A robust sample of submillimetre galaxies: constraints on the prevalence of dusty, high-redshift starbursts

R. J. Ivison†

Astronomy Technology Centre, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ
Institute for Astronomy, University of Edinburgh, Blackford Hill, Edinburgh EH9 3HJ

Abstract. We develop and apply a dual-survey extraction technique to published submm images of the Lockman Hole using the SCUBA and MAMBO bolometer arrays. Cut above 5σ, this catalogue of submm galaxies (SMGs) is significantly more robust than previous samples, typically selected above 3.0–3.5σ, with a much-reduced likelihood of real, faint SMGs (beneath and around the confusion limit) entering the catalogue via superposition with noise. Our selection technique yields 19 SMGs of which we expect at most two to be due to chance superposition of SCUBA and MAMBO noise peaks. The flux limit of the sample (~5 mJy at 0.85 mm), which is sensitive to luminous, dusty galaxies at extreme redshifts, is optimally matched to a deep 1.4-GHz image (4.6 μJy beam⁻¹ rms) which probes starbursts in the z < ~4 regime. A high fraction of these robust SMGs (~80%) have radio counterparts which, given the probable ~10% contamination by spurious sources, suggests that very distant SMGs (z > ~3–4) are unlikely to make up more than ~10% of the bright SMG population. Finally, the accurate radio positions are used to test the accuracy of positions determined in the submm, with important implications for upcoming surveys with SCUBA-2.

INTRODUCTION

Submm surveys pick up dusty objects at extreme redshifts, galaxies that drop out of surveys at shorter and longer wavelengths due to obscuration and unfavourable K corrections, respectively. The first cosmological surveys using SCUBA and MAMBO quickly and radically changed the accepted picture of galaxy formation and evolution, moving away from the optocentric view of the last century. The discovery of so-called ‘SCUBA galaxies’ [28] was greeted with surprise due to the remarkable evolution in the dusty, starburst galaxy population implied by such a large source density at the flux levels accessible to the first generation of bolometer arrays [2]. Excitement was replaced by pessimism with the first efforts to study SMGs at optical/IR wavelengths: early reports, backed by a study in the HDF-N [16], suggested that the majority of the submm population had no plausible optical counterparts. Attention was diverted to various redshift engines and broadband photometric techniques [1, 37]. As a result, only a handful of detailed studies were attempted, often for extreme, non-representative galaxies.

Recent progress has largely been the result of radio imaging of submm survey fields, though early radio follow-up data were not well matched to the depth of the submm imaging and only modest detection rates were achieved. Once sufficient depths were attained [19], radio maps were able to pinpoint half of the brightest known SMGs to ~0.3″ and, combined with the submm flux density, provide a rough estimate of
redshift. Radio data also enabled some refinement of submm samples, increasing the detection fraction to two thirds of SMGs at 0.85-mm flux levels in excess of $\sim$5 mJy. With positions in hand, these bright SMGs were found to be a diverse population — some quasar-like, with broad lines and X-ray detections [17, 21], some morphologically complex [18], some extremely red [29, 19, 36, 9], some with the signatures of obscured AGN and/or superwinds [32].

Spectroscopic redshifts have been extremely difficult to determine, requiring a significant investment of time on 10-m telescopes with highly efficient spectrometers. The first systematic survey was undertaken by Chapman et al. [5, 33, 7]. The median redshift is now known to be $\sim$2.2 for $S_{0.85\mathrm{mm}} \geq 5$ mJy galaxies selected using SCUBA and pinpointed at 1.4 GHz. A similar survey with a modern red-sensitive spectrometer has yet to be attempted. We must therefore accept that present knowledge may be biased by wavelength coverage and by the need to identify targets via radio imaging (which does not benefit from the same $K$ correction as submm observations and is thought to impose an artificial redshift cut-off at $z \sim 3–4$).

Accurate redshifts have facilitated the first systematic measurements of molecular gas mass for SMGs ($\sim10^{11} M_\odot$) via observations of CO [24, 15], as well as constraints on gas reservoir size and dynamical mass [34]. The results agree with earlier work on what were thought to be less representative systems [11, 12], suggesting that SMGs are massive systems and providing some of the strongest tests of galaxy-formation models to date [13].

Despite this progress, a detailed understanding of SMGs remains a distant goal. Confusion currently limits our investigations to the brightest quartile of the submm population (though surveys through lensing clusters have provided a handful of sources more typical of the faint population that dominates the cosmic background [31]). We must also recall that selection biases have potentially skewed our understanding — limited coverage of red and IR wavelengths in spectroscopic surveys, and a third of SMGs remain undetected in the radio and thus untargeted by existing spectroscopic campaigns.

