SHARC-2 350 μm Observations of Distant Submillimeter-Selected Galaxies and Techniques for the Optimal Analysis and Observing of Weak Signals

Thesis by

Attila Kovács

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

2006
(Defended May 19, 2006)

This one’s for the doubters...

Acknowledgements

I owe most to Mom and Dad for always letting me follow my dreams, and Tom Phillips, my adviser, for likewise allowing me to “do my thing” on this never-ending thesis research. I ought to mention, that it all started on my 7th birthday when Grandma presented me with my first astronomy book. She used to tell me that after she was gone, she would watch me from the stars. I am now looking back at her from down here! Thanks Grannie, you are a star! And many thanks to all who gave me love, caring, friendship, support and advice all these years (and may you find forgiveness if I left you out below...). You are my living (rock) stars! Especially my brother, the rest of my family, Connie, and...

Haroon & Martha Ahmed · Zahra Ahmed · Dominic Benford · Peter Bibring · Andrew Blain · Adwin Boogert · Colin Borys · Előd Both · Matt Bradford · Maša Čemažar · Scott Chapman & Janya Freer · Ed Chen · Søren “Ice” Christensen · Gergő Dávid · Darren Dowell · Moses Dili Duku · Mick Edgar · György Flórik · Ilja Friedel · Rosaria Fronteddu · Marisol Gerona · Carl Gottlieb · Thomas Greve · Yiannis Hadjiyiannis · Péter Horváth · Ákos Issekutz · Dean “Marvellas” Karvellas · Coley King & Yana Queen · Zsuzsi Kiss · Jacob Kooi · Edo Kussell · Melanie Leong · Darek Lis · Carole Lu & Alok Saldanha · Sara Magnusson · Mike McCarthy · Denis Medvedsek · Karl Menten · Stanimir Metchev & Anne Simon · Simona Miglietta · Jim Moran · Harvey Moseley · Frederique Motte · Neo Muyanga · Pia Oberoi · Bartek Ostdenda · Júlia Paraizs & the Paraizs family · Péter Péli · Paul Pénzes & Rocio Roman · the Pop-Eleches brothers, “Kiki” & Grigore · Anu Raunio · Frank Rice · Zoltán Rihmer · Morvan Salez · Ines Schwetz & the Schwetz family · Zsuzsa Serfőző · Rick Shafer · Gábor Siklós · Bastien & Corinne Solier · Johannes Staguhn · Mark Sylvester · Ramy Tadros · Patrick Thaddeus · Mike Travers · Dominik Treeck · Charlotte Vastel · Beate Wegscheider · Whisper · Bartek Wydrowski · Min Yang · Hiroshige Yoshida · Jonas Zmuidzinas · Judit Zombori · the dedicated proofreader, who had to suffer my many errors · and the many beers, wines, martinis, & other magical concoctions that kept me sane (?)...

Hungarian Goulash

This basic food of the Hungarian cattle-herder (i.e., “gulyás” in Hungarian) begins by saying that its proper name is “pörkölt” (meaning, braised) in Hungarian. Mix in some sour cream before serving, and you may call it “paprikás” instead. This, highlights an important point: goulash is a quintessential paprika dish—so better have the best grinds of that powder when preparing it. Good paprika is bright red (even orangy) in color and is very fragrant. Then, there’s also “gulyás leves” (i.e., goulash soup), which uses the same base for a soup version with many additional vegetables (almost to the point of too healthy for the average heart-clogged Hungarian) and flour dumplings (“csipetke”).

As the nomenclature suggests, it is normally made from beef (i.e., “gulya”), especially from the unique gray-colored variety that is (or used to be) native to the plains of Hungary but one, which is now on the brink of extinction (I must admit, I never had the pleasure of tasting one of those...). However, almost any dark meat (e.g., pork, lamb, mutton, deer, boar – but no chicken!), or not-quite-meat (e.g., rooster fries, tripe) works. The trick is always to cook it long enough (so that it falls apart under a gentle bite), and add substantial amounts of pork lard (or less tasty but more heart-friendly vegetable oil) and red wine. Most importantly, it really should be cooked outdoors in a cauldron hanging over an open fire. Of course, it can be prepared on the stove top also, but don’t expect the divine outcome when you allow yourself such profanity. Anyway, here’s my (slightly nontraditional) recipe:

Ingredients. (for 6-8 servings)

4 lb	beef (marbled), lamb, deer or else, cubed in $\gtrsim$1 in pieces (even mixed). Make it tough, it just becomes better. (Avoid clean cuts by all means!)
4 Tsp.	pork lard (or oil if you must...)
	slices of bacon diced (optional)
2	medium yellow onion, finely chopped.
4-6 tsp.	paprika (make sure to mix in some of the spicy kind!)
1/2 bottle	red wine
2-3	Hungarian or Anaheim peppers (optional) cut in 1/2 in cicles
1-2	ripe tomatoes (optional) cut in wedges
	chillies (red, green, or yellow, fresh or dry)
1/2 lb	mushrooms, preferably wild, diced (optional)
3-4	bay leafs (laurel)
1	sprig of rosemary (optional)
1	cinnammon stick (optional)
	salt to taste

Preparation. Sauté the onions in the lard over medium-low heat (with the bacon for extra goodness), so that it just sizzles, until the onions start to turn golden (about 15-20 minutes). Mix in peppers and chillies and stir in the paprika quickly (it will become bitter if you burn it—and you better start over), and cool it (and yourself!) off with a glass of red wine. With the relief that comes with successful completion of this step, all troubles are now behind! Add the meat, and turn the heat to medium-high. Stir until the meat is browned on all sides (i.e., braised). Add mushrooms, and cook until its juices are evaporated. Add salt, spices, and tomatoes. Simmer for at least 2 hours uncovered. Whenever the liquid is about to disappear under the meat, add more wine. At some point, the fat will come on top in a marbled layer—that is a good sign. It is finally done, when the meat falls apart from a gentle squeeze against the cauldron side.

Serve with boiled potatoes or spätzle (for which gnocchi makes a good substitute), and pickles.

Abstract

New $350\,\mu$m data constrain accurately the thermal far-infrared spectral energy distributions (SEDs) for 12 distant submillimeter selected galaxies (SMGs). The results confirm that the linear radio to far-infrared correlation, established for local infrared galaxies, holds out to high redshifts ($z \sim 1$–3). The low correlation constant $q \simeq 2.14$ is more indicative of star formation than AGN-fueled dust heating. The sample exhibits an apparent luminosity–temperature relation ($L_{FIR} \propto T_d^{2.89}$), possibly owing to selection effects. As a result, photometric redshifts in the radio or far-infrared may not be viable, but expressions may relate the observed quantities for current flux and volume limited SMG samples. These suggest that SED estimation may be possible, for objects similarly selected, based on a single radio or far-infrared flux measurement.

The detection of these faint objects ($\simeq$10 mJy at $350\,\mu$m) from the ground is complicated by a bright ($\simeq 10^3$ Jy) and highly variable ($\simeq$10 Jy RMS in 10 minutes of integration) atmosphere with a $1/f^2$ noise spectrum and by instrumental $1/f$ noise. To reach optimum sensitivities, a careful analysis of the data is required, and well-chosen observing strategies are helpful. The principal techniques that aid the extraction of weak signals from colored noise are presented. Close to optimal analysis is implemented effectively by the *CRUSH* software. Both the computing and storage requirement of the implementation scales linearly with the size of the data set, making this approach superior to the computationally expensive alternatives for handling the very large data volumes expected from future instruments.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 From Dust to Star Formation in the Distant Universe

Today's technological capabilities allow us to study the universe via nearly the full electromagnetic spectrum. The wavebands of astronomical interest are determined either by the physical processes that they probe, or by their reach into the distant universe. To study large volumes of space, it is best to look in a part of the spectrum where most energy is radiated.

Galaxies emit much or most of their light at far-infrared wavelengths ($\lambda \simeq 30\,\mu$m–1 mm), with their spectral energy distributions (SEDs) typically peaking in the neighborhood of $100\,\mu$m (fig. 1.1). A mostly unresolved milieu of distant galaxies thus give rise to the Cosmic Infrared Background (CIB, fig. 1.2). Thus, the wavebands of the submillimeter[1] and far-infrared offer an optimal window to study galaxy formation into the depths of the universe.

The shape of the CIB spectrum is consistent with cold ($\simeq$10–100 K) thermal of gray- or black-body radiation. Only free charges (i.e., plasma) and solids can produce such thermal continuum, since free neutral atoms and molecules absorb and emit light in narrow bands at specific energies only. The ionized plasma of the interstellar or intergalactic medium is much too hot (thousands to millions of kelvin) to explain the observed cool galactic continua. Therefore, solid objects must be responsible for radiating this dominant thermal component. And because it is the surface area, not the volume, of solids which gives rise to bulk the thermal radiation, the light we detect in the submillimeter or far-infrared comes mainly from great numbers of rather small particles, instead of large bodies (e.g., planets and asteroids). These particles, also known as the interstellar dust-grains, have typical sizes ranging between 1 nm and $1\,\mu$m. Thus, it is fair to say, that the study of thermal far-infrared is mainly concerned with the physical properties of dust.

Stars produce dust when they shed their metal-rich outer envelopes during the final stages of their life cycle, either as planetary nebulæ or as supernovæ. Dust also plays an important role in the formation of new stars by efficiently radiating large amounts of energy, thus enabling the

---

[1] The submillimeter band refers to the part of far-infrared spectrum with $\lambda \sim 100\,\mu$m–1 mm.

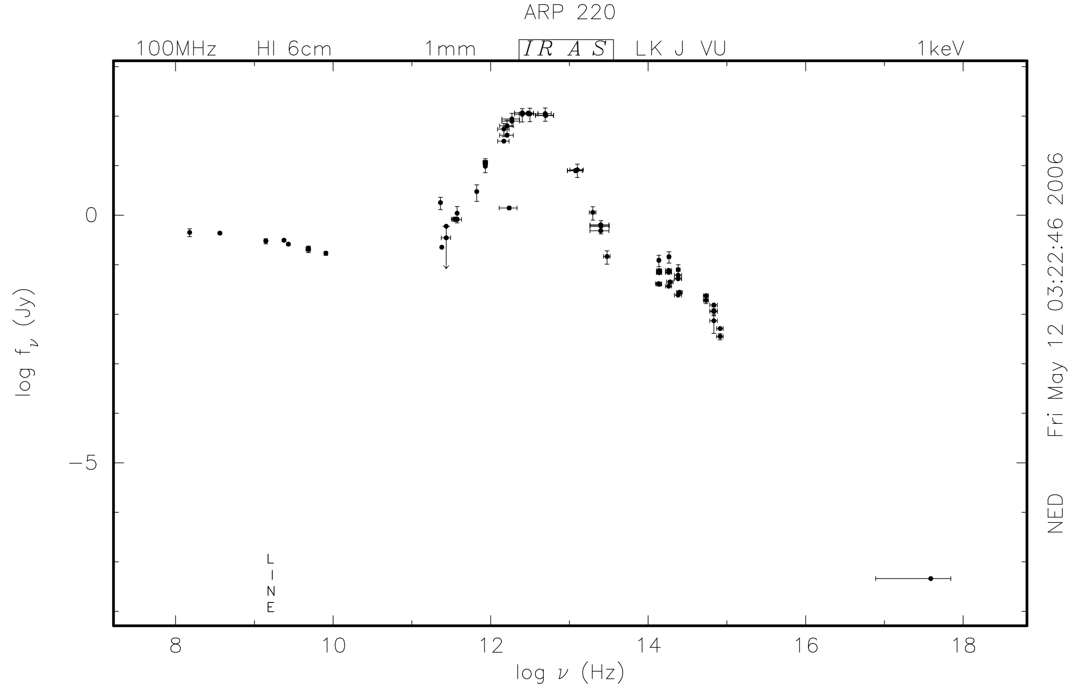


Figure 1.1 The spectral energy distribution of a dusty galaxy. Arp220 is a local universe example of what we think distant dusty galaxies may be like. This local starburst emits most of its energy at far-infrared wavelengths via warm dust (characteristically 37 K). This figure was obtained from the *NASA/IPAC* Extragalactic Database (NED).

gravitational collapse of molecular clouds to form young stellar cores. In this way, the quantities of dust trace both the formation and the death of stars, while dust luminosities and temperatures are fueled by the energetics of the active stellar populations whose light dust absorbs and re-emits in the far-infrared.

Since massive ($\gtrsim 8\,M_\odot$) stars are far more energetic than their smaller-sized counterparts (e.g, the Sun), they tend to dominate the heating of dust in their environments (cf. Salpeter 1955 IMF and L-M relation from Kuiper 1938; Schaller et al. 1992; Baraffe et al. 1998; Henry & McCarthy 1993; Henry et al. 1999). Massive stars are also short lived ($\sim 10^7$ yr), and the dust luminosities they fuel persist only as long as new stars are forming at the same rate as old ones die. As such, far-infrared luminosities provide an indirect measure of the present rate of high-mass star formation. Therefore, far-infrared astronomy is also the study of star formation, and its history, by reaching into the depth of a younger universe.

Chapter 2 presents a star-formation study for the most distant population of submillimeter-selected galaxies (SMGs). New data, acquired at $350\,\mu$m wavelengths from the SHARC-2 bolometer array (Dowell et al. 2003) at the Caltech Submillimeter Observatory (CSO), were critical in constraining accurately the thermal SEDs of these far-away galaxies. The derived dust properties provide valuable insight into the star-formation characteristics of these elusive objects, which often

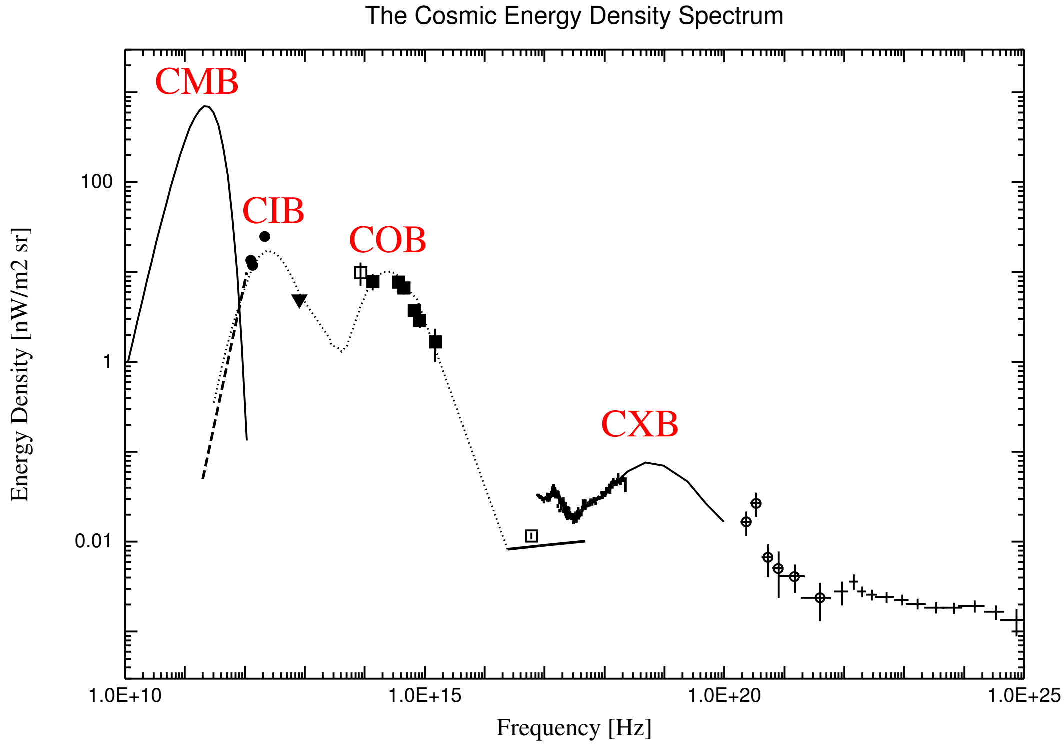


Figure 1.2 The cosmic energy density spectrum. The dominant Cosmic Microwave Background (CMB), with its 2.7 K Planck blackbody spectrum, carries a snapshot of the universe from the recombination era before galaxies were formed. In contrast, a field of distant galaxies contribute to the Cosmic Infrared (CIB) and Cosmic Optical (COB) Backgrounds, indicating that these wavelengths are optimal for the study of galaxy evolution. Galaxies, like Arp220 (fig. 1.1) at redshifts $z \sim 3$, could explain the observed CIB. Figure from Hasinger (2000).

lie hidden at optical wavelengths. Additionally, the new results help confirm an important relationship between radio and far-infrared luminosities, well studied locally, but not at the greater distances of SMGs where its impact could be most profound. The new data could hint at the exciting possibility of providing reasonable SED estimates for flux- and volume-limited radio and far-infrared galaxy samples based on photometry at the single wavelength of detection.

1.2 New Techniques for Observing and Analysis

Detecting faint and distant objects at submillimeter wavelengths is not an easy task. This is especially true, when attempted from ground-based telescope facilities (e.g., the CSO), because our own atmosphere is many orders of magnitude brighter than the signals sought from the galaxies of great distance. The variability of the atmospheric flux alone, over the time scales of a few hours integration, is typically thousands of times greater than the feeble $\simeq 20\,\mathrm{mJy}$ fluxes of the typical SMG at $350\,\mu\mathrm{m}$ wavelength. That these detections are, nevertheless, possible, owes to creative techniques for the collection and analysis of submillimeter data.

The traditional time-tested approach has been to switch (chop) rapidly between a "source" and a nearby patch of "blank" sky. With the same detectors observing both source and sky, the difference of the two signals provide viable source fluxes (Emerson 1988, 1995). When done sufficiently fast, such that the changing of the atmospheric flux can be neglected (at $350\,\mu\mathrm{m}$ this typically requires switching back-and-forth several times a second), sensitive photometry becomes possible. However, the signal differencing involved comes at a price of losing sensitivity to many types of source structures, and that image reconstructions require a fair amount of guessing (via maximum-entropy deconvolution, see Emerson 1995; Jenness et al. 2000a,b). Nonetheless, for instruments with but a single, or a just few, detectors chopping can mean the only viable way of collecting data.

Recent progress in detector technology heralded the arrival of new larger format, 2-dimensional, bolometer arrays with significantly increased detectors counts, and longer time-scale stabilities. The SHARC-2 instrument with its 384 detectors, arranged in a 32×12 format, replaced a 24 pixel predecessor in January 2003, and is currently the largest submillimeter array in operation. Even bigger arrays are already in production and expected in the near future (e.g., SCUBA-2 with ca. 2×5000 pixels, see Audley et al. 2004). These large arrays can measure the sky (and sources) simultaneously at many different positions, at all times. As such, it is reasonable to ask whether the old favorite technique of chopping and differencing could now yield to new, and superior, models of observing and data analysis.

With the SHARC-2 array we have pursued a novel approach of measuring total-power detector signals, while the array is scanned across the targeted area of sky with recurring, cross-linked, 2-dimensional patterns. Data is then analyzed with the *CRUSH* software package, developed by the

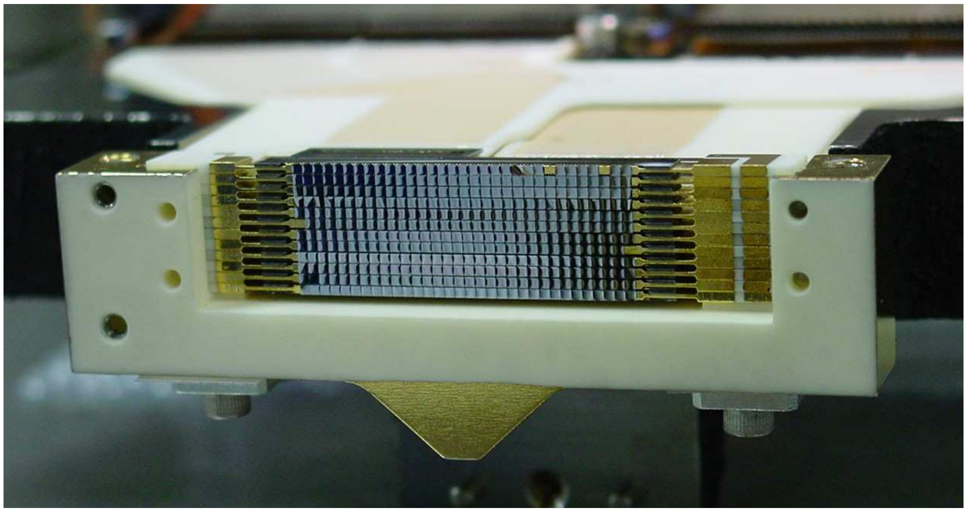


Figure 1.3 The SHARC-2 detector array.

author as a novel approach to data reduction. The faint submillimeter galaxies of Chapter 2 provide ample testimony to the success of this approach, as do a number of other papers[2] whose scientific results benefited from the new techniques.

Chapters 3 and 4 introduce both the observing and data reduction techniques, which go hand in hand for a successdul experiment. They are discussed in general terms, such that the techniques and their conclusions can be adapted to a wide variety of weak-signal experiments with adverse noise conditions.

The first of the two chapters focuses on the general principles of analyzing weak signals in the presence of non-Gaussian noise. It also introduces the concepts and challenges that specific implementations face and must meet. The discussion covers many of the often overlooked details. which prove critical in low signal-to-noise experiments, such as appropriate weighting and filtering of the data, and the necessity of accurately characterizing signal gains. This chapter also offers a quantitative treatment on how data can be observed and analyzed optimally with the least loss of information possible.

Chapter 4 introduces a particular implementation, that of *CRUSH* for the analysis of SHARC-2 data. The method analyzes data in the time domain through an iterated sequence of statistical estimators. A great advantage of the approach is that its computation cost grows linearly with the size of the data, in sharp contrast with other established methods based on the inversion of large matrices or Fourier transforms. The method approaches maximally optimal analysis in the presence of $1/f$ type noise (and especially $1/f^2$ Brownian noise), the greatest malady affecting typical ground-based submillimeter data and CMB experiments. It may even be better than its

---

[2] e.g., Houde et al. (2004); Marsh et al. (2005); Khan et al. (2005,a); Andrews & Williams (2005); Chen et al. (2005); Borys et al. (2006); Beelen et al. (2006); Laurent et al. (2006) and Greve et al. (2006, in preparation).

expensive alternatives, whose actual effects on weak signals can often remain mysteriously buried under their mathematical complexities and subtleties.

Finally, the specifics relating to the astronomical application are presented in Appendix B, while further information is available on the SHARC-2 web-page.[3]

Bibliography

Andrews, S. M., & Williams, J. P. 2005, ApJ, 631, 1134

Audley, M., et al. 2004, preprint (astro-ph/0410439)

Baraffe, I., Chabrier, G., Allard, F., & Hauschildt, P. H. 1998, A&A, 337, 403

Beelen, A., Cox, P., Benford, D. J., Dowell, C. D., Kovács, A., Bertoldi, F., & Omont, A. 2006, ApJ, in press

Borys, C. B., et al. 2006, ApJ, 636, 134

Chen, C. H., et al. 2005, ApJ, 634, 1372

Dowell, C. D., et al. 2003, SPIE Conf. Proc. Vol. 4855, Millimeter and Submillimeter Detectors for Astronomy, ed. T. G. Phillips, & J. Zmuidzinas, 73

Emerson, D. T. 1995, ASPC, 75, 309

Emerson, D. T., & Graeve, R. 1988, A&A, 190, 353

Hasinger, G. 2000, LNP, 548, 423

Henry, T. J., Franz, O. G., Wasserman, L. H., Benedict, G. F., Shelus, P. J., Ianna, P. A., Kirkpatrick, J. D., & McCarthy, D. W 1999, ApJ, 512, 864

Henry, T. J., & McCarthy, D. W. 1993, AJ, 106, 773

Houde, M., Dowell, C. D., Hildebrand, R. H., Dotson, J. L., Valillancourt, J. E., Phillips, T. G., Peng, R., & Bastien, P. 2004, ApJ, 604, 717

Jenness, T., Holland, W. S., Chapin, E., Lightfoot, J. F., & Duncan, W. D. 2000a, ASPC, 216, 559

Jenness, T., Lightfoot, J. F., Holland, W. S., Greaves, J. S, & Economou, F. 2000b, ASPC, 217, 205

Johnstone, D., Wilson, C. D., Moriarty-Schieven, G., Giannakopoulou-Creighton, J., & Gregersen, E. 2000, ApJS, 131, 505

Khan, S. A., Benford, D. J., Clements, D. L., Modeley, S. H., Shafer, R. A., & Sumner, T. J. 2005, MNRAS, 359, 10

Khan, S. A., et al. 2005, ApJ, 631, 9

Kuiper, G. P. 1938, ApJ, 88, 472

Laurent, G. T., Glenn, J., Egami, E., Rieke, G. H., Ivison, R. J., Yun, M. S., Aguirre, J. E., & Maloney, P. R. 2006, ApJ, 643, in press

Marsh, K. A., Velusamy, T., Dowell, C. D., Grogan, K., & Beichman, C. A. 2005, ApJ, 620, 47

Salpeter, E. E. 1955, Apj, 121, 161

Schaller, G., Schaerer, D., Meynet, G., & Maeder, A. 1992, Nature, 360, 320

---

[3] `http://www.submm.caltech.edu/~sharc`

Chapter 2

SHARC-2 350 μm Observations of Distant Submillimeter-Selected Galaxies

The text in this chapter is reproduced in part with permission from:

Kovács, A., Chapman S. C., Dowell, C. D., Blain A. W., Ivison R. J., Smail I, & Phillips, T. G. 2006, ApJ, in press

2.1 Introduction

Ever since the first submillimeter-selected galaxy (SMG) samples debuted from SCUBA $850\,\mu$m surveys (Smail et al. 1997; Barger et al. 1998; Hughes et al. 1998; Eales et al. 1999), the nature of the SMG population has been the focus of attention for the galaxy formation community, because the $850\,\mu$m selection is expected to pick similar sources almost independently of redshift ($z \sim$ 1–8), due to a negative K-correction that essentially compensates for the loss of flux from increasing distance. This allows unbiased, luminosity-selected studies of galaxy formation. The hunger for information on these sources spurred a flurry of follow-up studies at all wavelengths, long and short. Since then many of these sources have been identified at optical, UV (Borys et al. 2003; Chapman et al. 2003; Webb et al. 2003) and radio wavelengths (Smail et al. 2000; Ivison et al. 2002), providing accurate positions which allowed optical redshift measurements (Chapman et al. 2003, 2005). As a result we now know that these massive galaxies, with redshifts distributed around a median z of 2.3, are enshrouded with such quantities of dust that they often lie hidden at optical wavelengths, and therefore constitute a distinct population from the galaxies selected by optical surveys.

More recently, longer wavelength submillimeter surveys, at $1100\,\mu$m and at $1200\,\mu$m (Laurent et al. 2005; Greve et al. 2004), added to the pool of available information. However, the close proximity of the SCUBA, Bolocam and MAMBO wavelengths on the Rayleigh-Jeans side of the SED does not allow for an effective constraint on the thermal far-infrared spectral energy distributions (SEDs) at the relevant redshifts. Nor do the latest results from the *Spitzer Space Telescope* provide powerful constraints, since at the shorter mid-infrared wavelengths the emission is dominated by polyaromatic hydrocarbons (PAHs) and a minority population of hot dust. For these reasons, the best estimates of the characteristic temperatures and the integrated luminosities, to date, have relied on the assumption that the local radio to far-infrared correlation (Helou et al. 1985; Condon 1992; Yun et al. 2001) can be extended into the distant universe. There are hints that this may be appropriate (Garrett 2002; Appleton et al. 2004), but the assumption has remained largely unchecked.

Shorter wavelength submillimeter measurements, sampling near the rest frame peak of the emission, are thus essential to provide firm constraints to the far-infrared SED. Here, we present results at $350\,\mu$m, using the second generation Submillimeter High Angular Resolution Camera (SHARC-2; Dowell et al. 2003) at the Caltech Submillimeter Observatory (CSO). From these we derive the first direct measures of dust temperatures and far-infrared luminosities for a sample of SMGs, testing the radio to far-infared correlation. We also attempt to constrain dust emission properties, and investigate the implications of our findings for the viability of photometric redshifts based on far-infrared and radio measurements. Finally, we present a range of useful scaling relations that may apply to the SMG population.

2.2 SHARC-2 350 μm Observations

We conducted follow-up observations of SCUBA $850\,\mu$m detected sources with radio identifications and optical redshifts (Chapman et al. 2003, 2005). Seven of the of 15 targets were hand picked based on their predicted bright $350\,\mu$m fluxes ($S_{350\,\mu\mathrm{m}} > 45\,$mJy from Chapman et al. 2005), while the remaining were selected at random to partially compensate for any selection bias.

The observations were carried out during 8 separate observing runs between November 2002 and April 2005, in excellent weather ($\tau_{225\,\mathrm{GHz}} < 0.06$), reaching $1\,\sigma$ depths of 5–9 mJy in 2–4 hours of integration in 14 small fields ($\simeq 2.5 \times 1$ arcmin^2) around the targeted sources. Our scanning strategy was to modulate the telescope pointing with a small amplitude (15–20 arcsec) nonconnecting Lissajous pattern within the limits of reasonable telescope acceleration (with typical periods of 10–20 seconds). This pattern was chosen to provide fast, 2-dimensional, recurring but nonclosed, patterns with crossing paths—all of which are essential to allow the separation of the superposed source, atmospheric and instrumental signals. For the observations since February 2003, we have taken advantage of the CSO's Dish Surface Optimization System (Leong 2005) to improve beam shapes and efficiencies, and hence sensitivities, at all elevations.

Pointing corrections were retroactively applied to the observations by a careful fit to the point-

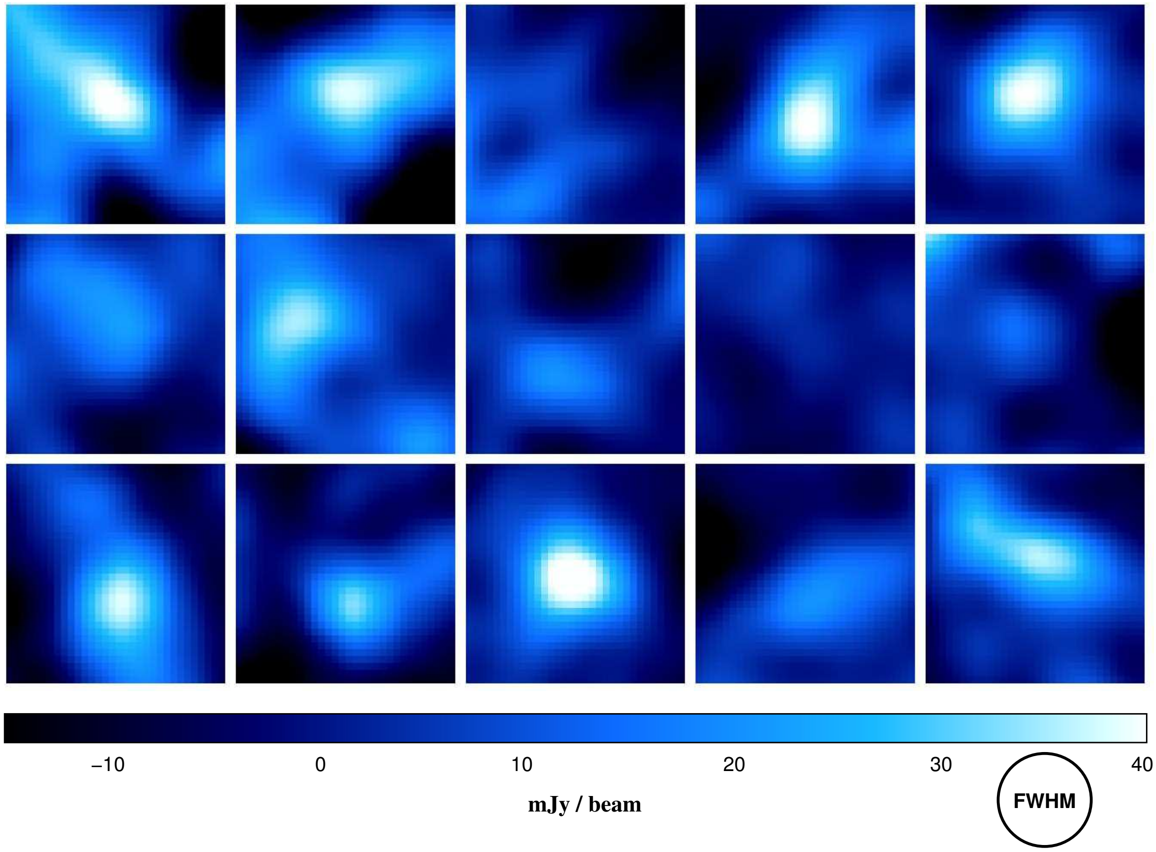


Figure 2.1 SHARC-2 $350\,\mu$m thumbnail images of the SMGs observed (sources $1-15$ in Tables 2.1–2.3) shown in reading order (left to right, top to bottom). Images are 30″×30″ regions centered on the radio positions, and displayed on the same flux scale for comparison. The $12.4''$ FWHM effective beams of the optimally filtered (convolved) images are also indicated. Sources 3, 9 and 10 are not detected.

ing data taken during each run. Additionally, our preliminary source identifications revealed a small, albeit significant ($\simeq$3 arcsec) systematic pointing shift in the negative RA direction. All maps were re-aligned accordingly. The reconstructed pointing is accurate to 3.5″ rms, in good agreement with CSO specifications.

