



**HAL**  
open science

## A massive core for a cluster of galaxies at a redshift of 4.3

T. B. B Miller, S. C. Chapman, M. Aravena, M. L. N. Ashby, C. C. C Hayward, J. D. Vieira, A. Weiss, A. Babul, M. Bethermin, C. M. Bradford, et al.

► **To cite this version:**

T. B. B Miller, S. C. Chapman, M. Aravena, M. L. N. Ashby, C. C. C Hayward, et al.. A massive core for a cluster of galaxies at a redshift of 4.3. *Nature*, 2018, 556 (7702), pp.469 - 472. 10.1038/s41586-018-0025-2. hal-01789108

**HAL Id: hal-01789108**

**<https://hal.science/hal-01789108>**

Submitted on 9 May 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A massive core for a cluster of galaxies at a redshift of 4.3

T. B. Miller<sup>1,2</sup>, S. C. Chapman<sup>1,3,4</sup>, M. Aravena<sup>5</sup>, M. L. N. Ashby<sup>6</sup>, C. C. Hayward<sup>6,7</sup>, J. D. Vieira<sup>8</sup>, A. Weiß<sup>9</sup>, A. Babul<sup>10</sup>, M. Béthermin<sup>11</sup>, C. M. Bradford<sup>12,13</sup>, M. Brodwin<sup>14</sup>, J. E. Carlstrom<sup>15,16,17,18</sup>, Chian-Chou Chen<sup>19</sup>, D. J. M. Cunningham<sup>1,20</sup>, C. De Breuck<sup>19</sup>, A. H. Gonzalez<sup>21</sup>, T. R. Greve<sup>22</sup>, J. Harnett<sup>23</sup>, Y. Hezaveh<sup>24</sup>, K. Lacaille<sup>1,25</sup>, K. C. Litke<sup>26</sup>, J. Ma<sup>21</sup>, M. Malkan<sup>27</sup>, D. P. Marrone<sup>26</sup>, W. Morningstar<sup>24</sup>, E. J. Murphy<sup>28</sup>, D. Narayanan<sup>21</sup>, E. Pass<sup>1,29</sup>, R. Perry<sup>1</sup>, K. A. Phadke<sup>8</sup>, D. Rennehan<sup>10</sup>, K. M. Rotermund<sup>1</sup>, J. Simpson<sup>30,31</sup>, J. S. Spilker<sup>26</sup>, J. Sreevani<sup>8</sup>, A. A. Stark<sup>6</sup>, M. L. Strandet<sup>9,32</sup> & A. L. Strom<sup>33</sup>

<sup>1</sup>Department of Physics and Atmospheric Science, Dalhousie University, Halifax, Canada

<sup>2</sup>Department of Astronomy, Yale University, 52 Hillhouse Avenue, New Haven, CT, 06511, USA

<sup>3</sup>Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK

<sup>4</sup>National Research Council, Herzberg Astronomy & Astrophysics, 5071 West Saanich Road, Victoria, BC, V9E 2E7, Canada

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

<sup>6</sup>Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA

<sup>7</sup>Center for Computational Astrophysics, Flatiron Institute, 162 Fifth Avenue, New York, NY 10010, USA

<sup>8</sup>Department of Astronomy, University of Illinois, 1002 West Green Street, Urbana, IL 61801, USA

<sup>9</sup>Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69 D-53121 Bonn, Germany

<sup>10</sup>Department of Physics and Astronomy, University of Victoria, Victoria, BC V8P 1A1, Canada

<sup>11</sup>Aix-Marseille Université, CNRS, LAM, Laboratoire d'Astrophysique de Marseille, Marseille, France

<sup>12</sup>California Institute of Technology, 1200 E. California Boulevard, Pasadena, CA 91125, USA

<sup>13</sup>Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, CA 91109, USA

<sup>14</sup>Department of Physics and Astronomy, University of Missouri, 5110 Rockhill Road, Kansas City, MO 64110, USA

<sup>15</sup>Kavli Institute for Cosmological Physics, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA

<sup>16</sup>Department of Physics, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA

<sup>17</sup>Enrico Fermi Institute, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA

<sup>18</sup>Department of Astronomy and Astrophysics, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA

<sup>19</sup>European Southern Observatory, Karl Schwarzschild Straße 2, 85748 Garching, Germany

<sup>20</sup>Department of Astronomy and Physics, Saint Mary's University, Halifax, Nova Scotia, Canada

<sup>21</sup>Department of Astronomy, University of Florida, Bryant Space Sciences Center, Gainesville, FL 32611, USA

<sup>22</sup>Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK

<sup>23</sup>School of Physics, University of Sydney, Sydney, NSW 2006, Australia

<sup>24</sup>Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, Stanford, CA 94305, USA

<sup>25</sup>Department of Physics and Astronomy, McMaster University, Hamilton, ON L8S4M1, Canada

<sup>26</sup>Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721, USA

<sup>27</sup>Department of Physics and Astronomy, University of California, Los Angeles, CA 90095, USA

<sup>28</sup>National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, VA 22903, USA

<sup>29</sup>Department of Physics and Astronomy, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada

<sup>30</sup>Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK

<sup>31</sup>Centre for Extragalactic Astronomy, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK

<sup>32</sup>International Max Planck Research School (IMPRS) for Astronomy and Astrophysics, Universities of Bonn and Cologne, Germany

<sup>33</sup>Observatories of The Carnegie Institution for Science, 813 Santa Barbara Street, Pasadena, CA 91101, USA

**Massive galaxy clusters are now found as early as  $\sim 3$  billion years after the Big Bang, containing stars that formed at even earlier epochs.<sup>1-3</sup> The high-**

redshift progenitors of these galaxy clusters, termed ‘protoclusters’, are identified in cosmological simulations with the highest dark matter overdensities.<sup>4-6</sup> While their observational signatures are less well defined compared to virialized clusters with a substantial hot intra-cluster medium (ICM), protoclusters are expected to contain extremely massive galaxies that can be observed as luminous starbursts.<sup>7</sup> Recent claimed detections of protoclusters hosting such starbursts<sup>8-11</sup> do not support the kind of rapid cluster core formation expected in simulations<sup>12</sup> because these structures contain only a handful of starbursting galaxies spread throughout a broad structure, with poor evidence for eventual collapse into a protocluster. Here we report that the source SPT2349-56 consists of at least 14 gas-rich galaxies all lying at  $z = 4.31$  based on sensitive observations of carbon monoxide and ionized carbon. We demonstrate that each of these galaxies is forming stars between 50 and 1000 times faster than our own Milky Way, and all are located within a projected region only  $\sim 130$  kiloparsecs in diameter. This galaxy surface density is more than 10 times the average blank field value (integrated over all redshifts) and  $>1000$  times the average field volume density. The velocity dispersion ( $\sim 410 \text{ km s}^{-1}$ ) of these galaxies and enormous gas and star formation densities suggest that this system represents a galaxy cluster core at an advanced stage of formation when the Universe was only 1.4 billion years old. A comparison with other known protoclusters at high redshifts shows that SPT2349-56 is a uniquely massive and dense system that could be building one of the most massive structures in the Universe today.

In a multi-band survey over  $2500 \text{ deg}^2$  of sky, the South Pole Telescope (SPT) discovered a population of rare ( $n \sim 0.04 \text{ deg}^{-2}$ ), extremely bright ( $S_{1.4 \text{ mm}} > 20 \text{ mJy}$ ) millimeter-selected sources.<sup>13,14</sup> The Atacama Large Millimeter Array (ALMA)  $870\text{-}\mu\text{m}$  imaging showed that more than 90% of these SPT-selected sources are single high-redshift submillimeter galaxies (SMGs)<sup>15</sup> with intrinsic flux densities of  $S_{870 \mu\text{m}} = 5 - 10 \text{ mJy}$ , but gravitationally lensed by factors of  $5 - 20$ ,<sup>16</sup> with a median redshift  $z \sim 4$ .<sup>17</sup> However,  $\sim 10\%$  of these sources show no evidence for lensing and may instead be intrinsically very luminous galaxies or even groups of multiple rapidly star-forming galaxies. The brightest such source in the SPT  $2500 \text{ deg}^2$  survey, SPT2349-56 ( $S_{1.4 \text{ mm}}=23.3 \text{ mJy}$ ), is revealed by LABOCA (a low resolution bolometer camera on the APEX telescope) observations at  $870 \mu\text{m}$  to consist of two elongated sources with a combined flux density  $S_{870 \mu\text{m}} \sim 110 \text{ mJy}$  (Fig. 1), with the brighter southern source comprising  $\sim 77 \text{ mJy}$  of this flux density. An ALMA redshift survey<sup>17</sup> further resolved SPT2349-56 into a pair of bright 3-mm sources associated with the southern LABOCA source, with both lying at  $z = 4.3$ . To better understand the nature of this structure, deep ALMA spectral imaging of the brighter southern peak of the extended LABOCA source was undertaken. A 358-GHz map containing the redshifted  $[\text{CII}]_{1900.5 \text{ GHz}}$  line was used to search for line-emitting galaxies. A blind spectral line survey (described in the Methods) was performed on the data cube, revealing

14  $z \sim 4.31$  line emitters at high significance ( $\text{SNR} > 7$ ). Twelve of these emitters are individually detected in the 1.1-mm continuum map at  $> 5\sigma$ , with 1.1-mm flux densities ranging from 0.2-5 mJy (Fig. 1). The remaining two line emitters (M,N) are both detected at lower significance in the 1.1-mm continuum map but have robust IRAC infrared counterparts (Extended Data Table 1, Extended Data Fig. 4). Eight of these sources are also detected ( $> 5\sigma$ ) in the CO(4-3) line. The ALMA spectra are shown in Fig. 1.

The measurements of both the continuum and spectral lines of the 14 galaxies allow us to estimate their star formation rates (SFRs) and gas masses (Tables 1 & Extended Data Table 1). The physical properties of these sources indicate that this protocluster already harbors massive galaxies that are rapidly forming stars from an abundant gas supply. The two brightest sources, A & B, have SFRs in excess of 1000 solar masses per year ( $M_{\odot} \text{ yr}^{-1}$ ) within their resolved  $\sim 3$ -kpc radii. The total SFR of the 14 sources is  $6000 \pm 600 M_{\odot} \text{ yr}^{-1}$ . Multi-colour imaging with Herschel-SPIRE (250, 350, 500- $\mu\text{m}$ ), in addition to the 870- $\mu\text{m}$  LABOCA map, show that the northern LABOCA structure is also consistent with lying at  $z = 4.3$  (see Methods). The sources detected in the ALMA 870- $\mu\text{m}$  imaging therefore comprise just 50% of the total flux density of the southern LABOCA source and 36% of the total LABOCA flux density, suggesting that the  $\sim 500$  kpc extent of the protocluster contains  $16,500 M_{\odot} \text{ yr}^{-1}$  of star formation. Modelling the spectral energy distribution based on this combined submillimeter photometry yields an IR luminosity (from 8-1100  $\mu\text{m}$ ) of  $(8.0 \pm 1.0) \times 10^{13}$  times the solar luminosity ( $L_{\odot}$ ). The gas masses of the 14 protocluster galaxies, estimated from CO(4-3), or [CII] if undetected in CO(4-3) (see Methods), range from  $1 \times 10^{10}$  to  $1 \times 10^{11} M_{\odot}$ , with a total gas mass of  $\sim 6 \times 10^{11} (X_{\text{CO}}/0.8) M_{\odot}$ . A follow-up survey of colder molecular gas in CO(2-1) with the ATCA radio telescope detects the bulk of this large gas repository, especially in the central region near sources B, C, & G, and confirms that the assumed line intensity ratio, CO(4-3) to CO(1-0), used in the Methods when calculating the total gas mass is consistent with the average measurements from ATCA.

The detected ALMA sources also enables an initial estimate of the mass of the protocluster. We determine the mean redshift using the biweight estimator<sup>18</sup> as  $\langle z \rangle_{\text{bi}} = 4.3040_{-0.0019}^{+0.0020}$ . The velocity dispersion of the galaxy distribution is  $\sigma_{\text{bi}} = 408_{-56}^{+82} \text{ km s}^{-1}$  according to the biweight method,<sup>18</sup> which is the standard approach for galaxy samples of this size. Other common methods (gapper,<sup>18</sup> Gaussian fit) agree to within 3% and provide similar errors. Under the assumption that SPT2349-56 is approximately virialized, the mass-dispersion relation for galaxy clusters<sup>19</sup> indicates a dynamical mass of  $M_{\text{dyn}} = (1.16 \pm 0.70) \times 10^{13} M_{\odot}$ , which is an upper limit if the system has not yet virialized. Given the possible selection effect of requiring a bright source ( $S_{1.4\text{mm}} > 15 \text{ mJy}$ ) within the 1' SPT beam for detection, we also further consider the possibility that our structure may represent an end-on filament being projected into a compact but unbound configuration, rather than a single bound halo. Our analysis in the Methods suggests this is not as likely as a relatively bound system in a massive halo, given the velocity dispersion measured as a function of position, and other supporting arguments. However we cannot rule this possibility out completely, and further analysis and observations of the larger angular scale of the structure will be required to more fully understand the nature of this system.



