this computation, we have used the [C II] luminosity function from Loiacono et al. (2021), and assumed that the quasar–C II cross-correlation function is the same as the quasar–CO cross-correlation function at  $z \sim 4$ .



efficiency, or they could have an excess of dust, thus becoming more difficult to detect in Ly $\alpha$  emission. Exploring these mentioned scenarios would only be possible with the acquisition of additional multiwavelength observations.

Finally, we use our quasar–CO cross-correlation to infer the clustering of CO-emitting galaxies at  $z \sim 4$ . Assuming a deterministic bias model, and extrapolating the observed small-scale cross-correlation up to larger scales, we find an autocorrelation length for CO emitters of  $r_{0,\text{CO}} = 3.14 \pm 1.71 h^{-1} \text{ cMpc}$  (for a fixed slope  $\gamma = 1.8$ ), which agrees well, within the  $1\sigma$  error bars, with the clustering of LBG and LAE at  $z \sim 4$ , but is lower than the clustering of SMGs at  $1.5 < z < 3$ . Assuming that there is not a strong evolution of the clustering of SMGs with redshift, this would imply that the population of CO emitters inhabit less massive halos compared to the general population of SMGs. Larger surveys of CO emitters are required in order to independently constrain their autocorrelation.

As the first quasar sample targeted for clustering studies of both optical and dusty galaxies, our study demonstrates the importance of tracing different galaxy populations, and it also opens new questions about environmental effects on galaxy evolution, highlighting the importance of characterizing galaxies in the vicinity of high- $z$  quasars.

This paper makes use of the following ALMA data: ADS/JAO.ALMA #2019.1.00441.S. ALMA is a partnership of ESO (representing its member states), NSF (USA), and NINS (Japan), together with NRC (Canada), MOST and ASIAA (Taiwan), and KASI (Republic of Korea), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO, and NAOJ. The authors acknowledge assistance from Allegro, the European ALMA Regional Center node in the Netherlands.

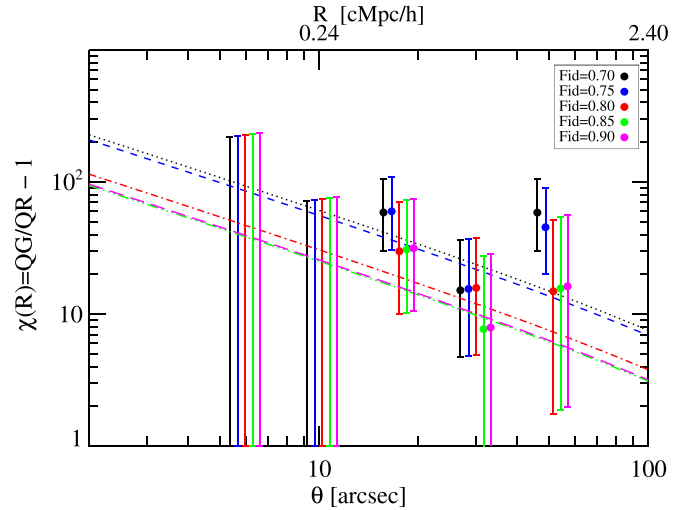
M.R. acknowledges the support of the NWO Veni project “Under the lens” (VI.Veni.202.225). M.R. and J.H. acknowledge support of the VIDI research program with project number 639.042.611, which is (partly) financed by the Netherlands Organisation for Scientific Research (NWO). M.A. acknowledges support from FONDECYT grant 1211951, “CONICYT + PCI + INSTITUTO MAX PLANCK DE ASTRONOMIA MPG190030” and “CONICYT+PCI+REDES 190194”.

## Appendix

### Impact of the Fidelity and $\delta v$ Choice on the Cross-Correlation Function

In this appendix, we explore how the fidelity and  $\delta v$  choices impact the cross-correlation measurement that we present in Section 4.

