



Ubiquitous Molecular Outflows in $z > 4$ Massive Dusty Galaxies. I. Sample Overview and Clumpy Structure in Molecular Outflows on 500 pc Scales

Justin S. Spilker^{1,18}, Kedar A. Phadke², Manuel Aravena³, Matthieu Béthermin⁴, Scott C. Chapman^{5,6,7},
 Chenxing Dong (董辰兴)⁸, Anthony H. Gonzalez⁹, Christopher C. Hayward⁹, Yashar D. Hezaveh^{8,10},
 Sreevani Jarugula², Katrina C. Litke¹¹, Matthew A. Malkan¹², Daniel P. Marrone¹¹, Desika Narayanan^{8,13,14},
 Cassie Reuter², Joaquin D. Vieira^{2,15,16}, and Axel Weiss¹⁷

¹ Department of Astronomy, University of Texas at Austin, 2515 Speedway Stop C1400, Austin, TX 78712, USA; spilkerj@gmail.com

² Department of Astronomy, University of Illinois, 1002 West Green St., Urbana, IL 61801, USA

³ Núcleo de Astronomía de la Facultad de Ingeniería y Ciencias, Universidad Diego Portales, Av. Ejército Libertador 441, Santiago, Chile

⁴ Aix Marseille Univ., CNRS, CNES, LAM, Marseille, France

⁵ Department of Physics and Astronomy, University of British Columbia, 6225 Agricultural Rd., Vancouver, V6T 1Z1, Canada

⁶ National Research Council Herzberg Astronomy and Astrophysics, 5071 West Saanich Rd., Victoria, V9E 2E7, Canada

⁷ Department of Physics and Atmospheric Science, Dalhousie University, Halifax, Nova Scotia, Canada

⁸ Department of Astronomy, University of Florida, 211 Bryant Space Sciences Center, Gainesville, FL 32611, USA

⁹ Center for Computational Astrophysics, Flatiron Institute, 162 Fifth Avenue, New York, NY 10010, USA

¹⁰ Département de Physique, Université de Montréal, Montréal, Québec H3T 1J4, Canada

¹¹ Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721, USA

¹² Department of Physics and Astronomy, University of California, Los Angeles, CA 90095-1547, USA

¹³ University of Florida Informatics Institute, 432 Newell Drive, CISE Bldg E251, Gainesville, FL 32611, USA

¹⁴ Cosmic Dawn Center (DAWN), DTU-Space, Technical University of Denmark, Elektrovej 327, DK-2800 Kgs. Lyngby, Denmark

¹⁵ Department of Physics, University of Illinois, 1110 West Green St., Urbana, IL 61801, USA

¹⁶ National Center for Supercomputing Applications, University of Illinois, 1205 West Clark St., Urbana, IL 61801, USA

¹⁷ Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, D-53121 Bonn, Germany

Received 2020 August 25; revised 2020 October 13; accepted 2020 October 22; published 2020 December 16

Abstract

Massive galaxy-scale outflows of gas are one of the most commonly invoked mechanisms to regulate the growth and evolution of galaxies throughout the universe. While the gas in outflows spans a large range of temperatures and densities, the cold molecular phase is of particular interest because molecular outflows may be capable of suppressing star formation in galaxies by removing the star-forming gas. We have conducted the first survey of molecular outflows at $z > 4$, targeting 11 strongly lensed dusty, star-forming galaxies (DSFGs) with high-resolution Atacama Large Millimeter/submillimeter Array observations of OH 119 μm absorption as an outflow tracer. In this first paper, we give an overview of the survey, focusing on the detection rate and structure of molecular outflows. We find unambiguous evidence for outflows in 8/11 galaxies (73%), more than tripling the number known at $z > 4$. This implies that molecular winds in $z > 4$ DSFGs must have both a near-unity occurrence rate and large opening angles to be detectable in absorption. Lensing reconstructions reveal that 500 pc scale clumpy structures in the outflows are common. The individual clumps are not directly resolved, but from optical depth arguments we expect that future observations will require 50–200 pc spatial resolution to resolve them. We do not detect high-velocity [C I] wings in any of the sources with clear OH outflows, indicating that [C II] is not a reliable tracer of molecular outflows. Our results represent a first step toward characterizing molecular outflows at $z > 4$ at the population level, demonstrating that large-scale outflows are ubiquitous among early massive, dusty galaxies.

Unified Astronomy Thesaurus concepts: [High-redshift galaxies \(734\)](#); [Galactic winds \(572\)](#); [Gravitational lensing \(670\)](#); [Galaxy formation \(595\)](#)

1. Introduction

Galactic feedback is now widely recognized as a key component in our modern understanding of galaxy formation and evolution. “Feedback” is an umbrella term for a wide range of physical processes enabling self-regulated galaxy growth, setting the efficiency of star formation and shaping fundamental correlations between galaxy properties such as stellar mass, metallicity, star formation rate (SFR), and supermassive black hole mass. One of the most striking observational windows into galactic feedback is the ubiquitous detection of massive outflows of gas and dust being launched from galaxies, generally thought to be powered by supernovae and/or supermassive black hole accretion. Outflows of ionized and

neutral atomic gas have been detected in galaxies over a wide range of mass and redshift for decades (e.g., Heckman et al. 1990; Rupke et al. 2005; Weiner et al. 2009; Chisholm et al. 2015), including in massive dusty galaxies at high redshifts (e.g., Banerji et al. 2011; Casey et al. 2017; Schechter & Casey 2018). More recently, observations that probe the cold molecular gas in outflows have been a focus of recent interest because molecular gas is the direct fuel for future star formation and is often the dominant phase in the outflow mass budget (see Veilleux et al. 2020 for a recent review of cold galactic winds).

In the high-redshift universe, spatially resolved studies of massive quenching galaxies at $z \sim 2$ typically find evidence for an inside-out suppression of star formation (e.g., Tacchella et al. 2015, 2018; Nelson et al. 2016; Spilker et al. 2019)

¹⁸ NHFP Hubble Fellow.

accompanying a sharp overall decrease in the molecular gas fraction compared to equally massive star-forming galaxies at the same epoch (e.g., Spilker et al. 2016a; Popping et al. 2017; Tadaki et al. 2017; Talia et al. 2018). Indeed, massive ($M_{\text{star}} \sim 10^{10} M_{\odot}$) quiescent galaxies have been identified in sizable numbers as early as $z \sim 4$ (e.g., Straatman et al. 2014; Guarnieri et al. 2019; Carnall et al. 2020; Valentino et al. 2020), implying a very rapid formation history with SFRs of hundreds of $M_{\odot} \text{ yr}^{-1}$ and subsequent rapid quenching of star formation (e.g., Glazebrook et al. 2017; Estrada-Carpenter et al. 2020; Forrest et al. 2020). The required high SFRs are generally only found in very infrared-luminous systems, in which the UV radiation from young stars is absorbed and reprocessed by dust. These observations paint an appealing picture in which initially gas-rich, dusty, star-forming galaxies (DSFGs) at least temporarily suppress star formation via the consumption, heating, and/or ejection of the molecular gas fuel, through a self-regulating feedback process or processes in order to create the early passive galaxy population (e.g., Narayanan et al. 2015).

There are two primary tracers of the cold molecular phase of galactic winds in both the nearby and distant universe, either of which is easily detectable: carbon monoxide (CO) and the hydroxyl molecule OH (e.g., Veilleux et al. 2020). Low-order transitions of CO can be used to detect cold molecular outflows just as they are often used to probe the overall molecular contents of galaxies more generally (e.g., Walter et al. 2002; Alatalo et al. 2011; Barcos-Muñoz et al. 2018). For unresolved observations, the outflow signature is an excess of CO emission at high velocities relative to systemic that is not plausibly related to rotational or non-rotational motions within the galaxies. CO observations have the benefit of sensitivity to gas at all distances and lines of sight to the host galaxy, but the line wings are very faint and the geometry of the emitting gas is difficult to constrain (for example, it is hard to distinguish outflowing from inflowing gas because the line-of-sight location of the emission is unknown, or to rule out that the emission is from a separate galaxy in a merger that may not be apparent even in deep imaging data). An alternative is far-infrared transitions of OH, demonstrated to be a very good tracer of outflowing and inflowing gas in dozens of nearby galaxies over the lifetime of the Herschel Space Observatory (e.g., Sturm et al. 2011; Spoon et al. 2013; Veilleux et al. 2013; Stone et al. 2016; González-Alfonso et al. 2017). In this case the outflow (or inflow) signature takes the form of broad blueshifted (redshifted) absorption profiles against the continuum emission of the host galaxies. Because the gas flows are seen in absorption, the geometrical interpretation of the line profiles is simplified, but OH studies consequently require galaxies with bright continuum emission, and are not sensitive to outflowing material that does not intersect the line of sight to the hosts.

In the best-studied examples in the local universe, the geometry and structure of the winds can be spatially resolved, allowing for detailed pictures of the location, kinematics, and conditions within the molecular gas contained in the outflows. The prototypical starburst-driven outflows in M82 and NGC 253, for example, are both seen nearly edge-on, and both show clumpy streamers of molecular gas extending a kiloparsec or more out of the disks (e.g., Walter et al. 2002, 2017; Leroy et al. 2015; Krieger et al. 2019). Meanwhile, very high-resolution observations of the outflow from the nuclear region

of the nearby Seyfert galaxy NGC 1068 in multiple molecular tracers reveal the impact of the wind from the active galactic nucleus (AGN) on the surrounding torus, also likely extending to larger spatial scales in the host galaxy (e.g., García-Burillo et al. 2019).

Such detailed views of molecular outflows have thus far been confined to the local universe. Despite a handful of past successes in merely detecting molecular outflows in the early universe at all continues to be extremely challenging: detecting the faint CO line wings associated with outflows requires substantial observational investments even with the Atacama Large Millimeter/submillimeter Array (ALMA) and is generally only possible out to $z \sim 2$ except in extreme cases, the detection of OH absorption requires bright continuum fluxes that limit the plausible target galaxies to very IR-luminous QSOs and DSFGs. Thus far only three objects at $z > 4$ have molecular outflows reported (Spilker et al. 2018; Jones et al. 2019; Herrera-Camus et al. 2020), and even in these cases the interpretation of the observations is not necessarily clear-cut given limitations in signal-to-noise ratio (S/N) and the complex galactic dynamics at play in the early universe. All told, at best a handful of galaxies at $z > 2$ have molecular outflows detected and while selected in very heterogeneous ways, all are limited to luminous dusty galaxies and/or AGN hosts (Weiß et al. 2012; George et al. 2014; Feruglio et al. 2017; Fan et al. 2018; Herrera-Camus et al. 2019).

The structure of molecular outflows is of special interest because the emergence of molecular winds with properties similar to those observed in real galaxies has proven to be especially challenging for hydrodynamical simulations. In particular, accelerating molecular gas initially at rest up to gas velocities of hundreds of km s^{-1} through direct ram pressure from a hot, fast wind or through entrainment in such a wind has proven extremely difficult. The cold and dense gas is shredded by hydrodynamical instabilities long before it reaches speeds like those observed in real galaxies (e.g., Klein et al. 1994; Scannapieco 2013; Schneider & Robertson 2017). One possible alternative is that molecules in outflows re-form at large galactocentric distances, cooling out of a hotter wind fluid having already reached velocities like those observed (e.g., Zubovas & King 2014; McCourt et al. 2018; Richings & Faucher-Giguère 2018; Schneider et al. 2018). Depending on the details of the simulation and the outflow energetics, this cold gas can show either kiloparsec-scale clumpy structures or a fine mist-like morphology on very small scales.

This is the first in a series of papers in which we present the first constraints on the occurrence, structure, and physical properties of molecular outflows in a sample of DSFGs at $z > 4$ targeting the OH 119 μm doublet. This sample expands on our work in Spilker et al. (2018, hereafter S18), in which we reported the highest-redshift detection of a molecular outflow toward a $z = 5.3$ galaxy. Our targets are gravitationally lensed by foreground galaxies, which allows us to spatially resolve both the rest-frame 120 μm dust continuum emission as well as the OH absorption at systemic and blueshifted velocities. While still not representative of the general population of high-redshift galaxies, our goal with this survey is to take a first step toward constraining the occurrence rate and properties of molecular outflows in the early universe in a statistical sense at the population level.

In this work we give an overview of the sample objects and present our new ALMA data. We focus here on the broad

sample properties, outflow detection rates, and the resolved structure of the molecular outflows we determine from gravitational lensing reconstructions of the sources. In a companion paper (Spilker et al. 2020, hereafter Paper II) we characterize the physical properties of the molecular outflows we detect, focusing on the outflow rates, energetics, and wind driving mechanisms. Section 2 gives an overview of the sample, objects, ALMA observations, and ancillary data for our objects and literature comparison samples. Section 3 describes our analysis methods for the OH spectra, how we classify whether or not objects show signs of outflow, and our lens modeling methodology and tests. Section 4 gives our main observational results, with additional discussion in Section 5. We summarize our principal findings and conclude in Section 6. We assume a flat Λ CDM cosmology with $\Omega_m = 0.307$ and $H_0 = 67.7 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (Planck Collaboration et al. 2016), and we take the total infrared and far-infrared luminosities L_{IR} and L_{FIR} to be integrated over rest-frame 8–1000 and 40–120 μm , respectively. Tables of the sample properties from this work, as well as the outflow properties from Paper I, are available in electronic form at <https://github.com/spt-smg/publicdata>.

2. Sample and Observations

2.1. Parent Sample and Source Selection

We designed an observing campaign targeting the OH $^2\Pi_{3/2}$ $J = 3/2 \rightarrow 5/2$ absorption. This transition is a Λ doublet with components at rest-frame 2509.9 and 2514.3 GHz (separated by $\sim 520 \text{ km s}^{-1}$) and additional hyperfine structure that remains spectrally unresolved. We selected sources for OH observations from the point-source catalog of the 2500 deg² SPT survey at 1.4 and 2 mm (Vieira et al. 2010; Carlstrom et al. 2011; Mocanu et al. 2013; Everett et al. 2020). From the survey data and subsequent observations using the APEX/LABOCA camera at 870 μm , a total of 81 objects were selected with spectral indices consistent with thermal dust emission (namely $S_{1.4 \text{ mm}}/S_{2 \text{ mm}} > 1.8$), a 1.4 mm flux density greater than 20 mJy, flux density at 870 μm greater than 25 mJy, and no detection in various shallow multiwavelength surveys to reject low-redshift interlopers. Given their extreme brightness, the vast majority of these sources were expected to be gravitationally lensed by foreground galaxies. High-resolution ALMA imaging confirmed the lensed nature of these sources with typical magnifications of 3–30 (Hezaveh et al. 2013; Vieira et al. 2013; Spilker et al. 2016b). Extensive spectroscopic campaigns subsequently measured spectroscopic redshifts for the entire sample, which range from 1.87 to 6.90 with a median of 3.9 (Weiß et al. 2013; Strandet et al. 2016; Marrone et al. 2018; Reuter et al. 2020), although not all sources had known redshifts at the time the present outflow survey was designed.

Our primary selection criterion for OH 119 μm observations was that the source redshift place the OH doublet lines at frequencies of relatively good atmospheric transmission in ALMA Band 8 (385–500 GHz), requiring $z_{\text{source}} \leq 4.02$ (to reach ALMA Band 7 requires $z_{\text{source}} > 5.8$, where we have few available targets). The atmospheric transmission at these frequencies is strongly affected by telluric water, oxygen, and ozone features, so OH observations are not feasible for all sources at $z > 4$. In particular, OH observations are not possible for redshift windows 4.07–4.30, 4.30–4.38, 4.38–4.68, 4.86–4.99, and 5.43–5.53 due to atmospheric

features. We restricted the sample to sources with redshifts that avoided frequencies of poor transmission, and predicted 119 μm continuum flux densities bright enough that ALMA would be able to reach sensitivities of 5% of the continuum level in $\sim 200 \text{ km s}^{-1}$ channels in less than an hour of observing time after resolving the source over 5–20 resolution elements. These predicted continuum flux densities were estimated using the available far-IR photometry, which provides very good sampling of the long-wavelength spectral energy distributions (SEDs; Section 2.3). We finally required that all targets have lens models from ALMA 870 μm observations (Spilker et al. 2016b) and chose objects to span a wide range in L_{IR} . The sources selected for OH observations are not obviously biased with respect to the full SPT sample of $z > 4$ DSFGs in intrinsic (lensing-corrected) dust mass, or effective dust temperature (Reuter et al. 2020), although this remains somewhat uncertain because we lack lens models for every SPT DSFG. Even after lensing correction these remain extremely luminous objects ($\log(L_{\text{IR}}/L_{\odot}) = 12.5\text{--}13.5$); they are certainly not “typical” galaxies at these redshifts by any conceivable definition.

The final sample consists of 11 objects at $0.9 < z < 5.30$ including SPT2319-55, previously published in S18. Basic properties of the sample are given in Tables 1 and 2, with a few salient properties shown in Figure 1.

2.2. ALMA Observations

ALMA observed our sample galaxies across several projects from 2016 to 2019, summarized in Table 1. For each object, we configured the correlator to observe the OH doublet with two slightly overlapping 1.875 GHz wide basebands and 3.9 MHz channels, providing contiguous coverage over 2200–2700 km s^{-1} around the OH lines, depending on the redshift of each source. These basebands were placed such that the lower-frequency edge corresponded to $\sim 1200 \text{ km s}^{-1}$ redward of the upper OH transition (or $\sim 680 \text{ km s}^{-1}$ redward of the lower-frequency transition), leaving $\sim 1000\text{--}1700 \text{ km s}^{-1}$ to the blueshifted side of the upper-frequency doublet transition. This setup was chosen to maximize the amount of blueshifted velocity coverage while still allowing both doublet transitions to be detected. Unfortunately it does not allow for the detection of strongly redshifted emission (or absorption), as expected for the classical P Cygni profile and sometimes observed in the OH spectra of local ultraluminous infrared galaxies (ULIRGs) and quasar hosts (e.g., Veilleux et al. 2013). We also placed an additional two basebands of 1.875 GHz width each for continuum coverage in the other sideband of the ALMA correlator. Given the fixed 4–8 GHz intermediate frequency of the ALMA Band 8 receivers, these continuum measurements are centered either 12 GHz above or 12 GHz below the OH frequencies in Table 1, depending on the atmospheric transmission. For SPT0459-59 and SPT2132-58 the atmospheric transmission is poor both above and below the OH observed frequencies, making half (SPT0459-59) or all (SPT2132-58) of the continuum bandwidth unusable.

The observing time and requested spatial resolution were estimated using the available far-IR and (sub)millimeter photometry and the lens models available from high-resolution ALMA imaging for each individual source (Spilker et al. 2016b; Reuter et al. 2020). The array configuration(s) varied for each source, with maximum baseline lengths ranging from 700 to 1600 m and minimum baseline lengths $\sim 15 \text{ m}$ for all observations. The shortest baselines lead to maximum

Table 1
Summary of ALMA Observations

Source	R.A.	Decl.	V_{obs} (GHz)	Program ID	t_{obs} (minutes)	Beam Size (arcsec)	σ_{cont} ($\mu\text{Jy beam}^{-1}$)	$S_{100\text{ km s}^{-1}}$ (mJy beam $^{-1}$)
SPT0202-61	02 ^h 02 ^m 58 ^s .86	- 61 [°] 21 ['] 11 ["] .1	417.8	2016.1.00089.S	43	0.37 $\square$ × $\square$ 0.45	64	0.41
SPT0418-47	04 ^h 18 ^m 39 ^s .67	- 47 [°] 51 ['] 52 ["] .7	481.2	2015.1.00942.S	12	0.11 $\square$ × $\square$ 0.16	250	1.28
				2018.1.00191.S	87	0.32 $\square$ × $\square$ 0.46	110	0.67
				2015.1.00942.S	11	0.14 $\square$ × $\square$ 0.18	352	2.10
SPT0441-46	04 ^h 41 ^m 44 ^s .08	- 46 [°] 05 ['] 25 ["] .5	459.1	2016.1.00089.S	17	0.25 $\square$ × $\square$ 0.33	205	0.89
				Combined	28	0.22 $\square$ × $\square$ 0.30	178	0.84
				2019.1.00253.S	40	0.33 $\square$ × $\square$ 0.40	82	0.69
SPT0459-58	04 ^h 58 ^m 59 ^s .80	- 58 [°] 05 ['] 14 ["] .3	429.4	2019.1.00253.S	41	0.33 $\square$ × $\square$ 0.41	120	0.56
SPT0544-40	05 ^h 44 ^m 00 ^s .80	- 40 [°] 36 ['] 31 ["] .1	477.2	2019.1.00253.S	32	0.28 $\square$ × $\square$ 0.31	103	1.00
SPT2048-55	20 ^h 48 ^m 22 ^s .86	- 55 [°] 20 ['] 21 ["] .3	493.7	2018.1.00191.S	48	0.37 $\square$ × $\square$ 0.44	118	0.72
SPT2103-60	21 ^h 03 ^m 30 ^s .85	- 60 [°] 32 ['] 40 ["] .5	462.6	2016.1.00089.S	23	0.46 $\square$ × $\square$ 0.49	238	1.10
SPT2132-58	21 ^h 32 ^m 43 ^s .23	- 58 [°] 02 ['] 46 ["] .2	435.9	2015.1.00942.S	17	0.32 $\square$ × $\square$ 0.50	345	1.64
SPT2311-54	23 ^h 11 ^m 23 ^s .97	- 54 [°] 50 ['] 30 ["] .2	476.2	2015.1.00942.S	45	0.15 $\square$ × $\square$ 0.20	157	1.27
				2018.1.00191.S	49	0.28 $\square$ × $\square$ 0.37	91	0.92
				Combined	94	0.23 $\square$ × $\square$ 0.30	80	0.75
SPT2319-55	23 ^h 19 ^m 21 ^s .67	- 55 [°] 57 ['] 57 ["] .8	399.4	2016.1.00089.S	30	0.27 $\square$ × $\square$ 0.39	71	0.52

Notes. All beam sizes and sensitivities are measured from naturally weighted images. The spectral line sensitivity is measured in a 100 km $^{-1}$ channel near the upper OH rest frequency.

