A complete distribution of redshifts for submillimetre galaxies in the SCUBA-2 Cosmology Legacy Survey UDS field

D. J. B. Smith, 1 C. C. Hayward, 2 M. J. Jarvis 3, 4 and C. Simpson 5

1 Centre for Astrophysics Research, School of Physics, Astronomy and Mathematics, University of Hertfordshire, Hatfield, Hertfordshire, AL10 9AB, UK
2 Center for Computational Astrophysics, Flatiron Institute, 162 Fifth Avenue, New York, NY 10010, USA
3 Oxford Astrophysics, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK
4 Department of Physics, University of the Western Cape, Bellville 7535, South Africa
5 Gemini Observatory, Northern Operations Center, 670 North A’ohoku Place, Hilo, HI 96720-2700, USA

Accepted 2017 July 3. Received 2017 June 30; in original form 2017 March 6

ABSTRACT
Submillimetre galaxies (SMGs) are some of the most luminous star-forming galaxies in the Universe; however, their properties remain hard to determine due to the difficulty of identifying their optical/near-infrared counterparts. One of the key steps to determining the nature of SMGs is measuring a redshift distribution representative of the whole population. We do this by applying statistical techniques to a sample of 761,850 µm sources from the Submillimetre Common User Bolometer Array 2 Cosmology Legacy Survey observations of the UKIDSS Ultra-Deep Survey (UDS) field. We detect excess galaxies around >98.4 per cent of the 850 µm positions in the deep UDS catalogue, giving us the first 850 µm selected sample to have virtually complete optical/near-infrared redshift information. Under the reasonable assumption that the redshifts of the excess galaxies are representative of the SMGs themselves, we derive a median SMG redshift of $z = 2.05 \pm 0.03$, with 68 per cent of SMGs residing between $1.07 < z < 3.06$. We find an average of $1.52 \pm 0.09$ excess K-band galaxies within 12 arcsec of an 850 µm position, with an average stellar mass of $2.2 \pm 0.1 \times 10^{10} M_\odot$. Whilst the vast majority of excess galaxies are star forming, 8.0 ± 2.1 per cent have passive rest-frame colours and are therefore unlikely to be detected at submillimetre wavelengths even in deep interferometry. We show that brighter SMGs lie at higher redshifts, and use our SMG redshift distribution – along with the assumption of a universal far-infrared Spectral Energy Distribution (SED) – to estimate that SMGs contribute around 30 per cent of the cosmic star formation rate density between $0.5 < z < 5.0$.

Key words: galaxies: high-redshift – submillimetre: galaxies.

1 INTRODUCTION
Submillimetre wavelengths are well suited to studying high-redshift galaxies, since redshift effects combine with the steep Rayleigh–Jeans tail of the dust spectral energy distribution resulting in negative K-corrections. This ensures that galaxies of a fixed far-infrared (far-IR) luminosity (a proxy for star formation rate) are approximately equally bright from $z = 1$ out to $z = 6$ (e.g. Blain et al. 2002).

The low resolution of submillimetre observations with even the most widely used cameras – the Submillimetre Common User Bolometer Array (SCUBA; Holland et al. 1999) on the 15 m James Clerk Maxwell Telescope, and the Large APEX BOlometer CAmera (LABOCA; Siringo et al. 2009) on the APEX 12 m telescope – and high-redshift nature of submillimetre galaxies (SMGs) make identifying their counterparts (i.e. the sources responsible for the submillimetre emission) challenging (e.g. Ivison et al. 2007), particularly in deep optical observations. Whilst the tight far-IR radio correlation (e.g. Helou, Soifer & Rowan-Robinson 1985; Yun, Reddy & Condon 2001; Jarvis et al. 2010; Smith et al. 2014) and low sky density of radio sources (e.g. Condon et al. 2012), allow the unambiguous identification of many SMG counterparts using high-resolution radio observations (e.g. Ivison et al. 2002, 2005, 2007), this is still a challenging undertaking, despite the current availability of sensitive radio data over degree-scale fields. This means that the most fundamental requirement for inferring the properties of SMGs – their redshift distribution – is hard to measure (e.g. Smail et al. 2002). Chapman et al. (2005) obtained spectroscopic redshifts to 73 SMGs based on following up 98 out of 104 radio-identified SMGs in a parent sample of $15 > 3 \sigma$ SCUBA 850 µm sources, whilst Wardlow et al. (2011) were able to robustly identify only 72 out of 126 SMGs (an identification rate of 57 per cent) detected...
in LABOCA observations of the ECDF-S field, using a combination of radio, 24 μm and Spitzer IRAC data. Recent advances have improved matters, but even the latest methods (e.g. the ‘tri-colour’ technique put forward by Chen et al. 2016) can identify reliable counterparts to only 69 ± 16 per cent of SMGs. This means that, although SMGs are undoubtedly some of the most rapidly star-forming systems in the Universe (e.g. Hayward et al. 2011, 2013a), optical/near-infrared studies of their counterparts and therefore redshifts have so far been incomplete.

At the same time, several studies had attempted to complete the redshift distribution of SMGs using longer wavelengths, for example, Koprowski et al. (2014), Koprowski et al. (2016) and Michałowski et al. (2017), which used an average SMG template from Michałowski, Hjorth & Watson (2010) to find the redshift giving the best fit to the far-IR/millimetre photometry. Of course, the redshifts produced using long wavelengths are in general less accurate than those produced from optical/near-infrared data (although the latter are also highly uncertain at the highest redshifts). There are several reasons for this: for example, that far-IR data are susceptible to the influence of the choice of dust Spectral Energy Distribution (SED), and are subject to errors arising from blended far-IR photometry (e.g. Scudder et al. 2016, although – as in the aforementioned studies – this may be mitigated using deblending techniques such as those from Merlin et al. 2015 or Hurley et al. 2017). On the other hand, this method of redshift estimation does benefit from not requiring an optical/near-infrared counterpart, is able to obtain individual redshifts for every source and is also potentially less susceptible to the influence of chance alignments between optical sources, e.g. lensing.