We first check the impact of the fidelity choice on the measurement. For this, we create source catalogs with different fidelity thresholds ranging from 0.7 to 0.9 in steps of 0.05 (to follow the same criteria as those described in Section 3.3, in all the cases we increase the fidelity threshold by 0.1 for sources selected at distances larger than  $50.14''$  from the central quasar, corresponding to the radius at which the telescope sensitivity is  $\geq 20\%$  of the maximum). We compute the cross-correlation function following the same procedure described in Section 4.1, and we fit the correlation with a fixed slope  $\gamma = 1.8$ . We show our results in Figure 7. We find that the cross-correlation decreases with increasing fidelity threshold, and it converges for fidelity  $\geq 0.8$ , which motivates the choice of this as the threshold for our study.

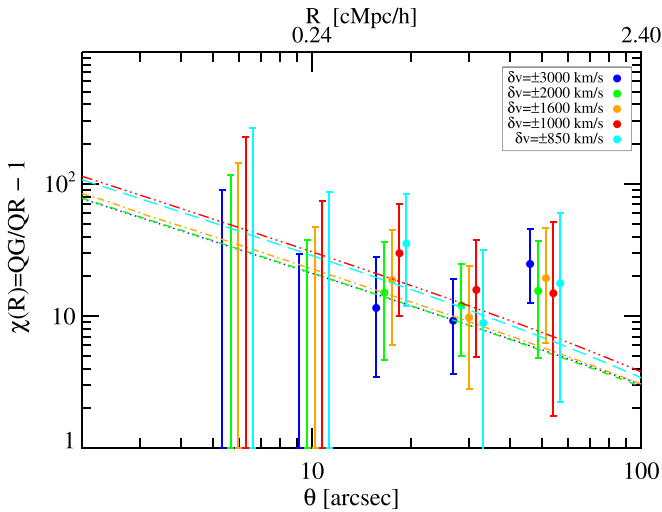


**Figure 7.** Impact of the fidelity threshold adopted to create the catalog of sources. The results converge at fidelity  $\geq 0.8$ . The clustering signal becomes flat when including sources with lower fidelities, which means that we are including contaminants (i.e., noise fluctuations are being detected as real sources by the algorithm).

We also note that, at lower fidelity ( $< 0.8$ ), we include more sources, but most of them are impacting the last bin of the measurement (at  $\theta \sim 50''$ ), making the cross-correlation flatter, with the fixed slope at  $\gamma = 1.8$  poorly constraining their shape. This suggests that we are including contamination in these samples (i.e., probably fake noise fluctuations that are detected as real sources by the line search algorithm), which dilute the power-law signal, causing the flattening.

We also explore the impact of the radial comoving range ( $\delta v$ ) choice. For this, we only focus on the sources with fidelity  $\geq 0.8$  (i.e., on the nine sources of Table 3), and compute the correlation function using different  $\delta v$  values ranging from  $850 \text{ km s}^{-1}$  to  $3000 \text{ km s}^{-1}$  (i.e., the whole cube). We note that some quasars show large redshift offsets, and thus their locations within the ALMA cube are not centered. Specifically, for the two cases where the ALMA redshift and optical redshifts have a difference of  $\sim 2000 \text{ km s}^{-1}$ , the quasar in the ALMA data cube is located at  $\sim 1000 \text{ km s}^{-1}$  from the edge of the cube (see Section 2.1). When exploring  $\delta v \leq 1000 \text{ km s}^{-1}$ , we are tracing the complete volume around all the quasars, but when exploring  $\delta v = \pm 2000 \text{ km s}^{-1}$ , we are only exploring an asymmetric volume covering  $+2000 \text{ km s}^{-1}$  and  $-1000 \text{ km s}^{-1}$  around these two quasars, affecting the observed  $\langle QG(R) \rangle$  value. Similarly, when exploring  $\delta v = \pm 3000 \text{ km s}^{-1}$ , most of the fields (13 out of 17) are affected by this issue. We have taken into account this effect when computing the  $\langle QR(R) \rangle$  term in Equation (4), such that we only integrate over the effective volume traced in each individual field. In this way, the volume traced to compute  $\langle QG(R) \rangle$  and  $\langle QR(R) \rangle$  is the same, resulting in a correct clustering computation.

