# The HASHTAG project: The First Submillimeter Images of the Andromeda Galaxy from the Ground.

Matthew W. L. Smith [ID],[1] Stephen A. Eales [ID],[1] Thomas G. Williams [ID],[2] Bumhyun Lee [ID],[3,4] Zongnan Li [ID],[5,6] Pauline Barmby [ID],[7] Martin Bureau [ID],[8] Scott Chapman,[9] Brian S. Cho,[10] Aeree Chung,[11] Eun Jung Chung [ID],[12] Hui-Hsuan Chung,[13] Christopher J. R. Clark [ID],[14] David L. Clements [ID],[15] Timothy A. Davis [ID],[1] Ilse De Looze [ID],[16,17] David J. Eden [ID],[18] Gayathri Athikkat-Eknath [ID],[1] George P. Ford,[1] Yu Gao [ID],[19,20] Walter Gear [ID],[21] Haley L. Gomez [ID],[1] Richard de Grijs [ID],[22,23] Jinhua He [ID],[24,25,26] Luis C. Ho [ID],[3,27] Thomas M. Hughes [ID],[25,28,29,30] Sihan Jiao,[31,32] Zhiyuan Li [ID],[5,6] Francisca Kemper [ID],[33,34] Florian Kirchschlager [ID],[17] Eric W. Koch [ID],[35] Albert K. H. Kong [ID],[13] Chien-Hsiu Lee [ID],[36] En-Tzu Lin [ID],[13] Steve Mairs [ID],[37] Michał J. Michałowski [ID],[38] Kate Pattle [ID],[21] Yingjie Peng,[3] Sarah E. Ragan [ID],[1] Mark G. Rawlings [ID],[37] Dimitra Rigopoulou [ID],[8] Amelie Saintonge [ID],[17] Andreas Schruba,[39] Xindi Tang [ID],[40] Junfeng Wang [ID],[19] Anthony P. Whitworth,[1] Christine D. Wilson [ID],[41] Kijeong Yim [ID],[12] and Ming Zhu[31]

[1]School of Physics & Astronomy, Cardiff University, The Parade, Cardiff, CF24 3AA, UK.
[2]Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117, Heidelberg, Germany.
[3]Kavli Institute for Astronomy and Astrophysics, Peking University, Beijing 100871, China.
[4]Department of Astronomy, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul, 03722, South Korea
[5]School of Astronomy and Space Science, Nanjing University, Nanjing 210023, China.
[6]Key Laboratory of Modern Astronomy and Astrophysics, Nanjing University, Nanjing 210023, China.
[7]Department of Physics & Astronomy, Institute for Earth & Space Exploration, University of Western Ontario, 1151 Richmond Street, London, Ontario, Canada.
[8]Sub-department of Astrophysics, Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford, OX1 3RH, UK.
[9]Dalhousie University, Halifax NS, Canada
[10]Department of Physics and Astronomy, Seoul National University, Gwanak-gu, Seoul 08826, Republic of Korea.
[11]Department of Astronomy, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul, 03722, South Korea.
[12]Department of Astronomy and Space Science, Chungnam National University, 99 Daehak-ro, Yuseong-gu, Daejeon 34134, Republic of Korea.
[13]Institute of Astronomy, National Tsing Hua University, Hsinchu, 30013, Taiwan.
[14]Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland 21218, USA.
[15]Astrophysics Group, Physics Department, Blackett Laboratory, Imperial College London, Prince Consort Road, London SW7 2AZ, UK.
[16]Sterrenkundig Observatorium, Ghent University, Krijgslaan 281 – S9, B-9000 Ghent, Belgium.
[17]Department of Physics and Astronomy, University College London, London WC1E 6BT, UK.
[18]Astrophysics Research Institute, Liverpool John Moores University, IC2, Liverpool Science Park, 146 Brownlow Hill, Liverpool, L3 5RF, UK.
[19]Department of Astronomy, Xiamen University, Xiamen, Fujian, 361005, People's Republic of China.
[20]Purple Mountain Observatory & Key Laboratory for Radio Astronomy, Chinese Academy of Sciences, 10 Yuanhua Road, Nanjing 210023, People's Republic of China.
[21]Centre for Astronomy, Department of Physics, National University of Ireland Galway, University Road, Galway H91 TK33, Ireland.
[22]Department of Physics & Astronomy, Macquarie University, Balaclava Road, Sydney, NSW 2109, Australia.
[23]Research Centre for Astronomy, Astrophysics and Astrophotonics, Macquarie University, Balaclava Road, Sydney, NSW 2109, Australia.
[24]Yunnan Observatories, Chinese Academy of Sciences, 396 Yangfangwang, Guandu District, Kunming, 650216, P. R. China.
[25]Chinese Academy of Sciences South America Center for Astronomy, National Astronomical Observatories, CAS, Beijing 100101, China.
[26]Departamento de Astronomía, Universidad de Chile, Casilla 36-D, Santiago, Chile.
[27]Department of Astronomy, School of Physics, Peking University, Beijing 100871, China
[28]Instituto de Física y Astronomía, Universidad de Valparaíso, Avda. Gran Bretaña 1111, Valparaíso, Chile.
[29]CAS Key Laboratory for Research in Galaxies and Cosmology, Department of Astronomy, University of Science and Technology of China, Hefei 230026, China.
[30]School of Astronomy and Space Science, University of Science and Technology of China, Hefei 230026, China.

Corresponding author: Matthew W. L. Smith
Matthew.Smith@astro.cf.ac.uk

[31] *National Astronomical Observatories, Chinese Academy of Sciences, 20A Datun Road, Chaoyang District, Beijing 100012, China.*

[32] *University of Chinese Academy of Sciences, 100049, Beijing, PR China.*

[33] *European Southern Observatory, Karl-Schwarzschild-Str. 2, 85748 Garching, Germany.*

[34] *Institute of Astronomy and Astrophysics, Academia Sinica, No. 1, Sec. 4, Roosevelt Rd., Taipei 10617, Taiwan.*

[35] *Center for Astrophysics | Harvard & Smithsonian, 60 Garden St., 02138 Cambridge, MA, USA.*

[36] *NSF's National Optical-Infrared Astronomy Research Laboratory, Tucson, AZ 85742, USA.*

[37] *East Asian Observatory, 660 N. A'ohōkū Place, University Park, Hilo, HI 96720, USA.*

[38] *Astronomical Observatory Institute, Faculty of Physics, Adam Mickiewicz University, ul. Słoneczna 36, 60-286 Poznań, Poland*

[39] *Max-Planck-Institut für extraterrestrische Physik, Giessenbachstraße 1, D-85748 Garching, Germany.*

[40] *Xinjiang Astronomical Observatory, Chinese Academy of Sciences, 830011 Urumqi, P. R. China.*

[41] *Department of Physics and Astronomy, McMaster University, 1280 Main St W, Hamilton, ON, L8S 4M1 Canada.*

## ABSTRACT

Observing nearby galaxies with submillimeter telescopes on the ground has two major challenges. First, the brightness is significantly reduced at long submillimeter wavelengths compared to the brightness at the peak of the dust emission. Second, it is necessary to use a high-pass spatial filter to remove atmospheric noise on large angular scales, which has the unwelcome by-product of also removing the galaxy's large-scale structure. We have developed a technique for producing high-resolution submillimeter images of galaxies of large angular size by using the telescope on the ground to determine the small-scale structure (the large Fourier components) and a space telescope (*Herschel* or *Planck*) to determine the large-scale structure (the small Fourier components). Using this technique, we are carrying out the HARP and SCUBA-2 High Resolution Terahertz Andromeda Galaxy Survey (HASHTAG), an international Large Program on the James Clerk Maxwell Telescope, with one aim being to produce the first high-fidelity high-resolution submillimeter images of Andromeda. In this paper, we describe the survey, the method we have developed for combining the space-based and ground-based data, and present the first HASHTAG images of Andromeda at 450 and 850 μm. We also have created a method to predict the CO($J$=3–2) line flux across M 31, which contaminates the 850 μm band. We find that while normally the contamination is below our sensitivity limit, the contamination can be significant (up to 28%) in a few of the brightest regions of the 10 kpc ring. We therefore also provide images with the predicted line emission removed.

## 1. INTRODUCTION

The Andromeda Galaxy (Messier 31) is possibly the most frequently observed galaxy in the sky. Galaxies in the Local Group are important for the obvious reason that they are closest, allowing us to study galaxies in the greatest possible detail, but they are also important because they are the only galaxies in which we can detect large numbers of individual stars. The ability to see stars adds a large number of investigative tools to the astronomer's toolkit, which are not possible to use on galaxies outside the Local Group.

There are only three spiral galaxies in the Local Group: our own, Andromeda, and the Triangulum (Messier 33). The Triangulum has a mass roughly ten times less than our own, but Andromeda has a mass

and other properties that are quite similar to our own (Yin et al. 2009). However, there are also some interesting differences. Andromeda has a larger bulge (Yin et al. 2009), less obvious spiral arms (Gordon et al. 2006; Kirk et al. 2015), and much of the star formation in the galaxy is occurring in a large ring (Ford et al. 2013). The cause of this ring is unknown. One interesting suggestion is that the ring may be the result of the dwarf galaxy M 32 passing through the center of the disk, generating a density wave, and thus a wave of star birth that propagates outwards through the disk (Block et al. 2006). This now seems unlikely since the star-formation history in the disk has no obvious radial gradient (Lewis et al. 2015), and the cause of the ring remains a mystery.

An iconic naked-eye object, Andromeda has now been observed by professional astronomers for over a thousand years. It was discovered, or at least first mentioned, in 964 by the Persian astronomer Abd al Rahman al-Sufi (*Book of Fixed Stars*). In the eighteenth century, it was observed by William Herschel, who noticed the red colors of the central bulge (Herschel 1785). In the twentieth century, Edwin Hubble used the Cepheid variables in the Andromeda Nebula to show that the nebula is actually a galaxy (Hubble 1929). In the modern era, Andromeda has been surveyed by virtually every modern observatory. A very incomplete list of the telescopes that have surveyed Andromeda includes XMM-Newton which surveyed the galaxy in X-ray (Stiele et al. 2011), GALEX in the ultraviolet (UV, Thilker et al. 2005), *Spitzer* in the mid- and far-infrared (Barmby et al. 2006; Gordon et al. 2006), *Herschel* in the far-infrared and submillimeter (Fritz et al. 2012; Smith et al. 2012; Draine et al. 2014), Westerbork & VLA in the radio 21 cm line (Braun et al. 2009; Koch et al. 2021), and the IRAM 30 metre telescope in the CO($J$=1–0) line (Nieten et al. 2006). The northern third of the galaxy has also been observed with *Hubble* in the Panchromatic Hubble Andromeda Treasury Survey (PHAT), which detected ≃117 million stars (Dalcanton et al. 2012).

A glaring omission on the list is a submillimeter telescope on the ground, from where it is possible to get much better angular resolution than is possible with the small mirrors of space telescopes. The angular resolution of the *Herschel* observations of Andromeda at 500 μm was 36 arcsec (FWHM), which is equivalent at the distance of Andromeda (780 kpc, de Grijs & Bono 2014) to a spatial resolution of about 136 pc, roughly the size of an association of GMCs. But with the SCUBA-2 camera on the James Clerk Maxwell Telescope (JCMT), the world's largest submm telescope, it should be possible to map Andromeda at 450 μm with a resolution of ≃8 arcsec, equivalent to a spatial resolution of ≃30 pc, slightly less than the size of a typical GMC. The reason that such a map has not previously been created is due to the atmospheric noise in the sub-millimeter waveband, which requires the data to be filtered strongly on angular sizes larger than that of the field-of-view of the camera, which is 45 arcmin$^2$ in the case of SCUBA-2 (Holland et al. 2013), much smaller than the $3 \times 1$ deg$^2$ ($\sim 10^4$ arcmin$^2$) that Andromeda occupies on the sky.

The solution to the problem is to combine data from a space observatory with data from a camera on the ground, using the camera on the ground to produce the high-resolution information and the observatory in space to determine the large-scale structure. In Fourier terms, we use the space data, from *Herschel* and *Planck* to pro-vide the low-$k$ Fourier components and the camera on the ground to provide the high-$k$ components.

We are using this technique to carry out a large survey of Andromeda with the JCMT: The HARP and SCUBA-2 High Resolution Terahertz Andromeda Galaxy Survey (henceforth HASHTAG). HASHTAG has been awarded 276 hours on the telescope, most of which is being used to carry out a survey with SCUBA-2 at 450 and 850 μm (221/275 hours). The rest of the time has been used to carry out a survey in the CO($J$=3–2) line with HARP in 12 regions covering a total area of 60 arcmin$^2$ within Andromeda's disk. By combining the data from SCUBA-2 with the *Herschel* images of Andromeda at six wavelengths, using an algorithm that does not require any smoothing of the data or assumptions about the temperature of the dust (Marsh et al. 2015), our goal is to produce maps of the bolometric dust emission and of the dust column density as a function of dust temperature and dust emissivity index ($\beta$) with a resolution of ≃25 pc at ≃70,000 independent positions within the galaxy, maps which will be used for a large range of scientific projects.

The CO part of HASHTAG has been completed and the results published in Li et al. (2020). The continuum part of HASHTAG is now about 70% complete and we have recently made the first full mosaics. The images cover the entire galaxy and have reached full sensitivity in the one third of the disk that has been covered by *Hubble* by PHAT (Dalcanton et al. 2012), by the Combined Array for Research in Millimeter wave Astronomy in the CO($J$=1–0) line (Caldú-Primo & Schruba 2016), and higher-resolution ($\sim$10″) VLA Hɪ observations (Koch et al. 2021). By using the *Herschel* image at 500 μm (Fritz et al. 2012) and the *Planck* image at 850 μm (Planck Collaboration et al. 2015) to fill in the low-frequency (low-$k$) Fourier components, we have produced the first high-fidelity images of Andromeda from the ground, at two wavelengths, 450 and 850 μm (Figure 1). These images are being used in the first round of HASHTAG science papers.

This paper gives an overview of HASHTAG and describes the observations and data reduction procedure used to generate the images shown in Figure 1, including a description of the technique we have developed to combine the space-based and ground-based continuum submillimeter observations, and the measures we have taken to optimize the pipeline parameters. All data products and codes presented here are available on the

HASHTAG website[1]. Future data releases will also be made available on this site.

The organisation of the paper is as follows. Section 2 gives an overview of the science programme that can be carried out with these images. Section 3 describes the observing method and mapping strategy. Section 4 describes the data-reduction pipeline and the method used to combine the space-based and ground-based data. Sections 5 and 6 describes the extensive simulations that we have carried out to optimize and test the data-reduction pipeline and combination method described in Section 4. Section 7 presents our final reduced maps, including some simple analysis of their properties. Finally, Section 8 describes how we estimate the contamination from CO($J$=3–2) in our continuum observations.

## 2. OVERVIEW OF SCIENCE PROGRAMME

In this section we give a brief overview of the scientific projects that will be possible with the HASHTAG dataset. These mostly fall into two categories: dust and star formation.