If the total halo mass represented by these 14 SMGs is indeed  $\sim 10^{13} M_{\odot}$ , then the proto-cluster is a viable progenitor of a  $> 10^{15} M_{\odot}$  galaxy cluster comparable to the Coma cluster at  $z = 0$  (Fig. 2), which we deduce by comparison to simulations that track the hierarchical halo growth of a  $\sim 10^{15} M_{\odot}$  galaxy cluster from early epochs.<sup>12</sup> The location of SPT2349-56 in this plane suggests a very massive descendant, but we caution that N-body simulations indicate that it is difficult to reliably predict  $z = 0$  halo mass from the halo mass at a given epoch due to the large halo-to-halo variation in dark matter halo growth histories.<sup>20</sup>

To study the relative overdensity and concentration of SPT2349-56, it is desirable to compare with other active protoclusters at high redshift. SPT2349-56 is highly overdense, as it harbors 10 SMGs with  $S_{1.1\text{mm}} \gtrsim 0.5$  (a level at which we are complete, with uniform sensitivity across our search area) located within a circle of diameter  $19''$  (130 kpc), corresponding to a number density of  $N(S_{1.1\text{mm}} > 0.5 \text{ mJy}) \approx 2 \times 10^4 \text{ deg}^{-2}$ . By comparison, the average number of field sources with  $S_{1.1\text{mm}} > 0.5 \text{ mJy}$  within this area across all redshifts is less than one;<sup>21</sup> thus, this field is overdense by more than a factor of 10. When we account for the fact that all sources are at the same redshift, the volume density is  $> 1000\times$  the field density, assuming a redshift binning of  $\Delta z=0.1$  and the redshift distribution for SMGs.<sup>22</sup> In Fig. 2, we plot ‘curves of growth’ of the total 870- $\mu\text{m}$  flux density versus on-sky area for SPT2349-56 and other SMG-rich protoclusters (see Methods for the details of the comparison sample). For SPT2349-56, we plot both the total flux density of the 14 confirmed protocluster members detected with ALMA and the total flux density of the extended LABOCA structure. The curve of growth for SPT2349-56 rises much more steeply than those of the other high-redshift protoclusters, demonstrating its extreme density. For SPT2349-56 the on-sky area encompassing the accumulated 870- $\mu\text{m}$  flux density (and thus approximately the total SFR) is as much as 3 orders of magnitude less than for other protoclusters at  $z > 2$ . SPT2349-56 clearly stands out as the densest collection of SMGs: although some other protoclusters contain as many SMGs, they extend over much larger areas on the sky, with separations often exceeding 10 arcmin (22 comoving Mpc (cMpc) to 15 cMpc at  $z = 4.3$  to  $z = 2$ ). This comparison demonstrates that SPT2349-56 is likely observed during a significantly more advanced stage of cluster formation than other high-redshift protoclusters, a cluster core in the process of assembly rather than an extended structure that may not even collapse to form a cluster by the present day.<sup>12</sup>

Also shown in Fig. 2 is the maximal curve of growth predicted by a theoretical model for submm-luminous protocluster regions at  $z \sim 4.5$  (see Methods for details). Except for SPT2349-56 and the recent Herschel discovery SMM J004224,<sup>23</sup> the comparison high- $z$  protoclusters exhibit  $S_{870\mu\text{m}}$  curves of growth fairly consistent with the model expectations. The model prediction for the region spanned by SPT2349-56 is  $\sim 10\%$  of the observed total flux density of the 14 ALMA sources. The under-prediction is more severe if we consider the extended LABOCA source: only  $\sim 5\%$  of the observed flux density is recovered. This discrepancy may suggest that environmental effects (such as enhanced galaxy interactions or gas accretion in high-density environments) that are not included in the theoretical model employed are responsible for the extremely high SFR density exhibited by SPT2349-56. An alternative theoretical approach, ‘zoom’ hydrodynamical simulations of protoclusters,<sup>24</sup> can potentially

capture such environmental effects, but to date, such simulations have been unable to reproduce the extremely high SFR inferred for SPT2349-56: of the 24 protocluster simulations presented by these authors, the maximum total SFR attained was  $\sim 1700 M_{\odot} \text{ yr}^{-1}$ , an order of magnitude less than that of SPT2349-56. However, the volume of the N-body simulation from which the 24 halos were selected was  $1 h^{-3} \text{ cGpc}^3$ , which may be too small to contain an object as rare as SPT2349-56. Nevertheless, the existence of SPT2349-56, which contains an unprecedented concentration of rapidly star-forming SMGs when the Universe was only 1.4 Gyr old, poses a formidable challenge to theoretical models seeking to explain the origin and evolution of galaxy (proto)clusters.

SPT2349-56 may represent a significantly more advanced stage of cluster formation than the typical  $z > 4$  protoclusters identified to date, as outlined above. Since the cores of present-day galaxy clusters are characterized by massive elliptical galaxies with old-to-intermediate-age stellar populations<sup>25</sup> and SMGs are thought to be the high-redshift progenitors of present-day ellipticals,<sup>22</sup> it is likely that at least some of the 14 SMGs located at the same redshift within a region  $< 130 \text{ kpc}$  in diameter will soon merge to form a massive elliptical galaxy at the core of a lower-redshift galaxy cluster.

Theoretical studies have shown that at  $z > 4$ , the progenitors of galaxy clusters should span  $> 5$  comoving Mpc,<sup>12,26</sup> corresponding to an angular scale of as much as a degree; we are thus possibly observing only a small part of a much larger structure. For SPT2349-56, it is unknown whether the overdensity extends over such a large scale, as more detailed observations are required to characterize the field surrounding SPT2349-56. We have demonstrated that the extended LABOCA-detected complex has submm colours similar to the core region identified by our ALMA observations and is thus likely all at  $z \sim 4.3$ . We have also identified five additional bright SPIRE sources in the surrounding  $\sim 18 \times 18 \text{ cMpc}$  field with similar red colours lying several arcmin from the core structure (see Methods). These are candidates for additional protocluster members located in an extended, collapsing structure, similar to the comparison SMG overdensities shown in Fig. 2. If all these sources are confirmed to lie at  $z = 4.31$ , this would approximately double the far-IR luminosity of the cluster, making it by far the most active system known in the Universe. Since SPT2349-56 was selected from a blind mm survey of  $2500 \text{ deg}^2$  (approximately 1/16th of the sky), it is unlikely there are more than approximately 16 such structures across the entire sky. A full analysis of other unlensed sources from the SPT survey to identify possible systems similar to SPT2349-56 will place stronger constraints on early structure formation in the Universe.

### Acknowledgments

This paper makes use of the following ALMA data: ADS/JAO.ALMA#2016.0.00236.T and ADS/JAO.ALMA#2015.1.01543.T. ALMA is a partnership of ESO (representing its member states), NSF (USA) and NINS (Japan), together with NRC (Canada) and NSC and ASIAA (Taiwan), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO and NAOJ. This work is also based in part on observations made with the

Spitzer Space Telescope, which is operated by the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA. The SPT is supported by the National Science Foundation through grant PLR-1248097, with partial support through PHY-1125897, the Kavli Foundation and the Gordon and Betty Moore Foundation grant GBMF 947. This publication is based on data acquired with the Atacama Pathfinder Experiment (APEX) under programme IDs E-299.A-5045A-2017 and ID M-091.F-0031-2013. APEX is a collaboration between the Max-Planck-Institut für Radioastronomie, the European Southern Observatory, and the Onsala Space Observatory. Supporting observations were obtained at the Gemini Observatory, which is operated by the Association of Universities for Research in Astronomy, Inc., under a cooperative agreement with the NSF on behalf of the Gemini partnership: the National Science Foundation (United States), the National Research Council (Canada), CONICYT (Chile), Ministerio de Ciencia, Tecnología e Innovación Productiva (Argentina), and Ministério da Ciência, Tecnologia e Inovação (Brazil). The Australia Telescope Compact Array (ATCA) is part of the Australia Telescope National Facility which is funded by the Australian Government for operation as a National Facility managed by CSIRO. D.P.M., J.S.S., J.D.V., K.C.L., and S.J. acknowledge support from the U.S. National Science Foundation under grant AST-1312950. S.C.C., T.B.M., and A.B. acknowledge support from NSERC. S.C.C. and T.B.M. acknowledge CFI and the Killam trust. M.A. acknowledges partial support from FONDECYT through grant 114009. The Flatiron Institute is supported by the Simons Foundation. J.D.V. acknowledges support from an A.P. Sloan Foundation Fellowship.

**Author Contributions** T.B.M. led the data analysis, and assembled the paper. S.C.C. designed the study, proposed the ALMA observations, reimagined the data, and analyzed the data products. C.C.H. developed the theoretical model and advised on the literature comparison. M.A. led the ATCA follow-up and the blind emission line studies. A.W. procured and analyzed the deep LABOCA imaging. M.B. provided the cluster mass and evolution context and discussion. J.S. reimagined the calibrated data. K.A.P. performed the SED fitting. T.B.M., S.C.C., M.A., K.A.P. and A.W. made the figures. S.C.C., T.B.M., M.A., C.C.H., J.D.V., and A.W. wrote the manuscript. All authors discussed the results and provided comments on the paper. The authors are ordered alphabetically after A.W.

**Author Information** The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.B.M. at [tim.miller@yale.edu](mailto:tim.miller@yale.edu)

# Methods

## 1 Observations

### 1.1 SPT, LABOCA, and Herschel discovery and ALMA follow-up

The South Pole Telescope<sup>27</sup> (SPT) possesses a unique combination of sensitivity, selection wavelengths (3, 2, and 1.4 mm), and beam size that potentially make it ideal for finding the active core regions of galaxy clusters forming at the earliest epochs. Finding very distant ( $z > 4$ ), gravitationally lensed millimetre sources in the SPT survey is relatively straightforward, where the contrast to such distant bright sources is high relative to the weak (generally undetected) galactic foregrounds (Extended Data Figure 1). However searching for the rare SMGs in the SPT 2500 deg<sup>2</sup> survey that are unlensed, and therefore candidates for active groups and protoclusters like SPT2349-56, involves sifting through the many gravitationally lensed sources, and typically involves multi-stage follow-up efforts using various facilities: a single dish mapping instrument like APEX-LABOCA to better localize the emission within the  $\sim 1'$  SPT beam, deep optical imaging to search for bright lensing galaxies, and high resolution ALMA mapping. The spatially extended sources in SPT2349-56 found with LABOCA span more than an arcmin. With deep upcoming surveys using the next generation SPT-3G receiver, this ‘extended-beam’ thermal source structure may present a unique signature of many early forming protoclusters, affording the first complete census in the early epochs of structure formation.

A shallow, wide field SPIRE image over a 100 deg<sup>2</sup> subregion of SPT-SZ<sup>28</sup> reveals the red colours of SPT2349-56, and that SPT2349-56 appears to reside in something of a void in the  $z \sim 1$  foreground that dominates the SPIRE galaxy population. However the high redshift of SPT2349-56 means that it is not significantly brighter than many other SPIRE sources in this field, and aside from its colours, SPT2349-56 does not stand out substantially from the field despite its extreme properties. SPT2349-56 is not detected in the all sky *Planck* survey,<sup>29</sup> the lower sensitivity of *Planck* compared with SPT being exacerbated by beam dilution in the 3' *Planck* beam.

Obtaining the redshift for SPT2349-56 was beyond the scope of the original SPT-SMG redshift survey, due to the faintness of the unlensed components relative to the typical bright, gravitationally lensed SMGs found in the bulk of the SPT-SMG sample. In ALMA Cy 0 & 1, SPT2349-56 was included in the 3-mm spectral scan redshift survey,<sup>15,30</sup> but no lines were detected in the short  $\sim 1$  min integrations with  $\sim 16$  ALMA antennae. In Cy 3, a deeper follow-up 3-mm spectral scan was able to tentatively identify two CO lines and a double source structure with a likely redshift  $z = 4.30$ , confirmed by APEX/FLASH C+ detection.<sup>17</sup>

### 1.1.1 APEX - LABOCA

We obtained 870- $\mu\text{m}$  imaging of SPT2349-56 using LABOCA on the APEX telescope. A shallow image with 1.6hr integration time was observed on 27 Sep 2010 reaching  $\sim 5$  mJy/beam rms. In August 2017 we obtained a deeper image (18.8h integration time, Project ID: E-299.A-5045A-2017, PI: Chapman) reaching a minimum noise level of 1.3 mJy/beam and  $< 2.0$  and  $< 1.5$  mJy/beam rms for 75.3 and 32.4 sq arcmin, respectively (shown in Figure 1 & Extended Data Figure 2). All observations were carried out using standard raster-spiral observations<sup>31</sup> under good weather conditions (PWV of 0.6 mm and 0.8 mm for the 2010 and 2017 observing campaigns, respectively). Calibration was achieved through observations of Uranus, Neptune and secondary calibrators and was found to be accurate within 8.5% rms. The atmospheric attenuation was determined via skydips every 2hr as well as from independent data from the APEX radiometer which measures the line of sight water vapor column every minute. The data was reduced and imaged using the BoA reduction package.<sup>32</sup> LABOCA's central frequency and beam size are 345 GHz and 19.2'', resolving the SPT 1.4-mm elongated source into two bright LABOCA sources.

Both LABOCA observations yield consistent calibration results with peak intensity at 21'' resolution of 50 mJy/beam for the brighter, southern component (RA 23:49:42.70, DEC -56:38:23.4). In addition the LABOCA map reveals a second source to the north at RA: 23:49:42.86, DEC: -56:37:31.02 with a peak flux density at 21'' resolution of 17 mJy/beam. Both sources are clearly extended even at LABOCA's relatively coarse spatial resolution with observed source size of 28'' $\times$ 25'' and 31'' $\times$ 24'' for the southern and northern source, respectively. These components are connected by a faint bridge emission. The total 870- $\mu\text{m}$  flux density of the SPT2349-56 system is  $110.0 \pm 9.5$  mJy, of which  $\sim 77$  mJy are associated with the southern component,  $\sim 25$  mJy with the northern component, and  $\sim 7$  mJy with the connection between the components (using the sub apertures shown in Extended Data Figure 2). One additional submm source is detected at  $> 5\sigma$  in the LABOCA image to the east of the primary source, but having blue colours inconsistent with  $z \sim 4$ , and not likely being a member of the extended protocluster.

## 1.2 ALMA

Observations using ALMA Band-3 targeted the CO(4-3) line in SPT2349-56 centred in the lowest frequency of the spectral windows adopted (86-88 GHz), taken under a Cycle 3 program 2015.1.01543.T (PI: K. Lacaille). Data was taken on June 24th, 2016 with a 47 min integration time. The array used 36 antennas with baselines ranging from 15 to 704 m, and provided a naturally weighted synthesized beam size of  $\sim 1''$ . Pallas and J2343-5626 were used to calibrate the flux and phase respectively. Data was processed using the standard ALMA pipeline using natural beam weighting.