The data were processed using the CRUSH[1] software package, developed at Caltech (see Chapter 4), which models total power detector signals using an iterated sequence of maximum likelihood estimators. The FWHM point-spread function of the final maps (fig. 2.1) are approximately 12″ after optimal filtering to search for beam-sized ($\simeq$9″) features in order to yield maximal signal-to-noise ratios on point sources. This degradation of the image resolution by around $\sqrt{2}$, from the nominal instrumental beam width of 8.5″, is because the fitting for beam sized features is equivalent to convolving the image with the slightly wider effective beam of around 9″, which accounts for the smearing of the nominal beam by pointing and focus variations on the long integrations.

Calibration was performed primarily against planets and asteroids, when available, or using stable galactic continuum sources and Arp220.[2] Trends in aperture efficiency, especially with elevation, were estimated and were taken into account when calibrating our science targets. The systematic flux filtering effect of the aggressive reduction parameters used in CRUSH to subtract noise signals and reach maximal image depths, was carefully estimated for the case of point sources, and the appropriate corrections were applied to the images. The final calibration is expected with high confidence to be more accurate than 15%, with systematic effects anticipated to be less.

The maps produced by CRUSH reveal Gaussian noise profiles with an anticipated tail at positive fluxes. Hence, in searching for $350\,\mu$m counterparts around the published radio positions, we calculate detection thresholds based on the effective number of beams $N_{\rm beam} \approx 1 + A/A_{\rm beam}$ (see Appendix A) inside a detection area A in the smoothed image. A detection confidence $C \approx 1$ is reached at the significance level SNR at which,

$$1 - C \approx \frac{N_{\rm beam}}{\sqrt{2\pi}} \int_{SNR}^{\infty} e^{-\frac{1}{2}x^2}\, dx. \tag{2.1}$$

As for the appropriate search radius, the probability that the peak of the smoothed image falls at some radius away from the true position is essentially the likelihood that the smoothed noise level at that radius is sufficient to make up for the deficit in the underlying signal. Therefore, for the case of Gaussian noise profiles, we expect the detection to lie inside of a $2\,\sigma$ noise peak, and therefore within a maximal radius given by the condition $SNR = 2/[1 - \exp(-r_{\rm max}^2/2\sigma_{\rm beam}^2)]$ for a Gaussian beam profile with $\sigma_{\rm beam}$ spread at an underlying signal-to-noise ratio of SNR. To this we must add the appropriate pointing tolerance (at $2\,\sigma$ level) in quadrature, arriving at an expression

---

[1] `http://www.submm.caltech.edu/˜sharc/crush`.

[2] `http://www.submm.caltech.edu/˜sharc/analysis/calibration.htm`.

for the detection radius around the actual position as a function of detection significance, that is,

$$r_{\max}^2 = 4\,\sigma_{\rm pointing}^2 - 2\,\sigma_{\rm beam}^2 \ln\left(1 - \frac{2}{SNR}\right). \tag{2.2}$$

Note, that this expression simplifies to $r_{\max} \rightarrow 2\sigma_{\rm beam}/\sqrt{SNR}$ for the case of $SNR \gg 1$ and negligible $\sigma_{\rm pointing}$.

The combination of $SNR = 2.30$ and $r_{\max} = 10.4''$ simultaneously satisfy both constraints (eqs. [2.1] and [2.2]) at $C = 95\%$ confidence level for a $9''$ $FWHM$ effective beam and $\sigma_{\rm pointing} = 3.5''$ pointing rms. Potential candidates thus identified are subsequently verified to lie within the expected distance from their respective radio positions. The resulting identifications are summarized in Table 2.1. When counterparts were not found, the peak measurement values inside the search area are reported.

The sources, collectively, are much smaller than the SHARC-2 beam (i.e., $d \ll 9''$), as neither (i) fitting larger ($12''$) beams or (ii) the filtering of extended structures produce systematically different fluxes for the sample as a whole. Source extents typically $\lesssim$30 kpc are therefore implied. While the partial resolution of a few objects cannot be excluded, the peak fluxes of the optimally filtered images are expected to be generally accurate measures of the total integrated flux for the objects concerned.

2.3 Spectral Energy Distributions of SMGs

We fitted the SHARC-2 $350\,\mu$m and SCUBA $850\,\mu$m fluxes, combined with Bolocam $1100\,\mu$m (Laurent et al. 2005) and MAMBO $1200\,\mu$m (Greve et al. 2004) data when available, with single temperature, optically thin grey-body models of the form $S(\nu, T) \propto \kappa(\nu) B(\nu, T)$, where $\kappa(\nu) \propto \nu^{\beta}$ is an approximation for the full emissivity term $(1 - \exp[(\nu/\nu_0)^{\beta}])$ for $\nu_{obs} \ll \nu_0$. Alternative SED models incorporating the full optical depth, or a distribution of temperatures and power-law Wien tails did not provide better description the data. Specifically, the flattening of the Rayleigh-Jeans slope due to optical depths approaching unity is not detectable with the $\simeq$10% uncertain relative calibration of the bands, while the Wien side of the spectra is not sampled by the observations. More complex SED models, e.g., the two temperature model used by Dunne & Eales (2001), were not considered, since these require greater number of parameters than can be determined from the few, often just two, photometric data points available for the typical SMG.

SED models, whether incorporating an emissivity slope, full optical depth or multiple temperature components, are simple parametrizations of complex underlying spectral distributions, produced by a range of dust properties inside the targeted galaxies. Therefore, T and β of the model describe not the physical temperature and emissivity of every single dust grain, but provide an

Table 2.1. Summary of Observations

ID	Name	Offset J2000[a] (arcsec)	z	S(350 μm) (mJy)	S(850 μm) (mJy)	S(1100 μm) (mJy)	S(1200 μm) (mJy)	S(1.4 GHz) (μJy)
1	SMM J030227.73+000653.5	+1.0,+0.8	1.408	42.2 ± 9.8	4.4 ± 1.3	···	···	217 ± 9
2	SMM J105207.49+571904.0	-1.0,+3.7	2.689	38.0 ± 7.2	6.2 ± 1.6	···	(0.4 ± 0.8)	277.8 ± 11.9
3 [b]	SMM J105227.77+572218.2	···	1.956	(11.3 ± 6.7)	7.0 ± 2.1	5.1 ± 1.3 [b]	3.1 ± 0.7	40.4 ± 9.4 [f]
4 [b]	SMM J105230.73+572209.5	+3.3,-1.8	2.611	41.0 ± 6.8	11.0 ± 2.6	5.1 ± 1.3 [b]	2.9 ± 0.7	86.3 ± 15.4 [f]
5	SMM J105238.30+572435.8	+1.4,+2.5	3.036	40.5 ± 6.5	10.9 ± 2.4	4.8 ± 1.3	4.8 ± 0.6	61.0 ± 22.0 [f]
6	SMM J123600.15+621047.2	-1.4,+2.0	1.994	22.3 ± 6.3	7.9 ± 2.4	···	···	131 ± 10.6
7	SMM J123606.85+621021.4	+6.7,+3.5	2.509	35.1 ± 6.9	11.6 ± 3.5	···	···	74.4 ± 4.1
8	SMM J131201.17+424208.1	···	3.405	21.1 ± 7.7	6.2 ± 1.2	···	···	49.1 ± 6.0
9	SMM J131212.69+424422.5	+2.2,-4.2	2.805	(3.7 ± 4.4)	5.6 ± 1.9	···	···	102.6 ± 7.4
10	SMM J131225.73+423941.4	···	1.554	(14.7 ± 7.4)	4.1 ± 1.3	···	···	752.5 ± 4.2
11	SMM J163631.47+405546.9	-1.3,-4.0	2.283	38.3 ± 5.5	6.3 ± 1.9	···	(1.1 ± 0.7)	99 ± 23
12	SMM J163650.43+405737.5	-1.0,-1.7	2.378	33.0 ± 5.6	8.2 ± 1.7	···	3.1 ± 0.7	221 ± 16
13	SMM J163658.19+410523.8	+1.1,-1.2	2.454	45.2 ± 5.3	10.7 ± 2.0	···	3.4 ± 1.1	92 ± 16
14	SMM J163704.34+410530.3	-1.8,-4.4	0.840	21.0 ± 4.7	11.2 ± 1.6	···	(0.8 ± 1.1)	45 ± 16
15	SMM J163706.51+405313.8	-2.3,+2.4	2.374	36.1 ± 7.7	11.2 ± 2.9	···	4.2 ± 1.1	74 ± 23
(16)	SMM J163639.01+405635.9	···	1.495	···	5.1 ± 1.4	···	3.4 ± 0.7	159 ± 27
(17)[e]	SMM J105201.25+572445.7	···	2.148	24.1 ± 5.5	9.9 ± 2.2	4.4 ± 1.3	3.4 ± 0.6	72.1 ± 10.2
(18)	SMM J105158.02+571800.2	···	2.239	···	7.7 ± 1.7	···	2.9 ± 0.7	98.1 ± 11.6
(19)[d,e]	SMM J105200.22+572420.2	···	0.689	15.5 ± 5.5	5.1 ± 1.3	4.0 ± 1.3	2.4 ± 0.6	57.4 ± 13.2
(20)[e]	SMM J105227.58+572512.4	···	2.142	44.0 ± 16.0	4.5 ± 1.3	4.1 ± 1.3	2.8 ± 0.5	39.2 ± 11.4 [f]
(21)[c]	SMM J105155.47+572312.7	···	2.686	···	5.7 ± 1.7	···	3.3 ± 0.8	46.3 ± 10.2

Note. — MAMBO 1200 μm fluxes are from Greve et al. (2004), upper limits (bracketed fluxes) are from T. Greve (priv. comm.), Bolocam 1100 μm fluxes are from Laurent et al. (2006), 850μm SCUBA and 1.4 GHz fluxes are taken from Chapman et al. (2005). Biggs & Ivison (2006) provide alternative radio fluxes for many of the objects listed here. The bracketed IDs indicate sources that have not been observed by the authors within the context of this paper, but archived data is used in the analysis.

[a] SHARC-2 detection offsets are with respect to the published radio positions.

[b] Both of these sources are likely contributors to the observed 1100 μm flux. The Bolocam data are ignored in the analysis.

[c] Chapman et al. (2005): "These SMGs have double radio identifications, both confirmed to lie at the same redshift."

[d] Chapman et al. (2005): "These SMGs have double radio identifications, one lying at the tabulated redshift and a second lying at $z < 0.5$." The higher redshift source is assumed to be the dominant contributor.

[e] SHARC-2 350 μm fluxes from Laurent et al. (2006).

[f] The alternative radio fluxes from Biggs & Ivison (2006) are significantly different from or more accurate than the values listed here.

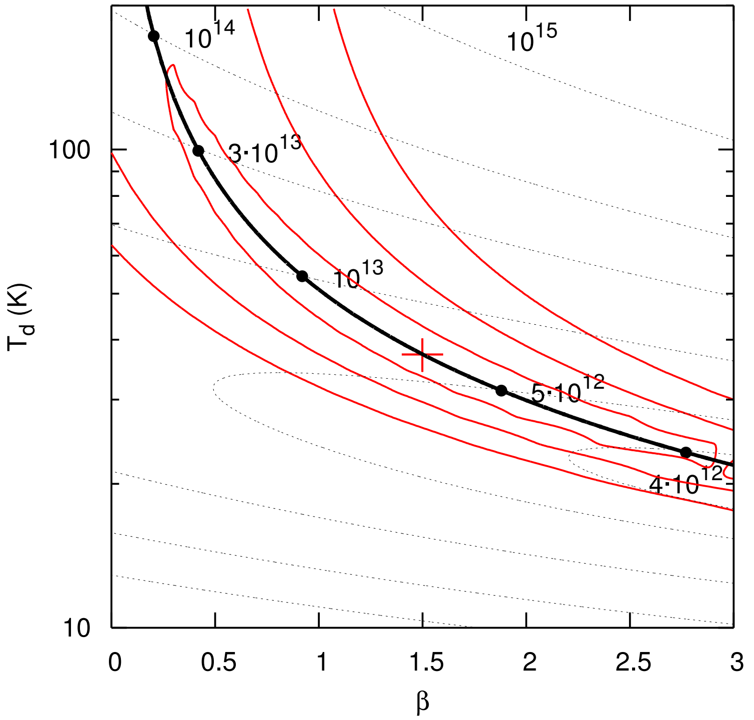


Figure 2.2 The characteristic dust temperature T_d and effective emissivity index β are highly correlated parameters of the SED fit for the typical SMG, even when three photometric data points allow simultaneous fitting of both T_d and β (source 13 from Table 2.2, shown with solid 1,2 and $3\,\sigma$ contours). However, by assuming $\beta = 1.5$, we can accurately constrain luminosities (dotted contours) and temperatures, which are listed in Tables 2.2 and 2.3. The derived properties may be scaled to other β values of the reader's preference using the relations of eq. (2.3) (thick solid line).

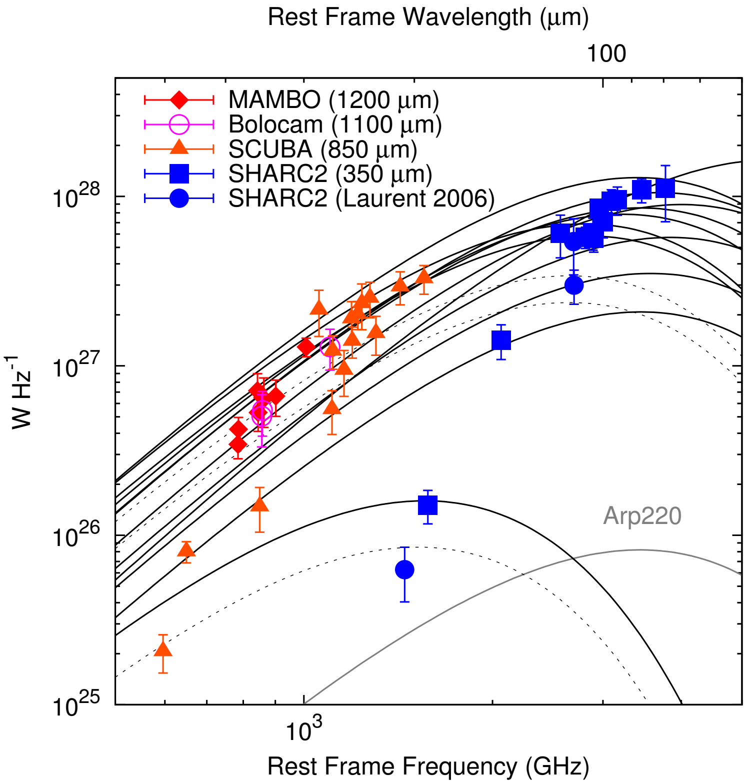


Figure 2.3 Optically thin grey-body SED models were fitted (solid lines) to the various available observed fluxes, and are transformed to the rest frames of the galaxies for comparison. Fits using (Laurent et al. 2006) data also shown (dotted lines). All the spectra peak in the neighbourhood of the redshifted $350\,\mu$m point, providing testimony of the pivotal importance of the SHARC-2 data point in constraining these models. The SEDs shown represent our best fit to the data for sources detected by SHARC-2, use an emissivity index $\beta = 1.5$, and incorporate radio fluxes via the far-infrared to radio correlation with $q_L = 2.14$ with an assumed 0.12 dex dispersion.

effective characterization of dust emission in a galaxy as a whole. In this sense, the characteristic β values, expected in the range $1-2$, reflect both the underlying grain emissivities, and a distribution of physical temperatures within the observed objects.

Simultaneous fitting of T and β requires at least three photometric data points (while many SMGs have only two), and even when permitted these are highly correlated parameters of the fit for the typical SMG (fig. 2.2). Therefore, we initially assume $\beta = 1.5$ for the SED fit, as it provides good characterization of actively star-forming environments, both in accurately modeled galactic clouds (Dupac et al. 2003) and galaxies of the local universe (Dunne & Eales 2001), and is broadly consistent with laboratory measurements on carbite and silicate grains (Agladze et al. 1996). The properties, thus derived, may be scaled to the β values of the reader's choosing, via the deduced spectral indices ε and λ listed in the Tables 2.2 and 2.3 as,

$$T_d \propto \beta^{-\varepsilon},\ L_{FIR} \propto \beta^{-\lambda},\ M_d \propto \beta^{\varepsilon(4+\beta)-\lambda}. \tag{2.3}$$

We included 15% calibration uncertainty in addition to the published statistical uncertainties for all submillimeter data. Nevertheless the luminosities are constrained accurately because the SHARC-2 $350\,\mu$m measurement falls near the emission peak for all of these sources (fig. 2.3).

Flux boosting or Eddington bias (Coppin et al. 2005), in the case of less significant detections, where the selection is in the submillimeter, induce a small bias in the derived SEDs. As the bias will not apply to follow-up measurements, and because the exact quantification of the Eddington bias is nontrivial, requiring a priori knowledge of the underlying source distributions, we left fluxes uncorrected for its effects.

For the first time we are able to use multiband photometry to determine accurately the characteristic dust temperatures and far-infrared luminosities for the SMG population. Luminosities (Tables 2.2 and 2.3) are calculated analytically from the fitted grey body model $S(\nu, T)$, following (De Breuck et al. 2003),

$$L_{FIR} = 4\pi\, D_L^2\, \Gamma(4+\beta)\, \zeta(4+\beta) \left[\frac{kT}{h\nu}\right]^{4+\beta} \left(e^{h\nu/kT} - 1\right)\, \nu\, S(\nu, T), \tag{2.4}$$

where luminosity distances (D_L) were obtained[3] for a Λ-CDM cosmology with $H_0 = 65\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_M = 0.3$ and $\Omega_\Lambda = 0.7$. The above expression provides the correct SED integral as long as the transition from the optically thin grey-body approximation to optically thick black body is above the emission peak where the contribution to the total luminosity is negligible. If power law Wien tails, with spectral slopes of $-\alpha$ (Blain et al. 2003), are assumed instead, the luminosities can be scaled by a constant $\eta_\beta(\alpha)$, which is around 1.5 for an Arp220 type template with $\alpha = 2$. More generally, in the range of $\alpha \sim 1.1$–4.0 and $\beta \sim 0$–3, the values of η are well approximated (with an rms of 5%

---

[3] `http://www.astro.ucla.edu/~wright/CosmoCalc.html`.

or 0.02 dex) by the empirical formula,

$$\eta_\beta(\alpha) \approx (1.44 + 0.07\,\beta)(\alpha - 1.09)^{-0.42}.$$

Similar corrections may be derived for the case of increasing optical depth, with η as a function of $h\nu_0/kT$.

Illuminated dust masses (Tab. 2.3) were also estimated from the SED model $S(\nu, T)$, using (De Breuck et al. 2003),

$$M_d = \frac{S(\nu,T) D_L^2}{(1+z)\,\kappa_d(\nu_{rest})\,B(\nu_{rest}, T_d)}.$$

Here, the normalization for the absorption efficiency was assumed to be $\kappa_{850\,\mu\mathrm{m}} = 0.15\,\mathrm{m^2\,kg^{-1}}$, representing the extrapolated average $125\,\mu$m value of $2.64 \pm 0.29\,\mathrm{m^2\,kg^{-1}}$ (Dunne et al. 2003) from various models by assuming β of 1.5. In comparison to the gas masses derived from CO measurements (Greve et al. 2005; Tacconi et al. 2006)[4], we find an average gas-to-dust ratio of 54^{+14}_{-11} $[\kappa_{850\,\mu\mathrm{m}}/0.15\,\mathrm{m^2\,kg^{-1}}]$, resembling the ratios seen in nuclear regions of local galaxies by Seaquist et al. (2004), who assume a $\kappa_d\,X_{\mathrm{CO}}$ product comparable to ours. The individual measurements indicate an intrinsic spread of $\simeq$40% around the mean value. The low ratios may be interpreted as an indication for the relative prevalence of dust in SMGs over the local population which typically exhibit Milky Way like gas-to-dust ratios around 120 (Stevens et al. 2005) or, alternatively, that absorption in SMGs is more efficient with $\kappa_{850\,\mu\mathrm{m}} \approx 0.33\,\mathrm{m^2\,kg^{-1}}$.

Our estimates of the dust temperatures, and hence luminosities, are systematically lower, than those anticipated based on the $850\,\mu$m and $1.4\,$GHz fluxes alone by Chapman et al. (2005), who overestimate these quantities by 13% and a factor of $\simeq$2 respectively, when assuming the local far-infrared to radio correlation. This discrepancy can be fully reconciled if a different constant of correlation is assumed for SMGs (see Section 2.4).

We confirm that the SMG population is dominated by extremely luminous (several times $10^{12}\,L_\odot$) systems with $\simeq 10^9\,M_\odot$ of heated dust, and characteristic 35 K dust temperatures typical to actively star forming ULIRGs. As anticipated, these objects resemble the nearby archetypal ULIRG, Arp220 (with $T_d \approx 37\,$K, similarly obtained), except that they are about 7 times more luminous on average.

2.3.1 Cold, Quiescent SMGs?

In addition, there appear to a be several cooler ($T_d \lesssim 25\,$K), less luminous (10^{11}–$10^{12}\,L_\odot$) objects, albeit with comparable dust masses (few times $10^9\,M_\odot$) present in the population (sources 14 and

---

[4] The CO fluxes from Tacconi et al. (2006) were converted into gas masses using the conversion factor X_{CO} of $0.8\,M_\odot\,(\mathrm{K\,km\,s^{-1}pc^2})^{-1}$ in Solomon et al. (1997) appropriate for ULIRGs (Downes & Solomon 1998). The results are therefore consistent with Greve et al. (2005).

Table 2.2. Far-Infrared Properties of SMGs from SHARC-2 Data

ID	T_d (K)	$\log L_{FIR}$ ($L_\odot$)	q_L	ε	λ
1	43.3 ± 18.7	12.83 ± 0.64	2.39 ± 0.64	1.36	2.94
2	66.8 ± 26.1	13.49 ± 0.57	2.31 ± 0.57	1.61	3.69
3	21.3 ± 3.8	11.95 ± 0.24	1.91 ± 0.26	0.50	0.21
4	40.1 ± 5.8	12.98 ± 0.19	2.33 ± 0.21	0.84	1.06
5	39.1 ± 4.2	13.03 ± 0.14	2.39 ± 0.21	0.69	0.66
6	25.7 ± 4.7	12.28 ± 0.18	1.71 ± 0.18	0.52	0.35
7	31.0 ± 5.3	12.73 ± 0.15	2.19 ± 0.15	0.54	0.39
8	40.9 ± 8.2	12.85 ± 0.25	2.20 ± 0.26	0.57	0.46
9	21.1 ± 6.3	11.95 ± 0.29	1.16 ± 0.29	0.31	0.04
10	24.3 ± 6.8	11.88 ± 0.35	0.79 ± 0.35	0.59	0.50
11	53.4 ± 16.2	13.18 ± 0.42	2.61 ± 0.43	1.38	2.88
12	34.5 ± 4.5	12.72 ± 0.17	1.75 ± 0.17	0.73	0.77
13	37.3 ± 5.0	12.93 ± 0.16	2.31 ± 0.18	0.77	0.91
14	16.8 ± 2.3	11.38 ± 0.15	2.13 ± 0.22	0.65	0.53
15	31.5 ± 4.2	12.71 ± 0.16	2.23 ± 0.21	0.65	0.57
(17)	27.4 ± 3.0	12.41 ± 0.15	2.02 ± 0.16	0.61	0.44
(19)	14.2 ± 2.2	10.97 ± 0.21	1.82 ± 0.23	0.58	0.41
(20)	40.6 ± 12.3	12.87 ± 0.48	2.76 ± 0.50	0.99	1.61

Note. — A summary of the derived far-infrared properties of the observed SMGs. All quantities were derived using an optically thin approximation with $\beta = 1.5$. Temperatures and luminosities may be scaled to other β values using the relations of eq. (2.3), and the corresponding indeces ε and λ listed in here. The q_L values derived from the radio data of Biggs & Ivison (2006) tend to be higher than these by 0.06 on average.

19, and possibly 3, 9 and 10 in Tables 2.2 and 2.3). While these resemble the Milky Way in temperatures, and hence probably in star formation densities, they are tens of times more luminous than the Galaxy.

This combination of extreme dust masses yet surprisingly low relative star formation rates, indicated by the lesser dust heating, allude to possibly incorrect low redshift identifications. Should these galaxies lie at the higher, more typical redshifts of the SMG population, their temperatures and luminosities would closely resemble those of the hotter population. However, while lensing of distant SMGs by massive foreground galaxies at the measured redshifts, is conceivable (Blain et al. 1999a; Chapman et al. 2002), these should be rare, and unlikely to account for all cold SMGs.

Alternatively, these seemingly normal type galaxies could, potentially, be remnants to the more remote, hotter population of SMGs, once the short lived starbursting activity subsides. Dust could persist and remain detectable in the more proximate universe, provided that radiation from the rapidly growing black holes (Borys et al. 2005; Alexander et al. 2005a) does not disperse dust entirely, e.g., dusty QSOs of Benford et al. (1999) and Beelen et al. (2006), or if the dust can later reassemble to form large disk galaxies, like source 14 which is consistent with the large spiral disk galaxy seen in HST images (Almaini et al. 2006; Borys et al. 2006), or the cold dust mergers suggested by the chain of $16\ \mu$m emission seen for source 19. This connection between the hot and cold SMGs is further hinted at by the apparent trend of dust temperatures increasing with redshift (fig. 2.4) even without the low redshift data. We, therefore, find the existence of massive cold SMGs plausible.

2.4 The Far-IR/Radio Correlation at High Redshifts

One of the most valuable uses of the new measurements is to test the local far-infrared to radio correlation at the redshifts of SMGs ($z \sim$ 1–3). The simple extension of the correlation into the distant universe was suggested by earlier studies based on a combination of $1.4\,$GHz radio with *ISO* $15\,\mu$m (Garrett 2002) and *Spitzer* $24\,\mu$m and $70\,\mu$m observations (Appleton et al. 2004), both of which take advantage of SED template fitting to extrapolate fluxes into the far-infrared regime. More recently, Beelen et al. (2006) hints at the validity of the correlation in the case of a handful of distant ($z \sim 2-6$) quasars with SEDs well constrained directly from far-infrared measurements.

A quantitative treatment of the correlation (Helou et al. 1985), has been formulated in terms of *IRAS* $60\,\mu$m and $100\,\mu$m fluxes, expressed in the rest frame as,

$$q = \log\left(\frac{FIR}{3.75\times10^{12}\,\mathrm{W\,m^{-2}}}\right) - \log\left(\frac{S_{1.4\,\mathrm{GHz}}}{\mathrm{W\,m^{-2}\,Hz^{-1}}}\right), \tag{2.5}$$

where the far-infrared parameter $FIR = 1.26\times10^{-14}(2.58\,S_{60\,\mu\mathrm{m}} + S_{100\,\mu\mathrm{m}})\,\mathrm{W\,m^{-2}}$ and the fluxes

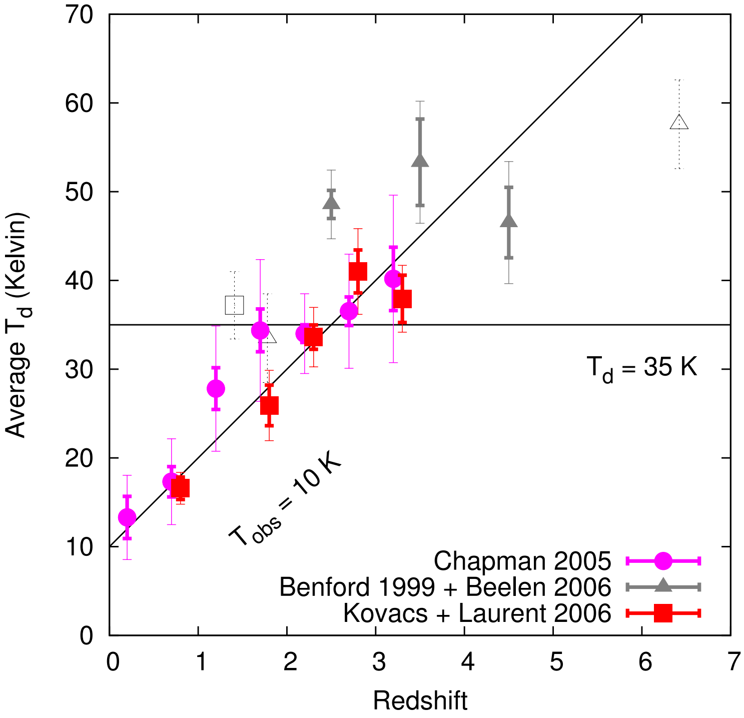


Figure 2.4 Median dust temperatures vs. redshift. Data is binned for this paper (squares), for the Chapman et al. (2005) sample (dots) with temperatures estimates incorporating the radio data, and for dusty quasars (triangles). $1\,\sigma$ error bars were estimated from the scatter inside the bins and the uncertainties of dust temperatures where available (thick bars). Dispersion within bins are also shown (thin bars), while dotted error bars indicate measurement error for single datum inside a bin. T_d of $35 \pm 3\,$K may be applicable for $z \sim 1.5$–3.5 which, if true, can provide photometric redshifts in that range. However, observing frame temperatures of around $10\,$K appear to fit the SMG data somewhat better than constant T_d in the rest frame. This similarity of the SEDs in the observing frame may owe to selection effects.

$S_{60\,\mu\mathrm{m}}$ and $S_{100\,\mu\mathrm{m}}$ are in Jy. The quantity essentially estimates the flux in a wide band centered at $80\,\mu$m, and is a good tracer of the far-infrared luminosity for temperatures in the range of 20–80 K and emissivity indices $\beta \sim 0$–2 (Helou et al. 1988).

Perhaps, a better measure of the correlation is obtained by directly comparing luminosities. Therefore, we propose to use,

$$q_L = \log\left(\frac{L_{FIR}}{4.52\,\mathrm{THz}\cdot L_{1.4\,\mathrm{GHz}}}\right). \tag{2.6}$$

Here the normalization frequency $4.52\,\mathrm{THz}$ has been derived for the adopted grey-body model[5] such that $q_L \to q$ for $T = 40\,\mathrm{K}$ and $\beta = 1.5$, thus approximating the original definition of q while extending its usefulness beyond the original restrictions in temperature and emissivity. Radio luminosities, in turn, are calculated as,

$$L_{1.4\,\mathrm{GHz}} = 4\pi D_L^2 S_{1.4\,\mathrm{GHz}}(1+z)^{\alpha-1}, \tag{2.7}$$

which includes a bandwidth compression by $(1+z)^{-1}$ and a K-correction $(1+z)^{\alpha}$ to rest frame $1.4\,$GHz. We assume a synchrotron power law of $S \propto \nu^{-\alpha}$, with the spectral index α for which a typical value of 0.7 (Condon 1992) of nonthermal radio sources, is assumed for the calculations.[6]

We confirm that the far-infrared to radio correlation appears to hold throughout the redshift range of $z \sim 1$–3 (figs. 2.5 and 2.6), with the notable exceptions of sources 9 and 10, which likely host radio loud AGNs[7]. The detections reveal $\overline{q} \approx 2.07 \pm 0.09$ and $\overline{q}_L = 2.12 \pm 0.07$ with intrinsic spreads estimated at $\sigma_q \approx 0.21$ and $\sigma_{q_L} \approx 0.12$ around the respective mean values. The alternative radio fluxes of Biggs & Ivison (2006) provide q values that are higher by 0.06 on average and scatter slightly more with $\sigma_q \approx 0.30$ and $\sigma_{q_L} \approx 0.23$. The reduced dispersion in q_L vs. q and the somewhat peculiar case of the two cold lower redshift objects (14 and 19), whose derived dust temperature of only $16\,$K fall outside the range where FIR traces luminosity (eq. [2.5]), highlight the practicality of the luminosity-based measure q_L vs. the original *IRAS*-based one. The derived mean values are significantly less than the locally observed median values of 2.75 ± 0.03 for local spirals and starbusts galaxies from the *IRAS* Faint Catalog (Condon & Broderick 1991), 2.34 ± 0.01 for radio identified, flux limited ($S_{60\,\mu\mathrm{m}} > 2\,\mathrm{Jy}$) *IRAS* sources (Yun et al. 2001) or $\simeq$2.3 for a range of normal galaxies (Condon 1992) that include spirals and irregulars, E/S0 types, *IRAS* and radio selected samples. The corresponding local scatters are also larger with $\sigma_q = 0.14$, $\sigma_q = 0.33$ and $\sigma_q \lesssim 0.2$ respectively.

---

[5] If SEDs include power-law Wien tails instead, the normalizing frequency should be appropriately scaled by $\eta_\beta(\alpha)$. Similar adjustements to the normalization can be derived for any other SED model, hence the shape of the particular SED model used has no effect on the correlation parameter q_L otherwise.

[6] The assumption of the radio spectral index α is not critical. Indeces different by $\delta\alpha$ may produce a small bias in q on the order of $\delta q \approx 0.5\delta\alpha$ in the redshift range of $z \sim 1$–3 of SMGs, with an undetectable the redshift dependence as $\delta q \ll \sigma_q$.

[7] Source 9 has been characterised as AGN by Chapman et al. (2005).

