SCUBA-2 DR Pipeline Project Office
University of British Columbia
6224 Agricultural Road
Vancouver, British Columbia
CANADA
V6T 1Z1

Tel: +1-604-822-2211
Fax: +1-604-822-5324
Email: jmolnar@ubc.ca
WWW: http://scuba2.jach.hawaii.edu

Document Title: SCAN Mode Data Reduction Strategies for SCUBA-2

Document Number: SC2/ANA/S210/006

Issue: 1.10

Date: 2005/05/27

Document Prepared By:	Douglas Scott	Signature and Date:	2005-05-26
Document Approved By:	Janos Molnar	Signature and Date:	2005-05-26
Document Released By:	Janos Molnar	Signature and Date:	2005-05-26

Contents

1 Introduction

This document summarizes ideas for the approach to data reduction algorithms for SCUBA-2 SCAN mode.

Detailed strategies for SCAN mode are discussed in an accompanying document SC2/ANA/S210/006.[26]

The basic reduction approach is to construct an estimate for the maximum likelihood image in the presence of several different sources of noise.[20] This can be done by fitting one model at a time, and then iterating through subsets of models if necessary.

2 Basic principles

2.1 A map is a nearly lossless compression of timestreams

Images are probably the most fundamental kind of astronomical information. Precisely how raw data are converted into an image depends on different detector technologies, different wavelengths and different science goals. We will here be assuming that the primary output of a data reduction pipeline is a 'map', i.e. a 2-dimensional representation of the data which contains all (or as much as possible) of the spatial information contained in the raw data. So, apart from the possibility of some time variability in the data, we should regard the map as essentially a lossless (or as nearly lossless as possible) compression of a set of timestreams coming from the detectors. In other words, it ought to be possible to derive all results about structures on the sky from the properties of the map, without the need to go back to the raw data. This means, for example, that we shall not be content to find an efficient method which only works for point sources, but shall be concerned to retain information on scales all the way up to those of the largest modes which are reliably probed in the data (which might be much smaller than the map if some filtering needs to be applied to reduce artefacts). Such compression of the timestreams into maps will be crucial for SCUBA-2, where it is expected that the data rate will be approximately $0.5\,$TB per night, while a single $1^\circ \times 1^\circ$ map will take only $\sim 1\,$GB of storage, even at full resolution and with a generous number of bits per pixel.

2.2 Data products

It will be important to estimate the noise as well as the signal for each of the sky pixels. In practice there may be more than one way to estimate this noise. The SCUBA-2 software will allow for at least a couple of possibilities, namely: the noise estimated simply from the time-weighted average of the rms's of the bolometers which contribute to a given pixel; and the noise estimated from the spread in the values which contribute to a given sky pixel[21] (and these are distinct from the 'spatial' noise calculated from an aperture in the map itself).

Another fundamental principle is that we shall *not* be attempting to deconvolve the sky with the beam. The 'map' that we are trying to estimate is the underlying astronomical sky convolved with the telescope beam. For a symmetric beam which is the same for all bolometers, the observed sky is a single simple convolution with the beam, i.e. it is a linear process which we can consider to be the first step in the signal processing chain. So we can treat the 'beam-convolved sky' as the signal which we are trying to recover.

For high enough signal-to-noise it may be possible to attempt to use techniques for deconvolving with the beam, in order to find higher resolution structures in the map. Moreover, since the beam-shape at $450\,\mu$m is less symmetric, it may be possible to apply some techniques to enhance those images in general. Similarly, it may be possible to remove any residual beam-smearing effects caused by scanning at the fastest rates (helped through having SCAN data obtained with pixels scanned in more than one direction). However, all of this will be left as a research exercise for interested observers. The SCAN reduction for SCUBA-2 will assume that the signal to be estimated is the symmetric beam-convolved sky.

This is a great simplification, since it means that each datum in the timestream of a particular detector can be assigned entirely to whichever sky pixel it is pointing at. This is sometimes called the 'zero footprint' approximation, in analogy with the idea of a 'drizzling footprint' used in algorithms to reconstruct fully sampled images from dithers of undersampled arrays.[7] There is no loss of resolution provided that the pixels are chosen to be significantly smaller than the beamsize (since we are just convolving the image with the pixel box-car function). Experience with SCUBA data (as well as a Nyquist argument) has shown that 3 arcsec pixels are sufficient for $850\,\mu$m images, while one may need 1 arcsec pixels in order not to lose any resolution in $450\,\mu$m maps. However, one cannot go to arbitrarily small pixels for 2 reasons: firstly the storage of the maps becomes expensive; and secondly one ends up with pixels for which the 'hit count' is small or even zero, making the reconstruction of the map and the estimation of the errors less reliable. This second argument is what leads to the 'footprint' concept in 'drizzling', but is less relevant for a large scanning array, since (unless the pixels are very small) almost all pixels will have a large number of hits. The exceptions to this are regions near the edge of the map or regions strongly affected by dead detectors. Clearly these pixels can be flagged as less reliable in the final map.