Given the low resolution of submillimetre data based on single-dish observations, interferometers have an important role to play for studying SMGs. Some of the first results using sensitive data from the Atacama Large Millimetre Array (ALMA) to identify the counterparts to SMGs by Hodge et al. (2013) showed using a combination of radio and mid-IR data that ~80 per cent of SMGs are identified correctly, but with completeness of just 45 per cent. Interferometry has also shown that many SMGs have multiple counterparts; whilst Barger et al. (2012) found that 37.5 per cent of bright SMGs are multiples (albeit with a 1σ confidence interval between 17 and 74 per cent), Hodge et al. (2013) found that at least 35 per cent of SMGs are multiples, whilst Chen et al. (2013) used the Submillimeter Array (SMA) to determine a lower multiple fraction of 12.5 ± 6.8 per cent, and this value is likely to be affected by the sample selection (e.g. the flux limit). Multiplicity is also a key parameter if we are to clarify the role of SMGs within current models of galaxy formation and evolution; however, it presents considerable challenges for traditional cross-identification efforts (e.g. using likelihood ratio methods; Wolstencroft et al. 1986; Sutherland & Saunders 1992; Smith et al. 2011; Fleuren et al. 2012).

The properties of SMGs are therefore still under investigation, using both observations (e.g. with large amounts of time allocated to investigate the source counts and multiplicity with the ALMA; Hodge et al. 2013; Karim et al. 2013; Carniani et al. 2015; Simpson et al. 2015; Fujimoto et al. 2016; Oteo et al. 2016; Dunlop et al. 2017) and simulations (e.g. can theory produce sufficiently star-forming galaxies, or are chance alignments required? Davé et al. 2010; Hayward et al. 2013b; Cowley et al. 2015; Muñoz Arancibia et al. 2015).

New technology used by the Submillimetre Common User Bolometer Array 2 (hereafter SCUBA-2; Holland et al. 2013) has given us unprecedented mapping speed at these wavelengths, meaning that we can now efficiently survey large areas. As a result, the SCUBA-2 Cosmology Legacy Survey (CLS; Geach et al. 2017) has surveyed ~5 deg² to better than mJy depth, increasing the total known sample of 850 μm sources by an order of magnitude. The extensive efforts to cross-identify these sources at other wavelengths highlight that they are critical if we are to understand the properties of SMGs, and their high redshift, dusty nature means that the K-band data in the UKIDSS UDS (Hartley et al. 2013) are ideal for this purpose. With sensitive ancillary data covering the rest-frame UV, optical and near-infrared (UV/OIR) bands, we can also estimate the redshifts and stellar masses of possible counterparts, as well as classify them on the basis of their rest-frame colours. Williams et al. (2009) identified regions in the rest-frame (U − V) versus (V − J) – hereafter ‘UVJ’ – space believed to contain predominantly passive and star-forming galaxy populations, and this method is widely used (e.g. Whitaker et al. 2012; Hatch et al. 2014; Leja et al. 2015; Mendel et al. 2015; Fumagalli et al. 2016; Rees et al. 2016).

In this paper, we statistically study the optical/near-infrared properties of galaxies around the positions of a large sample of submillimetre sources taken from the CLS. Wardlow et al. (2011) and Simpson et al. (2014) made attempts to statistically identify those SMGs that could not be cross-identified using other means in the ECDF-S field; however, the ancillary data in those works were not deep enough to detect counterparts to the full SMG sample (whilst the former work detected counterparts to 85 per cent of their SMG sample, the latter found only 0.61 ± 0.07 counterparts per ALMA-blank SMG). Given the availability of the largest samples of submillimetre positions from the CLS and the deepest K-band data from the Ultra-Deep Survey (UDS), we now extend this idea to infer the properties of the whole 850 μm selected sample, complementary to the aforementioned long-wavelength estimates of the complete SMG redshift distribution, most recently by Michałowski et al. (2017).

In Section 2, we introduce the data sets used in our analysis, and explain how our sample was selected, whilst Section 3 contains our results, which we discuss in Section 4, before concluding with Section 5. We adopt a standard cosmology with \( \Omega_M = 0.3, \Omega_{\Lambda} = 0.7, H_0 = 71 \, \text{km s}^{-1} \, \text{Mpc}^{-1} \), whilst all magnitudes and colours are quoted on the AB system (Oke & Gunn 1983).

### 2 DATA AND SAMPLE SELECTION

We base our analysis on the 850 μm data from the CLS observations of the UKIDSS UDS field. The CLS 850 μm data reduction is described in Geach et al. (2017), and the UDS field is centred on \( \alpha = 02^{h}17^{m}46^{s}, \delta = -05^\circ05'15'' \) (J2000), covering an area of 0.8 deg² with a median rms ~0.88 mJy beam⁻¹. We use the catalogue of SMGs in the CLS-UDS maps from Geach et al. (2017), which contains 1088 SMGs detected at \( > 3.5 \sigma \).

The UDS data release 8 (Hartley et al. 2013) has a 5σ depth of \( K = 24.6 \) within 2 arcsec apertures over a field of 0.8 deg². Over the central 0.6 degree² of this field, the DR8 catalogue includes matched photometry in the \( UBVRi'z'JKH \) bands (3.6 μm and 4.5 μm bands (measured using the method given in Simpson et al. 2012), ideal for estimating galaxy colours. We use the high-quality (\( \Delta z/ (1+z) \sim 0.031 \)) photometric redshifts from Hartley et al. (2013), alongside stellar mass estimates and rest-frame colours estimated using the FAST code (Kriek et al. 2009), which will be fully described in Smith et al. (in preparation). We derive these assuming an initial mass function from Chabrier (2003), based on the stellar SEDs from Bruzual & Charlot (2003), with exponentially declining star formation histories, a range of metallicities (0.2 < Z/Z⊙ < 2.5), a range of 0.0 < A_V < 5.0 (assuming a dust law from Calzetti et al. 2000) and correcting for extinction in our own galaxy.
Rest-frame colours for galaxies are estimated by convolving the best-fitting BC03 spectra found by \textsc{fast} with the Bessel $U$- and $V$-band curves (Bessell 1990) and with the Mauna Kea $J$-band filter curve (Tokunaga, Simons & Vacca 2002) in the rest frame. These values allow us to locate galaxies within the $UVJ$ diagram, in which passive galaxies are those which have $(U - V)_{\text{rest}} > 1.3$, $(V - J)_{\text{rest}} < 1.6$ and $(U - V)_{\text{rest}} > 0.88 \times (V - J)_{\text{rest}} + 0.49$.