^a SPT0418-47 was observed in 2015.1.00942.S at much higher spatial resolution than requested. Given the large extent of the source and the short observing duration, these data have too low an S/N to be usable and are excluded from all figures and models. We also exclude the observations in this table for completeness.

^b Reproduced from Spilker et al. (2018).

Table 2
Summary of Sample Properties

Source	z_{ens}	z_{source}	μ	$L_{\text{IR}} (10^{12} L_{\odot})$	$L_{\text{FIR}} (10^{12} L_{\odot})$	f_{AGN}	$M_{\text{H}_2} (10^9 M_{\odot})$	$r_{\text{cont}} (\text{kpc})$
SPT0202-61	L	5.0180	17.5	9.6 $\square$ ± $\square$ 1.5	4.6 $\square$ ± $\square$ 0.6	<0.25	25.1 $\square$ ± $\square$ 3.5	0.72
SPT0418-47	0.26	4.2248	37.2	3.0 $\square$ ± $\square$ 0.5	1.7 $\square$ ± $\square$ 0.2	<0.1	6.0 $\square$ ± $\square$ 0.5	0.74
SPT0441-46	0.88	4.4770	11.5	6.1 $\square$ ± $\square$ 1.3	3.5 $\square$ ± $\square$ 0.6	<0.15	12.3 $\square$ ± $\square$ 2.0	0.53
SPT0459-58	L	4.8560	7.3	8.1 $\square$ ± $\square$ 2.0	4.5 $\square$ ± $\square$ 0.8	<0.2	27.4 $\square$ ± $\square$ 3.3	1.22
SPT0459-59	0.94	4.7993	3.1	18.1 $\square$ ± $\square$ 5.7	9.9 $\square$ ± $\square$ 2.0	<0.3	79.9 $\square$ ± $\square$ 7.0	3.99
SPT0544-40	L	4.2692	10.5	7.3 $\square$ ± $\square$ 1.1	4.3 $\square$ ± $\square$ 0.6	<0.2	46.6 $\square$ ± $\square$ 4.3	0.69
SPT2048-55	L	4.0923	10.8	4.5 $\square$ ± $\square$ 1.2	2.6 $\square$ ± $\square$ 0.5	<0.2	16.0 $\square$ ± $\square$ 2.6	0.67
SPT2103-60	0.76	4.4357	20.9	2.9 $\square$ ± $\square$ 0.5	1.7 $\square$ ± $\square$ 0.3	<0.15	9.8 $\square$ ± $\square$ 1.7	1.02
SPT2132-58	L	4.7677	5.7	11.3 $\square$ ± $\square$ 4.3	6.2 $\square$ ± $\square$ 1.4	<0.25	27.6 $\square$ ± $\square$ 2.6	0.78
SPT2311-54	0.44	4.2795	2.5	29.8 $\square$ ± $\square$ 7.9	16.2 $\square$ ± $\square$ 2.9	<0.45	63.5 $\square$ ± $\square$ 4.9	1.08
SPT2319-55	0.91	5.2943	5.8	7.9 $\square$ ± $\square$ 3.0	4.3 $\square$ ± $\square$ 0.8	<0.3	11.8 $\square$ ± $\square$ 2.1	0.92

Notes. L_{IR} and L_{FIR} are integrated over rest-frame 8–1000 and 40–120 μm , respectively. All values have been corrected for the lensing magnification μ ; we estimate uncertainties of $\sim 15\%$ on the magnification. f_{AGN} is the fractional contribution of AGN-heated dust to the rest-frame 5–1000 μm luminosity; upper limits are 1σ from Aravena et al. (2016) and M. Aravena et al. (2020, in preparation) using updated magnifications from this work. Intrinsic dust continuum sizes r_{cont} are circularized radii of the regions where the continuum is detected at $\text{S/N} > 5$ in the lensing reconstruction. This table is available in machine-readable format at <https://github.com/spt-smg/publicdata>.

^a Excludes the faint source southwest of the lensed source (see Figure 2).

^b Reproduced from Spilker et al. (2018).

recoverable scales ~ 2.5 depending on observing frequency; attempted one or two rounds of phase-only self-calibration on we do not expect significant emission on larger spatial scales. the test and target sources, using solution intervals of the scan The data were reduced with the standard pipelines available for length or half the scan length. This self-calibration was each ALMA cycle, with additional manual calibration and flagging where necessary. Besides the typical bandpass flux, and complex gain calibrators, all observing blocks also successful for all sources except SPT2103-60, decreasing the image rms by up to a factor of two. We note that self-calibration makes absolute astrometry impossible, so the recorded data for a quasar near each DSFG used as a test the astrometry and calibration quality; this is standard for high- $\sim 0.''1$, as measured from observations of the test source before self-calibration. Self-calibration has no influence on the relative astrometry within the ALMA data (e.g., across the line profiles of residual atmospheric phase noise varying faster than the astrometry or between sidebands). We generate images of each target using natural weighting of the visibilities, which maximizes sensitivity at the expense

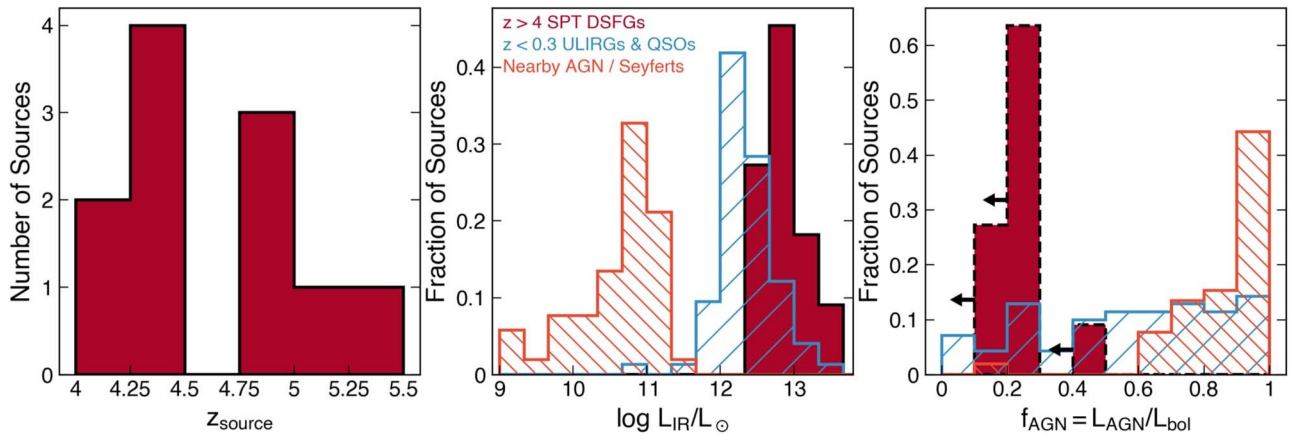


Figure 1. Summary of source properties for our SPT-selected sample of high-redshift DSFG outflows and literature comparison samples detailed in Section 2.4. All f_{AGN} values for the SPT sample are upper limits, although the rest-frame mid-IR data used to compute f_{AGN} fail in the event of extremely high column densities hiding highly obscured AGN.

of spatial resolution. Natural weighting is also the closest approximation to the visibility weighting used in our subsequent visibility-based lens modeling procedure (Section 3.3). We followed standard imaging procedures manually applying a clean mask over regions with clear, high-S/N emission, and stopped the image cleaning at five times the image noise level. Continuum images were created combining all available data, while image cubes of the OH lines were created with channel resolutions varying from 50 to 150 km s⁻¹ in order to maximize the S/N. To extract integrated OH spectra, we performed aperture photometry within the region where the continuum is detected at $>3\sigma$. We also performed a similar procedure on image cubes created by tapering the visibilities to resolutions ~ 2 – 3 times lower than the full data and find no evidence that significant flux has been resolved out in our observations. Continuum images of each source are shown in Figure 2 and integrated spectra in Figure 3.

2.3. Ancillary Data

In addition to the ALMA OH observations that are our primary focus, we also use a variety of ancillary photometric and spectroscopic data to aid in the interpretation of the OH data.

The systemic redshift and line profile of gas within each galaxy are key to our interpretation of the OH spectra where available (five sources). We use very high-S/N ALMA [C II] 158 μ m spectra, observed in program 2016.1.01499.S (see Litke et al. 2019 for a representative object from this sample). While these data have fairly high spatial resolution, $\sim 0.3''$, we use only the integrated [C II] line profile extracted similarly to the OH spectra. For those sources without high-quality [C II] data, we instead stack the spectra of all available transitions of CO for each source, weighted by the S/N of each line. These CO lines were observed with ALMA and the Australia Telescope Compact Array (ATCA), and were the primary features used to measure the redshift of each source. The CO lines include CO(2–1) and CO(5–4) for all sources and CO(4–3) for $z \lesssim 4.4$. Each is typically detected at S/N ~ 5 – 10 . For the sources with ALMA [C II] spectra, we find no evidence for a difference in line width compared to the CO lines with lower S/N.

We measure the IR (8–1000 μ m) and far-IR (40–120 μ m) luminosities by fitting to the available far-IR and submillimeter

photometry (Weiß et al. 2013; Strandet et al. 2016; Reuter et al. 2020), correcting for the lensing magnifications of each source as described further in Section 3.3. For all sources, the available data include Herschel/PACS and SPIRE data at 100, 160, 250, 350, and 500 μ m, APEX/LABOCA 870 μ m, the SPT 1.4, 2, and 3 mm photometry, and ALMA 870 μ m and 3 mm data. We find consistent results for the luminosities between simple modified blackbody fits and more complex modeling because the far-IR SED is very well sampled. We also make use of the PACS photometry at 100 and 160 μ m, which probes rest-frame mid-IR wavelengths ~ 15 – 30μ m, to constrain the contribution of hot dust heated by AGN activity. We use the CIGALE SED fitting code (Burgarella et al. 2005; Boquien et al. 2019) to place limits on the fractional AGN contribution to the total luminosity integrated over rest-frame 5–1000 μ m, f_{AGN} ; no source shows strong evidence for AGN-related mid-IR emission. For the low-redshift comparison samples (Section 2.4) it is more common to measure f_{AGN} using the rest-frame 30 μ m/15 μ m flux ratio assuming fixed mid-IR flux ratios for pure star formation and pure AGN emission. While we prefer the CIGALE fitting values for easier comparison with typical practice in the extragalactic literature, we have verified that we recover f_{AGN} to within ~ 0.2 – 0.3 using the mid-IR color definition. We note that it is possible that our target galaxies are optically thick at mid-IR wavelengths, which could hide very highly obscured AGN and result in less strict limits on f_{AGN} than we adopt here (e.g. Snyder et al. 2013).

Masses of molecular gas were measured from observations of CO(2–1) using ATCA. Five of the 11 sources in our sample were published in Aravena et al. (2016), while the remainder have been observed and analyzed using the same procedures and will be published elsewhere. We assume a line brightness temperature ratio of 0.9 between the CO(2–1) and (1–0) transitions and a CO–H₂ conversion factor $\alpha_{\text{CO}} = 0.8 \text{ M} (\text{K km s}^{-1} \text{ pc}^2)^{-1}$, both typical for highly star-forming DSFGs like our sample (Spilker et al. 2014, 2015; Aravena et al. 2016).

2.4. Literature Comparison Samples

Throughout this work we compare to a number of studies of OH absorption in low-redshift galaxies performed by the Herschel/PACS instrument. While detailed sensitivity metrics (e.g., the typical fractional contrast compared to the continuum reached per spectral element) are generally not given, the

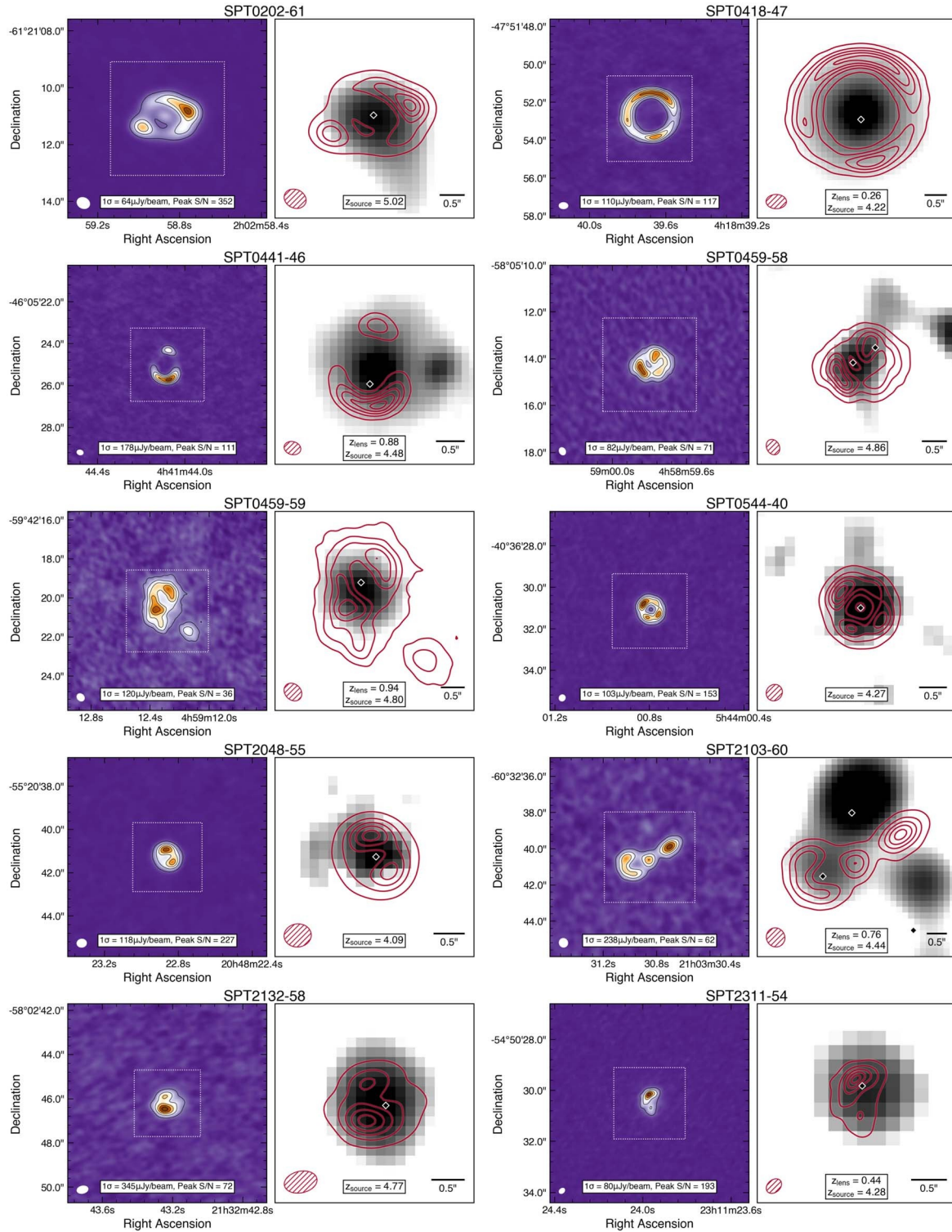


Figure 2. Left: $\square$ rest-frame 119 μm ALMA continuum images of each sample object. Contours are drawn at 10%, 30%, 50%, 70%, and 90% of the peak. Squares mark the zoomed regions in the right panels. Right: ALMA continuum contours overlaid on the best available near-IR images from several different facilities (Spilker et al. 2016b), which detect only the foreground lensing galaxies. Diamonds mark the best-fit positions of the lens(es); we do not use these images to constrain the positions because the astrometry is more uncertain than the uncertainties on the positions. Squares at lower left show the synthesized beam.

published spectra of these sources appear to be of broadly similar quality to our own. A brief description of these samples follows, and a few relevant quantities are summarized in Figure 1.

Veilleux et al. (2013) present OH spectra of 43 nearby galaxy mergers, mainly consisting of ULIRGs and IR-luminous

QSOs. The sample supplements 23 galaxies observed as part of the SHINING key program (Sturm et al. 2011) with 15 additional sources selected to have higher values of α and a further five chosen to be less IR-luminous than the full sample. Spoon et al. (2013) present an analogous sample of 24 ULIRGs from the HERUS program (Farrah et al. 2013) that largely

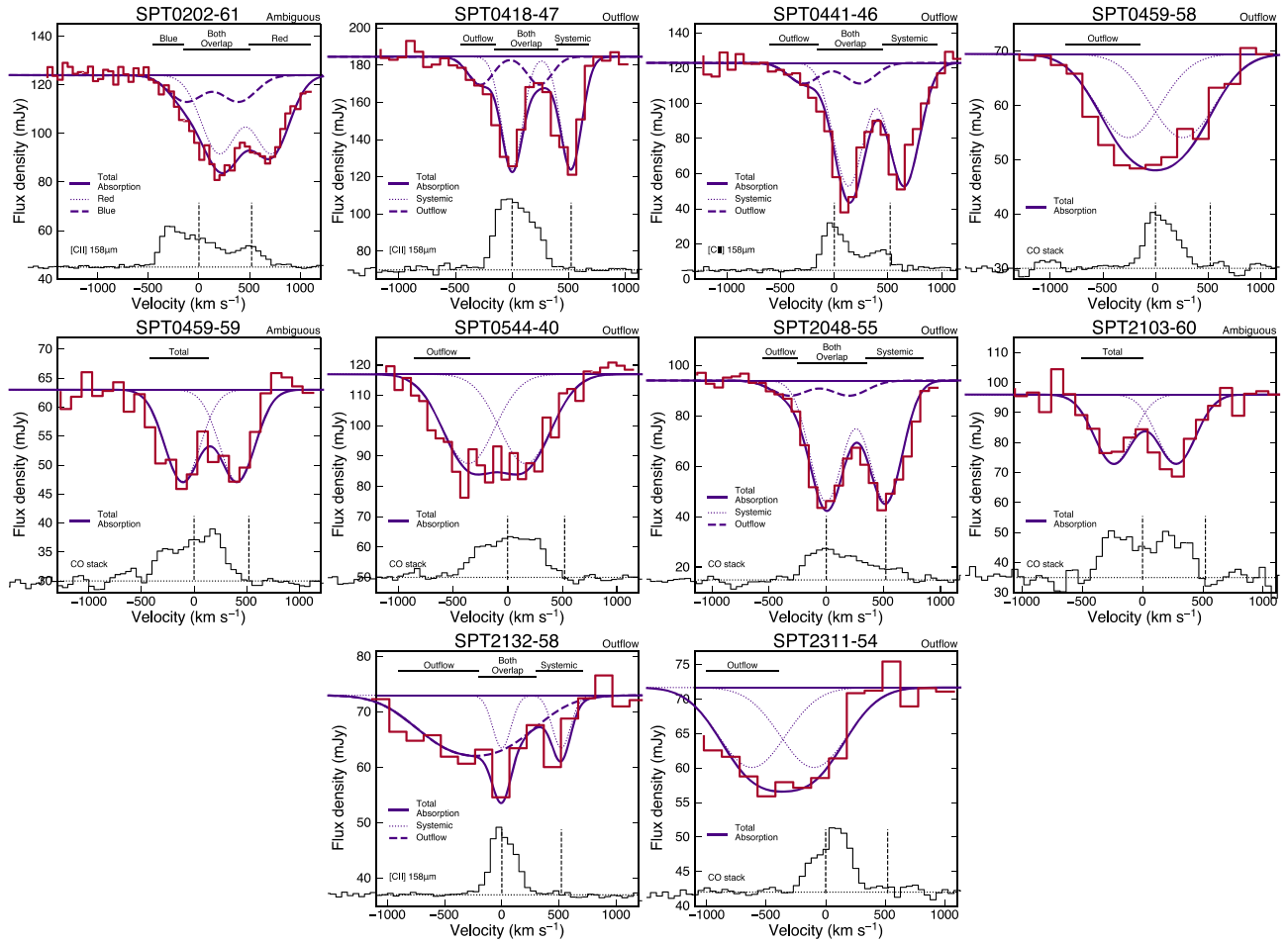


Figure 3. OH 119 μm spectra of each sample target (red), not corrected for lensing magnification. Also overplotted are fits to the spectra using either one or two pairs of Gaussians (navy lines), as detailed in Section 3.1. Vertical dashed lines show the rest velocities of the two OH doublet components, where we assign the higher frequency transition to zero relative velocity to help interpret the complex OH doublet spectra. We also show a “reference” line profile of [O] or CO, which is expected to be dominated by gas internal to the galaxies. Horizontal bars at top label the velocity ranges we selected for lens modeling, chosen to be dominated by single velocity components (we also label the velocity ranges where multiple velocity components overlap, where applicable). There is clearly a large diversity of line profile shapes, but at least 7/10 of these sources host molecular outflows as defined by absorbing components more blueshifted than the reference line emission. The remaining three sources typically show broad reference line profiles with multiple peaks themselves, making interpretation of the OH line profiles difficult.

overlap in source properties. Calderón et al. (2016) and Herrera-Camus et al. (2020) further expand on the aforementioned samples, presenting a combined total of nine ULIRGs with slightly higher redshifts ($z \sim 0.25$) and correspondingly higher typical luminosity ($\log(L_{\text{IR}}/L_{\odot}) \sim 12.5$ – 13.5) to ensure sufficiently bright continuum fluxes. We collectively refer to these samples as “ULIRGs and QSOs” throughout this work.