ALMA Band-7 imaging (276 GHz) were obtained under a Cycle 4 program (2016.0.00236.T; PI: S. Chapman) targeting the peak of the brightest LABOCA source. Observations were obtained on December 14th, 2016 in a 40-2 array configuration with baseline lengths of 15-459 m,

giving a naturally weighted synthesized beam size of  $\sim 1''$ . There were 40 antennas available, with total on source integration time of 22 minutes. Ceres and J2357-5311 were used as flux and phase calibrators respectively. The [CII] line ( $\nu_{rest} = 1900.5$  GHz) was observed as part of the same ALMA project on March 23rd, 2017, tuning in Band 7 to the redshifted line at  $\nu_{obs} = 358.3$  GHz in the upper sideband covering 356 to 360 GHz. These observations used the 40-2 array configuration with baselines of 16-459 m, giving a naturally weighted synthesized beam size of  $\sim 0.5''$ . An on-source integration time of 14 min was obtained, and J2357-5311 was used as both the flux and phase calibrator. The data were re-processed using CASA and the standard ALMA-supplied calibration using natural beam weighting to maximize sensitivity.

One dimensional spectra are extracted from the centroid of the line emission for each source and binned into  $75 \text{ km s}^{-1}$  channels. Spectra are presented in Figure 1, and are smoothed using a Gaussian filter with  $\text{FWHM} = 100 \text{ km s}^{-1}$  for presentation. A Gaussian line profile is fit using a least-squares method, providing errors to the velocity offsets from  $z = 4.304$  in Table 1 and line widths in Extended Data Table 2. The continuum level is left as a free parameter in the fitting function which is then subtracted to derive line fluxes and for presentation.

### 1.2.1 Blind search for [CII]

We performed a blind search for [CII] line emission in the ALMA band 7 data cube toward SPT2349-56. For this, we follow the procedure used to detect line emitters in the ASPECS survey.<sup>33</sup> We use a data cube channelized at  $100 \text{ km/s}$ , without primary beam correction and continuum subtraction. We used the Astronomical Image Processing System (AIPS) task SERCH. This task convolves the data cube along the frequency axis with a Gaussian kernel defined by different input linewidths, subtracts surrounding continuum, and reports all channels and pixels that have a signal-to-noise ratio (SNR) over a specified limit. The SNR is defined as the maximum significance level achieved after convolving over the Gaussian kernels. We used a set of different Gaussian kernels, from  $200$  to  $600 \text{ km/s}$  and searched for all line peaks with  $\text{SNR} > 4.0$ .

Once all peaks were identified, we used the IDL routine CLUMPFIND<sup>34</sup> to isolate individual candidates. A full list of 68 positive line peaks with  $\text{SNR} > 4.0$  were thus obtained. We quantified the reliability of our line search based on the number of negative peaks in our ALMA cube, using the same line procedure. We find 43 negative peaks with  $\text{SNR} < 5.8$  and none at a higher SNR. This means that all positive line candidates with  $\text{SNR} > 6.0$  are likely real (100% purity). Out of the 14 [CII] line candidates detected, all have  $\text{SNR} > 6.3$  and 12 are associated with continuum detections in the ALMA data.

## 1.3 ATCA CO(2-1)

### 1.3.1 Observations

We used the Australia Telescope Compact Array (ATCA) in its H168 hybrid array configuration to observe the CO(2-1) emission line ( $\nu_{rest} = 230.5380$  GHz) toward SPT2349-56 (with a primary beam size of  $53''$ ). The observations were performed as part of project ID C2818 during

2016 October 2,3 and 11 under good weather conditions (atmospheric seeing values 90-400 m) with five working antennas.

We used the ATCA 7-mm receivers, with the Compact Array Broadband Backend configured in the wide bandwidth mode.<sup>35</sup> This leads to a total bandwidth of 2 GHz per correlator window and a spectral resolution of 1 MHz per channel (6.9 km/s per channel). The spectral windows were centred at observing frequencies of 43.5 and 45.0 GHz, and aimed at observing the CO line and continuum emission, respectively.

Gain and pointing calibration were performed every 10 min and 1 h, respectively. The bright sources 1921-293, 1934-638 and 2355-534 were used as bandpass, flux and gain calibrators, respectively. We expect the flux calibration to be accurate to within 15 per cent, based on the comparison of the Uranus and 1934-638 fluxes. The software package MIRIAD<sup>36</sup> and the Common Astronomy Software Applications (CASA<sup>37</sup>) were used for editing, calibration and imaging.

The calibrated visibilities were inverted using the CASA task CLEAN using natural weighting. No cleaning was applied given the relatively low significance of the CO line detection in individual channels. The final data cube, averaged along the spectral axis, yields an rms of 0.23 mJy beam<sup>-1</sup> per 100 km/s channel with a synthesized beam size of 5.6'' × 4.5'' (PA=70.4 deg) at 43.5 GHz.

### 1.3.2 Results

One source formally detected at the centre, which corresponds to CII/continuum sources B+C+G. This central CO source (C) is unresolved at the resolution of the ATCA observations. Other two sources are marginally detected to the West (w) and North (N) of the central source, coinciding with the location of CII/continuum sources D+E and A+K, respectively. We extracted spectra at these locations and obtained integrated line intensities, by fitting Gaussian profiles to the identified line emission.

We compute CO luminosities using the integrated line intensities and compute gas masses by assuming a ULIRG  $X_{CO}$  factor of 0.8 ( $M_{\odot} (\text{K km/s pc}^2)^{-1}$ ) and that the CO gas is in local thermodynamic equilibrium thus  $L'_{(CO2-1)} \sim L'_{(CO1-0)}$ .<sup>38</sup> The results of the CO line observations are summarized in Extended Data Table 2. Collapsing the line-free spectral window along the spectral axis over the 2-GHz bandwidth, leads to a non-detection of the continuum emission down to 80  $\mu\text{Jy/beam}$  ( $3\sigma$ ).

These results confirm the finding from CO(4-3) line that that the main reservoir (72%) of molecular gas resides in the B+C+G system, with a smaller fraction hosted at the West and North locations.

## 1.4 Spitzer imaging

This field was twice observed at 3.6 and 4.5  $\mu\text{m}$  with the Infrared Array Camera (IRAC<sup>39</sup>) on board the *Spitzer Space Telescope*.<sup>40</sup> It was first observed in 2009 August as part of a large

program to obtain follow-up imaging of a large sample of SPT-selected SMGs sources (PID 60194, PI Vieira). The observing scheme used for PID 60194 was to obtain 36 dithered 100 sec integrations at  $3.6 \mu\text{m}$  and, separately, a much shallower  $12 \times 30$  sec integration at  $4.5 \mu\text{m}$ . Later, in Cycle 8, the field was covered serendipitously as part of the *Spitzer*-SPT Deep Field survey (PID 80032, PI Stanford; Ashby et al. 2013). PID 80032 surveyed  $92 \text{ deg}^2$  uniformly in both IRAC passbands to a depth of  $4 \times 30$  sec. Using established techniques, we combined all exposures covering the SPT target from PID 60194 and 80032 at  $3.6$  and  $4.6 \mu\text{m}$  to obtain the best possible S/N in our final mosaics, which were pixellated to  $0.5''$ . Nine of the 14 sources identified by ALMA are detected in the IRAC bands at  $> 3\sigma$  in at least one of the  $3.6$  or  $4.5 \mu\text{m}$  channels, as shown in Extended Data Figure 4.

## 1.5 Analysis of the surrounding field with SPIRE and LABOCA imaging

In Extended Data Figure 5, our deep SPIRE RGB image is shown with LABOCA contours overlaid. A source sample is culled from the  $250 \mu\text{m}$ -selected catalog (135 sources with  $\text{SNR}(250 \mu\text{m}) > 3$  in an area of  $52 \text{ arcmin}^2$ ), where the source peaks are best defined. To account for the large beam size difference with SPIRE (ranging from  $36''$  at  $500 \mu\text{m}$  to  $18''$  at  $250 \mu\text{m}$ ), we employed a deblending code, using the  $250 \mu\text{m}$  positions as spatial priors, which provides the standard parameters as well as the covariance matrices highlighting the degeneracies (almost none at  $250 \mu\text{m}$ , but significant at  $500 \mu\text{m}$ ). The code, FASTPHOT,<sup>41</sup> takes into account these degeneracies to estimate the flux measurement errors.

Colour-colour (CC) and colour-flux (CF) diagrams are shown in Extended Data Figure 5. The CC diagram shows a  $250 \mu\text{m}$ -selected sample with  $\text{SNR}(250 \mu\text{m}) > 3$  and is dominated by the  $z \sim 1$  cosmic infrared background (blue, green colours) in the *foreground* of SPT2349-56. The CF diagram shows an additional  $\text{SNR}(500 \mu\text{m}) > 3$  cut to highlight just the well detected  $500 \mu\text{m}$  source sub-sample. These diagrams highlight the extreme and red properties of SPT2349-56, but make clear that one of the three  $250 \mu\text{m}$ -peaks within the SPT2349-56 LABOCA structure is very likely a foreground galaxy (green symbol highlighted in the figure shows very blue colours). Nevertheless, a full ALMA mapping of the structure is warranted given the uncertainties involved in the SPIRE deconvolution procedure.

Five red sources consistent with  $z \sim 4$  ( $S_{500 \mu\text{m}} > S_{350 \mu\text{m}} > S_{250 \mu\text{m}}$ ) are found in the surrounding  $\sim 10' \times 10'$  field and are candidates for additional protocluster members in an extended, collapsing structure. If all these sources were bona fide  $z = 4.3$  sources, this would significantly increase the total  $870\text{-}\mu\text{m}$  flux density (and thus the far-IR luminosity) of the cluster beyond the  $110 \text{ mJy}$  found in the central structure, making it by far the most active system known in the Universe (see Figure 2). The deep LABOCA map marginally detects the closest of the five red SPIRE sources at  $\sim 3\sigma$ , consistent with expectations given the SPIRE flux densities. Full analysis of these surrounding SMGs will require additional follow-up efforts.



## 2 Properties, Comparisons, Simulations

### 2.1 Derivation of physical properties

We briefly describe our procedures for calculating various physical quantities from observables below. To derive SFR, we measure 870- $\mu\text{m}$  flux density directly in the lower sideband (line-free bands) of our ALMA Band-7 observations from Cycle 4, finding consistent measurements with those found in previous shallower observations.<sup>16</sup> We adopt an SFR-to- $S_{870\mu\text{m}}$  ratio of  $150 \pm 50 \text{ M}_{\odot} \text{ yr}^{-1}/\text{mJy}$ , which is typical for SMGs.<sup>42</sup> The uncertainty in this ratio owes to variations in the dust temperature distribution amongst the SMG population, which are primarily driven by differences in the ratio of the luminosity absorbed by dust to the total dust mass.<sup>43</sup> This combined with the measurement error dictates the error on the SFR shown in Table 1

Gas mass is calculated from the CO(4-3) line luminosity, which is converted to CO(1-0) luminosity using a ratio between the brightness temperatures of these lines  $r_{4,1} = 0.41 \pm 0.07$  found from the average of a sample of unlensed SMGs with multiple CO line transitions detected.<sup>44</sup> We use a conservative conversion factor  $\alpha_{CO} = 0.8 \frac{\text{M}_{\odot}}{\text{K km s}^{-1} \text{ pc}^2}$  and multiply by 1.36 to account for the addition of helium. When CO(4-3) is not significantly detected, we use our [CII] line luminosity and the average CO(4-3)/[CII] ratio for our detected sample; we denote these sources with asterisks in Table 1.

### 2.2 Spectral energy distribution of SPT2349-56

The SPT, LABOCA and SPIRE measurements resolve the SPT2349-56 structure to varying degrees, but none can isolate the core region resolved by our current ALMA observations with any confidence. We thus assemble a photometric catalog of the total SPT2349-56 flux density from 250  $\mu\text{m}$  to 850  $\mu\text{m}$  and model the resulting total SED to estimate some global properties of the system. Although SPT does measure the flux at 1.4, 2.0 and 3.0 mm ( $27.9 \pm 4.6$  mJy,  $5.2 \pm 1.1$  mJy and  $0.5 \pm 0.1$  mJy, respectively), we do not include these points in the SED fit since the measurements are uncertain due to the elongated structure of SPT2349-56 and difficulty with the filtering used to make the map. At IRAC wavelengths in the mid-infrared, we identify 9 SMGs detected at  $> 3\sigma$ , and include the sum in the SED fit. We defer further analysis to followup work with forthcoming deeper *Spitzer*-IRAC data and follow up optical and NIR photometry.

We have used Code Investigating GALaxy Emission (CIGALE)<sup>45,46</sup> for the SED fitting of the combined photometry of the source. The SED modelling assumes a single-component star formation history and solar metallicity.<sup>47</sup> A Chabrier<sup>48</sup> IMF is assumed. The resulting best-fitting SED is shown in Extended Data Figure 6. The IR luminosity (from 8–1100  $\mu\text{m}$ ) is  $(8.0 \pm 1.0) \times 10^{13} \text{ L}_{\odot}$ .

### 2.3 Protocluster comparison sample

To place SPT2349-56 in context and compare to other systems claimed to be protoclusters, we assemble from the literature various SMG-rich overdensities at  $2 < z < 5$ . Although a direct comparison of the number counts (number/deg<sup>2</sup>) of SMG-overdense systems can be performed, it involves making somewhat arbitrary choices of enclosed areas and redshift boundaries. We have opted in Figure 2 to instead show a curve of growth analysis of the 870- $\mu\text{m}$  flux density. Only galaxies confirmed to be protocluster members via spectroscopic redshifts are considered. The data are drawn from a recent compilation<sup>49</sup> and original references therein.