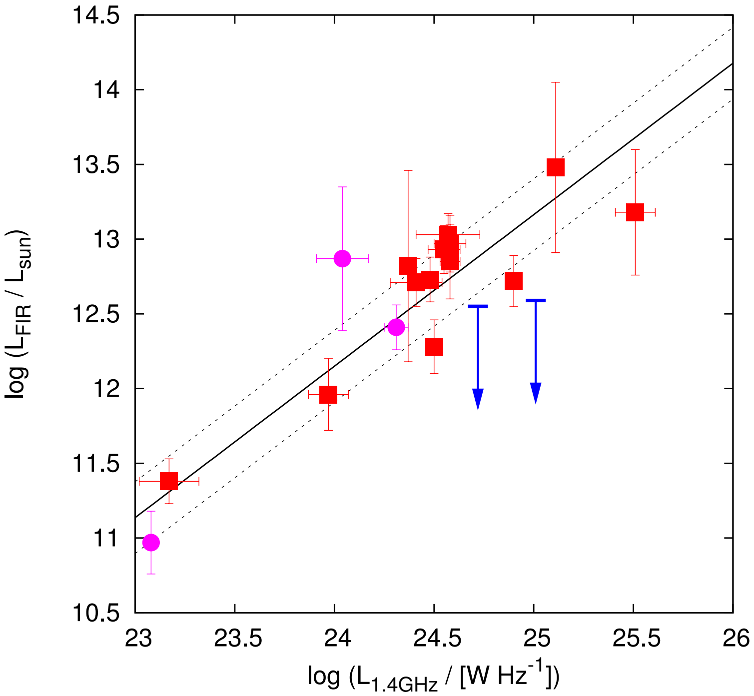


Figure 2.5 The radio to far-infrared correlation for SMGs. Galaxies observed by the authors shown with squares, data from Laurent et al. (2006) with dots. $2\,\sigma$ upper limits are indicated when appropriate (arrows). Far-infrared luminosities were calculated exclusively based on the available submillimeter measurements, with $\beta = 1.5$. The best fit model (solid line) reveals no deviation from linearity in the relationship. The deduced $2\,\sigma$ instrinsic scatters around the model are also shown (dotted lines).

Moreover, we find the relationship to be linear, within measurement uncertainty, over nearly three decades of luminosities of the observed SMGs, with a best fit FIR to radio luminosity index, $d\log L_{FIR}/d\log L_{1.4\,\mathrm{GHz}}$, of 1.02 ± 0.12. Nonetheless, small levels of nonlinearity, like those discussed by Fitt et al. (1988); Cox et al. (1988); Devereux & Eales (1989) and Condon et al. (1991) remain possible.

The low q values, the tightness of the correlation, the observed linearity, and the typically warm dust temperatures, all point to dust heating that is dominated by high mass ($>8\,M_{\odot}$) stars. In the two component model of Yun et al. (2001), short lived ($<10^7$ yr) high mass stars fuel the 'hot' thermal component that is strongly coupled to the nonthermal radio via Type II supernovae, whereas the 'cold' component heated by a different population of lower mass stars is expected to produce a different correlation with a weaker coupling. They estimate $q_1 = 2.3$ and $q_2 > 4$ respectively for the components. As most galaxies contain both components, intermediate q values are typically observed. The low qs that characterize SMGs are, therefore, indicative of the predominance of high mass star formation in the dust heating. Moreover, the SMG result forces the revision of the 'hot' component correlation to $q_1 \lesssim 2.1$, or lower still (by about 0.16), if steeper radio spectral slopes of $\alpha \simeq 1$ are assumed for these very luminous objects (Yun et al. 2001).

The presence of radio-quiet AGNs would contribute to the total far-infrared emission, biasing towards higher qs. Radio-loud AGNs, on the other hand would produce low q values, albeit to wildly varying extent, which should manifest as an increased spread in the correlation. The tightness of the observed correlation, therefore, confirms that radio loud AGNs are rare, while the low q values, otherwise observed, indicate that the AGN contribution to the far-infared luminosity is small when compared to heating from star formation-activity in SMGs. This is consistent with Alexander et al. (2005a,b), who used ultradeep Chandra X-ray data to find that the AGNs, present in at least 75% of SMGs, contribute little to the total bolometric luminosity of their hosts. Dust heating, therefore, is assumed to be star formation dominated.

Last but not least, the low q values observed may partially arise from selection bias as our initial sample would miss SMGs with low radio fluxes, i.e., sources with higher q values, in the secondary selection required for redshift identification. However, faint radio sources with undetectable submillimeter counterparts, representing a hypothetised hot extension to the SMG population, may be more typical (Chapman et al. 2004; Blain et al. 2004), and thus selection bias could be reversed, missing the low q values instead.

2.5 Dust Emissivity Index

It is possible to further constrain the dust properties, e.g., by measuring the effective emissivity index β. In principle, this may be attempted on an individual basis, since several galaxies have

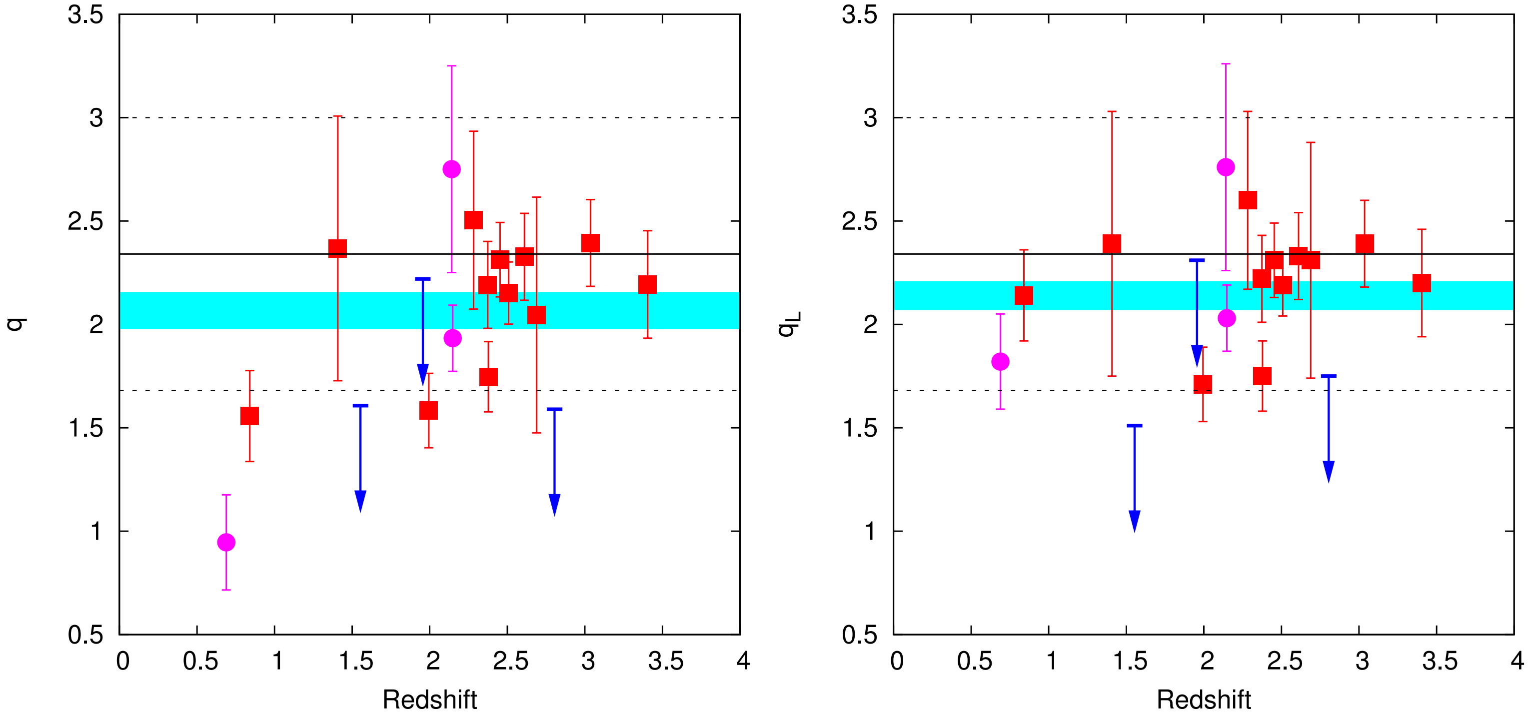


Figure 2.6 (a) Original definition of q from Helou et al. (1985), and (b) the modified, luminosity based definition q_L as a function of redshift. $2\,\sigma$ upper limits are indicated as appropriate (arrows). Data from this paper are shown with squares while those from Laurent et al. (2006) as dots. The observed SMGs seem to consistently fall below the locally observed mean of 2.34 (Yun et al. 2001) for local bright *IRAS* sources (solid line) whose $2\,\sigma$ scatters are also indicated (dashed lines). The values were all derived assuming $\beta = 1.5$. The derived distribution mean values of the nonradio loud subsample ($q > 1$) are also shown (shaded bands).

MAMBO (Greve et al. 2004) and/or Bolocam (Laurent et al. 2005) detections at longer wavelengths, providing justification for a 3-parameter fit. In practice, however, such an approach is unlikely to yield robust quantities for individual galaxies (see fig. 2.2) owing to the proximity of the observed bands and the modest significance of the detections. Therefore, we aimed to fit a single emissivity slope β for the entire sample hoping to provide a better constrained, ensemble averaged, dust emissivity index. Moreover, we extend our sample to Chapman et al. (2005) SMGs with multiband submillimeter data which were not observed in this paper. MAMBO associations (Greve et al. 2004) and the $350\,\mu$m follow-up of Bolocam and candidates (Laurent et al. 2006) provide 5 extra sources.

We analyze separately the data consisting of detections only, and that including measurement peaks inside the search area for nondetections. While the first set may appear more robust, it could be more susceptible to selection biases due to the rigorous detection requirement. The larger set, containing upper limits, is thus likely to be less biased of the two, and could provide a better characterisation of the whole SMG population.

We have used a modified, nested, version of the downhill simplex method (Press et al. 1986) to perform the fit, yielding $\beta = 2.42 \pm 0.71$, a value that is too uncertain to distinguish between the expected values in the range of $1 - 2$. However, the demonstrated validity of the linear far-infrared to radio correlation for the SMG sample, allows us to incorporate the radio fluxes to constrain the total integrated far-infrared luminosities, and improve the quality of the fit. Here we assumed 0.12 dex as the intrinsic spread in q_L, producing the expected model deviation with $\hat{\chi}^2 \to 1$, and consistent with the scatter observed for local spirals and starbursts (Condon et al. 1991).

To avoid radio loud objects biasing our analysis, we ignored sources 9 and 10 due to their a priori low q_L values. At the same time, we utilized a modified χ deviation definition to allow some of the sources to become "radio loud" in the fit, i.e. to significantly deviate from the correlation in the lower q_L direction. Specifically, when the derived q_L value for a source is more than $2\,\sigma$ below the average $\overline{q}_L$ of the fit, we use a relaxed deviation with $\sigma'_{q_L} = 2$, more characteristic of radio loud objects, while keeping χ^2 continuous over the entire parameter space. The fraction of normal sources falsely identified as "radio loud" by this method is expected to be around 2%, and therefore the effects of mischaracterizations are expected to remain limited.

With the local median q value of 2.34 we obtain $\beta = 0.81 \pm 0.19$ for the detections and $\beta = 1.00 \pm 0.19$ when all the available information is used in the analysis. Both these results imply an emissivity index significantly less than the usually quoted values around 1.5. The interpretation, however, relies entirely upon the assumption that the local radio–far-infrared correlation holds unaltered for higher redshifts.

For a more rigorous treatment, we obtained confidence contours for the likely combinations of both β and q_L (fig. 2.7), hence avoiding a priori assumptions on both of these quantities. The smaller set, with detection requirement imposed, favours somewhat lower β values, in combination with

Table 2.3. Properties of SMGs Incorporating Radio Data

ID	T_d (K)	$\log L_{FIR}$ ($L_\odot$)	$\log M_d$ ($M_\odot$)	ε	λ
1	37.2 ± 3.8	12.60 ± 0.12	8.64 ± 0.17	0.82	1.09
2	60.3 ± 6.1	13.34 ± 0.12	8.22 ± 0.16	0.86	1.15
3	23.5 ± 2.5	12.10 ± 0.15	9.24 ± 0.17	0.72	0.82
4	37.1 ± 3.4	12.86 ± 0.11	8.91 ± 0.14	0.77	0.89
5	36.8 ± 3.3	12.94 ± 0.12	9.01 ± 0.12	0.67	0.65
6	38.6 ± 7.0	12.66 ± 0.13	8.61 ± 0.33	0.94	1.29
7	30.3 ± 3.9	12.71 ± 0.09	9.24 ± 0.25	0.68	0.78
8	39.8 ± 4.5	12.81 ± 0.12	8.69 ± 0.19	0.71	0.93
9[a]	21.1 ± 6.3	11.95 ± 0.29	9.36 ± 0.48	0.31	0.04
10[a]	24.3 ± 6.8	11.88 ± 0.35	8.93 ± 0.37	0.59	0.50
11	41.5 ± 4.4	12.82 ± 0.13	8.60 ± 0.16	0.80	0.99
12	43.2 ± 4.7	13.01 ± 0.10	9.69 ± 0.17	0.96	1.40
13	34.8 ± 3.2	12.84 ± 0.10	9.04 ± 0.14	0.73	0.83
14	16.8 ± 1.9	11.38 ± 0.14	9.32 ± 0.19	0.72	0.74
15	30.7 ± 3.2	12.68 ± 0.12	9.18 ± 0.16	0.70	0.75
(17)	28.6 ± 2.3	12.47 ± 0.12	9.14 ± 0.13	0.75	0.85
(19)	16.4 ± 1.7	11.17 ± 0.13	9.17 ± 0.16	0.81	1.02
(20)	28.6 ± 3.5	12.31 ± 0.17	8.98 ± 0.15	0.76	1.02

Note. — Quantities are derived similarly to Table 2.2, except that the radio data are also incorporated with $q_L = 2.14$. The estimate of the dust masses additionally assumes $\kappa_d(850\,\mu\mathrm{m}) = 0.15\,\mathrm{m}^2\,\mathrm{kg}^{-1}$ for the dust absorbtion efficiency.

[a] These SMGs are identified as being radio loud. Therefore, the radio data are not used in constraining the far-infrared SEDs.

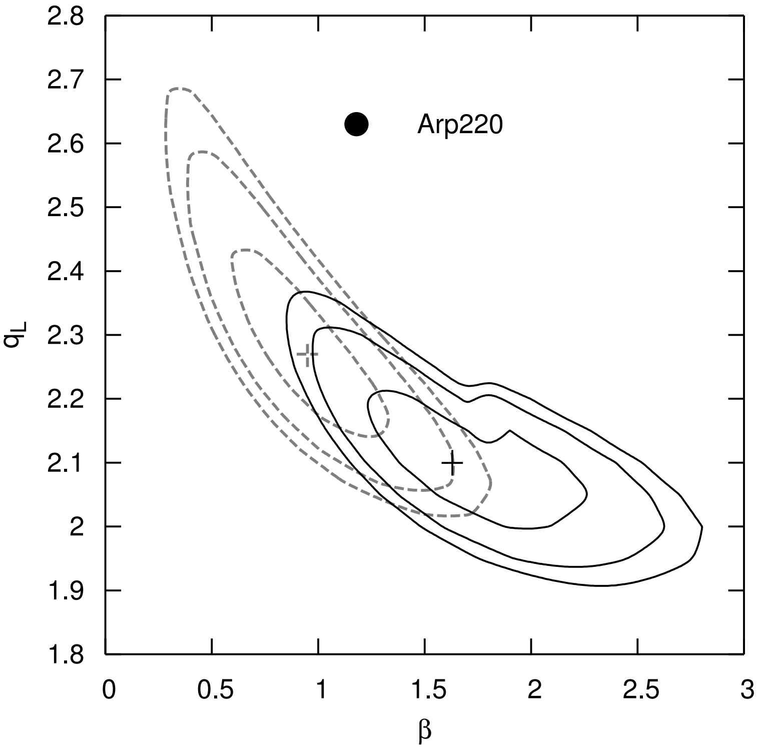


Figure 2.7 Likelihood contours (at 68%, 90% and 95% confidence) for the dust emissivity β and the radio to far-infrared correlation constant q_L, using only the available detections (dashed curves) and all information, including nondetections (solid curves). The best fit locii of both fits are indicated by crosses. The more inclusive set is expected to be a more reliable indicator of SMG properties, as it is less affected by the selection biases that are introduced by requiring detections in all bands. A similar fit to Arp220 is also indicated (black dot).

'normal' q_L, with the best fit located at the respective values of 0.95 and 2.27, whereas the more inclusive data set tends towards emissivity indices in line with expectations, but with q_L decidedly below the mean in the local universe. The best fit values of the extended set are at $\beta = 1.63$ and $q_L = 2.10$.

Accordingly, we assumed $\beta = 1.5$ and used the corresponding $q_L = 2.14$ (from fig. 2.7) to calculate most likely estimates of the far-infrared emission properties, which are listed in Table 2.3.

Systematic biases in the cross-calibration, between the SHARC-2 and the longer wavelength data points, would introduce bias into the estimated emissivity index. Fortunately, with expected 10% relative calibration error, the resulting bias is only $\delta\beta \simeq 0.08$, and our conclusions on the dust emissivity are nearly unaffected.

2.6 Photometric Redshifts

Obtaining accurate redshifts for distant SMGs has relied on optical measurements, guided by radio or optical associations (Chapman et al. 2003, 2005). This approach has been very successful, providing accurate redshifts for nearly a hundred SMGs to date, but it is unlikely to be practical for compiling much larger samples. Moreover, the identifications involving radio or optical counterparts could introduce selection biases into the samples. For both reasons a 'holy grail' of large submillimeter surveys has been to obtain redshifts directly from the radio, sub-mm and mid- to far-infrared (e.g.,*Spitzer*) if possible.

The simplest types of redshift estimators, such as Carilli & Yun (1999, 2000a,b), Wiklind (2003) and Laurent et al. (2006), assume that all SMGs have intrinsically similar rest-frame temperatures. In essence, redshifts are calculated as $(1+z) \approx T_d/T_{obs}$ (Blain 1999) for SEDs characterized by T_{obs} in the observing frame and an assumed representative dust temperature T_d for all objects. The methods only differ on how T_{obs} is estimated from the various available flux data. Only independent, and accurate, measurements of actual dust temperatures for SMGs with known redshifts, can test the validity of this approach. We offer this test for the first time.

A glance at Tables 2.2 and 2.3 reveals that dust temperatures are distributed around $35\,\mathrm{K}$, with some 80% of the data between 25 and 45 K. The same conclusion is suggested by the larger Chapman et al. (2005) sample, with the dust temperatures estimated from the SCUBA $850\,\mu$m and radio data using the appropriate far-infrared to radio correlation as discussed above. This majority population of SMGs is consistent, within measurement uncertainty, with a single representative temperature of $34.6 \pm 1.4\,\mathrm{K}$ in the rest frame, with an estimated 3 K intrinsic dispersion around the mean, lending some credibility to photometric redshifts derived as,

$$(1+z) \approx \frac{34.6 \pm 3.0\,\mathrm{K}}{T_{obs}} \left[\frac{1.5}{\beta}\right]^{0.71},$$

in the redshift the range of $z \sim 1.5 - 3.5$. However, the observations suggest that these photometric redshift indicators could be wholly inappropriate for as many as 20% of all SMGs, unless some of these outliers have incorrect redshifts. Curiously, the very low q_L values measured for sources 14 and 19 hint at 'hot'-type galaxies (Yun et al. 2001), in apparent contradiction with the cool dust temperatures that are implicated at their low redshifts. This mismatch would be reconciled if the temperatures were typical at the photometric z values of $2.96^{+0.51}_{-0.43}$ and $1.95^{+0.33}_{-0.28}$ respectively.

While quasars (Benford et al. 1999; Beelen et al. 2006) too can be characterised by rest-frame SED templates, with $T_d \approx 48 \pm 8\,\mathrm{K}$[8] more or less uniformly across $z \sim 2-7$, they are typically hotter than SMGs, perhaps due to additional dust heating by an AGN. However, quasars, selected from larger volumes than SMGs, could provide the rare extension of the SMG population into ever

---

[8] Temperatures from Beelen et al. (2006) were adjusted for $\beta = 1.5$ for consistency.

more extreme luminosities and hotter dust temperatures. If true, the different temperatures that characterize quasars vs. the bulk of SMGs, invalidates the single representative temperature assumption for all SMGs, already under fire from a minority population of cold SMGs, and therefore may oppose rather than support the applicability of temperature-based photometric redshifts.

Alternatively, the *entire* SMGs sample, *including those low redshift data points*, is better fit by temperature evolution of $T_d \propto (1+z)$, producing similar temperatures in the observed frame ($T_{obs} \simeq 10\,\mathrm{K}$), with an average deviation of $|\hat{\chi}| = 1.40$ vs. $|\hat{\chi}| = 2.67$ produced by the representative rest frame temperature model when assuming a 10% intrinsic scatter in T_d. Hence, the temperature range that fits most SMGs could arise from the similar fractional dispersion of the redshift distribution, i.e., objects selected near the median redshift of 2.3 tend to have temperatures of about $35\,\mathrm{K}$. As such, temperature-based photometric redshifts may not provide additional information beyond a reflection of the median z of the sample.

A similarity of SEDs in the observed frame, yielding a nearly constant T_{obs} and thus $T_d \sim (1+z)$, could arise simply from selection effects that result from the minimum flux requirement for detections (fig. 2.8), and the small dynamic range (of a factor $\simeq 3$) of flux densities that characterize current SMG samples. As such, selection effects alone could render far-infrared and radio based photometric redshift indicators wholly unsuitable for unbiased flux and volume limited samples of SMGs.

2.6.1 Luminosity-Temperature Relation

An alternative, luminosity based approach exploits the hypothetical relationship between dust temperatures and the luminosities that they fuel, to derive photometric redshifts by comparing the rest frame relationship to observed temperature indicators and fluxes. Like other scaling relations that characterize galaxies, the $L-T$ relationship would, presumably, be a genuine property of galaxies, reflecting the physics of galactic structure and dynamics. While the idea has been drafted by Yun & Carilli (2002), Blain et al. (2003) and Aretxaga et al. (2005), the particular implementation of such a method was left awaiting the determination of the actual form of such an $L-T$ relation.

We find, that the far-infrared luminosity is well approximated by a power law of the form,

$$L_{FIR} = \mathcal{L}_0 T_d^{\gamma} \times (1+z)^{\mu}, \tag{2.8}$$

for some exponents γ and μ. The expression incorporates the possibility of explicit redshift evolution, approximated by the term $(1+z)^{\mu}$ chosen for convenience. This independent redshift term is expected to absorb selection effects, and allow the determination of the underlying temperature dependence.

The possible combinations of L_{FIR} and T_d pairs are correlated (fig. 2.8), which adds difficulty

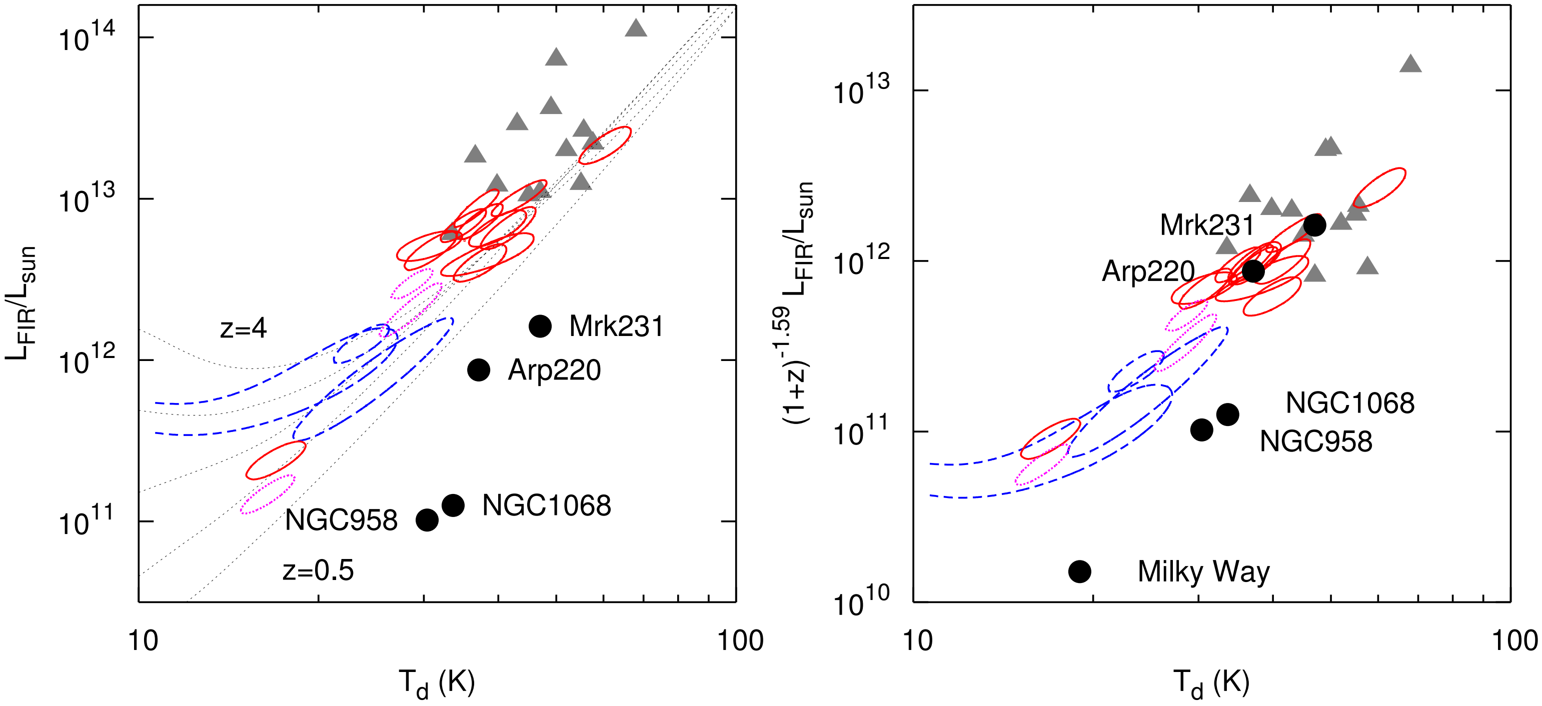


Figure 2.8 Luminosity–temperature (L–T) relation. $1\,\sigma$ likelihood contours for grey-body models fitted to submillimeter data (this paper with solid curves, Laurent et al. (2006) as dotted curves). Contours are also shown for objects not detected by SHARC-2 (dashed). The dusty quasars (triangles) of Benford et al. (1999) and Beelen et al. (2006), and some well-known objects with accurately constrained SEDs are also shown. The plot highlights the difficulty of measuring the underlying temperature dependence of luminosity, as the two parameters are correlated, most data points are heavily clustered and clearly affected by selection effects. When such a relation is established, however, it can dispute the idea of luminosity based photometric redshifts. The top graph shows the apparent L–T relation and the $850\,\mu$m selection functions ($S_{850\,\mu\mathrm{m}} > 5\,\mathrm{mJy}$) at redshifts of 0.5,1,2,3 and 4. (a) The observed relation is evidently biased by selection. (b) A bias- and evolution-free version shows the same relation after the observed redshift evolution (including selection bias) has been accounted for, effectively 'moving' luminosities into the present era. The case of the nearby Arp220 clearly supports the case of explixit redshift evolution of the L–T relationship.

to constraining the underlying relationship. With this warning in mind, our best estimates of the exponents are $\gamma = 2.76 \pm 0.29$ and $\mu = 1.74 \pm 0.38$ for all data points and $\gamma = 2.61 \pm 0.30$ with $\mu = 1.25 \pm 0.48$ excluding the lower redshift points. The corresponding normalizations, $\mathcal{L}_0$, are $3.4 \times 10^7\, L_\odot$ (± 0.33 dex) and $1.1 \times 10^8\, L_\odot$ (± 0.45 dex) respectively. The fits were obtained via a two variable linear regression of the logarithms with error bars (Table 2.3) on both the $\log L$ and $\log T$ values. As most galaxies lie in a narrow temperature and luminosity range, we point out that the derived temperature exponents are constrained by relatively few points, some or all of which could be peculiar, and therefore the true uncertainty of γ and μ are likely underestimated. Nonetheless, these scaling models successfully apply to the quasars of Benford et al. (1999) and Beelen et al. (2006), and the local ULIRGs Arp220 and Mrk231. If we consolidate our estimates with the inclusion of these onjects, we obtain $\gamma = 2.82 \pm 0.29$, $\mu = 1.59 \pm 0.18$ and $\mathcal{L}_0 = 3.6 \times 10^7\, L_\odot$ (± 0.46 dex). This relationship, derived for SMGs, quasars and ULIRGs, could hint at a possible connection among these types of objects, however, it may not apply to the less active galaxies of the local universe (fig. 2.8).

Assuming the exponents and scaling are properly determined, we may attempt to derive a photometric redshift indicator by comparing the expressions for the far-infrared luminosity from equations (2.4) and (2.8) and from the definitions (2.6) and (2.7). In order to facilitate the derivation of analytic solutions, we approximate the luminosity distance by $D_L \approx D_0\,(1+z)^\delta$, the effect of which is to simplify all redshift dependence to powers of $(1+z)$. This simplification is reasonable for the study of SMGs (leading to an rms $\delta z/z$ of $\simeq$ 6%) in the redshift range of $0.5 - 4$ with $D_0 = 1.63\,\mathrm{Gpc}$ and $\delta = 2.07 \pm 0.03$. We find that the far-infrared luminosity is then proportional to

$$S_{1.4\,\mathrm{GHz}}\,(1+z)^{2\delta+1-\alpha} \propto T_{obs}^{\gamma}\,(1+z)^{\gamma+\mu} \propto \frac{S_\nu T_{obs}^{4+\beta}}{e^{h\nu/kT}-1}\,(1+z)^{2\delta}. \tag{2.9}$$

Substituting the quantities assumed or determined so far, we find that all three expressions possess similar dependences on $(1+z)$, with exponents of 4.44 ± 0.12, 4.41 ± 0.34 and 4.14 ± 0.06 respectively. As the exponents are indistinguishable within the measurement uncertainties involved, we conclude that, in the redshift range of $z \sim 0.5$–4 of the approximation, it is very unlikely that a meaningful measure of redshift can be deduced in this way; bad news for luminosity-based photometric-redshift estimation. While it may simply be an unfortunate coincidence, it is more likely that the inability to determine redshifts in this manner is, once again, the product of the sample selection yielding candidates with similar observed SEDs. With these being nearly indistigushable in the observing frame, photometric redshifts of any kind remain elusive.

We can, nevertheless, derive a true luminosity based photometric redshift indicator in the low redshift limit $z \ll 1$, where the luminosity distance takes the asymptotic form $D_L \to cz/H_0$. In this limit,

$$z \approx 1\,\mathrm{Gpc}\,\frac{H_0}{c}\left(\frac{\mathcal{L}_0 T_d^{\gamma}}{L_{1\,\mathrm{Gpc}}}\right)^{1/2},$$

where $L_{1\,\mathrm{Gpc}}$, the luminosity as it were if the galaxy were at a distance of $1\,\mathrm{Gpc}$. While this is irrelevant to the SMG sample, it provides a useful expression for the study of the L–T relation in the nearby ($z \ll 1$) universe, when values of z are already known.

Comparing the L–T relationship (eq. [2.8]) with the expression for the integrated far-infrared luminosity (eq. [2.4]) we infer that $M_d \propto T_d^{-(4+\beta-\gamma)}(1+z)^{\mu}$, i.e., at any given redshift $T_d \propto M_d^{1/(4+\beta-\gamma)}$, that is $T_d \propto M_d^{-0.38\pm0.04}$, which is expected to be free of selection bias. The dust heating is predominantly fueled by the luminosity of massive ($>8\,M_{\odot}$) stars, and so we find that the number of high-mass stars $N_{HM} \propto T_d^{\gamma}$, assuming that the high-mass IMF and stellar content of all SMGs is similar. We may estimate the high-mass star-formation efficiency in SMGs by $\eta_{HM} \propto N_{HM}/M_d \propto M_d^{-\gamma/(4+\beta-\gamma)}$, i.e., $\eta_{HM} \propto M_d^{-1.10\pm0.13}$.

Surprisingly, this implies that more dusty galaxies produce, and thus contain, fewer massive stars relative to dust mass. Perhaps, this is due to more dusty environments favoring lower mass star formation. This conclusion is consistent with the evolution of q from lower values in SMGs to higher values in nearby dusty disk galaxies (Yun et al. 2001), and the dominance of high-mass stars in SMGs giving way to more radio-quiet low-mass stars heating dust in the local universe (Blain et al. 1999b; Baugh et al. 2005).

2.7 Scaling Relations

After cancelling the nearly identical redshift terms appearing in the luminosity scaling relations (eq. [2.9]), we obtain a useful set of expressions interrelating the observed quantities, $S_{1.4\,\mathrm{GHz}}$, S_{ν} and T_{obs} in the far infrared. For distant SMGs, in the redshift range $z \sim 0.5$–4, we find,

$$\begin{aligned} S_{1.4\,\mathrm{GHz}} &\propto T_{obs}^{\gamma}, \\ S_{\nu} &\propto T_{obs}^{\gamma-(4+\beta)}\left(e^{h\nu/kT}-1\right). \end{aligned}$$

The appropriate constants of proportionality are easily derived from the relevant expressions (eqs. [2.4], [2.7], [2.6] and [2.8]) applied at the median redshift of the sample. Figure 2.9 testifies to the predictive power of these relations. Excluding the radio-loud points, which clearly lie away from the far-infrared to radio correlation, the model is consistent with observations within the measurement uncertainties. We emphasize, that these relations are not fitted to the data, but rather a direct consequence of the radio to far-infrared correlation, and the deduced $L-T$ relation.[9] Therefore,

---

[9] A pure L–T relation without the evolution term yields a poor model of the observed relations which therefore directly support the inclusion of the explicit redshift evolution in eq. (2.8).