2.3 SCUBA-2 SCAN mode data

For SCUBA-2 we expect the data to be taken in one of 4 basic variants of SCAN mode.[26] It is expected that the analysis procedure described here will require most of the criteria listed for a good SCAN mode, described in SC2/ANA/S210/005, including sufficient cross-linking and avoidance of periodic repeats in the pattern. The telescope will be scanning across the sky in a well defined manner, with the pointing information well understood (at least for the bulk of the data). Each set of $200\,$Hz data from a detector will be referred to here as a 'timestream', and for SCUBA-2 (with all of the sub-arrays at both wavelengths) there are $\sim 10^4$ of these timestreams. Although the same data reduction procedure will be used for the two wavelengths, the $850\,\mu$m Pipeline will be separate from the $450\,\mu$m Pipeline.

Instead of considering the data as a set of timestreams for each detector We could equivalently think of the data as consisting of 'frames' of pixels coming in at a rate of $200\,$Hz, which is the conventional way of treating optical or near-IR data, for example. For CCD-type data (or for images coming from the DREAM reduction to some extent), there is a one-to-one correspondence between detector elements and pixels in the image. *However*, this is not the case for SCAN mode. Here we are trying to estimate the value of astronomical source emission in a set of 'sky pixels', which are on a regular grid in RA,Dec (or some other) coordinates. So there is no relationship between the grid of detector positions and the grid of map pixels. At any time the two are related by a 'pointing matrix', which rolls out the timestream onto the sky and for 'total power' bolometers consists of a single entry in each row and column.

2.4 Major effects in the data

The details of methods used to construct maps from astronomical data depend on the wavelength regime and specific detector technology. In particular it is important to consider the major sources of noise, which will lead to different data reduction approaches, as discussed in SC2/ANA/S210/001.[20] For SCUBA-2 we expect the raw timestreams to contain several effects which must be dealt with in order to make a map. Not all such effects have been fully investigated for SCUBA-2, and we must keep an open mind to the possibility that there may be surprises when we have real data to examine. But nevertheless, based on what we know from SCUBA and from other mm or sub-mm telescopes, we can make a reasonably well-informed guess at the sizes of each of these effects. In roughly decreasing order of amplitude they are:

1. Detector offsets, which are fairly stable over time.
2. Atmospheric emission fluctuations, which are largely common to all the detectors at once.
3. Atmospheric transmission, which will be very close to the same for the entire array.
4. Random (thermal) noise fluctuations from the instrument and the sky, which should integrate down as $t^{-1/2}$.
5. Gain and offset fluctuations of the individual bolometers.
6. Possible correlations between detectors.
7. Emission from structures in the heavens, which is what most astronomers are after.

As well as these effects it will be necessary to correct for cosmic rays and other glitches, as well as dead detectors, or perhaps whole rows of detectors. In addition there may be microphonic effects or other sources of synchronous noise, and other systematic effects, for example depending on the acceleration of the telescope.

2.5 Lessons from other instruments

SCUBA-2 has been designed to be the best sub-mm survey instrument in the world. Nevertheless, it is not setting out into entirely uncharted territory. We already know a great deal about the atmosphere at these wavelengths, and about how to deal with detector and other sources of $1/f$ noise when making maps.

A large amount of literature has built up in the last decade or so on how to make maps from ground-based, balloon-based and satellite-based Cosmic Microwave Background experiments.[30,24,3,23] Some of these ideas are certainly useful for our purposes here. However, CMB experiments do not typically have to contend with atmospheric emission (since they are usually at either high enough altitude or low enough frequency). And another major difference with SCUBA-2 is that, even if they have multiple detectors, the usual CMB approach is to treat each one separately and make individual detector maps to be combined at a later stage of the analysis. For these reasons one cannot simply take one of the available CMB analysis codes and modify it for SCUBA-2. However, one can certainly take many of the features of this general approach to map-making.