The GOODS-N overdensity at  $z = 1.99^{9,50,51}$  spans a  $\sim 10'$  by  $10'$  field in the Hubble Deep Field North containing 9 SMGs in  $\Delta z = 0.008$ . The probability of finding this large of an overdensity being drawn from the field distribution by chance is  $< 0.01\%$ . Interestingly, only a modest overdensity of Lyman-break galaxies is found in this GOODS-N structure.

The COSMOS  $z = 2.5$  SMG overdensity<sup>8</sup> is similar to the GOODS-N structure in terms of the numbers and luminosities of the component SMGs, the angular size of the system, and the modest overdensity of LBGs associated with it.

The MRC1138 overdensity was originally discovered as an overdensity of Ly $\alpha$  and H $\alpha$  emitters.<sup>52</sup> Follow-up observations<sup>53,54</sup> revealed the presence of 5 SMGs, in addition to an AGN known as the ‘Spiderweb galaxy’. This is a radio-loud AGN that resides in a large Ly- $\alpha$  halo.

The SSA22 protocluster was one of the first discovered by observing an overdensity of LBGs.<sup>55</sup> It is an extremely extended structure located at  $z = 3.09$ , with LAEs spanning greater than 50 comoving Mpc (cMpc).<sup>56</sup> Submm observations of the field have revealed a population of at least eight SMGs<sup>10,50,57-59</sup>.

The COSMOS  $z = 2.1$  protocluster<sup>60</sup> lacks sufficiently deep 850- $\mu\text{m}$  data to characterize the *Herschel*-SPIRE sources identified in the structure. We estimate 870- $\mu\text{m}$  flux densities by taking their published  $L_{\text{IR}}$  (integrated over 3–1100  $\mu\text{m}$ ) and use the SED of Arp 220 to estimate  $S_{870\mu\text{m}}$ , finding that  $L_{\text{IR}} = 2 \times 10^{12} L_{\odot}$  corresponds to  $S_{870\mu\text{m}} = 1 \text{ mJy}$  at  $z \sim 2$ . For the SSA22 protocluster, we use the measured 870- $\mu\text{m}$  flux density when available and otherwise estimate it from the 1.1-mm flux using a standard conversion at  $z \sim 3$  of  $S_{870\mu\text{m}} = 2 \times S_{1.1\text{mm}}$ . To create the curves of growth for Figure 2, the centre of each protocluster is defined by computing the median RA and DEC of all submm sources. We checked that adjusting the centres of the curve of growth tracks randomly by  $\sim 1'$  did not boost the curves by more than 10%, demonstrating that the curves of growth for the literature SMG overdensities are insensitive to the adopted centre.

Recently, there have also been detections of SMG overdensities at  $z > 4$ . The first, GN20, at  $z = 4.05$ , was discovered through the serendipitous detection of CO(4-3) from two SMGs,<sup>61</sup> with two further SMGs detected subsequently.<sup>62</sup> An excess of *B*-band dropouts is also observed in this structure, several of which are spectroscopically confirmed to lie at  $z \sim 4.05$ . HDF850.1 contains a single SMG, a QSO, and 11 spectroscopically confirmed galaxies. The SMG has a confirmed redshift of  $z = 5.18$ .<sup>63</sup> The *AzTEC-3* overdensity is centred on a single SMG at

$z = 5.3$ , with 12 spectroscopically confirmed optical galaxies at the same redshift. This is a relatively dense structure, with most of the galaxies residing within a circle  $\sim 1'$  in diameter. The most luminous example at  $z > 4$ , SMMJ004224, was recently found from the *Herschel* surveys,<sup>23</sup> with several additional  $870\mu\text{m}$  sources in the surrounding field. In Fig. 2 we plot, as an extension, the flux from the satellites which have observed SPIRE colors consistent with  $z = 4$ .<sup>64</sup>

Overdensities of SMGs and optical galaxies have also been found around high-redshift radio galaxies (HzRGs),<sup>65</sup> continuing to confirm HzRGs as useful beacons of structure forming in the early Universe. However none of these systems come close to the level of overdensity found in SPT2349-56, and furthermore, they suffer from the bias inherent in targeting these sources, namely, that one or more protocluster members have to be radio-luminous.

There have also been discoveries of compact binary HyLIRG systems, the most luminous of which is the  $z = 2.4$  source HATLAS J084933,<sup>66</sup> with others approaching this luminosity.<sup>67,68</sup> In each of these systems, the dynamics and SFRs are dominated by two SMGs, but there is no strong evidence of any surrounding protocluster in the form of an excess of galaxies selected optically or in the submm. In one case,<sup>68</sup> there is evidence for a relative void around the structure. These systems may simply be instances of very rare events in fairly typical (but still massive) halos,<sup>69</sup> analogous to hyper-luminous quasars.<sup>70</sup>

Theoretical studies of N-body simulations have shown that the progenitors of  $z = 0$  dark matter halos with masses  $> 10^{15} M_{\odot}$  should extend to effective radii of  $\gtrsim 5$  cMpc at  $z > 2$ .<sup>12,26</sup> Since the overdensities listed above are mostly concentrated in small areas, it is difficult to assess their exact evolution or compare them easily to simulated structures. Interpreting a small overdense region at high-redshift as a ‘protocluster core’ is certainly prone to misinterpretation, and small overdense regions at high redshift can evolve into halos spanning a range of masses at the current epoch.<sup>12</sup> These authors suggest investigating if an overdensity extends to larger scales ( $> 20$  cMpc) to better determine whether it will form an  $M \gtrsim 10^{15} M_{\odot}$  cluster. However this is difficult at high redshift because the excess of galaxies will be less pronounced on larger scales, and it is challenging to detect high-redshift, low-luminosity galaxies.

To assess the evolution of SPT2349-56 we also compare to a selection of the highest mass galaxy clusters at lower redshift ( $z \lesssim 2.5$ ). XMMU J2235.3<sup>71</sup> and XLSSC 122<sup>2</sup> were both discovered in x-ray surveys with masses derived from the x-ray light profile. IDCS J1426.5+3508,<sup>3,72</sup> JKCS5041,<sup>73</sup> CL J1449+0856<sup>74,75</sup> and CL J1001+0220<sup>1</sup> were discovered as overdensities of massive red galaxies and have corresponding x-ray detections. Masses for all are estimated using these x-ray detections and consistent with dynamical masses estimated using the velocity dispersion. We also show the sample of clusters discovered in the 2500 square degree SPT survey using the Sunyaev-Zel’dovich effect, with mass estimate based on the velocity dispersions.<sup>76</sup>

## 2.4 Geometry and Dynamics

In Extended Data Figure 7 we investigate the geometry and kinematics of the SPT2349-56 system. By analysing  $\Delta V$  vs. projected radius for our sample of 14 galaxies it is clear that, even with our most conservative mass estimate, at least 12 of the 14 galaxies appear bound. We show the escape velocity as a function of radius for a point source and NFW profile with a mass of  $1.16 \times 10^{13} M_{\odot}$ . The NFW profile assumes a virial radius of 200 kpc and concentration of 5, typical values of massive halos at this epoch, found in N-body simulations described in § 2.5. Even if the projected radius only accounts for a fraction of the true physical distance (i.e. separated along the line of sight) it is likely that most of the galaxies with low relative velocity ( $\Delta V < 500 \text{ km s}^{-1}$ ) will still be bound and eventually collapse into a single galaxy. The cumulative distribution of relative velocities is also shown, alongside that of a Gaussian distribution with  $\sigma = 408 \text{ km s}^{-1}$ . The observed distribution is smooth and well fit by a Gaussian. The relative kinematics of the 14 galaxies is consistent with most, if not all, members being mutually bound with SPT2349-56 following a Gaussian  $\Delta V$  distribution characteristic of a virialized system.

An alternative interpretation of this system is that we are observing individual galaxies, or separate less massive halos along a filament aligned with our line of sight. The probability of this occurring is potentially boosted by our selection technique, since the beam size of SPT is approximately  $1'$ . We are preferentially sensitive to structure on this scale, and require several of the most luminous galaxies found in the Universe together in the SPT beam to exceed our detection threshold.

To explore this possibility, we also show in Extended Data Figure 7 the most extreme interpretation of SPT2349-56 where none of the velocity offsets are peculiar, and the physical distribution of the SPT2349-56 galaxies is represented by the relative velocities being entirely due to cosmic expansion rather than peculiar motions. In this case the galaxies of course are maximally separated along a (proper distance) 3Mpc *filament* compared to their 130kpc maximum tangential extent. However, this cannot be the true distribution, and a more realistic interpretation would entail some  $\sim 20\text{-}30\%$  of velocities being Hubble flow (600-900pc extent), and the remainder representing peculiar motions. (Note that this would be true for any massive halo at this epoch given the expected velocity dispersion of  $\sim 400 \text{ km s}^{-1}$  and Hubble constant at this epoch of  $470 \text{ km s}^{-1} \text{ Mpc}^{-1}$ .) This explanation, if substantiated, would alter our interpretation of the system, specifically decreasing the mass estimate as the velocity offsets are not due to peculiar motions and the system is not virialized.

However, given the observed kinematics and spatial configuration of the galaxies discussed above, we argue that SPT2349-56 is less likely to consist of multiple groups, widely separated on a line of sight filament. To further back up this claim, we perform a systemic search for similar filamentary structures in the N-body simulations described in § 2.5. We first search for groups at  $z > 4$  with more than 8 galaxies that have a projected radius of  $< 150 \text{ kpc}$  but can be extended up to 5 Mpc along the line of sight. There are three such systems found in our  $1 \text{ cGpc}^3$  simulation, and they are also displayed in E.D. Fig. 7 showing their geometry. All three

of these analogous systems in the simulation have total halo masses of  $\sim 10^{13}M_{\odot}$ . None are extended significantly past 500 kpc, implying that at  $z \sim 4$ , there are no substantially extended *filaments* hosting massive galaxies comparable to even the gas masses our SPT2349-56 ALMA galaxies in the simulation. These simulated structures that we have found are not filaments – they are close to collapsed structures which are slightly cigar-shaped, and while SPT2349-56 could in principle be distributed like this, it does not fundamentally change our discussion in the paper. (We also note that we are plotting 2 different things in E.D. Fig. 7 velocity offsets for 2349 and actual geometry for the simulation galaxies.) These results further suggest that most of the velocity offsets in SPT2349-56 galaxies are truly mostly due to peculiar motion rather than line of sight projection. At least one of the simulation systems found appears to be chance projections of multiple groups, but it has a characteristically different  $\Delta V$  distribution. The system is found to have multiple smaller groups with  $\sigma \sim 200 \text{ km s}^{-1}$ , reflecting their smaller masses, separated in velocity space by  $500 \text{ km s}^{-1}$ .

We also note that a filamentary structure feeding a  $\sim 10^{13} M_{\odot}$  halo should have a characteristic width of 200 kpc - 400 kpc<sup>77,78</sup> compared to the 80 kpc (FWHM) width we find for our 14 ALMA galaxies. This is suggestive that even with a direct line-of-sight view down a filament, the configuration is far more concentrated than one would infer based on the typical size of filament. We then ask what sort of environment a  $\sim 80\text{kpc}$  wide filament (consistent with our SPT2349-56 galaxies) would connect, and infer a typical halo mass of  $\sim 2 \times 10^{12}M_{\odot}$  halo<sup>77</sup>. Given that the SPT2349-56 galaxies contain at least this much mass entirely in their cold gas (as we have adopted the lowest plausible  $\alpha_{CO} = 0.8$  conversion factor), and are likely hosted by haloes at least twice as large as this, it becomes somewhat contradictory that they could be found along a filament that is connecting a significantly less massive halo.

A further important piece of evidence against these 14 galaxies being a filament is that we find no additional sources in the surrounding spectral windows in either side band, both in the band 7 and band 3 ALMA data (Extended Data Figure 7 - d). While all 14 sources are clustered within  $1500 \text{ km s}^{-1}$  of each other, our full frequency coverage extends over relative velocities of  $6500 \text{ km s}^{-1}$  (for [CII]) and  $27,000 \text{ km s}^{-1}$  (for CO(4-3)), or many tens of Mpc line of sight distance. If SPT2349-56 was actually an elongated filament, one might expect to see a distribution of sources over a much larger fraction of our observed frequency bandwidth. Although we can not rule out the possibility that we are observing multiple galaxies along our line of sight, the observed geometry and kinematics suggest SPT2349-56 consists of a single group of mutually gravitationally bound galaxies.

Another way to assess the configuration of this system is by employing the Millennium Simulation<sup>4</sup> database, using the implementation of the<sup>79</sup> galaxy-formation recipe. We search the model output at  $z \sim 4$  for galaxies with a total baryonic mass in excess of  $\sim 10^{11} M_{\odot}$ , consistent with the combined mass of gas and stars estimated for the brightest ALMA-resolved galaxies in SPT2349-56. We use this total baryonic mass cut as it means we are less sensitive to the details of the early star-formation histories of galaxies in the model. We find only one galaxy in the  $3.2 \times 10^8$  comoving-Mpc<sup>3</sup> simulation volume at  $z = 4$  with a total baryonic mass above  $10^{11} M_{\odot}$ . Half of the baryonic mass in this model galaxy is in stars and it has a  $3 \times 10^8 M_{\odot}$  black

hole. Moreover this galaxy is the central galaxy of a  $1 \times 10^{13} M_{\odot}$  halo, the optimal environment to find merging galaxies according to simulations.<sup>80</sup> Searching the environment of this system we find another four massive galaxies are distributed across a  $\sim 0.5$ -comoving-Mpc-diameter region around the central galaxy, with baryonic masses of  $> 0.7 \times 10^{10} M_{\odot}$  ( $> 5\%$  of the primary). By redshift zero these galaxies are all predicted to reside in a  $1 \times 10^{15} M_{\odot}$  halo, consistent with our other assessments of the outcome of this system.