We limit the influence of bright stars and UDS image artefacts (e.g. detector cross-talk) by removing all sources which are flagged in the UDS masks, supplied on the UDS website.\footnote{\url{http://www.nottingham.ac.uk/astronomy/UDS}} In addition, we remove stellar sources from our sample by using the Bayesian classifications from Simpson et al. (2013). We limit our area to the region where the UDS data include photometric redshifts, and after removing regions masked in the UDS catalogue, the effective area of our survey is 0.527 deg$^2$ in which we find 761 SMGs from the Geach et al. (2017) catalogue. We also remove those sources with $\chi^2 > 20$ in either the Hartley et al. (2013) or \textsc{fast} SED fitting to limit the influence of bad fits on photometric redshifts, galaxy colours and stellar masses.

3 RESULTS

In contrast to other investigations on this subject, we do not attempt to identify individual counterparts to CLS submillimetre sources. Instead, we take a statistical approach by measuring the properties of galaxies detected within and immediately around the SCUBA-2 beam, after subtracting off the expected background population, as a function of search radius. This method has the benefit of not requiring any assumptions about the number of $K$-band sources around each individual 850 \mum source, and in the coming sections we will show that this enables us to generate a complete redshift distribution for 850 \mum sources based on the UDS optical/near-infrared photometric redshifts. However, the price that we pay for this is that we do not identify the counterparts to any individual 850 \mum source, and we also suffer from the issue that, because the environments of 850 \mum sources do not identify the counterparts to any individual \mum metric redshifts. However, the price that we pay for this is that we show that this enables us to generate a complete redshift distribution for 850 \mum sources around each individual \mum source, and in the coming sections we will show that this enables us to generate a complete redshift distribution for 850 \mum sources based on the UDS optical/near-infrared photometric redshifts. However, the price that we pay for this is that we do not identify the counterparts to any individual 850 \mum source, and we also suffer from the issue that, because the environments of 850 \mum sources have been shown to be overdense by a factor of $80 \pm 30$ times the background level (Simpson et al. 2015), the excess galaxies that we see in the near-infrared may not be responsible for the submillimetre emission themselves (see Section 3.4 below).

3.1 Astrometry and initial checks

Before searching around the 850 \mum source positions for $K$-band galaxies, we must first ensure that the CLS and UDS positions are consistent with each other. We do this by following the method used for sources in the Herschel-ATLAS survey by Smith et al. (2011), which involves creating a two-dimensional histogram of the difference between the 850 \mum and UDS positions for every source in each catalogue, which is shown in Fig. 1. Each bin of the histogram is 1 arcsec in size and reveals a residual offset in the 850 \mum astrometry relative to the UDS. To make the best estimate of the magnitude and direction of the offset, we model this distribution using a simple three-component model, consisting of two Gaussians with a common centre, plus a constant background, and find the best-fitting values for the centroid. The best-fitting offsets derived using this model are $\Delta \alpha = 1.41 \pm 0.15$ arcsec and $\Delta \delta = -0.95 \pm 0.14$ arcsec, giving an effective offset of 1.7 arcsec (i.e. smaller than the pixel size in the 850 \mum maps), which we apply to the CLS catalogue before continuing.

3.2 The fraction of 850 \mum sources detected in the UDS catalogue

Next, we determine how many of the 850 \mum sources have at least one potential counterpart in the UDS catalogue. We do this using the method of Fleuren et al. (2012, hereafter F12), in which this parameter is referred to as $Q_0$. It is difficult to estimate this value directly, due to the high probability of observing unrelated background sources around the 850 \mum positions (i.e. the observed number of apparent counterparts within a particular search radius is in fact the sum of the actual number of counterparts and the number of chance associations), and due to the possibility of finding multiple $K$-band galaxies per 850 \mum source. F12 showed that it is preferable to estimate $(1 - Q_0)$, by instead counting the number of ‘blanks’ (i.e. the number of sources without any possible counterpart) and statistically accounting for the unrelated background population. In the F12 formalism,

$$\frac{1 - Q_0}{S} = \frac{\tilde{R}}{R}$$

where $S$ is the fraction of 850 \mum sources without a possible counterpart, and we estimate $\tilde{R}$ (the fraction of random positions without a possible counterpart) by generating 100 model catalogues with 761 random positions within the survey area, and returning the number of times zero matches are obtained (accounting for the masked area of our survey). We estimate these values as a function of radius, and the results are shown in Fig. 2, in which the light blue dashed line and red crosses connected by a red solid line show the fraction of blanks around random positions and 850 \mum positions, respectively. The red crosses (showing the results around...
what would be expected at random positions, \( N_{\text{excess}} \). We do this using equation (2):

\[
N_{\text{excess}} = (S - R)N_{\text{850}} - \sum_{N_{\text{850}}} P(\text{false}),
\]

where \( N_{\text{850}} \) is the number of 850 \( \mu m \) sources in our sample, and \( P(\text{false}) \) is the probability that each 850 \( \mu m \) source is a false positive (as calculated by Geach et al. 2017, and these values indicate a total of 45.6 \( \pm \) 1.8 spurious sources in our sample). We then calculate the null probability of this excess by taking the probability that a position is a blank out to some separation, \( R \), and multiplying it by itself \( N_{\text{excess}} \) times, giving a null probability of \( P_{\text{null}} = R^{N_{\text{excess}}} \). The results are shown by the grey dot–dashed line in Fig. 2, and tabulated for separations \( \leq 10 \) arcsec in Table 1, which suggests that we can be confident that we detect a significant galaxy excess around \( >98.4 \pm 0.9 \) per cent of the 850 \( \mu m \) sample. This value is a little surprising, since the ‘false-positive’ probabilities imply that \( Q_0 \) should be no more than 94 per cent, and the uncertainty on \( Q_0 \) is insufficient to fully account for the discrepancy. If the false-positive probabilities are unduly conservative, this is relevant to other studies as well, for example, Michałowski et al. (2017) found that the faintest submillimetre sources have a lower ID fraction and attributed this partly due to the increased fraction of spurious sources at fainter flux limits and partly due to fainter submillimetre sources also being fainter in the radio maps (or other ancillary data) and thus harder to cross-identify. Forthcoming observations with ALMA will hold the key to determine how many of the sources in this catalogue are truly real.