### 2.1. *Dust*

Dust itself is interesting for two main reasons. First, it is of great intrinsic interest because it is a vital phase of the interstellar medium (ISM), containing half the heavy metals (James et al. 2002) and being a catalyst in the networks of chemical reactions in the ISM, including the vital one in which atomic hydrogen is transformed into molecular hydrogen. Second, mapping the continuum emission from dust grains is a promising method for both mapping the ISM in galaxies and estimating the total mass of the ISM (Hildebrand 1983; Eales et al. 2012; Magdis et al. 2012; Scoville et al. 2014). The standard tracer of the molecular phase, the CO molecule, has many well-known disadvantages (Bolatto et al. 2013). There is also now the more fundamental problem that one third of the molecular gas in the Galaxy appears to contain no CO (Abdo et al. 2010; Planck Collaboration et al. 2011; Pineda et al. 2013), and there are even galaxies in which the fraction of 'CO-dark' gas seems to be close to 100% (Dunne et al. 2021). Some of the advantages of using dust grains rather than CO molecules to trace the ISM are that the dust emission is optically thin, dust grains are robust and not liable to be destroyed by starlight, and the relationship between the gas-to-dust ratio and metallicity seems to be much simpler than between CO abundance and metallicity (Eales et al. 2012; Sandstrom et al. 2013; Rémy-Ruyer et al. 2014).

The biggest contribution that HASHTAG seems likely to make to our understanding of the dust itself is to show how the properties of dust vary within an individual galaxy. The earlier *Herschel* observations of Andromeda revealed systematic large-scale spatial variation in the properties of dust. The emission from interstellar dust follows a modified blackbody ($S_\nu \propto B_\nu(T_{\rm d})\nu^\beta$). The *Herschel* observations revealed that $\beta$ varies radially within Andromeda's disk (Smith et al. 2012; Draine et al. 2014; Whitworth et al. 2019), and radial variation in $\beta$ has subsequently been found in the Galaxy (Planck Collaboration et al. 2014a), M 33 (Tabatabaei et al. 2014) and in $\simeq$20 other galaxies (Hunt et al. 2015), although the form of the radial variation varies between galaxies (Hunt et al. 2015). There is also now some evidence that the global value of $\beta$ varies between galaxies (Lamperti et al. 2019). The cause of the variation must be caused by changes in the structure, physics or chemistry of the dust, although what the key changes are is currently unknown, one clue may be that in Andromeda $\beta$ does not appear to differ between low-density and high-density gas (Athikkat-Eknath et al. 2021).

HASHTAG will produce measurements of $\beta$ at $\simeq$70,000 positions in Andromeda's disk, effectively producing a dust atlas for Andromeda. The *Hubble* observations of one third of the disk have produced estimates, from the optical extinction, of the column-density of dust with the same spatial resolution that HASHTAG will provide (Dalcanton et al. 2012). The combination of measurements of the emission properties of dust from submm observations and the absorption properties of dust from optical observations is a powerful one for testing theoretical dust models. The emission and absorption properties of dust, derived from *Hubble* and *Herschel* data, are already inconsistent with all existing dust models (Whitworth et al. 2019). The combination of the high-resolution measurements of $\beta$ from HASHTAG, the *Hubble* dust measurements and the maps of the ISM phases, star formation, chemical abundances and other properties that are available for Andromeda offer at least the prospect of uncovering the physical/chemical causes of the variation in dust.

The use of the dust emission to trace the ISM in Andromeda offers a number of interesting possibilities. By comparing the dust emission to the CO and HI emission it will be possible to search for CO-dark gas in Andromeda (Planck Collaboration et al. 2011; Sandstrom et al. 2013). It will also be possible to produce catalogues of GMCs based on dust emission rather than CO. This technique has already been applied to the *Herschel* observations of Andromeda, producing a catalogue of 326 clouds with masses between $10^4\,{\rm M_\odot}$ and $10^7\,{\rm M_\odot}$
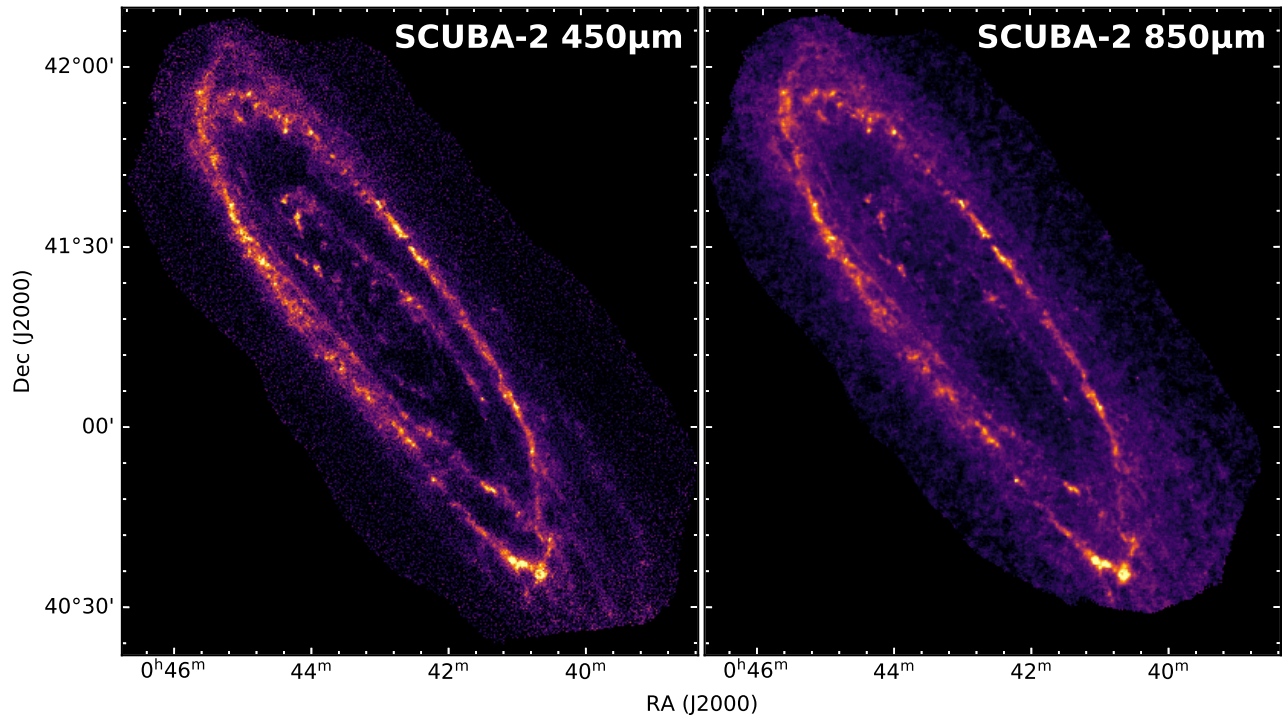
**Figure 1.** The HASHTAG images created from the first ∼70% of the final SCUBA-2 dataset. The 450 and 850 µm images have been smoothed with a 7.9, and 13″ FWHM Gaussian, respectively. For the raw-resolution images see Figure 16.

(Kirk et al. 2015). A more recent study suggests that clouds found by the dust method have much lower CO-to-dust ratio than the clouds found from their CO emission (Athikkat-Eknath et al. 2021), suggesting there is much more variation in the properties of clouds than one would expect from a CO-selected catalogue. The clouds in the *Herschel* catalogue are probably associations of GMCs rather than single GMCs, but with the extra resolution of HASHTAG it will be possible for the first time to produce a catalogue of dust-selected clouds that are likely to be GMCs rather than GMC associations.

## 2.2. *Star Formation*

One of the most important properties to measure within a galaxy is the star-formation rate, but there is still no gold-standard way of doing this. There are at least 12 different methods which use different techniques for tracing the obscured and unobscured star formation (Kennicutt & Evans 2012; Speagle et al. 2014; Davies et al. 2016), all of which have limitations, and none is clearly better than the others.

HASHTAG will produce high-resolution maps of the bolometric dust emission, which is a direct measurement of the emission from the obscured OB stars, although it is really an upper limit since much of the bolometric dust emission is re-radiated emission from the older stellar population (Bendo et al. 2012; Kennicutt & Evans

2012; Bendo et al. 2015; Viaene et al. 2017; Ford et al. 2013). However, since Andromeda is close enough to detect individual stars, there is at least the possibility of combining *Hubble* observations of the unobscured OB stars and the HASHTAG observations of the emission from the obscured stars to provide a direct measurement of the star-formation rate, rather than the ones produced by the current methods, which mostly rely on indirect tracers of the star formation. The PHAT team made a first attempt to do this (Lewis et al. 2017), using optical extinction measurements to correct for obscuration, but their method was unable to account for the OB stars that are still deep in GMCs and are completely hidden by dust. If it is possible to correct for the part of the bolometric dust emission that is re-radiated emission from the older stellar population, HASHTAG will provide estimates of numbers of these missing OB stars.

## 3. SURVEY STRATEGY

HASHTAG is a JCMT Large Programme (ID: M17BL005), and is split into two components: the continuum submillimeter observations of the entire galaxy and observations in the CO($J$=3–2) line of selected regions.

We used 55.3 hours with HARP (Buckle et al. 2009) to observe 11 $2' \times 2'$ fields, and one $4' \times 4'$ field. We selected these fields to cover a range of diverse ISM con-

ditions in M 31 and to maximize overlap with useful ancillary data e.g. *Herschel* far-infrared spectroscopy (Kapala et al. 2015). Our observations were carried out in 2017 using grade-3 weather defined as when the opacity at $225\,\mathrm{GHz}$ ($\tau_{225\,\mathrm{GHz}}$) is between 0.08 and 0.12, and reached a sensitivity of $\simeq 15\,\mathrm{mK}$ (antenna temperature, $\mathrm{T_a^*}$) with an angular resolution of $15''$ and spectral resolution of $2.6\,\mathrm{km\,s^{-1}}$. A full discussion of our CO observations, is given in Li et al. (2020). The CO($J$=3–2) line falls within the $850\,\mu\mathrm{m}$ continuum filter and so our CO spectroscopic mapping are useful for assessing the effect of line contamination on the continuum measurements.

The larger component of HASHTAG consists of the continuum observations, which were allocated 221 hours in the less-common grade-2 weather ($0.05 < \tau_{225\,\mathrm{GHz}} < -0.12$) and commenced in 2017 (expected to complete 2021). SCUBA-2 observes simultaneously at 450 and $850\,\mu\mathrm{m}$, producing images at the two wavelengths with the same field-of-view (Holland et al. 2013), and angular resolutions of 7.9 and $13.0''$ at 450 and $850\,\mu\mathrm{m}$, respectively (Dempsey et al. 2013). Our goal was to observe the entire galaxy at both wavelengths.

Our continuum observing strategy is based on our experience in a smaller project in 2015 (project M15BI082, P.I. Smith) and was effectively HASHTAG's pilot field. The SCUBA-2 Pong observing mode is used for sources greater than $\sim5'$ in size and can be used to map a circular region of diameter 15, 30 or 60 arcmin (since the design of HASHTAG a 45 arcmin mode has been added). For the pilot we made a long *Pong* exposure of a single circular region of diameter 30 arcmin. For this field, we chose the position of the center so that this circular region would also include a significant area in which there was no obvious sub-millimeter emission in the *Herschel* images from the galaxy itself, since we knew this 'background region' would help with the data reduction to converge. The maximum integration time for a SCUBA-2 observation is 45 minutes. We chose an integration time of 43 minutes, repeating it 37 times. We reached a sensitivity of 44.9 and $3.0\,\mathrm{mJy\,beam^{-1}}$ using 2.0 and $4.0''$ pixels, at 450 and $850\,\mu\mathrm{m}$, respectively.

Given the success of these observations, and our development of a technique to use space-telescope data to replace the large-scale structure severely suppressed by the filtering necessary to remove atmospheric noise (see Sections 4 and 5), we realized it would be practical to use a similar strategy to observe the whole of Andromeda.

We continued to use the 30-arcmin *Pongs*, choosing the center of each field so that the field contained a similar amount of blank sky as the pilot field. This decision required us to have two rows of *Pongs* along the major axis. To achieve fairly uniform sensitivity we chose the positions of the centers so the circles overlapped, as is shown in Figure 2. At each position we made 17 repeat observations of 43 minutes each. Since each position in the galaxy will be covered by at least two *Pongs*, every point in the galaxy will be observed, when the survey is complete, at least 34 times, with a total integration times of 24.4 hours per 30 arcmin diameter region, the same as the pilot survey. We also include data from two significantly shallower projects (M12BU26 & M13BU18) who also observed the entirety of M 31.

There are two advantages of the design of the fields shown in Figure 2. There is a very large amount of overlap in the observations, especially as the area covered by a *Pong* is somewhat larger than the nominal 30-arcmin circle. This overlap and redundancy in the data is a considerable help in the data reduction, in particular for distinguishing real emission from atmospheric emission. The second advantage is that the two rows of *Pongs* overlap, so the sensitivity of the survey will be much greater along the major axis of the galaxy, which helpfully covers the central regions of Andromeda where the dust emission is much weaker than in the star-forming ring.

Inspired by the discovery of luminous transients in the mid-infrared (Kasliwal et al. 2017), one of our science goals is to determine whether there are any luminous transient sources in the submillimeter waveband. We have therefore split our 17 observations in each field into two sets of 9 and 8 observations, with the aim of eventually producing two images of Andromeda separated in time, so we can search for transient phenomena. We have tried to prioritize our observations to ensure that there is at least a six-month gap between the two sets for each field, but due to the vagaries of the weather, and the flexible observing queue at the JCMT, it is impossible to do this perfectly. However, we do achieve some time cadence in the observations, which we will investigate in future works.

## 4. DATA REDUCTION: 1. THE METHOD

Big sub-millimeter datasets can be challenging to reduce, but HASHTAG is particularly difficult. Large cosmology programs for example produce datasets as large as ours but they have the advantage that the data can be reduced piecemeal; the individual datasets are reduced separately and then the images added together. We were not able to follow this approach as we want to maximize our sensitivity to extended low signal-to-noise emission, which required us to reduce all the data together.

In this section we describe the elements of our method. In Sections 5 and 6 we describe the sky simulations we carried out to optimize the method. Unless stated oth-
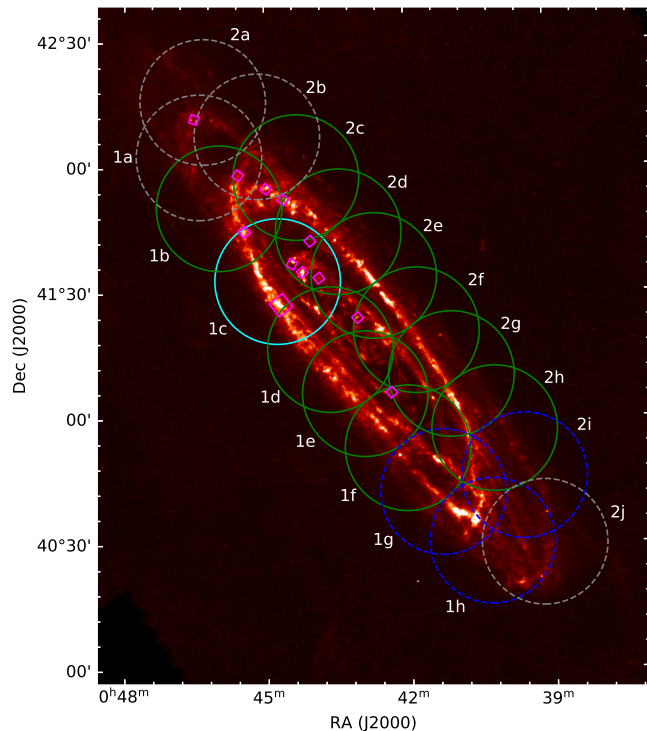
**Figure 2.** The HASHTAG observing plan superimposed on the *Herschel* 250 μm image (Smith et al. 2012). Each circle represents one of our 30′ *Pong* observations, each identified by a white label. The color of the circle represents the observing status of the *Pong* when the DR1 products were constructed (November 2020): green shows pongs for which all 17 observations have been completed; dashed blue shows roughly half (eight or nine), have been completed; and dashed grey indicates no observation has yet been made (1a has one observation). The cyan circle (1c) is our 'pilot' field from 2015 which has 37 observations, which is why there is less overlap with the other circles. The magenta squares show the regions covered by our CO($J$=3–2) observations.

erwise, the methods we used for the observations and for the data reduction are the standard ones used for SCUBA-2 (Dempsey et al. 2013; Chapin et al. 2013).

