figure overlap considerably, but are not identical because not all of the requisite information is available for each source.

Dust mass. Mass estimates are unmodified from literature values[4,15,56–62], owing to the heterogeneity of the data available across the sample. The dust masses are generally derived from the far-infrared continuum emission, using one to several wavelengths. Differences between the cosmology assumed here and previously result in unimportant corrections and are ignored.

Gas mass. Following standard observational practice, the primary source for the gas masses[4,27,56–58,61,63–67] shown in Fig. 2b is measurement of the luminosity of rotational transitions of CO. The lowest available rotational transition is typically used; any translation between the observed transition and the $J=1$–0 line, which is most commonly used as a molecular gas indicator, is taken from the original source. Rather than accepting the varying coefficients for the conversion of CO luminosity to gas mass, we re-calculate all masses using a common value of $\alpha_{\rm CO}=1.0M_\odot$ (K km s^{-1} pc$^2)^{-1}$, which is a typical value for actively star-forming galaxies[68,69]. For one source the gas mass is estimated through the star-formation surface density[59].

Halo mass. The halo masses of Fig. 3 are derived from the gas mass sample above. Each halo mass is represented using a range of values, starting with a conservative and hard lower limit found by dividing the measured gas mass by the universal baryon fraction[10] $f_{\rm b}=0.19$. This lower limit ignores any baryonic mass that has been converted into stars or hot or cool atomic gas phases, which would increase the inferred halo mass. A more realistic, but still conservative, lower limit is represented by the top of the plotted symbols in Fig. 3. Here we assume that the ratio of baryonic mass to halo mass is $M_{\rm b}/M_{\rm halo}=0.05$. This value is a factor of about four less than the universal baryon fraction but still higher than the typical stellar-to-halo mass ratio inferred for haloes of any mass and redshift via subhalo abundance matching[3]. Given that we do not expect high-mass galaxies such as SPT0311−58 to expel a large fraction of their molecular gas content[70] or to later accrete dark matter without also accreting gas in proportion to the universal baryon fraction, it is reasonable to expect that the baryon-to-halo mass ratio should be less than this inferred upper limit on the stellar-to-halo mass ratio across all masses and redshifts.

HFLS3 and SPT0311−58 masses. For the two most distant DSFGs, HFLS3[4] and SPT0311−58, which have extensive far-infrared photometry and atomic and molecular line measurements, we also compute the gas mass using a joint continuum–line radiative transfer model[6,29]. The mass for SPT0311−58 has been computed previously[6] without spatially resolved (CO and [C I]) line emission. For Fig. 3, only the total gas mass of the two SPT0311−58 sources is important for estimating the halo mass. For Fig. 2, the dust mass is divided between the two sources according to the ratio of dust continuum emission in our resolved observations. The gas mass is similarly divided, although the velocity profile of the CO lines provides weak evidence that the molecular gas is concentrated in SPT0311−58 W, which would increase the gas mass for this source by 15%.

Calculation of halo rareness. Figure 3 demonstrates the 'rareness' of SPT0311−58 by considering its position in the dark-matter halo mass–redshift plane compared with other extreme high-redshift objects (DSFGs, QSOs and an LBG) that are believed to be hosted by massive dark-matter haloes. To quantify the rareness of these extreme objects we use a previously reported method[28], including a MATLAB script (https://bitbucket.org/itrharrison/hh13-cluster-rareness) that we modified slightly to extend the calculation to $z=10$. This method enables us to compute (z, $M_{\rm halo}$) contours ('exclusion curves') above which the Poisson probability of such an object being detected in the standard ΛCDM cosmology is less than $\alpha<1$; the existence of a single object above such an exclusion curve is sufficient to rule out ΛCDM at the $100(1-\alpha)$% confidence level. In Fig. 3, we plot 1σ exclusion curves ($\alpha=0.32$). Of the three different statistical measures of rareness proposed[28], we use the '$>\nu$' measure, which quantifies the rareness according to the minimum height of the primordial density perturbation from which a halo of mass $M_{\rm halo}$ and redshift z could have formed: $\nu(M_{\rm halo}, z)\propto[D_+(z)\sigma(M_{\rm halo})]^{-1}$, where $D_+(z)$ is the normalized linear growth function and $\sigma^2(M_{\rm halo})$ is the variance of the matter power spectrum smoothed on the co-moving spatial scale that corresponds to the mass $M_{\rm halo}$. This statistic is sensitive to changes in the ΛCDM initial conditions, such as primordial non-Gaussianity (which would lead to more high-mass dark-matter haloes at a given redshift than expected in the standard ΛCDM cosmology). For the purposes of this calculation, we assume a ΛCDM cosmology with parameters[10] $\Omega_{\rm m}=0.309$, $\Omega_\Lambda=0.691$, $h_0=0.677$ and $\sigma_8=0.816$ and use a previously reported halo mass function[71].

The $>\nu$ rareness statistic (and the corresponding exclusion curves) depends on the region of the $M_{\rm halo}$–z plane to which the survey is sensitive. We assume that the SPT sample of lensed DSFGs is complete for $z>1.5$. At lower redshift, the probability of lensing is strongly suppressed[30,72], which means that the galaxy (or galaxies) associated with a halo mass of more than about $10^{15}M_\odot$ (the $M_{\rm halo}$ value of the exclusion curves for $z=1.5$) would have to have a very high intrinsic (that is, unlensed) millimetre-wavelength flux density (more than about 20 mJy) to be included in the sample. Because of the effects of downsizing (that is, star formation is terminated at higher redshift in higher-mass galaxies than in lower-mass galaxies), it is unlikely that massive galaxies at $z<1.5$ would have sufficiently high infrared luminosity to be detected by the SPT[73]. We furthermore assume that the survey is complete for $M_\odot>10^{11}M_\odot$. The assumption that the sample is complete to $M_{\rm halo}>10^{11}M_\odot$ is a conservative one because the galaxies hosted by such haloes (which would have $M_{\rm b\,\odot}\,10^{11}M_\odot$) are unlikely to be sufficiently luminous to be detected without being very strongly lensed ($\mu>10$); erring on the side of overestimating the completeness yields a lower limit on the rareness. Substituting a minimum halo mass of, for example, $10^{12}M_\odot$ would make the value of the $>\nu$ rareness statistic less than that found for $10^{11}M_\odot$; that is, SPT0311−58 would be inferred to be even rarer.