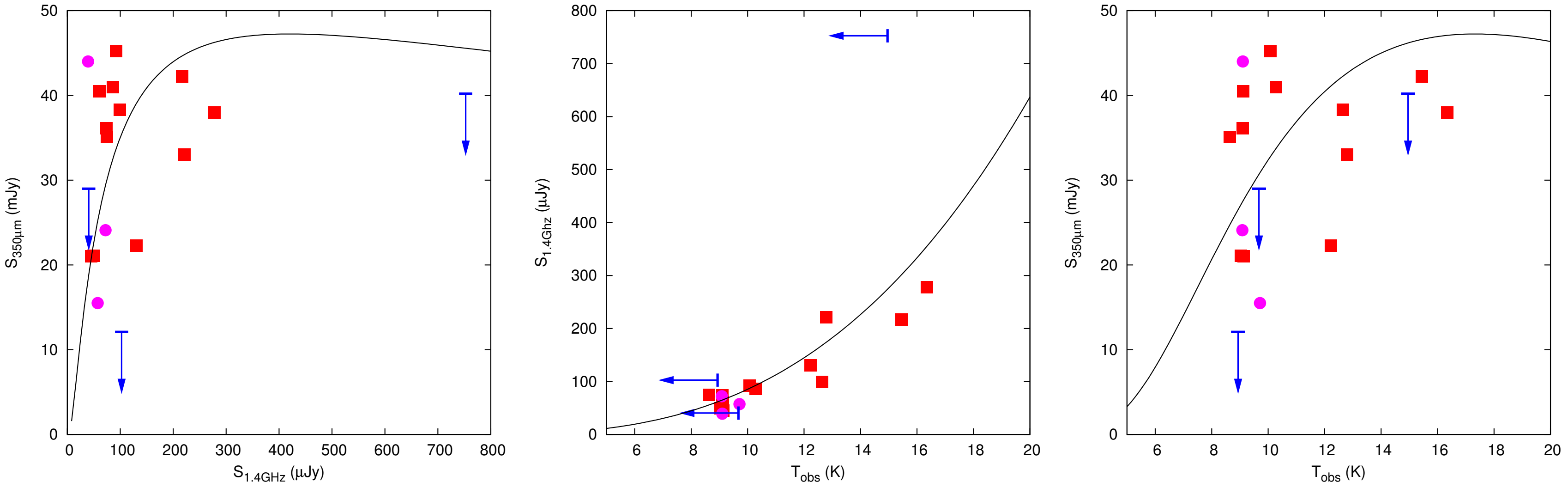


Figure 2.9 Relations between observed quantities. (a) Radio and far-infrared fluxes, (b) radio fluxes, and (c) far-infrared fluxes vs. observed temperatures. Results from this paper are plotted as squares with upper limits indicated as appropriate (arrows), Laurent et al. (2006) data is shown as dots. The model deriving from the radio-FIR correlation and the observed *L*–*T* relation is also shown (solid curve). The data shows excellent agreement with predictions within the measurement uncertainties involved. A single radio or submillimeter measurement may, therefore, suffice to predict the other quantities. As such, these relationship can offer a useful tool for all radio or submm surveys.

the excellent agreement with data is remarkable.

2.8 Conclusions

New $350\,\mu$m data leads to the first direct measures of the characteristic dust temperatures and far-infrared luminosities of SMGs, and the following valuable conclusions.

1. The linear radio to far-infared correlation remains valid out to the redshifts of $z \sim 1 - 3$ for SMGs, with the exception of rare radio-loud objects. The power-law index in the correlation is 1.02 ± 0.12, and tighter than observed locally, with an intrinsic dispersion of only 0.12 dex.
2. Either the far–infrared–radio correlation constant q is lower than locally measured ($q_L \approx 2.14 \pm 0.07$), or the effective dust emissivity index is less, with $\beta \to 1$ consistent with observations. The lower q value indicates dust heating dominated by high-mass stars, and that any AGN contribution is minor in comparison.
3. Compared with low-redshift galaxies, SMGs are characterized either by low gas-to-dust ratios, around 54^{+14}_{-11} $[\kappa_{850\,\mu\mathrm{m}}/0.15\,\mathrm{m}^2\,\mathrm{kg}^{-1}]$, indicating dust-rich environments, or by efficient dust absorption of $\kappa_{850\,\mu\mathrm{m}} \gtrsim 0.33\,\mathrm{m}^2\,\mathrm{kg}^{-1}$.
4. Far-infrared and radio based photometric redshifts might be appropriate for up to 80% of SMGs in the redshift range of of $z \sim 1.5 - 3.5$, with a rest frame temperature assumption of $34.6 \pm 3.0\,\mathrm{K}\,[1.5/\beta]^{0.71}$. However, photometric redshift indicators may not be appropriate for an unbiased SMG sample, as existing selection effects tend to favor measuring SEDs that are similar in the observed frame.
5. We deduce an $L - T$ relationship which we argue to be a property of star-forming galaxies. When recast as mass scaling, we find, that ISM heating scales inversely with the quantity of dust in the galaxies, i.e., the galaxies with more dust are less active, possibly due to the formation of more low-mass stars being favoured in dustier environments.
6. From the observed $L - T$ relation, possibly biased by selection, and the radio to far-infrared correlation, we derive scaling relations among the observed quantities $S_{1.4\,\mathrm{GHz}}$, S_ν in the submillimeter or far-infrared, and the observing frame dust temperature T_{obs}, applicable to the redshift range of $z \sim 0.5 - 4$. A determination of one of these quantities may sufficiently characterise the observed far-infrared properties of SMGs.

Bibliography

Agladze, N. I., Sievers, A. J., Jones, S. A., Burlitch, J. M., & Beckwith, S. V. W. 1996, ApJ, 462, 1026

Alexander, D. M., Bauer, F. E., Chapman, S. C., Smail, I., Blain, A. W., Brandt, W. N., & Ivison, R. J. 2005a, ApJ, 632, 736

Alexander, D. M., Smail, I., Bauer, F. E., Chapman, S. C., Blain, A. W., Brandt, W. N., & Ivison, R. J. 2005b, Nature, 434, 738

Almaini, O., Dunlop, J. S., Conselice, C. J., Targett, T. A., & Mclure, R. J. 2006, submitted to MNRAS

Appleton, P. N., et al. 2004, ApJs, 154, 147

Aretxaga, I., Hughes, D. H., & Dunlop, J. S. 2005, MNRAS, 358, 1240

Barger A. J. Cowie, L. L., Sanders, D. B., Fulton, E., Taniguchi, Y., Sato, Y., Kawara, K., & Okuda, H. 1998, Nature, 394, 248

Baugh, C. M., Lacey, C. G., Frenk, C. S., Granato, G. L., Silva, L., Bressan, A., Benson, A. J., & Cole, S. 2005, MNRAS, 356, 1191

Beelen, A., Cox, P., Benford, D. J., Dowell, C. D., Kovács, A., Bertoldi, F., & Omont, A. 2006, ApJ, in press

Benford, D. J., Cox, P., Omont, A., Phillips, T. G., & McMahon, R. G. 1999, ApJ, 518, L65

Bigss, A. D., Ivison, R. J. 2006, submitted to MNRAS

Blain, A. W. 1999, MNRAS, 309, 955

Blain, A. W., Barnard, V., & Chapman, S. C. 2003, MNRAS, 338, 733

Blain, A. W., Chapman, S. C., Smail, I., & Ivison, R. J. 2004, ApJ, 611, 52

Blain, A. W., Möller, O., & Maller, A. M. 1999, MNRAS, 303, 423

Blain, A. W., Smail, I., Ivison, R. J., & Kneib J.-P. 1999, MNRAS, 302, 632

Borys, C., Chapman, S., Halpern, M., & Scott, D. 2003, MNRAS, 344, 358

Borys C., Smail, I., Chapman, S. C., Blain, A. W., Alexander, D. M., & Ivison, R. J. 2006, ApJ, 635, 853

Borys, C., et al. 2005, ApJ, 636, 134

Carilli, C. L., & Yun, M. S. 1999, ApJ, 513, L13

Carilli, C. L., & Yun, M. S. 2000a, ApJ, 530, 618

Carilli, C. L., & Yun, M. S. 2000b, ApJ, 539, 1024

Chapman, S. C., Blain, A. W., Ivison, R. J., & Smail, I. 2003, Nature, 422, 695

Chapman, S. C., Blain, A. W. Smail, I., & Ivison, R. J. 2005, ApJ, 622, 772

Chapman, S. C., Smail, I., Blain, A. W., & Ivison, R. J. 2004, ApJ, 614, 671

Chapman, S. C., Smail, I., Ivison, R. J., & Blain, A. W. 2002, MNRAS, 335, L17

Condon, J. J. 1992, ARA&A, 30, 575

Condon, J. J., Anderson, M. L., & Helou, G. 1991, ApJ, 376, 95

Condon, J. J., & Broderick, J. J. 1991, AJ, 102, 1663

Coppin, K., Halpern, M., Scott, D., Borys, C., & Chapman, S. 2005, MNRAS, 357, 1022

Cox, M. J., Eales, S. A. E., Alexander, P., & Fitt, A. J. 1988, MNRAS, 235, 1227

De Breuck, C., et al. 2003, A&A, 401, 911

Devereux N. A., & Eales S. A. 1989, ApJ, 340, 708

Dowell, C. D., et al. 2003, SPIE Conf. Proc. Vol. 4855, Millimeter and Submillimeter Detectors for Astronomy, ed. T. G. Phillips, & J. Zmuidzinas, 73

Downes, D., & Solomon, P. M. 1998, ApJ, 507, 615

Dunne, L., & Eales, S. A. 2001, MNRAS, 327, 697

Dunne, L., Eales, S. A., & Edmunds, M. G. 2003, MNRAS, 341, 589

Dupac, X., et al. 2003, A&A, 404, L11

Eales, S., Lilly, S., Gear, W., Dunne, L., Bond, R. J., Hammer, F., Le Fèvre, O., & Crampton, D. 1999, ApJ, 515, 518

Fitt, A. J., Alexander, P., & Cox, M. J. 1988, MNRAS, 233, 907

Garrett, M. A. 2002, A&A, 384, L19

Greve, T. R., Ivison, R. J., Bertoldi, F., Stevens, J. A., Dunlop, J. S., Lutz, D., & Carilli, C. L. 2004, MNRAS, 354, 779

Greve, T. R., et al. 2005, MNRAS, 359, 1165

Helou, G., Khan, I. R., Malek, L., & Boehmer, L. 1988, ApJs, 68, 151

Helou, G., Soifer, T., & Rowan-Robinson, M. 1985, ApJ, 298, L7

Hughes, D. H., et al. 1998, Nature, 394, 241

Ivison, R. J., et al. 2002, MNRAS, 337, 1

Laurent, G. T., Glenn, J., Egami, E., Rieke, G. H., Ivison, R. J., Yun, M. S., Aguirre, J. E., & Maloney, P. R. 2006, ApJ, 643, in press

Laurent, G. T., et al. 2005, ApJ, 623, 742

Leong, M. M. 2005, URSI Conf. Sec. J3-10, 426, http://astro.uchicago.edu/ursi-comm-J/ursi2005/rf-telescope-fabrication/leong

Press W. H., Flannery B. P., & Teukolsky, S. A. 1986, Numerical Recipes. The Art of Scientific Computing (Cambridge: University Press)

Seaquist, E., Yao, L., Dunne, L., & Cameron, H. 2004, MNRAS, 349, 1428

Smail, I., Ivison, R. J., & Blain, A. W. 1997, ApJ, 490, L5

Smail, I., Ivison, R. J., Owen, F. N., Blain, A. W., & Kneib, J.-P. 2000, ApJ, 528, 612

Solomon, P. M., Downes, D., Radford, S. J. E., & Barrett, J. W. 1997, ApJ, 478, 144

Stevens, J. A., Amure, M., & Gear, W. K. 2005, MNRAS, 357, 361

Tacconi, L. J., et al. 2006, ApJ, in press

Webb, T. M. A., Lilly, S. J., Clements, D. L., Eales, S., Yun, M., Brodwin, M., Dunne, L., & Gear, W. K. 2003, ApJ, 597, 680

Wiklind, T. 2003, ApJ, 588, 736

Yun, M. S., & Carilli, C. L. 2002, ApJ, 568, 88

Yun, M. S., Reddy, N. A., & Condon, J. J. 2001, ApJ, 554, 803

Chapter 3

Fundamentals of Weak-Signal Observing and Analysis for Non-Gaussian Noise

3.1 Introduction

Noise is something the experimental scientist will always have to live with, at least to a degree. It is the understanding of how noise limits a particular measurement that is critical to producing optimal results.

This is most important when structured (i.e., colored or nonstationary) noise dominates the experiment, and signal-to-noise ratios of the individual measurements (i.e., digitized samples) are extremely low ($S/N \ll 1$). Unfortunately, many familiar techniques of data analysis (e.g., Press et al. 1986) and data collection strategies presume appreciable signal-to-noise ratios ($S/N \gtrsim 1$) in the measurements, and can fail spectacularly when applied to time-series signals, which are buried deep beneath a complex structure of noise. Therefore, we aim to provide a comprehensive framework for the optimal treatment of data with low ($S/N < 1$) and extremely low ($S/N \ll 1$) signal-to-noise conditions. As such, the discussion is directly relevant to the subject of ground-based submillimeter astronomy, which is often characterized by signal-to-noise ratios as low as 10^{-4} in individual time samples (taken at $36\,$ms intervals for SHARC-2)[1] due to the overwhelming nature of a bright and highly variable atmosphere at these wavelengths. While submillimeter astronomy provides the motivation for this discussion, the realm of possible applications goes beyond, with implications for many weak-signal experiments. Therefore, the discussion presented henceforth is offered in full generality (nonetheless, often using examples from submillimeter astronomy as illustration), while the specifics of astronomical application are discussed separately (Appendix B).

When dealing with weak signals, special attention must be afforded not only to analysis tech-

---

[1] The second generation Submillimeter High Angular Resolution Camera (Dowell et al. 2003).

niques but also to data collection strategies. These go hand-in-hand to provide the critical observational results sought under specific noise conditions. Whereas every measurement, taken at any time, in any configuration, is equally valuable with white noise, that is no longer assured when excess noise, such as narrow band noise, $1/f^{\alpha}$ type noise, or transient noise, is additionally present in an experiment. Strategies that distribute repeated measurements along temporal, spatial or other configuration coordinates (e.g., frequency), will provide improved isolation of the critical signals from excess noise interference (c.f. Emerson 1995; Tegmark 1997; Comito & Schilke 2002; Dupac & Tauber 2005). Thus, some of optimization of observing strategies is evidently necessary for constructing data streams that are ideally suited for the analysis methods presented.

3.2 Signals, Measurements, and Observing Strategies

Experiments are designed to probe underlying source structures S_u, which is projected into the space spanned by the detectors. This space can be parametrized by some appropriate coordinate vectors $\vec{\xi}$ (e.g., a projected flux distribution on the sky, parametrized by its right-ascension α and declination δ). Therefore, the underlying source can be considered a function $S_u(\vec{\xi})$, which may also be represented, or as approximated, by a finite number of discrete model parameters S_i (e.g., pixels in a model map of the emission).

The underlying source signal $S_u(\vec{\xi})$ is observed through detectors which occupy a volume in the sampled space (e.g., a detector pixel area projected to the sky, an instrument beam or the bandwidth of a spectrometer channel). Thus, measurements by the detector will represent the integrated source signals inside that detector volume. In other words, the structure $S_{obs}(\vec{\xi})$ is related to $S_u(\vec{\xi})$ via a convolution (eq. [3.14]) with the detector response $R(\vec{\xi})$ as,

$$S_{obs}(\vec{\xi}) = S_u(\vec{\xi}) \otimes R(\vec{\xi}). \tag{3.1}$$

The underlying signal often has a naturally limiting resolution (such as a telescope beam), and if the width of the detector response is sufficiently small in comparison ($R(\vec{\xi}) \to \delta(\vec{\xi})$) the above simplifies to $S_{obs}(\vec{\xi}) \to S_u(\vec{\xi})$, i.e., the measured source structure will closely resemble the underlying structure of the source. Otherwise, the underlying structures may be estimated through an appropriate deconvolution (e.g., Clark 1980; Narayan & Nityananda 1986), although this may not be well defined or trivial, especially in the presence of noise.

The typical experimental apparatus is equipped with a number of detectors, distributed along physical coordinates $\vec{x}$. Each detector produces a continuous signal, which is sampled and digitized to produce the discrete time-series data. Accordingly, the data are parametrized either by a coordinate variable in the configuration space of the experiment which includes the time coordinate t

(we shall use $\mathbf{x} = \{t, \vec{x}\}$ from here on), or by discrete indeces that identify a measurement time and a detector channel for every sample in the time-series data (t and c, or simply a measurement serial number k henceforth).

$$D_k \equiv D(\mathbf{x}_k). \tag{3.2}$$

Therefore, D may be used without parametrization to refer to both the signals $D(\mathbf{x})$ and the discrete measurements D_k.

The detector signals S, corresponding to the observed source structures $S_{obs}(\vec{\xi})$ are, unfortunately, corrupted by the presence of noise N in the experiment, yielding observed signals:

$$D = S + N. \tag{3.3}$$

In an effort to reduce measurement uncertainties, the source may be observed repeatedly, often through altered experimental configurations (e.g., by detectors scanning across a flux distribution) to produce a time series of source signals $S(\mathbf{x})$. $S(\mathbf{x})$ is generated from the source structure by the mapping function $\mathcal{M}$, as $S(\mathbf{x}) = \mathcal{M}(S_{obs}(\vec{\xi}))$, or in the alternative notation

$$S(\mathbf{x}) = \mathcal{M} \circ S_{obs}(\vec{\xi}). \tag{3.4}$$

Here $\mathcal{M}$ fully encapsulates the observing strategy (see Section 3.6).

3.2.1 A Data Reduction Challenge

For a noiseless signal, the observed source structure $S_{obs}(\vec{\xi})$ can be recreated from the time-series data $S(\mathbf{x})$, within boundaries of data coverage, by the an appropriate 'inversion' of eq. (3.4). Thus for a mathematically perfect signal without noise, one can express the source structures $S_{obs}(\vec{\xi})$ from the time-series source signal $S(\mathbf{x})$ via the reverse mapping function $\mathcal{A}$ (as in 'analysis') of the signals as $S_{obs}(\vec{\xi}) = \mathcal{A}(S(\mathbf{x}))$, or

$$S_{obs}(\vec{\xi}) = \mathcal{A} \circ S(\mathbf{x}). \tag{3.5}$$

For the reverse mapping to reproduce the same structure $S_{obs}(\vec{\xi})$ from the noiseless time-series, $\mathcal{A}$ applied on top of $\mathcal{M}$ must produce the identity map $\mathbf{1}$,

$$\mathcal{A} \circ \mathcal{M} = \mathbf{1}. \tag{3.6}$$

The trouble is that $\mathcal{A}$, is generally not determined uniquely by the above, especially when source components (e.g., map pixel) are mapped multiple times into the time stream in the effort to integrate down on noise. In other words, the dimension of the data vector is typically much larger then the number of discrete source components sought from them (i.e., $\dim \mathbf{S} \gg \dim \mathbf{S}_{obs}$, hence it is a data *reduction*). The inversion problem, overdetermined by the data, yields a family of reverse mapping solutions, all of which are equally plausible for a noiseless mathematical signals. In the presence of noise, however, some solutions are more accurate, and thus the challenge is to identify which one this may be.

Clearly, the simple reverse mapping of the noisy signals (eq. [3.3]),

$$\mathcal{A} \circ D(\mathbf{x}) \longrightarrow S_{obs}(\vec{\xi}) + \mathcal{A} \circ N(\mathbf{x}), \tag{3.7}$$

produces a model of the source that is corrupted by the noise re-mapped from the time stream as $N_{obs}(\vec{\xi}) = \mathcal{A} \circ N(\mathbf{x})$. There are established ways for finding suitable reverse mappings $\mathcal{A}$ (especially if the $\mathcal{M}$ is linear) that provide critical extrema to the inversion problem (e.g., maximum-likelihood or maximum-entropy estimates). When these satisfy eq. (3.6) exactly, the solution is lossless, since the source structure are recovered completely (along with the re-mapped noise, of course). However, for weak signals this may not be good enough.

When the level of the noise pollution is comparable or greater than the expected level of source components, $\langle |N_{obs}(\vec{\xi}_i)| \rangle \gtrsim \mathbf{E}(|S_{obs}(\vec{\xi}_i)|)$, the observer is forced to explore ways of reducing noise further such that the scientific objective of the experiment can be met. This may be achieved, in part, by improving the observing patterns $\mathcal{M}$ to such that $\mathcal{A} \circ N(\mathbf{x})$ is minimized. Additionally, instead of seeking exact (i.e., lossless) solutions to the inversion problem (eq. [3.5]), one may try find approximate answers, which recover the source structures *mostly*, alas not perfectly. By allowing for an imperfect (i.e., lossy) source reconstructions, the noise pollution of the estimate may be substantially reduced.

This is the essence of filtering. While the filtering may be absorbed completely into the reverse mapping ($\mathcal{A} \rightarrow \mathcal{A}'$), it generally convenient to separate it explicitly, thus leaving $\mathcal{A}$ in its lossless form. Then, $\mathcal{A}$ then operates on a filtered time stream $S'(\mathbf{x}) = \mathcal{F}(S(\mathbf{x}))$ (e.g., Tegmark 1997). Further filtering is possible on the resulting source estimate $\widehat{S}_{filtered}(\vec{\xi})$ also i.e., $\widehat{S}'_{filtered}(\vec{\xi}) = \mathcal{F}'(\widehat{S}_{obs}(\vec{\xi}))$. With filtering both prior and after inversion, the lossy estimate can be expressed most generally as,

$$\mathcal{F}' \circ \mathcal{A} \circ \mathcal{F} \circ D(\mathbf{x}) \longrightarrow \widehat{S}'_{filtered}(\vec{\xi}) + \mathcal{F}' \circ \mathcal{A} \circ \mathcal{F} \circ N(\mathbf{x}). \tag{3.8}$$

When optimized, the noise thus remapped, i.e.,

$$N_{filtered}(\vec{\xi}) = \mathcal{F}' \circ \mathcal{A} \circ \mathcal{F} \circ N(\mathbf{x}), \tag{3.9}$$

can be substantially reduced compared to that in eq. (3.5), while the reconstituted filtered source structures are adequately preserved at the same time, i.e., $\widehat{S}'_{filtered}(\vec{\xi}) \approx S_{obs}(\vec{\xi})$. Alternatively, the lossy inversion $\mathcal{A}'$ is composed of the lossless (exact) inversion $\mathcal{A}$ and the filters $\mathcal{F}$ and $\mathcal{F}'$ as,

$$\mathcal{A}' = \mathcal{F}' \circ \mathcal{A} \circ \mathcal{F}. \tag{3.10}$$

Therefore, the challenge of weak signal data reduction can be summarized: to recover source structures $S_{obs}(\vec{\xi})$, or even better, the underlying structure $S_u(\vec{\xi})$, as faithfully as possible, *and* with as little noise pollution as possible, given the noisy data $D(\mathbf{x})$.

3.2.2 Representations of Data

Time-series data may be transformed into other representations via invertible transformations. Such alternative representations of the time series are often convenient and insightful and may highlight different aspects of the data set that manifest in the time stream alone. Equivalent representations of data must all contain an equal number of independent components, which is the number of independent measurements in the time series.

A particularly useful set of transformations for the subject of signal processing and data analysis is that provided by Fourier transformations (e.g., Press et al. 1986), defined for complex functions of coordinate vectors by the transform pairs,

$$\begin{aligned} \widetilde{A}(\mathbf{f}) &\propto \int e^{-2\pi i \mathbf{f}\cdot\mathbf{x}} A(\mathbf{x})\, d\mathbf{x}, \\ A(\mathbf{x}) &\propto \int e^{+2\pi i \mathbf{f}\cdot\mathbf{x}} \widetilde{A}(\mathbf{f})\, d\mathbf{f}. \end{aligned} \tag{3.11}$$

Since the integrals are over the volume of space, $d\mathbf{x}$ and $d\mathbf{f}$ here (and throughout the discussion) denote volume elements (not line segments!). In a convenient shorthand notation we shall mark Fourier transformed quantities with a tilde on top (e.g., $\widetilde{A}$) and use thick arrows to indicate transform pairs as,

$$A(\mathbf{x}) \Longleftrightarrow \widetilde{A}(\mathbf{f}). \tag{3.12}$$

The transformations can also be expressed as summations over a discrete set of coordinates and frequencies in an equispaced grid of samples, providing analogous Discrete Fourier Transforms (DFTs, see Press et al. 1986) or Fast Fourier Transforms (FFTs, see Cooley & Tukey 1965; Frigo & Johnson 2005), which are better suited for digitized data. The forward and backward transforms are essentially symmetric, apart from a phase inversion (i.e., complex conjugate transform) between the directions, and therefore the direction of the transform is generally irrelevant when considering the transformation properties.

The principal importance of Fourier transforms in the field of signal processing is because signals are often generated in dynamical systems with sinusoidal solutions. Fourier representations also offer a convenient way of describing most forms of stationary noise (see Section 3.3.3) that commonly distort the physical measurement process.

There are various conventions for the normalizations of the transforms (Press et al. 1986), which is why this was left implicit in expressing these only as proportionalities. However, for the case of signal processing, a a noteworthy choice is when the transformed function $\widetilde{A}(\mathbf{f})$ is used for estimating the power spectrum density (PSD) of A as $|\widetilde{A}(\mathbf{f})|^2 \to P(\mathbf{f})$. In a related practical normalization for discrete transforms, the frequency components $\widetilde{A}_f$ are the effective amplitudes for the sinusoids into which the time-series data is decomposed.

The spectral parameter $\mathbf{f}$ is a frequency vector whose components carry units of $[f_j] = 1/[x_j]$. While the above transform represents transformation along all configuration coordinates, one may selectively transform the data for subspaces (parametrized by $\mathbf{x}_T$) also while leaving the conjugate subspace (parametrized by $\mathbf{x}_\perp$) in the untransformed representation. Then, transforming the time series $A(\mathbf{x}_T, \mathbf{x}_\perp)$ into $\widetilde{A}(\mathbf{f}_T, \mathbf{x}_\perp)$ is done by substituting $\mathbf{x}_T$ and $\mathbf{f}_T$ in eqs. [3.11] in place of $\mathbf{x}$ and $\mathbf{f}$ respectively.

Fourier transforms are accompanied by a number of handy theorems. Of these, we will make much use of the convolution theorem (Press et al. 1986),

$$A \cdot B \Longleftrightarrow \widetilde{A} \otimes \widetilde{B}, \tag{3.13}$$

where the convolution $A \otimes B$ is defined in integral form as,

$$A \otimes B = \int A(\mathbf{x}')B(\mathbf{x} - \mathbf{x}')\, d\mathbf{x}', \tag{3.14}$$

a corollary of which is that the autocorrelation of $A(\mathbf{x})$ transforms into the power spectrum $P(\mathbf{f})$, i.e.,

$$R(\mathbf{x}) = A(\mathbf{x}) \otimes A^*(-\mathbf{x}) \quad \Longleftrightarrow \quad P(\mathbf{f}) = |\widetilde{A}(\mathbf{f})|^2. \tag{3.15}$$

Parseval's theorem (Press et al. 1986) is equally pivotal, relating the integrated signal powers in the configuration and Fourier domains as,

$$\int |A(\mathbf{x})|^2\, d\mathbf{x} \propto \int |A(\mathbf{f})|^2\, d\mathbf{f} \tag{3.16}$$

or,

$$\sum_k |A_k|^2 \propto \sum_f |A_f|^2, \tag{3.17}$$

both in terms of the integrated functions, or as sums over the discrete time-series and spectral components. The constants of proportionality depend on the particular normalization, and on the transforming subspace. Let us note, that the term ′signal power′ always denotes $|A|^2$, independent of what physical units the signal A might have. As such, the signal power may, or may not, be related to the physical power carried by the signals.

Additionally, Fourier transforms of signals are characterized by an uncertainty principle i.e., an inequality relating the localization of signals in the two domains. This may expressed in terms of the the signal spreads (standard deviations) σ_x and σ_f calculated according to,

$$\sigma_x^2 = \left\langle (x - \langle x \rangle)^2 \right\rangle , \tag{3.18}$$

as if the $|A(x)|$ were a probability density function (after the appropriate normalization) along the chosen x direction in the configuration and spectral domains respectively. More explicitly the signal spread can be calculated in terms of the amplitudes as,

$$\sigma_x^2(A) = \frac{\int x^2 |A(\mathbf{x})| \, dx}{\int |A(\mathbf{x})| \, dx} - \left(\frac{\int x \, |A(\mathbf{x})| \, dx}{\int |A(\mathbf{x})| \, dx} \right)^2 . \tag{3.19}$$

Equivalently, the same may be expressed for discrete data A_i with appropriate sums replacing the integrals. The uncertainty principle then requires that,

$$\sigma_x \, \sigma_f \geq \frac{1}{2\pi}, \tag{3.20}$$

with equality holding for structures with Gaussian shapes only. Alternatively, one can express the uncertainty principle in terms of the full-width half-maxima (FWHMs) Δx and Δf along the chosen direction in both domains, by observing that the $\mathrm{FWHM} \approx 2.35\,\sigma$ for Gaussians and tend to be larger for arbitrary shapes. Thus arriving at an approximate inequality between the configuration and spectral signal full-width half-maxima as

$$\Delta x \, \Delta f \gtrsim 1. \tag{3.21}$$

Note, that the uncertainty principle of quantum mechanics is but a manifestation of the more general uncertainty principle relating to Fourier transforms. The wave functions expressed in terms of conjugate pairs of variables (such as x and p_x or t and E) are Fourier transform pairs in the Schrödinger representation (e.g., $\psi(\mathbf{x})$ and $\phi(\mathbf{p})$ are the Fourier transforms of each another).

3.2.3 Fourier Transforms of Noise

Fourier transforms provide a practical representation of noise. When one considers the Fourier transform of an independent (i.e., uncorrelated) complex stationary random variable $X(\mathbf{x})$, with a projection into an independent real stationary random variable $X_r = \mathrm{Re}(X)$,

$$\widetilde{X}(\mathbf{f}) \propto \int e^{-2\pi i \mathbf{f}\cdot\mathbf{x}} X(\mathbf{x})\, d\mathbf{x} = \int X'_f(\mathbf{x})\, d\mathbf{x}, \tag{3.22}$$

it becomes clear that the transform is effectively an integral over another independent complex random variable X'_f with identical properties to X, because the transforming exponential term is just a phase rotation. When expressed this way, the integral in terms of X'_f no longer has an explicit frequency dependence, other than each frequency corresponding to an integral of a different but identical looking random variable. Therefore, one concludes that the transforms of independent noise should have no direct functional dependence on **f**, other than random fluctuations around the zero mean. The transforms of independent stationary noise yield a flat (invariant) power spectrum with $\langle |\widetilde{X}(\mathbf{f})|^2 \rangle = \sigma_f^2$, with independent (uncorrelated) frequency components. The simplest way to summarize this is to state that independent white noise transforms into independent white noise.

Noise that is independent but not stationary can be composed from white noise n_w via multiplication by a profile $\pi(\mathbf{x})$, which according to the convolution theorem (eq. [3.13]) will transform as,

$$\pi(\mathbf{x}) n_w(\mathbf{x}) \Longleftrightarrow \widetilde{\pi}(\mathbf{f}) \otimes n'_w(\mathbf{f}). \tag{3.23}$$

It follows that independent noise has a flat (invariant) noise spectrum, but with correlated components (via $\widetilde{\pi}$). A corollary of this, by reversal of the configuration and spectral coordinates, is that noise with a spectral profile but independent spectral components will result in stationary noise with correlated samples. In summary,

$$\text{invariant} \Longleftrightarrow \text{independent}.$$

For noise, the term 'stationary' is used more often to describe noise that is translation invariant (i.e. uniform) in time, and more generally along other configuration space parametrizations, while 'transient' provides the conjugate to that to describe noise whose properties vary in configuration space. Therefore, considering separately the forward and backward transforms and the negated noise transform relations, we can explicitly spell out how the various noise properties in configuration space transform into spectral properties, and vice versa, as,

$$\begin{aligned} \text{stationary} &\Longleftrightarrow \text{independent} \\ \text{transient} &\Longleftrightarrow \text{correlated} \\ \text{independent} &\Longleftrightarrow \text{uniform} \\ \text{correlated} &\Longleftrightarrow \text{structured.} \end{aligned}$$

These pairs of (configuration and spectral) properties, which characterize the Fourier transformations of noise, are useful connections to bear in mind.

3.3 Noise Weighting

Due to the nature of the data reduction problem, the inverse mapping $\mathcal{A}$ (in eqs. [3.5] and [3.8]) must combine several measurements D_k for coming up with every single source structure component $S_i = S_{obs}(\vec{\xi}_i)$. The combining of data may be done brute force via global optimization, e.g., by maximizing the likelihood of the estimate provided by $\mathcal{A}$. However, such maximum-likelihood estimates are usually derived from the minimization of χ^2 (eq. [3.36]) and are not applicable to colored noise (see Section 3.3.2). Such brute force approaches also fail to highlight the more general statistical principles which guide how repeated measurements ought to be recombined into single estimates (e.g., S_i). The linear inversion problem using the normal-equations approach (see Press et al. 1986) does, in fact, produce weighted averages, which are approached from a more general perspective below. Identifying proper noise weights, either explicitly, or implicitly though global optimization, is a natural step in obtaining optimal reverse maps for source reconstructions.