### 4.1. *Initial Processing and Quality Review*

As soon as one of our observations was made, we processed it using a quick-look script which produced images at the two wavelengths (450 and 850 μm) and saved the cleaned data from the individual bolometers. We used these images to check for any severe problems (e.g., array failures or low sensitivity), which for SCUBA-2 data is extremely rare (only 2 out of 235 observations had issues). We used the saved bolometer data from this initial processing as the input for the next stage, which resulted in a considerable rationalization of the data, since the initial processing converted the hundreds of raw-data files into just eight (one per array).

### 4.2. *An adaptable skyloop: scubaDuperSkyloop*

To reduce SCUBA-2 data, the observatory provides an iterative data-reduction procedure called MAKEMAP, which at the end of every iteration provides a better estimate of the astronomical signal and the noise until no further improvement is made. Here we provide a brief summary of the MAKEMAP algorithm but refer the reader to Chapin et al. (2013) for full details. MAKEMAP is provided as part of the STARLINK software package (Currie et al. 2014) and throughout this paper we use a recent development build of STARLINK (version 9072c4434 from April 2020), as we required some updates since the last 2018A stable release.

Initially MAKEMAP splits the raw data into individual observations (sub-divided into chunks if enough RAM is not available, although for HASHTAG 'chunking' was not required). There is an initial cleaning step for each observation in which bad bolometers are masked and artifacts (e.g. glitches) are removed from the timelines. MAKEMAP then starts the iterative process. At the beginning of each iteration, the common-mode signal (common to all bolometers) is removed. The data is then corrected for atmospheric extinction, and a high-pass filter is applied to the timeline data to remove any residual slowly varying signal. Since in the SCUBA-2 arrays are moving across the sky, the removal of a slowly varying signal is equivalent to removing emission on a large angular scale. An image is then made from the data. Any real astronomical signal is then identified in the image, and this astronomical signal is then removed from the timelines (an optional signal-to-noise cut or mask can be applied). The process is then repeated on the new timeline data, with the astronomical signal being updated in each iteration. The process stops when the pixel variations in the map at the end of each iteration fall below a set threshold (i.e., when the map has 'converged'). If we had used the standard implementation of MAKEMAP, our final image of Andromeda would have been a mosaic of the images made by MAKEMAP from each individual observation.

However, a weakness of MAKEMAP is that a single ∼43 minute *Pong* observation does not have the sensitivity to detect the low-surface-brightness emission in Andromeda's disk. Recognising this limitation, the STARLINK team created a script called SKYLOOP[2], which runs MAKEMAP on all the observations, one iteration at a time, combining the individual images at the end of each iteration to produce the best estimate of the as-

---

[2] https://starlink.eao.hawaii.edu/docs/sun258.htx/sun258ss72.html

tronomical signal, thus maximizing the signal-to-noise ratio in the faint extended structure. However, the volume of our data is so large that SKYLOOP would take too long to run. We built on our previous work (Smith et al. 2019) to develop a new version of SKYLOOP that we nick-named 'scubaDuperSkyloop'. The main difference from the old script is that while SKYLOOP is given all the observations and internally processes them individually, mosaiking the images at the end of every iteration, our modified script calls MAKEMAP separately for each iteration and each observation. We found that the new script was more stable on our system (possibly due to more regular memory clearance),however, the main advantage of our approach is that each observation can be processed in parallel, or even on separate machines. Although MAKEMAP can process individual observations using multiple-threads, there are diminishing improvements in processing time.

To make the final images shown in this paper, we reduced all the data using a new processing machine with $768\,\mathrm{GB}$ of RAM and 32 cores/64 threads, which allowed us to run five invocations of MAKEMAP in parallel. The main factor limiting the speed of 'scubaDuperSkyloop' was the speed of our hard-disk drives. We increased the speed of the simulations described in section 5 by running the script in parallel on two separate smaller machines (with common storage access).

### 4.3. Restoring the Large-Scale Structure

Ground-based sub-mm surveys of sources with extended emission (greater than a few arcminutes) face the challenge of slow variations in both the atmospheric emission and within the camera which need to be removed from the data. This is overcome using harsh high-pass filtering, which also removes real astronomical signal on large angular scales. However, there are sub-millimeter observations of many objects with the space telescopes *Planck* and *Herschel*. The images made with these telescopes do preserve the large-scale structure; however, because of the small sizes of the telescopes' mirrors they do not have as good resolution as images made with telescopes on the ground. In principle, we can now produce high-fidelity images of submillimeter sources with large angular sizes by combining observations with telescopes on the ground, which provide the small-scale structure, with observations made with telescopes in space, which provide the missing large-scale structure. This technique has often been used in radio astronomy to combine single-dish and interferometer measurements, as the latter sparsely samples the UV-plane the case here is potentially simpler.

We have produced the high-fidelity image of Andromeda at $850\,\mu\mathrm{m}$ by combining the SCUBA-2 image at $850\,\mu\mathrm{m}$ and the 353-GHz *Planck*, which are at virtually the same wavelength. At $450\,\mu\mathrm{m}$, we have combined the SCUBA-2 image at this wavelength with the *Herschel* image at $500\,\mu\mathrm{m}$ (Smith et al. 2012). To combine the low- and high-resolution images we have written a python module which applies a 'feathering' technique (Bajaja & van Albada 1979). The module performs the following steps:

1. The low-resolution FITS image is re-projected so that the pixel scale and the celestial coordinates of the pixels are the same as in the high-resolution image. The units of both images are converted into $\mathrm{Jy\,beam^{-1}}$ using information contained in the header or supplied by the user.

2. We apply color-corrections to the high and low resolution images to correct for the effect of different instrumental filters, calibration schemes (e.g., different reference spectra), and differing central frequencies. To apply this correction the user can specify a fixed dust temperature and $\beta$, provide maps of the dust parameters, or a PPMAP cube with the surface-density of dust for a grid of dust temperatures and $\beta$'s. A pre-computed grid specific to each far-infrared/sub-millimeter instrument is used to perform the corrections, and is provided with the task.

3. The median value in each image is subtracted from the image and 'NaN' pixels are replaced with zeros to avoid artifacts.

4. Both images are Fourier transformed and shifted so a spatial frequency of zero is assigned to the center. The values in the Fourier Transform (FT) of the low-resolution image are scaled by the ratio of the beam areas of the high-resolution and low-resolution images.

5. A filter is then created to weigh the FT images by the selected amount when they are combined. The standard filter in the module, which we used for HASHTAG, is a Gaussian filter in Fourier space. We chose the value of the filter's standard deviation using the simulations described in the next section. The FT of the high-resolution image is multiplied by one minus the Gaussian filter. The FT of the low-resolution image by default is not weighted, as the image is effectively already weighted by the point spread function of the image; this is the same method employed

by CASA (McMullin et al. 2007). In our simulations we also try the alternative where the low-resolution FT image is weighted by the Gaussian filter, which is applicable when the filter is on significantly larger spatial scales than the resolution of the images. For a full discussion of combining images in the Fourier plane see Stanimirovic (2002). The weighted FT images are then added together. There is also an option in the module to use either a Butterworth (Csengeri et al. 2016) or Sigmoid filter, in which case both FTs are multiplied by the filter. We tried these filters for HASHTAG but found they did not produce appreciable benefit over the Gaussian filter.

6. The combined FT image is then inverse Fourier transformed (with appropriate inverse shifts applied). The NaN pixels are restored, the median background from the original low-resolution image is added back to the new image, and the keywords in the image header are updated[3].

7. As many SED fitting procedures apply a color-correction step in their processing, we remove the color corrections performed in step 2, so the flux densities in the final images are based on the same assumptions as regular SCUBA-2 images.

One key parameter in the 'feathering' algorithm is the scale of the Gaussian filter. As the feathering step is a relatively quick process (compared to the SCUBA-2 pipeline) we keep this as a free parameter which we optimize in Section 5 and 6. For the color corrections we assume the spectral energy distribution (SED) in each pixel estimated by Whitworth et al. (2019), who applied the PPMAP algorithm to the *Herschel* dataset to generate SEDs with the angular resolution of the highest resolution *Herschel* image. For the *Herschel* and *Planck* images, we calculated the color corrections using these SEDs and the filter curves available on the observatories' websites, after removing the standard SED used to estimate the *Planck* and *Herschel* flux densities ($F_\nu \propto \nu^{-1}$). SCUBA-2 flux densities, on the other hand, are calibrated relative to Mars and Uranus, which means that they are based on a very different assumption about SEDs (roughly $F_\nu \propto \nu^{1.7}$, Lellouch & Amri 2008; Orton et al. 2014). We calculated the color corrections for the SCUBA-2 images after removing this assumption and then using the PPMAP SED in each pixel. For SCUBA-

2, an additional complication is that the effective filter function is the product of the actual filter function and the atmospheric transmission. We calculated the effective filter function from the filter function available on the observatory's website and a model for the transmission of the atmosphere[4] with $\tau_{225\,\mathrm{GHz}} = 0.065$ to match the weather for our survey. The color corrections are small ($\leq 3\%$) for the two SCUBA-2 filters and for the *Herschel* 500 μm filter, apart from the large correction needed to change the flux at 500 μm to one at 450 μm, which ranges from a factor of ∼1.5 in the center to ∼1.34 in the ring. The color correction for the *Planck* filter is ≃10%.

Figure 3 shows the results of the feathering technique when applied to SCUBA-2 simulation (see Section 5), illustrating how it is effective at restoring the structure on all spatial scales. The red line shows the power spectrum of a simulated 'true' image of the sky at 850 μm. The green and blue lines show models of what *Planck* and SCUBA-2 would see, respectively, in one case missing the high-$k$ and in the other case the low-$k$ Fourier components. The orange line shows the power spectrum of our reconstruction of the sky by applying our feathering technique to the artificial *Planck* and SCUBA-2 images.
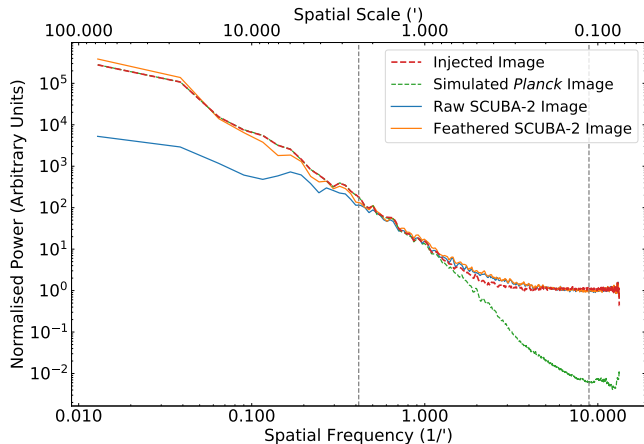
We performed a sanity-check of the method by applying it to a *Herschel* 250 μm image of the Milky Way (Molinari et al. 2016), in which the large-scale emission is detected with high signal-to-noise. We created an artificial image at the *Planck* resolution by smoothing the *Herschel* image, and we created a rough simulation of what a camera like SCUBA-2 would see by using the NEBULISER algorithm developed by the Cambridge Astronomical Survey Unit[5] to remove the structure on large scales. When we produced our combined image by applying our feathering technique to the two artificial images, we found a good agreement with the original image (to within a few percent), although in the brightest regions there were differences up to the 10% level.

There are also some parameters to tune in MAKEMAP, in particular the scale of the high-pass filter used to remove the signal from the atmosphere and the camera itself. To produce a reliable map of Andromeda it is crucial to get these values right. This is a particular challenge at the longer of the two SCUBA-2 wavelengths because of the need to ensure that emission on the SCUBA-2 images is preserved on all angular scales up to the an-

---

[3] There is an option in the module to add back a background by calculating the offset between the original low-resolution image and the new image smoothed to the same resolution.

[4] Atmospheric model from from CSO (Pardo et al. 2001), http://www.submm.caltech.edu/cso/weather/atplot.shtml.

[5] http://casu.ast.cam.ac.uk/surveys-projects/software-release/background-filtering

**Figure 3.** A demonstration of the ability of our feathering technique to recover the structure of the sky on all spatial scales, using the final simulation in Section 5. The red line shows the power spectrum of a simulated 'true' image of the sky at 850 μm (with artificial noise added to match the SCUBA-2 image). The green line shows the power spectrum of what *Planck* should see, obtained by convolving the true image to the *Planck* resolution. The blue line shows what SCUBA-2 should see, obtained by passing the true image through the SCUBA-2 data-reduction pipeline. As expected, the *Planck* spectrum is missing the high-$k$ Fourier components and the SCUBA-2 spectrum the low-$k$ Fourier components. The orange line shows the power spectrum of our reconstruction of the sky by applying our feathering technique to the artificial *Planck* and SCUBA-2 images. For reference the location of the half-width half-maximum (HWHM) for *SCUBA-2* and *Planck* is shown by the grey vertical lines. Note in all cases we have restricted the images to the central 30′ region.

gular resolution of *Planck* ($\simeq$5 arcmin, FWHM). In the next section, we describe the simulations of the sky that we used to determine the best values of the parameters.

## 5. THE SIMULATIONS AT 850 μm

We carried out simulations of the sky to optimize the data-reduction procedure described in the previous section. The SCUBA-2 pipeline has many parameters which can be tweaked to optimize the reduction process depending on the angular extent of the source, observing strategy and the atmospheric conditions. Two of the big unknowns are the scale of the Gaussian filter used in combining the low- and high-resolution data and the angular scale of the high-pass filter that should be used in the SCUBA-2 data reduction to remove the noise on large angular scales caused by the atmosphere and the camera. Too harsh a filter would remove the noise but also remove too much astronomical signal, too weak a filter would leave the astronomical signal alone but not remove the noise. In this section we create a

'simulation' to test the effects of the various pipeline parameters so we can obtain the most accurate map of M 31's sub-millimeter emission. In this process we have restricted ourselves to the SCUBA-2 pipeline, rather than attempt alternative methods (e.g., Scanamorphos, Roussel 2013) or complex atmospheric modelling.