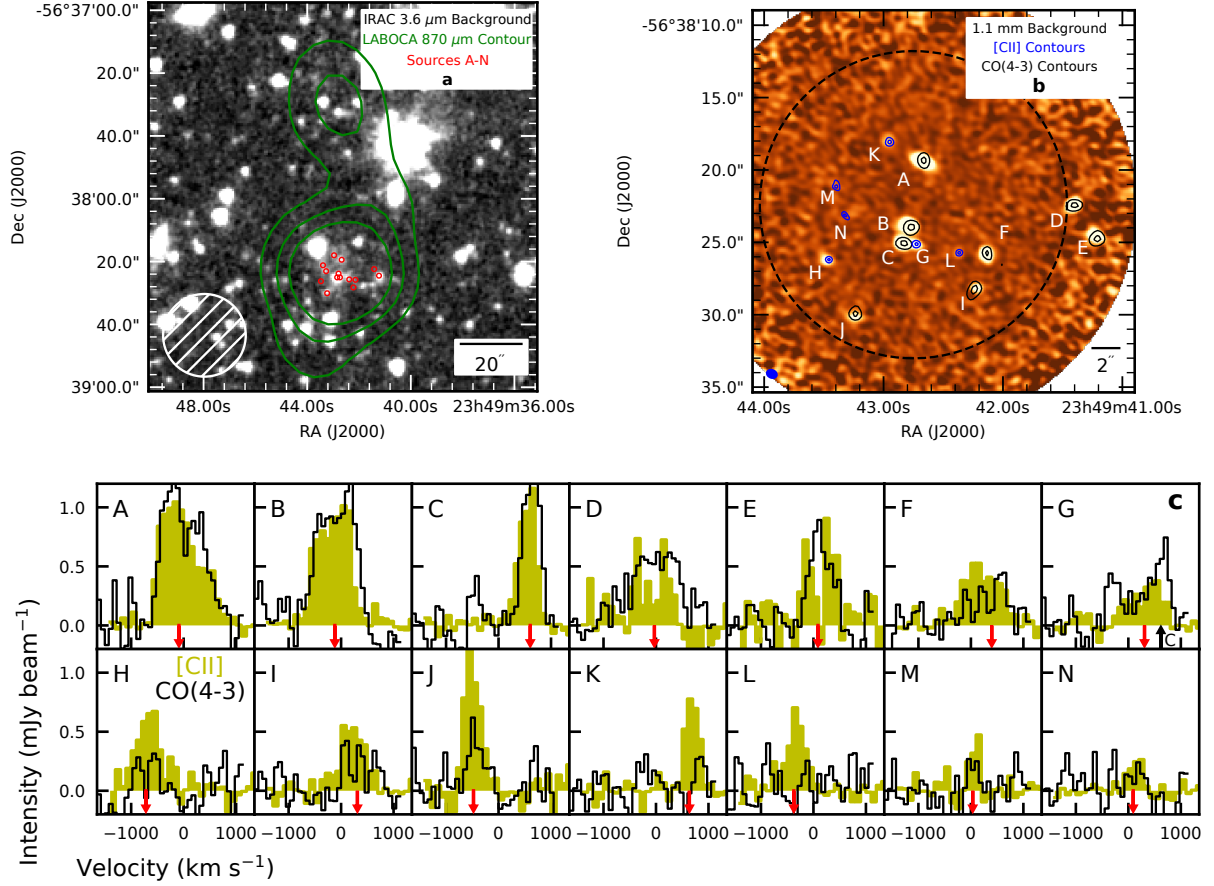
## 2.5 Simulations

To further place SPT2349-56 in context, we compare with the predictions of a theoretical model for SMG overdensities.<sup>69,81</sup> We employ the MultiDark<sup>82</sup> N-body simulation, which is one of the largest ( $2.91 \text{ Gpc}^3$ ) available N-body simulations that still resolves SMG-like halos ( $M_{\text{halo}} \gtrsim 10^{12} M_{\odot}$ ). The  $z = 4.68$  and  $z = 4.1$  snapshots, which are the available snapshots closest in redshift to SPT2349-56, are analyzed. Halo catalogs were created using the Rockstar halo finder,<sup>83</sup> and stellar masses are assigned to dark matter halos using a relation derived based on sub-halo abundance matching relation.<sup>84</sup> To assign SFRs, it is assumed that the distribution of specific SFR (SFR per unit stellar mass, hereafter SSFR) is the sum of two Gaussians, corresponding to quiescently star-forming and starburst galaxies.<sup>85</sup> The median SSFR value is based on the abundance-matching-derived relation,<sup>84</sup> and the starburst fraction and the widths of the Gaussian distributions are set based on observations of massive, high-redshift star-forming galaxies similar to the members of SPT2349-56.<sup>85</sup>  $M_d$  is estimated from stellar mass using empirical gas fraction and metallicity relations.<sup>86</sup> Once SFR,  $M_*$ , and  $M_{\text{dust}}$  values are assigned to each halo,  $S_{870\mu\text{m}}$  is calculated using the following fitting function, which was derived based on the results of performing dust radiative transfer on hydrodynamical simulations of both isolated and interacting galaxies:<sup>81,87</sup>

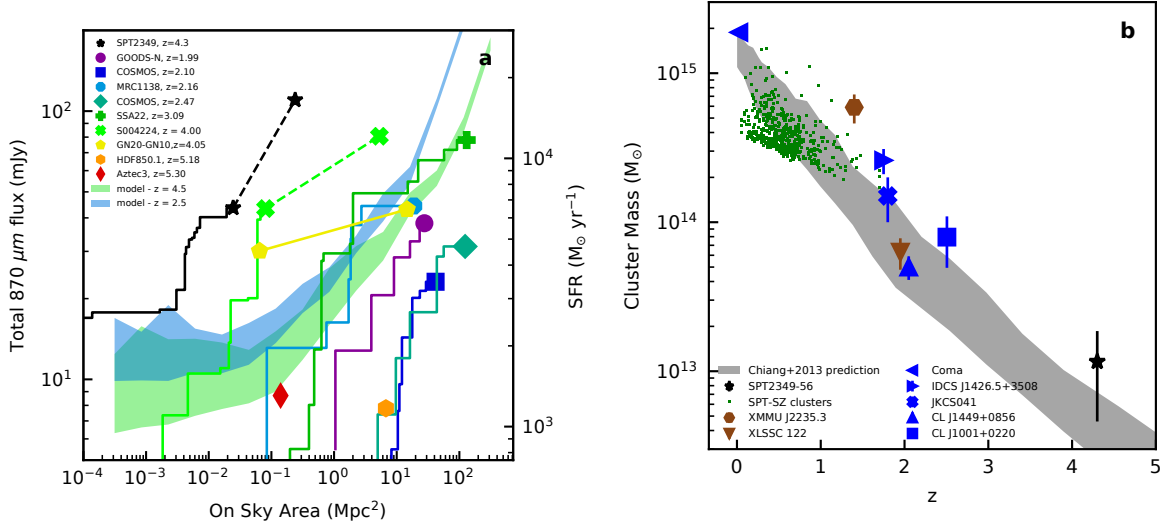
$$S_{870\mu\text{m}} = 0.81 \text{ mJy} \left( \frac{\text{SFR}}{100 M_{\odot} \text{ yr}^{-1}} \right)^{0.43} \left( \frac{M_d}{10^8 M_{\odot}} \right)^{0.54}, \quad (1)$$

where  $S_{870\mu\text{m}}$  is the 870- $\mu\text{m}$  flux density, SFR is the star formation rate, and  $M_d$  is the dust mass. Scatter of 0.13 dex is included when applying the relation.

Once  $S_{870\mu\text{m}}$  has been assigned to each halo, we search the entire simulation volume for the most luminous regions. We begin at each independent halo and calculate the total  $S_{870\mu\text{m}}$  of all halos within a given projected radius  $r$  and LOS distance along a given axis of this halo. We use a line of sight distance of 1 Mpc for the  $z = 4$  snapshot and 2 Mpc for the  $z = 2.5$  snapshot reflecting the Hubble constant at each epoch and the expected velocity dispersion of  $\sim 400 \text{ km s}^{-1}$ . For each value of  $r$ , we record the largest total  $S_{870\mu\text{m}}$  obtained (across all halos). One hundred Monte Carlo iterations are performed for each snapshot; in each iteration, galaxy properties are re-assigned, drawing from the distributions described above. The shaded region in Figure 2 shows the entire region spanned by the 100 realizations of the maximum  $S_{870\mu\text{m}}$  vs. area curves. To compare to SPT2349-56 to lower redshift proto-clusters we perform a similar analysis on a snapshot at  $z = 2.49$  with 20 Monte Carlo iterations.



**Figure 1: The SPT2349-56 field and spectra of the constituent galaxies.** (a) The LABOCA 870- $\mu\text{m}$  contours of SPT2349-56 overlaid on the IRAC 3.6- $\mu\text{m}$  image; the 26'' beam at 870  $\mu\text{m}$  is shown. Contours represent SNR = 3, 7 and 9. The small circles show the locations of the 14 protocluster sources. (b) ALMA band 7 imaging (276 GHz, 1.1 mm) displaying the 14 confirmed protocluster sources, labeled A-N. Black (blue) contours denote the points 75% and 90% of the peak flux for each source from the CO(4-3) ([CII]). The dashed black line shows where the primary beam is 50% the maximum. The filled blue ellipse shows the 0.4'' naturally weighted synthesized beam. (c) CO(4-3) spectra (black lines) and [CII] spectra overlaid (shaded yellow bars) for all 14 sources centred at the biweight cluster redshift  $z = 4.304$ . The [CII] spectra are scaled down in flux by a factor of ten for clarity of presentation. The red arrows show the velocity offsets determined by fitting a Gaussian profile to the CO(4-3) spectra for all sources except for G, H, K, L, M, and N, for which we used [CII] because these are not detected in CO(4-3).



**Figure 2: Comparison of SPT2349-56 to other cluster and proto-cluster systems.** (a) The cumulative 870- $\mu\text{m}$  flux density vs. on-sky area for SPT2349-56, compared to other SMG-rich overdensities at high redshift (see Methods for details). The solid black line shows the ALMA-identified sources in SPT2349-56, while the dashed line includes the wider-field LABOCA-detected structure. The blue (green) shaded region denotes the maximum flux density vs. area curves obtained in 100 Monte Carlo realizations of a theoretical model for submm-luminous protoclusters at  $z = 4.5$  ( $z = 2.5$ ) based on an N-body simulation (see Methods for full details). Most of the literature SMG overdensities are consistent with the model expectations, whereas SPT2349-56 lies vastly above the region spanned by the model. A recently discovered  $z = 4$  protocluster from *Herschel*,<sup>23</sup> SMM J004224, is quite a unique system but  $> 10$  times less dense than and likely only  $\sim 50\%$  the total luminosity of SPT2349-56. (b) The cluster mass versus redshift is shown for SPT2349-56 and other massive galaxy clusters from the literature with detected ICM and well-defined masses. The colour scheme highlights the different methods for selecting massive clusters employed (brown = X-ray, blue = optical, green = Sunyaev-Zeldovich effect). Error bars represent the  $1\sigma$  standard deviation. We also show the mean protocluster most-massive-progenitor mass vs.  $z$  relation predicted by N-body simulations.<sup>12</sup> The location of SPT2349-56 in this plane suggests a very massive descendant (halo mass of  $> 10^{15} M_\odot$  at  $z = 0$ ), although we caution that the complex growth histories of dark matter halos make it difficult to reliably predict the  $z = 0$  halo mass from the halo mass at a given epoch.<sup>20</sup>



**Table 1:** Derived physical properties of SPT2349-56 protocluster members.

Source	$\Delta V^\ddagger$ [km s <sup>-1</sup> ]	SFR [M <sub>⊙</sub> yr <sup>-1</sup> ]	$M_{\text{gas}}$ [10 <sup>10</sup> M <sub>⊙</sub> ]
A	-90 ± 35	1170 ± 390	12.0 ± 2.1
B	-124 ± 31	1227 ± 409	11.2 ± 2.0
C	603 ± 12	907 ± 302	6.7 ± 1.2
D	-33 ± 40	530 ± 182	8.4 ± 1.5
E	84 ± 21	497 ± 179	4.8 ± 0.9
F	395 ± 82	505 ± 169	3.4 ± 0.7
G	308 ± 42	409 ± 137	2.9 ± 1.3 <sup>†</sup>
H	-719 ± 28	310 ± 105	4.4 ± 2.0 <sup>†</sup>
I	310 ± 78	268 ± 91	2.2 ± 0.5
J	-481 ± 35	243 ± 85	2.2 ± 0.5
K	631 ± 12	208 ± 71	3.1 ± 1.4 <sup>†</sup>
L	-379 ± 18	122 ± 43	3.3 ± 1.5 <sup>†</sup>
M	34 ± 21	75 ± 34	1.2 ± 0.6 <sup>†</sup>
N	90 ± 25	64 ± 29	1.0 ± 0.5 <sup>†</sup>

<sup>‡</sup> Velocity offsets relative to the mean redshift,  $z = 4.304$

<sup>†</sup> [CII] line used to derive  $M_{\text{gas}}$  as CO(4-3) is not detected