The CMB approach is usually to derive the formal linear algebra solution for the map estimator in the presence of correlated noise. However, this means inverting huge matrices (at least $N_{\rm pix} \times N_{\rm pix}$). These methods have generally not been applied to multiple detectors (except effectively to co-add the individual detectors' maps) and so certainly a different approach should be used for huge detector arrays, such as SCUBA-2. One could imagine taking combinations of bolometer signals which are orthogonal, i.e. 'rotating' the full data matrix (of bolometer index versus time index) and focussing on the sub-space of orthogonal modes, or equivalently carrying out a Principal Component Analysis (PCA). However, in practice this is very computationally expensive and tends to result in complicated pixel-pixel correlations in the final map.

We can obviously also learn a great deal from the SCUBA experience. However, SCAN mode with SCUBA was rarely used; in practice a SURF reduction of JIGGLE data produced decent maps a large fraction of the time, whereas most users found SCAN reductions to be awkward to deal with. The map-making method which was mainly used for SCUBA involved 'Emerson-II' type reduction of data taken in multiple chops with a single scan direction.[6,5,9] This method provides acceptable results for some applications, but is far from optimal. It is based on patching Fourier-space methods (adapted from radio interferometry) onto real space differencing, and hence cannot be applied to total power measurements at all. So, although there are many important lessons from SCUBA SCAN-mode which we can bring to SCUBA-2, the basic reduction approach cannot be reused. It is also worth noting that other ideas, involving iterative and CMB-type map-making approaches, were tried out on SCUBA data with at least partial success.[2,11,27]

Most of the issues that we will have to face for SCUBA-2 have already been confronted by other bolometer array instrument teams, particularly total power instruments such as MAMBO, BOLOCAM, BLAST and SHARC-II. MAMBO[17] has 37 and 117 detector arrays operating at IRAM, with scanning and chopping in azimuth to create 'On-The-Fly' maps and an iterative procedure used to remove the atmosphere. BOLOCAM has an array of 144 bolometers operating at $1.1\,$mm on the CSO, using a variety of scan strategies and an IDL-based PCA reduction scheme.[18] BLAST (the Balloon-borne Large Aperture Sub-mm Telescope)[4] has much in common with SCUBA-2, since it works in total power mode, and although it flies above much of the atmosphere, its detectors go down to $250\,\mu$m where the atmospheric emission is still an issue at balloon altitudes. There are several members of the BLAST team who are also working on SCUBA-2, and so there will be direct transfer of know-how between the two efforts. However, at this point it is the SHARC-II experience which has been of particularly use to the SCUBA-2 DR effort. This is partly because of direct contacts between people working on SCUBA-2 and people working on SHARC-II, but perhaps more importantly because they are a facility instrument, and hence have an open policy towards sharing information. This is of tremendous benefit to the SCUBA-2 Project, since we can directly adapt what the SHARC-II Team have already spent several years developing.

3 Atmospheric emission

The most important issue to deal with for ground-based sub-mm experiments is variable atmospheric emission. So before we go any further we should outline the expected situation here.

The atmosphere represents a signal of thousands of Janskys, with variations at the level of hundreds of Janskys. This is *many* orders of magnitude larger than most astronomical signals, and hence careful removal of the atmosphere is crucial if one is to be able to reach faint detection levels.

Fortunately we know a great deal about the atmosphere at these wavelengths. There have been studies at several telescopes around the world, including JCMT,[1] CSO, IRAM and a vigourous campaign of testing and simulations for ALMA.[19,8]

Using total power data obtained by SHARC-II at the CSO, we investigated the nature of the atmospheric emission on Mauna Kea at $350\,\mu$m. The first result[28] we found is that the power spectrum has a $1/f$ (actually a little stronger even) plus white noise form, with a knee frequency at around a few Hz. It is therefore clear that if one can 'chop' faster than this, one has a good chance of removing the noise. The second result[28] is that the atmospheric emission for SHARC-II is very nearly common-mode (i.e. the same for all the bolometers in the array). Hence removing an appropriately computed average of the array at every time sample can act as a way of 'chopping out' the atmospheric emission. The SHARC-II array is about 3 arcmin across, and takes data every 36 ms, while for comparison, SCUBA-2 will be about 8 arcmin across and will take data every 5 ms. Hence if 'chopping' at 36 ms is adequate for SHARC-II, then even although SCUBA-2 has a larger array, one would expect that the proportionally greater increase in sampling rate will make this work well for SCUBA-2 also.

There is anecdotal evidence for atmospheric emission gradients (or even features moving across the array) in data from JCMT, CSO and IRAM. However, we are aware of no statistical study which shows that a significant effect remains after one subtracts a simple DC term from the array. Nevertheless, we studied some SHARC-II data to search for such an effect, and could find none.[28] This is perhaps not surprising given the expected angular speed of many thousands of arcseconds per second at which low level atmospheric structure is expected to blow through the telescope beam.