### 3.3 The \( K < 24.6 \) galaxy population around 850 \( \mu m \) sources

Having ascertained that the UDS data are deep enough that almost every CLS 850 \( \mu m \) source has at least one detectable (if not identifiable) and associated \( K < 24.6 \) galaxy, we now wish to estimate how many galaxies we find for each. We do this by calculating the number of possible counterparts within a range of search radii between 0 and 20 arcsec, and subtract off the number of unassociated background sources that we would expect based on scaling the source density in the UDS \( K \)-band catalogue to the relevant search area.

The results are shown in Fig. 3, in which the red points with Poisson error bars show the radial distribution of the number of excess galaxies around the 761 850 \( \mu m \) sources. For comparison, we also overlay the corresponding plot based on 100 Monte Carlo simulations of the hypothetical scenario with one counterpart galaxy for each 850 \( \mu m \) position, assuming normally distributed positional uncertainties depending solely on the signal-to-noise ratio (SNR) of each source in the CLS catalogue (following the method in Ivison et al. 2007), added in quadrature with the 2 arcsec uncertainty resulting from pixel-quantized positions in the public CLS catalogue. We overlay the median-likelihood values (black) on the 16th and 84th percentiles of the cumulative frequency distribution (in grey) as a function of separation. We observe that the number of excess galaxies increases steadily out to around 12 arcsec, giving an average of 1.52 \( \pm 0.09 \) excess \( K < 24.6 \) galaxies around each 850 \( \mu m \) source. These values highlight that in many cases we are detecting an excess of nearby \( K \)-band galaxies rather than the sources responsible for the 850 \( \mu m \) emission; this number is therefore not directly analogous to the SMG ‘multiplicity’ discussed above, but assuming that it is not dominated by a handful of sources with many components (and our estimate of \( Q_0 \) derived in Section 3.2 suggests that this is a reasonable assumption), finding 1.52 \( \pm 0.09 \) excess galaxies per 850 \( \mu m \) position is in good agreement with more direct studies of

---

**Figure 2.** Illustration of the procedure of Fleurieu et al. (2012) to estimate \((1 - Q_0)\), the fraction of 850 \( \mu m \) source positions that are not detected in the UDS catalogue. The red crosses show the fraction of blanks (i.e. those 850 \( \mu m \) source positions with no possible counterpart within the search radius), whilst the light blue dashed line shows the corresponding values based on random positions, both relative to the right-hand vertical axis. The black filled circles represent the values obtained dividing the number of blanks by the randomly generated values, which is our estimate of \((1 - Q_0)\), relative to the left-hand axis. The grey dot–dashed line linking the triangles indicates the probability of obtaining the number of blanks around the 850 \( \mu m \) source positions by chance, and this is \(<1\) per cent at all separations \( \leq 10 \) arcsec. The values for \( Q_0 \) and the probability of the null hypothesis are shown in Table 1.

**Table 1.** Tabulated values of \( Q_0 \) as a function of maximum separation from the 850 \( \mu m \) source positions, with the right-hand column showing the probability of obtaining our results by chance, after accounting for the fraction of sources in the \( >3.5\sigma \) 850 \( \mu m \) catalogue which are spurious.

| Max separation (arcsec) | \( Q_0 \)  | \( P_{\text{null}} \) |
|-------------------------|-----------|----------------------|
| 1                       | 0.103     | 0.000                |
| 2                       | 0.346     | 0.000                |
| 3                       | 0.602     | 0.000                |
| 4                       | 0.802     | 0.000                |
| 5                       | 0.882     | 0.000                |
| 6                       | 0.947     | 0.000                |
| 7                       | 0.967     | 0.000                |
| 8                       | 0.984     | 0.000                |
| 9                       | 0.975     | 0.205                |
| 10                      | 1.000     | 0.780                |

The black filled circles show our best estimate of \((1 - Q_0)\) for CLS 850 \( \mu m \) sources, estimated using equation (1) and detailed in Table 1. To determine whether seeing an excess \( K \)-band galaxy around all of the 850 \( \mu m \) sources is real (or just due to the large source density in the UDS catalogue), we first estimate the average number of excess non-blanks (i.e. detections) that we observe above the 850 \( \mu m \) positions) differ considerably from the results around random positions.

The black filled circles show our best estimate of \((1 - Q_0)\) for CLS 850 \( \mu m \) sources, estimated using equation (1) and detailed in Table 1.
3.4 The properties of galaxies near 850 µm sources

We characterize these galaxies by extending the method of Section 3.3 to produce background-subtracted distributions of their photometric redshifts, $K$-band magnitudes and stellar masses, and these are shown in Fig. 4. To determine the extent to which the properties of these sources are representative of the SMG host galaxy population, we perform an identical analysis using realistic mock catalogues based on lightcones from the Bolshoi simulation (Klypin, Trujillo-Gomez & Primack 2011), containing galaxies with stellar mass $>10^9$ $M_\odot$ over $0.5 < z < 8.0$, which were produced by Hayward et al. (2013b) to study SMGs. Whilst the stellar masses of excess galaxies within 12 arcsec of 850 µm sources in the Bolshoi simulation (selected to the same flux criterion as in the real CLS catalogues) are significantly different from the stellar masses of the SMGs themselves, a Kolmogorov-Smirnov (K-S) test shows that we are unable to rule out the hypothesis that the redshift distributions are drawn from the same parent distribution (we perform this simulation 1000 times and find a median $P \approx 0.3$). We therefore conclude that the photometric redshift distribution derived in this manner is representative of the full SMG population.