Stone et al. (2016) present OH observations of a sample of 52 nearby hard X-ray-selected AGN. These sources are the least similar to the other samples or to the $z > 4$ SPT DSFGs, consisting of objects more than an order of magnitude less luminous ($L_{\text{IR}} \sim 0.1 L_{\odot}$) than those in the other samples but whose bolometric luminosities are dominated by the AGN power (typically $f_{\text{AGN}} > 0.8$). Over half of this sample shows OH only in emission, a common feature among AGN-dominated systems, and only 17 show OH in absorption, which Stone et al. use to define their outflow detection rate.

For these low-redshift samples, we adopt the sample properties as published by the original authors. Although for maximal consistency we would remeasure, for example, L_{IR} and f_{AGN} using the same methods as for our own sample, we prefer the literature values because they are publicly and

centrally available. We spot-checked a few sources from the various samples using far-IR photometry from the literature and found results consistent with the published values.

A handful of additional high-redshift objects besides our sample have published OH 119 μm spectra. George et al. (2014) present OH data for SMM J2135-0102 (the “Eyelash” DSFG) at $z = 2.3$. Those authors argue that the OH absorption is associated with one of several spatio-kinematic “clumps” seen in the dust continuum emission in early interferometric data, and is outflowing with respect to that clump. Those clumps have recently been shown to be false (Ivison et al. 2020), so it is unclear how to interpret the OH spectrum in light of the new understanding of the source structure. By our adopted definition (Section 3.2) this source would not be classified as an outflow because the absorption is fully contained within the cores of the bright CO and [CII] emission lines. Additionally, Herrera-Camus et al. (2020) present ALMA spectroscopy of a $z = 6.1$ quasar with a tentative 3σ detection of blueshifted OH absorption. Given the tentative nature of this detection we also consider this case to be inconclusive. Finally, Zhang et al. (2018) detect OH 119 μm in a stack of 45 lensed DSFGs at $1 < z < 3.6$ and $\sim 3\sigma$ detections in two individual

objects, but the spectral resolution was too low to measure any velocity shifts or search for blueshifted line wings.

3. Analysis

3.1. Spectral Analysis

While OH is clearly detected in absorption toward each source, the integrated spectra in Figure 3 show a diverse range of absorption depths and line profiles. In about half the sample the OH doublet lines are sufficiently narrow to be individually resolved, while in the rest they are wide enough to be blended, creating a single wide absorption profile. Some sources show obvious signs of multiple velocity components contributing to the overall line profile (e.g., SPT0418-47), while others are adequately fit by a single velocity component (e.g., SPT0459-58).

To help interpret these line profiles, we fit the integrated spectra with one or two pairs of Gaussians depending on the complexity of the line profile for each source. These fits are by no means unique, but they do capture the information available in the spectra. As a Λ doublet, the two 119 μm OH lines are expected to have equal amplitude, and to have a fixed separation in velocity of $\approx 520 \text{ km s}^{-1}$. The free parameters are the continuum flux density and either one or two of the absorption depth, the velocity offset relative to the systemic, and the Gaussian line width.

We assume a constant continuum flux density across the relatively narrow bandwidth of these observations. In a handful of sources there are few absorption-free channels to constrain the continuum level (e.g., SPT0544-40), or there is blueshifted absorption that extends out to and beyond the edge of the band (e.g., SPT2311-54). For these sources we also make use of the continuum data in the alternate sideband of the ALMA data and/or a global fit to the long-wavelength SED. Rest-frame 119 μm is near the peak of the dust SED where the slope is nearly flat, and we estimate that taking the continuum level from the alternate sideband introduces an uncertainty of at most a few per cent in the continuum level of the OH sideband due to the sideband wavelength separation. We tested this procedure using the sources with sufficient line-free bandwidth in both sidebands and find no appreciable differences in the fit parameters.

We derive several other parameters from the best-fit Gaussian profiles. The OH equivalent width is straightforward to derive from the fitting results. We also report several velocity-related quantities to facilitate comparison with literature samples, including v_{50} and v_{84} , the velocities above which 50% or 84% of the absorption takes place, and v_{max} , the estimated terminal outflow velocity. As in S18, here we take v_{max} to be the velocity above which 98% of the absorption occurs. Various definitions of v_{max} have been used in the literature, and its value depends on the S/N of the data (and, for SPT2311-54, the assumption that the absorption continues to follow a Gaussian profile beyond the edge of the ALMA bandwidth). We note that these fit parameters are largely immune to gravitational lensing, since both the continuum and absorption must be magnified by nearly the same factor. The line profile best-fit and derived parameters are given in Tables 3 and 4.

Finally, we use the fits to the OH line profiles and the ancillary spectral data to define velocity ranges dominated by either outflowing or systemic absorption that we later use in our

lensing reconstruction to measure the structure of the absorbing components. The wide variety of OH line profiles makes it difficult to define these ranges with unambiguous criteria, and we have no way to separate low-velocity outflowing material from high-velocity systemic material. For sources with multiple velocity components, we take the systemic and outflowing components to be those velocity ranges where each component dominates the total absorption profile. For sources with single broad absorption profiles we use the fits to try to avoid double-counting gas at the same source-frame velocity due to the Λ doublet splitting. In both cases, we prefer to define the outflows as beginning at the velocity where we no longer detect CO or [C] emission, but this is not always possible given the relatively weak absorption seen in some sources. In total, we are confident that the velocity ranges we select are at least dominated by outflowing or systemic gas, although in some cases not exclusively so.

3.2. Molecular Outflow Classification

The wide diversity in OH absorption profiles among our sample raises the obvious question of how to determine whether or not a particular source has a molecular outflow. Various definitions to answer this question have been used in the literature. Perhaps the most common method is to classify any source with $v_{50} < -50 \text{ km s}^{-1}$ as an outflow, as done by Rupke et al. (2005) and subsequently adopted by several studies of OH in low-redshift galaxies with Herschel (e.g., Veilleux et al. 2013; Herrera-Camus et al. 2020). For a variety of reasons we find this definition unsatisfying for our sample. First, it is clear that many of our sources have very broad and/or double-peaked [C] or CO lines that are hundreds of km s^{-1} wide, which makes the redshift used to define systemic (zero) velocity somewhat arbitrary. In other words, the relevant metric is not whether the absorption appears to be blueshifted based on the assigned systemic velocity, but whether the absorption is blueshifted relative to the emission line profiles of the gas within the galaxies. Second, this definition ignores the fact that the total absorption profiles are often a superposition of a component at systemic velocities and a second blueshifted component with a typically much weaker absorption depth. There is no reason to expect that the presence or strength of absorption at systemic velocities has any bearing on whether or not an outflowing component is also present. Objects with extremely deep systemic absorption (e.g., SPT0441-46) have v_{50} biased by this very strong systemic absorption since it is measured from the total absorption profile.

Instead, we define a source as containing an outflow if it shows OH absorption more blueshifted than the detected [C] or CO emission, which we expect to be a conservative definition. The benefit of this definition is that outflows defined this way are unambiguous—no fitting technique or analysis method can change the fact that a source shows absorption blueshifted more than any gas in the host galaxy. This definition has the drawback of being dependent on S/N, which may exclude weak outflows (although this is also true of outflow classification based on v_{50}). It may also exclude sources in which the outflow shows strong emission in [C] or CO. However, studies of local objects in multiple outflow tracers typically find that the high-velocity emission outside the line cores is indeed very weak (Lutz et al. 2020), and we see no evidence for high-velocity wings of emission in any of our sources. In such cases additional information (such as

Table 3
Spectral Fitting Results and Lens Model Velocity Ranges

Source	$S_{119\ \mu\text{m}}$ (mJy)	Component	v_{cen} (km s ⁻¹)	S_{abs} (mJy)	FWHM (km s ⁻¹)	Eq. Width (km s ⁻¹)	Model v_{cen} (km s ⁻¹)	Model Δv (km s ⁻¹)
SPT0202-61	124.4 $\pm$ 0.6	Red	+200 $\pm$ 30	-31.7 $\pm$ 2.9	410 $\pm$ 50	120 $\pm$ 9	+800	600
		Blue	-130 $\pm$ 80	-10.9 $\pm$ 2.8	400 $\pm$ 120	34 $\pm$ 9	-300	300
SPT0418-47	184.4 $\pm$ 1.2	Systemic	0	-58.8 $\pm$ 2.7	220 $\pm$ 20	83 $\pm$ 8	+535	270
		Outflow	-280 $\pm$ 30	-14.8 $\pm$ 2.9	260 $\pm$ 90	18 $\pm$ 5	-300	300
SPT0441-46	123.0 $\pm$ 1.5	Systemic	+130 $\pm$ 10	-68.0 $\pm$ 2.8	330 $\pm$ 30	197 $\pm$ 20	+700	500
		Outflow	-280 $\pm$ 80	-11.7 $\pm$ 2.6	350 $\pm$ 160	38 $\pm$ 10	-375	450
SPT0459-58	69.6 $\pm$ 1.1	Outflow	-260 $\pm$ 20	-15.4 $\pm$ 1.3	730 $\pm$ 120	176 $\pm$ 22	-500	700
SPT0459-59	63.0 $\pm$ 0.6	Total	-110 $\pm$ 20	-15.8 $\pm$ 1.3	400 $\pm$ 30	107 $\pm$ 8	-145	550
SPT0544-40	116.9 $\pm$ 1.5	Outflow	-360 $\pm$ 20	-29.4 $\pm$ 2.5	560 $\pm$ 60	165 $\pm$ 27	-600	500
SPT2048-55	94.0 $\pm$ 1.3	Systemic	0	-48.3 $\pm$ 2.2	340 $\pm$ 20	213 $\pm$ 8	+600	500
		Outflow	-320 $\pm$ 50	-6.1 $\pm$ 3.2	380 $\pm$ 160	26 $\pm$ 5	-400	300
SPT2103-60	96.2 $\pm$ 1.2	Total	-240 $\pm$ 20	-23.6 $\pm$ 2.2	380 $\pm$ 30	97 $\pm$ 7	-250	500
SPT2132-58	73.1 $\pm$ 1.9	Systemic	0	-9.5 $\pm$ 3.0	190 $\pm$ 40	19 $\pm$ 7	+500	400
		Outflow	-490 $\pm$ 70	-7.7 $\pm$ 1.6	750 $\pm$ 320	99 $\pm$ 23	-550	700
SPT2311-54	71.7 $\pm$ 1.2	Outflow	-620 $\pm$ 40	-11.6 $\pm$ 1.9	660 $\pm$ 160	118 $\pm$ 12	-700	600
SPT2319-55	52.1 $\pm$ 0.5	Systemic	0	-7.8 $\pm$ 1.2	330 $\pm$ 80	52 $\pm$ 15	+500	400
		Outflow	-440 $\pm$ 50	-7.0 $\pm$ 1.2	450 $\pm$ 60	64 $\pm$ 14	-450	500

Notes. OH spectral component fits are labeled as in Figure 3. Velocities are relative to the higher-frequency OH doublet transition. Systemic profiles centered on 0 km s⁻¹ were fixed to the systemic redshift of those sources. Equivalent widths are given for only one of the OH doublet transitions (i.e., they should be multiplied by 2 for the total equivalent width). The final two columns give the center velocity and width that we use for lens modeling, selected to be dominated by each absorbing component (see Figure 3).

^a Reproduced from Spilker et al. (2018).

Table 4
OH Absorption Profile Characteristics

Source	Outflow?	v_{50} (km s ⁻¹)	v_{84} (km s ⁻¹)	v_{max} (km s ⁻¹)
SPT0202-61	L	+135 $\pm$ 10	-115 $\pm$ 25	-370 $\pm$ 55
SPT0418-47	Y	-35 $\pm$ 10	-230 $\pm$ 50	-430 $\pm$ 80
SPT0441-46	Y	+100 $\pm$ 10	-120 $\pm$ 45	-440 $\pm$ 75
SPT0459-58	Y	-260 $\pm$ 20	-560 $\pm$ 60	-890 $\pm$ 100
SPT0459-59	L	-110 $\pm$ 20	-280 $\pm$ 25	-460 $\pm$ 40
SPT0544-40	Y	-360 $\pm$ 20	-590 $\pm$ 90	-830 $\pm$ 170
SPT2048-55	Y	-25 $\pm$ 10	-210 $\pm$ 30	-480 $\pm$ 85
SPT2103-60	L	-240 $\pm$ 20	-400 $\pm$ 20	-570 $\pm$ 30
SPT2132-58	Y	-360 $\pm$ 60	-750 $\pm$ 120	-1110 $\pm$ 220
SPT2311-54	Y	-620 $\pm$ 35	-900 $\pm$ 65	-1200 $\pm$ 105
SPT2319-55	Y	-215 $\pm$ 40	-525 $\pm$ 40	-760 $\pm$ 60

Notes. See Section 3.2 for our metrics for whether or not a given source shows a definite molecular outflow. The quantities v_{50} , v_{84} , and v_{max} refer to the velocities above which 50%, 84%, and 98% of the total absorption takes place, see Section 3.1.

^a For SPT2311-54 the given v_{max} is an extrapolation of our fit to the spectrum because the absorption profile continues beyond the end of the ALMA bandwidth.

kinematics at high spatial resolution) would be needed to determine whether the bright emission is associated with gas in the host galaxy, a merging partner, or a bona fide outflow. With this classification we consider 8/11 sources in our sample to show unambiguous signs of molecular outflow (detailed further in Section 4.1).

Compared to some literature classifications, our outflow detection rate could be considered a conservative limit. In particular at high redshift we would call the outflows in the

$z=6.1$ quasar J1319+0950 (Herrera-Camacho et al. 2020), the $z=2.3$ DSFG SMM J2135-0102 (George et al. 2014), and the $z=5.7$ DSFG SPT0346-52 (Jones et al. 2019) ambiguous cases instead of confirmed outflows. In the first case the OH spectrum has too low an S/N to be confident in its classification (and was also noted as tentative by those authors), while in the latter two cases the absorption lines used to claim outflow are fully contained within the bright [O] line profile cores. In the case of SPT0346-52, it et al. (2019) suggest that the galaxy is actually a major merger based on modeling of the [C II] data. The absorption is well aligned with one of the [C II] line peaks and is instead most likely simply systemic absorption within one of the merging pair. This particular object highlights that in such cases kinematics at high spatial resolution can clarify whether or not a given absorption profile is indicative of an outflow.

We note that using an alternative outflow definition, $v_{50} < -50$ km s⁻¹, would still result in 7/11 sources being classified as showing outflows, with five sources exhibiting outflows both by this metric and by our preferred definition. The three sources that we identify as showing outflows but with $v_{50} \geq -50$ km s⁻¹, as expected, all have very strong absorption at systemic velocities (or even slightly redshifted from systemic), which results in a biased value of v_{50} . On the other hand, the two sources that would be classified as outflows based on the v_{50} criterion, but which we label as ambiguous, are sources with broad CO emission, where the OH absorption profile is still fully contained within the bright emission from gas inside the galaxies. The OH could simply be absorption internal to the galaxies, and we are not confident enough to label them outflows despite their v_{50} values. This comparison highlights the value of having high-quality reference emission

line spectra and the peril of accepting the results of OH spectral fitting in the absence of additional information.

3.3. Lens Modeling Methods and Tests

We create gravitational lens models to reconstruct the intrinsic structure of each source utilizing the pixellated modeling code described in detail in Hezaveh et al. (2016), as in S18. For each source we fit to the available continuum data, consisting of the data in the line-free sideband of the ALMA data. For SPT2132-58 which has no useful alternate sideband data due to atmospheric opacity we instead simply use the full OH-containing sideband. Once the best-fit parameters of the lensing potential have been determined following the above procedure, we then use these parameters to reconstruct the OH absorption components using the velocity ranges shown in Figure 3 and listed in Table 4. In principle a joint fit to the continuum and absorption components would provide the optimal constraints on the lens model parameters, but this becomes computationally expensive due to the large number of visibilities.

Briefly, the code fits directly to the interferometric visibilities, which we average temporally unless doing so would cause a binned visibility to span more than 10 m in the uv plane. While the code also has the ability to marginalize over residual time-variable antenna-based phase calibration errors, we neglect this capability for computational efficiency. In any case, the phase self-calibration performed as part of the data reduction largely supplants the need for further control over the antenna phases. We fit for the lensing potential using Markov Chain Monte Carlo (MCMC) sampling algorithms. In practice, we first use a code that represents the source plane with one or more simple parametric light profiles (Spilker et al. 2016b) to get a reliable estimate of the lens parameters before refitting with the pixellated code in order to minimize the number of MCMC iterations required for the chains to converge.

The lensing potentials described by one or more singular isothermal ellipsoid (SIE) mass profiles (e.g. Kormann et al. 1994). Each SIE potential is described by two positional coordinates, a strength related to the lensing mass, and two orthogonal components of the lens ellipticity. Where necessary we allow for additional angular structure in the lensing potential with external shear and low-order multipoles in the main lens (up to $m = 4$), as parameterized in Hezaveh et al. (2016). The best-fit lens model parameters are given in the Appendix.

The source plane is represented by a grid of pixels that is regularized by a linear gradient prior on the source, which minimizes pixel-to-pixel variations in the source plane in order to avoid over- or underfitting the data (Warren & Dye 2003; Suyu et al. 2006; Hezaveh et al. 2016). The best-fit strength of the regularization is determined by maximizing the Bayesian evidence given a fixed set of parameters for the lensing potential. Because the regularization strength is only for a fixed set of lens parameters, we perform an iterative process of fitting for the regularization strength, MCMC fitting for the lens parameters, and refitting for the regularization strength until all parameters have converged.

3.3.1. Model Selection and Tests

We generally begin each modeling process assuming a simple lens potential parameterization, adding complexity where necessary. Our models begin under the assumption that the lensing potential can be described adequately by a single SIE profile and an external shear component. If this model does not satisfactorily reproduce the data, we introduce additional complexity as needed. For SPT0459-58 and SPT2103-60 the morphology of the lensed images is clearly inconsistent with a simple SIE mass profile, in agreement with the near-IR imaging that shows multiple plausible lensing galaxies near the main lens. For these sources, we fit models using two (SPT0459-58) or three (SPT2103-60) lensing galaxies. SPT0441-46 also shows a second object $\sim 1''$ west of the main lens that may influence the lensing potential, but our models do not require the inclusion of a second lens to reproduce the data.