An outline of our method is as follows. (1) We used real SCUBA-2 data from a cosmology Large Programme with similar noise properties to our own dataset. (2) We then created a 'true' image of Andromeda from a *Herschel* image and inserted this into the timelines for the cosmology programme. (3) We convolved the true image to produce an artificial *Planck* image. (4) We ran the SCUBA-2 dataset (cosmology timelines injected with our model galaxy) through the SCUBA-2 data-reduction pipeline (Section 4.2). (5) We combined the reduced SCUBA-2 image and the *Planck* image to try to recover the original image using the method of Section 4.3. (6) We measured the statistical differences between our recovered image of Andromeda and the original true image. We ran thousands of simulations, trying different variants of the SCUBA-2 data-reduction pipeline, in particular trying different values of the scale of the high-pass filter, and trying a range of values for the scale of the Gaussian filter used to combine the low-resolution and high-resolution data (Section 4.3). In this section we describe the simulations at 850 μm, which were more critical because *Planck* at 850 μm has a much lower resolution than *Herschel* at 500 μm. We describe what we did for the 450 μm data in Section 6.

### 5.1. *Simulation Setup*

We chose to carry out a simulation of a single 30′ *Pong* observation with a similar sensitivity to our pilot field. While ideally we would have simulated observations of the whole galaxy, the processing time would have been too long and there was no suitable SCUBA-2 data we could use for a simulation of the entire galaxy at our depth. For our simulation we used data from the SCUBA-2 Cosmology Legacy Survey (Geach et al. 2017). We chose to use the data from the survey of the Lockman Hole because it was carried out in similar weather conditions and consisted of 35 30′ *Pongs*, the same as we used in our observations of our pilot field (Section 3), reaching a similar sensitivity.

We made our 'true' image out of the *Herschel* 250 μm image, which has a resolution (18 arcsec, FWHM) which is not very different from that of SCUBA-2 at 850 μm (13.5 arcsec, FWHM). We first reprojected the *Herschel* image onto a 4″ pixel grid and then multiplied the intensity values in each pixel by the ratio of the global 250 μm and 850 μm fluxes (Planck Collaboration et al.

2015). The Lockman-Hole data are actually slightly less sensitive than the data for our pilot field (3.4 versus 3.0 mJy beam$^{-1}$), so to make the signal-to-noise ratio of our artificial image the same we multiplied the intensity values by the ratio of the noises. We also applied color corrections. We then converted the intensity values into instrumental units of pico watts (pW) and ran MAKEMAP with the options "fakemap" and "exportclean" set so that our artificial 850 μm image was added to the SCUBA-2 timelines for the Lockman Hole survey. We used the data files exported by this process, which also did the initial cleaning of the timelines, to carry out our simulations. We produced our artificial *Planck* image by convolving our input image to the *Planck* resolution.

We performed five different sets of simulations, to optimize different aspects of the method. While in an ideal world every possible combination of parameters in the SCUBA-2 data-reduction pipeline would be tested, each individual simulation required ∼8–10 hours to run. Therefore, we varied one parameter at a time. However, our reconstruction process in which we combined the processed SCUBA-2 image with our artificial *Planck* image was quite quick to run, so at the end of every run of the pipeline, we applied our reconstruction technique many times, each time with a different value for the scale of the Gaussian filter, with scales ranging from 160 to 840″.

We measured the success of a simulation by the differences between our final image and the original true image. We assessed the significance of the differences using the noise image produced by the pipeline, but since changes to the pipeline also change the noise in the final image we also used a reference noise image (from a run using typical values of all the pipeline parameters). For each final SCUBA-2 image, we created several 'difference maps' of the difference between the image and the true image: (1) a basic residual image, i.e., the final image minus the 'true' image; (2) the residual image divided by the noise image; (3) the residual image divided by the reference noise image; (4) the residual image divided by the true image.

To assess the agreement, we measured statistics in two different regions of these difference maps: (1) the full-depth region of the *Pong*; (2) the full-depth region of the *Pong* but only for pixels where the flux in the pixel in the *Herschel* 500 μm image is above a critical value (500 μm is used as its closest in wavelengths to both SCUBA-2 bands). The point of the second region was to stop the more numerous pixels outside the galaxy with little emission biasing the results, as a method producing a flat map (e.g., harsh Fourier filtering) would be preferred. We inspected all of these methods of assessing the agreement at one time or another. The statistic that we found most useful was the mean of the absolute difference between the final image and the true image divided by the reference noise image, for pixels in the deep region above the 500 μm threshold. This statistic using a reference noise map has the advantage that changes in the noise map do not bias the estimate of how well we recover the galaxy, while still accounting for variations in the sensitivity across the map.

### 5.2. *Filter-scale and PCA-components*

The most important parameter in the SCUBA-2 data-reduction pipeline is the scale of the high-pass filter used to remove residual emission from the atmosphere or the instrument. We started our simulations with the expectation that we would need to set the scale to roughly the angular resolution of *Planck*. In our early results we found that with a filter scale of 340″ we were able to reproduce the true image well.

In April 2019, however, the observatory released a new mode for the SCUBA-2 pipeline in which principal component analysis (PCA) is used to remove residual atmospheric and instrumental noise. The advantage of PCA is that it makes it possible to increase the angular scale of the high-pass filter, reducing the attenuation of the emission on the SCUBA-2 images on large angular scales. Of course, if one allows PCA to remove too many components, it is also possible to remove real astronomical emission from the image. In the SCUBA-2 pipeline the default number of PCA components to remove is 20 components per sub-array, but after inspecting the final images made with this setting we decided it led to the removal of some real emission. We therefore decided to run a suite of simulations in which we varied both the scale of the high-pass filter and the number of PCA components.

In our initial simulations without PCA we had found an optimum filter scale of 340 arcsec. We realized that with PCA we should be able to increase this scale. We therefore ran simulations with filter scales between 340″ and 560″ and the number of PCA components between 0 and 20, running 91 simulations to cover this 2D parameter space (20 PCA components is the default in the new pipeline mode).
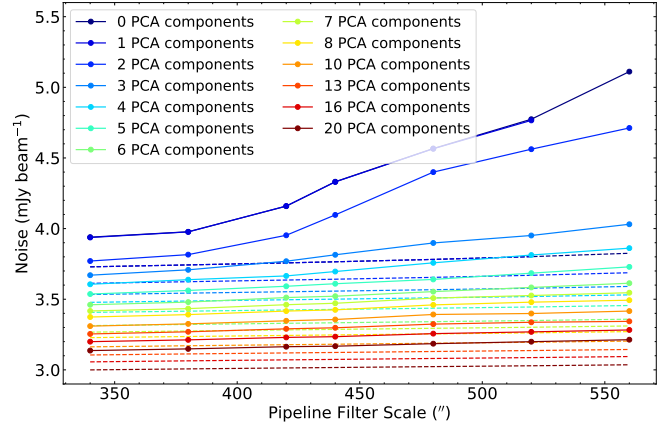
Figure 4 shows the effect of changing the number of PCA components while keeping a constant filter scale of 520 arcsec. The top-left hand panel shows what happens if no PCA components are removed. The filter may not remove much large-scale emission from the galaxy but much of the emission visible in the picture is clearly spurious. Increasing the number of PCA components leads to better removal of these artifacts, but the removal

## 520″ Filter Scale
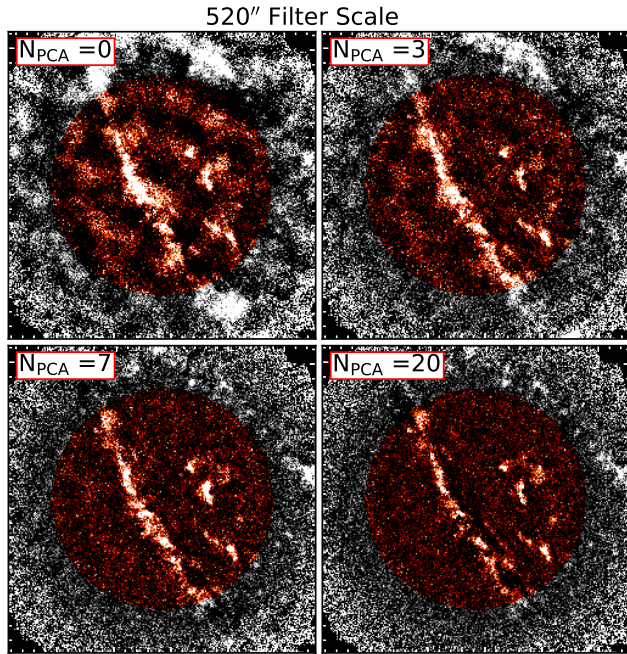


**Figure 4.** Reconstructed SCUBA-2 images made with a high-pass filter in the SCUBA-2 pipeline of 520 arcsec but with different numbers of PCA components removed (0, 3, 7 and 20). The red area is the circular region with a diameter of $30'$ in which the *Pong* reaches full sensitivity. The color scale has been chosen to enhance faint features. The removal of too few PCA components leads to large-scale artifacts in the image (top left panel); too many PCA components removes real astronomical signal, leading to the negative regions close to the bright structure (bottom right panel).



**Figure 5.** The relationship between the average noise in an image as a function of the the scale of the high-pass filter and the number of PCA components that have been removed. The solid lines show the results when the noise is measured from empty areas of the image and the dashed lines show when it is measured from the noise image produced by the SCUBA-2 data-reduction pipeline. The difference between the two sets of lines shows that the noise values from the variance map produced by the pipeline are unreliable because the pipeline mistakenly identifies large-scale noise as real astronomical signal and therefore underestimates the noise. Even when these artifacts are removed with a harsh high-pass filter or by removing a large number of PCA components, there is still a small difference of $\sim 0.2\,\mathrm{mJy\,beam^{-1}}$, which is possibly caused by faint sources in the apparently empty regions of the image.

of too many PCA components leads to the removal of real astronomical signal, producing the negative regions around the image in the bottom-right panel, which has had 20 PCA components removed.

An important point to note is although the eye tells us that the patches in the top left panel in Figure 4 are artifacts, the SCUBA-2 data-reduction pipeline treats these as real emission, which means that the values in the noise image produced by the pipeline are too low. Figure 5 shows how the average noise on an image depends on the number of PCA components and on the scale of the high-pass filter. The solid lines show the results if the noise is measured from empty areas of the final image; the dashed lines show the result if the noise is measured from the noise image produced by the SCUBA-2 data-reduction pipeline. The pipeline estimates are much lower, showing the effect of the pipeline treating the artifacts as real astronomical signal (Users of the SCUBA-2 pipeline beware!). The more reliable noise values, measured from the images themselves, show that the noise in an image can be reduced either by decreasing the scale of the high-pass filter or by increasing the number of PCA

components. In either case, of course, one also runs the risk of removing real astronomical signal.

We chose the best combination of filter scale and number of PCA components based on a combination of visual inspection of the residual maps and the statistical estimates of the difference between the recovered image and the true image. Figure 6 shows the mean absolute residual for the difference map made by dividing the residual image by the reference noise image (number three in the list above). The statistic has been calculated for the pixels in the full-depth region and which are above the $500\,\mu\mathrm{m}$ threshold value (see above).

The figure shows that there is an advantage in increasing the scale of the high-pass filter from the 340 arcsec that we had originally considered (see above) if removal of PCA components is included in the analysis. The best agreement between the recovered and true image is obtained for a filter scale of $\simeq 520$ arcsec. The figure also shows that there is an optimum number of PCA components of $\simeq 8$ — removing more increases the difference between the recovered and true images. Based on these results, we visually inspected the recovered images for a filter scale of $480''$ and 5 or 6 PCA components and
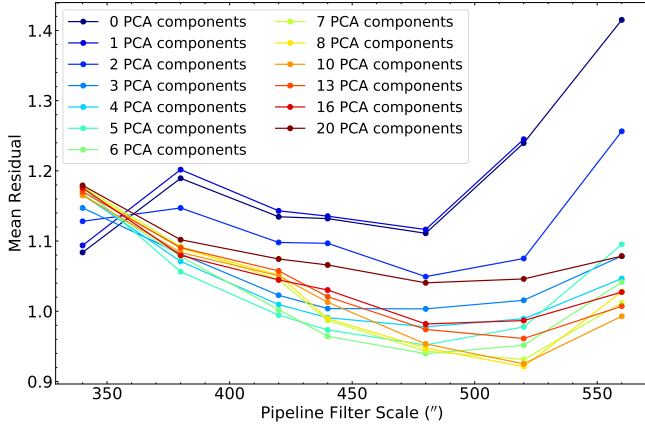
**Figure 6.** The mean absolute residual in the difference map versus the angular scale of the high-pass filter. Each line shows the results when a different number of PCA components are removed. The difference map is the residual image (true – recovered) divided by the reference noise image and pixels have only been included if they lie in the central region (30′ diameter) of the *Pong* and if the 500 μm flux in the *Herschel* image of Andromeda (Smith et al. 2012) is greater than 200 mJy beam$^{-1}$. The figure shows the best combination (smallest difference between input and output images) is a filter scale of ∼480–520″ and 5–10 PCA components.

for a filter scale of 520″ with 7 or 8 PCA components, concluding that the best results came with a filter scale of 520″ with 7 PCA components. However, in the next stage of the simulations we also tested the method with a filter scale of 480″ and 5 PCA components, in order to check whether the optimum values shifted if other parameters in the method were varied.

In the simulations, we also checked the number of iterations in the pipeline required for each set of parameters, since the computer processing time is directly proportional to the number of iterations. We found, as expected, that there is a processing cost to setting a larger filter scale (from 8 to 19 iterations) but including PCA analysis can reduce this.

### 5.3. *The Mask*

An important element in the SCUBA-2 data-reduction pipeline is a mask provided by the user as their best guess of the area in which real astronomical signal will be found. This greatly helps the convergence of the iterative procedure as it makes it easier for the algorithm to distinguish between real astronomical emission and extended structures that are actually the result of atmospheric emission or noise in the camera (see Figures 4 & 5). This does not mean that a pixel in the final image that is not within the mask will necessarily contain no astronomical signal. The software only subtracts astronomical signal from samples in the timelines that will

contribute to image pixels within the mask, but once all data-reduction stages are completed (subtraction of PCA components etc.) even pixels outside the mask, which are the averages of many samples in the timeline, may still contain astronomical signal.

The mask we used was created from the *Herschel* 500 μm image of Andromeda (Smith et al. 2012), the *Herschel* image closest in wavelength to our 850 μm image. We defined the mask as all pixels above a 500 μm flux-density threshold, with this threshold providing another knob we could twiddle in our analysis. Too high a threshold makes it possible for the algorithm to treat real astronomical signal as noise and remove it from the timelines; too low a threshold leads to slower convergence of the algorithm and higher noise. In the simulations described in the previous section, we set the 500 μm flux-density threshold at 200 mJy beam$^{-1}$, which seemed a reasonable compromise as the mask then included most of the disk and some inner regions of the galaxy while excluding some regions between the rings where the emission is faint.