Measurements can be combined to provide the statistical descriptions of data. When a quantity A is expressed from the measurements D_k i.e., $A = f(\mathbf{D})$, the effective noise level of A can be calculated according to the rules of error propagation,

$$\sigma_A^2 = (\partial_{\mathbf{D}} A)^T \cdot \mathbf{C}_D \cdot (\partial_{\mathbf{D}} A), \tag{3.24}$$

where $\partial_{\mathbf{D}} A$ is the vector composed of the partial derivatives of $\partial A / \partial D_k$, and $\mathbf{C}$ is the covariance matrix,

$$[C_D]_{k,l} = \mathbf{E}\left[(D_k - \langle D \rangle)(D_l - \langle D \rangle)\right], \tag{3.25}$$

whose entries are the expected covariance of the zero mean measurements.

Consider then, forming weighted averages from a set of multiply observations of a quantity X,

$$\langle X \rangle = \frac{\sum_i w_i X_i}{\sum_i w_i}. \tag{3.26}$$

Here, X_i are point estimates of X, i.e., $\mathbf{E}(X_i) = X$, and may be composed of the raw data D_k by some functions $X_i = f_i(\mathbf{D})$. Thus, the normalization of the weighted average by $\sum_i X_i$ is necessary to force the expectation value of the weighted average $\mathbf{E}(\langle X \rangle)$ to be the underlying value X. According to the rules of error propagation, the error in $\langle X \rangle$ is then calculated as,

$$\sigma_{\langle X \rangle}^2 = \frac{\mathbf{w}^T \cdot \mathbf{C}_X \cdot \mathbf{w}}{\left(\sum_i w_i\right)^2} = \frac{\sum_i \sum_j w_i w_j [C_X]_{i,j}}{\left(\sum_i w_i\right)^2}. \tag{3.27}$$

With X_i deriving from D_k, the covariance matrix $\mathbf{C}_X$ is expressed from $\mathbf{C}_D$ by application of the chain rule,

$$C_X = (\partial_I \mathbf{X})^T \cdot \mathbf{C}_D \cdot (\partial_I \mathbf{X}). \tag{3.28}$$

An optimal set of weights can be derived for weighted means, by requiring that the resulting uncertainty of $\langle X \rangle$ is minimized such that $\partial \sigma^2_{\langle X \rangle} / \partial w_i = 0$. The minimization condition yields a matrix equation, which can be expressed in component notation as the set of equations

$$\sum_j w_j [C_X]_{i,j} = \Lambda^2, \tag{3.29}$$

for every value of i, with $\Lambda = \sigma_{\langle X \rangle} \sum_i w_i$ being a constant of choice that can be adjusted to arbitrary positive values by the appropriate scaling of weights. Thus, optimal weights for X_i are provided from the inverse covariance matrix C^{-1} as,

$$w_i \propto \sum_j [C_X^{-1}]_{i,j}. \tag{3.30}$$

A practical set of noise weights are obtained by converting this to an equality.

3.3.1 Weighted Fourier Transforms

Consider the Fourier transform pairs of eq. (3.11) as an application for noise weights. The transform pairs resemble weighted means (esp. when considering discrete transforms) with $\exp(-2\pi i \, \mathbf{x} \cdot \mathbf{f}) \, A(\mathbf{x})$ taking the place of X in eq. (3.26) for a transforming quantity $A(\mathbf{x})$. Therefore, the spectral density estimates, which are normally provided by the direct Fourier transforms under uniform noise, ought to be replaced by the weighted transform of the form,

$$\widetilde{A}_w(\mathbf{f}) \propto \int w_A(\mathbf{x}) e^{-2\pi i \mathbf{f} \cdot \mathbf{x}} A(\mathbf{x}) \, d\mathbf{x}, \tag{3.31}$$

for providing the most accurate (also most likely when noise is independent, see Section 3.3.2) PSD estimate under nonuniform noise conditions (e.g., when detectors have different noise levels or if noise is transient).

Recalling the convolution theorem, one can also express the weighted transform in terms of the unweighted transform $\widetilde{A}(\mathbf{f})$ (eq. [3.13]), and a convolving window function $W(\mathbf{f})$, itself the Fourier transform of the weights $w(\mathbf{x})$ as,

$$\widetilde{A}_w(\mathbf{f}) = W(\mathbf{f}) \otimes \widetilde{A}(\mathbf{f}). \tag{3.32}$$

The weighted transforms is simply the convolved version of the unweighted transform. As $\widetilde{A}_w$ is the estimate of the underlying spectrum with the least error possible, in practice this translates to variable noise conditions effectively limiting the resolution at which one is able to recover the underlying signal spectrum.

With localized in the spectral domains, weighted transforms should be used in producing spectral estimates that are faithful to the underlying signals. If the simpler, unweighted, transforms of eqs. (3.11) are used instead, features will end up unnecessarily scattered or smeared, resulting in an imperfect spectral estimate.

3.3.2 Independent (Uncorrelated) Noise

When noise is independent among samples, i.e., $C_{i,j} = 0$ for $j \neq i$, the covariance matrix has entries $C_{i,i} = \sigma_i^2$ only. The inversion for optimal weights is readily provided as,

$$w(X_i) = \frac{1}{\sigma^2(X_i)}, \tag{3.33}$$

and the expression for the uncertainty of the weighted mean values (derived according to eq. [3.26]) becomes simply,

$$\sigma^2_{\langle X \rangle} = \frac{1}{\sum_k w(X_k)}. \tag{3.34}$$

It ought to be emphasized that $1/\sigma^2$ noise weights are applicable *if and only if* samples are truly independent (i.e., uncorrelated).

3.3.2.1 Gaussian Noise and χ^2

Arguably, the most common form of independent noise encountered by experimentalists is Gaussian. It is white (i.e., has a flat spectrum) and characterized by the normal probability density distribution $p(x)$, expressed in terms of a single complex variable x as,

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{|x|^2}{2\sigma^2}\right). \tag{3.35}$$

The universal occurrence of Gaussian noise owes to the *Central Limit Theorem*, which observes that the compound distributions of a large number of random variables tend to produce normal distributions. Because microscopic fluctuations add up to produce macroscopic noise properties, Gaussian noise is omnipresent in nature.

When noise is Gaussian, the most probable underlying value μ_X producing the observed set of repeated measurements X_k is the one which minimizes the noise normalized square deviation (cf. Section 4.3.1),

$$\chi^2 = \sum_k \left| \frac{X_k - \mu_X}{\sigma_{X,k}} \right|^2. \tag{3.36}$$

While χ^2 is normally expressed in terms of the noise variances in this way, it may also be written in terms of the noise weights as,

$$\chi_w^2 = \sum_k w_{X,k} |X_k - \mu_X|^2. \tag{3.37}$$

since these two definitions are equivalent for all types of independent noise (cf. $w = 1/\sigma^2$). Interestingly, he weighted mean of eq. (3.26) provides the μ_X value that minimizes χ^2, and thus yields the maximum-likelihood estimate for the underlying value when noise is Gaussian (see Section 4.3.1).

The method of χ^2 minimization is commonly used in data analysis, even outside Gaussian noise. However, χ^2 minimizing estimates are guaranteed to be most likely only when Gaussian noise. More generally, the maximum-likelihood property may arise for independent noise also, for a sufficiently large number ($N \gg 1$) of samples, owing to the *Central Limit Theorem*. When noise is correlated among samples, the definitions of eqs. (3.36) and (3.37) are no longer equivalent. Then, the χ^2 minimization may be performed using either of these, although we suggest that the weighted χ_w^2 of eq. (3.37) provides a more natural extension of the arithmetic *sum-of-squares* deviation measure for weighted data. There is no truly compelling argument for this choice which is a matter of personal preference. However, let us re-emphasize that χ^2 minimization does not, in general, lead to maximum-likelihood estimates for correlated noise.

3.3.3 Stationary Noise

Stationary noise is translation invariant (uniform) in the configuration space, but not necessarily independent among samples. As a result all measurements from the same stationary noise process must carry equal weights. The translation invariance allows the approximation of the covariance matrix C_D from the data themselves, since the entries of C can depend only on the separation vector $\mathbf{\Delta} = |\mathbf{x}' - \mathbf{x}|$ of measurements not explicit positions of those (c.f. circulant matrices in Tegmark 1997).

$$\begin{aligned} C_D(\mathbf{\Delta}) &\approx \langle\, D_0(\mathbf{x})\, D_0(\mathbf{x} - \mathbf{\Delta})\,\rangle \\ &\approx N_k\, R_I(\mathbf{\Delta}). \end{aligned} \tag{3.38}$$

Approximately correct weights may be estimated, when sufficient data was taken for providing an accurate representation of the covariances. As the relevant terms are effectively expressed by the autocorrelation $R(\mathbf{\Delta})$ (cf. eq. [3.15]) of the zero mean signal $D_0 = D - \langle D \rangle$, it can also be estimated from the back transform of the PSD estimate, $P'_D(\mathbf{f}) = |\widetilde{D}_0(\mathbf{f})|^2$ owing to the convolution theorem (eq. [3.13]). Therefore,

$$\widehat{C}(\mathbf{\Delta}) \Longleftrightarrow |\pi(\mathbf{f})|^2.$$

A sufficient and complete description of stationary noise is thus provided by its spectral profile $\widetilde{\pi}(\mathbf{f})$ alone.

Many types of noise that affect physical measurements are stationary in nature. Otherwise, they may be slowly varying—in which case small enough subsets of the data can still be treated individually as stationary bits. Or, the noise may affect only a fraction of the time-stream data (e.g., electronic glitches, chirps, cosmic rays or someone accidentally jolting the measurement apparatus), and the affected measurements may be discarded without jeopardizing the effectiveness of the experiment. Therefore, even when noise is not quite stationary, often it can still be treated as such after the necessary precautions are taken. In this way, the assumption of stationary noise, in discussing data analysis, can be made without loss of generality.

When noise is stationary along some but not all of the coordinate directions, the above can be extended by considering a split of the coordinates $\mathbf{x}$ into a the subspace with stationary noise $\mathbf{x}_S$ and its orthogonal counterpart $\mathbf{x}_\perp$ (c.f. subspace Fourier transforms of Section 3.2.2). Then,

$$C\big(\{\mathbf{x}_S, \mathbf{x}_\perp\}, \{\mathbf{x}'_S, \mathbf{x}'_\perp\}\big) \rightarrow C\big(\mathbf{\Delta}_S, \mathbf{x}_\perp, \mathbf{x}'_\perp\big), \tag{3.39}$$

and the estimation involves the cross-correlation between various stationary signals located at $\mathbf{x}_\perp$ and $\mathbf{x}'_\perp$ respectively, i.e.,

$$C_D\big(\mathbf{\Delta}_S, \mathbf{x}_\perp, \mathbf{x}'_\perp)\}\big) \approx \langle\, D_0(\mathbf{x}_S, \mathbf{x}_\perp)\, D_0(\mathbf{x}_S - \mathbf{\Delta}, \mathbf{x}_\perp)\,\rangle\,, \tag{3.40}$$

which can be estimated also using the product spectra from selective Fourier transforms along $\mathbf{x}_S$, of the stationary signals at $\mathbf{x}_\perp$ and $\mathbf{x}'_\perp$,

$$\widehat{C}_D(\mathbf{\Delta}_S, \mathbf{x}_\perp, \mathbf{x}'_\perp) \iff \widetilde{D}^*_0(\mathbf{f}_S, \mathbf{x}_\perp) \times \widetilde{D}_0(\mathbf{f}_S, \mathbf{x}'_\perp). \tag{3.41}$$

3.3.3.1 $1/f$ Type Noise

The PSD profile $P(\mathbf{f})$ gives the name to $1/f$ noise, which is used more generally to describe noise with spectral power profiles $|\widetilde{\pi}(f)|^2 \propto 1/f^\alpha$ for various positive values of α. The occurrence of $1/f$-type noise is most typical in the time direction of data, but examples of spatial $1/f$ noise are also known.

When $\alpha = 2$ the noise is Brownian or red. It is produced in random walk processes (i.e., Brownian motion) while the redness is in reference to the heavy distribution of power towards the lower (redder) frequencies. The spectrum of atmospheric variations in radio or submillimeter astronomy is an example of Brownian noise (Bester et al. 1992). The profile corresponding to $\alpha = 1$ belongs to the canonical $1/f$ noise. Also referred to as pink noise (because its spectral distribution is intermediate between that of white noise and red noise) or flicker noise, it is a common property of electronic devices (e.g., bulk resistors). Other values of α also arise in nature, and may be attributed to clustering (e.g., $\alpha \simeq 2.9$ in galactic cirrus in Gautier et al. 1992; Miville-Deschenes et al. 2002, $\alpha \simeq 1.6$ for APM galaxies in Padilla & Baugh 2003, or $\alpha \simeq 0.83$ for the CIB in Miville-Deschenes et al. 2002) or may arise from a distribution of feature sizes (e.g., spatial Kolgomorov spectrum of eddies in atmospheric turbulence with $\alpha = 5/3$, see Bester et al. 1992).

The trouble with $1/f$-type noise is that deviations grow in configuration space (i.e., the time series), along the affected coordinate direction x. To find out how rapidly this happens consider N consecutive integrated samples of data taken at intervals Δ in the affected direction. These explore a frequency range from the lowest nonzero frequency component $1/N\Delta$ to the Nyquist cutoff at $1/2\Delta$, after the constant component is used for the mean subtraction allowing for covariance matrix estimation. By Parseval's theorem one can then relate the expected variances of the data in both configuration and spectral domains as (after the necessary normalization involving N),

$$\sigma^2 \propto \int_{1/(N\Delta)}^{1/(2\Delta)} P(f)\, df. \tag{3.42}$$

Deriving arithmetic averages for the samples one finds that the variances $\sigma^2 = [C_I]_{i,i}$ grow for sufficiently large number of samples $N \gg 1$ as,

$$\sigma^2 \propto \begin{cases} N^{1-\alpha} & \text{for } \alpha < 1 \\ \sqrt{\ln(N)} & \text{for } \alpha = 1 \\ N^{\alpha-1} & \text{for } \alpha > 1 \end{cases} . \tag{3.43}$$

Variances cannot grow indefinitely in the physical world. Common sense requires that $1/f$-type power spectra plateau at the very low frequencies ($f < f_c$) causing a leveling in the variances for data spans $N > 1/f_c\Delta$. However, the underestimation of relevant covariance terms should be expected for smaller data spans, resulting in higher than fair weights for data under $1/f$ conditions. Therefore, it is better to filter (Section 3.4) $1/f$ noise rather than just attempt weighting it.

The variances of averages is $\sigma^2/(N-1)$. Thus, $1/f$ noise integrates down with averaging only for $\alpha < 2$. When α is 2 (Brownian noise), the effective noise variance remains constant no matter how many samples are combined to form averages. Single measurements carry just as much Brownian noise as the average from any number of consecutive samples.

In practice, the structure of noise tends to be complex, with various types of stationary noise overlaid on white noise. Thus, when Brownian noise and white noise are simultaneously present, the latter will integrate down whereas the Brownian component stays invariant as data is averaged.

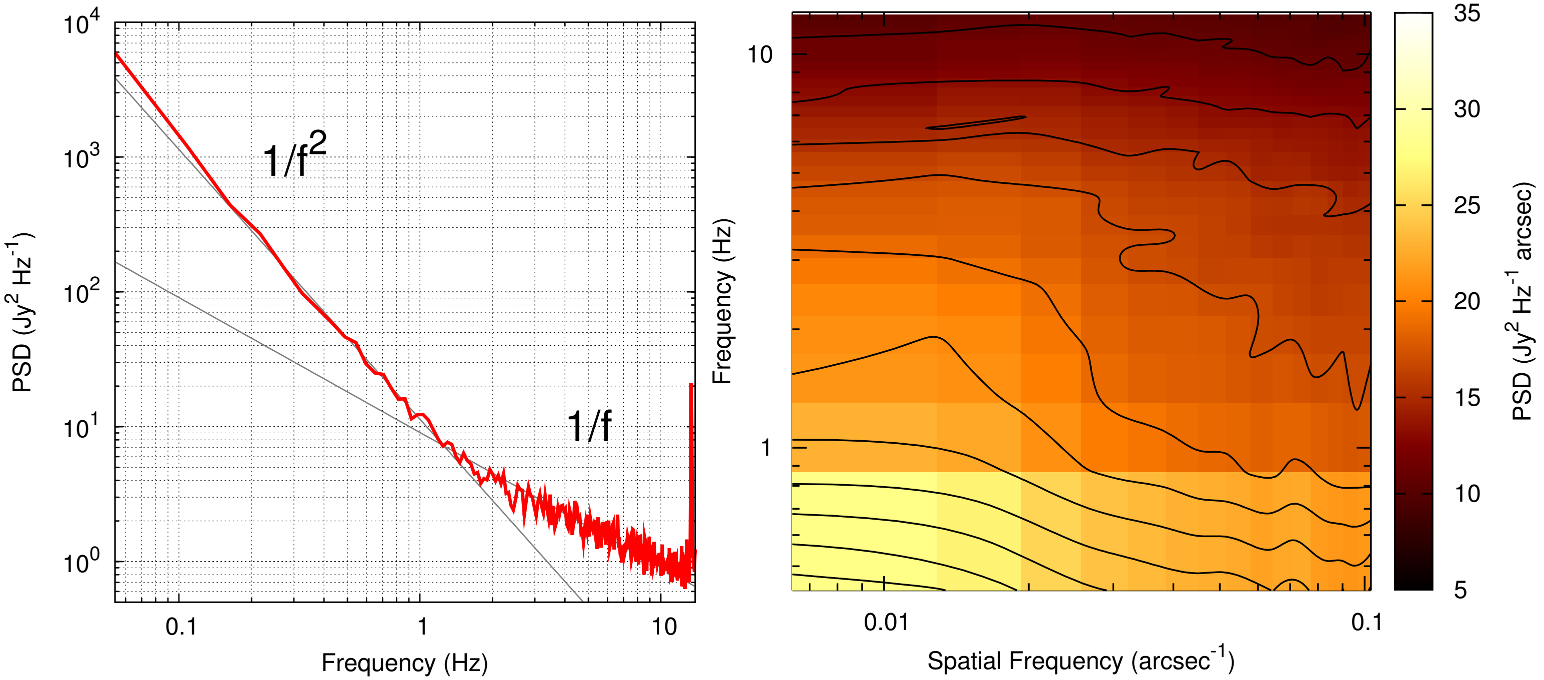


Figure 3.1 The spatial and temporal atmospheric spectrum at 350 μm. The averaged spectrum from a 10 minute integration is shown for a single SHARC-2 detector (left). The signal exhibits both $1/f^2$ and $1/f$ behaviors as well as a narrow band pickup aliased around 13.4 Hz. After the effective removal of the $1/f^2$ common atmospheric noise component and instrumental noise terms, the spatial-temporal spectrum of the atmosphere is estimated using the entire SHARC-2 detector array (shown to the right). The 350μm sky exhibits $1/f$ like behavior along both the spatial and temporal spectral axes.

This causes the experimentalist to reach a noise floor once the $1/f$ component becomes dominant. Such noise floors are not uncommon to experiments (e.g., heterodyne spectrometers), and will impose practical sensitivity limits unless noise weighting is accompanied by appropriate filtering (see Section 3.4) in the analysis of data. Even when $1/f$ noise is less harmful, it will take a tow on the integration times necessary to reach deeper in the data once the $1/f$ component becomes dominant.

3.3.3.2 Weighting Composite Spectra of Stationary Bits

The above example of Brownian noise demonstrates that weighting alone does not always handle stationary noise in a satisfactory manner. At least, not when it is applied in the configuration (time-series) domain of the data. However, as it was already hinted earlier for the case of slowly varying noise properties, the time series can normally be split (sometimes arbitrarily) into smaller stationary chunks. When such a division of data can be done such that the source spectra are stackable (e.g., via linear transformations, see Section 3.6.6.1), it becomes possible to consider an averaged source spectrum, constructed from the stacked components much the same way as time series averages. As each of the stacking spectra measures the same spectral components independently, it implies that $1/\sigma^2$ noise weights can be used for combining components into statistically meaningful, maximum-likelihood averages.

For a stacking set of spectra, indexed by m, each with frequency components indexed by f, i.e., for a set of measured spectral components $\widetilde{D}_{m,f} = \widetilde{S}_f + \widetilde{N}_{m,f}$, we can construct the weighted spectral source estimate, $\langle \widetilde{S}_f \rangle$ as,

$$\langle \widetilde{S}_f \rangle = \frac{\sum_m \widetilde{D}_{m,f} \widetilde{\sigma}_{m,f}^{-2}}{\sum_m \widetilde{\sigma}_{m,f}^{-2}}, \tag{3.44}$$

where $\widetilde{\sigma}_{m,f}$ is the standard deviation of the spectral noise component $\widetilde{N}_{m,f}$. The resulting spectral source estimate can be then transformed into the time-series source estimate $\langle S_k \rangle$ by the appropriate backward Discrete Fourier Transform (DFT) or Fast Fourier Transform (FFT).

3.3.4 Flagging Data

Before concluding the discussion on weighting, one ought to discuss noise that is approximately stationary but with the occasional transient flukes that may occur in any experiment. Often such outlier data may be simply discarded, without a negative impact (e.g., induced bias) on the remaining data set. Most data analysis approaches should run consistency checks on data and flag measurements which do not pass. The successive analysis ought to ignore the flagged measurement. The method of weighting offers an obvious and practical way for discarding questionable data, simply by assigning zero weights ($w_k = 0$) to the flagged measurements I_k. This way, all the statistical estimates involving weights (e.g., summations used in averages and propagates noise estimates) remain valid, such that no further treatment of discarded data is necessary. The use of zero weighting can thus be extended as a generic tool for flagging.

3.4 Filtering (Noise Isolation)

Filtering provides a complement approach to weighting, by discarding noisy bits of data, when these are not critical to the measurement, or simply deemed unreliable. Filtering is especially effective for treating excess noise which is localized in a representation of the data (i.e., affecting only a few components therein) but, nevertheless, 'pollutes' (cf. eq. [3.7]) much or all of the measurements in the time series. By downsizing or discarding the problem components one can, at times, dramatically reduce noise in the time series, at the minimal cost of axing the corresponding source components along the way. Often, the loss of information from filtering is generously compensated by reduced levels of background noise. However, it is a for the observer to decide what trade-off between conservation of the source signals (i.e., completeness) and lower noise levels is desirable or acceptable. Such a trade-off is unavoidable when filtering overlapping signals and noise, and therefore one is presented with the unattractive dilemma between,

$$\text{more complete but noisier} \xrightarrow{\text{more filtering}} \text{cleaner but more incomplete}$$

signal reconstructions (cf. eqs. [3.5] and [3.8]).

Spectral filtering presents a particularly effective way for dealing with stationary noise, such as $1/f$ noise or narrow band resonance (e.g., 50 Hz or 60 Hz pickup from the electric power grid). The following discussion primarily addresses filtering in the spectral domains provided by the appropriate Fourier transformations. Nevertheless, one can easily generalize and apply the same arguments (and results) to other representations of the data.

Multiplicative spectral filtering is performed via a filter pass function $\Phi(\mathbf{f})$ as,

$$\widetilde{D}_{filtered}(\mathbf{f}) = \Phi(\mathbf{f})\widetilde{D}(\mathbf{f}). \tag{3.45}$$

Comparing the above to the convolution theorem (eq. [3.13]), the filter can also be expressed in the configuration space, as an equivalent convolution filter $\phi(\mathbf{x})$, itself the backward Fourier transform of $\Phi(\mathbf{f})$, i.e.,

$$D_{filtered}(\mathbf{x}) = \phi(\mathbf{x}) \otimes D(\mathbf{x}). \tag{3.46}$$

Whereas Φ may take all values, in principle, including complex numbers, practical noise filters are real valued in the range of 0–1, thus preserving, reducing or discarding spectral amplitudes, but never altering their phase, or enhancing them beyond the measured levels. These then correspond to configuration space convolution filters which are symmetric in the coordinates (i.e., depend only on the separation of samples but not on their relative order).

Since the measured signal D can be decomposed explicitly as a sum of the filtered and rejected signals, $D = D_{filtered} + D_{rejected}$, one can equivalently characterize a filter response by its rejection signal spectrum $\widetilde{D}_{rejected}(\mathbf{f}) = \rho(\mathbf{f})\widetilde{D}(\mathbf{f})$ where $\rho(\mathbf{f}) = 1 - \Phi(\mathbf{f})$, or alternatively by the time-series rejection function $\widetilde{\rho}(\mathbf{x}) = \delta(\mathbf{x}) - \phi(\mathbf{x})$ calculated as the Fourier transform of $\rho(\mathbf{f})$ and applied by convolution. Since the ultimate goal of filtering is the rejection of noise, the rejected signal $D_{rejected} \approx N$ should represent an approximate model of the noise signals, while the signal passed by the filter, $D_{filtered} \approx S$ provides a cleaner (albeit incomplete) model of the source signal. The method of filtering, therefore provides a *de facto* separation of noise from the source, by assigning signals either as source or as noise (or as some statistical combination of both).

3.4.1 Filters vs. Weights

There is something puzzling about filters at first. In the discussion above, filters were described as operating in the spectral domains much the same way as weights were used in the configuration space,

$$\Phi(\mathbf{f}) \longleftrightarrow w(\mathbf{x}). \tag{3.47}$$

Consider, as an example the weighted Fourier transform and the filtering in the two representations. Both of these can be be expressed in either domain with help from the convolution theorem (eq. [3.13]), which is expressed for weighting and filtering respectively as,

$$\begin{aligned} w(\mathbf{x})A(\mathbf{x}) &\Longleftrightarrow W(\mathbf{f}) \otimes \widetilde{A}(\mathbf{f}), \\ \phi(\mathbf{x}) \otimes A(\mathbf{x}) &\Longleftrightarrow \Phi(\mathbf{f})\widetilde{A}(\mathbf{f}). \end{aligned} \tag{3.48}$$

Clearly, these are fully symmetric operations. Yet, we have previously proclaimed that weights can be used to provide accurate and unbiased (i.e., complete) statistical estimates from the data whereas filters are typically associated with information loss (i.e., incomplete). The different behaviors seem to contradict the underlying symmetry, unless the asymmetry arises from the data (the quantity A) themselves. This is, in fact, the critical detail which decides definitively the respective domains of weighting vs. filtering.

Weights provide unbiased estimates because they are used for combining *repeated measurements* of the *same source structure* (i.e., components) in configuration space, whereas filters operate on components (i.e., structures) *independently* (cf. independent noise) by simply throwing away the less valuable parts. It is therefore, the asymmetric nature of the data (the quantity A above), redundant in one domain but not necessarily in the other, which is responsible for the very different effect that filters and weights have.

In light of the above, it makes sense to generally refer to the *configuration space* as the domain of choice in which repeated measurements are (statistically) combined (e.g., averaged by a weighted mean) into providing the critical experimental results, while the *spectral domain* is consequently refers to the conjugate Fourier space. However, for simplicity we will continue to assume that the time-series is the configuration domain (especially since we already used these interchangeably). This can be done without loss of generality, as the raw data may always be appropriately transformed into the convenient *configuration space* representation thus providing an alternative *time-stream* description.

Weights are primarily useful in the configuration space while multiplicative filters are more typical in the spectral domains. This is exactly how these were introduced in the first place. However, when redundancies simultaneously exists in both domains (all the better!), weighting can be used in both (cf. spectral weighting in Section 3.3.3.2). Multiplicative filters also may be used in both domains, and the discarding of unreliable measurements (see Section 3.3.4) serves as an example of configuration space filters.

It was argued that filtering is ideally performed on the weighted Fourier transforms of the time-series data for providing the best estimate of the underlying signal spectrum. On the other hand, weights ought to be derived from the filtered time streams, such that the best statistical estimates of the source structure can be made. Thus, effective filtering requires weights while effective weighting requires filtering first—presenting a chicken-and-egg problem if both are to be determined in the analysis.

However, imagine starting with only a very crude set of weights (e.g., uniform weights) to perform suboptimal filtering of the data. Even crude filtering tends to result in cleaner time streams, which then can provide better noise weights, and in turn better filters. In a leap of faith, it is not hard to imagine that after iterating the weighting–filtering cycle a number of times, both weights and filters will converge close enough to a self-consistent and optimal pairing, that these may be accepted as practically final. An iterative data reduction approach may be unavoidable when noise weights are not determined externally (e.g., by separate noise characterization of the instrument).

3.4.2 Noise Whitening

Encouraged by the resemblance between weights and filters (eq. [3.48]), it is natural to extend this connection, by constructing noise filters which have the look and feel of noise weights. Since stationary noise, we argued, leads to independent (uncorrelated) spectral components, it should not be surprising to suggests noise filters as $\Phi(\mathbf{f}) \propto 1/\sigma^2(\mathbf{f})$. Such filters are *noise-whitening filters*, since its application converts a stationary noise spectrum into a white noise spectrum. Under the practical normalization (which leaves the white noise component unfiltered), the noise-whitening filter takes the form,

$$\Phi(\mathbf{f}) = \frac{\sigma^2}{|\widetilde{N}(\mathbf{f})|^2} = \frac{\sigma^2}{\sigma^2 + |\widetilde{N}_x(\mathbf{f})|^2}, \tag{3.49}$$

expressed in terms of the total noise $N(\mathbf{f})$, or the excess noise $N_x(\mathbf{f})$, and the RMS spectral white noise level σ. Noise whitening is an attractive choice because it converts correlated noise into a time series with independent samples (see Section 3.2.3). With whitened noise, simple $1/\sigma^2$ noise weights (see Section 3.3.2) can be used for the various statistical estimates deriving from the filtered data. This is useful, especially when repeated observations of the same source component are distributed in the time series such that the covariance matrix may be difficult to estimate.

Another practical attribute of noise whitening, especially when compared to optimal filtering (see below), is that the required filter function is uniquely determined by the noise profile alone (which needs to be known, guessed or otherwise assumed by the observer), but is entirely independent of the shapes and sizes of source signals.

Because of its independence on the source signals, and because it greatly simplifies the task of providing reliable statistics on the data, the noise whitening filter is one of the most useful filters for the treatment of signals under stationary noise (e.g., Tegmark 1997).

3.4.3 Optimal Filters

For a more rigorous approach, a different class of practical filters seek to optimize the filtered signal according to chosen criteria. While the Wiener filter (below) is often referred to as *the* optimal filter, it is 'optimal' only under the particular criterion (i.e., providing least squares deviation from the underlying signal) that is used in its derivation. If one considers different criteria different 'optimal' filters result. Some of these may be better suited for low signal-to-noise conditions than others, from an observer's point of view.

3.4.3.1 The Wiener Filter

The Wiener filter (Press et al. 1986) provides the least-squares estimate to the underlying source by minimizing the integrated squared deviation between the filtered-deconvolved data and the underlying source signal as,

$$\int [\phi(\mathbf{x}) \otimes D_u(\mathbf{x}) - S_u(\mathbf{x})]^2 d\mathbf{x} \propto \int \frac{|\Phi(\mathbf{f})\widetilde{D}(\mathbf{f}) - \widetilde{S}(\mathbf{f})|^2}{|\widetilde{R}(\mathbf{f})|^2} d\mathbf{f}, \tag{3.50}$$

The above makes use of Parseval's theorem (eq. [3.17]) to express the integrated (or summed) least squares in terms of the spectral Fourier components. Using that stationary noise signals are uncorrelated among components, and also to any other signal components (e.g., source), the above can

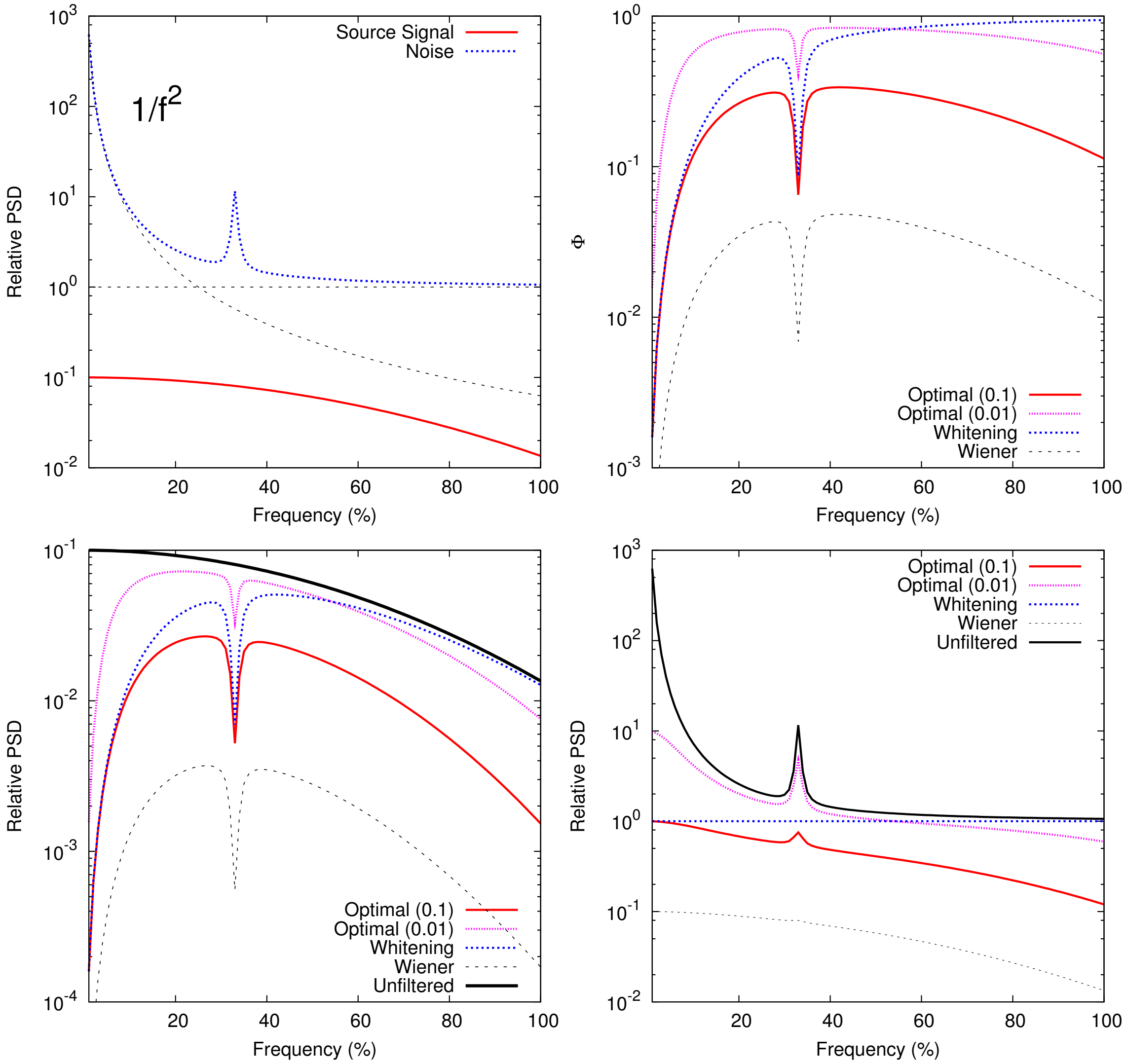


Figure 3.2 A weak-signal spectrum and some practical filters. The spectrum of a weak Gaussian source signal (solid red) in the presence of noise (dashed blue) with $1/f^2$, white and narrow-band components, is shown (top left), with peak signal-to-white-noise ratio of 0.1. Various practical spectral filter responses are plotted for this signal configuration (top right). Also shown are the effect of these filter on the source (bottom left) and the noise (bottom right) spectra. The Wiener filter (long dashes) is clearly not optimal in this low signal-to-noise regime. Noise whitening and modified Wiener filtering offer more optimal alternatives for weak signals.

be further manipulated to yield,

$$\int \frac{|\widetilde{S}(\mathbf{f})|^2\{1-\Phi(\mathbf{f})\}^2 + \Phi(\mathbf{f})^2|\widetilde{N}(\mathbf{f})|^2}{|\widetilde{R}(\mathbf{f})|^2} d\mathbf{f}, \tag{3.51}$$

for filter functions $\Phi(\mathbf{f})$ which are real. The minimizing condition with respect to the filter component $\Phi(\mathbf{f})$ provides the Wiener filter as,

$$\Phi(\mathbf{f}) = \frac{|\widetilde{S}(\mathbf{f})|^2}{|\widetilde{S}(\mathbf{f})|^2 + |\widetilde{N}(\mathbf{f})|^2}. \tag{3.52}$$

Thus, Wiener filters will closely preserve source components ($\Phi \approx 1$) whenever $|\widetilde{S}(\mathbf{f})| \gg |\widetilde{N}(\mathbf{f})|$, while effectively discard ($\Phi \ll 1$) components that are noise dominated i.e., when $|\widetilde{S}(\mathbf{f})| \ll |\widetilde{N}(\mathbf{f})|$.