We use several metrics to determine whether these simple models are sufficient to capture the information in the data or whether further complexity is warranted. We first compare the deviance information criterion of different models (Spiegelhalter et al. 2002), preferring the models with greater likelihood if the additional free parameters from additional lens potential complexity legitimately provide a better fit to the data. Second, in reality we know that the reconstructions of dust continuum emission should be positive, but nothing in our methodology forces positivity. If a given model yields a source-plane reconstruction with large negative “bowls,” we take this as an indication that the parameterization of the overall lensing potential probably requires additional complexity to exclude an unphysical source reconstruction. In practice we flagged models where the peak negative pixels in the source reconstruction had an absolute value $\geq 10\%$ of the peak positive pixels for additional scrutiny.

We perform extensive tests of the effective sensitivity and resolution of the source reconstruction. It is not straightforward to infer an effective source-plane resolution or sensitivity from the observed (image plane) data. For example, the effective resolution and sensitivity vary with location in the source plane based on the local lensing magnification, and the source regularization strength depends on both the resolution and sensitivity of the original data. This becomes even more complicated when considering absorption components because the detection of absorption requires the presence of continuum emission, but the continuum brightness, absorption depth, and effective resolution and sensitivity all vary across the source plane. Following S18, we perform a series of reconstructions of mock data to test the resolution and sensitivity of the source reconstructions.

Briefly, we create many mock observations of pointlike background sources tiled across the source plane, analyzing these fake data in an identical way to the real data. The intrinsic flux density of the artificial sources is set such that the total apparent (magnified) flux density matches that of the real sources (since these sources were selected in part based on their apparent brightness). We then fit the source reconstructions of each set of fake data with a two-dimensional Gaussian, taking the FWHM as an empirical estimate of the resolution of the reconstruction. We also measure the differences between the input and best-fit positions of the artificial sources, which are small in all cases except when the input source lies very near the lensing caustics with magnifications ≥ 50 . We associate

these failed solutions with too-poor resolution in the source- and image-plane pixel grids, because the pixel sizes of these grids were not optimized for such extreme cases of high magnification and compact source sizes. Finally, we repeat the entire procedure but change the input flux density of the source to be a factor of 2–3 weaker than the faintest component in the lens emission. This final test shows that even these weak signals are recoverable to $\approx 25\%$ accuracy.

In summary, these tests lead us to conclude that the structures seen in the source reconstructions are real and that absorption signals a few times weaker than those actually observed can be successfully recovered. We find no evidence that the lensing reconstruction procedure introduces artificial clumpy structure. The sources in our sample are resolved over ≈ 5 –20 independent resolution elements, in agreement with the image-plane data (Figure 2).

For the purposes of Figure 6 we illustrate the effective source-plane resolution with an ellipse based on the fits to the artificial data at the position corresponding to the peak of the actual reconstructed source continuum emission. All maps of the effective resolution are provided in the Appendix.

3.4. Notes on Individual Sources

Our sample shows very diverse characteristics in terms of both lensing geometry and OH 119 μm line profiles, in some cases requiring special treatment. Here we give a brief summary of these particularities and comment on the conclusions we draw from the OH spectra.

3.4.1. SPT0202-61

The OH spectrum of this source clearly requires two velocity components to reproduce the data. The [C II] profile of this source is also very broad and shows two peaks, likely indicating a major merger. The deepest OH absorption is redshifted compared to one of the peaks and blueshifted relative to the other. It is thus possible that we are seeing a molecular outflow (launched from the fainter [C] component) or a molecular inflow (falling toward the brighter [C] peak), or simply strong systemic absorption from the interaction/overlap region. The [C] data will be analyzed in future work and a detailed comparison between the extended [C] and the OH absorption (confined to the continuum emitting region by definition) is complicated, but initial modeling does not conclusively point to an outflow. We thus consider this source ambiguous and do not claim a molecular outflow.

Additionally SPT0202-61, nearly uniquely among the SPT DSFG sample, shows submillimeter emission at the center of the Einstein ring of background source emission. This is most clearly visible in Figure 2. The available data make clear that the central emission is not a (demagnified) lensed image of the background source. There is an additional unlensed continuum source located $\approx 1/5$ southwest of the lensed source, also noted by Spilker et al. (2016b). These sources will be explored in more detail in future work. The pixelated reconstruction tool we employ does not have the ability to model lensed and unlensed emission simultaneously, so we model and subtract the lens and secondary source before lens modeling. The lens emission does not cleanly separate from the lensed background emission, so the subtraction is imperfect. To mitigate this, we

carefully define the image-plane and source-plane pixel grids in the lensing code such that no source-plane pixels map to the center of the Einstein ring. We have verified that this choice has no impact on our reconstructions apart from leaving residuals in the model-subtracted data from the imperfectly subtracted lens emission.

3.4.2. SPT0418-47

This source shows a clear but not especially deep outflow component extending well beyond the high-S/N [C] emission. We note that this source was observed in ALMA projects 2015.1.00942.S and 2018.1.00191.TS. The 2015.1.00942.S data were taken with the array in a much more extended configuration than we originally requested, yielding a synthesized beam $\sim 0''.15$. These data proved too shallow given their high spatial resolution, and we excluded them from all of our analysis.

3.4.3. SPT0441-46

The OH absorption at systemic velocities is deeper in this source than in any other in our sample, but the spectrum is better fit with an additional blueshifted velocity component. While the [C II] spectrum of this source is also broad and double-peaked like SPT0202-61, unlike that source the OH absorption continues beyond the bluest [C] emission. We can thus unambiguously confirm that this source hosts a molecular outflow. The deepest OH absorption is slightly redshifted compared to the [C I] peak, possibly indicating a molecular inflow toward the [C II] peak.

The lens model of this source required low-order multipoles to adequately reproduce the data. The best near-infrared image of the lens galaxy, from Hubble Space Telescope/WFC3 imaging, shows a second source $\sim 1''$ west of the main lens galaxy. We do not know whether this second source is associated with the main lens galaxy, but it may be the cause of the additional complexity in the lens model required to fit the data.

3.4.4. SPT0459-58

The OH spectrum of this source shows very deep blueshifted absorption, and does not require multiple velocity components to fit the spectrum (mostly because the absorption is so broad that any additional velocity components are indistinct).

The continuum image of this source shows a morphology that is clearly inconsistent with a single simple lensing potential, which went unrecognized in earlier analysis due to the much lower sensitivity and resolution of the earlier data (Spilker et al. 2016b). The northernmost lensed image in particular requires that a second lens potential be placed in its vicinity in order to reproduce the image splitting at that location. Unfortunately the quality of the best available near-IR image of this source from Very Large Telescope/ISAAC is too poor to confirm or refute optical counterparts to the best-fit lens potentials we find.

3.4.5. SPT0459-59

The OH spectrum of this source shows no obvious evidence of additional blueshifted absorption beyond the broad CO emission, and we do not classify it as an outflow source. In addition to the lensed galaxy, this source shows at least two other weakly lensed sources, one just south of the lensed

emission and the other $\sim 1.5''$ southwest of the lensed emission. We do not detect OH absorption from either of these sources but they are faint in the continuum so it is unclear whether they are physically associated with the lensed source. Observations of an emission line at high spatial resolution could clarify the structure of the galaxy and allow for a better characterization of the OH absorption. In our modeling of this source, we subtracted the southwestern source prior to lens modeling in order to allow for more computationally feasible image- and source-plane pixel grids. This also enables better source-plane regularization, since the fitting for the regularization strength need not be influenced by both the strongly lensed main source and the very weakly lensed southwestern source.

3.4.6. SPT0544-40

The OH spectrum of this source shows a straightforward, albeit broad, blueshifted absorption profile. While this galaxy appears to be a standard quadruply imaged background source at first glance, both the flux ratios and the spatial extent of the images make clear that it must instead contain two continuum components. It is possible that these two continuum components both contribute to cause the overall broad absorption line profile.

3.4.7. SPT2048-55

Much like SPT0441-46, this source shows very strong systemic absorption and a weak blueshifted absorption component, and the stacked CO lines used as a reference do not have particularly high S/N. The outflow in this source is the weakest of those we consider unambiguous.

3.4.8. SPT2103-60

Similarly to SPT0459-59, the OH absorption in this source does not show an unambiguous outflowing component. While the absorption troughs are blueshifted compared to the flux-weighted mean redshift of the CO emission, it is still possible that the absorption is simply at systemic velocities with respect to the bluest part of the CO emission.

This source is also known to be lensed by a small group of galaxies, three of which are required in order to reproduce the data (Spilker et al. 2016b). We note that we do not require the positions of these lens galaxies to align with galaxies detected in the near-IR in either absolute or relative astrometry, since the baryonic and dark-matter masses can become spatially decoupled in overdense environments. In particular the best-fit mass and position of the southwestern lens are degenerate.

3.4.9. SPT2132-58

This source shows very broad blueshifted absorption as well as narrow systemic absorption, only barely reaching back to the continuum level at the blue end of the ALMA bandwidth. This source was also studied in detail by Béthermin et al. (2016), who found a high excitation in the CO(12–11) transition that could be due to the presence of an AGN.

3.4.10. SPT2311-54

This source shows very broad and extremely blueshifted absorption. This galaxy hosts the fastest outflow of our sample, although the maximum outflow velocity is obviously uncertain

since the limited ALMA bandwidth does not extend far enough to capture the full absorption profile.

4. Results

4.1. Molecular Outflow Detection Rate

We detect 119 μm OH absorption in 100% of our sample, and in 8/11 cases associate this absorption with an unambiguous outflow (including SPT2319-55 published by S18). The remaining three sources all have broad and/or double-peaked CO or [C II] line profiles that make it difficult to interpret the OH absorption. Assuming binomial statistics, the outflow detection rate is $73\% \pm 13\%$; 3% of the sample shows evidence of OH in emission at systemic velocities; our ability to detect redshifted emission as in the classic P Cygni profile is limited due to the small ALMA bandwidth.

The overall high outflow detection rate demonstrates that molecular winds in these highly star-forming objects are very common. Additionally, because OH as an outflow tracer manifests in absorption and because the absorption is highly optically thick (e.g., González-Alfonso et al. 2017), our high detection rate also implies that the opening angle of outflowing material must also be high (otherwise the outflows would not be detectable in absorption because most lines of sight would not intersect outflowing gas). Our detection rate is therefore a lower limit on the true occurrence rate of molecular winds in $z > 4$ DSFGs. Because there are presumably some sources driving outflows that do not intersect the line of sight toward the galaxy continuum.

Even without performing any lensing reconstructions or other spatial analysis, we can infer some details about the spatial structure of these outflows. In one possible scenario, large galaxy-scale outflows are being driven with a high opening angle so that most sources have a wind detectable in absorption due to the high covering fraction of outflowing material. Alternatively, it may be that the outflows are launched such that they are viewable along most lines of sight even with a small covering fraction. For example, if spherical outflows are preferentially launched from the nuclear regions of the host galaxies, we would nearly always detect an outflow even if the molecules are halted or destroyed before reaching kiloparsec-scale distances from the nucleus, resulting in low covering fractions but a high detection rate. These two scenarios are distinguishable using our lensing reconstructions of the wind-absorbing material.

Figure 4 compares the outflow detection rate of our $z > 4$ DSFG sample with the low-redshift comparison samples (Section 2.4). Our detection rate is very similar to that of $z < 0.3$ IR-luminous galaxies but much higher than that of lower-luminosity AGN. We find a slightly higher detection rate when considering only the low-redshift ULIRGs with the lowest AGN fraction, $f_{\text{AGN}} < 0.3$. Although this difference is not statistically significant, Figure 4 also shows the outflow detection rate as a function of L , combining all available low-redshift OH samples. We do not find a decrease in the detection rate among the most luminous galaxies, a marginally significant difference compared to the low-redshift samples. Although the high outflow detection rate in our very luminous high-redshift sample was not unexpected, our observations place the first statistical constraints on the outflow occurrence rate in the early universe. Following Veilleux et al. (2013) in assuming that all galaxies in our sample have a biconically

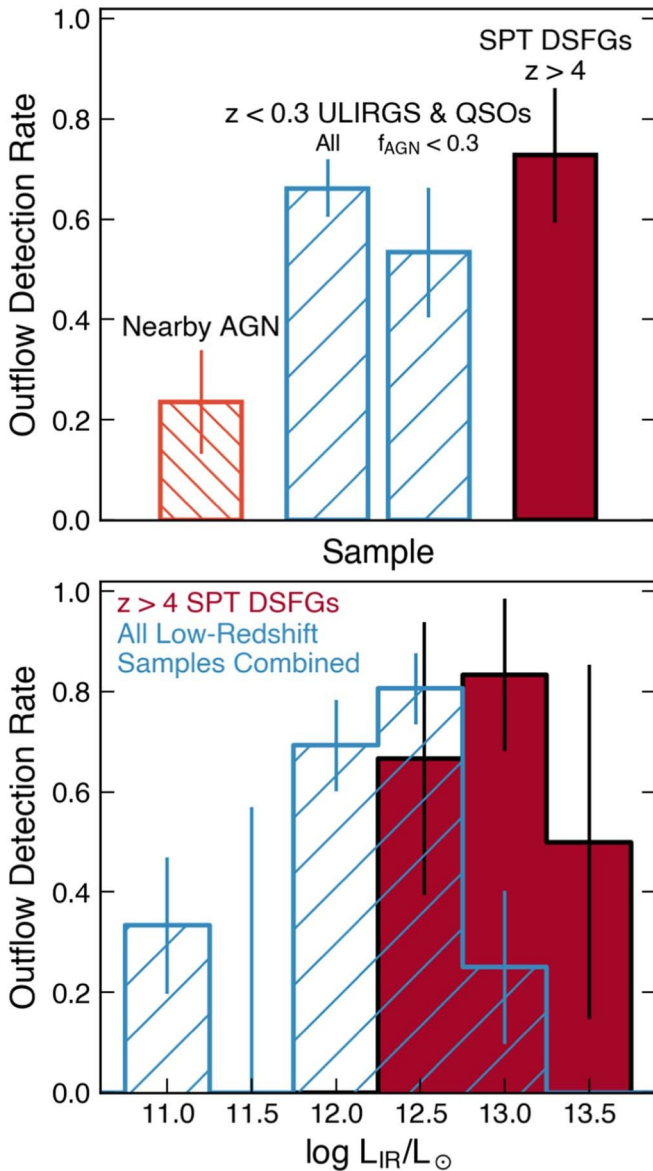


Figure 4. Top: comparison of outflow detection rates via blueshifted OH absorption between our high-redshift DSFGs and a compilation of low-redshift samples (Section 2.4). The high detection fraction we find is very similar to that in low-redshift ULIRGs. Uncertainties assume binomial statistics. Bottom: outflow detection rates as a function of $\log L_{\text{IR}}/L_{\odot}$. Although the number of sources is small, we do not find a decrease in the detection rate in the most luminous galaxies.

expanding outflow, our detection rate would correspond to an opening angle of $\sim 150^\circ$, again very similar to the 145° inferred by Veilleux et al. (2013) or the 125° we infer for the low- f_{AGN} subsample of all local objects.

Finally, we note that no source shows unambiguous evidence for molecular inflows, which would manifest as redshifted absorption profiles. Two sources show $v_0 \approx \pm 50 \text{ km s}^{-1}$, sometimes used to classify sources as showing evidence for inflows; our criticisms of this metric when applied to outflows also hold here (Section 3.2). In our sample, both sources have clearly double-peaked profiles, often a sign of mergers. It is unclear whether the slightly redshifted absorption we see is due to inflow toward one of the velocity components, outflow from the other, or simply systemic absorption from the putative interaction region between them.

The inflow detection rate in low-redshift ULIRGs is $\sim 10\%$ (Veilleux et al. 2013; Herrera-Camus et al. 2020) using the v metric, far lower than the outflow detection rate. Given the ambiguity in our data and small sample size we cannot draw strong conclusions on this point but there is no immediately obvious difference between the inflow detection rates at $z \sim 0$ and $z > 4$.

4.2. Basic OH Absorption Properties

We detect OH in absorption in all 10 target DSFGs (11 including the source in S18); no source shows OH in emission or P Cygni profiles. While observational restrictions preclude us from detecting P Cygni profiles, the lack of OH in emission in any source is interesting because OH $119 \mu\text{m}$ appears in emission primarily in AGN-dominated galaxies ($f_{\text{AGN}} \approx 0.8$; e.g., Veilleux et al. 2013; Stone et al. 2016; Runco et al. 2020). For example, in the sample of Stone et al. (2016) of nearby AGN-dominated galaxies, $>70\%$ of the detected objects showed either pure emission or emission/absorption composite spectra, with 60% purely in emission. A similar conclusion applies to the more IR-luminous ULIRGs and QSOs with $f_{\text{AGN}} \approx 0.8$ typically show OH in emission (Veilleux et al. 2013). This is presumably because the dense nuclear regions are able to excite the $119 \mu\text{m}$ energy levels in spite of the very high gas densities required for collisional excitation, $\sim 10^8 \text{ cm}^{-3}$ (Spinoglio et al. 2005; Runco et al. 2020). This may also be an evolutionary effect, where wide-angle outflows have cleared the sightline to the dense nuclear region and already subsided (e.g., Veilleux et al. 2013; Stone et al. 2016; Falstad et al. 2019). The fact that we do not see OH in emission in any of our sources is a secondary empirical indication that AGN are probably not dominant in these galaxies, unless the column densities are so high as to be optically thick in the mid-IR in the direction of the emitting regions.

Figure 5 shows histograms of three different metrics of the OH absorption velocity v_{50} , v_{84} , and v_{max} , for our sample and literature sources. We exclude those sources for which OH was detected only in emission. We find very similar distributions of these quantities between the low- and high-redshift samples. A two-sided Kolmogorov–Smirnov (K-S) test confirms that we cannot reject the hypothesis that the two samples are drawn from the same underlying distribution in any of the three velocity metrics. While it appears that the low-redshift sample has a tail to extremely fast maximum outflow velocities $v_{\text{max}} \approx 1400 \text{ km s}^{-1}$ that is not present in our data, the limited ALMA bandwidth means we could not probe such high velocities if they were present in our sample. Indeed, two of our sources show absorption that continues essentially all the way to the edge of the observed bandwidth and could plausibly reach higher outflow velocities than the extrapolations from our fits to the spectra suggest depending on the true line profiles.

Figure 5 also shows the equivalent width distributions of these samples and the relationships between the equivalent widths and the velocity metrics. The equivalent widths are the total values (systemic plus any red- or blueshifted absorption) for one of the two doublet lines (i.e., should be multiplied by 2 for the total equivalent widths). The equivalent width distribution does show a significant difference between the low- and high-redshift samples; a two-sided K-S test rejects the hypothesis that the two samples are drawn from the same underlying distribution ($p = 0.004$). This appears to be because the SPT DSFGs show stronger OH absorption than

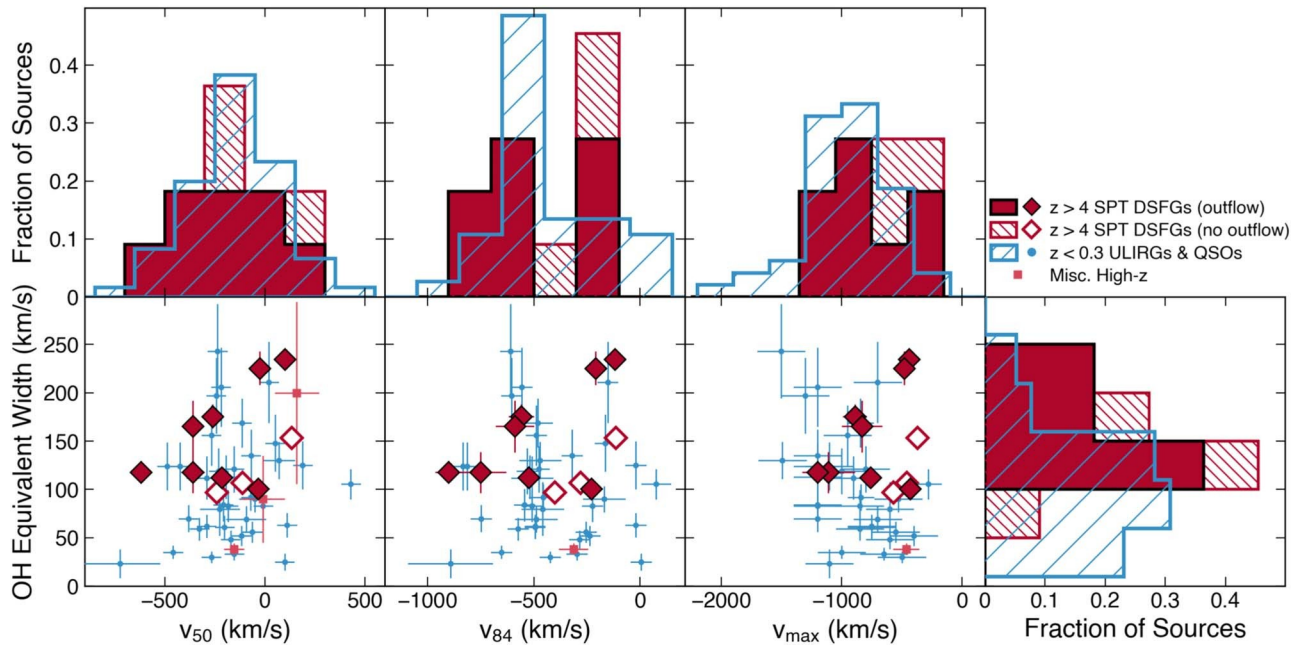


Figure 5. Distributions of the OH absorption median, 84th percentile, and maximum velocities and the equivalent widths of the OH doublet lines for our sample and literature sources. Histogram bins have been slightly offset between the samples for clarity. The limited ALMA bandwidth would prevent us from detecting the fastest outflows seen in the low-redshift samples. Not all literature studies report all quantities in this figure, so there are some discrepancies between the histograms and scatter plots. In particular Spoon et al. (2013) (a subset of the “ $z \leq 0.3$ ULIRGs and QSOs”) do not provide equivalent widths; this sample is noticeably absent from the plot of equivalent width vs $v_{\max}$ compared to the $v_{\max}$ histogram.

the low-redshift samples being underrepresented at low equivalent widths and overrepresented at high equivalent widths.