Once we had identified the best combinations of filter scale and number of PCA components (520 arcsec and 7 PCA components or 480 arcsec and 5 PCA components), we used these to test the effect of changing the flux threshold used to construct the mask. We ran simulations with values of the 500 μm flux threshold between 120 and 520 mJy beam$^{-1}$. Figure 7 shows masks created with a range of flux thresholds that are representative of the ones we used in the simulations. Figure 8 shows the results of the simulations. The agreement between the input and output images is clearly best for a 500 μm flux-threshold of 280 mJy beam$^{-1}$ (this includes ∼25% of the total flux of M 31), which is the threshold we adopted to create the mask for the real 850 μm observations of Andromeda. We found that changing the flux threshold and thus the mask had a negligible effect on the noise of the final image, a very slight increase (< 1%) for masks generated with a 500 μm flux-threshold below 200 mJy beam$^{-1}$.

### 5.4. *Tolerance Level*

The next parameter we investigated was the map-tolerance parameter which the SCUBA-2 data-reduction pipeline uses to decide whether the algorithm has converged or whether more iterations are required. The default tolerance value is 0.05, which means that the iterations stop when the average change in the flux in a pixel from the last iteration is less than 0.05 σ, σ being the noise in that pixel (calculated from the distribution of instrument samples contributing to that pixel). There have, however, been some studies (Mairs et al. 2015;
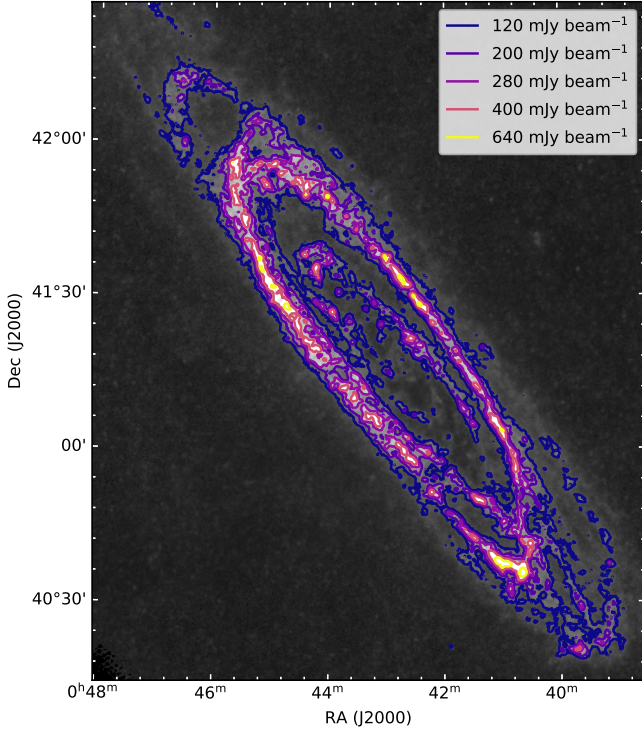
**Figure 7.** The grey-scale image is the 500 μm *Herschel* image which we used to create the mask used in the SCUBA-2 data-reduction pipeline. We defined the mask as all pixels with a 500 μm flux density greater than a threshold value. In the simulations described in Section 5.3, we tested the effect of changing this threshold value. The five contours show the masks created from flux thresholds that are representative of the ones we used in the simulations.

Smith et al. 2019) that suggest a lower tolerance value might improve results, something we wanted to explore in our simulations.

Given our huge volume of data, another important consideration was processing time, which increases if the tolerance value is reduced because a greater number of iterations are then needed to reach a lower tolerance value (processing time scales linearly with number of iterations). We therefore carried out simulations over a fairly small range of tolerance values: 0.0075–0.05. In Section 5.3 we found that we achieved the best results with a mask generated with a 500 μm threshold of 280 mJy beam$^{-1}$. In the simulations described in this section, we also experimented with masks created with 500 μm thresholds of 200 and 240 mJy beam$^{-1}$ to see whether the choice of best mask changed if we also changed the tolerance parameter. We also tried both winner and runner-up from the competition between filter-scale/PCA combinations of Section 5.2 to see if the order might be reversed with a different value of the tolerance parameter.
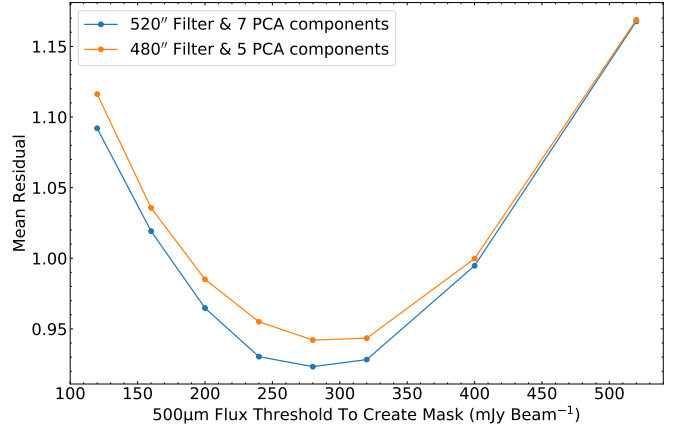


**Figure 8.** Mean absolute residual in the difference map versus the 500 μm flux-threshold used to define the mask used in the SCUBA-2 data-reduction pipeline (see the caption of Figure 6 for details of the difference map and of the region used to measure the statistic). The two lines are the results from using the best combinations of filter scale and number of PCA components identified during the simulations described in Section 5.2. The plot shows the best 500 μm flux-threshold for creating a mask is 280 mJy beam$^{-1}$ for both combinations of filter scale and PCA number.
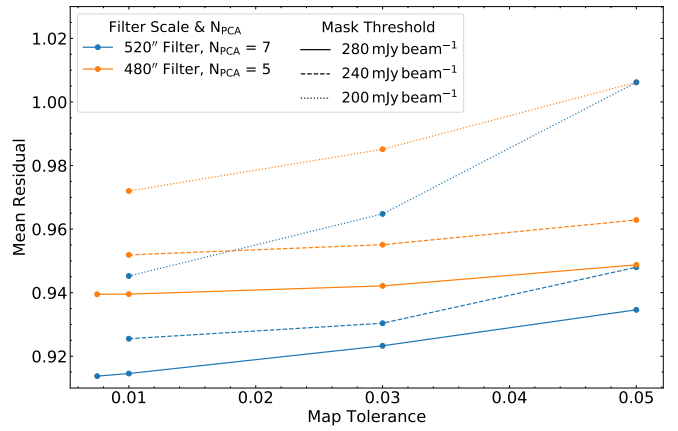


**Figure 9.** Mean absolute residual in the difference map versus the value of the tolerance parameter used in the SCUBA-2 data-reduction pipeline (see the caption of Figure 6 for details of the difference map and of the region used to measure the statistic). The blue lines show the results from simulations with a filter scale of 520″ and 7 PCA components, and the orange lines show the results from simulations with a filter scale of 480″ and 5 PCA components. The solid, dashed and dotted lines show the results from simulations with a mask made with a 500 μm flux threshold of 280, 240 and 200 mJy beam$^{-1}$, respectively. The plot shows that there is an improvement made by reducing the value of the tolerance parameter, but the improvement is modest, and there is a trade-off with an increase in computer processing time.
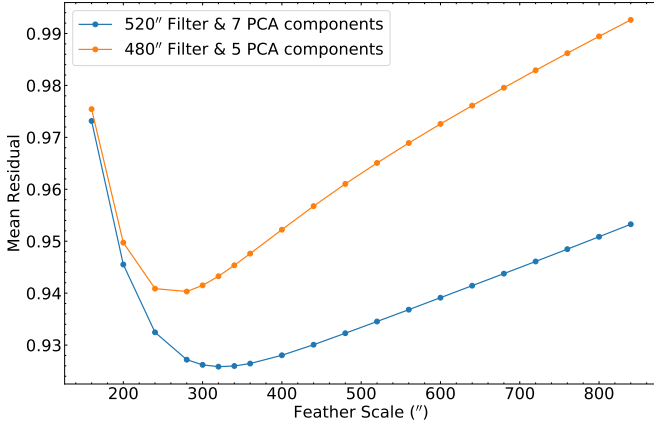
**Figure 10.** The mean absolute residual in the difference map versus the scale of the Gaussian filter, the 'feather scale', used to combine the SCUBA-2 and *Planck* image (see the caption of Figure 6 for details of the difference map and of the region used to measure the statistic). The two lines show the results of the simulations for our winner and runner-up combinations of high-pass filter scales in the SCUBA-2 data-reduction pipeline and number of PCA components. The plot shows we get the best results for our winner combination (520 arcsec and 7 PCA components) when we combine the SCUBA-2 and *Planck* images with a Gaussian filter with a scale of 320 arcsec.

Figure 9 shows that decreasing the value of the tolerance parameter does improve the agreement between the true and recovered images. They also show that the best choices for mask and filter-scale/PCA combination generally remain the best choices at all values of the tolerance parameter. The improvement with decreasing tolerance parameter is fairly slow, and there is a high price in increased processing time. Therefore, as a trade-off, we adopted a tolerance parameter of 0.03 for the real observations. Even with this tolerance parameter, reducing the current HASHTAG 850 μm dataset, which is only 70% of the final dataset, required 7.5 days of computer processing time.

### 5.5. *Feather Scale*

The final parameter we investigated was the scale of the Gaussian filter, the 'feather scale', used to combine the SCUBA-2 and *Planck* images. In this final round of simulations, we changed the pixel scale from 4.0 arcsec to 4.5 arcsec, which makes the pixels in the final 850 μm image close to one third of the full-width half maximum of the point spread function, the value that was eventually adopted after some experimentation for making *Herschel* images. Figure 10 shows the results from the simulations. The agreement between the 'true' input image and the recovered output image is best when the scale of the filter is 320″, which is roughly what we

expected; the resolution of *Planck* is 4.8′ (290 arcsec) (Planck Collaboration et al. 2014b) and so the *Planck* image should supply all Fourier components on angular scales larger than this. In all the 850 μm simulations the feathering method where the low-resolution image is not weighted is preferred (see Section 4.3).

### 5.6. *The final values of the parameters*

As the result of these simulations, we adopted the following values for all of the parameters when reducing the real SCUBA-2 data:

- We set the scale of the high-pass filter in the SCUBA-2 data-reduction pipeline to 520″ (FLT.FILT_EDGE_LARGESCALE=520).

- We set the number of PCA components per array in the SCUBA-2 data-reduction pipeline to 7 (PCA.PCATHRESH=-7) with the values of all the other parameters in the PCA analysis having their default values.

- We use an input mask to define the region where there is likely to be astronomical signal created from the *Herschel* 500 μm image (Smith et al. 2012). This is used by the SCUBA-2 data-reduction pipeline (the AST model in the pipeline). We defined the mask as all pixels with a 500 μm flux greater than 280 mJy beam$^{-1}$.

- We set the tolerance parameter in the SCUBA-2 data-reduction pipeline to 0.03.

- We use a pixel scale for the final 850 μm image of 4.5 arcsec.

- We used a Gaussian filter with a scale (the 'feathering scale') of 320″ to combine the final SCUBA-2 image with the *Planck* image.

These pipeline parameters are optimized for the observing strategy, source properties, and weather for HASHTAG and M 31. For other SCUBA-2 datasets with extended structure, we would recommend performing a similar simulation to optimize the processing, however, these results should provide a useful initial guess.

### 5.7. *A test of the image fidelity*

The final stage in the simulations was to assess the fidelity of the final image produced with the values of the parameters listed in the previous section. How close is the structure in the final image to the structure in the original true image? This analysis gives us a useful estimate of the fidelity of our real image. Note, however, that analysis will yield an upper limit to the errors on
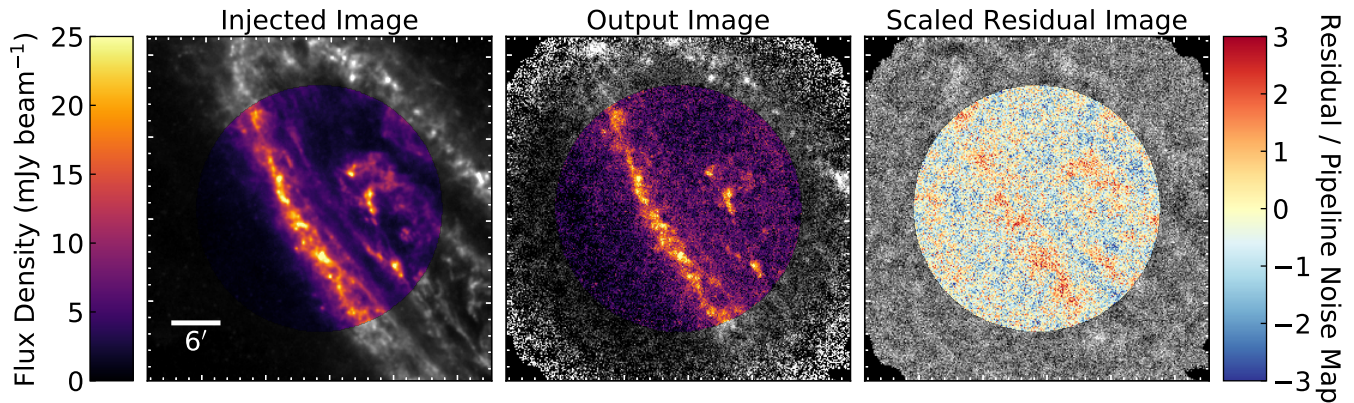
**Figure 11.** The 'true' input image we used in our simulations (left panel), the output image we recovered using the data-reduction parameters listed in Section 5.6 (middle panel) and the difference between the two divided by the noise estimate from the SCUBA-2 data-reduction pipeline (right panel). The colored area is the central circular region (diameter 30 arcmin) in which the observations have their full sensitivity.

the real image because the simulations have been carried out for only one *Pong* (Section 3); the spatial overlaps of the many *Pong* fields for the real observations (Figure 2) should improve the fidelity of the final real image.

Figure 11 shows the input 'true' image, the recovered output image, and the difference between the two divided by the noise image produced by the SCUBA-2 data-reduction pipeline. If our recovery method was perfect, the final panel should simply show random noise, but in fact there is some faint structure in the noise that is clearly correlated with bright structures. We therefore need to assess the importance of these systematic errors.

We estimated the random and systematic errors in the flux densities from a plot of $D = (F_{\rm o} - F_{\rm i})/\sigma_{\rm pipe}$ versus $F_{\rm i}$, in which $F_{\rm o}$ is the flux in a pixel in the output image, $F_{\rm i}$ is the flux in that pixel in the input image, and $\sigma_{\rm pipe}$ is the estimate from the SCUBA-2 data-reduction pipeline for the noise in that pixel. The top panel of Figure 12 is a surface-density plot showing how the number of pixels depends on $D$ and $F_{\rm i}$. We have only included pixels in the full-sensitivity central circular region of the image (diameter of 30 arcmin). If the fidelity of the final image was perfect, and if our noise estimates were correct, $D$ should have a Gaussian distribution around zero with a standard deviation of one. In reality, the plot shows that there is a clear bias in $D$, which is systematically higher than zero at high flux densities in the input true image. The standard deviation of $D$ is also slightly higher than one, showing the estimate of the noise produced by the pipeline is too low.