In the top panel of Fig. 4, we show the photometric redshift distribution of the 1154.6 ± 67.5 excess galaxies around all 761 >3.5$\sigma$ CLS 850 µm sources in our survey area as the black histogram, derived from the best-fitting redshifts from the Hartley et al. (2013) catalogue, and overlaid (in red) with the corresponding distribution derived from the individual redshift probability distribution functions. The median redshift of the 850 µm sources is $z = 2.05 \pm 0.03$, which makes for an interesting comparison with many independent studies despite their incomplete redshift information (e.g. Wardlow et al. 2011; Casey et al. 2013; Michałowski et al. 2017, shown as the grey shaded region), and those that we have produced for brighter subsets of the SMG population, detected at $>4\sigma$, $>5\sigma$ and $>6\sigma$, as shown by green, blue and purple histograms, respectively. We also overlay the corresponding summed $P(z)$ derived based on the individual redshift probability distributions for each source (in red). In the middle panel, the black filled circles show the same data as the error bars, but corrected for residual completeness in the UDS catalogue, derived using the corrections from Mortlock et al. (2015).
Simpson et al. 2014; Chen et al. 2016) and despite the important role played by selection (which we will discuss in Section 4.2). We find that 18.2 per cent of the SMGs in our sample lie at \( z > 3 \), in good agreement with Wardlow et al. (2011). We also overlay the results of performing an identical analysis on the same number of random positions as the grey error bars, which recover a redshift distribution consistent with zero excess galaxies, reassuring us that our results are not caused by some undiagnosed bias. We have overlaid our results on the redshift distribution for the \( \geq 4\sigma \) sample of SMGs selected from the CLS observations of the COSMOS and UDS fields from Michalowski et al. (2017), who (as discussed in Section 1) completed the redshift distribution for the \( \sim \)third of sources without cross-identified counterparts in their sample using long-wavelength data. The Michalowski et al. (2017) redshift distribution (which has been arbitrarily rescaled for clarity) has a higher median \( (z) = 2.4_{-0.1}^{+0.1} \), and this is most readily apparent in the increased high-redshift tail in the shaded grey distribution in the top panel of Fig. 4. This apparent disagreement between the two \( >4\sigma \) source redshift distributions (shaded grey from Michalowski et al. 2017, corresponding to the green distribution from this work) is of particular interest given the complementary strengths and weaknesses of the two methods.

Though the Bolshoi simulation shows that the mass distribution of the galaxies in the simulation around the 850 \( \mu \)m source positions differs from the mass distribution of the model SMG counterparts, it is nevertheless instructive to calculate the \( K \)-band magnitude and stellar mass distributions of these sources. In the middle panel of Fig. 4, we show the distribution of \( K \)-band magnitude for the excess galaxies, overlaid with the results of the analogous experiment using the same number of random positions (grey error bars). Interestingly, rather than increasing towards the limit of the UDS catalogue, the magnitude distribution peaks around \( K = 23.4 \) mag \( (\sim 1.2 \text{ mag above the } K = 24.6 \text{ mag limit}) \) before turning down. This reinforces our result from Section 3.2 that the UDS data are sensitive enough to detect an excess of galaxies around virtually all of the 761 850 \( \mu \)m sources in the Geach et al. (2017) catalogue for which we have coverage. If the excess \( K \)-band galaxies that we detect are not the sources responsible for the 850 \( \mu \)m emission, the magnitude distribution of SMG counterparts must be bimodal, as expected based on the extreme dust columns around bright SMGs found by Simpson et al. (2015, \( A_V \approx 540_{\text{mag}}^{+80}_{-40} \)). The fact that we are detecting galaxies around 850 \( \mu \)m sources not necessarily responsible for the 850 \( \mu \)m emission offers a possible means to reconcile our results with the fact that 5 out of the 30 \( > 8 \) mJy CLS submillimetre sources in Simpson et al. (2016) do not have a \( K \) counterpart to at least one of the multiple blended components resolved by ALMA in this field.

To test for the possible influence of incompleteness in our \( K \)-band catalogue on our magnitude distribution, we correct the \( K \)-band magnitude distribution using the completeness curve for the parent UDS DR8 catalogue from Mortlock et al. (2015), and the results of applying this correction are overlaid as the black filled circles, which are consistent with the original results. In the bottom panel of Fig. 4, we show the corresponding mass distribution. We find that galaxies around SMGs have an average stellar mass of \( 2.2 \pm 0.1 \times 10^{10} \text{ M}_\odot \) around a factor of 2 lower than the value found by Wardlow et al. (2011, after accounting for the different initial mass function used in that work), though as noted above, that work assumed a standard \( H \)-band mass-to-light ratio for the stellar mass estimation, and was unable to account for the whole 850 \( \mu \)m source population, whilst we are likely biased towards lower stellar masses by including SMG neighbours in our calculation.

Figure 5. Left: the distribution of \( K_{\text{AB}} \leq 24.6 \) galaxies within the \( UVJ \) colour–colour space from Williams et al. (2009), which is designed to separate passive and star-forming galaxies. The histograms are normalized to the area around all 761 850 \( \mu \)m sources in our sample and coloured according to the log of the bin occupancy, with values ranging from 5 to 100 galaxies per bin. The right-hand panel shows the \( UVJ \) colours of those \( K < 24.6 \) galaxies within 12 arcsec of a CLS 850 \( \mu \)m source, with the same colour scale as the left-hand panel.