This difference cannot be explained by simple observational selection effects. While it is plausible that weak absorption would be more easily detected in the bright nearby samples than in our very distant targets, OH was strongly detected in every SPT source. We could have detected equivalent widths $\sim 5\times$ lower than even the weakest absorption actually seen; the dearth of weak OH absorption in the high-redshift sample is genuine. The differences in typical L_{IR} and f_{AGN} between the two samples also do not explain the different distributions. The same two-sided K-S test returns $p = 0.02$ when considering only the low-redshift sources with higher L_{IR} than our least-luminous source ($\log(L_{\text{IR}}/L_{\odot}) \geq 12.5$) and $p = 0.04$ for only the low-redshift sources with $f_{\text{AGN}} \leq 0.4$ (the number of low-redshift objects that are both very luminous and have low f is too small for a meaningful comparison). We conclude that OH absorption at high redshift does appear to be genuinely stronger than in similar low-redshift galaxies, but a more complete sampling of objects at lower L_{IR} and/or higher f_{AGN} at high redshift is required to make a more thorough comparison.

4.3. Continuum and Absorption Reconstructions

Figure 6 shows the lensing reconstructions of the continuum and the OH absorption component(s) for each source, presenting the absorption maps in terms of both integrated absorbed flux and equivalent width. While absorption is a multiplicative and not an additive process, our ability to detect absorption is not. Consider for example a region with continuum $S/N \approx 10$. We would likely not confidently detect absorption even in the case of 100% absorption, while far weaker absorption is detectable where the continuum is brightest. We mask the maps of absorbed flux in regions

where the continuum is detected at $S/N \leq 5$. Even above this threshold in most sources we find some pixels in the absorption maps with positive flux (or negative equivalent width), which is simply a reflection of differencing two moderately uncertain measurements. These positive regions are mostly eliminated if we change the threshold to continuum $S/N > 10$, which we consider to be an empirical indication of the level where the absorption reconstructions are most reliable.

The continuum emission in our sources is mostly smooth at the resolution of these observations, with effective circularized radii $r_{\text{cont}} \approx 0.5\text{--}4$ kpc (the largest source, SPT0459-59, clearly consists of multiple sources poorly resolved by the present data). These sizes are similar to or somewhat smaller than those found by Hodge et al. (2016) in a sample of unlensed DSFGs at lower redshift but typical for the lensed SPT DSFGs (Spilker et al. 2016b), although the methodology for estimating size differs between these works. A few sources show multiple peaks in the continuum reminiscent of mergers, but we see little evidence for distinct sub-kiloparsec substructure. Recent ALMA observations of lensed and unlensed galaxies have demonstrated that identifying clumps in dust emission can be precarious even in data with $S/N \approx 30$ because the relevant metric is the contrast between the clumps and the underlying smooth emission (e.g. Hodge et al. 2016, 2019; Rujopakarn et al. 2019; Ivson et al. 2020). Our observations typically reach peak S/N values far higher than these studies, but the substantial freedom afforded by the pixelated lensing reconstructions reduces the effective peak S/N in the reconstructions to $\sim 25\text{--}40$ for all sources. That is, there is no direct correspondence between image- and source-plane S/N because our lensing reconstructions fully account for the uncertainties in the data rather than directly map from image to source. We rule out significant 500 pc scale clumpiness in our sample at $\leq 10\%$ of the total flux density as seen in some ALMA

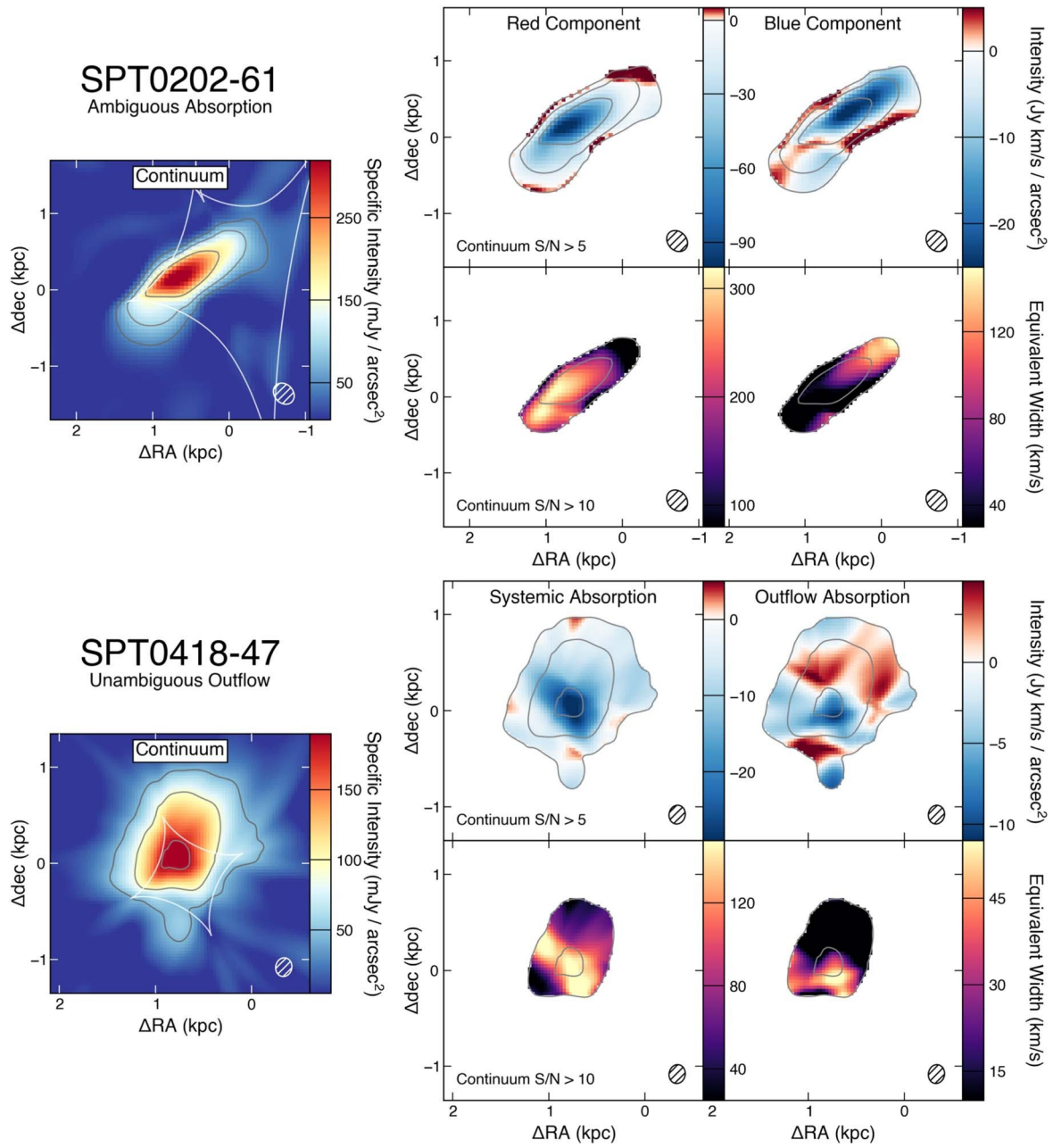


Figure 6. Lensing reconstructions of the OH data for each target DSFG. Axes are relative to the ALMA phase center. Left shows the dust continuum emission with gray contours overlaid at S/N of 5, 10, and 20; these contours are repeated in the other panels. The lensing caustics are shown in white. Right panels show the reconstructions of the OH absorption component(s) as indicated in Figure 3 and listed in Table 3. Upper rows show maps of the integrated absorbed flux density with color scale in Jy km/s / arcsec². Lower rows show maps of the equivalent width. We mask the upper rows for continuum S/N < 5 and the lower rows for continuum S/N < 10. At lower right of each panel we show an ellipse representing the effective spatial resolution of the reconstructions at the peak of the reconstructed continuum emission. The spatial resolution varies across the source plane; see Section 3.3.1 and the Appendix.

studies (Hodge et al. 2019) but cannot rule out weaker and/or smaller clumps.

In sources with systemic absorption, the absorption generally peaks near the same location as the continuum, with an equivalent width distribution that can be either centrally concentrated or more widespread and uniform. The systemic absorption tends not to show signs of distinct clumps, like the continuum emission. This is altogether unsurprising, given that OH is sensitive to even fairly small column densities of molecular gas, which is satisfied over large regions of these gas-rich galaxies but subject to line-of-sight geometric effects (e.g., González-Alfonso et al. 2017).

The outflow maps show more interesting structure when visualized in terms of either absorbed flux or equivalent width. In contrast to the systemic absorption, the blueshifted outflowing gas rarely peaks in equivalent width at the locations of the continuum peaks, instead often showing the largest equivalent widths significantly offset from the brightest continuum emission. Also in contrast to the continuum and systemic absorption, distinct clumps or small-scale substructures are common in the molecular outflows. Combined, these two facts may indicate that the high-velocity gas has already moved away from the nuclear regions and now appears offset (if indeed the outflows were launched from the nuclear

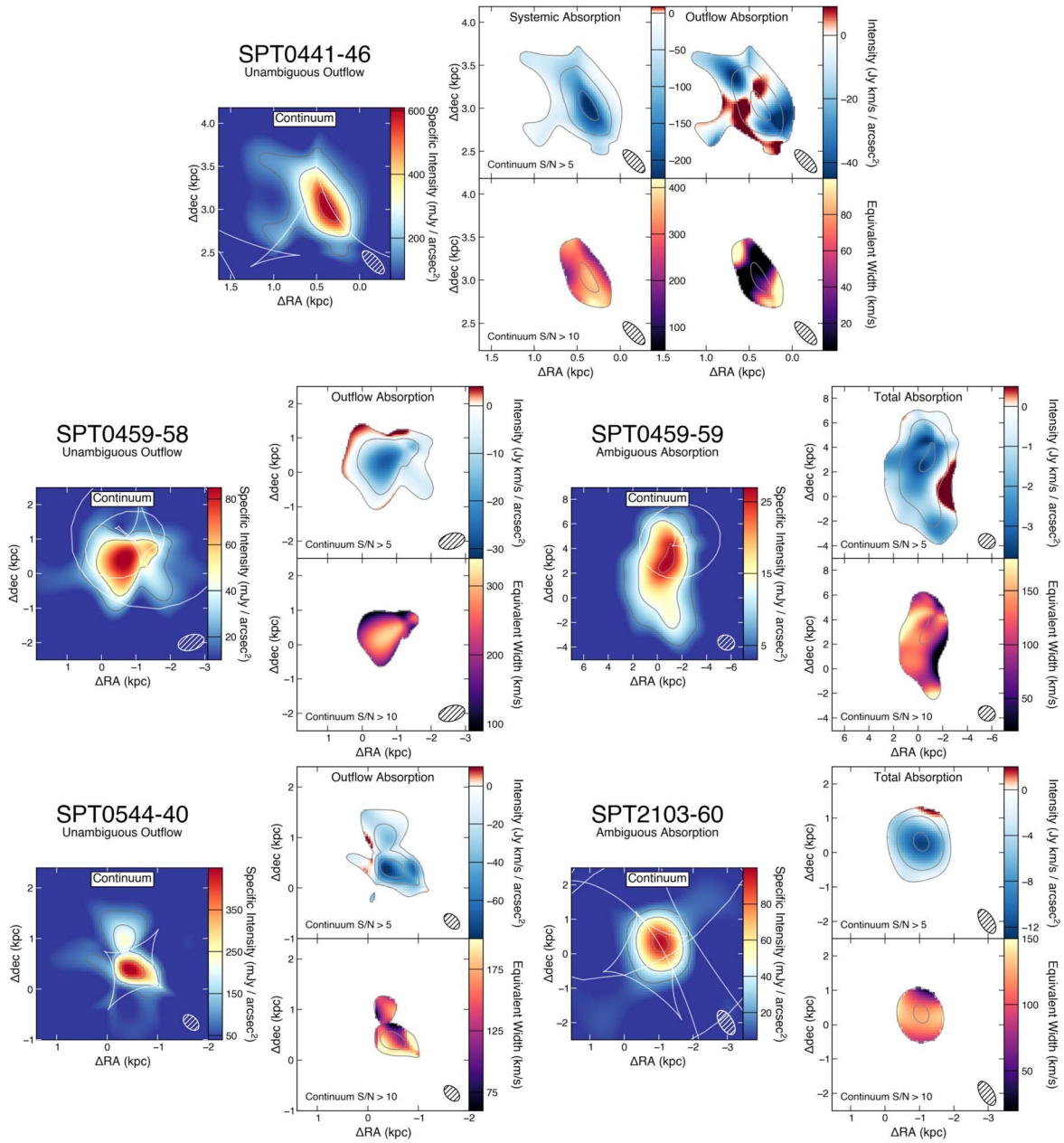


Figure 6. (Continued.)

regions), similar to the offsets commonly seen between the CO and high-velocity CO emission in low-redshift outflows (e.g., Lutz et al. 2020). Alternatively, it could be that high-velocity outflows are more easily launched in regions with lower column densities than the very dense nuclear regions (e.g., Thompson & Krumholz 2016; Hayward & Hopkins 2017).

Figure 7 shows the absorption covering fraction for the reconstructions in each source, defined as the fraction of pixels in the reconstruction of equivalent width with detectable absorption where the continuum $\text{S/N} \geq 10$. We find generally high covering fractions ranging from 30% to 85% for the outflows or 40% to 95% for the systemic absorption. These values are typically somewhat lower than the overall sample detection rate of 73% for either component but significantly higher than the fractional absorption depths in the apparent OH spectra. This may be somewhat surprising given that OH

$19 \mu\text{m}$ absorption is expected to be very optically thick. We return to this comparison in Section 5.1.

Figure 8 searches for trends of absorption covering fraction with selected other quantities we have measured. First, we find no correlation between the covering fraction of either systemic or outflow components with the actual physical size of the dust continuum emitting region (Figure 8 left). This lack of correlation rules out that the absorbing material has a typical constant physical size. If this were so, we would expect the covering fraction and continuum emitting size to be inversely correlated, with the typically sized absorbing regions covering a larger fraction of intrinsically smaller galaxies.

Second, we do find intriguing evidence for a correlation between the outflow covering fraction and L_{FIR} , with the most luminous sources showing higher covering fractions (Figure 8 right). Taken at face value, this result implies that more

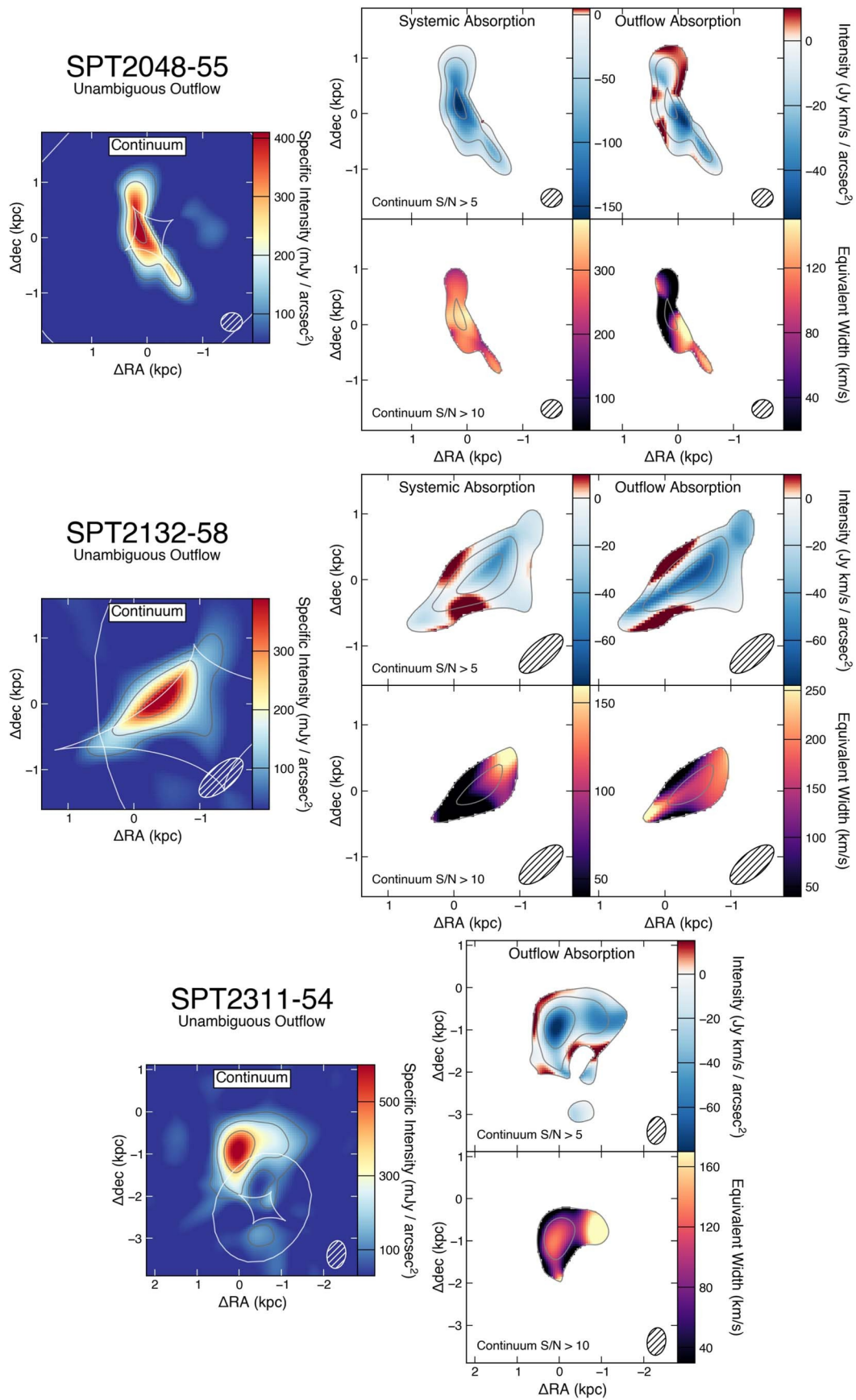


Figure 6. (Continued.)

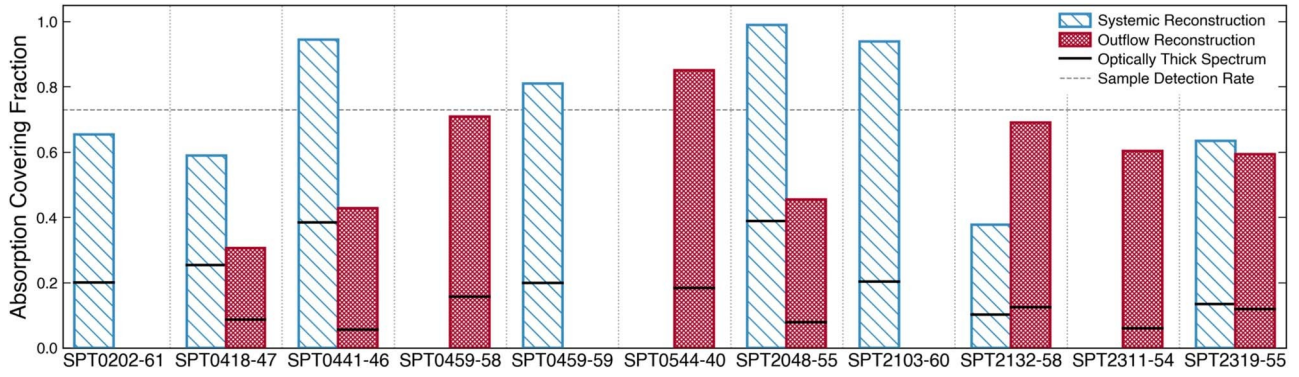


Figure 7. Covering fractions of OH absorption from the reconstructions of each source defined as the fraction of pixels with detectable absorption where the continuum $S/N > 10$. Horizontal black lines show the expected covering fraction derived from the absorption depth over the velocity range of each modeled component assuming optically thick absorption; these represent lower bounds on the covering fractions. The dashed gray line shows the sample-averaged detection rate of 73% (which is the same for the outflows and systemic absorption). The differences between these values may imply the presence of substructure absorption on scales several times smaller than the resolution our data provides; see Section 5.1.