Here, we present a robust sample of bright SMGs selected using SCUBA and MAMBO in one of the ‘8-mJy Survey’ regions: the Lockman Hole. Our goal is to provide a bright sample which we would expect to detect in well-matched radio imaging ($\sigma_{0.85\mathrm{mm}}/\sigma_{1.4\mathrm{GHz}} \sim 500$) whilst minimising, so far as is practicable, the possibility that sources are spurious or anomalously bright. We may thus determine the true fraction of radio drop-outs ($z > 4$) amongst SMGs, as well as practical information such as the intrinsic positional uncertainty for SMGs in the absence of radio/IR counterparts.

**SAMPLE SELECTION**

Our objective here is to avoid one of the most common criticisms levelled at submm surveys — the potentially large number of spurious sources present in submm catalogues. This has limited our ability to address the true recovery fraction in the radio, and hence the corrections that must be made to the redshift distributions that are used to determine star-formation histories and galaxy-formation models.
Greve et al. [14] presented a 1.2-mm survey of the ELAIS N2 and Lockman Hole regions, centred on the coordinates mapped at 0.85 mm by SCUBA in the ‘8-mJy Survey’ [26]. Mortier et al. [23] present a refined analysis of SCUBA data for a subset of the Lockman Hole MAMBO image.

Greve et al. argue that several maps with low signal-to-noise ratio (SNR) of the same region, with only marginal differences in frequency, produce several visualisations of essentially the same sky, tracing the same population of luminous, dusty galaxies. We adopt this philosophy here, starting from the 1.2-mm catalogue of Greve et al., looking for matches at 0.85 mm from Mortier et al., substituting SMGs from Scott et al. where blends are evident in the Mortier et al. catalogue.

Existing surveys have typically employed a SNR threshold of 3.0–3.5 and false detections are dominated by ‘flux boosting’, possibly at the 10–40% level [26]. To realise a significant improvement, reducing the probability of an SMG being spurious, with a lower chance of flux boosting, we must limit our sample to higher SNRs. Initially, we take those sources that have a combined statistical significance equivalent to a signal-to-noise ratio (SNR) in excess of 4.5 in a map with Gaussian noise properties, leaving aside the potential clustering of SMGs [3]. We restrict ourselves to the Lockman Hole because these data are of better quality. We then refine the selection criteria by comparing the properties of this sample with catalogues of spurious sources generated by the introduction of arbitrary positional offsets.

To accomplish this dual-survey extraction, where there are numerous permutations of SNRs that yield the desired SNR, we have developed a simple algorithm. We take the
FIGURE 2. Flux density versus redshift for an Arp 220 spectral energy distribution, at 5× Arp 220’s luminosity. The variation with redshift is shown for flux densities at 1 mm and 1.4 GHz (solid and dotted lines, respectively), with the 5-σ detection limits indicated as dashed horizontal lines. As is widely known, mm and submm surveys are sensitive to dusty starbursts out to extreme redshifts (z ∼ 6) but the deepest current radio surveys lose sensitivity to these galaxies at z > 3–4.

2-σ catalogue from the MAMBO image, extracted as described by Greve et al. [14]. We then search for SCUBA sources within 14″ (roughly the area of five beams), applying a correction for separation and checking for a combined significance above my initial threshold, 4.5σ. We include sources that exceed 4.5σ in either dataset, as long as there is a valid reason why the source is not seen in the other image, and sources with a combined probability exceeding 4.5σ with a minimum 2-σ detection in both images, within 14″ of one another. To account for the distance between sources, we divide the probabilities by the number of beams inside a circle with a radius equal to the separation of the sources.

The noise properties of submm images are not Gaussian due to the miriad of real, faint sources near the confusion limit. The most important effect of increasing the SNR threshold should be to drastically reduce the false detections due to faint SMGs that have been boosted above the detection threshold by noise — flux boosting [26, 14]. However, with so many ≥2-σ sources in the catalogues, it is important to investigate how many SMGs may contaminate the combined sample due to random coincidence of faint MAMBO and SCUBA peaks. To this end, we performed simulations, offsetting the MAMBO sample by ±30″ and ±45″ (in R.A. and Dec.). Each simulation typically yields ten SMGs — far too many to hope that a statistically robust catalogue would emerge from the process. The results of these simulations, and for the real samples, are illustrated in the left-hand panels of Fig. 1. The left-hand panels show the scatter of R.A. and Dec. offsets between each MAMBO and SCUBA pair, with symbol size
coded to match the SNR of the highest peak. The right-hand panels show histograms of $\sqrt{\Delta \alpha^2 + \Delta \delta^2}$ — the radial offset, $r$. The simulated catalogues clearly yield a very different distribution of offsets from the position-matched data, exactly the $n \propto r^2$ form expected for randomly scattered peaks, and without the expected concentration of high-SNR pairs at low $r$.