Given the discrepancy between the actual distribution of $D$ and its predicted distribution, we have assumed that the true uncertainty of the flux in each pixel is given by three uncertainties added in quadrature:

$$\sigma_{\rm tot}^2 = (a)^2 + (b\,\sigma_{\rm pipe})^2 + (f\,S_{\rm o})^2 \qquad (1)$$

in which $a$ is a constant, $b$ is a multiplicative factor for the error given by the SCUBA-2 data-reduction pipeline ($\sigma_{\rm pipe}$), and $f$ is a multiplicative factor for the flux in the output image ($S_{\rm o}$). The third error in Equation 1 is effectively a photometric calibration error, which exists for all astronomical observations, although in this case it is an error on top of the standard SCUBA-2 photometric calibration error, which we have not included in the equation. We estimated the values of $a$, $b$ and $f$ by applying the minimization package LMFIT (Newville et al. 2016) so that the standard deviation of $D$ approached as closely as possible a value of 1. We carried out the minimisation on a rolling group of 800 pixels ranked in flux. We found $a = 0$, $b = 1.04$, showing the SCUBA-2 data-reduction pipeline had slightly underestimated the random errors in the fluxes, and that $f = 0.12$, showing that there is a systematic error that depends on the brightness of the emission, confirming the qualitative impression produced by Figure 11.

The bottom-panel of Figure 12 shows the same surface-density plot of pixels as in the top panel but with the noise value predicted by the pipeline ($\sigma_{\rm pipe}$) replaced by the noise value calculated from Equation 1 ($\sigma_{\rm tot}$). The distribution shows that while the width of the distribution of $D$ now is roughly correct, the distribution is still distorted, in the sense that as the output flux density ($F_{\rm o}$) increases, the output flux density becomes progressively higher than the input flux density ($F_{\rm i}$), although there is a suggestion above $30\,{\rm mJy\,beam^{-1}}$ this reduces. We could have corrected the flux densities in the real HASHTAG image using the red curve in the bottom panel of the figure. We decided not to do this for two reasons. First, the real HASHTAG image is made of a large number of spatially overlapping datasets, so it is possible this effect is less for the real image. Second,
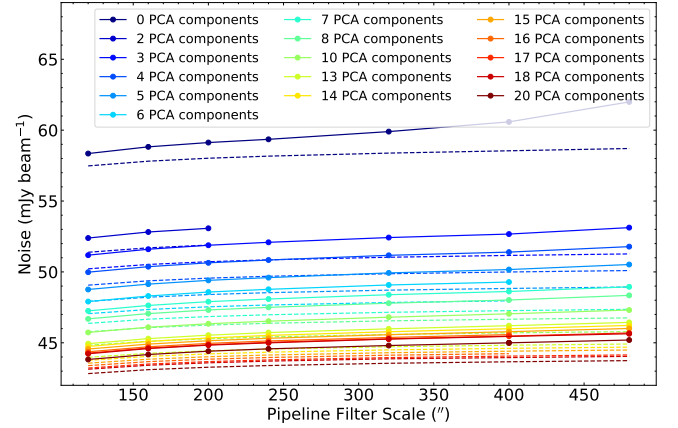
**Figure 13.** The same plot as Figure 5, but for 450 μm data rather than at 850 μm. Increasing the number of PCA components leads to a significant reduction in the noise measured in the image.



**Figure 14.** The same plot as Figure 6, but for 450 μm data rather than at 850 μm. See the caption of the earlier figure for details.

**Figure 12.** A surface-density plot showing how the number of pixels depends on $D$ and the input 'true' flux in a pixel. $D$ is the difference between the flux density in a pixel in the recovered image and the flux density in the true image divided by the noise in that pixel. The distribution has been renormalised, column by column, so that the figure is not dominated by the number of pixels with faint flux densities. The top panel shows the distribution if the noise that is used is the noise produced by the SCUBA-2 pipeline. The red and cyan lines are the rolling (800 pixels) mean and $\pm 1\sigma$ standard deviation, respectively. For a perfect observation the red line would follow a mean of zero (black line) and the standard deviation lines would follow the grey lines at $\pm 1$. The green vertical lines show the average $1\sigma$ noise in the true image. The bottom panel shows the same distribution when the noise has been rescaled using the method described in the text.

this systematic effect is fairly small compared with the statistical error: only $0.6\,\sigma$ at $F_{\mathrm{o}} = 21\,\mathrm{mJy\,beam}^{-1}$.

## 6. THE SIMULATIONS AT 450 μm

Optimising our data-reduction was much simpler at 450 μm than at 850 μm because at the shorter wavelength the space-based image has structure down to a much smaller angular scale (*Herschel* — 36 arcsec) than at 850 μm (*Planck* — 5 arcmin). The ranges of Fourier components of the space-based and SCUBA-2 observations are therefore much closer than at the long wave-

length. Nevertheless, we performed the same set of simulations as at 850 μm, and we summarize the results in this section.

Figure 13 shows how the noise in the 450 μm image varies with the filter-scale and the number of principal components used in the reduction. The number of PCA components is found to be particularly important with the noise changing from ∼60 to ∼45 mJy beam$^{-1}$ with increasing number of PCA components. As before, we investigated the effect on the mean absolute residual in the difference map of changing the filter scale in the pipeline and the number of PCA components. We tried filter scales between 120″ and 480″, with the lower bound chosen because it is used by the Cosmology Legacy Survey (Geach et al. 2013) which was optimized for detecting point sources. The number of PCA components was again from 0 to 20. Figure 14 shows that the
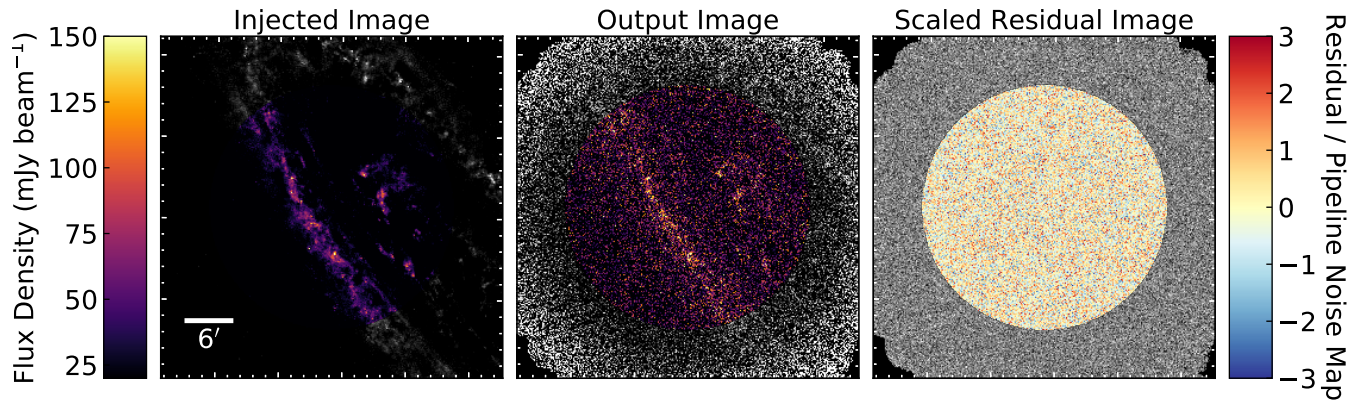
**Figure 15.** As for Figure 11, but for the 450 µm simulations.

filter scale has little effect on the mean absolute residual, but that it is reduced by increasing the number of PCA components. We decided to use a filter scale of $320''$ with 14 PCA components; there was little improvement by increasing the number of PCA components further and there was a marginal sign that using a filter with a smaller angular scale increased the mean absolute residual.

As before, we tested the effect of varying the 500 µm threshold used to make the mask and of varying the tolerance value, both of which are used in the SCUBA-2 pipeline. We found that the mean absolute residual in the difference map varied very little when either parameter was adjusted. We therefore decided to use the same values as at 850 µm.

We tested the effect of varying the feather scale, finding that the optimum feather scale was 40 arcsec, similar to the size of the *Herschel* beam at 500 µm. We found very little difference in the resulting image when using feathering scales up to 100 arcsec (above 50 arcsec the low-resolution data must be weighted in the feathering, see Section 4.3).

The final stage in the simulations was to test the fidelity of the image made with the values of the parameters above. We used exactly the same procedure as at 850 µm (Section 5.7). We found that at 450 µm the noise estimate from the pipeline ($\sigma_{\rm pipe}$) was a much better estimate of the true noise ($\sigma_{\rm tot}$) than at 850 µm. We found that the noise scaling term, $b$ in Equation 1, was only 1.02, and the other two terms ($a$ and $f$) were both zero. The input and the output images and the residual map divided by the noise value from the pipeline are shown in Figure 15. The facts that the residual map shows no structure at all demonstrates that at 450 µm the random errors are much greater than the systematic errors.

It is known *Herschel* can miss emission on the very largest scales, predominately due to the finite size of the images. Clark et al. (2021) have investigated this for the

Local Group, and found in the case of M 31 that very little extended dust emission is missing from the SPIRE 500 µm image. We therefore have chosen to feather with the normal SPIRE map, but have provided all the tools and instructions on our website[1] if users wish to feather with an alternative map.

## 7. THE REAL DATA

### 7.1. *Calibration*

The SCUBA-2 Makemap routine produces maps in instrumental units of picowatts, and so a flux conversion factor (FCF) is used to convert these units into either $\rm Jy\,beam^{-1}$ or $\rm Jy\,arcsec^{-2}$. There have recently been two changes at the JCMT which has adjusted the standard FCF values used at the observatory. First, in November 2016 a filter set in SCUBA-2 was changed, which predominantly affected the 850 µm FCF. Second, in May 2018 the maintenance of the secondary mirror resulted in a change in the value of the FCF for 450 µm. We have observations both before and after these changes. Based on an interim analysis by observatory staff (private communication), we assume FCF values of 3.62 and $2.14\,\rm Jy\,pW^{-1}\,arcsec^{-2}$ at 450 and 850 µm, respectively, for observations post May 2018[6]. For our older observations, we multiply the 'cleaned' data (see Section 4.1) by a correction factor, so that the final data can be calibrated with the same FCF. We adopted values for this correction factor of 1.21053 and 1.06481, for 450 and 850 µm, respectively.

---

[6] Since we calibrated the data, new (but still preliminary) values of the FCF have been released https://www.eaobservatory.org/jcmt/instrumentation/continuum/scuba-2/calibration/, which are slightly different from the values we use (resulting in approximately 3.5% and 5% lower flux-density at 450 and 850 µm, respectively). As the correction depends on individual observing conditions, we will apply the new calibration in the next data release.

The final factor we have to consider is that the standard JCMT calibration scheme is not designed for such extended objects as Andromeda. The JCMT calibration scheme (Dempsey et al. 2013) uses the flux of a calibrator source within a $30''$ radius aperture, after subtracting a background measured in an annulus around the calibrator between radii of 45 and 60 arcsec. This scheme is fine for calibrating images that contain point sources. But we are trying to calibrate very extended emission, and the beam of the telescope extends to much larger radii than the radii used in the standard calibration scheme.

We have adopted the calibration scheme we devised for the JINGLE Large Programme (Smith et al. 2019), where we multiply the FCFs so that the flux densities in the images matched the convention of other telescopes, that an aperture centered on a galaxy would only include the entire flux of the galaxy if the radius was increased to infinity. By integrating the beam model in Dempsey et al. (2013), we calculated that the standard $850\,\mu$m FCF should be multiplied by 0.91, which agreed with the curve of growth found by Dempsey et al. (2013). Doing the same calculation at $450\,\mu$m, we obtain a correction factor of 0.99. The curve of growth given in Dempsey et al. (2013), however, suggests a slightly smaller factor of 0.97. We suspect the difference is caused by extra non-Gaussian features in the beam at large radii. We therefore adopt the smaller value of 0.97. The final values of the FCF used to create these HASHTAG images were therefore the FCF values given above multiplied by these correction factors. These are 3.51 and $1.95\,\mathrm{Jy\,pW^{-1}\,arcsec^{-2}}$ at 450 and $850\,\mu$m, respectively.

### 7.2. Final Maps

The real HASHTAG data was reduced using the method outlined in the previous subsections, using the best values of the parameters found from the simulations. Our final maps (Data Release 1, DR1) are composed of two multi-extension fits files which contain the flux-density, uncertainty, and sensitivity maps for the 450 and $850\,\mu$m images, respectively.

The uncertainty map contains the true uncertainty values we derived in Section 5.7, which includes both the statistical uncertainty in each pixel and the systematic uncertainty as the result of the flux density in that pixel (the third term in Equation 1). The sensitivity map is the same except that this map does not include the systematic term ($f$ in Equation 1 is set to zero). At both wavelengths, the sensitivity of our final images exceeds our targets. In the 10 kpc ring, the typical sensitivity is $\sim$2.0 and $\sim$30 mJy beam$^{-1}$ at 850 and $450\,\mu$m, respectively, with peak sensitivities in the ring of 1.5 and

$20.6\,\mathrm{mJy\,beam^{-1}}$ at 850 and $450\,\mu$m, respectively. As a rough comparison the *Herschel* $500\,\mu$m observations have an instrumental sensitivity of $\sim$11 mJy beam$^{-1}$ (Smith et al. 2017), and the point source sensitivity of *Planck* at $850\,\mu$m is $\sim$69 mJy beam$^{-1}$ (Planck Collaboration et al. 2014c).

The flux-density and sensitivity maps are shown in Figure 16. The sensitivity is not quite as uniform at $450\,\mu$m, which we attribute to variations in sub-mm opacity during the periods we took the data, since opacity variations have a bigger effect at the shorter wavelength.

As well as the images shown in Figure 16, we also provide versions that have three different levels of Gaussian smoothing (for example, Figure 1), so that users can balance resolution versus signal-to-noise ratio. In these smoother images, the raw images have been smoothed with a Gaussian with a FWHM of 4, 5 and $7.9''$ at $450\,\mu$m and 7, 10 and $13''$ at $850\,\mu$m (this equates to effective resolutions of 8.9, 9.3, and $11.2''$ at $450\,\mu$m, and 15.2, 16.8, and $19.1''$ at $850\,\mu$m).

## 8. CO($J$=3–2) SUBTRACTION

A possible contamination in the $850\,\mu$m image is the contribution of the CO($J$=3–2) line which has a rest frequency of 345.796 GHz, putting it within the $850\,\mu$m passband. This line has been shown to contribute anything from 0.7% to 41% of the $850\,\mu$m emission in nearby galaxies, although it is normally less than 15% (Smith et al. 2019), and in the Milky Way the contamination is typically small ($<$5%) but can be 30% (Moore et al. 2015). The contamination seems likely to be low in M 31 because of the low fraction of molecular gas. We can get a rough idea of the likely scale of the contamination using our CO($J$=3–2) survey of selected regions within the galaxy (Li et al. 2020). In our CO survey the strongest line flux we found was $I_{\mathrm{CO}(J=3-2)} \simeq 5\,\mathrm{K\,km\,s^{-1}}$, which corresponds, using the relationship given in Parsons et al. (2018), to an $850\,\mu$m flux density of $\simeq$3 mJy beam$^{-1}$, $\simeq$1.5 times the typical noise (Section 7.2).