We show our results in Figure 8. As mentioned in Section 4.1, the choice of a larger radial comoving range increases the statistics, reducing the error bars in the measurement, but it also dilutes the strong small-scale signal, because we are integrating over larger volumes and at large distances from the quasar where the number counts of the background start to be reached. This effect is relatively small, but can be seen to happen at  $\delta v \geq 1600 \text{ km s}^{-1}$ . Our choice of  $\delta v = \pm 1000 \text{ km s}^{-1}$  (corresponding



**Figure 8.** Impact of the radial comoving range around the quasar ( $\delta v$ ) used to measure the clustering. The clustering slightly decreases when increasing the  $\delta v$ , because small-scale signal starts to be diluted when integrating over a large volume, covering distances from the quasar at which the background number density is close to being reached.

to  $8.19 h^{-1} \text{ cMpc}$  at  $z = 3.87$ ) is a balance between having good statistics but not strongly diluting the signal.

### ORCID iDs

Cristina García-Vergara <https://orcid.org/0000-0002-4811-0677>  
 Matus Rybak <https://orcid.org/0000-0002-1383-0746>  
 Jacqueline Hodge <https://orcid.org/0000-0001-6586-8845>  
 Joseph F. Hennawi <https://orcid.org/0000-0002-7054-4332>  
 Roberto Decarli <https://orcid.org/0000-0002-2662-8803>  
 Jorge González-López <https://orcid.org/0000-0003-3926-1411>  
 Fabrizio Arrigoni-Battaia <https://orcid.org/0000-0002-4770-6137>  
 Manuel Aravena <https://orcid.org/0000-0002-6290-3198>  
 Emanuele P. Farina <https://orcid.org/0000-0002-6822-2254>