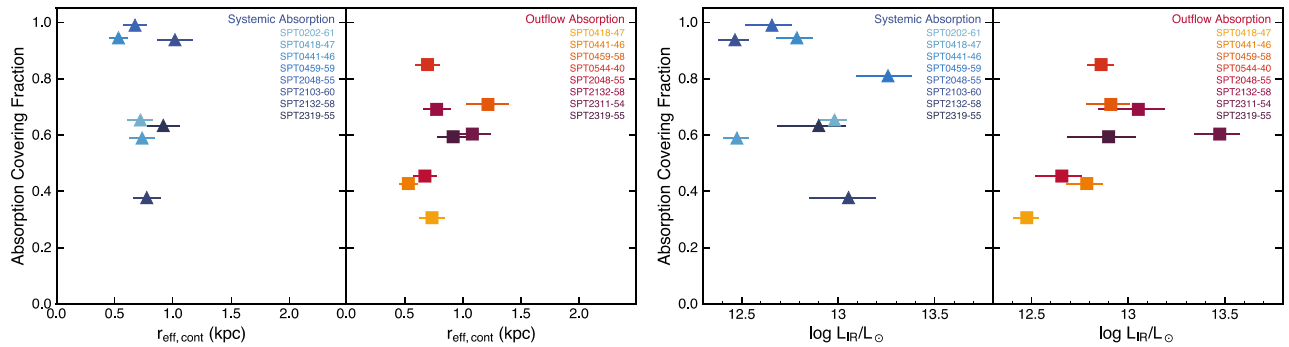


Figure 8. The most L_{IR} -luminous sources in our sample have higher outflow covering fractions, implying that these galaxies are able to launch more widespread outflows. The figure shows the OH absorption covering fractions from the lensing reconstructions against the effective dust continuum emitting sizes (left) and L_{IR} (right). The lack of an inverse correlation with the continuum size rules out that the absorbing structures have a typical constant physical size, but there is an intriguing correlation between covering fraction and L_{IR} .

luminous galaxies are able to launch more widespread outflows, with either larger opening angle or more and/or larger clumps. At some level this result is also consistent with the trend of outflow detection rate rising with L_{IR} , which extends over a larger dynamic range in L_{IR} than probed by our sample alone (Figure 4). In order to see such a trend, one can invoke either a rising outflow occurrence rate or an increasing outflow covering fraction with L_{IR} , or both. The same trend does not exist for the systemic absorption components, implying (unsurprisingly) that the covering fraction of gas at rest is dependent on the line-of-sight geometry but not necessarily any intrinsic physical properties of the galaxies.

For the sources that show unambiguous outflows only in Figure 9 we additionally find correlations between the outflow covering fraction and the outflow velocities v_0 and v_{max} . In addition to being driven by the most luminous sources, the galaxies with high covering fractions also show the fastest outflows. Some of this is no doubt driven by underlying trends we find between L_{IR} and v_0 or v_{max} for the sources in our sample with outflows, as we explore in more detail in Paper II. It is intriguing nonetheless that the outflow covering fraction appears to be correlated with either or both quantities. Whether these trends hold with increased sample sizes that span a wider dynamic range in host properties is an interesting motivation for future investigation.

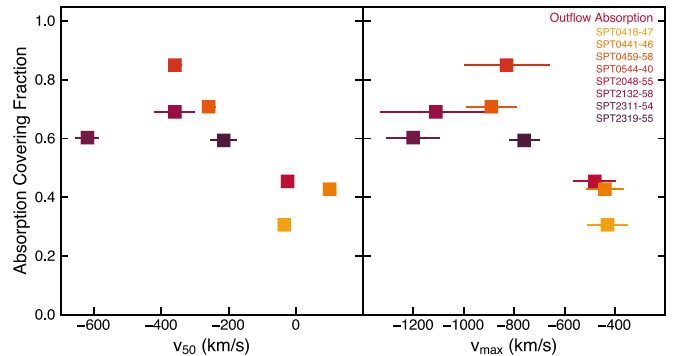


Figure 9. The galaxies with the fastest outflows also show the highest covering fractions or most widespread outflows. We show the lens model covering fractions against v_0 and v_{max} , plotting only those sources with unambiguous outflows.

4.4. Ubiquitous Clumpy Molecular Outflows

In an effort to further quantify the degree of small-scale “clumpiness” in the reconstructions, we borrow several metrics from the extensive literature on morphological analyses of clumps and mergers in high-redshift galaxies (e.g. Lotz et al. 2004; Law et al. 2007; Förster Schreiber et al. 2011; Wuyts et al. 2012; Guo et al. 2018). Specifically, for the continuum and absorption component(s) of each source, we calculate:

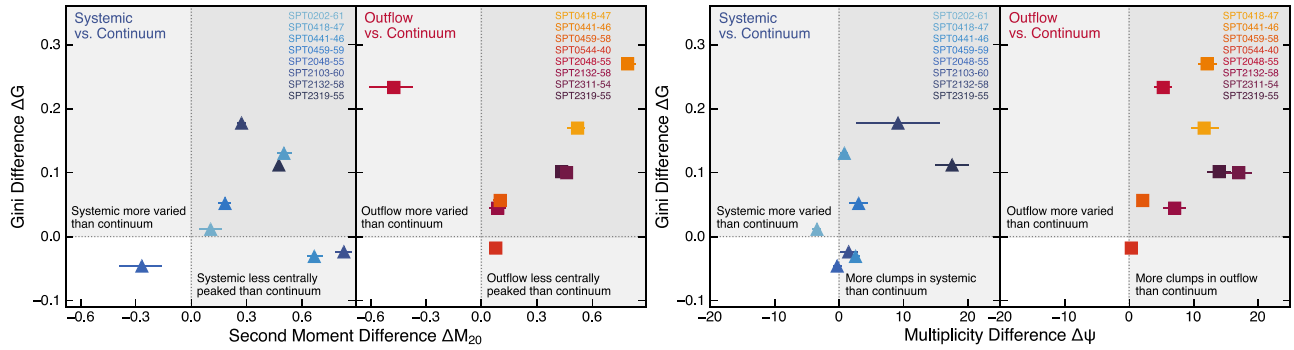


Figure 10. The molecular outflows in our sample are clumpier than the continuum emission, and generally clumpier than the systemic absorption as well. We show the difference (absorption minus continuum) of the Gini coefficient G , second-moment parameter M_{20} , and multiplicity Ψ for all sources. Qualitative descriptions of these differences are stated on the plots; sources in the upper right quadrant can be considered to have clumpier structure in the absorption than the continuum by multiple metrics. The molecular outflows are significantly clumpier in general than the continuum, with 6/8 sources considered so by all three metrics. The systemic absorption, on the other hand, is generally less clumpy, with only 3/8 sources showing significant differences in all three metrics.

(1) the Gini coefficient G (Abraham et al. 2003), which quantifies how uniformly the pixel values in an image are distributed, ranging from 0 if all pixels are equal, to 1 if a single pixel contains all the flux; (2) the second-order moment of the 20% brightest pixels M_{20} (Lotz et al. 2004), which measures how centrally peaked the pixels in a source are, with high values indicating the presence of off-nuclear clumps; and (3) the “multiplicity” Ψ (Law et al. 2007), sensitive to the presence of multiple clumps of flux with higher dynamic range than M_{20} , where higher values indicate that an image contains more and/or brighter clumps. As a reference for the dynamic range of these quantities G is by definition confined to the range 0–1, M_{20} typically ranges from -3 to -1 in optical images of nearby galaxies from the Sloan Digital Sky Survey, and Ψ ranges from 0 to ≈ 30 in rest-frame UV images of $z \sim 2$ galaxies (Lotz et al. 2004; Law et al. 2007).

We prefer these metrics as opposed to other methods for several reasons. First, they do not rely on subjective visual identification of clumps, and are deterministic and reproducible. Second, they are non-parametric and do not rely on further modeling of the reconstructions (which are themselves already models of the original data). Third, it is straightforward to compare these quantities between the absorption reconstructions and the continuum emission.

We calculate each of these quantities for the absorption and continuum reconstructions measured from the maps of equivalent width and for pixels where the continuum $S/N > 10$ (our conclusions are unchanged if we use the maps of absorbed flux instead). We calculate uncertainties for these measurements through a Monte Carlo procedure, adding normally distributed noise to the reconstructions and remeasuring G , M_{20} , and Ψ , and taking the 16th–84th percentile range of the distributions as the uncertainty on the measured values. We measure the difference between the absorption and the continuum for each of these quantities to mitigate effects from the varying spatial resolution across the source plane, because the same pixels are considered for both absorption and continuum.

Figure 10 shows the results of this procedure for the systemic and outflow absorption components of each source. This figure confirms what was apparent by eye from the reconstructions: the molecular outflows are significantly

clumpier than the continuum emission. All outflows would be considered as clumpier by at least two metrics, and six of eight outflows by all three metrics. The systemic absorption components on the other hand are less clearly clumpier than the continuum: only 3/8 sources show significant differences in all three metrics. The multiplicity Ψ shows the largest differences between outflow and systemic absorption, with 6/8 outflows showing large differences $\Delta\Psi > 5$ between the outflow and continuum, compared to only 2/8 of the systemic reconstructions.

From this analysis, we conclude that the molecular outflows in $z > 4$ DSFGs are much more irregular on 500 pc scales than the continuum emission. These metrics and the reconstructions themselves make clear that the outflows are not uniform, as expected in a simple spherically expanding geometry. The maps of equivalent width show that the absorption is rarely strongest where the continuum is brightest, even though absorption is easiest to detect in those regions. Instead the equivalent width often reaches its maximum values in small clumps near the outskirts of the continuum emission, as might be expected in the case of either a clumpy expanding shell or a scenario in which the lowest column densities in the interstellar medium are more easily removed (e.g., Thompson & Krumholz 2016; Hayward & Hopkins 2017).

5. Discussion

5.1. Expectations and Prospects for Detecting Small-scale Clumpiness

We have presented the first sample of spatially resolved molecular outflows in the early universe, reaching typical spatial resolutions ≈ 200 – 800 pc. The physics of OH absorption allows some predictions to be made for the smaller-scale structure of the outflows, beyond our current resolution limits.

OH 119 μm absorption is highly optically thick in local galaxies, $\tau_{\text{OH}119} \gg 10$ even in the line wings (e.g., Fischer et al. 2010); in effect, wherever OH is present along the line of sight $\sim 100\%$ of the 119 μm continuum light is absorbed. Alternatively, if we observe that, say, 10% of the continuum light has been absorbed from the OH spectra, we can infer that $\approx 10\%$ of the continuum source is covered by absorbing OH molecules. For our sample, the outflow absorption components

show depths of 5%–20%, which implies that only 5%–20% of the continuum emitting regions are covered by outflowing gas. These values are much smaller than either the overall detection rate ($73\% \pm 13\%$) or the covering fractions from the lensing reconstructions¹⁹. Figure 7 compares these three quantities. How can these values be reconciled?

The difference with the outflow detection rate is easiest to consider if molecular outflows typically have some fortuitous geometry, then the detection rate and the covering fractions need not be similar. Consider a simple scenario in which small spherical outflows are launched from the nuclear regions of a population of galaxies. In this case an outflow would be detected in virtually every source and from every viewing angle even if the molecular outflows cover only a small fraction of the galaxies before being decelerated or destroyed. While our reconstructions do not necessarily support this geometry, the argument is the same. There need not be a correspondence between the typical outflow covering fraction and a sample-averaged detection rate.

The comparison between the covering fractions inferred from the OH spectra and what we find from the lensing reconstructions has more interesting potential implications because now two different estimates of covering fraction are being compared for individual galaxies. The differences between the two estimates in our sample span factors of 2–10, with the lensing reconstructions always showing larger covering fractions. The reconstructions also show that the outflowing gas tends to be distributed in several distinct clumps at the spatial resolutions we achieve, but the individual clumps themselves are generally not spatially resolved. While we cannot conclusively reconcile these two methods for covering fraction with the data at hand, we briefly consider a few possibilities.

First, it is important to remember that the lensing reconstructions use data that span a wide range of velocities, using 300–700 km s^{-1} of bandwidth from the ALMA spectra. The reconstructions could thus represent the superposition of absorption clouds at many different velocities that cover only small fractions of the sources at any individual velocity but a much larger fraction of the source when considered as an ensemble. This scenario would be eminently testable with more sensitive data at the same spatial resolution as we currently have, allowing more detailed lensing reconstructions that subdivide the blueshifted line wings into narrower velocity bins.

Second, the continuum could be “filled in” by dust emission from within the outflows, which would be particularly relevant if a large fraction of the total dust mass is contained in the outflows and/or if the dust in the outflows is substantially warmer than that in the galaxies. Cold dust has been detected in a handful of local outflows (e.g. Leroy et al. 2015; Meléndez et al. 2015; Barcos-Muñoz et al. 2018), comprising $\sim 10\%$ of the total dust mass in the galaxies when it has been possible to separate the contributions. This is insufficient to explain the gap between the outflow covering fractions from our OH spectra and reconstructions unless the dust in the outflows is much warmer than that in the galaxies. Meléndez et al. (2015) find that either higher dust temperatures or a steeper dust emissivity index could explain the differences they observe between the dust emission in the disk and the outflow in the

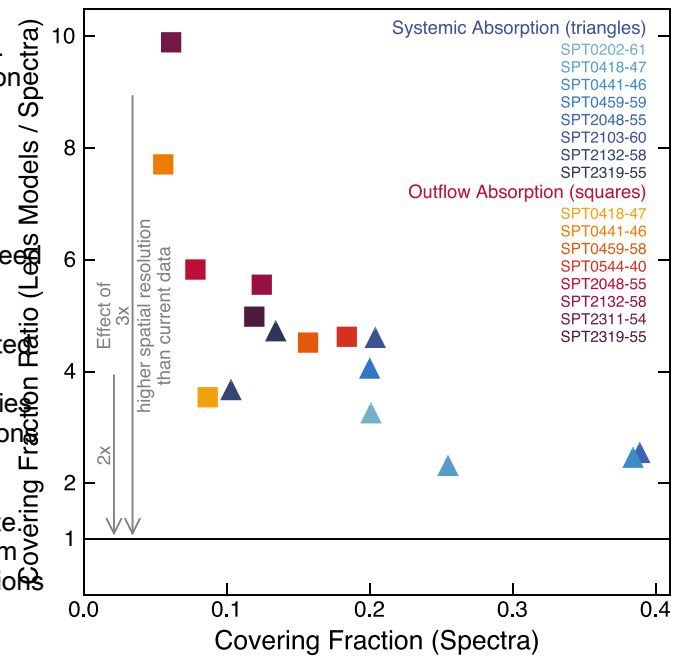


Figure 11. We estimate that the substructures in the molecular outflows would be directly resolved at $\sim 3\times$ better resolution than the current data provide, corresponding to 50–200 pc scales. The ratio of our measured covering fractions from the lensing reconstructions to those from the OH spectra provides an estimate of the spatial resolution that would be required to directly resolve the absorbing substructure because OH is expected to be highly optically thick. We implicitly assume that the spectra-based covering fractions are the “true” values and neglect any emission from dust in the outflows, which would move the points downward on this plot.

nearby galaxy NGC 4631. However, even under the assumption that the emissivity plays no role, the dust in the outflow is only $\sim 10\%$ warmer than the dust in the disk, well short of the $\sim 50\%$ – 100% higher temperatures that would be required to reconcile the estimates of covering fraction in our sample. Thus, while we expect that some portion of the absorption profiles has been filled in by emission from dust in the outflows, it is not likely to be enough to reconcile the difference in covering fractions we find.

Finally, it is very likely that the outflows we have observed contain substructure on spatial scales smaller than our current resolution limits. Most of the clumpy structures in the lensing reconstructions are not individually spatially resolved; we only identify distinct clumps because they are separated from each other by more than our effective resolution. Thus, we would expect that if these outflows were observed at higher spatial resolution then the covering fractions in the lensing reconstructions would decrease, most likely in conjunction with any decrease resulting from modeling narrower velocity bins in the outflows as well. Note that this implicitly assumes that the spectra-based covering fractions are the “true” values.

Ignoring any re-emission of continuum from dust in the outflows, the ratio between the two estimates of covering fraction corresponds to the expected size scale required to resolve the substructures in the outflows. If we take the spectrum-based covering fractions as the “true” values, then the differences with our lens modeling results must be due to insufficient spatial resolution; higher-resolution observations would then yield lower covering fractions until in agreement with the spectrum-based values. This is shown more clearly in Figure 11. This figure shows the ratio of the covering fractions

¹⁹ There is some wiggle room at the factor-of-two level allowed in this comparison if one allows significant turbulent velocity dispersion in the absorbing material at a given covering fraction, which effectively redistributes some amount of absorption to adjacent velocities; see González-Alfonso et al. (2017), their Figure 3.

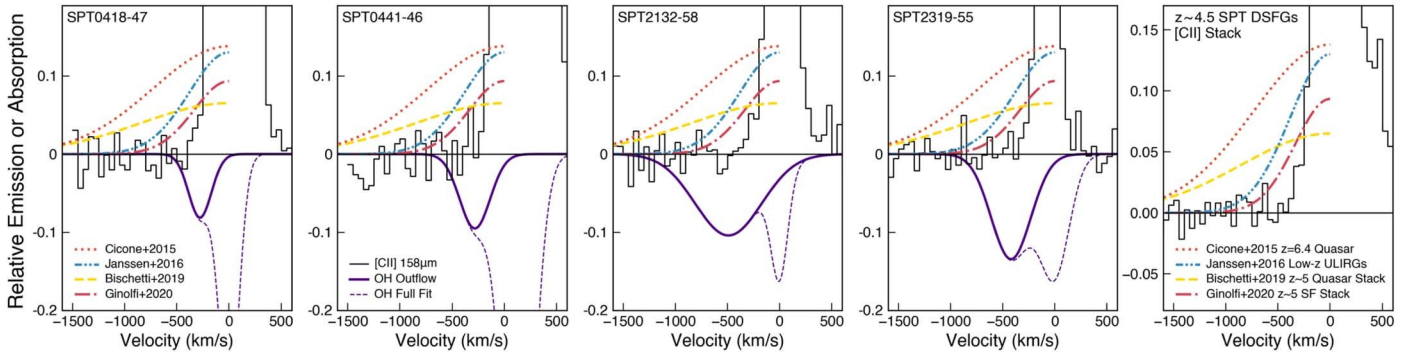


Figure 12. [C II] is at best an unreliable indicator of whether molecular outflows are present. We compare [C II] and OH for the four sources in our sample with unambiguous OH outflows and high-quality ALMA [C II] data (left four panels) and the stacked [C II] spectra of these four sources (rightmost panel). The [C II] spectra have been normalized to peak at 1. The OH best-fit absorption profiles have been normalized by the 119 μm continuum level and show only one of the double components for clarity. Colored lines show several claimed outflows seen as broad line wings in a $z \sim 6.4$ quasar (Cicone et al. 2015), the median result for a sample of 22 nearby ULIRGs (Janssen et al. 2016), a stack of 22 $z \sim 5.5$ quasars (Bischetti et al. 2019), and a stack of 25 $z \sim 5$ star-forming galaxies (Ginolfi et al. 2020). We find no evidence for excess [C II] emission at the velocities with clear blueshifted OH absorption. The S/N in our data rules out outflows similar to these studies in individual sources and strongly excludes them in the stacked spectrum.

derived from the lens models to those from the spectra (or the ratio of the bars to black lines in Figure 7). Assuming the OH absorption is optically thick, this ratio is a proxy for the spatial resolution that would be required to directly resolve the absorbing structures. We estimate that linear spatial resolutions 2–3 times better than the current data would be capable of resolving the absorption. Given the resolution and covering fractions measured in the current data, we would predict that the true size scale of clumps in the molecular outflows we have observed is of the order of 50–200 pc. We note that a similar estimate can be obtained simply by combining the covering fractions from the OH spectra with the observed continuum sizes. Even such small spatial scales are accessible with ALMA in lensed objects given sufficient observational investment.