Using the information in these plots we have taken a number of approaches to minimise the number of spurious SMGs in our sample. Most of the sources in the simulated catalogues are the result, perhaps unsurprisingly, of near-coincident low-SNR peaks. We found that raising the minimum catalogue threshold to $\geq 2.5\sigma$ reduced the number of false detections whilst having no effect on the real catalogue. Lowering the search radius to $11''$ reduced the number of false positives in line with the ratio of the respective search areas, removing only one source from the real catalogue. Insisting that the higher of the two peaks is $\geq 3.5\sigma$ removes a further quarter of the simulated SMGs, whilst reducing the real catalogue by half that amount. Finally, we increased the combined threshold to $\geq 5\sigma$ to further reduce the number of rogue sources. The effects of this approach on the SCUBA–MAMBO positional offsets are shown in the right-hand panels of Fig. 1. We are left with a situation where the final catalogue consists of 19 SMGs of which we expect at most two to be the result of coincidence. It is clear that the effect of confusion is less apparent than in previous catalogues, without the sample size shrinking dramatically, though the sample is not entirely immune. We list the resulting 19 sources in Table 1.

Note that all five of the SCUBA sources detected individually above $5\sigma$ are also found in the MAMBO image with a SNR of at least 2.0, though one is below $2.5\sigma$ and has thus been excised from my final sample. Of the eight $\geq 5\sigma$ MAMBO sources, only one was not seen by SCUBA at $\geq 2.5\sigma$: LE-1200.4, which lies off the SCUBA map, to the East, but was subsequently detected by Scott et al. (in preparation).

IDENTIFICATION OF THE SMGS IN THE RADIO

The process of identifying counterparts to SMGs has been refined in a series of studies [17, 18, 19, 20, 29, 30, 35, 36, 25]. The most effective methods employ a combination of radio, optical and IR imaging, searching for the red rest-frame optical light expected of a dust-enshrouded galaxy and the synchrotron emission expected of a starburst or radio-loud active galaxy. Radio sources and EROs are sufficiently rare that associations of either with SMGs can be viewed as robust in most cases.

Fig. 2 shows how the flux density of a luminous, dusty starburst varies with redshift at both 1 mm and 1.4 GHz, adopting the SED of Carilli & Yun [4]. To identify a $z \leq 3$ 5-mJy SMG (the average of the 0.85- and 1.2-mm flux densities) in a 1.4-GHz image, it is clear that the radio data must achieve a sensitivity of at least 5 $\mu$Jy beam$^{-1}$.

Even at this radio sensitivity, the lack of a robust radio identification could have at least five origins: i) the SMG could be spurious; ii) the SMG could be flux boosted significantly; iii) the radio/far-IR emission could be significantly larger than the 1.4-GHz synthesised beam [19]; iv) the characteristic dust temperature could be low [6]; or, most interestingly, v) the SMG could lie at $z \gtrsim 3$–4 [10]. It is difficult, given the quality
FIGURE 3. Postage stamps (30′ × 30′) of the fields surrounding the first nine >5σ SMGs in the Lockman Hole. Optical (R) data are shown as a grayscale upon which 1.4-GHz contours are plotted. Open crosses mark SCUBA galaxies; solid crosses mark MAMBO galaxies; 8′-radius circles mark the average positions; dashed lines represent the position of GMOS, LRIS and NIRSPEC slits [5, 33, 7]; squares mark radio counterparts with $P < 0.05$, solid squares being the most probable in each case.

of existing samples, and the mis-matched radio/submm datasets, to determine which of these are important, though Ivison et al. [19] showed that (i) affects at least 15% of a 3.5σ sample.

Deep, high-resolution, wide-field radio images were obtained at the Very Large Array and employed to pinpoint the SMGs. The data used here are described in detail by Ivison et al. [19]. Fig. 3 shows postage stamps around the positions given in Table 2 (see the printed version of this document) for the nine most significant SMGs. A radio source peaking at $\geq 4 \sigma$ in the 1.4′′ or smoothed images, with an integrated flux density in excess of 15 $\mu$Jy, is considered a robust detection. Fainter sources, where the definition
is relaxed to only the integrated flux, were also catalogued.

For each SMG we have searched for a potential radio (1.4-GHz) counterpart out to a radius of $8''$ from the mid-point of the 0.85- and 1.2-mm emission. This relatively large search area ($201''^2$ around each source) represents a ~99% positional confidence region (see later) and should ensure that few real associations are missed.