**Extended Data Table 1:** Observed properties of SPT2349-56 protocluster members

Source	RA (J2000) [h:m:s]	Dec (J2000) [d:m:s]	$S_{1090\mu\text{m}}$ [mJy]	$S_{870\mu\text{m}}$ [mJy]	CO(4-3) $\int S dv$ [Jy km s <sup>-1</sup> ]	CO(4-3) $\sigma_V$ [km s <sup>-1</sup> ]	[CII] $\int S dv$ [Jy km s <sup>-1</sup> ]	[CII] $\sigma_V$ [km s <sup>-1</sup> ]
A	23:49:42.67	-56:38:19.3	4.63 ± 0.04	7.8 ± 0.1	0.99 ± 0.03	376 ± 46	8.81 ± 0.26	354 ± 30
B	23:49:42.79	-56:38:24.0	4.35 ± 0.04	8.2 ± 0.1	0.92 ± 0.03	341 ± 38	7.53 ± 0.22	314 ± 28
C	23:49:42.84	-56:38:25.1	2.69 ± 0.04	6.0 ± 0.1	0.55 ± 0.02	154 ± 13	4.43 ± 0.17	160 ± 10
D	23:49:41.42	-56:38:22.6	2.20 ± 0.08	3.5 ± 0.3	0.69 ± 0.04	485 ± 64	3.62 ± 0.78	346 ± 129
E	23:49:41.23	-56:38:24.4	2.12 ± 0.11	3.3 ± 0.4	0.39 ± 0.02	199 ± 23	3.47 ± 1.24	310 ± 137
F	23:49:42.14	-56:38:25.8	1.69 ± 0.05	3.4 ± 0.1	0.28 ± 0.03	396 ± 103	4.28 ± 0.35	353 ± 35
G	23:49:42.74	-56:38:25.1	1.11 ± 0.04	2.7 ± 0.1	-	-	2.45 ± 0.23	305 ± 50
H	23:49:43.46	-56:38:26.2	0.85 ± 0.05	2.1 ± 0.1	-	-	3.63 ± 0.30	236 ± 31
I	23:49:42.22	-56:38:28.3	0.78 ± 0.05	1.8 ± 0.1	0.18 ± 0.03	277 ± 90	3.18 ± 0.32	236 ± 24
J	23:49:43.22	-56:38:30.1	0.61 ± 0.06	1.6 ± 0.2	0.19 ± 0.02	151 ± 38	3.79 ± 0.29	138 ± 15
K	23:49:42.96	-56:38:17.9	0.34 ± 0.04	1.4 ± 0.1	-	-	2.54 ± 0.17	129 ± 12
L	23:49:42.38	-56:38:25.8	0.23 ± 0.04	0.8 ± 0.1	-	-	2.78 ± 0.20	176 ± 20
M	23:49:43.39	-56:38:21.1	0.21 ± 0.05	0.5 ± 0.2	-	-	1.04 ± 0.14	87 ± 23
N	23:49:43.27	-56:38:22.9	0.18 ± 0.04	0.4 ± 0.1	-	-	0.86 ± 0.16	128 ± 26

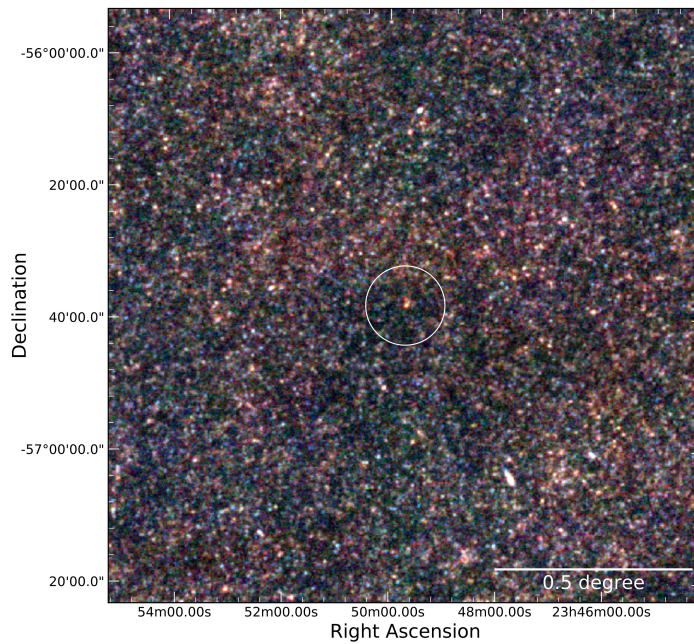
**Extended Data Table 2:** Properties of the 3 ATCA CO(2-1) sources

ATCA source	ALMA ID	$\int S dv$ [Jy km s <sup>-1</sup> ]	$\sigma_V$ [km s <sup>-1</sup> ]	L'(CO 2-1) 10 <sup>11</sup> [K km s <sup>-1</sup> pc <sup>2</sup> ]	$M_{\text{gas}}$ [10 <sup>11</sup> M <sub>⊙</sub> ]
Central (C)	B, C, G	0.69 ± 0.076	372 ± 47	1.22 ± 0.12	1.33 ± 0.15
West (W)	D, E	0.16 ± 0.04	166 ± 47	0.29 ± 0.07	0.32 ± 0.08
North (N)	A, K	0.085 ± 0.0028	175 ± 68	0.15 ± 0.05	0.16 ± 0.05

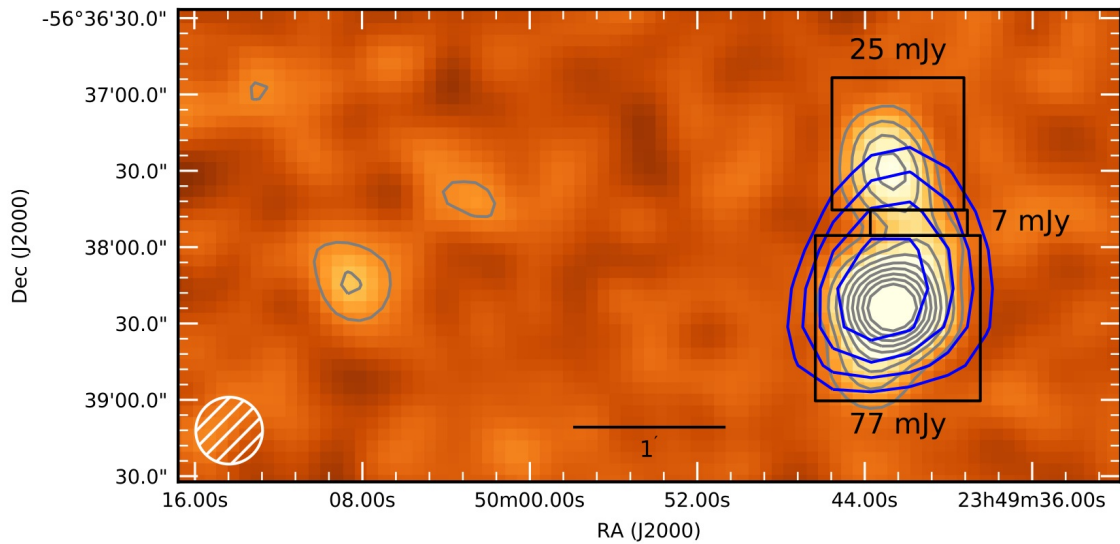
**Extended Data Table 3:** Observed properties of all red ( $S_{500\mu\text{m}} > S_{350\mu\text{m}} > S_{250\mu\text{m}}$ ) SPIRE sources in the field surrounding SPT2349-56. The LABOCA sources corresponding to SPT2349-56 are listed first, and the red SPIRE sources in the surrounding field follow. All sources listed are highlighted in Extended Data Figure 5.

RA (J2000) [h:m:s]	Dec (J2000) [d:m:s]	$S_{250\mu\text{m}}$ [mJy]	$S_{350\mu\text{m}}$ [mJy]	$S_{500\mu\text{m}}$ [mJy]	$S_{850\mu\text{m}}$ [mJy]	d <sup>†</sup> [arcmin]
23:49:42	-56:38:25	45 ± 3	71 ± 3	96 ± 3	77.0 ± 2.9	-
23:49:43	-56:37:31	21 ± 3	37 ± 3	43 ± 3	25.0 ± 2.8	0.9
23:49:25	-56:35:27	23 ± 4	26 ± 4	32 ± 4	2.9 ± 1.7	5.2
23:49:39	-56:36:33	12 ± 3	16 ± 3	23 ± 4	3.9 ± 1.3	2.1
23:49:36	-56:41:17	7 ± 3	14 ± 3	19 ± 3	3.2 ± 1.6	3.2
23:49:55	-56:34:17	6 ± 4	10 ± 3	20 ± 5	4.8 ± 1.8	5.3
23:49:12	-56:40:31	11 ± 5	16 ± 5	22 ± 5	6.8 ± 2.6	7.7

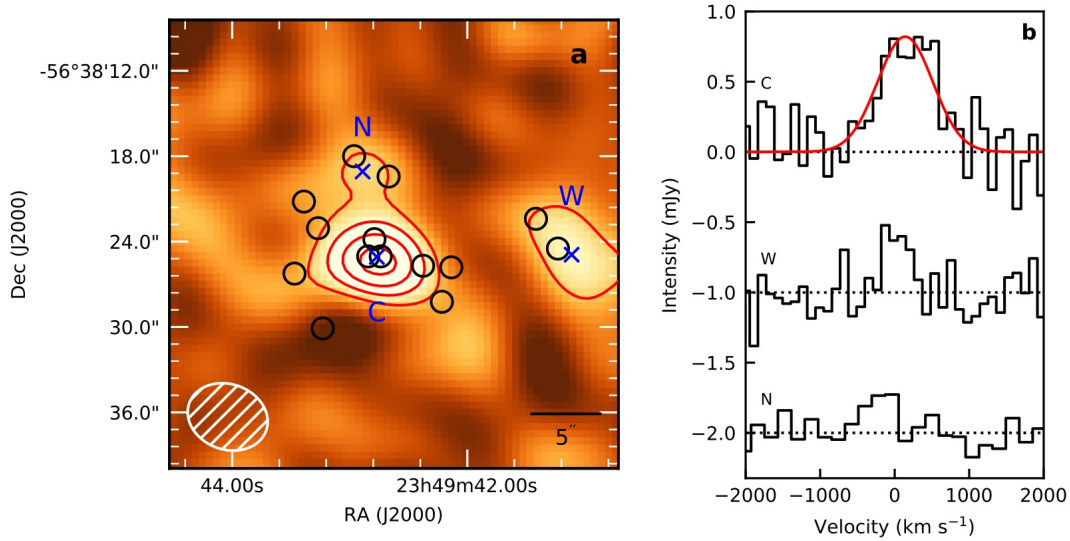
† Distance from central SPT2349-56 source.



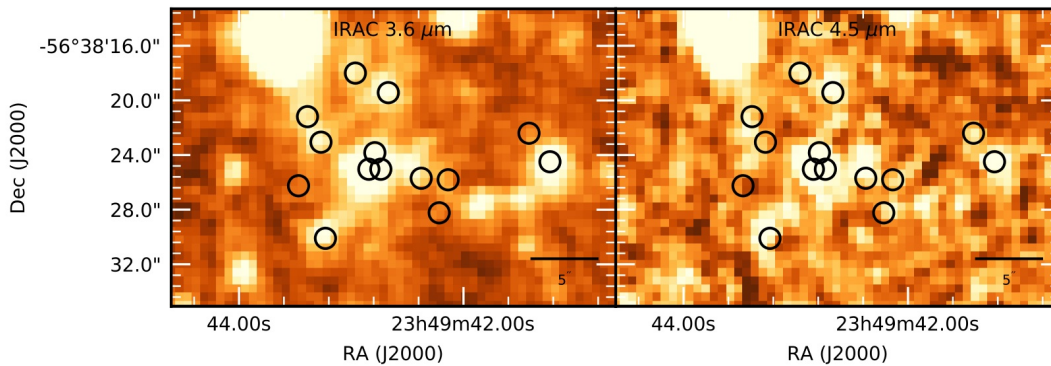
**Extended Data Figure 1: *Herschel*-SPIRE image.** A RGB scale is used to represent 500, 350, and 250  $\mu\text{m}$ , with the red SPT2349-56 extended complex clearly visible in a relative void in the foreground  $z \sim 1$  cosmic infrared background (blue to green-coloured galaxies).



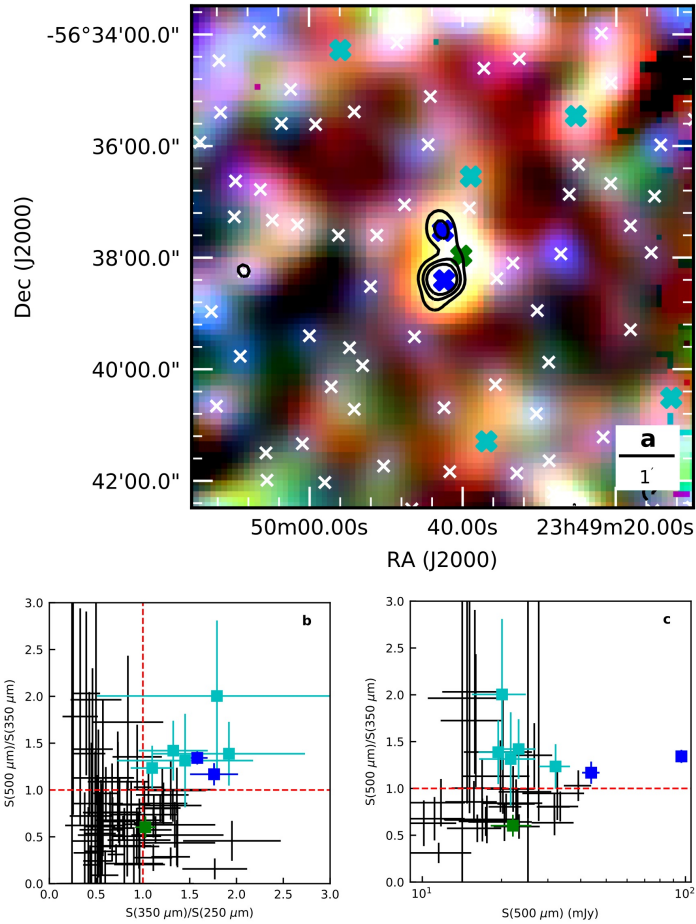
**Extended Data Figure 2: Wide field 870- $\mu\text{m}$  image and photometry.** A wide-field LABOCA image (21'' beam size) of SPT2349-56. The image rms noise is 1.3 mJy at center, increasing to 2mJy at the edges of the region shown. The total flux density recovered is  $110.0 \pm 9.5$  mJy. Subapertures are drawn showing three different regions and their recovered flux densities. Grey contours start at  $3\sigma$  and increase in steps of  $3\sigma$ . SPT 1.4-mm contours are also shown (blue), revealing that even with the 1' beam of SPT, SPT2349-56 is resolved. One additional submm source is detected at  $> 5\sigma$  in the LABOCA image to the east of the primary source, though *Herschel*-SPIRE photometry indicates that it is unlikely to be at  $z \sim 4$ .