The current version of the SCUBA-2 simulator[14,15] uses a simple wind-blown Kolmogorov screen model for the atmospheric emission.[12] While this is probably an oversimplistic picture of the atmosphere, and more sophisticated approaches are feasible,[8,22] nevertheless it appears that there is no evidence (based on the SHARC-II data) that it is inadequate for the specific purpose in hand, i.e. the simulator appears to make atmospheric variations which are no more nor less complex than what is seen in real data. After removing the average of the array there is no evidence of lingering atmosphere, for example a slope across the size of the array. The only effect seen in the SHARC-II data was a correlation in the final maps on a scale of $\simeq 35$ arcsec. A brief analysis of some SCUBA $450\,\mu$m data did not find the same effect, strongly suggesting that it is a problem related to the SHARC-II instrument, rather than the atmosphere.[25]

So the current belief about the atmosphere is that removing the array average at 200 Hz will leave essentially no residual variation. Even if this proves not to be completely borne out with real SCUBA-2 data, it should be feasible to fit for a plane or low-order polynomial across the array, *or* to use the idea of using a 'local' average of the sky around each bolometer. This second approach was used for 'fast-scanning' at IRAM and SEST.[29] However, it suffers from the disadvantage of effectively acting as a spatial filter on the data, and this would have to be investigated further before this method was adopted for SCUBA-2.

Nevertheless, at the moment there is no evidence that simply removing the array average *doesn't* work. The only tricky issue is to be careful about how to estimate the average in the presence of bolometers with different noise properties, as well as individual detectors which are suffering from cosmic ray glitches or are otherwise misbehaving.

4 The CRUSH approach

The reduction software used for SHARC-II data is 'CRUSH', the Comprehensive Reduction Utility for SHARC-2, developed mainly by Attila Kovacs at Caltech.[16]

The basic approach is to carry out least squares fitting (see also SC2/SOF/S200/015[13]) for a series of 'models' one at a time, with the final model being the best estimated map of the astronomical sky. The models are fit in order of amplitude, with the biggest (or brightest) one first. Iteration among sub-sets of the models allows for complicated solutions to be found without the need to invert huge matrices. There is no *proof* that such an approach will work, and indeed it would be expected to fail when models are completely degenerate (e.g. variation of gains and baseline with simple 'On-The-Fly' mapping). However, if the SCAN mode data-taking strategy is sufficiently complex and the noise effects are sufficiently separated in timescale, then one might expect it to have a good chance of converging to a sensible solution.

There are several advantages to this overall approach: it is intuitive, in the sense that one can understand the meaning of each of the models used; it can deal with non-linearities, e.g. gain variations; it can be adapted to deal with new effects as they arise; it can be optimised to treat different science requirements, e.g. more iterations for deeper maps or more care for very extended structures; it scales linearly with the number of pixels in the map; and, most importantly, it has already been demonstrated to converge for real maps of interesting size on modest workstations.

In practice one writes down an expression for the timestream of each bolometer as a sum of all the individual models, then solves for each model in turn. This requires finding the maximum likelihood (minimum χ^2 solution) for the model amplitude using the residuals to the model along with appropriate weights. More mathematical details can be found in the current version of the CRUSH manual.[16]

5 The plan for SCUBA-2 SCAN mode reduction

For reduction of SCUBA-2 SCAN mode data, we will follow the basic approach which has been developed through a great deal of trial and error for SHARC-II at the CSO, namely the CRUSH approach. Although code is publicly available for CRUSH, we will develop our own software, tailored for SCUBA-2 – we will be inspired by the approach used for SHARC-II reduction rather than directly adapting CRUSH. We will also be using ideas taken from the reduction approaches used for SCUBA, BLAST and other relevant instruments.

The set of bolometer timestreams will be fit to find the parameters of a series of models using direct maximum likelihood estimation (which has already been described in the SCUBA-2 context[13]). These models will be of quite general form for use in the Off-line Data Reduction software, but the simplest and most robust forms (consistent with achieving good results) will need to be implemented for the Data Reduction Pipeline. Some of these models may be stable over sufficiently long periods of time that they can be fitted nightly (or weekly, or whatever is required) and the relevant coefficients stored in a file, and hence it may be that some of the models can be removed without actually being fit to the data. It is also expected that ancillary data (e.g. the Water Vapour Radiometer) will be useful in this regard. However, the reduction facility also needs to be able to cope with a wide variety of effects which can only be fitted out using the data themselves.