We used the results of our CO($J$=3–2) survey, which covered small regions but over a range of environments (Figure 2), as the ground truth for developing a method to estimate the CO contamination at all points across M 31. As our starting point, we used the CO($J$=3–2) cubes presented by Li et al. (2020). However, we created a new set of integrated intensity maps (moment-0 maps), using a new VLA H I dataset (Koch et al. 2021). The VLA dataset provides a $58''$ resolution (FWHM) image of the entire galaxy, and a higher resolution $18''$ image of the region covered by *Hubble*. To create the CO
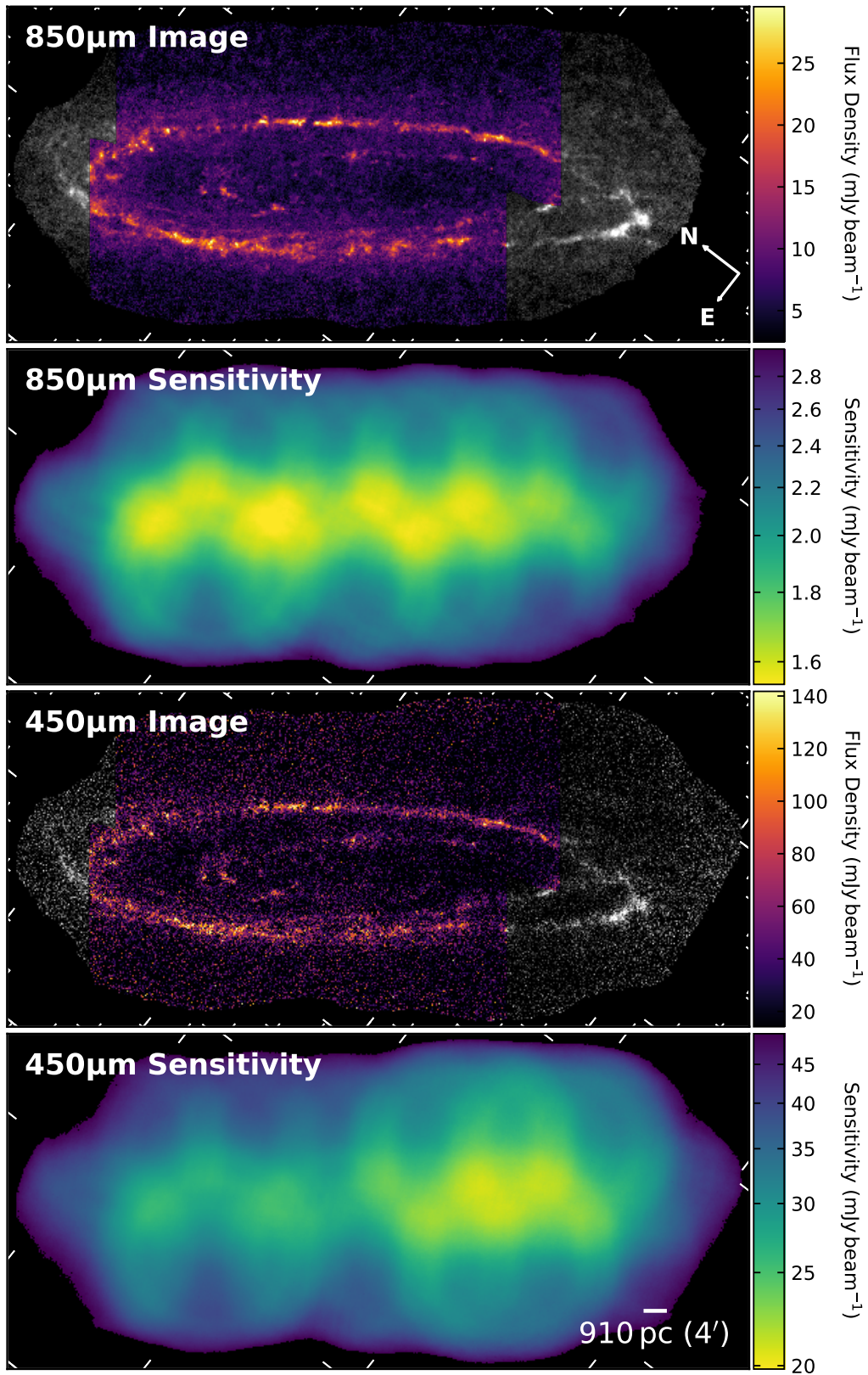
**Figure 16.** Our final 450 and 850 μm images and sensitivity maps for HASHTAG Data Release 1. The colored regions of the images are approximately where our observations and complete and at our full sensitivity (equivalently, the grey-scale shows regions where observations are still ongoing). The images have a resolution of $7.9''$ and $13.0''$(FWHM), at 450 and 850 μm, respectively, and are available in both $\mathrm{mJy\,beam^{-1}}$ and $\mathrm{mJy\,arcsec^{-2}}$ units. The sensitivity maps are shown on a log scale and are described in Section 7.2.

integrated intensity maps we use the H I data as a prior to mask the channels in the cube that are not expected to contain any CO emission, using the moment-1 map to predict the center of the line with a width based on the H I line-width (with a minimum $40 \, \mathrm{km \, s^{-1}}$ adopted). Using the H I as a prior is similar technique to that applied by Schruba et al. (2011).

Our overall approach was to use a combination of several datasets to predict the CO($J$=3–2) emission, using our own CO($J$=3–2) survey to determine the combination that gives the best prediction. We included maps of the dust emission (column-density, temperature, and emissivity index; Whitworth et al. 2019), images in the WISE bands (Wright et al. 2010; Cutri et al. 2013), in the UV (Thilker et al. 2005), in the mid-infrared (*Spitzer*, MIPS; Gordon et al. 2006), a survey of part of the galaxy with CARMA in CO($J$=1–0) (Caldú-Primo & Schruba 2016, A. Schruba et al., *in preparation*), and a map of the estimated star-formation rate in the galaxy calculated using UV and 24 µm (Ford et al. 2013). While some of these datasets may be degenerate, our aim was to find the best model to predict the CO($J$=3–2) line flux, rather than understanding the physical meaning of the obtained model.

The most useful dataset for predicting the CO($J$=3–2) emission is a map in another CO line. We decided not to use the map in the CO($J$=1–0) line over the whole galaxy (Nieten et al. 2006) because its resolution (23″) is significantly lower than that of our 850 µm image (13″), and we found a model derived at 23″ resolution when applied on 13″ scales did not perform as well as one derived at the higher resolution. The CARMA survey was our key dataset because it was a survey in the CO($J$=1–0) line with a resolution of 5.5″ (we use the version corrected for missing large-scale CO emission using the IRAM single-dish data). The survey, however, only covers $\sim$323 arcmin$^2$ (see Figure 19) and our CO($J$=3–2) survey covers an even smaller region (see Figure 2). We were therefore forced to develop a two-stage method.

In the first stage of the approach, we restricted our analysis to the small portion of the galaxy covered by our CO($J$=3–2) survey that was also within the region covered by CARMA. We do not explicitly include the CO($J$=3–2)/CO($J$=1–0) ratio which has been found to vary across M 31 (Li et al. 2020), but implicitly include it in the model. We first performed a background subtraction on the input continuum images where necessary, convolved these images to the same resolution as the 850 µm/JCMT CO products, and finally re-projected them to match each of the six JCMT CO($J$=3–2) integrated intensity maps that overlap with the CARMA field. We took as our inputs the logarithms of all the

**Table 1.** Parameters of CO($J$=3–2) Models

| 'Feature' Image | Gradient Coefficient ($m_i$) | |
| --- | --- | --- |
| | Including CARMA | Excluding CARMA |
| CARMA | $0.403 \pm 0.020$ | — |
| Dust Surface-Density | $0.076 \pm 0.023$ | $0.220 \pm 0.004$ |
| Dust Temperature | $0.060 \pm 0.019$ | $0.160 \pm 0.005$ |
| Dust $\beta$ | $0.032 \pm 0.009$ | $0.030 \pm 0.003$ |
| MIPS 24 µm | $0.014 \pm 0.049$ | $0.133 \pm 0.010$ |
| SFR Surface-Density | $0.011 \pm 0.039$ | $-0.115 \pm 0.009$ |
| WISE W1 | $-0.173 \pm 0.053$ | $-0.153 \pm 0.011$ |
| WISE W2 | $0.046 \pm 0.058$ | $0.079 \pm 0.012$ |
| WISE W3 | $0.030 \pm 0.044$ | $-0.076 \pm 0.008$ |
| WISE W4 | $0.022 \pm 0.042$ | $0.063 \pm 0.010$ |
| Constant ($c$) | $-0.288 \pm 0.016$ | $0.420 \pm 0.0003$ |

NOTE—The best-fit parameters from the model described by Equation 2, for both the model including CARMA observations ('stage 1') or excluding CARMA observations ('stage 2'). See Section 8 for more details.

input maps except for the dust temperature and emissivity index. We assigned the pixels randomly into a training set (80% of the data) and a testing set (20% of the data). We then used the SCIKIT-LEARN (Pedregosa et al. 2011) 'standard scaler' routine which standardizes each of the inputs by removing the mean and dividing by the standard deviation. To perform the fitting, we tried both linear and non-linear methods (random forest, and the Multi-layer Perception neural network) but found the linear model performed as well as the more complicated routines. As the JCMT data is relatively noisy, to incorporate the uncertainties we built a model using PYMC3 (Salvatier et al. 2016), in which the predicted CO value is given by:

$$y_{\mathrm{model}} = 10^{c \cdot \sum_i m_i x_i} \qquad (2)$$

in which $c$ is a constant, $m_i$ is the gradient of each input, and $x_i$ is each input 'feature' (i.e., each input image). For both the constant and gradients we assumed a weakly-informative Gaussian prior with $\mu = 0$ and $\sigma = 10$ for the intercept and $\sigma = 20$ for the gradient. Table 1 provides the best-fitting values of $m_i$ and and $c$.

Figure 17 shows a plot of the predicted versus the measured CO($J$=3–2) emission. Above $1.0 \, \mathrm{K \, km \, s^{-1}}$ the average accuracy is $\sim$28%, although the true uncertainty may be lower as there is significant uncertainty associated with some of the data points.
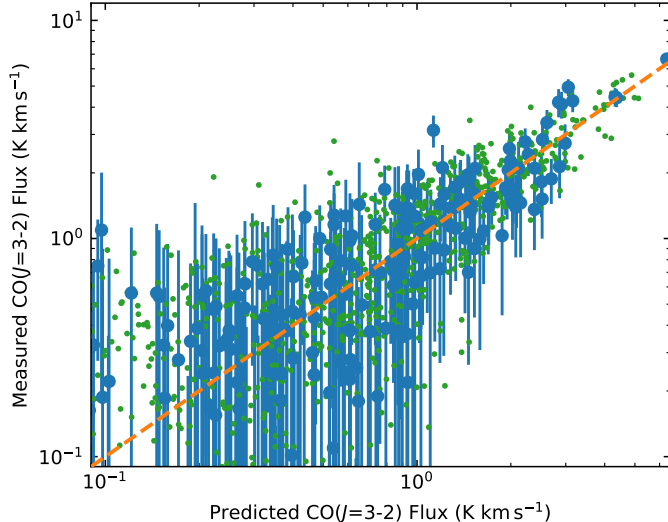
**Figure 17.** The measured CO($J$=3–2) flux versus the flux predicted by our linear model for the six regions of our CO($J$=3–2) survey that fall within the region of the CARMA($J$=1–0) survey. The blue data points are for the 20% of pixels that are in our test dataset, which we did not use in determining the best combination of parameters. The green points are from our training dataset. At low line-fluxes the scatter is dominated by the uncertainty in the JCMT CO($J$=3–2) observations. The orange line shows the 1-to-1 line that we would achieve if our prediction method was perfect.

In the second stage of our method, we extend our analysis to create a model for regions of M 31 where we do not have CARMA CO($J$=1–0). To create this model we extend our analysis to the entire CARMA region, for which we have CO($J$=1–0) observations but not CO($J$=3–2) observations. In this much larger region we use the linear combination we derived in stage 1 to predict the CO($J$=3–2) fluxes. We then use these predictions as the 'measurements' in this stage of the analysis, as well as our JCMT CO($J$=3–2) measurements not used in 'stage 1' (e.g., outside the CARMA footprint). In this stage we use equation 2 as above to determine the combination of inputs that makes the best prediction of the CO($J$=3–2) 'measurements', except this time we do not use the CARMA CO($J$=1–0) measurements as one of the inputs.

We found, as before, that there was no advantage when using the non-linear methods, so we used the simpler linear method. CARMA covers a large continuous region, so instead of assigning pixels randomly to the training and test data, we used slices in declination, which avoids pixels in the same cloud being assigned to both datasets. Since predicted CO($J$=3–2) fluxes below $0.5\,\mathrm{K\,km\,s^{-1}}$ correspond to 850 µm fluxes significantly less than the statistical noise in the HASHTAG image,
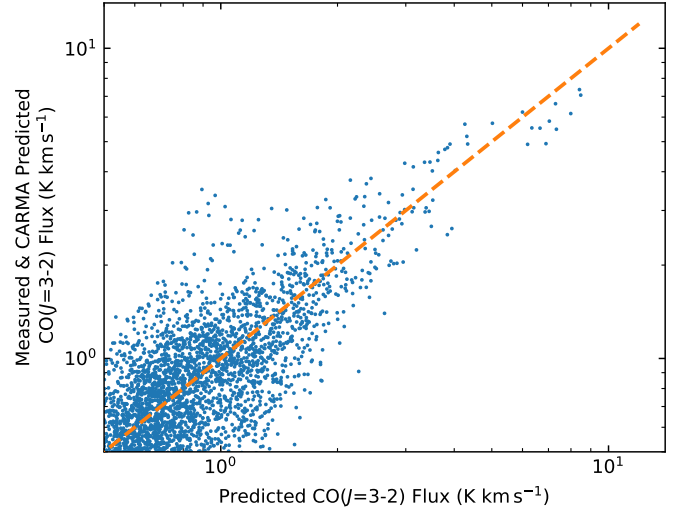


**Figure 18.** The CO($J$=3–2) line flux predicted by our final model versus CO($J$=3–2) measurements (either our real measurements or the predictions from stage 1). The blue data points are from our 'test' dataset (20% of the pixels), which was not used to derive the model. Error bars have not been included for clarity. The orange line shows the 1-to-1 line for a perfect prediction.

we only trained our model on regions with CO($J$=3–2) 'measurements' greater than this value. Figure 18 shows the relationship between the prediction of this new model versus the CO($J$=3–2) measurements (either our real measurements or the predictions from stage 1).