Since we derived rest-frame colours for every source in the UDS catalogue, we are able to apply the same technique to determine the locations of the excess galaxies around SMGs in the \( UVJ \) colour space. The results of doing so are shown in Fig. 5, in which the left-hand panel shows the distribution of galaxies in the UDS catalogue, with a dashed line to highlight the division between the regions occupied by star-forming and passive galaxies at \( z > 1 \), taken from Hartley et al. (2013). The colour of each bin is chosen to indicate the logarithm of the number of sources in that bin, ranging from 5 to 100 galaxies. In the right-hand panel, we show the corresponding colour–colour space for galaxies within 12 arcsec of a CLS 850 \( \mu \)m source, with the same colour scaling, highlighting that whilst the vast majority are classified as star forming on the basis of their rest-frame colours, a small subset (an excess of \( 92.8 \pm 23.7 \), or \( 8.0 \pm 2.1 \) per cent of the total galaxy excess) are consistent with being passive galaxies, in agreement with previous work on smaller samples of the SMGs themselves (Smail et al. 2004).

These results make a particularly interesting comparison with the results of Smith et al. (in preparation), who find that the average 850 \( \mu \)m flux density of \( z > 1 \), \( K \)-selected and \( UVJ \) star-forming galaxies increases with increasingly red \( (V - J)_{\text{rest}} \) colour, meaning that we expect many SMG neighbours to be submillimetre faint, and therefore very difficult to identify even using, e.g. sensitive high-resolution submillimetre observations with ALMA, and the same is of course true of the passive galaxies. This highlights the need for deep optical/near-infrared data alongside sensitive interferometry if we are to truly understand the relationship between the properties of SMG host galaxies and their environments.

4 DISCUSSION

4.1 Can we believe photometric redshifts for sources this faint?

In deriving these results, we have of course assumed some degree of reliability in the UDS photometric redshifts, which Hartley et al.

\(^{2}\) To convince ourselves that our background subtraction is working as expected, we inverted the data and replotted with the same colour scheme, revealing no significant galaxy excess.
(2013) showed to be very precise. However, this can only be verified using spectroscopy, which is extremely challenging to obtain for large samples of faint sources such as those studied here, and the lower-left panel of Fig. 4 indicates that ~39 per cent of the galaxies associated with 850 μm positions are fainter than $K = 23.0$, the limit of the UDS spectroscopic sample selection (Almaini et al., in preparation).

Meanwhile, some works in the literature (e.g. Simpson et al. 2014) have been reluctant to quote photometric redshifts for sources with formal detections (defined as having an SNR $> 3σ$) in fewer than four optical/near-infrared bands, taking the conservative view that photometric redshifts derived for such sources are unreliable. However, other works have shown the value of measurements for deriving galaxy parameters even when they are not formally significant (e.g. Smith et al. 2013, 2014), and this is one of the reasons why photometric redshift codes such as FAST (Kriek et al. 2009) and LEPHARE (Ilbert et al. 2006) ideally require measured flux densities rather than magnitudes in their input catalogues (since fluxes are symmetric even for faint sources, and allow noisy negative flux estimates to be naturally included in the fitting).

Since we have not directly identified any counterparts to the 850 μm sources, it is difficult to directly investigate the number of photometric detections for any individual source. However, we can repeat our analysis, this time including only those sources which have at least four $>3σ$ detections in the UDS catalogue, and see what difference this makes to our results. We are still able to detect at least one galaxy associated with every SMG (i.e. our estimate of $Q_0$ is still consistent with unity), and we still detect a significant excess of $1091.8 ± 66.8$ galaxies, as compared with $1154.6 ± 67.5$ when the full sample is considered. This change indicates that around 5 per cent of the galaxies around SMGs have fewer than four $3σ$ detections, but the significance of the change is $<1σ$, based on propagating the Poisson error bars. The average photometric redshift of the excess galaxies around SMGs in the remaining sample is $1.99 ± 0.03$, which is consistent with the original value of $2.05 ± 0.03$, whilst the UVJ passive fraction in the excess galaxies is unchanged. We therefore suggest that whilst this aspect may play a minor role, it does not give us any cause to mistrust our results more than we would any other study using photometric redshifts derived for such sources.

4.2 Do the brightest SMGs differ from the rest?

Several previous investigations (e.g. Ivison et al. 2002; Pope et al. 2005; Biggs et al. 2011; Chen et al. 2016) have found evidence for differences between the brightest and faintest SMGs in terms of their redshifts and the number of sources contributing to the single-dish 850 μm flux density (though Wardlow et al. 2011, found no evidence for this). In Fig. 6, we show how the median redshift varies as a function of the 850 μm flux density as red error bars, ranging from the full sample to increasingly brighter subsets (note that the leftmost error bar is derived based on the whole population, with increasingly smaller subsets as we move towards brighter SNRs). Also overlaid are results from other selected observational works in the literature (Chapman et al. 2005; Wardlow et al. 2011; Casey et al. 2013; Vieira et al. 2013; Simpson et al. 2014; Chen et al. 2016; Michałowski et al. 2017) along with their quoted error bars. We have also overlaid an estimate based on the latest measurements of the far-IR luminosity function using SCUBA-2 and ALMA data from Koprowski et al. (2017), shown as the dot–dashed light blue line) and theoretical results for bright ($>5$ mJy) SMGs based on the GALFORM semi-analytic model presented in Cowley et al. (2015), and based on the simulations from Hayward et al. (2013a).

We find strong evidence for a higher median redshift for increasingly bright 850 μm sources in our sample, with error bars estimated from the redshift histograms using the median statistics method of Gott et al. (2001). The average redshift ranges from $2.05 ± 0.03$ for the full SNR $> 3.5σ$ sample (equivalent to a 3 mJy selected sample due to the highly uniform 850 μm map) of 761 galaxies to $3.13 ± 0.18$ for the 45 sources brighter than $8σ$. This result becomes even clearer if we consider the values quoted by other surveys, although the scatter in those works is considerably larger due to the smaller sample sizes with the exception of Michałowski et al. (2017), who calculated the average redshift in bins of 850 μm flux density, including the ultrafaint sample from Koprowski et al. (2016); these values are also overlaid in Fig. 6. The brightest samples, including the $S_{1.1\text{mm}} > 4.2$ mJy SMGs from Koprowski et al. (2014) over the COSMOS field (corresponding to 850 μm flux $>9.8$ mJy assuming a standard dust SED) with $(z) = 3.44 ± 0.16$ and $S_{1.4\text{mm}} > 20$ mJy SMGs (corresponding to 850 μm fluxes $>100$ mJy) observed by the South Pole Telescope, which have an average redshift of 3.5 according to both Vieira et al. (2013) and Weiß et al. (2013) further underline this trend.