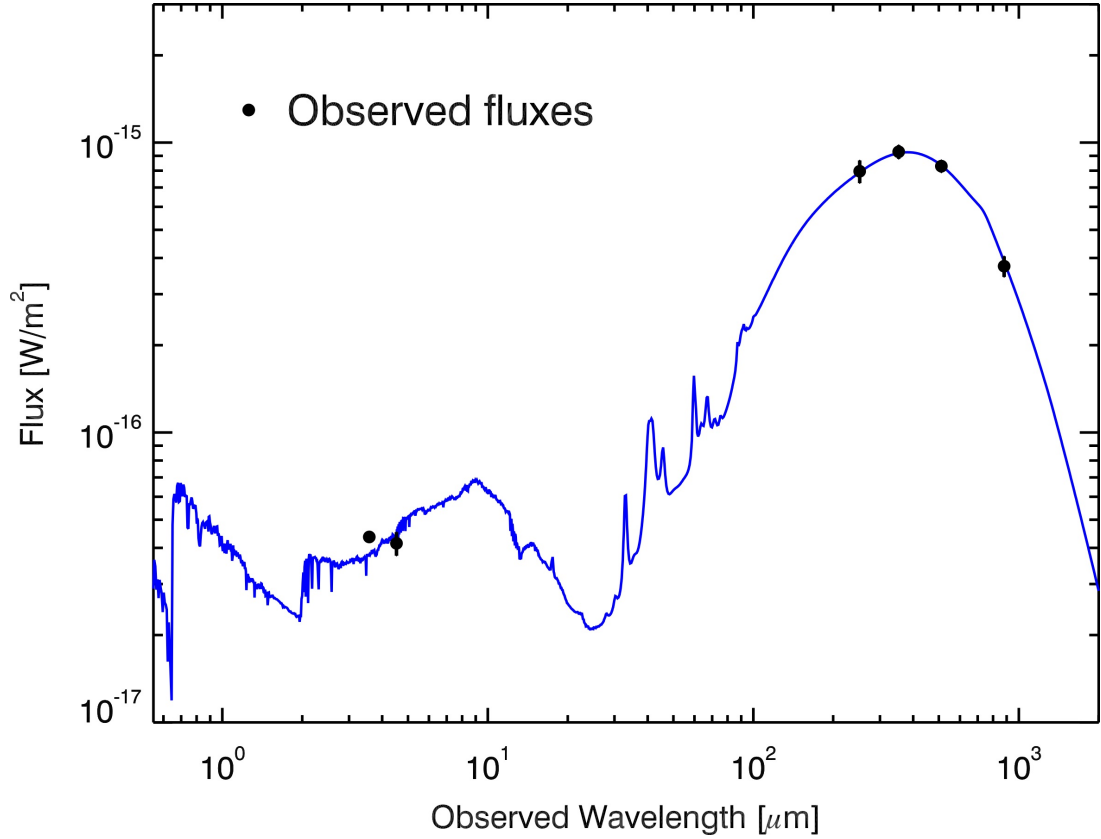
**Extended Data Figure 3: CO(2-1) observations of SPT2349-56.** (a): The colormap and red contours trace the CO(2-1) line integrated over the central  $830 \text{ km s}^{-1}$ , with contours spaced by  $2\sigma$  starting at  $2\sigma$ . The grey contours show the 1.1-mm ALMA continuum detections. (b): One-dimensional spectra extracted at the positions indicated in (a).



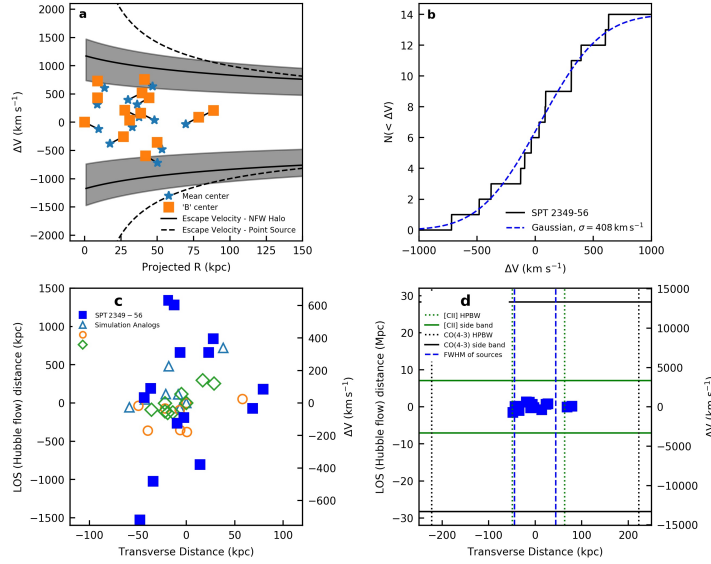
**Extended Data Figure 4: IRAC observations of SPT2349-56.** Circles display the location of the 14 sources detected in ALMA band 7 described above. Nine of the 14 ALMA sources are detected in at least one of the IRAC bands with at least  $3\sigma$  confidence, including the two faintest [CII] sources (M & N) from the blind line survey.



**Extended Data Figure 5: SPIRE RGB image and source colours in field surrounding SPT2349-56** (a) Deep SPIRE false colour image is shown with LABOCA contours overlaid. Locations of  $250\text{-}\mu\text{m}$  peaks used for analysis are marked with crosses (the faintest are not visible because of the contrast adopted in the image). Colour-colour (CC, (b)) and colour-flux (CF, (c)) diagrams for  $250 \mu\text{m}$  sources are also shown. Error bars represent the  $1 \sigma$  standard deviation. The CC diagram shows sources with  $\text{SNR}(250 \mu\text{m}) > 3$  and is dominated by the  $z \sim 1$  cosmic infrared background in the *foreground* of SPT2349-56 (sources with colours ranging from blue to green). The CF diagram applies an additional  $\text{SNR}(500 \mu\text{m}) > 3$  cut. The CC and CF diagrams show that one of three peaks associated with SPT2349-56 is likely a lower-redshift interloper (green symbol), but also that there are five additional sources (blue symbols) in the surrounding region with colours ( $S_{500 \mu\text{m}} > S_{350 \mu\text{m}} > S_{250 \mu\text{m}}$ ) that are suggestive of  $z=4.3$ .



**Extended Data Figure 6: Spectral energy distribution of SPT2349-56.** The SED of the extended SPT2349-56 source is shown, including the summed deconvolved *Herschel*-SPIRE flux densities, the total 870- $\mu\text{m}$  LABOCA flux density, and the summed IRAC flux densities. Error bars represent the  $1\sigma$  measurement errors. We do not include the SPT 1.4, 2.0, and 3.0 mm points because the source is elongated and flux measurements are difficult with the filtering used to make the map. Fitting the SED yields an IR luminosity of  $(8.0 \pm 1.0) \times 10^{13} L_{\odot}$ .



**Extended Data Figure 7: Geometry and dynamics of the SPT2349-56 system.** (a) The velocity offsets versus projected (physical) distance are shown, compared to the escape velocity of a  $1.16 \times 10^{13} M_{\odot}$  NFW halo ( $r_{\text{vir}} \sim 200$  kpc and concentration  $c = 5$ ), the grey shaded region showing our estimated halo mass uncertainty. Also shown is the escape velocity assuming a point mass halo of the same mass. Centres for the galaxy distribution are shown at the mean of the distribution, and centred on the ‘B’ galaxy. All galaxies are bound for all but the lowest range of NFW halo masses. (b) The cumulative velocity distribution of the SPT2349-56 galaxies, compared to a Gaussian distribution with our estimated dispersion ( $408 \text{ km s}^{-1}$ ), is at least consistent with expectations for a relatively bound system. (c) The physical distribution of the SPT2349-56 galaxies is shown, assuming the redshifts are due to cosmic expansion rather than peculiar motions. This gives an extreme (but unlikely) possibility that SPT2349-56 galaxies are stretched out along a *filament* compared to their 130kpc maximum tangential extent, but this requires that none of the velocity offsets are peculiar motions. The open markers show analogues of SPT2349-56 found in a search specifically for maximally extended filamentary structures in the N-body simulations. These simulated structures that we have found are not filaments – they are close to collapsed structures which are slightly cigar-shaped, and while SPT2349-56 could in principle be distributed like this, it does not fundamentally change our discussion in the paper. We also note that we are plotting 2 different things: velocity offsets for SPT2349-56 and actual geometry for the simulation galaxies, i.e. the 3-D positions. Even though the search allowed for structures extending  $\sim 5$  Mpc along the line of sight, none are found to stretch beyond 1 Mpc. (d) Similar to panel (c) except the full extent of our ALMA band 3 & 7 observations is shown. No structures are observed in the sidebands surrounding the 14 observed sources.



## References

- [1] Wang, T. *et al.* Discovery of a galaxy cluster with a violently starbursting core at  $z = 2.506$ . *The Astrophysical Journal* **828**, 56 (2016).
- [2] Mantz, A. B. *et al.* The XXL Survey: XVII. X-ray and Sunyaev-Zel'dovich Properties of the Redshift 2.0 Galaxy Cluster XLSSC 122. *ArXiv e-prints* (2017). 1703.08221.
- [3] Stanford, S. A. *et al.* IDCS J1426.5+3508: Discovery of a Massive, Infrared-Selected Galaxy Cluster at  $z = 1.75$ . *The Astrophysical Journal* **753**, 164 (2012).
- [4] Springel, V. *et al.* Simulating the joint evolution of quasars, galaxies and their large-scale distribution. *Nature* **435**, 629–636 (2005).
- [5] Overzier, R. A. *et al.*  $\Lambda$ CDM predictions for galaxy protoclusters - I. The relation between galaxies, protoclusters and quasars at. *Monthly Notices of the Royal Astronomical Society* **394**, 577–594 (2009).
- [6] Chiang, Y.-K., Overzier, R. A., Gebhardt, K. & Henriques, B. Galaxy Protoclusters as Drivers of Cosmic Star Formation History in the First 2 Gyr. *The Astrophysical Journal Letters* **844**, L23 (2017).
- [7] Miley, G. & De Breuck, C. Distant radio galaxies and their environments. *Astronomy and Astrophysics Review* **15**, 67–144 (2008).
- [8] Casey, C. M. *et al.* A massive, distant proto-cluster at  $z = 2.47$  caught in a phase of rapid formation? *The Astrophysical Journal Letters* **808**, L33 (2015).
- [9] Chapman, S. C. *et al.* Do Submillimeter Galaxies Really Trace the Most Massive Dark-Matter Halos? Discovery of a High- $z$  Cluster in a Highly Active Phase of Evolution. *The Astrophysical Journal* **691**, 560–568 (2009).
- [10] Tamura, Y. *et al.* Spatial correlation between submillimetre and Lyman- $\alpha$  galaxies in the SSA22 protocluster. *Nature* **459**, 61–63 (2009).
- [11] Ma, J. *et al.* Stellar Masses and Star Formation Rates of Lensed, Dusty, Star-forming Galaxies from the SPT Survey. *The Astrophysical Journal* **812**, 88 (2015).
- [12] Chiang, Y.-K., Overzier, R. & Gebhardt, K. Ancient Light from Young Cosmic Cities: Physical and Observational Signatures of Galaxy Proto-Clusters. *The Astrophysical Journal* **779**, 127 (2013).
- [13] Vieira, J. D. *et al.* Extragalactic Millimeter-wave Sources in the South Pole Telescope Survey Data: Source Counts, Catalog, and Statistics for an 87 Square-Degree Field. *The Astrophysical Journal* **719**, 763–783 (2010).

- [14] Mocanu, L. M. *et al.* Extragalactic millimeter-wave point source catalog, number counts and statistics from 771 square degrees of the SPT-SZ Survey. *The Astrophysical Journal* **779**, 61 (2013).
- [15] Vieira, J. D. *et al.* Dusty starburst galaxies in the early Universe as revealed by gravitational lensing. *Nature* **495**, 344–347 (2013).
- [16] Spilker, J. *et al.* ALMA Imaging and Gravitational Lens Models of South Pole Telescope-Selected Dusty, Star-Forming Galaxies at High Redshifts. *The Astrophysical Journal* **826**, 112 (2016).
- [17] Strandet, M. L. *et al.* The redshift distribution of dusty star forming galaxies from the SPT survey. *The Astrophysical Journal* **822**, 80 (2016).
- [18] Beers, T. C. *et al.* Measures of location and scale for velocities in clusters of galaxies - A robust approach. *The Astronomical Journal* **100**, 32 (1990).
- [19] Evrard, A. E. *et al.* Virial Scaling of Massive Dark Matter Halos: Why Clusters Prefer a High Normalization Cosmology. *The Astrophysical Journal* **672**, 122–137 (2008).
- [20] Cole, S., Helly, J., Frenk, C. S. & Parkinson, H. The statistical properties of Lambda cold dark matter halo formation. *Monthly Notices of the Royal Astronomical Society* **383**, 546–556 (2008).
- [21] Ono, Y., Ouchi, M., Kurono, Y. & Momose, R. Faint Submillimeter Galaxies Revealed by Multifield Deep ALMA Observation: Number Counts, Spatial Clustering, and a Dark Submillimeter Line Emitter. *The Astrophysical Journal* **795**, 5 (2014).
- [22] Simpson, J. M. *et al.* An ALMA Survey of Submillimeter Galaxies in the Extended Chandra Deep Field South: The Redshift Distribution and Evolution of Submillimeter Galaxies. *The Astrophysical Journal* **788**, 125 (2014).
- [23] Oteo, I. *et al.* An extreme proto-cluster of luminous dusty starbursts in the early Universe. *ArXiv e-prints* (2017). 1709.02809.
- [24] Granato, G. L. *et al.* The early phases of galaxy clusters formation in IR: Coupling hydrodynamical simulations with GRASIL-3D. *Monthly Notices of the Royal Astronomical Society* **450**, 1320–1332 (2015).
- [25] Renzini, A. Stellar Population Diagnostics of Elliptical Galaxy Formation. *Annual Review of Astronomy and Astrophysics* **44**, 141–192 (2006).
- [26] Oñorbe, J. *et al.* How to zoom: Bias, contamination and lagrange volumes in multimass cosmological simulations. *Monthly Notices of the Royal Astronomical Society* **437**, 1894–1908 (2013).