### References

- Adams, S. M., Martini, P., Croxall, K. V., Overzier, R. A., & Silverman, J. D. 2015, *MNRAS*, **448**, 1335
- Angulo, R. E., Springel, V., White, S. D. M., et al. 2012, *MNRAS*, **425**, 2722
- Appenzeller, I., & Rupprecht, G. 1992, *Mnscr*, **67**, 18
- Aravena, M., Decarli, R., González-López, J., et al. 2019, *ApJ*, **882**, 136
- Bacon, R., Accardo, M., Adjali, L., et al. 2010, *Proc. SPIE*, **7735**, 773508
- Balmaverde, B., Gilli, R., Mignoli, M., et al. 2017, *A&A*, **606**, A23
- Bañados, E., Venemans, B., Walter, F., et al. 2013, *ApJ*, **773**, 178
- Bañados, E., Venemans, B. P., Mazzucchelli, C., et al. 2018, *Natur*, **553**, 473
- Bertin, E., & Arnouts, S. 1996, *A&AS*, **117**, 393
- Bigiel, F., Leroy, A., Walter, F., et al. 2008, *AJ*, **136**, 2846
- Blain, A. W., Smail, I., Ivison, R. J., Kneib, J. P., & Frayer, D. T. 2002, *PhR*, **369**, 111
- Bruns, L. R., Wyithe, J. S. B., Bland-Hawthorn, J., & Dijkstra, M. 2012, *MNRAS*, **421**, 2543
- Buchner, J., Treister, E., Bauer, F. E., Sartori, L. F., & Schawinski, K. 2019, *ApJ*, **874**, 117
- Cantalupo, S., Lilly, S. J., & Haehnelt, M. G. 2012, *MNRAS*, **425**, 1992
- Capak, P. L., Riechers, D., Scoville, N. Z., et al. 2011, *Natur*, **470**, 233
- Carilli, C. L., & Walter, F. 2013, *ARA&A*, **51**, 105
- Coatman, L., Hewett, P. C., Banerji, M., et al. 2017, *MNRAS*, **465**, 2120
- Cole, S., & Kaiser, N. 1989, *MNRAS*, **237**, 1127
- Croom, S. M., Boyle, B. J., Shanks, T., et al. 2005, *MNRAS*, **356**, 415
- da Ângela, J., Shanks, T., Croom, S. M., et al. 2008, *MNRAS*, **383**, 565
- Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, *AJ*, **145**, 10
- Decarli, R., Aravena, M., Boogaard, L., et al. 2020, *ApJ*, **902**, 110
- Decarli, R., Walter, F., Aravena, M., et al. 2016, *ApJ*, **833**, 69
- Decarli, R., Walter, F., González-López, J., et al. 2019, *ApJ*, **882**, 138
- Decarli, R., Walter, F., Venemans, B. P., et al. 2017, *Natur*, **545**, 457
- Dodelson, S. 2003, *Modern Cosmology* (Amsterdam: Academic Press)
- Eftekharzadeh, S., Myers, A. D., White, M., et al. 2015, *MNRAS*, **453**, 2779
- Eisenstein, D. J., Weinberg, D. H., Agol, E., et al. 2011, *AJ*, **142**, 72
- Farina, E. P., Venemans, B. P., Decarli, R., et al. 2017, *ApJ*, **848**, 78
- Fossati, M., Fumagalli, M., Lofthouse, E. K., et al. 2021, *MNRAS*, **503**, 3044
- Francis, P. J., & Bland-Hawthorn, J. 2004, *MNRAS*, **353**, 301
- Fukugita, M., Shimasaku, K., & Ichikawa, T. 1995, *PASP*, **107**, 945
- García-Vergara, C., Hennawi, J. F., Barrientos, L. F., & Arrigoni Battaia, F. 2019, *ApJ*, **886**, 79
- García-Vergara, C., Hennawi, J. F., Barrientos, L. F., & Rix, H.-W. 2017, *ApJ*, **848**, 7
- García-Vergara, C., Hodge, J., Hennawi, J. F., et al. 2020, *ApJ*, **904**, 2
- Gardner, J. P., Mather, J. C., Clampin, M., et al. 2006, *SSRv*, **123**, 485
- Gehrels, N. 1986, *ApJ*, **303**, 336
- González-López, J., Decarli, R., Pavesi, R., et al. 2019, *ApJ*, **882**, 139
- Goto, T., Utsumi, Y., Kikuta, S., et al. 2017, *MNRAS*, **470**, L117
- Hayashi, M., Tadaki, K.-i., Kodama, T., et al. 2018, *ApJ*, **856**, 118
- Hennawi, J. F., Strauss, M. A., Oguri, M., et al. 2006, *AJ*, **131**, 1
- Husband, K., Bremer, M. N., Stanway, E. R., et al. 2013, *MNRAS*, **432**, 2869
- Kashikawa, N., Kitayama, T., Doi, M., et al. 2007, *ApJ*, **663**, 765
- Kennicutt, R. C. J. 1998, *ApJ*, **498**, 541
- Kikuta, S., Imanishi, M., Matsuoka, Y., et al. 2017, *ApJ*, **841**, 128