In Fig. 7, we show a similar analysis, this time calculating the average number of galaxies around 850 μm positions as a function of SNR/fux. Taking the error bars into account, the data reveal little or no evidence for an increasing number of excess $K < 24.6$ galaxies towards brighter flux densities. This is reassuring, since if brighter 850 μm sources had a significantly larger number of excess galaxies, this might have skewed our photometric redshift distribution given that brighter sources lie at higher redshifts (Fig. 6).

It is of course possible that strongly gravitationally lensed sources within the 850 μm selected population could artificially increase the number of excess galaxies observed around the 850 μm positions,
and previous works have shown that bright submillimetre flux selection can efficiently identify such systems (e.g. Negrello et al. 2010). However, according to the latest results, lensed systems only contribute appreciably to the 850 µm number counts above fluxes of 20 mJy (Geach et al. 2017), and there are no sources this bright in our sample. Since Fig. 7 also shows little evidence for an increase in the average number of excess galaxies for the brightest sources in our sample, we therefore suggest that strongly lensed galaxies do not have a major influence on our results.

4.3 The contribution of SMGs to cosmic star formation

In Fig. 8, we bring all of this information together, in an attempt to measure the contribution of SMGs to the cosmic star formation rate density (SFRD) estimates from Behroozi et al. (2013) and Madau & Dickinson (2014), shown as the dashed and dotted lines, respectively. The raw SFRs scaled from the individual deboosted 850 µm flux densities from Geach et al. (2017), assuming an isothermal far-IR SED with $T = 30$ K and $\beta = 1.5$, are shown as the blue crosses, whilst the red asterisks show the results when we fold the Geach et al. (2017) estimates of completenesses and false detection rates for each 850 µm source into our calculations. The error bars are derived by measuring the 16th and 84th percentiles of the results of performing 1000 Monte Carlo realizations of sampling from the redshift distributions shown in Fig. 4. Also overlaid are the results from the 4 mJy SMG sample from Michałowski et al. (2017), shown as grey diamonds, the obscured contribution estimated from ALMA observations of the Hubble Ultra Deep Field by Dunlop et al. (2017), shown as grey triangles, and the SFRD estimated based on radiative transfer modelling of Herschel galaxies from Rowan-Robinson et al. (2016), shown as light grey squares.

Figure 7. The variation in the average number of $K < 24.6$ galaxies within 12 arcsec of an 850 µm position, as a function of the 850 µm SNR. As in Fig. 6, the upper horizontal axis has been converted to 850 µm flux density assuming a uniform uncertainty of 0.88 mJy, the median value of sources in the CLS 850 µm map of the UDS field.

Figure 8. The contribution of the CLS 850 µm sources to the cosmic SFRD calculated by Behroozi et al. (2013) and Madau & Dickinson (2014), shown as the dashed and dotted lines, respectively. The raw SFRs scaled from the individual deboosted 850 µm flux densities from Geach et al. (2017), assuming an isothermal far-IR SED with $T = 30$ K and $\beta = 1.5$, are shown as the blue crosses, whilst the red asterisks show the results when we fold the Geach et al. (2017) estimates of completenesses and false detection rates for each 850 µm source into our calculations. The error bars are derived by measuring the 16th and 84th percentiles of the results of performing 1000 Monte Carlo realizations of sampling from the redshift distributions shown in Fig. 4. Also overlaid are the results from the 4 mJy SMG sample from Michałowski et al. (2017), shown as grey diamonds, the obscured contribution estimated from ALMA observations of the Hubble Ultra Deep Field by Dunlop et al. (2017), shown as grey triangles, and the SFRD estimated based on radiative transfer modelling of Herschel galaxies from Rowan-Robinson et al. (2016), shown as light grey squares.

Figure 8. The contribution of the CLS 850 µm sources to the cosmic SFRD calculated by Behroozi et al. (2013) and Madau & Dickinson (2014), shown as the dashed and dotted lines, respectively. The raw SFRs scaled from the individual deboosted 850 µm flux densities from Geach et al. (2017), assuming an isothermal far-IR SED with $T = 30$ K and $\beta = 1.5$, are shown as the blue crosses, whilst the red asterisks show the results when we fold the Geach et al. (2017) estimates of completenesses and false detection rates for each 850 µm source into our calculations. The error bars are derived by measuring the 16th and 84th percentiles of the results of performing 1000 Monte Carlo realizations of sampling from the redshift distributions shown in Fig. 4. Also overlaid are the results from the 4 mJy SMG sample from Michałowski et al. (2017), shown as grey diamonds, the obscured contribution estimated from ALMA observations of the Hubble Ultra Deep Field by Dunlop et al. (2017), shown as grey triangles, and the SFRD estimated based on radiative transfer modelling of Herschel galaxies from Rowan-Robinson et al. (2016), shown as light grey squares.
work; however, this will be possible in future studies with complete x-IDs and fully debianded Herschel photometry.

Also overlaid in Fig. 8 are the SFRDs calculated based on the obscured star formation attributed to galaxies with SFR > 300 M$_{\odot}$ yr$^{-1}$ from Michalowski et al. (2017, grey diamonds), and the corresponding larger estimate of the total dust-obscured star formation based on the latest ALMA results (reconstructing the relationship between obscured SFR and stellar mass) using deep observations of the Hubble Ultra Deep Field by Dunlop et al. (2017, grey triangles). The comparison between the latter study and our results is of particular interest, since together the two studies imply that roughly 60 per cent of the total dust-enshrouded star formation over $1 < z < 4.5$ is captured in SMGs detected by SCUBA-2.

We have also included the results derived based on a 500 µm selected sample of sources obtained from 20 deg$^2$ of Herschel data from Rowan-Robinson et al. (2016, as light grey squares), although as the authors note, they detect a far larger cosmic SFRD than in other works. This discrepancy underlines the power of SCUBA-2 for high-redshift studies.