This exposes an important weakness of Wiener filtering, when applied to data with low signal-to-noise ratios in the time stream as in the signal spectra. When signal-to-noise ratios are low (i.e., $S/N \ll 1$), Wiener filters flatten drastically the source signals in a bid to reduce noise amplitudes to the expected source signal level.

3.4.3.2 Optimal Low Signal-to-Noise Filters

The shortcoming of Wiener filters, therefore, is the direct consequence of the least squares condition that the filter is derived from. The changing of that condition is necessary to come up with different and otherwise 'optimal' filters, which are better tailored for low signal-to-noise data streams.

In the search for alternative optimal filters, let us reexamine, therefore, the objectives of optimal filtering. The primary goal should be the more or less accurate preservation of source signals, no matter how deeply buried under noise they may be in the time series. This is achieved by minimizing the square deviation between the underlying and filtered-deconvolved (but noiseless!) source signals i.e., by the minimization in the spectral domain (cf. Parseval's theorem) of,

$$\int \frac{|\widetilde{S}(\mathbf{f})|^2 |1-\Phi(\mathbf{f})|^2}{\widetilde{R}^2(\mathbf{f})} d\mathbf{f}. \tag{3.53}$$

Secondly, filtering should also be effective at ridding the data stream of excess noise, much like it does in the case of Wiener filters. As such we would like to minimize the integrated filtered noise power, expressed in the spectral domain as,

$$\int |\Phi(\mathbf{f})|^2 |\widetilde{N}(\mathbf{f})|^2 d\mathbf{f}. \tag{3.54}$$

The combined goal of filtering is to reduce noise while leaving the source signals intact, as much as possible, in the process. As such, one ought to minimize a linear combination of the two expressions (in a Lagrangian construct), such that the relative weight of one condition with respect to the other can be adjusted according to preference. To really make these two conditions comparable, nevertheless, it is advisable to minimize the deconvolved-filtered noise, instead of the observed-filtered noise, since the first condition has been formulated in terms of deconvolved signals already. This modification remains faithful to the principle of noise reduction. Therefore, we would like to find the optimal filter function $\Phi(\mathbf{f})$ which provides the minimum of

$$\int \frac{|\widetilde{S}(\mathbf{f})|^2 |1-\Phi(\mathbf{f})|^2}{\widetilde{R}^2(\mathbf{f})}\, d\mathbf{f} + \lambda^2 \int \frac{|\Phi(\mathbf{f})|^2 |\widetilde{N}(\mathbf{f})|^2}{\widetilde{R}^2(\mathbf{f})}\, d\mathbf{f}. \tag{3.55}$$

Here, the positive definite λ^2 parameters (which thus retains the signs of both components the same such that they are simultaneously minimized), determines the relative importance assigned to the two conditions. Note, that $\lambda = 1$ reproduces the least-squares condition leading to the Wiener filter (eq. [3.51]), making the Wiener filter a welcome entry in the expanded family of optimal filters thus derived. In general the simultaneous minimization of noise and model deviation, with respect to the filter component $\Phi(\mathbf{f})$, yields optimal filters:

$$\Phi(\mathbf{f}) = \frac{|\widetilde{S}(\mathbf{f})|^2}{|\widetilde{S}(\mathbf{f})|^2 + \lambda^2 |\widetilde{N}(\mathbf{f})|^2}. \tag{3.56}$$

These were used by Walden (1988) without detailed justification. As $\lambda = 1$ reproduces the Wiener filter (an optimal filter for appreciable signal-to-noise ratios $|S|^2/|N|^2 \gg 1$), and smaller λ values will provide 'optimal' filters for the low signal-to-noise ($\lesssim 1$) regime, typical λ values should lie in the range 0–1. Greater $\lambda > 1$ values can be meaningful also, for these provide filters which reject noise more aggressively than the Wiener filter would. Such heavy filtering may, at times, be desirable even when signal-to-noise ratios are generously high. Note, that this extended family of optimal filters will have well-behaved filter components strictly in the range of $0 \leq \Phi \leq 1$.

While the choice of a λ value is a reflection of the observer's preference as to how much filtering is desirable, useful values should provide filter functions near unity for most critical source components such that the filtered source signal remains representative of the underlying source signal (c.f. the first criterion). For noise that is mostly white, i.e., a flat spectrum characterized by the standard deviation σ_f^2, with localized spectral features in excess of the white level, λ value on the order of

$$\lambda \lesssim \frac{\left\langle |\widetilde{S}(\mathbf{f})| \right\rangle}{\sigma_f} = \frac{\left\langle |S(\mathbf{x})| \right\rangle}{\sigma_x}, \tag{3.57}$$

will be typically useful. Once again recalling Parseval's theorem, this can also be expressed in terms of the time-series white noise level σ_x^2.

It is interesting to note, that equality in the above expression i.e., when $\lambda = \left\langle |S(\mathbf{x})| \right\rangle / \sigma_x$, optimal

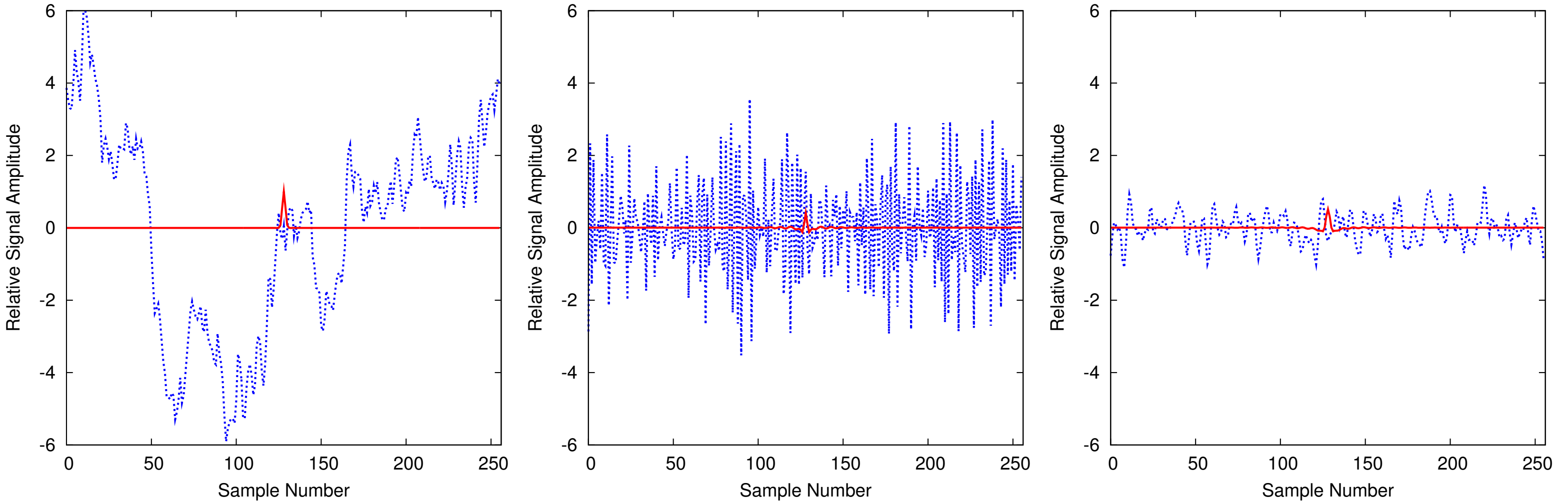


Figure 3.3 The effect of filtering on the time-series signals. Corresponding to the spectra of fig. 3.2, here the time series is shown (left), composed of a marginally resolved Gaussian source (solid red), and the $1/f^2$ dominated noise (blue dashes). Also shown are the effects of noise whitening (middle), and an optimal weak-signal filter with $\lambda^2 = 0.01$. Clearly, multiple observing of the source is still necessary for noise to integrate down for the signal to become detectable. Noise whitening is practical for filtering data prior to inversion, as it allows simple $1/\sigma^2$ weighting for combining measurements. At the same time, optimal filtering is, perhaps, better suited for filtering of the source image postinversion, resulting in a clean looking signal reconstruction.

filtering asymptotically approaches the noise whitening filter (see Section 3.4.2) for those component where noise greatly exceeds the white noise level i.e., $|\widetilde{N}(\mathbf{f})|^2 \gg \sigma_f^2$, and the source spectra is flat ($|\widetilde{S}(\mathbf{f})| \approx const.$). This connection between optimal and noise whitening filters means, that noise whitening provides a nearly optimal way for reducing excess noise, which is especially useful when one does not wish to assume an underlying source spectrum (perhaps because it is not known a-priori), necessary for optimal filter design. As such, it reaffirms the method of noise whitening as a crude way of quasi-optimal filtering for the most generic purposes.

3.4.4 Incorporating Prior Knowledge

As long as the excess noise dominates only a fraction of the source components, filtering can be an attractive and rewarding choice for obtaining cleaner, even if less complete, signals. The source components sacrificed in the process, may be partially or fully reconstituted by imposing constraints deriving from prior assumptions on the source structure. Examples of such corrections are offered by the *CLEAN* algorithm (Clark 1980) or maximum-entropy methods (Narayan & Nityananda 1986) used in radio astronomy for filling in the unsampled parts of the u–v plane in interferometric data.

At times, less sophisticated corrections will suffice. As an example, consider the integrated signal power $\int |S(\mathbf{x})|^2 d\mathbf{x}$ (of certain source features), often being the object of the experiment. One needs only to recall Parseval's theorem (eq. [3.17]), to recognize that the effect of filtering on the integrated signal power is,

$$\int |\widetilde{S}_{filtered}(\mathbf{x})|^2 \, d\mathbf{x} \propto \int |\Phi(\mathbf{f})\widetilde{S}(\mathbf{f})|^2 \, d\mathbf{f}. \tag{3.58}$$

The reduction (or change) in the integrated signal power is thus well characterized when the source distribution $\widetilde{S}(\mathbf{f})$, or equivalently $S(\mathbf{x})$, is priorly known, or can be reasonably guessed from the filtered spectrum itself (e.g., the astronomer often expects sources unresolved by the telescope, all of which therefore produce closely beam-sized images). If all components of the filtered source model are scaled by the correction factor,

$$Q = \frac{\int |\widetilde{S}_{exp}(\mathbf{f})|^2 \, d\mathbf{f}}{\int |\Phi(\mathbf{f})\widetilde{S}_{exp}(\mathbf{f})|^2 \, d\mathbf{f}} = \frac{\int |S_{exp}(\mathbf{x})|^2 \, d\mathbf{x}}{\int |\phi(\mathbf{x}) \otimes S_{exp}(\mathbf{x})|^2 \, d\mathbf{x}}, \tag{3.59}$$

then the peaks of the corrected source structures $Q \cdot S_{filtered}$ will be accurately represented as long as these are indeed resemblant of their expected profiles S_{exp}.

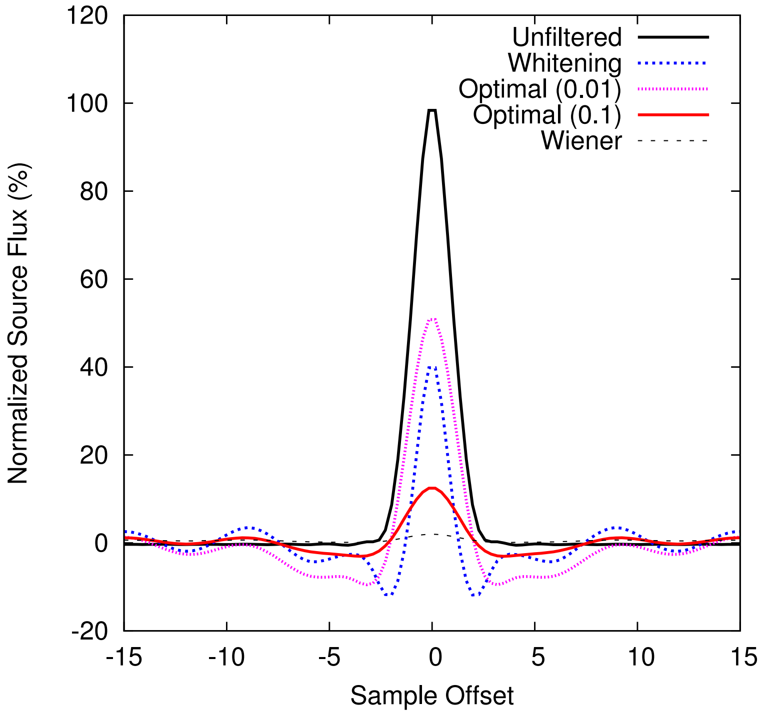


Figure 3.4 Source filtering and corrections. A close-up of the source signals, discussed in Figures 3.2 and 3.3. The different filters each have predictable effect on structures with known shapes. Thus, one can simply scale the filtered image by a correction factor (eq. 3.59), calculated for the expected source shapes, to recover peak signals strengths postfiltering.

3.5 Gain Knowledge

Noise filtering and noise weighting provide the essential tools of a successful data reduction approach. For these to be effective, further understanding of the measurement process and the signal pathways is required. Specifically, the typical experimental apparatus has several signal stages at every one of which interfering noise may be superposed on top the propagated source signals, producing the final noisy output at the readout stage.

$$D = g_S S + g_1 N_1 + \cdots + g_m N_m. \tag{3.60}$$

Here, the indeces (1–m) refer to the gain stages, while each quantity may be additionally indexed by measurement serials or may be thought of as continuous functions in configuration space, true to the convention of Section 3.2. Thus, g_1 and N_1 are shorthands that stand for the underlying noise contributions $[N_1]_k$ to the measurements D_k from the first gain stage with an effective $[g_1]_k$ gain. Alternatively, the gain function $g_1(\mathbf{x})$ affects the mapping of noise signal $N_1(\mathbf{x})$ into the output signal $D(\mathbf{x})$. The stage S, for the source signal, is one of the stages that are indexed 1–m.

3.5.1 Gain Accuracy Requirement

Suppose that the noise signals N_i, were known independently, such that these may be completely removed from the data stream—in effect creating an ideal signal D' with source only. However, when the noise is subtracted with a set of gains different by δg_i from the true values g_i, then a noise residual from N_i will remain in the 'cleaned' data D' i.e.,

$$D' = g_S S + \sum_{i=1}^{m} \delta g_i N_i. \tag{3.61}$$

The residual noise pollution is negligible, only if the gains g_i are accurate to δg_i as given by $|g_S| \langle |S| \rangle \gg |\delta g_i| \langle |N_i| \rangle$, leading to a relative gain accuracy requirement that is expressed as,

$$\left| \frac{\delta g_i}{g_S} \right| \ll \frac{\langle |S| \rangle}{\langle |N_i| \rangle}. \tag{3.62}$$

To appreciate what this pertains to in practice, let us illustrate the importance of characterizing signal (and noise) gains. As is the case in submillimeter astronomy, atmospheric noise (sky noise) can dominate signals from faint sources (e.g., distant galaxies) by several orders of magnitude. At $350\,\mu$m wavelength, sky variations in the few hours of observing required to detect galaxies at the $10\,$mJy level are typically on the order of $100\,$Jy. With signal-to-noise ratios of $\simeq 10^{-4}$ in the time stream, the relevant gains to atmospheric signals need to be known a to minimum of 4–5 significant figures, such that sky noise can be removed from the observed source effectively!

In this breakdown of the signals (eq. [3.60]), every term is a gain-signal product. Therefore, the terms are reproduced identically if gains are scaled by a constant factor while the corresponding noise signals are rescaled by its inverse. As a result of this degeneracy, the absolute scaling of gains is generally not important (unless one is interested in also learning the physical noise level), rather it is the relative scale-free structure gains that have critical impact.

3.5.2 Gains and Filters

In the discussion of filtering, we noted that the rejected signal $I_{rejected}$ can be treated as an approximate model of the noise. This makes the above argument, based on the knowledge of noise signals, directly relevant for the application of noise filters. As a result, filters require accurate gain knowledge (in line with eq. [3.62]) to perform effectively.

Because the filtering of stationary noise sources is most easily undertaken in an appropriate Fourier representation (see Section 3.4), it is necessary to investigate the effect of gains on the required transformations. According to the convolution theorem (eq. [3.13]), the Fourier transform of the data stream (eq. [3.11]), can be expressed as,

$$\widetilde{D} = \widetilde{g}_S \otimes \widetilde{S} + \widetilde{g}_1 \otimes \widetilde{N}_1 + \cdots + \widetilde{g}_m \otimes \widetilde{N}_m, \tag{3.63}$$

i.e., by convolution of the underlying signal (and noise) spectra with the transformed gain functions. In other words, noise signals, which may have tightly clustered spectral profiles at the signal stage of origin, tend to get spread out in the direct spectral representation of the readout signals. The spreading can be avoided, one noise component at a time.

Consider the gain corrected measurements D_i, which express what the signal D would look like if it were measured at the i^{th} gain stage i.e.,

$$D_i = \frac{D}{g_i} = \frac{g_S}{g_i} S + \frac{g_1}{g_i} N_1 + \cdots + N_i + \cdots + \frac{g_m}{g_i} N_m. \tag{3.64}$$

Accordingly, the Fourier transform of this gain corrected signal D_i will decompose into components as,

$$\widetilde{D_i} = \widetilde{g}_{S,i} \otimes \widetilde{S} + \widetilde{g}_{1,i} \otimes \widetilde{N}_1 + \cdots + \widetilde{N}_i + \cdots + \widetilde{g}_{m,i} \otimes \widetilde{N}_m, \tag{3.65}$$

where $g_{k,i} = g_k / g_i$. The noise produced at the i^{th} stage, is left undisturbed. Thus, this representation suited ideally for effective filtering by an appropriate Φ (see Section 3.4). However, the other noise signals, and the source signal, are left convoluted in this representation. Appropriate filtering at one gain stage will result in imperfect filtering at the other stages.

What becomes painfully clear from this, is that every noise signals needs to be filtered in the spectral representation, which is appropriate for the gains affecting that specific noise source. In practice, this means, that if Fourier transforms are used for the filtering, as many forward and backward transformations pairs (and corresponding filters) are necessary, as many different gain stages with excess noise are effectively present under the constraint of eq. (3.62). (Gains that are less different than the accuracy requirement, can be treated as practically identical.)

3.5.2.1 Filtering Order

In this way, when several gain stages are present, one needs to think about a sequence of filters (one for every relevant gain stage), instead of just a single pass of filtering. These filters must be applied one at a time to avoid unnecessary interference. The issue then becomes, what order, if any, among the filters is best. An answer may be provided by comparing the expression of the signals at a particular gain stage (eq. [3.64]), with eq. (3.60), and considering the residuals of improperly applied gains (eq. [3.61]). While filtering at one gain stage, the untreated (as yet) and overlapping noise signals remaining at other stages can be treated as if these were source signals. Thus, the choice for the remaining noise signals is likewise between insensitive completeness, or incomplete sensitivity (see Section 3.4).

Therefore, if (and only if) accurate (mostly complete and sensitive) recovery of the noise signals itself is critical to the experiment (e.g., submillimeter wavelength sky noise may be used to estimate small variations in atmospheric extinction, see Appendix B), it is best to filter the most dominant signals first. In this way, cross-pollution of noise signals (i.e., filter rejects) is minimized relative to the scale of the recovered noise amplitudes (or power).

3.5.3 Gains and Noise Weights

The incorporation of gain knowledge in providing weights appropriate to each of the signal stages is quite a bit simpler if the effective residual noise is can be treated as if white. The corresponding $1/\sigma^2$ weights of the gain transformed signals ($D_i \sim R/g$) relate to the weights derived for the readout signal D as,

$$w(R/g_i) = |g_i|^2 w(R). \tag{3.66}$$

As a result, weights from any signal stage can be easily converted into weights at any other stage, where statistical descriptions (e.g., means or variances) may thus be derived (see Section 3.3). Therefore, as a convention, we suggest that unqualified weights are derived at the readout i.e., w is synonymous to $w(R)$, whereas weights relevant to other signal stages are accordingly labeled.

3.6 Data Collection Strategies

Data collection (observing) strategies refer to the way the source distribution $S_{obs}(\vec{\xi})$, in the space spanned by the detectors, is translated into the time series of source signals $S(\mathbf{x})$. The observing patterns is captured by the mapping $\mathcal{M}$ in eq. (3.4).

$$S(\mathbf{x}) = \mathcal{M} \circ S_{obs}(\vec{\xi}). \tag{3.67}$$

Designing good observing strategies is often more art than science. The goal here is to turn this around and provide prominence to a scientific approach although, in the end, some intuition may be necessary in identifying optimal configurations. The best patterns provide data which allows the high *fidelity* modeling of the source with optimal *sensitivity*, while at the same time are able to withstand unforeseen adversities of the measurement process. Moreover, actual patterns must be physically realizable to be of practical use. These criteria each impose constraints independently and are, therefore, examined in objective detail below.

These concerns are not new in the field of ground based astronomy. Tweaks were suggested to the simple position-switched observations used at submillimeter wavelengths early on (Emerson 1995), also in scanning modes (Emerson 1988). Chopped observations were perfected also for the SCUBA[2] instrument (Johnstone et al. 2000). More recently, Cosmic Microwave Background (CMB) experiments, have called on new patterns to improve sensitivities (e.g., Dupac & Tauber 2005). These are also directly relevant to present day submillimeter astronomy. A good discussion of CMB mapping strategies is offered by Tegmark (1997), who introduced a quantitative comparison of various strategies based on an effective sensitivity measure (the 'figure of merit') under $1/f$ noise.

3.6.1 Fidelity

It was shown (in Section 3.4) that filtering for excess noise, used in the bid to provide optimal sensitivities for weak signals, can affect the source signals adversely, leaving one with incomplete source reconstructions. Observing strategies can help, with providing better isolation of the signal components (in the filter domains of the data) from the regions which are affected by excess noise. In this way, the desire for improved source completeness is analogous to arranging the source components such that they are measured with optimum sensitivities (cf. the dilemma of filtering between completeness and noise). This is a reflection of better strategies yielding data where both more sensitive and more complete signal reconstructions are possible. Therefore, refer to discussion on sensitivity (further below). Nonetheless, there is another, sampling related, aspect to consider here.

---

[2] Submillimeter Common User Bolometer Array.

3.6.1.1 Spacing of the Samples

When the source distribution $S_{obs}(\vec{\xi})$, as measured by the detectors (eq. [3.1]), is sampled at regular intervals (spacings) of Δ_s, the highest frequency component measured will be the Nyquist frequency at $f_N = 1/2\Delta_s$. Signal components at frequencies higher than the Nyquist frequency, are aliased into the sampled spectral range. To avoid aliasing signals, it is best to choose observing patterns, which sample the source structures $S(\vec{\xi})$ at intervals that are sufficiently small, such that the signal components above the cutoff frequency can be neglected. As a rule of thumb, the sampling interval should be chosen such that the Nyquist frequency f_N is over $\sim 2\ \sigma_f(I)$, or equivalently, $f_N \gtrsim \Delta I(f)$, leading to an approximate condition of,

$$\Delta_s \lesssim \frac{1}{2}\Delta S_{obs}(\vec{\xi}), \tag{3.68}$$

in terms of the typical widths (Δ) of the measured source signals $S_{obs}(\vec{\xi})$ in the physical space spanned by the detectors. Therefore, in designing observing patterns, one should sample the observed space at intervals of, or finer than, half the typical size of the (Gaussian) features which are to be accurately measured. That is the Nyquist sampling condition (Nyquist 1928; Shannon 1949; Press et al. 1986).

3.6.2 Sensitivity

Without qualification, which (if any) of the signal components are deemed more important for the measurement, it is desirable to maximize signal-to-noise ratios in all source components at once. Therefore, the average signal-to-noise ratio of source components (in the convenient representation of noise) presents itself as a generalized sensitivity measure. When noise components are independent (i.e., stationary noise) the average signal-to-noise ratio is identical to the averaged χ^2 (eq. [3.36]). Therefore, one ought to maximize the average, or equivalently the total χ^2, for which the weighted definition of χ^2 (eq. [3.37]) is suggested here i.e., $\chi^2 = \sum_i w_i |S_i|^2$.

Clearly, optimal sensitivities are reached, when source signals occupy the least noisy components, while conveniently all distributions of the signal power among components with identical noise levels (and weights) provide equal levels of sensitivity, leaving substantial freedom for the accommodation of other critical considerations in optimizing the signal distribution of observing strategies.

The problem with using just χ^2 is that it changes as the relative noise and signal levels are varied, even if their structures represent equivalent configurations. If to be applied as an absolute sensitivity measure, useful for comparing different strategies under similar (but not necessarily identical) signal and noise conditions, and additional normalization is required. Consider, therefore,

$$\eta_S = N_i \frac{\sum_i w_i p_i}{\sum_i w_i} = N_i \left\langle p_i \right\rangle , \tag{3.69}$$

in terms of the number of components N_i and the realized power fraction p_i,

$$p_i = \frac{|S_i|^2}{\left\{\sum |S|^2\right\}_{\max}}, \tag{3.70}$$

which expresses the measured component power in relation to the maximum potentially realizable signal power by any strategy (or representation), thus providing a pattern independent normalization. For stationary sources, this can be calculated effectively as the total observable integrated signal power in the detector space, multiplied by the number of time samples N_t in the data and appropriately scaled (cf. Parseval's theorem, eq. [3.17]) to the given signal representation. Simply stated, the utmost that can be expected from any data collection strategy is to represent in the data all the observable signals at all times. Thus,

$$\left\{\sum |S|^2\right\}_{\max} \propto N_t \int \left|S(\vec{\xi})\right|^2 d\vec{\xi}. \tag{3.71}$$

Once again, $d\vec{\xi}$ represents a volume element for the integration. Because, the normalization of p_i is such that it is $\leq$1, η_S is also restricted to the range of 0–1. As such, it is appropriate to think of it as a sensitivity efficiency measure.

3.6.3 Robustness

An observing strategy (pattern) is robust if the chance of unforeseen circumstances rendering the data unusable is kept at a minimum. Under anticipating the unexpected, any signal component is equally likely to become fragile or incapacitated by freak of nature. Data is more robust if the results are less dependent on each and every signal components. This suggests that entropy should serve as a measure as to how widely signals are distributed in the domain of a representation. Arguing that the realized fraction p_i of the highest attainable aggregate source signal power (defined in eq. 3.70), carried by a component i (in some representation of the data) is analogous to the imagined probability that the source signal power ends up in that component, the familiar form of entropy from statistical mechanics is directly suggested i.e., $H \propto -\sum p_i \ln p_i$.

Unfortunately, the maximum values of entropy reachable depend on the number of components in the representation, and therefore, on the volume of the data set, making it unsuitable for comparing collection strategies with different scopes. A normalization is required. Comparison to the entropy H_u arising from uniform valued components (i.e., each with $p_u = 1/N_i$), provides a relative, volume-independent, standard of the robustness of signal power distribution.

Because the various components S_i, and hence p_i, can carry uneven levels of noise a better measure ought to be provided by converting the summations involved into noise weighted ones, arriving at the noise weighted relative pattern entropy (or pattern entropy efficiency), expressed as,

$$\eta_R = \frac{H}{H_u} = -\frac{1}{\ln N_i} \frac{\sum w_i p_i \ln p_i}{\sum w_i}. \tag{3.72}$$

Here w_i are noise weights calculated from the variances σ_i^2 for the components S_i. This makes perfect sense, as the less noisy components, with the heavier weights, contribute more to the weighted relative entropy than those noisier components with the same signal power. It follows, that the highest levels of entropy efficiency are reached when signals are uniformly distributed among components with comparable noise, alas with preference for more power in the sensitive components than in noisier ones. As both a wider signal scope, and better noise avoidance, result in increased levels of relative entropy, it is ideal for providing an objective evaluation of robustness under the prerequisite of noise isolation.

Note, that the fraction on the right provides the weighted average entropy $\langle H \rangle$ (cf. eq. [3.26]) for a signal component, and thus the relative entropy is but a 'volume normalized' average entropy contribution per component, or

$$\eta_R = \frac{\langle H \rangle}{\ln N_i}. \tag{3.73}$$

When the expected noise is uniform, η_R provides a robustness measure that is restricted the range of $0 - 1$. However, when noise is uneven, the optimal weighted entropy levels will exceed that from uniform distributions (i.e., H_u), thus values of $\eta_R > 1$ become possible.

Ideally, the source is present in the detector signals at all times because there is little point to collecting data unless it contains the information on the source itself. If the source presence is structured (e.g., can be described as a smooth function of the configuration coordinates $\mathbf{x}$), some form of the uncertainty principle (eqs. [3.20], [3.21]) will be applicable to it, relating the volumes it can occupy in the spectral representation. Nonetheless, observing patterns which distribute source signals in a randomly can span both domains uniformly (see Section 3.2.3). Therefore, the most robust observing patterns will be those that are either entirely randomized (see also Comito & Schilke 2002), or have a random appearance (i.e., quasi-random).

3.6.4 Feasibility

Strictly speaking, it is possible to identify the absolute best i.e., maximally optimal, distribution of the source signals, for the desired combination of fidelity, sensitivity and robustness. One can always find, in a mathematical sense, the observing pattern(s) which distribute signals in such an ideal way among the components of a representation, by cleverly arranging the samples in configuration space. In theory, there is always an optimal pattern. Detectors, however, are subject to physical constraints of movement (e.g., detectors can be shifted along smooth trajectories, in space, or other domains, while arrays often lack the freedom if individualized detector movement and can be translated only as a whole), imposing limits as to what patterns can be implemented. Often, the mathematical optimum cannot be realized physically.

Instead, one is typically left with an approach based on trial and error. Useful observing patterns, within the physical limits of the experiment, may be suggested from guiding principles (e.g., cross-linking requirement or scanning speed limits). These can be compared objectively using the various criteria (or combination thereof) described above. In this way, a 'best' observing strategy can always be identified from a pool of suggested ideas. How close can such a strategy come to the (unknown) realizable optimum is a matter limited by imagination only.

Here are some of the guiding principles that may aid finding good data collection strategies. The more robust patterns are generally successful under most noise conditions. As a general principle, higher pattern entropies are achieved with observing strategies that move the source signal rapidly across a wide range of detector channels in all available directions. The faster the movement, and the more uniformly the source is distributed across the detector channels, the better source reconstruction is typically achieved (c.f. Weferling et al. 2002). Patterns which are less predictable (more random) tend to yield higher entropies (c.f. Comito & Schilke 2002). It should also be plainly obvious, that when noise is not uniform in the configuration space of detectors (e.g., some detectors, or some periods of observing are noisier than others) then it is best to avoid sampling the source signal in the noisiest configurations (e.g., by bad pixels or in bad weather).