- Kim, S., Stiavelli, M., Trenti, M., et al. 2009, *ApJ*, **695**, 809
- Leroy, A. K., Walter, F., Brinks, E., et al. 2008, *AJ*, **136**, 2782
- Loiacono, F., Decarli, R., Gruppioni, C., et al. 2021, *A&A*, **646**, A76
- Mazzucchelli, C., Bañados, E., Decarli, R., et al. 2017, *ApJ*, **834**, 83
- McMullin, J. P., Waters, B., Schiebel, D., Young, W., & Golap, K. 2007, in *ASP Conf. Ser. 376, Astronomical Data Analysis Software and Systems XVI*, ed. R. A. Shaw, F. Hill, & D. J. Bell (San Francisco, CA: ASP), **127**
- Mo, H. J., & White, S. D. M. 1996, *MNRAS*, **282**, 347
- Morselli, L., Mignoli, M., Gilli, R., et al. 2014, *A&A*, **568**, A1
- Myers, A. D., Brunner, R. J., Richards, G. T., et al. 2006, *ApJ*, **638**, 622
- Noble, A. G., McDonald, M., Muzzin, A., et al. 2017, *ApJL*, **842**, L21
- Oke, J. B. 1974, *ApJS*, **27**, 21
- Ota, K., Venemans, B. P., Taniguchi, Y., et al. 2018, *ApJ*, **856**, 109
- Ouchi, M., Hamana, T., Shimasaku, K., et al. 2005, *ApJL*, **635**, L117
- Ouchi, M., Shimasaku, K., Furusawa, H., et al. 2010, *ApJ*, **723**, 869
- Ouchi, M., Shimasaku, K., Okamura, S., et al. 2004, *ApJ*, **611**, 685
- Padmanabhan, T. 2006, in *AIP Conf. Proc. 843, Graduate School in Astronomy: X*, ed. S. Daflon et al. (Melville, NY: AIP), **111**
- Pâris, I., Petitjean, P., Aubourg, É., et al. 2014, *A&A*, **563**, A54
- Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, *A&A*, **641**, A6
- Porciani, C., & Norberg, P. 2006, *MNRAS*, **371**, 1824
- Richards, G. T., Strauss, M. A., Fan, X., et al. 2006, *AJ*, **131**, 2766
- Richards, G. T., Vanden Berk, D. E., Reichard, T. A., et al. 2002, *AJ*, **124**, 1
- Rudnick, G., Hodge, J., Walter, F., et al. 2017, *ApJ*, **849**, 27
- Schneider, P. 2015, *Extragalactic Astronomy and Cosmology: An Introduction* (Springer: Berlin)
- Shen, Y., Brandt, W. N., Richards, G. T., et al. 2016, *ApJ*, **831**, 7
- Shen, Y., Hennawi, J. F., Shankar, F., et al. 2010, *ApJ*, **719**, 1693
- Shen, Y., Strauss, M. A., Oguri, M., et al. 2007, *AJ*, **133**, 2222
- Sheth, R. K., & Tormen, G. 1999, *MNRAS*, **308**, 119
- Simpson, C., Mortlock, D., Warren, S., et al. 2014, *MNRAS*, **442**, 3454
- Skelton, R. E., Whitaker, K. E., Momcheva, I. G., et al. 2014, *ApJS*, **214**, 24
- Solomon, P. M., Downes, D., Radford, S. J. E., & Barrett, J. W. 1997, *ApJ*, **478**, 144
- Springel, V., White, S. D. M., Jenkins, A., et al. 2005, *Natur*, **435**, 629
- Stach, S. M., Smail, I., Amvrosiadis, A., et al. 2021, *MNRAS*, **504**, 172
- Stiavelli, M., Djorgovski, S. G., Pavlovsky, C., et al. 2005, *ApJL*, **622**, L1
- Swinbank, J., Baker, J., Barr, J., Hook, I., & Bland-Hawthorn, J. 2012, *MNRAS*, **422**, 2980
- Timlin, J. D., Ross, N. P., Richards, G. T., et al. 2018, *ApJ*, **859**, 20
- Trakhtenbrot, B., Lira, P., Netzer, H., et al. 2017, *ApJ*, **836**, 8
- Utsumi, Y., Goto, T., Kashikawa, N., et al. 2010, *ApJ*, **721**, 1680
- Venemans, B. P., Walter, F., Neelaman, M., et al. 2020, *ApJ*, **904**, 130
- Walter, F., Decarli, R., Aravena, M., et al. 2016, *ApJ*, **833**, 67
- White, M., Myers, A. D., Ross, N. P., et al. 2012, *MNRAS*, **424**, 933
- Willott, C. J., Percival, W. J., McLure, R. J., et al. 2005, *ApJ*, **626**, 657
- Yang, J., Wang, F., Fan, X., et al. 2020, *ApJL*, **897**, L14
- York, D. G., Adelmann, J., Anderson, J. E., Jr., et al. 2000, *AJ*, **120**, 1579
- Zheng, W., Overzier, R. A., Bouwens, R. J., et al. 2006, *ApJ*, **640**, 574