- [27] Carlstrom, J. E. *et al.* The 10 Meter South Pole Telescope. *PASP* **123**, 568 (2011).
- [28] Ashby, M. L. N. *et al.* The Spitzer South Pole Telescope Deep Field: Survey Design and Infrared Array Camera Catalogs. *The Astrophysical Journal Supplement Series* **209**, 22 (2013).
- [29] Collaboration, P. *et al.* Planck 2013 results. XXII. Constraints on inflation. *Astronomy & Astrophysics* **571**, A22 (2014).
- [30] Weiß, A. *et al.* ALMA Redshifts of Millimeter-Selected Galaxies From the SPT Survey: The Redshift Distribution of Dusty Star-Forming Galaxies. *The Astrophysical Journal* **767**, 88 (2013).
- [31] Siringo, G. *et al.* The Large APEX BOlometer CAmera LABOCA. *Astronomy & Astrophysics* **497**, 945–962 (2009).
- [32] Weiß, A. *et al.* LABOCA observations of nearby, active galaxies. *Astronomy & Astrophysics* **490**, 77–86 (2008).
- [33] Aravena, M. *et al.* The ALMA Spectroscopic Survey in the Hubble Ultra Deep Field: Search for [CII] Line and Dust Emission in  $6 < z < 8$  Galaxies. *The Astrophysical Journal* **833**, 71 (2016).
- [34] Williams, J. P. *et al.* Clumpfind: Determining Structure in Molecular Clouds. *Astrophysics Source Code Library* ascl:1107.014 (2011).
- [35] Wilson, W. E. *et al.* The Australia Telescope Compact Array Broad-band Backend: Description and first results. *Monthly Notices of the Royal Astronomical Society* **416**, 832–856 (2011).
- [36] Sault, R. J., Teuben, P. J. & Wright, M. C. H. A retrospective view of Miriad. *Astronomical Data Analysis Software and Systems IV* **77**, 433 (1995).
- [37] McMullin, J. P., Waters, B., Schiebel, D., Young, W. & Golap, K. CASA Architecture and Applications. In *Astronomical Data Analysis Software and Systems XVI*, vol. 376, 127 (2007).
- [38] Solomon, P. M. *et al.* The Molecular Interstellar Medium in Ultraluminous Infrared Galaxies. *The Astrophysical Journal* **478**, 144–161 (1997).
- [39] Fazio, G. G. *et al.* The Infrared Array Camera (IRAC) for the Spitzer Space Telescope. *The Astrophysical Journal Supplement Series* **154**, 10–17 (2004).
- [40] Werner, M. W. *et al.* The Spitzer Space Telescope Mission. *The Astrophysical Journal Supplement Series* **154**, 1–9 (2004).

- [41] Béthermin, M., Dole, H., Cousin, M. & Bavouzet, N. Submillimeter Number Counts at 250, 350 and 500 microns in BLAST data. *Astronomy and Astrophysics* **516**, 13 (2010).
- [42] Barger, A. J. *et al.* Is There a Maximum Star Formation Rate in High-Redshift Galaxies? *The Astrophysical Journal* **784** (2014).
- [43] Safarzadeh, M., Hayward, C. C., Ferguson, H. C. & Somerville, R. S. What Shapes the Far-infrared Spectral Energy Distributions of Galaxies? *The Astrophysical Journal* **818**, 62 (2016).
- [44] Bothwell, M. S. *et al.* A survey of molecular gas in luminous sub-millimetre galaxies. *Monthly Notices of the Royal Astronomical Society* **429**, 3047–3067 (2013).
- [45] Burgarella, D. *et al.* Star formation and dust attenuation properties in galaxies from a statistical ultraviolet-to-far-infrared analysis. *Monthly Notices of the Royal Astronomical Society* **360**, 1413–1425 (2005).
- [46] Noll, S. *et al.* Analysis of galaxy spectral energy distributions from far-UV to far-IR with CIGALE: studying a SINGS test sample. *Astronomy and Astrophysics* **507**, 1793–1813 (2009).
- [47] Bruzual, G. & Charlot, S. Stellar population synthesis at the resolution of 2003. *Monthly Notices of the Royal Astronomical Society* **344**, 1000–1028 (2003).
- [48] Chabrier, G. The Galactic Disk Mass Function: Reconciliation of the Hubble Space Telescope and Nearby Determinations. *The Astrophysical Journal* **586**, L133–L136 (2003).
- [49] Casey, C. M. The Ubiquity of Coeval Starbursts in Massive Galaxy Cluster Progenitors. *The Astrophysical Journal* **824**, 36 (2016).
- [50] Chapman, S. C., Blain, A. W., Smail, I. & Ivison, R. J. A Redshift Survey of the Submillimeter Galaxy Population. *The Astrophysical Journal* **622**, 772–796 (2005).
- [51] Blain, A. W., Chapman, S. C., Smail, I. & Ivison, R. Clustering of Submillimeter-selected Galaxies. *The Astrophysical Journal* **611**, 725–731 (2004).
- [52] Kurk, J. D. *et al.* A Search for clusters at high redshift. I. Candidate Ly $\alpha$  emitters near 1138-262 at  $z = 2.2$ . *Astronomy and Astrophysics* **358**, L1–L4 (2000).
- [53] Kuiper, E. *et al.* A SINFONI view of flies in the Spiderweb: A galaxy cluster in the making. *Monthly Notices of the Royal Astronomical Society* **415**, 2245–2256 (2011).
- [54] Dannerbauer, H. *et al.* An excess of dusty starbursts related to the Spiderweb galaxy. *Astronomy & Astrophysics* **570** (2014).

- [55] Steidel, C. *et al.* A Large Structure of Galaxies At Redshift  $z \sim 3$  and its Cosmological Implications. *The Astrophysical Journal* **492**, 428–438 (1997).
- [56] Hayashino, T. *et al.* Large-Scale Structure of Emission-Line Galaxies at  $z = 3.1$ . *The Astronomical Journal* **128**, 2073–2079 (2004).
- [57] Chapman, S. C. *et al.* Submillimeter Imaging of a Protocluster Region at  $z = 3.09$ . *The Astrophysical Journal* **548**, L17–L21 (2001).
- [58] Geach, J. E. *et al.* A submillimetre survey of Lyman  $\alpha$  haloes in the SA 22 protocluster at  $z = 3.1$ . *Monthly Notices of the Royal Astronomical Society* **363**, 1398–1408 (2005).
- [59] Umehata, H. *et al.* ALMA Deep Field in SSA22: A concentration of dusty starbursts in a  $z = 3.09$  protocluster core. *The Astrophysical Journal Letters* **815** (2015).
- [60] Hung, C.-L. *et al.* Large scale structure around a  $z = 2.1$  cluster. *The Astrophysical Journal* **826** (2016).
- [61] Daddi, E. *et al.* Two Bright Submillimeter Galaxies in a  $z = 4.05$  Protocluster in GOODS-North, and Accurate Radio-Infrared Photometric Redshifts. *The Astrophysical Journal* **694**, 1517–1538 (2009).
- [62] Daddi, E. *et al.* A CO Emission Line From the Optical and Near-IR Undetected Submillimetre Galaxy GN10. *The Astrophysical Journal* **695**, L176–L180 (2009).
- [63] Walter, F. *et al.* The intense starburst HDF 850.1 in a galaxy overdensity at  $z \sim 5.2$  in the Hubble Deep Field. *Nature* **486**, 233–236 (2012).
- [64] Lewis, A. J. R. *et al.* Ultra-Red Galaxies Signpost Candidate Proto-Clusters at High Redshift. *ArXiv e-prints* (2017). 1711.08803.
- [65] Noirot, G. *et al.* HST Grism Confirmation of Two  $z \sim 2$  Structures From the Clusters Around Radio-Loud AGN (CARLA) Survey. *The Astrophysical Journal* **830**, 90 (2016).
- [66] Ivison, R. J. *et al.* Herschel-ATLAS: a binary HyLIRG pinpointing a cluster of starbursting proto-ellipticals. *The Astrophysical Journal* **772** (2013).
- [67] Fu, H. *et al.* The rapid assembly of an elliptical galaxy of 400 billion solar masses at a redshift of 2.3. *Nature* **498**, 338–341 (2013).
- [68] Chapman, S. C. *et al.* A millimetre-wave redshift search for the unlensed HyLIRG, HS1700.850.1. *Monthly Notices of the Royal Astronomical Society* **453**, 951–959 (2015).
- [69] Miller, T. B., Hayward, C. C., Chapman, S. C. & Behroozi, P. S. The bias of the submillimetre galaxy population: SMGs are poor tracers of the most-massive structures in the  $z \sim 2$  Universe. *Monthly Notices of the Royal Astronomical Society* **452**, 878–883 (2015).

- [70] Trainor, R. F. & Steidel, C. C. The Halo Masses and Galaxy Environments of Hyperluminous QSO's at  $z \sim 2.7$  in the Keck Baryonic Structure Survey. *The Astrophysical Journal* **752**, 39 (2012).
- [71] Rosati, P. *et al.* Multi-wavelength study of XMMU J2235.3-2557: the most massive galaxy cluster at  $z > 1$ . *Astronomy and Astrophysics* **508**, 583–591 (2009). arXiv: 0910.1716v2.
- [72] Brodwin, M. *et al.* IDCS J1426.5+3508: The Most Massive Galaxy Cluster at  $z > 1.5$ . *The Astrophysical Journal* **817**, 122 (2016).
- [73] Andreon, S. *et al.* JKCS 041: a Coma cluster progenitor at  $z = 1.803$ . *A & A* **565**, A120 (2014). 1311.4361.
- [74] Gobat, R. *et al.* A mature cluster with X-ray emission at  $z = 2.07$ . *A & A* **526**, A133 (2011). 1011.1837.
- [75] Gobat, R. *et al.* WFC3 GRISM Confirmation of the Distant Cluster Cl J1449+0856 at  $\langle z \rangle = 2.00$ : Quiescent and Star-forming Galaxy Populations. *The Astrophysical Journal* **776**, 9 (2013).
- [76] Bocquet, S. *et al.* Mass Calibration and Cosmological Analysis of the SPT-SZ Galaxy Cluster Sample Using Velocity Dispersion and X-ray Measurements. *The Astrophysical Journal* **799** (2014). 1407.2942.
- [77] Mandelker, N., van Dokkum, P. G., Brodie, J. P., van den Bosch, F. C. & Ceverino, D. Cold Filamentary Accretion and the Formation of Metal Poor Globular Clusters and Halo Stars. *ArXiv e-prints* (2017). 1711.09108.
- [78] Cautun, M., van de Weygaert, R., Jones, B. J. T. & Frenk, C. S. Evolution of the cosmic web. *MNRAS* **441**, 2923–2973 (2014). 1401.7866.
- [79] Bower, R. G. *et al.* Breaking the hierarchy of galaxy formation. *MNRAS* **370**, 645–655 (2006). astro-ph/0511338.
- [80] Hopkins, P. F., Hernquist, L., Cox, T. J. & Kereš, D. A Cosmological Framework for the Co-Evolution of Quasars, Supermassive Black Holes, and Elliptical Galaxies. I. Galaxy Mergers and Quasar Activity. *ApJS* **175**, 356–389 (2008). 0706.1243.
- [81] Hayward, C. C. *et al.* Spatially unassociated galaxies contribute significantly to the blended submillimetre galaxy population: predictions for follow-up observations of ALMA sources. *Monthly Notices of the Royal Astronomical Society* **434**, 2572–2581 (2013).

- [82] Prada, F., Klypin, A. A., Cuesta, A. J., Betancort-Rijo, J. E. & Primack, J. Halo concentrations in the standard  $\Lambda$  cold dark matter cosmology. *Monthly Notices of the Royal Astronomical Society* **423**, 3018–3030 (2012).
- [83] Behroozi, P. S., Wechsler, R. H. & Wu, H.-Y. The Rockstar Phase-Space Temporal Halo Finder and the Velocity Offsets of Cluster Cores. *The Astrophysical Journal* **762** (2011).
- [84] Behroozi, P. S., Wechsler, R. H. & Conroy, C. The Average Star Formation Histories of Galaxies in Dark Matter Halos From  $z = 0 - 8$ . *The Astrophysical Journal* **770**, 57 (2013).
- [85] Sargent, M. T. *et al.* The Contribution of Starbursts and Normal Galaxies to Infrared Luminosity Functions at  $z < 2$ . *The Astrophysical Journal* **747**, L31 (2012).
- [86] Hayward, C. C. *et al.* Submillimetre galaxies in a hierarchical universe: Number counts, redshift distribution and implications for the IMF. *Monthly Notices of the Royal Astronomical Society* **428**, 2529–2547 (2013).
- [87] Hayward, C. C. *et al.* What Does a Submillimeter Galaxy Selection Actually Select? The Dependence of Submillimeter Flux Density on Star Formation Rate and Dust Mass. *The Astrophysical Journal* **743**, 159 (2011).