Figure 19 shows the CO($J$=3–2) line flux predicted by our models. Inside the CARMA region, where we have CO($J$=1–0) measurements, we have used the stage 1 model. Outside the CARMA region, we used our stage 2 model. The statistical noise in the HASHTAG 850 µm image is $\simeq 2\,\mathrm{mJy}$, which corresponds to a CO($J$=3–2) line flux of $\simeq 3\,\mathrm{K\,km\,s^{-1}}$. The figure shows that generally the line contamination is not a problem. In bright cores, though, it can be important. If $\sigma_{850\mu\mathrm{m}}$ is the statistical uncertainty in the 850 µm without the inclusion of the systematic term (the third term in equation 1), the maximum CO($J$=3–2) signal is $\simeq 5.7\sigma_{850\,\mu\mathrm{m}}$. But if the systematic term is included, this reduces to $\simeq 1.5\,\sigma$. If only pixels are included where the signal-to-noise ratio of the 850 µm image is greater than $3\,\sigma$ (not including the systematic term), the maximum contamination in a pixel is 28%, but in 80% of the pixels the CO line flux is less than $0.5\,\mathrm{K\,km\,s^{-1}}$, which is only $\simeq 16\%$ of the statistical noise in the continuum map. In general, then, contamination by the CO($J$=3–2) line is not a significant problem. We have provided 850 µm images in the data release both with and without a correction for line contamination, allowing users to either ignore the effect
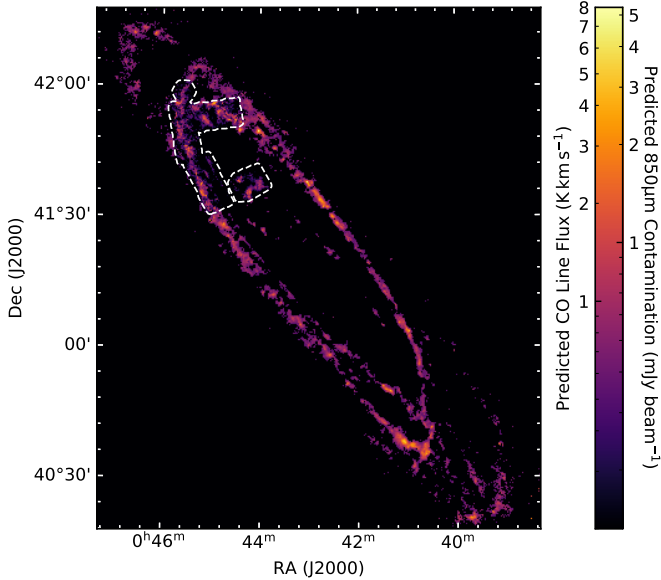
**Figure 19.** Our map of the CO line predicted (or equivalently 850 μm continuum contamination) using the method described in Section 8. Outside the CARMA region, any pixels with predicted line fluxes $<0.5\,\mathrm{K\,km\,s^{-1}}$ have been set to zero, because these pixels fall outside the range of CO line fluxes where the model was trained. The peak predicted line flux is $\sim 15\,\mathrm{K\,km\,s^{-1}}$, but we have capped the color-bar at $8\,\mathrm{K\,km\,s^{-1}}$ to aid visibility. The white-dashed line shows the region covered by CARMA CO($J$=1–0) observations.

of line contamination completely, use our corrected image, or make their own correction.

## 9. CONCLUSIONS

We have presented submillimeter images of the Andromeda Galaxy obtained at 450 and 850 μm, the first images made from the ground that properly represent the structure of the galaxy on all spatial scales. We have described the method we have developed to optimize the SCUBA-2 pipeline for M 31 and how we use a feathering technique to combine the small-scale structure (high-$k$ Fourier components) from SCUBA-2 and data from space observatories (*Herschel* and *Planck*) to provide the large-scale structure (low-$k$ Fourier components).

We describe the maps that comprise the HASHTAG DR1 data release which have a typical sensitivity of $\sim 2.0$ and $\sim 30\,\mathrm{mJy\,beam^{-1}}$ at 850 and 450 μm, respectively (at native SCUBA-2 resolution). As the CO($J$=3–2) line falls with the bandpass of the 850 μm band we derive a method to predict the line flux across M 31, and find while generally the contamination is small compared with the uncertainty in our continuum measurements, for some bright regions of the ring the contamination

is significant. We provide data products both with and without the CO correction, and at different resolutions.

## REFERENCES

Abdo, A. A., Ackermann, M., Ajello, M., et al. 2010, ApJ, 710, 133, doi: 10.1088/0004-637X/710/1/133

Athikkat-Eknath, G., Eales, S. A., Smith, M. W. L., et al. 2021, MNRAS, submitted

Bajaja, E., & van Albada, G. D. 1979, A&A, 75, 251

Barmby, P., Ashby, M. L. N., Bianchi, L., et al. 2006, ApJL, 650, L45, doi: 10.1086/508626

Bendo, G. J., Boselli, A., Dariush, A., et al. 2012, MNRAS, 419, 1833, doi: 10.1111/j.1365-2966.2011.19735.x

Bendo, G. J., Baes, M., Bianchi, S., et al. 2015, MNRAS, 448, 135, doi: 10.1093/mnras/stu1841

Block, D. L., Bournaud, F., Combes, F., et al. 2006, Nature, 443, 832, doi: 10.1038/nature05184

Bolatto, A. D., Wolfire, M., & Leroy, A. K. 2013, ARA&A, 51, 207, doi: 10.1146/annurev-astro-082812-140944

Braun, R., Thilker, D. A., Walterbos, R. A. M., & Corbelli, E. 2009, ApJ, 695, 937, doi: 10.1088/0004-637X/695/2/937

Buckle, J. V., Hills, R. E., Smith, H., et al. 2009, MNRAS, 399, 1026, doi: 10.1111/j.1365-2966.2009.15347.x

Caldú-Primo, A., & Schruba, A. 2016, AJ, 151, 34, doi: 10.3847/0004-6256/151/2/34

Chapin, E. L., Berry, D. S., Gibb, A. G., et al. 2013, MNRAS, 430, 2545, doi: 10.1093/mnras/stt052

Clark, C. J. R., Roman-Duval, J. C., Gordon, K. D., Bot, C., & Smith, M. W. L. 2021, arXiv e-prints, arXiv:2107.14302. https://arxiv.org/abs/2107.14302

Csengeri, T., Weiss, A., Wyrowski, F., et al. 2016, A&A, 585, A104, doi: 10.1051/0004-6361/201526639

Currie, M. J., Berry, D. S., Jenness, T., et al. 2014, in Astronomical Society of the Pacific Conference Series, Vol. 485, Astronomical Data Analysis Software and Systems XXIII, ed. N. Manset & P. Forshay, 391

Cutri, R. M., Wright, E. L., Conrow, T., et al. 2013, Explanatory Supplement to the AllWISE Data Release Products, Explanatory Supplement to the AllWISE Data Release Products

Dalcanton, J. J., Williams, B. F., Lang, D., et al. 2012, ApJS, 200, 18, doi: 10.1088/0067-0049/200/2/18

Davies, L. J. M., Driver, S. P., Robotham, A. S. G., et al. 2016, MNRAS, 461, 458, doi: 10.1093/mnras/stw1342

de Grijs, R., & Bono, G. 2014, AJ, 148, 17, doi: 10.1088/0004-6256/148/1/17

Dempsey, J. T., Friberg, P., Jenness, T., et al. 2013, MNRAS, 430, 2534, doi: 10.1093/mnras/stt090

Draine, B. T., Aniano, G., Krause, O., et al. 2014, ApJ, 780, 172, doi: 10.1088/0004-637X/780/2/172

Dunne, L., Maddox, S. J., Vlahakis, C., & Gomez, H. L. 2021, MNRAS, 501, 2573, doi: 10.1093/mnras/staa3526

Eales, S., Smith, M. W. L., Auld, R., et al. 2012, ApJ, 761, 168, doi: 10.1088/0004-637X/761/2/168

Ford, G. P., Gear, W. K., Smith, M. W. L., et al. 2013, ApJ, 769, 55, doi: 10.1088/0004-637X/769/1/55

Fritz, J., Gentile, G., Smith, M. W. L., et al. 2012, A&A, 546, A34, doi: 10.1051/0004-6361/201118619

Geach, J. E., Chapin, E. L., Coppin, K. E. K., et al. 2013, MNRAS, 432, 53, doi: 10.1093/mnras/stt352

Geach, J. E., Dunlop, J. S., Halpern, M., et al. 2017, MNRAS, 465, 1789, doi: 10.1093/mnras/stw2721

Gordon, K. D., Bailin, J., Engelbracht, C. W., et al. 2006, ApJL, 638, L87, doi: 10.1086/501046

Herschel, W. 1785, Philosophical Transactions of the Royal Society of London Series I, 75, 213

Hildebrand, R. H. 1983, QJRAS, 24, 267

Holland, W. S., Bintley, D., Chapin, E. L., et al. 2013, MNRAS, 430, 2513, doi: 10.1093/mnras/sts612

Hubble, E. P. 1929, ApJ, 69, 103, doi: 10.1086/143167

Hunt, L. K., Draine, B. T., Bianchi, S., et al. 2015, A&A, 576, A33, doi: 10.1051/0004-6361/201424734

James, A., Dunne, L., Eales, S., & Edmunds, M. G. 2002, MNRAS, 335, 753, doi: 10.1046/j.1365-8711.2002.05660.x

Kapala, M. J., Sandstrom, K., Groves, B., et al. 2015, ApJ, 798, 24, doi: 10.1088/0004-637X/798/1/24

Kasliwal, M. M., Bally, J., Masci, F., et al. 2017, ApJ, 839, 88, doi: 10.3847/1538-4357/aa6978

Kennicutt, R. C., & Evans, N. J. 2012, ARA&A, 50, 531, doi: 10.1146/annurev-astro-081811-125610

Kirk, J. M., Gear, W. K., Fritz, J., et al. 2015, ApJ, 798, 58, doi: 10.1088/0004-637X/798/1/58

Koch, E. W., Rosolowsky, E. W., Leroy, A. K., et al. 2021, MNRAS, 504, 1801, doi: 10.1093/mnras/stab981

Lamperti, I., Saintonge, A., De Looze, I., et al. 2019, MNRAS, 489, 4389, doi: 10.1093/mnras/stz2311

Lellouch, E., & Amri, H. 2008, Mars Brightness Model, http://www.lesia.obspm.fr/perso/emmanuel-lellouch/mars/

Lewis, A. R., Dolphin, A. E., Dalcanton, J. J., et al. 2015, ApJ, 805, 183, doi: 10.1088/0004-637X/805/2/183

Lewis, A. R., Simones, J. E., Johnson, B. D., et al. 2017, ApJ, 834, 70, doi: 10.3847/1538-4357/834/1/70

Li, Z., Li, Z., Smith, M. W. L., et al. 2020, MNRAS, 492, 195, doi: 10.1093/mnras/stz3409

Magdis, G. E., Daddi, E., Béthermin, M., et al. 2012, ApJ, 760, 6, doi: 10.1088/0004-637X/760/1/6

Mairs, S., Johnstone, D., Kirk, H., et al. 2015, MNRAS, 454, 2557, doi: 10.1093/mnras/stv2192

Marsh, K. A., Whitworth, A. P., & Lomax, O. 2015, MNRAS, 454, 4282, doi: 10.1093/mnras/stv2248

McMullin, J. P., Waters, B., Schiebel, D., Young, W., & Golap, K. 2007, in Astronomical Society of the Pacific Conference Series, Vol. 376, Astronomical Data Analysis Software and Systems XVI, ed. R. A. Shaw, F. Hill, & D. J. Bell, 127

Molinari, S., Schisano, E., Elia, D., et al. 2016, A&A, 591, A149, doi: 10.1051/0004-6361/201526380

Moore, T. J. T., Plume, R., Thompson, M. A., et al. 2015, MNRAS, 453, 4264, doi: 10.1093/mnras/stv1833

Newville, M., Stensitzki, T., Allen, D. B., et al. 2016, Lmfit: Non-Linear Least-Square Minimization and Curve-Fitting for Python. http://ascl.net/1606.014

Nieten, C., Neininger, N., Guélin, M., et al. 2006, A&A, 453, 459, doi: 10.1051/0004-6361:20035672

Orton, G. S., Moses, J. I., Fletcher, L. N., et al. 2014, Icarus, 243, 471, doi: 10.1016/j.icarus.2014.07.012

Pardo, J. R., Cernicharo, J., & Serabyn, E. 2001, IEEE Transactions on Antennas and Propagation, 49, 1683, doi: 10.1109/8.982447

Parsons, H., Dempsey, J. T., Thomas, H. S., et al. 2018, ApJS, 234, 22, doi: 10.3847/1538-4365/aa989c

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Journal of Machine Learning Research, 12, 2825

Pineda, J. L., Langer, W. D., Velusamy, T., & Goldsmith, P. F. 2013, A&A, 554, A103, doi: 10.1051/0004-6361/201321188

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2011, A&A, 536, A19, doi: 10.1051/0004-6361/201116479

Planck Collaboration, Abergel, A., Ade, P. A. R., et al. 2014a, A&A, 571, A11, doi: 10.1051/0004-6361/201323195

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014b, A&A, 571, A7, doi: 10.1051/0004-6361/201321535

—. 2014c, A&A, 571, A28, doi: 10.1051/0004-6361/201321524

—. 2015, A&A, 582, A28, doi: 10.1051/0004-6361/201424643

Rémy-Ruyer, A., Madden, S. C., Galliano, F., et al. 2014, A&A, 563, A31, doi: 10.1051/0004-6361/201322803

Roussel, H. 2013, PASP, 125, 1126, doi: 10.1086/673310

Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. 2016, PyMC3: Python probabilistic programming framework. http://ascl.net/1610.016

Sandstrom, K. M., Leroy, A. K., Walter, F., et al. 2013, ApJ, 777, 5, doi: 10.1088/0004-637X/777/1/5

Schruba, A., Leroy, A. K., Walter, F., et al. 2011, AJ, 142, 37, doi: 10.1088/0004-6256/142/2/37

Scoville, N., Aussel, H., Sheth, K., et al. 2014, ApJ, 783, 84, doi: 10.1088/0004-637X/783/2/84

Smith, M. W. L., Eales, S. A., Gomez, H. L., et al. 2012, ApJ, 756, 40, doi: 10.1088/0004-637X/756/1/40

Smith, M. W. L., Ibar, E., Maddox, S. J., et al. 2017, ApJS, 233, 26, doi: 10.3847/1538-4365/aa9b35

Smith, M. W. L., Clark, C. J. R., De Looze, I., et al. 2019, MNRAS, 486, 4166, doi: 10.1093/mnras/stz1102

Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. 2014, ApJS, 214, 15, doi: 10.1088/0067-0049/214/2/15

Stanimirovic, S. 2002, in Astronomical Society of the Pacific Conference Series, Vol. 278, Single-Dish Radio Astronomy: Techniques and Applications, ed. S. Stanimirovic, D. Altschuler, P. Goldsmith, & C. Salter, 375–396. https://arxiv.org/abs/astro-ph/0205329

Stiele, H., Pietsch, W., Haberl, F., et al. 2011, A&A, 534, A55, doi: 10.1051/0004-6361/201015270

Tabatabaei, F. S., Braine, J., Xilouris, E. M., et al. 2014, A&A, 561, A95, doi: 10.1051/0004-6361/201321441

Thilker, D. A., Hoopes, C. G., Bianchi, L., et al. 2005, ApJL, 619, L67, doi: 10.1086/424816

Viaene, S., Baes, M., Tamm, A., et al. 2017, A&A, 599, A64, doi: 10.1051/0004-6361/201629251

Whitworth, A. P., Marsh, K. A., Cigan, P. J., et al. 2019, MNRAS, 489, 5436, doi: 10.1093/mnras/stz2166

Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868, doi: 10.1088/0004-6256/140/6/1868

Yin, J., Hou, J. L., Prantzos, N., et al. 2009, A&A, 505, 497, doi: 10.1051/0004-6361/200912316