Iteration is important because models can be partially degenerate, i.e. variation in one model can play off against another. So when a ‘bright’ model has been fitted it may be necessary to go back and refit it after some fainter models have been solved for. Precisely which steps need to be iterated against each other and the number of iterations required is something which can only be determined using real data at the telescope. The crucial thing is that the Data Reduction facility is set up to allow for general levels of iteration.

A central part of using the maximum likelihood estimator is to carefully estimate the weights to use (in order to have the minimum variance estimator). These are basically the $1/\sigma^2$ weights from the bolometer timestreams, corrected for the time interval, and also in principle for the number of of degrees of freedom in the fit.

A schematic outline of the Data Reduction procedure for SCUBA-2 SCAN mode might then be:

- Procedure for Model 1
  - Flag ‘bad’ data
  - Calculate weights
  - Estimate model parameters
  - Characterize ‘goodness of fit’
  - Remove best-fit model
- Procedure for Model 2
  - Flag ‘bad’ data
  - Calculate weights
  - Estimate model parameters
  - Characterize ‘goodness of fit’
  - Remove best-fit model
  - Iterate with Model 1 fit if necessary
- Procedure for Model 3
  - Flag ‘bad’ data
  - Calculate weights
  - Estimate model parameters
  - Characterize ‘goodness of fit’
  - Remove best-fit model
  - Iterate with Models 1 and 2 fits if necessary
- . . .

At each step one needs to determine some measure of the ‘goodness of fit’, which will indicate whether further models need to be incorporated. Simulations will also need to be done to determine whether there are any biases introduced by any of the fitting steps.

During the model fitting one needs to carefully assess which parts of the data are behaving pathologically. Bolometers with anomalous behaviour in a given time interval should be flagged and ignored in the fitting process. Investigation of this procedure will have to be tuned using real data, and may require some effort, since it was found to be a limiting step in the SHARC-II experience.

The details of the models and iterative loops required for SCUBA-2 SCAN mode reduction will evolve as we learn more about the behaviour of the arrays and the instrument on the telescope. However, the set of models appropriate for SCUBA-2 will be similar to those outlined in Section 2.4. We now comment on each of the main modelling steps in turn.

5.1 Residual pixel offsets

These are not expected to vary very quickly and so they may stored be in a regularly updated look-up table. Their values may also come directly from the Data Acquisition system, which, in SCAN mode, will report some simple statistics along with the timestreams. Hence, for SCUBA-2 this may not be an issue at all, but we do need to be ready to deal with unexpectedly large offsets which might vary on shorter timescales than would have been hoped.

5.2 Atmospheric emission

At the moment it appears to be a good approximation to assume that the atmosphere is a common mode signal for all the bolometers.[28] However, as we stressed in Section 3 this signal needs to be removed very precisely if SCUBA-2 is going to integrate down into the sub-mJy regime. Hence, one needs to be open to the possibility that the atmosphere is not entirely common mode, since it is still not clear if the atmospheric emission is a function that depends on time only, and not on space. So the atmospheric model may include the possibility of a gradient, or 'local' atmospheric emission effects (using the nearest few hundred bolometers rather than the whole array, for example).

5.3 Atmospheric extinction

In principle one could use the data themselves to estimate the line of sight opacity. There may also be some partial correlation with the atmospheric emission which could be used here. However, trying to estimate the extinction using the data is tricky, because it is quite degenerate with a combination of atmospheric emission and gain variations. Fortunately SCUBA-2 will have the Water Vapour Radiometer measuring the opacity rapidly and accurately, and so this can be used directly to extinction-correct the SCUBA-2 data. Nevertheless, the reduction package should allow for the possibility of self-correcting the extinction in SCAN mode, in case this proves necessary.

5.4 Gain variations

Gains of individual detectors will certainly vary with time, particularly because background-limited detectors have a gain which depends on the loading. It is hoped that the timescale for these gain drifts will be much longer than that for atmospheric variations. Hence a set of piecewise gains can be fit over some time interval (the length of which remains to be determined). Provided there is a reasonable separation between the timescales for gain and atmospheric variation, then it should be easy to fit for gain variations in SCAN data. However, these gains will become more degenerate with atmospheric offsets as the timescales become closer. The best situation, of course, would be gains that vary so slowly that they can be in a table and measured once per night with a 'flat-field'. However, we must be ready to confront the possibility of having to estimate the gains from the data themselves.