3.6.5 Balanced Strategies

Observing patterns can be optimized for best sensitivities or the highest level of robustness (with some built-in regard for sensitivities also). In practice, however, the observer should be able to simultaneously maximize both, with some adjustable level of importance assigned to each condition. Therefore, consider a Lagrangian construct $\mathcal{L}$, composed of the robustness and sensitivity measures as,

$$\mathcal{L} = \eta_R + \lambda \eta_S, \tag{3.74}$$

where λ is a positive scalar expressing the relative importance of providing sensitivity vs. robustness. Examining the suggested expressions in eq. (3.73) and eq. (3.69) for both components, one can equivalently express the very same goal, in a simplified, albeit unnormalized, form as

$$\mathcal{L} = \sum_i w_i p_i (\Lambda - \ln p_i), \tag{3.75}$$

for some positive Λ value which is related to λ appearing in the expression before. Observing strategies which maximize $\mathcal{L}$, subject to feasibility constraints, will yield the desired combination of sensitivity and robustness.

3.6.6 Stationary and Metastationary Noise and Patterns

Because stationary noise calls for a spectral description, the above measures of observing patterns are best evaluated in Fourier domains. Often though, noise or the observing patterns are nonstationary. Then data may be subdivided into smaller bits, each of which is approximately stationary in nature (hence the meta-stationary terminology). The full observing pattern $\mathcal{M}$ is then the combined set of the sub-patterns $\mathcal{M}^m$ i.e.,

$$\mathcal{M} = \{\mathcal{M}^m : m = 1...N^m\}. \tag{3.76}$$

Accordingly, instead of a single spectrum, one has to consider a set of N^m spectra in total, together which provide the full spectral representation of the data. Summation over all components i, therefore, translates to summing over all frequency channels f of all spectra m as,

$$\sum_i X_i \longrightarrow \sum_m \sum_f X_f^m. \tag{3.77}$$

Such subdivision is often natural, as data is not always taken in a single contiguous integration, rather it is sometimes composed from several disjoint observations that can span days, months or years—noise conditions and instrumental configurations during which can change, not mentioning the particulars of the observing strategies that may be used at different times. In this sense, the concept of a single spectrum adequately describing an entire dataset is already unrealistic. When the nonstationary aspects of an observation arise at time scales shorter than the typical contiguous observation block, the 'scissoring' of the data into shorter observing bits becomes somewhat arbitrary. Nonetheless, it should never be too strange to regard any observation as a set of shorter integrations. However, if such sub-divisions are allowed, one should make certain that sub-dividing of stationary patterns (into smaller bits) will leave the pattern characterization (e.g., the sensitivity and robustness measures) uniquely determined. This is veryfied for the η_S and η_R sensitivity and robustness measures derived above.

3.6.6.1 Transforming Patterns and Anisotropic Noise

A specific case of nonstationary observing patterns is a set of observing patterns $\widetilde{\mathcal{M}}_m$ whose source spectra can be transformed into one another via invertible linear transformations, or equivalently into some particular pattern $\widetilde{\mathcal{M}}$ by the appropriate set of invertible linear transformations $\mathbf{T}_m$ as,

$$\widetilde{\mathcal{M}}_m = \mathbf{T}_m \cdot \widetilde{\mathcal{M}}. \tag{3.78}$$

Then, $\widetilde{\mathbf{S}} = \widetilde{\mathbf{T}}_m^{-1} \cdot \widetilde{\mathbf{S}}_m$, i.e. the source spectra from the various bits can be stacked once all spectral bits are transformed with $\mathbf{T}_m^{-1}$. In this way, the spectral bits behave like repeated measurements and, therefore, the combining of the spectral bits, using noise weights, into a statistical spectral estimate is appropriate (see Section 3.3.3.2).

$$\langle \widetilde{S}_f \rangle = \frac{\sum_m \left[\widetilde{\mathbf{T}}_m^{-1} \cdot \mathbf{w}_m \widetilde{\mathbf{S}}_m \right]_f}{\sum_m \left[\widetilde{\mathbf{T}}_m^{-1} \cdot \mathbf{w}_m \right]_f}. \tag{3.79}$$

(Here, the summation over the frequency components of $\mathbf{T}_m^{-1}$ were left implicit.) Consequently, the noise properties of the combined spectra follow from the noise structure of its components, and can be expressed by the transformed sum of weights as,

$$\mathbf{w} = \sum_m \widetilde{\mathbf{T}}_m^{-1} \cdot \mathbf{w}_m. \tag{3.80}$$

I.e., when the noise of a transformed component is independent among the various spectral bits (true for noise that is independent over the time scales between the observing bits), then the uncertainty of the stacked spectrum is expressed as $1/w_f$ i.e.,

$$\sigma^2_{\langle \widetilde{S}_f \rangle} = \frac{1}{\sum_m \left[\widetilde{\mathbf{T}}_m^{-1} \cdot \mathbf{w}_m \right]_f}. \tag{3.81}$$

This result is particularly useful when the observing patterns, which are geometric transforms of one another. These can be amplitude scalings, a scaling of the scanning speeds or rotation of the pattern along the physical spatial coordinates of a multidimensional detector, many of which are easily implemented in an experiment. The most obvious application of the above result is when noise is anisotropic i.e., it has spatial orientation. This condition may, therefore, apply to multi-dimensional detector arrays (e.g., SHARC-2). Then, signal-to-noise ratios can be improved for those high noise (low weight) spectral components that can be transformed (e.g., rotated) into low noise (high weight) components, and vice versa.

3.7 Conclusions

We showed data collection strategies may be optimized to provide better isolation of source signals from noise. The best strategies simultaneously provide nearly optimal sensitivities and robustness against unforeseen disturbances in the measurement process. Since stationary noise is characteristic to many experiments, strategies are compared by the spectral distributions of the source signals that they provide. We provide normalized measures of both sensitivity (η_S in eq. [3.69]) and robustness (η_R in eq. [3.73]) which can be combined to serve as objective comparison in evaluating the usefulness of possible strategies under the anticipated noise conditions.

When either the noise or the observing pattern is only meta-stationary i.e., can be approximated by a set of stationary bits, one needs to consider a set of spectra instead of just a single spectrum. If noise is anisotropic (i.e., directional), data analysis can benefit from the reconfiguration of observing patterns, such that the directional noise affects different source components each time.

Once a successful observing strategy is identified, the data has to be carefully analyzed, using gain knowledge, noise weighting and filtering, such that the weak signals are reconstructed from the data to the desired combination of achievable fidelity and completeness. The critical aspects of low signal-to-noise data analysis are:

1. Accurate knowledge of signal and noise gains is required, according to,

$$\left|\frac{\delta g_i}{g_S}\right| \ll \frac{\langle|S|\rangle}{\langle|N_i|\rangle}. \tag{3.82}$$

   When multiple distinct gain stages are present under the above condition, analysis must be performed sequentially, in at least as many steps as there are distinct noise associated gain stages, in order to provide the most effective noise isolation. If knowledge of the rejected noise components is also critical, additional attention must be given to filtering order, with the more dominant noise sources filtered first as a general principle.

2. Sufficient separation of noise and source signals is achieved through filtering. Low signal-to-noise optimal filters are derived from simultaneously requiring the preservation of source signals while effectively reducing the noise levels by application of the filter, yielding a family of modified Wiener filters of the form,

$$\Phi(\mathbf{f}) = \frac{|\widetilde{S}(\mathbf{f})|^2}{|\widetilde{S}(\mathbf{f})|^2 + \lambda^2 |\widetilde{N}(\mathbf{f})|^2}. \tag{3.83}$$

   In comparison, noise whitening has the advantage of rendering noise residuals independent in the time stream, allowing simple $1/\sigma^2$ type noise weights in that domain (facilitating the inversion step), and of not requiring prior knowledge of the source structure in its design. Moreover, noise whitening is nearly optimal (under specific λ parameters), for the spectral components with the highest levels of excess noise ($|N_f| \gg \sigma_w$). Thus, noise-whitening is practical for the prefiltering step, while more optimal filters may be used postinversion (cf. eq. 3.8).

3. Noise weights, estimated from the time-stream covariance matrix, are useful in performing the statistical inversion step. When noise is Gaussian, or more generally independent, the noise-weighted averages will minimize χ^2 and will provide maximum-likelihood estimates of underlying quantities. Covariance matrices may be estimated from the data themselves in the presence of stationary noise.

When noise properties of the experiment are not provided externally to the data (e.g., by separate noise measurements) but have to be determined from therein, some number of iterations of the weighting–filtering cycle may be required to arrive at optimal weights and filtering.

In conclusion, in the observing, as in the analysis of data, it is the understanding of limiting noise which provides the key to a successful experiment.

Bibliography

Bester, M., Danchi, W. C.. Degiacomi C. G., Greenhill, L. J., & Townes, C. H. 1992, ApJ, 392, 357

Clark, B. G. 1980, A&A, 89, 377

Cooley, J. W., & Tukey, J. W. 1965, Math. Comput. 19, 297

Comito, C., & Schilke, P. 2002, A&A, 395, 357

Dowell, C. D., et al. 2003, SPIE Conf. Proc. Vol. 4855, Millimeter and Submillimeter Detectors for Astronomy, ed. T. G. Phillips, & J. Zmuidzinas, 73

Dupac, X., Tauber, J. 2005, A&A, 363

Emerson, D. T., & Graeve, R. 1988, A&A, 190,353

Emerson, D. T. 1995, ASPC, 75, 309

Frigo, M., & Johnson S. G. 2005, in Proceedings of the IEEE 93 (2), Special Issue on Program Generation, Optimization, and Platform Adaptation, 216

Gautier, T. N., III, Boulanger, F., Perault, M., & Puget, J. L. 1992, AJ, 103, 1313

Johnstone, D., Wilson, C. D., Moriarty-Schieven, G., Giannakopoulou-Creighton, J., & Gregersen, E. 2000, ApJS, 131, 505

Miville-Deschênes, M.-A., Lagache, G., & Puget, J.-L. 2002, A&A, 393, 749

Narayan, R., & Nityananda, R. 1986, ARA&A, 24, 127

Nyquist, H. 1928, Trans. AIEE, 47, 617

Padilla, N. D., & Baugh, C. M. 2003, MNRAS, 343, 796

Press W. H., Flannery B. P., & Teukolsky, S. A. 1986, Numerical Recipes. The Art of Scientific Computing (Cambridge: University Press)

Shannon, C. E. 1949, Proc. Intitute of Radio Engineers, vol. 37, no.1, pp. 10-21

Tegmark, M. 1997, PhRvD, 56, 4514

Walden, A. T. 1988, Geophysics, 53, 186

Weferling, B., Reichertz, L. A., Schmid-Burgk, J., & Kreysa, E. 2002 A&A, 383, 1088

Chapter 4

CRUSH: A Linearly Scalable Implementation of Weak Signal Analysis via a Sequence of Estimators

4.1 Introduction

The preceding chapter showed that weak signals can be observed optimally in the presence of non-gaussian noise, then analyzed through a combination of noise weights, filters, and accurate gain knowledge using both the time-series and spectral representations of data. The focus of the discussion here is to demonstrate how a particular implementation—analysis of the signals in the time domain alone, through a sequence of statistical estimators,—can provide quasi-optimal data analysis in presence of $1/f$ noise (and possibly other colored noise sources).

This approach has been implemented in the *CRUSH*[1] software (by the author) for reduction of total-power data from the SHARC-2[2] bolometer array at the Caltech Submillimeter Observatory (CSO), providing maps with a sensitivities, clarity and detail that has been almost unprecedented in the field submillimeter astronomy with the simpler, more rudimentary approaches that were used previously (e.g., the differencing of signals from position-switched observations, see Emerson 1988, 1995; Johnstone et al. 2000).

As such, this chapter is also a summary of the mathematical basis of *CRUSH* providing a general understanding of the effect of various configurations from a mathematical point of view. The discussion is offered in generality with application to other similar data analysis challenges, some of which may be in unrelated fields. For this reason, the specifics relating to an astronomical implementations are discussed separately, in the Appendix B, while more information on the SHARC-2 specifics can be found online.[3]

---

[1] Comprehensive Reduction Utility for SHARC-2.

[2] The second generation of Submillimeter High Angular Resolution Camera (Dowell et al. 2003).

[3] `http://www.submm.caltech.edu/~sharc/crush`.

Since the method's computation cost and storage requirements scale linearly(!) with data size, it is a strong candidate for providing data analysis for future large bolometer arrays (e.g., SCUBA-2; Audley et al. 2004; Jenness et al. 2004; Gibb et al. 2005). The approach is also presents a fast and effective alternative to existing map-making methods for Cosmic Microwave Background (CMB) experiments (e.g., Tegmark 1997a,b; Ferreira & Jaffe 2000; Dupac & Giard 2002).

4.1.1 Time-Domain Analysis

There are several advantages to performing the analysis of data in the time domain vs. in spectral representations. Unlike the integral Fourier transform for continuous signals, Discrete Fourier Transforms (DFTs) and especially fast Fourier Transforms (FFTs), used for obtaining spectral representations of discrete data, are restricted for particular sample arrangements. Whereas DFTs are generally calculable for samples on a regular grid (i.e., regularly spaced detectors and time samples), these can be extremely slow and computationally very costly (the number of necessary operations increase as $\mathcal{O}(N^2)$ in terms of the number of samples N). Fast Fourier transforms offer computationally more feasible alternative (with operations $\gtrsim \mathcal{O}(N \log_2 N)$) for large data sets, but can be performed exactly only when data has very particular element counts in every direction of a (rectangular) grid. Most commonly this requires powers of 2 for traditional FFTs, although other prime factorizations can be used also. However, performance will drop as the factors involved become larger. Techniques such as re-griding and zero-padding can be used to mold any data into the required shapes and sizes, making the transforms more easily computed. However, such approaches represent an extrapolation from the actual measurements. Thus the resulting transforms become *estimates not exact representations* of data. In case of weak signal analysis, this could mean imperfect filtering resulting from the inexactitude of the resulting spectral estimates.

The other trouble is that even when FFTs are exact (because of a 'fortunate' number of detectors and their arrangement as well as the appropriate number of time samples), these can still take too long to compute. Even with the fastest (powers of 2) FFTs, the computation time requirement grows as $\mathcal{O}(\log_2 N)$ relative to the needs of linear processing with $\mathcal{O}(N)$ operations. As such there will always exist a limiting data size, no matter what amount of computer power is conjured up for the task, where FFTs become forbidding, whereas other (linear) analysis may remain feasible.[4]

4.2 Compound Signals and Models

Suppose, that the observed signal D is composed of several contributing 'komponents' K_m (with indeces $m = 1...M$), which may also include excess noise, the source signals (as one among the

---

[4] Consider that observations of distant galaxies with the SHARC-2 bolometer camera with its 32×12 pixels, typically have 10^8 elements. Similar data from SCUBA-2 would be a thousand times larger with up to 10^{11} elements due it is larger pixel count and the anticipated higher sampling rate (Audley et al. 2004).

K_ms) and noise n. The time stream is the superposition of these i.e.,

$$D_k = \sum_m K_{m,k} + n_k. \tag{4.1}$$

Here each of the components signals K_m may be parameterized by $A_{m,j}$. When the the total signal is the superposition of the various bits that each $A_{m,j}$ produce, the mapping of parameters for each signal component is described by a set of linear transformations $\mathcal{M}_m$, expressed in vector notation as,

$$\mathbf{K}_m = \mathcal{M}_m \cdot \mathbf{A}_m, \tag{4.2}$$

in terms of the component signal vector $\mathbf{K}_m$, and the parameter vector $\mathbf{A}_m$.

From this point of view, the source signal with its particular mapping that is provided by the observing strategy, is just one among the many signal components simultaneously present in the data stream (cf. eq. [4.1]). In principle, excess noise, such as $1/f$ noise, can also be parametrized by the appropriate set of spectral components (e.g., the rejected components $A_i \sim \rho_f \widetilde{D}_f$ of an 'optimal' filter, see Section 4.4), which are mapped into the time stream by the relevant Fourier transforms. The parametrization of noise prior to source signal estimation is, therefore, an equivalent approach to that of noise filtering.

noise modeling ≡ noise filtering

Alternatively, for simplicity of notation, consider the composite parameter vector $\mathbf{A}$, glued together from all the component signal parameters $\mathbf{A}_m$ (including the source parameters) and expressed in the (transposed) row-vector form as,

$$\mathbf{A}^T = (\mathbf{A}_1^T|\mathbf{A}_2^T|...|\mathbf{A}_M^T). \tag{4.3}$$

The corresponding composite mapping matrix $\mathcal{M}$ is assembled accordingly from $\mathcal{M}^m$ as,

$$\mathcal{M} = (\mathcal{M}_1|\mathcal{M}_2|...|\mathcal{M}_M). \tag{4.4}$$

The aggregated parameter vector $\mathbf{A}$ and the matrix $\mathcal{M}$ encapsulating its mapping into the time stream, allow dropping the m signal component indeces, and write the mapping of all parameters (eqs. [4.1] and [4.2] combined) with unparameterized noise $\mathbf{n}$ in the short-hand form,

$$\mathbf{D} = \mathcal{M} \cdot \mathbf{A} + \mathbf{n}, \tag{4.5}$$

From the data D_k, one may try to estimate the underlying parameters $A_{m,i}$ for the component signals by a mirrored set of model parameters $\widehat{A}_{m,i}$,

$$\mathbf{M}_m = \mathcal{M}_m \cdot \widehat{\mathbf{A}}_m. \tag{4.6}$$

The goal of the data analysis is to identify a best fitting set of model components $\widehat{A}_{m,i}$ which yield signal models such that $M_m \to K_m$. The aggregate signal model M is formed from the superposition of the components M_m,

$$M_k = \sum_m M_{m,k}, \tag{4.7}$$

for the measurement data D. The above two equations can be combined like before in short as,

$$\mathbf{M} = \mathcal{M} \cdot \widehat{\mathbf{A}} + \mathbf{n}. \tag{4.8}$$

At times, more than a single combination of parameters can provide the same 'best' solution. This reflects a potential inherent degeneracy among the model parameters. While the parametrization usually arises from physical considerations of the possible signal components (eq. [4.2]), it is possible that different combinations of these can produce identical signal structures in the time stream. This is especially true when data was not observed in an optimal pattern.

With a composite signal model at hand, it makes sense to define the model residuals as the difference between the observed data and the model,

$$\mathbf{R} = \mathbf{D} - \mathbf{M}. \tag{4.9}$$

Data analysis is effective when the total number N_i of parameters is far fewer than the number of measurements N_k from which they are derived. I.e., the dimension of the aggregate parameter vector $\mathbf{A}$ ought to be but a small fraction of the dimension of the data $\mathbf{D}$. In this way, the analysis that characterizes a large data set by relatively few parameters, is a *reduction* of the data.

4.2.1 Pipeline Reduction

In the linear pipeline reduction scheme presented here, the model parameters A_i^m are estimated independently from the residuals, one at a time, in a predetermined order of choice. Since each successive parameter estimate is derives from the residuals left behind by the preceding estimates, any degeneracies among the model parameters are naturally resolved with overlapping signal powers assigned to the first estimated parameter within the degenerate set. This means, that from the perspective of a particular model (e.g., the source), all those models whose estimation precede it act as filters, whereas those that follow in the pipeline will not impact on its solution.

Thus an intuitive way to think of the single parameter estimators that these treats the residuals

R at every turn as if they contained but a single signal component K_m. The corresponding model residuals are already filtered by all the preceding component estimates, affecting the completeness to which K_m may be determined. At the same time, the yet unmodeled signals are treated as if these were mere noise affecting mainly the accuracy of the resulting estimate of K_m but no so much its underlying form. Often, this latter assumption is granted by an inherent separation of the component signals (e.g., source signal isolation by a cleverly designed observing strategy), but it is also reasonable when the signal components with dominant signal power are estimated first, such that the remaining signal components amount to a small perturbation on that estimate (cf. eq. [3.7]).

A natural modeling order of the signal components is often suggested by the relative signals strengths of these, with the stronger signals typically preceding the weaker ones in the estimation. However, in determining the optimal filtering (modeling) order, further considerations between completeness and sensitivity (cf. Section 3.4) of every model estimate must play a role. As such, the final word on ordering of model components is that of the observer, who must determine that based on the required level of filtering that the models should impose on one another (see Section 4.4). In this way, the filtering of the source model is clearly the most critical, although other models may additionally be sensitive in the reduction (see Appendix ch:guide on absolute gains and direct line-of-sight transmissivities).

Convergence is achieved as even a rudimentary and noisy estimate of a signal component is better than none at all. By removing these crude models from the time stream one is left with cleaner residuals which, in turn, yield better estimates for the other signal components. The cleaner residuals additionally allow for improved weighting of the data, which then allows the further refinement of the signal models etc., creating a virtuous estimator cycle. Thus, an iterated pipeline can converge to a solution, where the residuals approach the noise limit (i.e., $R \to n$), exactly as was suggested for any weighting-filtering cycle for a general reduction scheme in Chapter 3.

In this way, the reduction pipeline most generally refers to a certain order of sequential parameter estimates, in which each model parameter A_i can be estimated several times,

$$\widehat{A}_{\mu_1} \to \widehat{A}_{\mu_2} \to \widehat{A}_{\mu_3} \to ... \text{ where } \mu_l \in \{1...N_i\}. \tag{4.10}$$

Each new attempt at the estimation of a given parameter A_i yields a new parameter model generation n denoted as $\widehat{A}_i^{(n)}$, derived through the incremental refinements from the residuals (the hat is used to distinguish estimators $\widehat{A}_i$ from the underlying parameters $\widehat{A}$). I.e.,

$$\{\mathbf{R}, \widehat{A}_i^{(n)}\} \longrightarrow \widehat{A}_i^{(n+1)} = \widehat{A}_i^{(n)} + \delta\widehat{A}_i^{(n+1)}. \tag{4.11}$$

The 0^{th} generation models parameters $A_i^{(0)}$ thus represent the a-priori signal models (if these are independently available), which can be left blank i.e., zero models such that $A_i^{(0)} = 0$. In effect, this is equivalent to applying the direct matrix inversion method via the normal equations (see Section 4.7.1) to each signal component separately, in sequence, yielding incremental refinements as,

$$\delta \widehat{\mathbf{A}}_m \approxeq [\mathcal{M}_m^T \mathcal{M}_m]^{-1} \mathcal{M}_m^T \cdot \mathbf{R}. \tag{4.12}$$

A more detailed proof of how the iterated solutions 'improve' is offered in Section 4.4, where the filtering properties of time-series modeling are also discussed, especially in relation to $1/f$ noise.

The idea of signal components being modeled sequentially as opposed to being modeled all at once (e.g., matrix inversion method or singular value decomposition, see Section 4.7), should not come as a surprise. In Chapter 3 we showed that proper filtering of signals, emanating from various different gain stages, are best performed in spectral domain and in turns to avoid unnecessary over-filtering. That alone leads a sequence of filters, more or less equivalently to signal components here.

4.3 Statistical Estimators

When constructing an estimator $\widehat{A}_i$ for an underlying parameter A_i, clearly, the estimate should be representative of the underlying quantity, notwithstanding some noise. I.e.,

$$\widehat{A}_i \longrightarrow A_i. \tag{4.13}$$

Beyond that, the estimate that is best depends on the qualities it is required to have. A number of different statistical estimators can be derived from the critical optimization condition that they are to fulfill. E.g., one may seek estimates which represent the most probable values given the data—these are maximum-likelihood estimates. Or, one may require that it is least dependent on uncharacterized or poorly measured quantities (e.g., maximum-entropy estimates), or not be biased by occasional interference from other sources, such as electronic glitches (robust estimates). Here we discuss these few instances, but many other types of 'optimal' estimates may be constructed from optimization of other properties.

4.3.1 Maximum Likelihood Estimators

The goal of maximum-likelihood modeling is to find the model set $\mathbf{M}$ (i.e., the set of model parameters A_i) which represents the most likely configuration of the underlying signals giving rise to the observed data set $\mathbf{D}$. When noise is independent among the samples the probability $p(\mathbf{I} \,|\, \mathbf{M})$ that the observed data set $\mathbf{D}$ resulted from the underlying processes modeled by $\mathbf{M}$ is calculated as the product of the (independent) probabilities that measurement D_k arose from a model value M_k, over all the relevant data points i.e.,

$$p(\mathbf{D}\,|\,\mathbf{M}) = \prod_i p(D_k\,|\,M_k). \tag{4.14}$$

Furthermore, when the underlying noise distribution is Gaussian, this likelihood is expressed (cf. eq. [3.35]) as,

$$p(\mathbf{D}\,|\,\mathbf{M}) = \prod_k e^{-\frac{|D_k - M_k|^2}{2\sigma_i^2}} = \exp\Big\{-\frac{1}{2}\sum_k w_k\,|D_k - M_k|^2\Big\} = e^{-\frac{1}{2}\chi^2}, \tag{4.15}$$

where χ^2 is defined, in the usual way, according to (eq. [3.36]) as,

$$\begin{aligned}\chi^2 &= \sum_k w_k |D_k - M_k|^2 \\ &= \sum_k w_k \Big|D_k - \sum_i \mathcal{M}_{ik} A_i\Big|^2.\end{aligned} \tag{4.16}$$

The extremum producing condition on the probability is given by,

$$\frac{\partial p(\mathbf{D}\,|\,\mathbf{M})}{\partial A_i} = 0 \;\longrightarrow\; \frac{\partial \chi^2}{\partial A_i} = 0. \tag{4.17}$$

I.e., maximizing the model likelihood, under Gaussian noise, is equivalent to minimizing the χ^2 deviation of the data from the model. Focusing on a single model parameter A_i, the maximum-likelihood condition leads to,

$$\hat{A}_i^{(n+1)} = \hat{A}_i^{(n)} + \frac{\sum_k w_k\, \mathcal{M}_{ik}^*\, R_k}{\sum_k w_k\, |\mathcal{M}_{ik}|^2}. \tag{4.18}$$

Here $\mathcal{M}_{ik}^*$ is the complex conjugate of the appropriate matrix component $\mathcal{M}_{ik}$. Observe, that the typical parameter A_i has limited scope on the data, in the sense that it affects significantly only a small fraction of the entire data set. Thus most components of the matrix column $\mathcal{M}_i$ thus tend to be zero. Therefore, from a computational perspective, it is impractical to perform the matrix multiplication in full when the transformation can be equally well described by the range $\mathcal{R}_i$ of A_i, meaning the set of k indices for which $\mathcal{M}_{ik} \neq 0$, and by the corresponding set of nonzero signal gains $g_{ik} = \mathcal{M}_{ik}$. In this way the the matrix multiplication can be written as,

$$\sum_k \mathcal{M}_{ik} R_k = \sum_{k \in \mathcal{R}_i} g_{ik} R_k. \longrightarrow \sum_{\mathcal{R}_i} g_i R \tag{4.19}$$

To simplify notation, here and throughout, the summation indices k are dropped entirely whenever these appear all terms and are sufficiently implied by the summation range $\mathcal{R}_i$ alone. In this more practical notation, the incremental maximum-likelihood model is then calculated as,

$$\delta \hat{A}_i = \frac{\sum_{\mathcal{R}_i} w\, g_i^*\, R}{\sum_{\mathcal{R}_i} w\, |g_i|^2}. \tag{4.20}$$

now explicitly in terms of the data range of the parameter A_i and the corresponding signal gains g_{ik} replacing the mapping matrix according to eq. (4.19). The uncertainty σ_i^2 in the model parameter A is obtained from $\hat{A}_i \pm \sigma_i$ at which the increase in χ^2 is unity i.e., $\Delta\chi^2 = 1$, and is accordingly estimated as,

$$\hat{\sigma}_i^2 = \frac{1}{\sum_{\mathcal{R}_i} w\, |g_i|^2}. \tag{4.21}$$

Note, that the maximum-likelihood estimate and its variance are identical to the weighted averages of Chapter 3 (eqs. [3.26] and [3.27]).

The Gaussian noise assumption is reasonable for many physical processes. The maximum-likelihood property is further applicable for many instances of independent (i.e., white) noise also, due to the *Central Limit Theorem* (see Section 3.3.2.1) which observes that compound noise distributions tend to be normal. More generally, arbitrary stationary noise can be whitened (i.e., made independent among samples) by the appropriate filtering (cf. noise whitening filters, see Section 3.4.2). In this light, true maximum-likelihood estimates can be derived for a wide variety of data in combination with noise whitening filters. Otherwise, the estimators will simply represent χ^2 minimizing values, which are not necessarily linked to most probable models for the general case.

4.3.2 Maximum Entropy Estimators

Instead of maximizing the probability that the data was produced by a given model, one may seek models that are the least dependent on the missing information. This may be desirable at time to correct for the filtered source components. It is achieved by models with maximal entropy. A definition of entropy used for astronomical images is from (Narayan & Nityananda 1986),

$$\mathcal{H} = -\sum_i A_i \ln \frac{A_i}{T_i\, e}, \tag{4.22}$$

where A is the signal model and T is an a-priori model target. Note, that while this has the familiar form of entropy from statistical mechanics, the entropy here is not a physical quantity, rather it serves as a constraint requiring signals to be smooth around the target model. Clearly, the definition is well behaved only for positive (or negative) definite target and signal models. An enforcement of positivity (e.g., by an appropriate offset, clipping of nonphysical negatives or converting a bidirectional signal X to its norm $A \sim |X|^2$ etc.) should, therefore, precede the maximum-entropy estimation.

The divergence at zero targets is rectified, while the requirement that $\mathcal{H}$ be differentiable at zero is provided, if a small positive constant is is added to both terms, which we designate to be σ_i, the characteristic uncertainty of the model parameter A_i, for convenience. Furthermore, $\mathcal{H}$ is made dimensionless by the normalization of the fluxes, outside the logarithm, by σ_i. Therefore, propose to use

$$\mathcal{H} = -\sum_i \left(1 + \frac{A_i}{\sigma_A}\right) \left\{ \ln\left(\frac{\sigma_i + A_i}{\sigma_i + T_i}\right) - 1 \right\} \tag{4.23}$$

as the model entropy. This expression yields a smooth $\mathcal{H} \to 0$ for signal and target models both near zero ($A_i, T_i \ll \sigma_i$), and will asymptotically retain a scaled version of the original definition (eq. [4.22]) for significant positive (or negative) signal levels ($A_i, T_i \gg \sigma_i$) and uniform noise conditions ($\sigma_i = \sigma_j$ for all i, j). The, somewhat arbitrary, modification in the way we define entropy is allowed since the image entropy has no strict physical interpretation, leaving the exact definition relatively free for choosing.

Entropy is typically maximized subject to the constraint that χ^2 is also near its minimum, i.e. that the model is also reasonably probable (see eq. 4.17). This is achieved by minimizing,

$$X = \chi^2 - 2\lambda\,\mathcal{H} \tag{4.24}$$

for some value of λ. Note, that $\mathcal{H}$ and thus λ are both defined to be dimensionless. The minimization condition $\partial X/\partial A_i \to 0$ leads to,

$$\delta A_i^{\mathrm{MEM}} = \delta A_i^{\mathrm{ML}} - \lambda\,\sigma_A \ln\left(\frac{\sigma_{A_i} + A_i}{\sigma_{A_i} + T_i}\right). \tag{4.25}$$

I.e., maximum entropy method (MEM) incremental solution is a modification of the maximum-likelihood (ML) estimator. In the iterated scheme presented, convergent algebraic solutions are facilitated if the right-hand side is evaluated using the data model from the previous iteration.

Here, we observe, that $\lambda = 0$ retains the maximum-likelihood solution (discussed above), while the larger λ values assign increasing prominence to the maximum-entropy condition relative to the χ^2 minimizing one. In the limit of $A - T \ll T, A$ i.e., solutions that are close to the a-priori target, the incremental solution is approximated by,

$$\delta A_i^{\mathrm{MEM}} \approx \delta A_i^{\mathrm{ML}} - \lambda\,\sigma_A \frac{A_i - T_i}{\sigma_{A_i} + T_i}, \tag{4.26}$$

which, in the low signal-to-noise limit, simplifies to

$$\delta A_i^{\mathrm{MEM}} \approx \delta A_i^{\mathrm{ML}} + \lambda\,(T_i - A_i). \tag{4.27}$$

In this limit, therefore, $\lambda = 1$ will essentially reproduce the preset target signal $A_i \to T_i$, as convergence ($\delta A^{\mathrm{ML}} \to 0$) is achieved. As useful solutions tend to lie between the target model T and the maximum likelihood solution A^{ML}, for all signal levels, λ values between 0 and $\simeq 1$ are typically useful for obtaining practical maximum-likelihood solutions.

In the high signal-to-noise limit, where targets and signals have the same sign, and solutions are also close to the target (see above) we find the approximation,

$$\delta A_i^{\mathrm{MEM}} \approx \delta A_i^{\mathrm{ML}} + \frac{\lambda}{\mathrm{SNR}} (T_i - A_i). \tag{4.28}$$

That is, the highly significant signals of the solution will approach the maximum likelihood (χ^2 minimizing) limit, while the weak signals will be closer to the a-priori model.

Typical target models can be uniform value models ($T_i = const.$), which will attempt equalization of the fluxes among all model parameters, with zero value models ($T_i = 0$) serving a special case. Alternatively target models can be constructed as a low-resolution versions (e.g., by smoothing) of the maximum-likelihood solution, if the target is to be less predetermined.

Since maximum-entropy estimators are obtained as a correction to maximum likelihood ones, the uncertainties can be estimated the same way as they were for maximum-likelihood estimators (eq. [4.21]).