5.2. [C II] Is Not a Reliable Molecular Outflow Tracer

We have detected molecular outflows in a large fraction of our sample, but OH is an inconvenient outflow tracer. Because the outflows are detected in absorption, galaxies bright in the far-IR continuum are required, limiting us to dusty objects. The rest frequency of the 119 μm doublet feature is also rather inconvenient because the atmospheric transmission limits observations to particular redshift ranges for $z < 6$ galaxies and is inaccessible from the ground for $z > 6.2$.

One potential alternative outflow tracer is [C II] 158 μm , where the outflow manifests as excess emission in broad line wings at high relative velocities or as emission that does not follow the overall galactic rotation curve if high-resolution data are available. [C II] is excited in a wide variety of gas conditions and arises from warm ionized neutral atomic, and molecular gas (e.g., Langer et al. 2014). In the nearby universe, Janssen et al. (2016) find tentative correlations between the OH outflow velocity and the broad [C II] line width and between the mass outflow rates derived from each tracer in a sample of ULIRGs and QSOs, suggesting that [C II] could prove a useful tracer of outflows at high redshift as well.

In the distant universe, a [C II] outflow has been reported in an individual $z \sim 6.4$ quasar (Maiolino et al. 2012; Cicone et al. 2015), although such bright [C II] line wings as in those works have not been replicated in any other individual

high-redshift source or confirmed through deeper observations of the original target. Now that ALMA observations of [C II] at $z > 4$ are fairly routine, several groups have stacked the spectra of various galaxy samples in an effort to detect broad line wings indicative of outflows, with mixed success. Gallerani et al. (2018), Bischetti et al. (2019), and Ginolfi et al. (2020) each report [C II] outflows in stacked spectra of, respectively, nine $z \sim 5.5$ UV-selected star-forming galaxies from Capak et al. (2015), 48 $4.5 < z < 7.1$ quasars assembled from the ALMA archive, and 50 $z \sim 5$ star-forming galaxies from the ALPINE survey (Le Fèvre et al. 2020; Béthermin et al. 2020; Faisst et al. 2020). Other quasar studies, however, have found no or only marginal evidence for [C II] outflows despite many sources overlapping with the sample of Bischetti et al. (e.g., Decarli et al. 2018; Stanley et al. 2019), suggesting that perhaps residual systematic uncertainties in the stacking make the detection and interpretation of broad [C II] wings less straightforward. [C II] wings are also not seen toward an individual $z \sim 6$ quasar with a tentative OH outflow, although the [C II] sensitivity is too low to be conclusive (Shao et al. 2017; Herrera-Camus et al. 2020).

Four galaxies in our sample with unambiguous OH outflows also have high-quality ALMA observations of [C II] (i.e., S/N that rivals or exceeds that of the stacking results above). These data will be analyzed in detail in future work. Figure 12 compares the OH absorption and [C II] emission profiles of these four objects. No individual source shows evidence of excess [C II] emission at the velocities where we detect OH outflows. Moreover, [C II] wings similar to those reported in the literature would have been easily detected in all individual targets²¹ with the possible exception of the stacking result of Ginolfi et al. (2020); the median SFR of those objects is more than an order of magnitude lower than that of our sample.

A still more stringent constraint comes from stacking the [C II] spectra of our four targets, which again shows no evidence for high-velocity [C II] emission, strongly excluding line wings similar to those reported in the literature (Figure 12, right). We construct this stack by subtracting a linear continuum from the integrated spectrum of each source

²⁰ That is, doubling or tripling the spatial resolution would result in a decrease in the beam area by a factor of 4–9.

²¹ To make the fairest comparison in each case, we compare to the median properties of the sources of Janssen et al. (2016), the 2" aperture fit of Cicone et al. (2015), an average of the FWHM > 400 km/s subsets of Bischetti et al. (2019), and the SFR_{median} $\sim 50 \text{ M}_{\odot} \text{ yr}^{-1}$ result of Ginolfi et al. (2020).

excluding velocities $|v| < 500 \text{ km s}^{-1}$, regridding each spectrum to a common velocity frame, and normalizing each spectrum by the peak of the $[O]$ emission before averaging. We tried a variety of other stacking procedures and continuum subtraction methods and found that our result is not sensitive to these details because the individual input spectra all have very high S/N.

At the very least, then, we must conclude that even if $[O]$ can be used as a molecular outflow tracer, it does not trace it reliably given our non-detections in sources with clear OH outflows. We note that any vagaries of gravitational lensing or differential magnification cannot explain the lack of high-velocity $[C II]$ emission: each source shows high-velocity OH absorption, which must necessarily be down-the-barrel and share essentially the same magnification as the dust continuum. Even if the $[C II]$ traces a more extended or diffuse outflow component, the portion along our line of sight would still have been detectable just as the OH was. We also note that the high-velocity $[C II]$ found by Ginolfi et al. (2020) in stacked images is compact and centered on the galaxies, while the core systemic emission is more extended, as expected for a shell-like geometry. We consider a few other possible explanations for our lack of evidence for $[O]$ outflows.

First, none of the literature stacking results targeted nor necessarily includes DSFGs such as our objects; no clean comparison result for DSFGs is currently available. The stacking result of Ginolfi et al. (2020) is the most difficult to rule out with our sample due to the intrinsically broader line profiles of our sources. The galaxies in that sample have both lower mass and lower SFR than our targets, and it seems unlikely that such objects would drive higher outflow rate or brighter outflows than our more extreme targets. Meanwhile, the quasars of Bischetti et al. (2019) are probably fairly similar to our target galaxies in mass, but we have no evidence for AGN in any of our targets. Perhaps the additional energy and momentum from an accreting black hole are required to boost the $[C II]$ line wings to observable fluxes.

Second, most recent $[O]$ outflow detections have relied on stacking low-S/N spectra of many objects which introduces additional complexity and uncertainty in the results. Given the faint line wings to be searched for and heterogeneity in the input objects, great care must be taken in the details of the stacking, including continuum subtraction, relative scaling, non-Gaussian line shapes in the input spectra, and the treatment of differing line widths among input sources. These details may also explain the differences in the quasar sample stacking results in the literature even though these studies have many individual sources in common (Decarli et al. 2018; Bischetti et al. 2019; Stanley et al. 2019). It could also be that $[O]$ is rarely apparent in outflows, but when it is present it is unusually bright, such that the stacked signal seen in literature studies is dominated by a few atypical objects.

Third, OH is in principle a far more sensitive tracer of weak outflows than $[C II]$, mostly due to the fact that it is an absorption tracer sensitive to column (rather than volume) density and the very high Einstein A coefficients of the doublet lines. While we consider it highly unlikely given the wealth of evidence for high line opacities at low redshift, if the OH absorption were optically thin, the OH data are sensitive to outflow masses as low as $\sim (2-6) \times 10^7 M_\odot$ (Paper II). On the other hand, following the calculations of Hailey-Dunsheath et al. (2010), our $[C II]$ data are sensitive to outflow masses

$\sim (5-15) \times 10^7 M_\odot$, or $\sim 5\%$ of the total molecular gas masses of these galaxies. These values are highly uncertain due to the unknown abundances of OH and C and the ionization fraction of C^+ but would reconcile the OH and $[C II]$ observations. More importantly, it is almost certainly not true that OH is optically thin (even the $[C II]$ may have non-negligible line opacity, though still far lower than expected for OH; e.g., Gullberg et al. 2015).

Finally, $[C II]$ is often stated to be a tracer of neutral atomic outflowing gas, and not necessarily of molecular material (although $[O I]$ is known to be emitted in both these phases as well as ionized gas). Perhaps, then, the molecular gas dominates the mass budget of these outflows, with atomic gas a smaller contribution to the total outflow mass. In Paper II, we find best estimates for the molecular outflow masses for these four sources $\sim (40-80) \times 10^7 M_\odot$. If we assume that the sensitivity to outflow mass from $[C II]$ in the previous paragraph is solely for the atomic phase, this would require molecular-to-atomic outflow mass ratios of $\sim 5-10$ to reconcile the $[C II]$ and OH observations. While the distribution of mass across outflow phases is completely unknown for high-redshift DSFGs from observations, this range is not infeasible based on estimates from chemical modeling in idealized hydrodynamic simulations (Richings & Faucher-Giguère 2018) or the few lower-redshift observational estimates (Feruglio et al. 2010; Rupke & Veilleux 2013; Herrera-Camus et al. 2019).

Regardless of the reason, our data make clear that $[C II]$ should not be considered a reliable molecular outflow tracer. We find no evidence for excess $[C]$ emission at the velocities where we see blueshifted OH absorption or excess emission at the level seen in literature stacking experiments. Further work is needed to determine exactly why $[C II]$ does not always appear in outflows.

6. Conclusions

In this work we have presented the first statistical sample of molecular outflows observed in $4 < z < 5.5$ DSFGs using ALMA observations of the ground-state OH $119 \mu\text{m}$ doublet lines toward 11 DSFGs selected from the SPT DSFG sample. Our target galaxies, all of which are gravitationally lensed, are IR-luminous, $\log(L_{\text{IR}}/L_\odot) > 12$, but do not show obvious signs of AGN activity in rest-frame mid-IR data. We detect OH in absorption in all objects. The observations also spatially resolve the targets, and we create source reconstructions of the rest-frame $120 \mu\text{m}$ dust continuum emission and the OH line absorption. These galaxies represent the largest sample at $z > 4$ with constraints on the molecular outflow properties, as well as the largest (and so far only) sample to spatially resolve the structure of said outflows in the early universe. Our main conclusions can be summarized as follows.

1. We find unambiguous evidence for molecular outflows in 8 of the 11 sample targets; the remaining sources have broad CO or $[O]$ line profiles that make interpretation of the OH absorption complicated. No source shows unambiguous evidence for molecular inflows. The outflow detection rate, $73\% \pm 13\%$, is similar to that found from OH observations of nearby ULIRGs and QSOs and significantly higher than for lower-luminosity AGN-dominated galaxies at low redshift. Because outflows with OH are only detectable in absorption, this detection rate is a lower limit on the true occurrence rate of

molecular outflows in $z > 4$ DSFGs; molecular outflows must be ubiquitous and widespread in such objects (Figure 4 and Section 4.1).

2. The outflows reach maximum velocities of 430–1200 km s⁻¹ and possibly faster. The distribution of outflow velocities in the $z > 4$ sample is indistinguishable from that of low-redshift IR-luminous galaxies (Figure 5 and Section 4.2).
3. Using our lensing reconstructions of the sources, we measure the structure of the outflows on ~500 pc spatial scales. The covering fraction of the outflowing molecular gas is correlated with both L and outflow velocity, such that more luminous sources and those with the fastest outflows also host the outflows with the highest covering fractions (Figures 8 and 9).
4. The molecular outflows show significantly more clumpy structure than the (generally smooth) dust continuum emission on scales of ~500 pc, which we quantify with metrics borrowed from the literature on high-redshift star formation (Figure 10 and Section 4.4). While the clumps are not directly resolved, from optical depth arguments we expect that higher-resolution observations of the outflows would reveal substantially more clumpy structure on 50–200 pc spatial scales (Figure 11 and Section 5.1).
5. We find no evidence of high-velocity wings in the [C I] line profiles of the four sources with high-quality [C I] spectra and obvious OH outflows; [C II] is at best an unreliable indicator of molecular outflows. We strongly rule out [C II] line wings at the level reported in several literature stacking results (though none explicitly targeted high-redshift DSFGs similar to our sample). This may be due to lingering systematics in the stacking results or genuine differences between the [C II] outflow properties of DSFGs like our sample and the variety of objects for which [C II] outflows have been claimed (Figure 12 and Section 5.2).

This work has largely focused on the structural properties of the OH outflows we have detected. In the second paper in this series we explore the physical properties of these outflows in much greater detail, including outflow rates and masses and implications for the outflow driving mechanisms. Moreover, the outflow occurrence rate, velocity distributions, and detailed structural properties we find place novel observational constraints on simulations of galactic feedback and winds in the early universe. While clearly a larger high-redshift sample that spans a wider range in galaxy properties will be required to fully understand galactic winds in the distant universe, this sample represents a first step toward understanding outflow properties among the general high-redshift galaxy population and the utility of OH absorption for characterizing these properties in statistical samples.

We thank the referee for a thorough and constructive report that improved the quality of this paper. J.S.S. thanks the McDonald Observatory at the University of Texas at Austin for support through a Harlan J. Smith Fellowship and the Texas Advanced Computing Center (TACC) for providing

high-performance computing resources. S.S. is supported by NASA Hubble Fellowship grant #HF2-51446 awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS5-26555. K.C.L., S.J., D.P.M., K.P., and J.D.V. acknowledge support from the US NSF under grants AST-1715213 and AST-1716127. This work was performed in part at the Aspen Center for Physics, which is supported by National Science Foundation grant PHY-1607611.

This paper makes use of the following ALMA data: ADS/JAO.ALMA#2015.1.00942.S, ADS/JAO.ALMA#2016.1.00089.S, ADS/JAO.ALMA#2016.1.01499.S, ADS/JAO.ALMA#2018.1.00191.S, and ADS/JAO.ALMA#2019.1.00253.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 National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc.

This research has made use of NASA's Astrophysics Data System.

Facility: ALMA.

Software: CASA (McMullin et al. 2007), visilens (Spilker et al. 2016b), ripples (Hezaveh et al. 2016), astropy (Astropy Collaboration et al. 2018), matplotlib (Hunter 2007).

Appendix

Supplementary Lens Model Results

This appendix provides additional results and diagnostic plots from the lens modeling procedure.

Table A1 lists the best-fit lens model parameters. As described in Section 3.3, the lensing potential in each source is parameterized as one or more SIE mass profiles with optional external tidal shear and low-order angular multipoles (see Hezaveh et al. 2016 for a more thorough description of these parameters).

Figure A1 shows the data, model, and residuals from the lens modeling for the 119 μ m continuum.

Figure A2 shows the effective FWHM resolution of the lensing reconstructions as a function of position in the source plane. These maps were made from reconstructions of mock data analyzed in the same way as the real data; see Section 3.3.1. These images show the resolution for the 119 μ m continuum reconstructions. A similar exercise with the noise properties of the frequency ranges of the OH absorption components in each source gives nearly identical results. These maps make intuitive sense in that regions near the caustics with high magnification correspond to better source-plane resolution. We find good agreement with the usual rule of thumb that source-plane resolution is approximately the data resolution divided by the square root of the local magnification.

Table A1
Best-fit Lens Model Parameters

Source	Δx	Δy	$\log(M(< 10 \text{ kpc})/M_\odot)$	e_x	e_y	Source	γ_x	γ_y	A_3
SPT0202-61	$+0.016 \pm 0.004$	$+0.070 \pm 0.002$	11.221 ± 0.010	$+0.22 \pm 0.07$	-0.53 ± 0.13	SPT0202-61	$+0.281 \pm 0.007$	$+0.074 \pm 0.004$	0.087 ± 0.005
SPT0418-47	$+0.097 \pm 0.004$	-0.019 ± 0.004	11.537 ± 0.005	$+0.09 \pm 0.02$	-0.23 ± 0.10	SPT0418-47	0.000 ± 0.005	-0.011 ± 0.005	L
SPT0441-46	$+0.079 \pm 0.010$	$+0.338 \pm 0.007$	11.385 ± 0.013	$+0.03 \pm 0.04$	-0.19 ± 0.06	SPT0441-46	-0.074 ± 0.019	-0.019 ± 0.025	-0.086 ± 0.013
SPT0459-58	-0.068 ± 0.022	$+0.045 \pm 0.018$	10.994 ± 0.025	$+0.27 \pm 0.05$	$+0.01 \pm 0.05$	SPT0459-58	L	L	L
	-0.378 ± 0.023	$+0.341 \pm 0.026$	10.563 ± 0.058	$+0.53 \pm 0.07$	-0.31 ± 0.09				
SPT0459-59	-0.280 ± 0.016	$+0.699 \pm 0.016$	11.246 ± 0.034	$+0.27 \pm 0.09$	-0.12 ± 0.07	SPT0459-59	-0.031 ± 0.022	-0.032 ± 0.043	L
SPT0544-40	-0.075 ± 0.002	$+0.056 \pm 0.003$	11.055 ± 0.010	$+0.40 \pm 0.02$	$+0.26 \pm 0.02$	SPT0544-40	-0.030 ± 0.010	-0.085 ± 0.012	L
SPT2048-55	-0.006 ± 0.017	$+0.007 \pm 0.016$	11.020 ± 0.007	$+0.17 \pm 0.21$	0.21 ± 0.15	SPT2048-55	$+0.067 \pm 0.119$	$+0.080 \pm 0.082$	L
SPT2103-60	$+0.851 \pm 0.005$	-0.485 ± 0.006	11.151 ± 0.001	$+0.42 \pm 0.02$	-0.24 ± 0.02	SPT2103-60	L	L	L
	$+0.117 \pm 0.020$	$+1.113 \pm 0.021$	11.161 ± 0.002	$+0.43 \pm 0.02$	-0.45 ± 0.02				
	-1.423 ± 0.043	-1.839 ± 0.149	11.022 ± 0.003	$+0.84 \pm 0.02$	0.11 ± 0.02				
SPT2132-58	-0.174 ± 0.023	-0.031 ± 0.014	10.874 ± 0.003	-0.15 ± 0.04	-0.74 ± 0.05	SPT2132-58	$+0.101 \pm 0.050$	-0.035 ± 0.036	L
SPT2311-54	-0.081 ± 0.018	-0.345 ± 0.026	10.657 ± 0.003	$+0.17 \pm 0.06$	-0.45 ± 0.08	SPT2311-54	L	L	L

Note. Parameter descriptions are as follows. Δx , Δy : lens position in arcseconds relative to the ALMA phase center. $\log(M(< 10 \text{ kpc})/M_\odot)$: mass contained within 10 kpc components. γ_x , γ_y : external tidal shear components. A_3 : $m=3$ multipole components. A_4 : $m=4$ multipole components. Entries with ellipses were fixed to 0 degrees (SPT0459-58 and SPT2103-60) list each lens; no shear or multipole parameters were used for these sources.

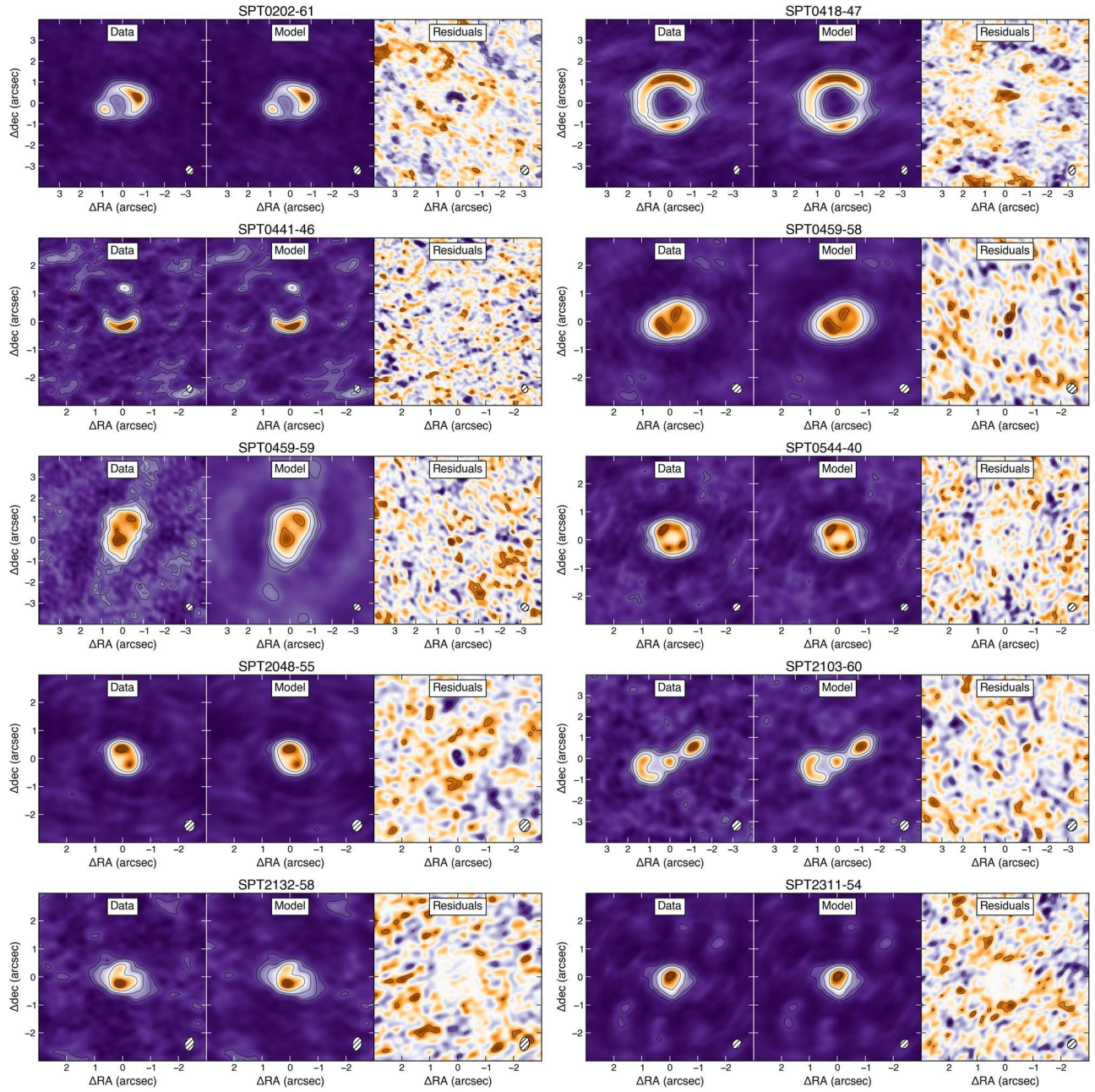


Figure A1. Continuum lens modeling results showing the data, model, and residuals. Contours in the data and models are 5%, 10%, 20%, ... of the peak, and contours in the residuals are in steps of $\pm 2\sigma$. Axes are relative to the ALMA phase center. Data and model images are dirty (not deconvolved) because the fitting is performed in the Fourier domain; the sidelobe structure resulting from the uv coverage of the observations in the data should be reproduced in the model. Emission from the lens in SPT0202-61 and southwestern source in SPT0459-59 has been subtracted in the visibilities prior to the lens modeling (visible in Figure 2); it is clear from the residual maps that this subtraction is imperfect.