To quantify the formal significance of each of the potential submm/radio associations, we have used the method of Downes et al. [8] to correct the raw Poisson probability, $P$, that a radio source of the observed flux density could lie at the observed distance from the SMG for the number of ways that such an apparently significant association could have been uncovered by chance.

Of the 19 SMGs in the sample, only four lack robust radio counterparts. The flux densities and positions of all candidate radio counterparts are listed in Table 2 (see the printed version of this document), along with the search positions. Of the four sources which have more than one potential radio counterpart, we find that the correct identification is never statistically obvious. The formal probability of the second candidate association occurring by chance is low, $P \leq 0.05$. The obvious interpretations of such multiple statistical associations are either gravitational lensing (implausible in most of the cases here), or clustering of star-forming objects/AGN at the source redshift (already proven in several cases — [22, 7]); another possibility is that sources with multiple radio counterparts are boosted into bright submm catalogues by virtue of comprising multiple, faint, physically unrelated SMGs, i.e. by conventional confusion.

This calculation has yielded statistically robust radio counterparts for 15 of the 19 SMGs. The plausibility of this figure can be checked by noting the ratio of areas inside and outside the circles in Fig. 3, 4.5:1. Ten random ‘field’ radio sources are detected in the outer areas so, given that $\mu$Jy radio sources are expected to be over-dense around SMGs, we expect only a handful (at most 2–3) of the counterparts to be spurious. Based on the $P$ values given in Table 2 (see the printed version of this document), the most likely candidates for spurious associations are LH-1200.007 and LH-1200.019, leaving 15/19 robust radio identifications.

**Implications for submm positional uncertainty**

Fig. 4 shows a histogram of R.A. and Dec. positional offsets between the mm/submm centroids and the most likely radio counterparts. A Gaussian fit to the distribution yields a FWHM of $5.2 \pm 1.2''$, which translates into an r.m.s. separation of $5.2''/2.354 = 2.2''$ between the (sub)mm and radio positions. The sample can thus be employed to re-calibrate the rule-of-thumb relationship between positional accuracy, beam size and SNR (typically, SNR $\sim 6$ here). We must acknowledge a mild circularity to the logic, given that positional offsets have been used to calculate $P$, though this may be offset by the lack of correction for radio sources in the field.

The conventional positional uncertainty, $\sigma_r$, occurs where the distribution of radial offsets peaks; this is the same as the the r.m.s. separation deduced earlier. Within this radius we expect to find 39.3% of the population, with 86.5 and 98.9% within 2 $\sigma_r$ and 3 $\sigma_r$ (from $1 - e^{-r^2/2\sigma_r^2}$). In the absence of radio counterparts, it would seem from our
analysis that around 39% of SMGs can be located within a radial distance of $\sim \theta / \text{SNR}$, where $\theta$ is the FWHM beam size, in arcsec). Thus the 1-σ$_r$ positional uncertainty for 3–3.5σ SCUBA-selected SMGs is 4–5′′, cf. the rule of thumb quoted by Hughes et al. [16]. The most secure SMGs, at $\sim 10\sigma$, representative of those expected in upcoming, confusion-limited, wide-field (tens of square degrees) 0.85-mm surveys using SCUBA-2, with lower significance, simultaneous detections at 0.45 mm, will be located with a uncertainty of $\sigma_r \sim 1''$, at which level the precision of the telescope pointing and SCUBA-2 flatfield may become important. Assuming these sources of uncertainty can be minimised, the current requirement for deep radio coverage to identify counterparts and enable spectroscopic follow-up may not be as urgent, particularly if $3'' \times 3''$ deployable integral field units are employed (e.g. KMOS [27]).

ACKNOWLEDGMENTS

Many thanks to my trusty collaborators: Thomas Greve, James Dunlop, Mark Swinbank, Jason Stevens, Angela Mortier, Steve Serjeant, Frank Bertoldi, Ian Smail, Andrew Blain and Scott Chapman.
REFERENCES