5.5 Individual detector drifts

$1/f$-type drifting of individual detectors is much less of a concern than the atmospheric variations (probably by a couple of orders of magnitude), but still potentially much larger than astronomical signals. If correlated structure remains in the detector timestreams even when the above models have been removed, then a method will need to be adopted to fit it out. If this behaves just like correlated noise, then one can adopt the CMB-inspired approach for estimating the noise power spectrum and the map simultaneously.[23] A similar method has been tried on SCUBA data, where it was assumed that the noise power spectra for has the same overall functional form for all the bolometers, varying only in normalization (i.e. different rms for each detector).[27]

5.6 Drifting along rows

Correlations along rows of bolometers are a significant effect for SHARC-II, due to fact that they are read out together in the electronics. Since SCUBA-2 uses TES detectors rather than traditional bolometers, and multiplexed read-outs, we do not expect to see the same effects. Nevertheless, we need to be ready to model such correlations if they arise in SCUBA-2 data. Correlations along rows may be the most obvious thing to look for, but this is just one example of a set of possible systematics involving correlations in the data. Whether such effects need to be modelled before or after individual detector variations will need to be determined when the output of the arrays is characterized.

5.7 Source model

This is a constant flux value in each pixel of the final map, and is what we are ultimately interested in measuring. Some expectation of what will be in this image may affect details of the reduction procedure (e.g. if there is faint extended structure then one would work harder on gain modelling for example, since they might be partially degenerate). As was described in Section 2.2, the map which is produced by the Data Reduction Software will *not* have been deconvolved with the telescope beam.

The final map could of course be convolved *again* with the beam if the observer is interested in finding point sources, since a weighted beam-convolution is equivalent to a maximum likelihood fit to the telescope PSF. Other statistics should be easy to derive directly from the final maps.

The signal map will be accompanied by error maps (two types of 'temporal' rms), maps of pixel 'hit count' and perhaps goodness of fit estimates for each pixel. The observer can use these quantities to set thresholds for clipping the edges of the map, etc.

5.8 Final residuals

After fitting all of the models, there ought to be only white noise left in the timestreams. Tests should be performed to check that the remaining residual timestreams are indeed consistent with white noise. If so the fitting was successful and we are done!

6 Future work

A great deal of investigative research needs to be carried out in order to tune the reduction procedure sketched out here. Precisely how many SCAN mode Recipes there are for reducing the data, and what the individual Primitives might be remains to be determined.[10]

The SCUBA-2 simulator will need to be continuously improved to keep up with the sophistication required to test the SCAN reduction procedure (realizing that at the time this document is written we have no firm knowledge of any systematic effects which would not allow a simple co-add of the data once the common-mode atmospheric emission has been subtracted).

Correlations need to be searched for at every step of the reduction process to ensure that there are no unexpected systematics. This should include 'regional correlations' of the sort found lingering in SHARC-II data.[28] Correlations with 'housekeeping' data should also be searched for, e.g. temperature, telescope acceleration, etc. Goodness of fit estimators will also be useful in this regard. Not all such diagnostics need to be available in the working summit Pipeline of course.

Simulations will need to be done to determine the order in which the models should be fitted, and this will be further honed when we have real data. Simulations are also important to ensure that the results are not biased (i.e. you get out what you put in).

Another issue to be determined is what to use for stopping criteria for the iterative loops, and in general the ability to determine whether the results are converging to anything useful or are going out of expected range.

Flagging anomalous parts of the data residuals is a necessary task to perform during the model fitting steps. One needs to check for any residual spikes in real space *and* in Fourier space (i.e. periodic of time-synchronous signals). Because Fourier Transforms will be carried out it is crucial to be able to take large contiguous chunks of data. Gaps may thus need to be filled, perhaps with 'constrained realisations', as is done routinely in CMB timestream analysis.[23]

One can also imagine investigating variants on the models, such as whether it makes sense to fit each of the sub-arrays separately, and whether there are options to set for mosaicing (e.g. can some model parameters be carried over between sub-maps, so that the whole of the raw data does not need to be loaded into memory at once?). Another possibility is using information from one wavelength to help fit the data at the other wavelength, or running the reduction in parallel with connection points (since there will be some models which are related for the $450\,\mu$m and $850\,\mu$m data). One more idea to consider would be whether pointing corrections could be included as part of the modelling.

It is important to realise that CRUSH has evolved over more than 2 years of tuning with real data at the CSO. The SCUBA-2 SCAN-mode Data Reduction Software needs to learn from this by being ready with a versatile (but robust) reduction facility. CRUSH has became an extensive package with at least 120 different options; SCAN reduction for SCUBA-2 needs to be simple and robust for the Pipeline, while allowing for greater flexibility in the the Off-line mode.