The total area from which the SPT DSFG sample was selected is 2,500 deg^2. However, the fact that most of the SPT DSFGs are strongly lensed implies that the effective survey area is potentially much less than 2,500 deg^2 because not only must a galaxy have a high intrinsic millimetre-wavelength flux density to be included in the sample but it also must be gravitationally lensed so that it exceeds the approximately 20-mJy threshold for inclusion in redshift follow-up observations. Properly accounting for the effects of lensing on the sample completeness would require defining an effective survey area as a function of halo mass and redshift: $A_{\rm eff}(M_{\rm halo}, z)=2{,}500$ deg$^2\times P(\mu_{\rm min}\,|\,M_{\rm halo}, z)$, where $P(\mu\,|\,M_{\rm halo}, z)$ is the probability of a galaxy hosted by a halo of mass $M_{\rm halo}$ at redshift z being lensed by a factor $\mu_{\rm min}$, the minimum magnification necessary for a halo of mass $M_{\rm halo}$ and redshift z to be detectable. However, given the large uncertainties in determining such a function, we opt for a simpler approach. Instead, in Fig. 3 we plot exclusion curves for the full sky (dotted line), for an area of 2,500 deg^2 (dashed line), which corresponds to the assumption that all haloes in the mass and redshift range specified above would be detected even if they were not lensed, and for an area of 25 deg^2 (solid line), which corresponds to the assumption that the survey area corresponds to only the approximately 1% of the SPT fields over which the magnification for sources at $z>1.5$ will be at least[30,72] $\mu=2$, such as SPT0311−58.

Code availability. The lensing reconstruction for the ALMA data was initially performed using the visilens code (https://github.com/jspilker/visilens). Pixelated reconstructions were performed using a proprietary code developed by a subset of the authors and additional non-authors, and we opt not to release this code in connection with this work. The rareness calculation was performed using publicly available code (https://bitbucket.org/itrharrison/hh13-cluster-rareness). The image de-blending for the Spitzer images used GALFIT (https://users.obs.carnegiescience.edu/peng/work/galfit/galfit.html). The SED modelling used the CIGALE code (https://cigale.lam.fr/), version 0.11.0. The photometric redshift of the lens galaxy was estimated using EAZY (https://github.com/gbrammer/eazy-photoz).

Data availability. This paper makes use of the following ALMA data: ADS/JAO.ALMA#2016.1.01293.S and ADS/JAO.ALMA#2015.1.00504.S, available at http://almascience.org/aq?projectcode=2015.1.00504.S and http://almascience.org/aq?projectcode=2016.1.01293.S. The HST data are available online at the Mikulski Archive for Space Telescopes (MAST; https://archive.stsci.edu) under proposal ID 14740. Datasets analysed here are available from the corresponding author on reasonable request.

30. Weiß, A. *et al.* ALMA redshifts of millimeter-selected galaxies from the SPT survey: the redshift distribution of dusty star-forming galaxies. *Astrophys. J.* **767,** 88 (2013).
31. Strandet, M. L. *et al.* The redshift distribution of dusty star-forming galaxies from the SPT survey. *Astrophys. J.* **822,** 80 (2016).
32. Hezaveh, Y. D. *et al.* ALMA observations of SPT-discovered, strongly lensed, dusty, star-forming galaxies. *Astrophys. J.* **767,** 132 (2013).
33. Fazio, G. G. *et al.* The Infrared Array Camera (IRAC) for the Spitzer space telescope. *Astrophys. J. Suppl. Ser.* **154,** 10–17 (2004).
34. Ashby, M. L. N. *et al.* SEDS: the Spitzer extended deep survey. Survey design, photometry, and deep IRAC source counts. *Astrophys. J.* **769,** 80 (2013).
35. Schuster, M. T., Marengo, M. & Patten, B. M. IRACproc: a software suite for processing and analyzing Spitzer/IRAC data. *Proc. SPIE* **6270,** 627020 (2006).
36. Bertin, E. & Arnouts, S. SExtractor: software for source extraction. *Astron. Astrophys. Suppl. Ser.* **117,** 393–404 (1996).
37. Hook, I. M. *et al.* The Gemini-North multi-object spectrograph: performance in imaging, long-slit, and multi-object spectroscopic modes. *Publ. Astron. Soc. Pacif.* **116,** 425–440 (2004).
38. Eikenberry, S. *et al.* FLAMINGOS-2: the facility near-infrared wide-field imager and multi-object spectrograph for Gemini. *Proc. SPIE* **8446,** 84460I (2012).
39. Warner, C., Packham, C., Eikenberry, S. S. & Gonzalez, A. GPUs and Python: a recipe for lightning-fast data pipelines. *ASP Conf. Ser.* **461,** 53–56 (2012).
40. Warner, C., Eikenberry, S. S., Gonzalez, A. H. & Packham, C. Redefining the data pipeline using GPUs. *ASP Conf. Ser.* **475,** 79–82 (2013).
41. Ma, J. *et al.* Stellar masses and star formation rates of lensed, dusty, star-forming galaxies from the SPT survey. *Astrophys. J.* **812,** 88 (2015).