26. Scott S.E. et al. 2002, 331, 817
<table>
<thead>
<tr>
<th>Name</th>
<th>Position at 0.85mm</th>
<th>Position at 1.2mm</th>
<th>S/N</th>
<th>S/N Sep Final</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \alpha_{J2000} ) h m s</td>
<td>( \delta_{J2000} ) ° ′ ′′</td>
<td>( S_{0.85\text{mm}} ) mJy</td>
<td>( \delta_{1.2\text{mm}} ) ° ′ ′′</td>
</tr>
<tr>
<td>LH-1200.001 = LE850.02</td>
<td>10:52:38.6 +57:24:38</td>
<td>10.9 ± 2.1</td>
<td>5.1</td>
<td>10:52:38.3 +57:24:37</td>
</tr>
<tr>
<td>LH-1200.002</td>
<td>10:52:38.6 +57:23:19</td>
<td>5.2 ± 2.0</td>
<td>2.6</td>
<td>10:52:38.8 +57:23:21</td>
</tr>
<tr>
<td>LH-1200.003 = LE850.14</td>
<td>10:52:04.2 +57:27:01</td>
<td>10.5 ± 2.0</td>
<td>5.2</td>
<td>10:52:04.1 +57:26:57</td>
</tr>
<tr>
<td>LH-1200.004</td>
<td>10:52:03.9 +57:27:10</td>
<td>1.3 ± 0.6</td>
<td>2.2</td>
<td></td>
</tr>
<tr>
<td>LH-1200.005 = LE850.01</td>
<td>10:52:01.5 +57:24:43</td>
<td>8.6 ± 1.1</td>
<td>8.0</td>
<td>10:52:01.3 +57:24:48</td>
</tr>
<tr>
<td>LH-1200.006 = LE850.16</td>
<td>10:52:27.6 +57:25:17</td>
<td>7.0 ± 1.6</td>
<td>4.3</td>
<td>10:52:27.4 +57:25:15</td>
</tr>
<tr>
<td>LH-1200.007</td>
<td>10:52:03.4 +57:18:12</td>
<td>5.8 ± 1.3</td>
<td>4.3</td>
<td>10:52:04.7 +57:18:12</td>
</tr>
<tr>
<td>LH-1200.008</td>
<td>10:51:40.7 +57:19:53</td>
<td>4.6 ± 1.6</td>
<td>2.8</td>
<td>10:51:41.9 +57:19:51</td>
</tr>
<tr>
<td>LH-1200.009</td>
<td>10:52:28.2 +57:22:17</td>
<td>6.3 ± 1.8</td>
<td>3.5</td>
<td>10:52:27.5 +57:22:20</td>
</tr>
<tr>
<td>LH-1200.010 = LE850.06</td>
<td>10:52:30.4 +57:22:13</td>
<td>10.8 ± 2.4</td>
<td>4.6</td>
<td>10:52:29.9 +57:22:05</td>
</tr>
<tr>
<td>LH-1200.011 = LE850.03</td>
<td>10:51:58.3 +57:18:00</td>
<td>5.0 ± 1.3</td>
<td>3.9</td>
<td>10:51:58.2 +57:17:53</td>
</tr>
<tr>
<td>LH-1200.012 = LE850.18</td>
<td>10:51:55.9 +57:23:13</td>
<td>4.3 ± 1.0</td>
<td>4.2</td>
<td>10:51:55.4 +57:23:10</td>
</tr>
<tr>
<td>LH-1200.014 = LE850.08</td>
<td>10:51:59.6 +57:24:21</td>
<td>4.8 ± 1.1</td>
<td>4.3</td>
<td>10:52:00.0 +57:24:24</td>
</tr>
<tr>
<td>LH-1200.017</td>
<td>10:51:22.5 +57:18:42</td>
<td>14.6 ± 4.1</td>
<td>3.5</td>
<td>10:51:21.4 +57:18:40</td>
</tr>
<tr>
<td>LH-1200.019</td>
<td>10:51:27.8 +57:19:47</td>
<td>6.7 ± 2.3</td>
<td>2.9</td>
<td>10:51:28.3 +57:19:46</td>
</tr>
<tr>
<td>LH-1200.022</td>
<td>10:52:03.3 +57:15:37</td>
<td>7.3 ± 2.2</td>
<td>3.3</td>
<td>10:52:03.0 +57:15:46</td>
</tr>
<tr>
<td>LH-1200.042 = LE850.29</td>
<td>10:52:16.2 +57:25:05</td>
<td>7.2 ± 1.5</td>
<td>4.7</td>
<td>10:52:16.0 +57:25:06</td>
</tr>
<tr>
<td>LH-1200.096 = LE850.07</td>
<td>10:51:51.4 +57:26:38</td>
<td>6.7 ± 1.7</td>
<td>4.0</td>
<td>10:51:51.4 +57:26:40</td>
</tr>
<tr>
<td>LH-1200.104 = LE850.27</td>
<td>10:51:53.9 +57:18:38</td>
<td>6.3 ± 1.2</td>
<td>5.2</td>
<td>10:51:53.7 +57:18:39</td>
</tr>
</tbody>
</table>