The SCAN mode Pipeline needs to be tuned so that it is efficient and robust, and should allow only a finite set of 'Recipes'. There may thus be different options depending on whether bright or faint sources are expected, point sources versus very extended structures, etc. The Off-line version should allow for much more versatility. Users should have the ability to tune things, increase the number of iterations, change the convergence criteria, the clipping thresholds, the pixel size, and other parameters. The Pipeline version, on the other hand, will need a set of default values (although these will probably vary for each of the 4 variants on SCAN mode, as well as for the basic science goals), affecting how aggressively some of the steps are performed or whether some of them are missed entirely.

7 Conclusions

SCUBA-2 data will have 2 major advantages over existing mm and sub-mm detectors: the data will be taken much more rapidly than the atmosphere varies; and there will be a huge number of detectors in the arrays. Reduction of similar 'total power' bolometer array data has already been carried out successfully at other telescopes, and we have the particularly useful experience of the SHARC-II team to draw upon. Scaling from SHARC-II analysis suggests that a similar SCAN mode data reduction approach should work even better for SCUBA-2.

At this point it appears that simply removing the array average should result in essentially no remaining effects from sky emission. Right now we also have very little information about the behaviour of the real detectors, and so no details of how to model potential correlations and other systematics. Nevertheless we must be ready for the situation to be much more complicated in practice. Hence we must have a reduction Pipeline in place which allows us to fit out all the additional effects that we can reasonably imagine. With a linear fitting approach, which models effects in decreasing order of importance, and allows for iterative solution, there can be some confidence that we will be ready to deal with real SCUBA-2 data.

References

[1] E. N. Archibald, T. Jenness, W. S. Holland, I. M. Coulson, N. E. Jessop, J. A. Stevens, E. I. Robson, R. P. J. Tilanus, W. D. Duncan, and J. F. Lightfoot. On the atmospheric limitations of ground-based submillimetre astronomy using array receivers. *Monthly Notices of the Royal Astronomical Society*, 336:1–13, October 2002.

[2] C. Borys, S. C. Chapman, M. Halpern, and D. Scott. A SCUBA scan-map of the Hubble Deep Field : measuring the bright end of the submillimetre source counts. *Monthly Notices of the Royal Astronomical Society*, 330:L63–L67, March 2002.

[3] J. Delabrouille. Issues and methods for CMB anisotropy data reduction. *New Astronomy Review*, 45:313–320, March 2001.

[4] M. J. Devlin, P. A. R. Ade, I. Aretxaga, J. J. Bock, J. Chung, E. Chapin, S. R. Dicker, M. Griffin, J. Gundersen, M. Halpern, P. Hargrave, D. Hughes, J. Klein, G. Marsden, P. Martin, P. D. Mauskopf, B. Netterfield, L. Olmi, E. Pascale, M. Rex, D. Scott,

C. Semisch, M. Truch, C. Tucker, G. Tucker, A. D. Turner, and D. Weibe. The Balloon-borne Large Aperture Submillimeter Telescope (BLAST). In *Astronomical Structures and Mechanisms Technology. Edited by Antebi, Joseph; Lemke, Dietrich. Proceedings of the SPIE, Volume 5498, pp. 42-54 (2004).*, pages 42–54, October 2004.

[5] D. T. Emerson. Approaches to Multi-beam Data Analysis. In *ASP Conf. Ser. 75: Multi-Feed Systems for Radio Telescopes*, pages 309–+, 1995.

[6] D. T. Emerson, U. Klein, and C. G. T. Haslam. A multiple beam technique for overcoming atmospheric limitations to single-dish observations of extended radio sources. *Astronomy & Astrophysics*, 76:92–105, June 1979.

[7] A. S. Fruchter and R. N. Hook. Drizzle: A Method for the Linear Reconstruction of Undersampled Images. *Publications of the Astronomical Society of the Pacific*, 114:144–152, February 2002.

[8] M.A. Holdaway. Effects of Atmospheric Emission Fluctuations and Gain Fluctuations on Continuum Total Power Observations with ALMA. ALMA Memo 490, 2004.

[9] T. Jenness, W. S. Holland, E. Chapin, J. F. Lightfoot, and W. D. Duncan. Dual-Beam Rastering and Deconvolution Techniques for SCUBA. In *ASP Conf. Ser. 216: Astronomical Data Analysis Software and Systems IX*, pages 559–+, 2000.