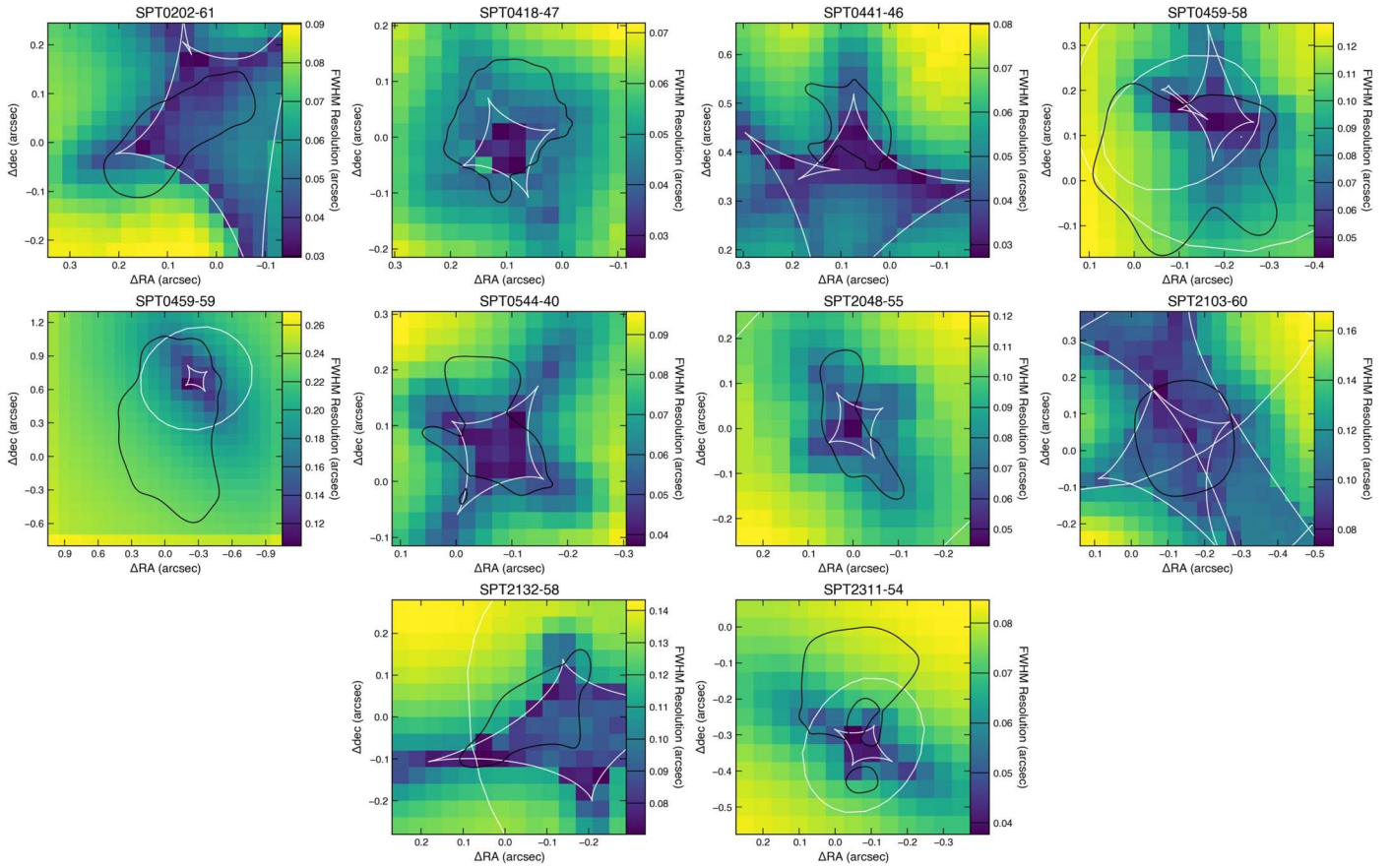


Figure A2. Maps of the effective source-plane resolution for each source made from reconstructions of mock data; see Section 3.3.1. Each image shows the FWHM of a 2D Gaussian fit to the reconstruction of a pointlike artificial source at the center of each map pixel, “observed” and analyzed identically to the real data. Axes are relative to the ALMA phase center. The black contour shows $S/N = 5$ for the actual source as in Figure 1. The white contours show the lensing caustics.

ORCID iDs

Justin S. Spilker <https://orcid.org/0000-0003-3256-5615>
 Kedar A. Phadke <https://orcid.org/0000-0001-7946-557X>
 Manuel Aravena <https://orcid.org/0000-0002-6290-3198>
 Matthieu Béthermin <https://orcid.org/0000-0002-3915-2015>
 Chenxing Dong (董辰兴) <https://orcid.org/0000-0002-5823-0349>
 Anthony H. Gonzalez <https://orcid.org/0000-0002-0933-8601>
 Christopher C. Hayward <https://orcid.org/0000-0003-4073-3236>
 Yashar D. Hezaveh <https://orcid.org/0000-0002-8669-5733>
 Sreevani Jarugula <https://orcid.org/0000-0002-5386-7076>
 Katrina C. Litke <https://orcid.org/0000-0002-4208-3532>
 Matthew A. Malkan <https://orcid.org/0000-0001-6919-1237>
 Daniel P. Marrone <https://orcid.org/0000-0002-2367-1080>
 Desika Narayanan <https://orcid.org/0000-0002-7064-4309>
 Cassie Reuter <https://orcid.org/0000-0001-7477-1586>
 Joaquin D. Vieira <https://orcid.org/0000-0001-7192-3871>
 Axel Weiss <https://orcid.org/0000-0003-4678-3939>

References

- Abraham, R. G., van den Bergh, S., & Nair, P. 2003, *ApJ*, **588**, 218
 Alatalo, K., Blitz, L., Young, L. M., et al. 2011, *ApJ*, **735**, 88
 Aravena, M., Spilker, J. S., Béthermin, M., et al. 2016, *MNRAS*, **457**, 4406
 Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, *AJ*, **156**, 123
 Banerji, M., Chapman, S. C., Smail, I., et al. 2011, *MNRAS*, **418**, 1071
 Barcos-Muñoz, L., Aalto, S., Thompson, T. A., et al. 2018, *ApJL*, **853**, L28
 Béthermin, M., De Breuck, C., Gullberg, B., et al. 2016, *A&A*, **586**, L7
 Béthermin, M., Fudamoto, Y., Ginolfi, M., et al. 2020, *A&A*, **643**, A2
 Bischetti, M., Maiolino, R., Carniani, S., et al. 2019, *A&A*, **630**, A59
 Boquien, M., Burgarella, D., Roehlly, Y., et al. 2019, *A&A*, **622**, A103
 Burgarella, D., Buat, V., & Iglesias-Páramo, J. 2005, *MNRAS*, **360**, 1413
 Calderón, D., Bauer, F. E., Veilleux, S., et al. 2016, *MNRAS*, **460**, 3052
 Capak, P. L., Carilli, C., Jones, G., et al. 2015, *Natur*, **522**, 455
 Carlstrom, J. E., Ade, P. A. R., Aird, K. A., et al. 2011, *PASP*, **123**, 568
 Carnall, A. C., Walker, S., McLure, R. J., et al. 2020, *MNRAS*, **496**, 695
 Casey, C. M., Cooray, A., Killi, M., et al. 2017, *ApJ*, **840**, 101
 Chisholm, J., Tremonti, C. A., Leitherer, C., et al. 2015, *ApJ*, **811**, 149
 Ciccone, C., Maiolino, R., Gallerani, S., et al. 2015, *A&A*, **574**, A14
 Decarli, R., Walter, F., Venemans, B. P., et al. 2018, *ApJ*, **854**, 97
 Estrada-Carpenter, V., Papovitch, C., Momcheva, I., et al. 2020, *ApJ*, **898**, 171
 Everett, W. B., Zhang, L., Crawford, T. M., et al. 2020, *ApJ*, **900**, 55
 Faist, A. L., Schaerer, D., Lemaux, B. C., et al. 2020, *ApJS*, **247**, 61
 Falstad, N., Hallqvist, F., Aalto, S., et al. 2019, *A&A*, **623**, A29
 Fan, L., Knudsen, K. K., Fogasy, J., & Drouart, G. 2018, *ApJL*, **856**, L5
 Farrah, D., Leboutteiller, V., Spoon, H. W. W., et al. 2013, *ApJ*, **776**, 38
 Feruglio, C., Ferrara, A., Bischetti, M., et al. 2017, *A&A*, **608**, A30
 Feruglio, C., Maiolino, R., Piconcelli, E., et al. 2010, *A&A*, **518**, L155
 Fischer, J., Sturm, E., González-Alfonso, E., et al. 2010, *A&A*, **518**, L41
 Forrest, B., Annunziatella, M., Wilson, G., et al. 2020, *ApJL*, **890**, L1
 Förster Schreiber, N. M., Shapley, A. E., Erb, D. K., et al. 2011, *ApJ*, **731**, 65
 Gallerani, S., Pallottini, A., Feruglio, C., et al. 2018, *MNRAS*, **473**, 1909
 García-Burillo, S., Combes, F., Ramos Almeida, C., et al. 2019, *A&A*, **632**, A61
 George, R. D., Ivison, R. J., Smail, I., et al. 2014, *MNRAS*, **442**, 1877
 Ginolfi, M., Jones, G. C., Béthermin, M., et al. 2020, *A&A*, **633**, A90
 Glazebrook, K., Schreiber, C., Labbé, I., et al. 2017, *Natur*, **544**, 71
 González-Alfonso, E., Fischer, J., Spoon, H. W. W., et al. 2017, *ApJ*, **836**, 11
 Guarnieri, P., Maraston, C., Thomas, D., et al. 2019, *MNRAS*, **483**, 3060
 Gullberg, B., De Breuck, C., Vieira, J. D., et al. 2015, *MNRAS*, **449**, 2883
 Guo, Y., Rafelski, M., Bell, E. F., et al. 2018, *ApJ*, **853**, 108
 Hailey-Dunsheath, S., Nikola, T., Stacey, G. J., et al. 2010, *ApJL*, **714**, L162

- Hayward, C. C., & Hopkins, P. F. 2017, *MNRAS*, **465**, 1682
- Heckman, T. M., Armus, L., & Miley, G. K. 1990, *ApJS*, **74**, 833
- Herrera-Camus, R., Sturm, E., Graciá-Carpio, J., et al. 2020, *A&A*, **633**, L4
- Herrera-Camus, R., Tacconi, L., Genzel, R., et al. 2019, *ApJ*, **871**, 37
- Hezaveh, Y. D., Dalal, N., Marrone, D. P., et al. 2016, *ApJ*, **823**, 37
- Hezaveh, Y. D., Marrone, D. P., Fassnacht, C. D., et al. 2013, *ApJ*, **767**, 132
- Hodge, J. A., Smail, I., Walter, F., et al. 2019, *ApJ*, **876**, 130
- Hodge, J. A., Swinbank, A. M., Simpson, J. M., et al. 2016, *ApJ*, **833**, 103
- Hunter, J. D. 2007, *CSE*, **9**, 90
- Ivison, R. J., Richard, J., Biggs, A. D., et al. 2020, *MNRAS*, **495**, L1
- Janssen, A. W., Christopher, N., Sturm, E., et al. 2016, *ApJ*, **822**, 43
- Jones, G. C., Maiolino, R., Caselli, P., & Carniani, S. 2019, *A&A*, **632**, L7
- Klein, R. I., McKee, C. F., & Colella, P. 1994, *ApJ*, **420**, 213
- Kormann, R., Schneider, P., & Bartelmann, M. 1994, *A&A*, **284**, 285
- Krieger, N., Bolatto, A. D., Walter, F., et al. 2019, *ApJ*, **881**, 43
- Langer, W. D., Velusamy, T., Pineda, J. L., Willacy, K., & Goldsmith, P. F. 2014, *A&A*, **561**, A122
- Law, D. R., Steidel, C. C., Erb, D. K., et al. 2007, *ApJ*, **656**, 1
- Le Fèvre, O., Béthermin, M., Faisst, A., et al. 2020, *A&A*, **643**, A1
- Leroy, A. K., Walter, F., Martini, P., et al. 2015, *ApJ*, **814**, 83
- Litke, K. C., Marrone, D. P., Spilker, J. S., et al. 2019, *ApJ*, **870**, 80
- Lotz, J. M., Primack, J., & Madau, P. 2004, *AJ*, **128**, 163
- Lutz, D., Sturm, E., Janssen, A., et al. 2020, *A&A*, **633**, A134
- Maiolino, R., Gallerani, S., Neri, R., et al. 2012, *MNRAS*, **425**, L66
- Marrone, D. P., Spilker, J. S., Hayward, C. C., et al. 2018, *Natur*, **553**, 51
- McCourt, M., Oh, S. P., O'Leary, R., & Madigan, A.-M. 2018, *MNRAS*, **473**, 5407
- 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**
- Meléndez, M., Veilleux, S., Martin, C., et al. 2015, *ApJ*, **804**, 46
- Mocanu, L. M., Crawford, T. M., Vieira, J. D., et al. 2013, *ApJ*, **779**, 61
- Narayanan, D., Turk, M., Feldmann, R., et al. 2015, *Natur*, **525**, 496
- Nelson, E. J., van Dokkum, P. G., Förster Schreiber, N. M., et al. 2016, *ApJ*, **828**, 27
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, *A&A*, **594**, A13
- Popping, G., Decarli, R., Man, A. W. S., et al. 2017, *A&A*, **602**, A11
- Reuter, C., Vieira, J. D., Spilker, J. S., et al. 2020, *ApJ*, **902**, 78
- Richings, A. J., & Faucher-Giguère, C.-A. 2018, *MNRAS*, **474**, 3673
- Rujopakarn, W., Daddi, E., Rieke, G. H., et al. 2019, *ApJ*, **882**, 107
- Runco, J. N., Malkan, M. A., Fernández-Ontiveros, A., Spinoglio, L., & Pereira-Santaella, M. 2020, arXiv:2009.07293
- Rupke, D. S., Veilleux, S., & Sanders, D. B. 2005, *ApJ*, **632**, 751
- Rupke, D. S. N., & Veilleux, S. 2013, *ApJL*, **775**, L15
- Scannapieco, E. 2013, *ApJL*, **763**, L31
- Schechter, A. L., & Casey, C. 2018, *RNAAS*, **2**, 228
- Schneider, E. E., & Robertson, B. E. 2017, *ApJ*, **834**, 144
- Schneider, E. E., Robertson, B. E., & Thompson, T. A. 2018, *ApJ*, **862**, 56
- Shao, Y., Wang, R., Jones, G. C., et al. 2017, *ApJ*, **845**, 138
- Snyder, G. F., Hayward, C. C., Sajina, A., et al. 2013, *ApJ*, **768**, 168
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. 2002, *Journal of the Royal Statistical Society Series B*, **64**, 583
- Spilker, J. S., Aravena, M., Béthermin, M., et al. 2018, *Sci*, **361**, 1016
- Spilker, J. S., Aravena, M., Marrone, D. P., et al. 2015, *ApJ*, **811**, 124
- Spilker, J. S., Aravena, M., Phadke, K. A., et al. 2020, *ApJ*, **905**, 86
- Spilker, J. S., Bezanson, R., Marrone, D. P., et al. 2016a, *ApJ*, **832**, 19
- Spilker, J. S., Bezanson, R., Weiner, B. J., Whitaker, K. E., & Williams, C. C. 2019, *ApJ*, **883**, 81
- Spilker, J. S., Marrone, D. P., Aguirre, J. E., et al. 2014, *ApJ*, **785**, 149
- Spilker, J. S., Marrone, D. P., Aravena, M., et al. 2016b, *ApJ*, **826**, 112
- Spinoglio, L., Malkan, M. A., Smith, H. A., González-Alfonso, E., & Fischer, J. 2005, *ApJ*, **623**, 123
- Spoon, H. W. W., Farrah, D., Leboutteiller, V., et al. 2013, *ApJ*, **775**, 127
- Stanley, F., Jolly, J. B., König, S., & Knudsen, K. K. 2019, *A&A*, **631**, A78
- Stone, M., Veilleux, S., Meléndez, M., et al. 2016, *ApJ*, **826**, 111
- Straatman, C. M. S., Labbé, I., Spitler, L. R., et al. 2014, *ApJL*, **783**, L14
- Strandet, M. L., Weiss, A., Vieira, J. D., et al. 2016, *ApJ*, **822**, 80
- Sturm, E., González-Alfonso, E., Veilleux, S., et al. 2011, *ApJL*, **733**, L16
- Suyu, S. H., Marshall, P. J., Hobson, M. P., & Blandford, R. D. 2006, *MNRAS*, **371**, 983
- Tacchella, S., Carollo, C. M., Förster Schreiber, N. M., et al. 2018, *ApJ*, **859**, 56
- Tacchella, S., Carollo, C. M., Renzini, A., et al. 2015, *Sci*, **348**, 314
- Tadaki, K.-i., Genzel, R., Kodama, T., et al. 2017, *ApJ*, **834**, 135
- Talia, M., Pozzi, F., Vallini, L., et al. 2018, *MNRAS*, **476**, 3956
- Thompson, T. A., & Krumholz, M. R. 2016, *MNRAS*, **455**, 334
- Valentino, F., Tanaka, M., Davidzon, I., et al. 2020, *ApJ*, **889**, 93
- Veilleux, S., Maiolino, R., Bolatto, A. D., & Aalto, S. 2020, *A&ARv*, **28**, 2
- Veilleux, S., Meléndez, M., Sturm, E., et al. 2013, *ApJ*, **776**, 27
- Vieira, J. D., Crawford, T. M., Switzer, E. R., et al. 2010, *ApJ*, **719**, 763
- Vieira, J. D., Marrone, D. P., Chapman, S. C., et al. 2013, *Natur*, **495**, 344
- Walter, F., Bolatto, A. D., Leroy, A. K., et al. 2017, *ApJ*, **835**, 265
- Walter, F., Weiss, A., & Scoville, N. 2002, *ApJL*, **580**, L21
- Warren, S. J., & Dye, S. 2003, *ApJ*, **590**, 673
- Weiner, B. J., Coil, A. L., Prochaska, J. X., et al. 2009, *ApJ*, **692**, 187
- Weiß, A., De Breuck, C., Marrone, D. P., et al. 2013, *ApJ*, **767**, 88
- Weiß, A., Walter, F., Downes, D., et al. 2012, *ApJ*, **753**, 102
- Wuyts, S., Förster Schreiber, N. M., Genzel, R., et al. 2012, *ApJ*, **753**, 114
- Zhang, Z.-Y., Ivison, R. J., George, R. D., et al. 2018, *MNRAS*, **481**, 59
- Zubovas, K., & King, A. R. 2014, *MNRAS*, **439**, 400