Finally, we note that despite the uncertainty over the far-IR SED, we have reached our plausible results with the largest SMG sample with complete redshift information, and having naturally accounted for the fraction of 850 µm sources which are broken up into multiple components in high-resolution submillimetre interferometry, even if those components are widely separated in redshift.

5 CONCLUSIONS

We have used 761 850 µm sources from the SCUBA-2 CLS observations of the UKIDSS UDS field (Geach et al. 2017) alongside the 8th UDS data release (Hartley et al. 2013) to statistically measure the complete redshift distribution of 850 µm sources. Whilst our method is unable to identify the counterparts to individual SMGs, we are able to account for multiple neighbouring galaxies and for the distribution of background sources using the rich ancillary data that are available over this field.

We find that the UDS data are sufficiently deep to detect excess galaxies around virtually all of the 761 850 µm sources in our survey area, though we do not identify them, and they are not necessarily the same galaxies responsible for the 850 µm emission (some of which are too faint to be detected even in the UDS K-band data; Simpson et al. 2016). Simulations based on model lightcones suggest that the redshift distribution of excess galaxies within 12 arcsec of an 850 µm source is indistinguishable from that of the SMGs themselves, and we use this fact to measure their complete optical/near-infrared photometric redshift distribution, finding that the SMGs in our $>3.5\sigma$ sample have an average redshift of 2.05 ± 0.03. We find an average of 1.52 ± 0.09 excess galaxies within 12 arcsec of an 850 µm position, and use the rest-frame colours of these galaxies to show that whilst the majority are likely to be dusty star-forming galaxies, a substantial fraction are likely to be submillimetre faint, passive galaxies, and therefore difficult to detect even in deep high-resolution submillimetre imaging, e.g. with ALMA. We show that the brightest SMGs lie at higher redshifts than the rest of the SMG population, and under the assumption of a universal isothermal far-IR dust spectral energy distribution, we find that SMGs contribute around 30 per cent of the cosmic SFRD between 0.5 < z < 5.0.

ACKNOWLEDGEMENTS

The authors would like to thank the referee, Jim Dunlop, for useful suggestions that have increased the quality of this paper. We are also grateful to Will Hartley and Omar Almaini for providing the UDS catalogue used in this work, as well as to Kristen Coppin, Jim Geach and Wendy Williams for valuable discussions. The James Clerk Maxwell Telescope has historically been operated by the Joint Astronomy Centre on behalf of the Science and Technology Facilities Council of the United Kingdom, the National Research Council of Canada and the Netherlands Organisation for Scientific Research. The James Clerk Maxwell Telescope is operated by the East Asian Observatory on behalf of The National Astronomical Observatory of Japan, Academia Sinica Institute of Astronomy and Astrophysics, the Korea Astronomy and Space Science Institute, the National Astronomical Observatories of China and the Chinese Academy of Sciences (Grant No. XDB09000000), with additional funding support from the Science and Technology Facilities Council of the United Kingdom and participating universities in the United Kingdom and Canada. Additional funds for the construction of SCUBA-2 were provided by the Canada Foundation for Innovation. The UKIDSS project is defined in Lawrence et al. (2007) and uses the UKIRT Wide Field Camera (Casali et al. 2007). The Flatiron Institute is supported by the Simons Foundation.

REFERENCES

Barger A. J., Wang W.-H., Cowie L. L., Owen F. N., Chen C.-C., Williams J. P., 2012, ApJ, 761, 89
Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57
Bessell M. S., 1990, PASP, 102, 1181
Blain A. W., Smail I., Ivison R. J., Kneib J.-P., Frayer D. T., 2002, Phys. Rep., 369, 111
Bradley G., Charlot S., 2003, MNRAS, 344, 1000
Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682
Casali M. et al., 2013, MNRAS, 436, 1919
Chabrier G., 2003, PASP, 115, 763
Chapman S. C., Helou G., Lewis G. F., Dale D. A., 2003, ApJ, 588, 186
Chapman S. C., Blain A. W., Smail I., Ivison R. J., 2005, ApJ, 622, 772
Chen C.-C., Cowie L. L., Barger A. J., Casey C. M., Lee N., Sanders D. B., Wang W.-H., Williams J. P., 2013, ApJ, 776, 131
Chen C.-C. et al., 2016, ApJ, 820, 82
Condon J. J. et al., 2012, ApJ, 758, 23
Cowley W. I., Lacey C. G., Baugh C. M., Cole S., 2015, MNRAS, 446, 1784
Davé R., Finlator K., Oppenheimer B. D., Fardal M., Katz N., Kereš D., Weinberg D. H., 2010, MNRAS, 404, 1355
Dunlop J. S. et al., 2017, MNRAS, 466, 861
Fleuren S. et al., 2012, MNRAS, 423, 2407 (F12)
Fujimoto S., Ouchi M., Ono Y., Shibuya T., Ishigaki M., Nagai H., Momose R., 2016, ApJS, 222, 1
Fumagalli M. et al., 2016, ApJ, 822, 1
Geach J. E. et al., 2017, MNRAS, 465, 1789
Gott J. R., III, Vogeley M. S., Podariu S., Ratra B., 2001, ApJ, 549, 1
Hartley W. G. et al., 2013, MNRAS, 431, 3045
Hatch N. A. et al., 2014, MNRAS, 445, 280
Hayward C. C., Kereš D., Jonsson P., Narayanan D., Cox T. J., Hernquist L., 2011, ApJ, 743, 159
Hayward C. C., Narayanan D., Kereš D., Jonsson P., Hopkins P. F., Cox T. J., Hernquist L., 2013a, MNRAS, 428, 2529
Hayward C. C., Behroozi P. S., Somerville R. S., Primack J. R., Moreno J., Wechsler R. H., 2013b, MNRAS, 434, 2572
Helou G., Soifer B. T., Rowan-Robinson M., 1985, ApJ, 298, L7
Hildebrand R. H., 1983, QJRAS, 24, 267
Hodge J. A. et al., 2013, ApJ, 768, 91
Holland W. S. et al., 1999, MNRAS, 303, 659