[10] Tim Jenness. SCUBA-2 Data Reduction Recipes and Primitives. SCUBA-2 Project SC2/SOF/S210/004, 2003.

[11] D. Johnstone, C. D. Wilson, G. Moriarty-Schieven, J. Giannakopoulou-Creighton, and E. Gregersen. Large-Area Mapping at 850 Microns. I. Optimum Image Reconstruction from Chop Measurements. *Astrophysical Journal Supplement*, 131:505–518, December 2000.

[12] B. D. Kelly. Modelling the sub-mm atmospheric emission. SCUBA-2 Project SC2/ANA/S100/013, 2001.

[13] B. D. Kelly. Image Processing and Least-Squares Reconstructions. SCUBA-2 Project SC2/SOF/S200/015, 2003.

[14] B. D. Kelly. SCUBA-2 data simulation. SCUBA-2 Project SC2/SOF/S200/016, 2003.

[15] B. D. Kelly. SCUBA-2 Data Simulation Progress. SCUBA-2 Project SC2/SOF/S200/024, 2004.

[16] Attila Kovács. Crush: Comprehensive reduction utility for sharc-ii. URL: http://www.submm.caltech.edu/∼sharc/crush/, 2004.

[17] E. Kreysa, H. Gemuend, J. Gromke, C. G. Haslam, L. Reichertz, E. E. Haller, J. W. Beeman, V. Hansen, A. Sievers, and R. Zylka. Bolometer array development at the Max-Planck-Institut fuer Radioastronomie. In *Proc. SPIE Vol. 3357, p. 319-325, Advanced Technology MMW, Radio, and Terahertz Telescopes, Thomas G. Phillips; Ed.*, pages 319–325, July 1998.

[18] G. T. Laurent, J. E. Aguirre, J. Glenn, P. A. R. Ade, J. J. Bock, S. F. Edgington, A. Goldin, S. R. Golwala, D. Haig, A. E. Lange, P. R. Maloney, P. D. Mauskopf, H. Nguyen, P. Rossinot, J. Sayers, and P. Stover. The Bolocam Lockman Hole Millimeter-Wave Galaxy Survey: Galaxy Candidates and Number Counts. *Astrophysical Journal*, 623:742–762, April 2005.

[19] S. Radford and L.-A. Nyman. Site Characterization and Monitoring. ALMA Project Book, Chapter 14, URL: http://www.alma.nrao.edu/projectbk/construction/chap14/chap14.html, 2001.

[20] Douglas Scott. Map making in different noise regimes. SCUBA-2 Project SC2/ANA/S210/001, 2003.

[21] Douglas Scott. SCUBA-2 Data Reduction Software Requirements. SCUBA-2 Project SC2/SRE/S210/001, 2004.

[22] A. Stirling, J. Richer, R. Hills, and A. Lock. Turbulence Simulations of Dry and Wet Phase Fluctuations at Chajnantor. Part I: The Daytime Convective Boundary Layer. ALMA Memo 517, 2005.

[23] R. Stompor, A. Balbi, J. D. Borrill, P. G. Ferreira, S. Hanany, A. H. Jaffe, A. T. Lee, S. Oh, B. Rabii, P. L. Richards, G. F. Smoot, C. D. Winant, and J. P. Wu. Making maps of the cosmic microwave background: The MAXIMA example. *Physical Review D*, 65:22003–+, January 2002.

[24] M. Tegmark. CMB mapping experiments: A designer's guide. *Physical Review D*, 56:4514–4529, October 1997.

[25] Alex van Engelen and Douglas Scott. A Correlation Study using SCUBA Data, with implications for SCUBA-2 data reduction. SCUBA-2 Project SC2/ANA/S210/004, 2005.

[26] Alex van Engelen and Douglas Scott. SCAN Mode Strategies for SCUBA-2. SCUBA-2 Project SC2/ANA/S210/005, 2005.

[27] Alexander van Engelen. Undergraduate Honours Thesis: New map-making ideas for submillimetre experiments. Simon Fraser University, 2001.

[28] Alexander van Engelen. Analysis of Atmospheric Emission using Simulations and SHARC-II Data. SCUBA-2 Project SC2/ANA/S210/002, 2004.

[29] B. Weferling, L. A. Reichertz, J. Schmid-Burgk, and E. Kreysa. Principles of the data reduction and first results of the fastscanning method for (sub)millimeter astronomy. *Astronomy & Astrophysics*, 383:1088–1099, March 2002.

[30] E.L. Wright. Scanning and Mapping Strategies for CMB Experiments. astro-ph/9612006, 1996.