4.3.3 Robust Estimators

Maximum-likelihood estimators, and hence also maximum-entropy estimators to a certain degree, are most strongly affected by data that lie far from the distribution mean. Occasional glitches, or noise spikes, may occur in any data stream (e.g., cosmic rays, static discharges), and can produce strong biases in these estimates. Nor does the tail clipping of the distributions, at which measurements are expected to be rare, will offer comfortable relief, since no matter how naturally rare extreme values may be, they should be included in truly unbiased estimates. The clipping of any naturally occurring noise tails will also lead to a bias in the those statistical estimates. In this sense, neither maximum-likelihood or maximum-entropy estimators are robust.

There are ways of characterizing distributions in more robust ways. Using medians instead of averages is an example of a robust estimation, as medians are much less affected either by the presence or absence of outliers, no matter how far from the true distribution mean they may lie. For symmetric distributions, medians, not averages, will provide the more reliable estimates of the underlying distribution mean. However, in what way should medians be calculated for weighted data i.e., one in which measurements may have wildly different uncertainties, such that they provide the truest possible estimate of the underlying distribution?

The median value is effectively defined, for uniform noise data, as the value above which lies half of the data and below which is the other half. In an ordered set of the data, the median is simply the middle point for a odd number of samples, and it is defined to be the arithmetic mean of the middle two measurements for even number of data.

This can be translated to weighted data by equalizing the left and right weight integrals around the median value. That is, identifying the i^{th} value X_i, for which

$$\sum_{k}^{X_j<X_i} w_j \approx \sum_{j}^{X_j>X_i} w_j, \tag{4.29}$$

or, in terms of the sorted values $X_\beta \geq X_\alpha$ for $\beta \geq \alpha$ (with the Greek letter indeces belonging to sorted sets), the median value is X_μ is given by the index μ for which,

$$\sum_{\alpha=1}^{\mu-1} w_\alpha \approx \sum_{\alpha=\mu+1}^{N_X} w_\alpha, \tag{4.30}$$

Under approximate equality, we mean that the median is the weighted measurement with index μ, for which the left and right sums are the closest. While there are several examples of equivalent definitions in literature, one may go a step further in refinement, by approximating the sorted values X_α by a monotonous function f of the integrated weights W of the sorted set, such that,

$$X_\alpha \to X(W_\alpha) \to f(W). \tag{4.31}$$

where, $W_\alpha = \sum_{\beta<\alpha} w_\beta + 1/2\, w_\alpha$ is the integrated weights assigned to data that is lesser or equal to X_α. The median value is then identified by the condition,

$$\int_0^{W_{\mathrm{med}}} dW\ f(W) = \int_{W_{\mathrm{med}}}^{W_{\mathrm{max}}} dW\ f(W) \tag{4.32}$$

A simple choice for $f(W)$ is one for which $f(W_\alpha) = X_\alpha$, and is connected by linear segments in between. Then, if the median lies between the sorted values X_α and $X_{\alpha+1}$, the weighted median is given by,

$$\hat{X}_{\mathrm{med}} = \left(\frac{\sum_{\beta<\alpha+1} w_\beta - \sum_{\beta>\alpha+1} w_\beta}{w_\alpha + w_{\alpha+1}} \right) X_\alpha + \left(\frac{\sum_{\beta>\alpha} w_\beta - \sum_{\beta<\alpha} w_\beta}{w_\alpha + w_{\alpha+1}} \right) X_{\alpha+1}. \tag{4.33}$$

When estimating a parameter A_i, the X should be replaced by the single point estimator of A_i from the residual R_k. Assuming the residuals are a good (if noisy) representation of the signals resulting from A_i i.e., $R_k \approx \mathcal{M}_{i,k} A_i$ this is produced by $X_j \to R_k/g_{i,k}$ for $k \in \mathcal{R}_i$, and therefore the corresponding weights $w_j \to w_k |g_{i,k}|^2$ should be used in concert.

As for estimating the uncertainty in the median value, let us point out that for the case of measurements characterized by symmetric probability distributions and no spurious outliers, both medians and averages are equivalent estimators of the distribution mean, and therefore, these should come with identical properties for the same data set. Hence, using eq. (4.21) for the uncertainties is reasonably justified.

4.4 Filtering Properties

The modeling of noise through estimators, like those described above, effectively represents a time-domain filter, as the parametrized noise components are removed from the time stream. The critical question that remains is how such filtering compares to the spectral filters which were described earlier in Chapter 3.

Consider the somewhat special, albeit common, case when the successive parameters $A_{m,j}$ of a noise signal component K_m map into the time stream via regularly spaced blocks which are closely resemble one another (i.e., the matrix columns $\mathcal{M}_{m,i} \sim \mathcal{M}_{m,j}$ but offset on a regular grid). More specifically, in the spirit of describing the mapping of parameters $A_{m,j}$ via their data range $\mathcal{R}_{m,j}$ and the corresponding nonzero mapping matrix elements or gains, consider $\mathcal{R}_{m,j}$ which are all just shifted versions of the representative range $\mathcal{R}_m$, laid out in an grid with regular $\mathbf{\Delta}_m$ offsets in configuration space i.e.,

$$\mathbf{x}_{l\in\mathcal{R}_{m,j}} = \mathbf{x}_{k\in\mathcal{R}_m} + (j-1)\mathbf{\Delta}_m \tag{4.34}$$

Then, the regular spacing of the representative range $\mathcal{R}_m$ is analogous to moving data windows W_m of the same 'shape,'

$$W_m(\mathbf{x}) \sim W_{m,k} \propto \begin{cases} 1 & \text{if } k \in \mathcal{R}_m \\ \\ 0 & \text{otherwise.} \end{cases} \tag{4.35}$$

The incremental maximum-likelihood estimators of eq. (4.18) invoke the (discrete) gain inverted weighted convolution (eq. [3.14]) of the residual signal with the window function $W_m(\mathbf{x})$. In fact, the incremental estimators $\delta\widehat{A}_{m,j}$ can be identified as regular samples at intervals of $\mathbf{\Delta}_m$ of the convolved product, i.e.,

$$\delta\widehat{A}_{m,j} \propto \frac{(W_m \otimes R)_l}{g_{i,l}} \tag{4.36}$$

for data indeces l which satisfy the constraint of eq. (4.34) in terms of j.

Since the modeled noise components are also the rejected noise components, it follows that $\widetilde{\rho}(\mathbf{x}) \propto W_m(\mathbf{x})$ and thus directly related to $\mathcal{R}_m$, while comparison to eq. (4.19) yields the configuration space convolution coefficients $\widetilde{\rho}_k$ of the rejection as,

$$\widetilde{\rho}_{m,k} = \begin{cases} \dfrac{1}{N(\mathcal{R}_m)} & \text{if } k \in \mathcal{R}_m \\ \\ 0 & \text{otherwise.} \end{cases} \tag{4.37}$$

where $N(\mathcal{R}_m)$ is the number of data elements in the representative range $\mathcal{R}_m$. The spectral filter components ρ_f are simply the direct (unweighted) Fourier transform of $\widetilde{\rho}_k$ which are applied multiplicatively to the weighted transform of (R/g).

While the above argument is somewhat abstract, it provides a very useful result. The shifted time windows such as those described above, can be used for parameterizing stationary noise. By reversal of the above argument, any spectral rejection function $\rho(\mathbf{f})$, and therefore filter pass function $\Phi(\mathbf{f})$, can be realized by assigning an appropriate representative range $\mathcal{R}_m$ and gains to the modeled parameters $A_i \sim \rho(\mathbf{f})$. The mapping $\mathcal{M}$ of the parameters is constructed by pasting those gains at regularly offset locations among the matrix columns. In this way, the time-domain estimators can always provide the desired filtering of the data.

Most importantly, when a set of parameters $A_{m,j}$ are estimated from nonoverlapping data windows (or parameter ranges $\mathcal{R}_{m,j}$) the resulting estimates are both *independent* and their computation costs grow *linearly* with data size. As such, it is desirable to choose parametrization of every signal component K_m such that its components $A_{m,j}$ can be estimated from distinct data regions $\mathcal{R}_{m,j} \sim \mathcal{R}_m$ for the various indeces j.

$$\mathcal{R}_{m,i} \cap \mathcal{R}_{m,j} = \delta_{i,j} \mathcal{R}_{m,i} \tag{4.38}$$

Then, the *independent* component parameters $A_{m,j}$ can be derived uniquely, through $\mathcal{O}(N_k)$ operations, and in arbitrary order (for the index j), as long as in a single block covering the entire set of these i.e., the full parameter vector $\mathbf{A}_m$ is estimated in a single go. Because such estimates of $\mathbf{A}_m$ by its independent individual components are both unique and optimal (most probable, most disordered or least affected by spurious measurements—depending on the type of estimator used) these estimates *must be* equivalent estimates to those provided by the submatrix inversion methods via normal equations i.e.,

$$\widehat{\mathbf{A}}_m \approxeq [\mathcal{M}_m^T \mathcal{M}_m]^{-1} \mathcal{M}_m^T \cdot \mathbf{R}. \tag{4.39}$$

for one signal component at a time (not to be confused with the full matrix inversion via the composite $\mathcal{M}$!). As such, the two are effectively interchangeable pictures of the pipeline reduction.

Since each of the different components K_m tends to affect all measurements, but via different shifted windows $\mathcal{R}_m$, the various parameters $A_{m,j}$ and $A_{n,i}$ of the separate component signals

K_m and K_n are generally not independent. Thus, the ordering among the estimation of the full parameter vectors $\mathbf{A}_m$ and $\mathbf{A}_n$ will matter. As such the data reduction is just a sequence of the parameter vector estimates for the signal components, where every parameter vector $\mathbf{A}_m$ may be estimated and re-estimated a number of times, until the desired 'optimal' filtering of the source signals ($\mathbf{S} = \mathbf{A}_m$ for some m) is achieved.

$$\mathbf{A}_{\mu_1} \rightarrow \mathbf{A}_{\mu_2} \rightarrow \cdots \rightarrow (\mathbf{A}_{\mu_p} = \mathbf{S}) \tag{4.40}$$

While, much of the above was based on the go-between argument for spectral filters and time domain estimates which was derived for the maximum-likelihood estimators, it is possible to generalize the argument to all other types of equivalent estimators.

4.4.1 Filter Evolution and Convergence

The above description of filtering is applicable only when each parameter is derived only once in the analysis. When modeling (i.e., filtering) is iterated instead, there maybe nontrivial interactions among the parameters, which alter the effective filtering properties with each new model set $\mathbf{A}$ generation. Consider the simple scenario, when the signal is composed of only two components, the source $\mathbf{S}$, and excess stationary noise $\mathbf{N}$ which here includes white noise term $\mathbf{n}$ from eq. (4.5) i.e.,

$$\mathbf{I} = \mathcal{M} \cdot \mathbf{S}_{obs} + \mathbf{N}. \tag{4.41}$$

The mapping of $\mathbf{N}$ is simply via the backward Fourier transform of the relevant independent components $\widetilde{N}_f$, leaving only the source mapping (i.e., observing pattern) written out explicitly via $\mathcal{M}$. Suppose that the correct weights and accurate gain knowledge are also granted, and that both $\mathbf{S}_{obs}$ and the parameters of $\mathbf{N}$ are estimated from respective sets of nonoverlapping data windows such that all parameters of $\mathbf{S}_obs$ are estimated together (in any order) and so are the components of $\mathbf{N}$.

In constructing the source estimator, one ought to consider what that estimator should be were the residual signal is the source structure $\mathbf{S}_{obs}$ alone. The signal is then inverted for the source structure using the reverse mapping $\mathcal{A}$ (eq. [3.5]),

$$\widehat{\mathbf{S}}_{obs} \approxeq \mathcal{A} \cdot \mathbf{R}, \tag{4.42}$$

which is an equivalent representation of what the time sequence of source estimators from nonoverlapping data windows would achieve, and also generally applicable to all methods of data reduction under linear mapping (e.g., matrix inversion via normal equations or SVD, see Section 4.7). A mapping can be expressed equivalently from the residual signal spectrum $\widetilde{\mathbf{R}}$ instead of the time-series residuals, by $\widetilde{\mathcal{M}}$ which takes source structures directly into the Fourier representation of the measured signals. Clearly, this requires that the columns of $\widetilde{\mathcal{M}}$ be the Fourier transforms of the corresponding columns of $\mathcal{M}$. Consequently, there is also such a $\widetilde{A}$ for the reverse mapping from Fourier space, with rows being the Fourier transforms of those of $\mathcal{A}$. As such, an equivalent expression of approximate equivalence above, but for Fourier spectra, is written as,

$$\widehat{\mathbf{S}}_{obs} \approxeq \widetilde{\mathcal{A}} \cdot \widetilde{\mathbf{R}}. \tag{4.43}$$

In a similar line of arguing, were the residual are entirely stationary noise i.e., its spectrum is that of the noise components $\widetilde{R}_f = N_i$. The trivial inversion for the (spectral) noise parameters $\widehat{N}_i \approxeq \widetilde{R}_f$ is suggested. However, for the case when both noise and source signals are simultaneously present in the spectrum, one clearly does not wish to assign all data as noise (that would be plain silly!), as such a totalistic noise estimator would tend to suggest. Instead, one should like to characterize the relevant noise signals by just a few components (e.g., those with excess noise levels), such that nearly all degrees of freedom are retained. In general, the dataset should be much larger than the number of model parameters that are derived from it i.e., $N_k \gg N_A$. For stationary noise, it makes sense to use the discrete rejection components ρ_f of the spectral filter. These are provided by the appropriate nonoverlapping time series windows for the sequence of estimators. This argues for the effective description of excess noise estimation in Fourier space, via the spectral parametrization,

$$\widehat{\mathbf{N}} \approxeq \boldsymbol{\rho} \cdot \widetilde{\mathbf{I}}. \tag{4.44}$$

Here the total effective number of noise model parameters is $\sum_f |\rho_f| \ll N_k$ i.e., $\rho_f \approx 0$ for most spectral channels f.

Alternating the estimators for source structure and noise components for the case composite signals, one can summarize the solutions for the n^{th} iteration as an incremental adjustment to the prior model generation through re-estimation from the up-to-date model residuals,

$$\begin{aligned} \mathbf{S}_{obs}^{(n)} &= \mathbf{S}_{obs}^{(n-1)} + \widetilde{\mathcal{A}} \cdot \widetilde{\mathbf{R}} \\ \widetilde{\mathbf{N}}^{(n)} &= \boldsymbol{\rho} \cdot \big(\widetilde{\mathbf{N}}^{(n-1)} + \widetilde{\mathbf{R}}\big), \end{aligned} \tag{4.45}$$

where the bracketed indeces (n) are used to mark model generation in the iterated scheme (with redundant estimator symboling via hats or "est" subscripts now dropped). The residuals $\mathbf{R}$ need to be recalculated between the two sets to correspond to the difference between the measurements and the modeled signals as,

$$\mathbf{R} = \mathbf{I} - \mathcal{M} \cdot \mathbf{S}_{obs}^{(m)} - \mathbf{N}^{(n)} \tag{4.46}$$

for the current generations of source and noise models i.e., $\mathbf{S}_{obs}^{(m)}$ and $\mathbf{N}^{(n)}$ respectively.

Let the initial models be $\mathbf{N}^{(0)}, \mathbf{S}_{obs}^{(0)} = \mathbf{0}$, i.e., make no prior assumptions on the structures present. Clearly, if such assumptions, or educated guesses, can be made, the following discussion remains valid, except that $\mathbf{S}_{obs}$ and $\mathbf{N}$ are then interpreted as the respective differences between the actual source and spectral noise components and their assumed forms.

With the two interrelated estimators at hand, the choice remains on which one to evaluate first. Accordingly, we discuss both possibilities. The results can be extended to arbitrary orderings of various stationary noise estimates some of which precede while others follow the source estimation. The interaction of each of these with the source model is provided by the relative placement of that and the source estimation in the analysis pipeline, while the interaction among the various noise estimates is generally inconsequential to the solution.

4.4.1.1 Source Estimation First

When source estimation, precedes the estimation of noise signals, one immediately arrives to the stable solution at the first evaluation of the estimators i.e.,

$$\begin{aligned} \mathbf{S}_{obs}^{(n)} &= \mathbf{S}_{obs} + \mathcal{A} \cdot \mathbf{N} \\ \widetilde{\mathbf{N}}^{(n)} &= \boldsymbol{\rho} \cdot (\mathbf{1} - \widetilde{\mathcal{M}}\widetilde{\mathcal{A}}) \cdot \widetilde{\mathbf{N}}. \end{aligned} \tag{4.47}$$

Further iterations are not necessary as these produce identical solutions. As expected, the source estimate will be unfiltered (i.e., complete), but will carry the footprint of all the noise from the raw data. This is effectively the the naïve map reconstruction method resulting from the simple reverse mapping of the raw data (cf. eq. [3.5]), and is generally not suitable for weak signals, although it may be of use when the source signal $\mathcal{M} \cdot \mathbf{S}_{obs}$ is dominant relative to N.

In a strict sense, the second step of noise estimation is unnecessary, unless one desires to retain (filtered) knowledge of these (e.g., for gain estimation, see Section 4.6.1), or wishes to continue to clean-up the time stream (e.g., for better re-estimation of weights, see Section 4.5).

For simplicity of interpretation, consider the special case when $\mathcal{M}$ is both stochastic and ergodic (cf. robust observing strategies of Section 3.6.3), such that any source component S_i is randomly mapped into the time streams, hence also into all spectral channels $\widetilde{I}_f$, then all derivatives of $\mathcal{M}$, including $\widetilde{\mathcal{M}}$, $\mathcal{A}$, $\widetilde{\mathcal{A}}$, etc. will also be stochastic and ergodic. The matrices $\mathcal{A}$ and $\widetilde{\mathcal{A}}$ then effectively map the time series and spectral noise components into random map noise. Provided that the mapping $\widetilde{\mathcal{M}}$ conserves signal power (which one should hope!) we can write

$$\mathcal{A} \cdot \mathbf{N} \rightsquigarrow \frac{N_k}{N_S} \mathbf{n}(\mathbf{N}), \tag{4.48}$$

where $\mathbf{n}(\mathbf{N})$ is a Gaussian noise vector with the same integrated noise power as $\mathbf{N}$ (i.e. $\sum_i |n_i|^2 = \sum_i |N_i|^2$). Therefore, for stochastic, ergodic mapping we can approximate the solutions of the iterated scheme as,

$$\mathbf{S}_{obs}^{(n)} \rightsquigarrow \mathbf{S}_{obs} + \frac{N_k}{N_S}\mathbf{n}(\mathbf{N}) \tag{4.49a}$$

$$\widetilde{\mathbf{N}}^{(n)} \rightsquigarrow \boldsymbol{\rho}\cdot\widetilde{\mathbf{N}} - \widetilde{\mathbf{n}}(\mathbf{N}) \tag{4.49b}$$

Here, the equivalent spectral noise $\widetilde{\mathbf{n}}(\mathbf{N})$ is the appropriate Fourier transform of $\mathbf{n}(\mathbf{N})$.

4.4.1.2 Noise Estimation First

When noise is dominant, its estimation ought to precede the source, to provide a level of filtering which allows more sensitive, although less complete reconstruction of the source structures (see Section 3.4). Therefore, the more interesting case is when noise modeling precedes the source estimation. Here the solutions can be most easily expressed in terms of the actual source signal spectrum $\widetilde{\mathbf{S}} = \widetilde{\mathcal{M}}\cdot\mathbf{S}_{obs}$ as,

$$\begin{aligned}\widetilde{\mathcal{M}}\cdot\mathbf{S}_{obs}^{(n)} &= (\mathbf{1}-\boldsymbol{\rho}^n)\widetilde{\mathcal{M}}\cdot\mathbf{S}_{obs} + \widetilde{\mathcal{M}}\widetilde{\mathcal{A}}(\mathbf{1}-\boldsymbol{\rho}^n)\cdot\widetilde{\mathbf{N}}\\ \widetilde{\mathbf{N}}^{(n)} &= \boldsymbol{\rho}\cdot\widetilde{\mathbf{N}} - \widetilde{\mathcal{M}}\widetilde{\mathcal{A}}(\boldsymbol{\rho}-\boldsymbol{\rho}^n)\cdot\widetilde{\mathbf{N}} + \boldsymbol{\rho}^n\widetilde{\mathcal{M}}\cdot\mathbf{S}_{obs}.\end{aligned} \tag{4.50}$$

As anticipated, the source signal spectrum is now filtered via the effective spectral filter function on the signals,

$$\boldsymbol{\Phi}^{(n)} = \mathbf{1} - \boldsymbol{\rho}^n \tag{4.51}$$

The noise that makes its way into the source signals is filtered similarly with the same $\boldsymbol{\Phi}$. Correspondingly, the noise model is also messier than before, with extra noise spawned from the rejected source signals, and an effective redistribution of noise flux from the back-and-forth mapping cycle.

A clearer picture can be provided for stochastic, ergodic mapping when solutions can be simplified to,

$$\mathbf{S}_{obs}^{(n)} \rightsquigarrow \boldsymbol{\phi}_{\mathcal{M}}^{(n)} \otimes \mathbf{S}_{obs} + \frac{N_k}{N_S}\mathbf{n}\big([\mathbf{1}-\boldsymbol{\rho}^n]\cdot\mathbf{N}\big) \tag{4.52a}$$

$$\widetilde{\mathbf{N}}^{(n)} \rightsquigarrow \boldsymbol{\rho}\cdot\widetilde{\mathbf{N}} - \widetilde{\mathbf{n}}\big([\boldsymbol{\rho}-\boldsymbol{\rho}^n]\cdot\widetilde{\mathbf{N}}\big) + \boldsymbol{\rho}^n\cdot\frac{N_S}{N_k}\widetilde{\mathbf{n}}'(\mathbf{S}_{obs}). \tag{4.52b}$$

Some immediate conclusions can be drawn either from either the exact or the approximate expressions. For a large number of iterations, the respective source and noise components will converge to (i) those of the naïve solution when the corresponding $\rho_f < 1$, whereas (ii) spectral components with $\rho_f = 1$ will be filtered completely, retaining all corresponding spectral component flux in the noise model. In this way, the convergent limit is either the full preservation or the total rejection of spectral signal components.

$$\widetilde{S}_f^{(\infty)} \longrightarrow \begin{cases} \widetilde{S}_f + \widetilde{N}_f & \text{if } |\rho_f| < 1 \\ \\ 0 & \text{if } \rho_f = 1 \end{cases} \tag{4.53}$$

However, final convergence is not necessarily what one should be looking for here. Instead, observe that the effective filtering (eq. [4.51]) can be adjusted to varying levels through proceeding with either fewer or more iterations, or choosing different estimator windows. The most aggressive filtering is provided at the very first round of (source) estimation, while the effective filtering becoming less the longer one iterates (and more successive source generations are derived). Gradually, increasing levels of noise will make the way into the source estimate, as the filtering becomes more and more relaxed.

4.4.1.3 Meta-Stationary Noise and Transforming Observing Patterns

So far, we have presented a complete assessment of the filtering properties in the iterated scheme for the case of stationary noise and source. Now, consider the case when noise (or equivalently the mapping pattern) is stationary in subsets of the data, which provide the meta-stationary full data. Then, the signal K_m for each stationary subset m is

$$\mathbf{I}_m = \mathcal{M}_m \cdot \mathbf{S}_{obs} + \mathbf{N}_m. \tag{4.54}$$

Each data subset can now have the appropriate source $\mathbf{S}^{(n)}_{obs,m}$ and noise estimators $\widetilde{\mathbf{N}}^{(n)}_m$ according to eqs. (4.45). However, while the noise estimates are mutually independent from set to set (and also component to component!), the observed source structure is the same for all subsets. Therefore, averaging of the source estimates is clearly suggested, in combining the N_m individuals $\mathbf{S}_{obs,m}$ into a single source estimate. For simplicity, let one use a simple arithmetic average here with

$$\widehat{\mathbf{S}}_{obs} = \frac{1}{N_m} \sum_m \widehat{\mathbf{S}}_{obs,m}. \tag{4.55}$$

Furthermore, let us consider the case of transforming observing patterns (Section 3.6.6.1) such that through an appropriate linear transformation all spectral source mappings $\widetilde{\mathcal{M}}_m$ can be replaced by the same $\widetilde{\mathcal{M}}$ i.e., the source spectra are stackable in the transformed signals (primed quantities) as,

$$\mathbf{I}'_m = \mathcal{M} \cdot \mathbf{S}_{obs} + \mathbf{N}'_m. \tag{4.56}$$

In this particular case, the set of $(m+1)$ relevant incremental estimators are expressed by,

$$\begin{aligned} \mathbf{S}_{obs}^{(n)} &= \mathbf{S}_{obs}^{(n-1)} + \widetilde{\mathcal{A}} \cdot \mu(\widetilde{\mathbf{R}}') \\ \widetilde{\mathbf{N}}^{(n)} &= \boldsymbol{\rho}'_m \cdot \big(\widetilde{\mathbf{N}}'^{(n-1)}_m + \widetilde{\mathbf{R}}'_m\big), \end{aligned} \tag{4.57}$$

where $\mu(\widetilde{\mathbf{R}})$ is the arithmetic mean of the stacking residual spectra.

Clearly, the more interesting scenario is that when filtering of the source is provided by the preceding estimates of the noise, leading to convergent solutions, for the source structure, in the limit of a large number of iterations,

$$\mathcal{M} \cdot \mathbf{S}_{obs}^{(n)} = \big[\mathbf{1} - \mu(\boldsymbol{\rho}')^n\big] \widetilde{\mathcal{M}} \cdot \mathbf{S}_{obs} + \widetilde{\mathcal{M}}\widetilde{\mathcal{A}}\big[\mathbf{1} - \mu(\boldsymbol{\rho}')^{n-1}\big] \cdot \big\langle \widetilde{\mathbf{N}}' \big\rangle \tag{4.58}$$

in terms of the effective average noise $\big\langle \widetilde{\mathbf{N}} \big\rangle$ whose spectral spectral components are,

$$\big\langle \widetilde{N}'_f \big\rangle = \frac{\sum_m (1 - \rho'_{m,f}) \widetilde{N}'_{m,f}}{\sum_m (1 - \rho'_{m,f})} = \frac{\sum_m \Phi'_{m,f} N'_{m,f}}{\sum_m \Phi'_{m,f}}. \tag{4.59}$$

These are the *weighted* average noise component from the transformed stationary bits, under the effective spectral weighting function,

$$\mathbf{w}(\mathbf{f}) = \boldsymbol{\Phi}'(\mathbf{f}) \tag{4.60}$$

provided by the spectral filter–pass function for the transformed, stacking, spectra. Thus, in the limit of a large number of iterations, only the the weighted average noise is mapped into the source, which, in the stochastic–ergodic limit can be simply approximated as,

$$\mathbf{S}_{obs}^{(n\to\infty)} \rightsquigarrow \widetilde{\mathbf{S}}_{obs} + \widetilde{\mathcal{A}} \cdot \mathbf{n}\Big(\big[\mathbf{1} - \mu(\boldsymbol{\rho}')^n\big] \cdot \big\langle \widetilde{\mathbf{N}}' \big\rangle\Big). \tag{4.61}$$

The interpretation of the above result is a rather important one for the case when anisotropic noise is present in the experimental setup. What eq. (4.58) signifies, is that noise anisotropies can be effectively dealt with in an iterated pipeline reduction by a reconfiguration of observing patterns which superpose the anisotropic noise components over different parts of the source spectrum (i.e., transforming patters, see Section 3.6.6.1). The analysis pipeline of estimators, achieves the optimal weighting of the spectra (Section 3.3.3.2) to provide a statistical average, in which every spectral component is recovered to the best accuracy possible. That is, if the filtering is noise whitening (Section 3.4.2), since these will act as proper $1/\sigma^2$ noise weights for the independent noise components (Section 3.3.2). Then, the resulting average spectral estimate is also a maximum-likelihood one (Section 4.3.1). This holds approximately even for the family of 'optimal' weak signal filters (Section 3.4.3.2) also, since these are not all that different from noise whitening ones under the typically useful choices for the filter parameter λ.

In the example of SHARC-2, which has unfortunate $1/f$-type noise source along the rows of detectors, affecting all 32 pixels in each of the 12 detector rows, this means that this imperfection of the instrument electronics can be treated by providing adequate rotation of the detector array relative to the source structures observed. Such a rotation is normally granted by the rotation of the field itself (for Cassegrain focus mount) or by the elevation change (for Nasmyth mount) for sources observed over long periods ($\gtrsim$1 hour) of time around or near transit.

When noise is isotropic, the effect is less dramatic, but still results in the parts of data observed though the better mapping patterns to contribute more to the recovery of those source signal components which were less well covered by the other parts of the data.

4.4.2 1/*f*-Type noise

Consider estimating stationary noise from consecutive blocks of sample along a single data dimension. (Estimators for such contiguous data ranges are most easily implemented from a computational perspective!) This is analogous to applying a convolving rejection function in the time series of the form,

$$\widetilde{\rho}(x) = \begin{cases} \frac{1}{a} & \text{if } -\frac{a}{2} < x \leq \frac{a}{2} \\ \\ 0 & \text{otherwise,} \end{cases} \tag{4.62}$$

where a is the width of the characteristic data range $\mathcal{R}$ along the x direction. The corresponding spectral rejection function, given by the appropriately normalized Fourier transform of $\widetilde{\rho}(x)$ is:

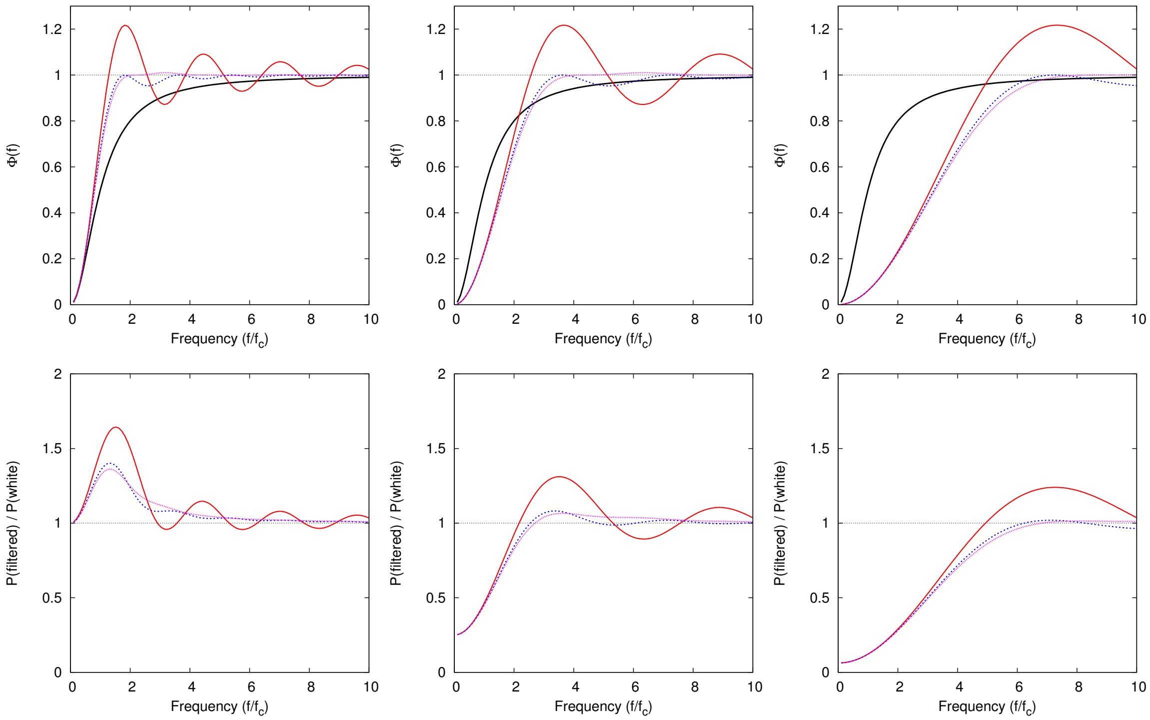
$$\rho(f) = \frac{\sin(\pi a f)}{\pi a f}. \tag{4.63}$$

According to eq. (4.51), the effective filter pass function $\Phi^{(n)}$ after n iterations becomes,

$$\Phi^{(n)}(f) = 1 - \left[\frac{\sin(\pi a f)}{\pi a f}\right]^n. \tag{4.64}$$

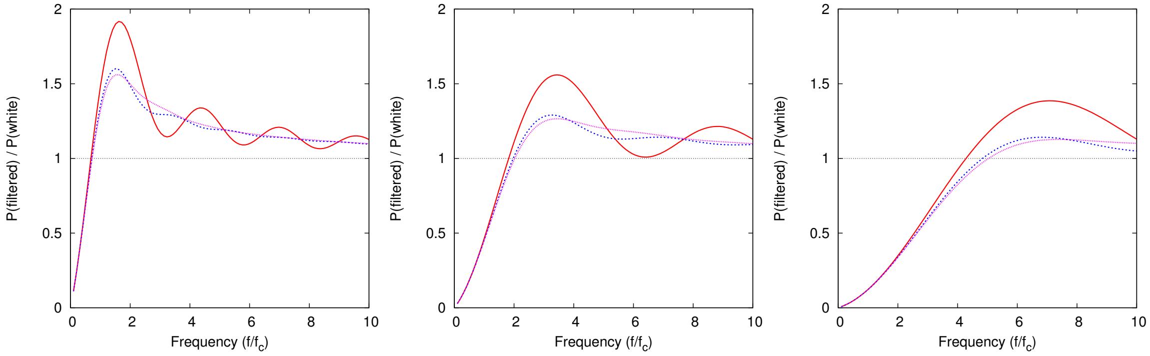
Two properties of this a filter that are worth noting (Fig. 4.2). Firstly, the initial rejection 'sidelobes' quickly disappear after a few iterations, yielding close to perfect spectral pass for frequencies $f \gtrsim 1/a$. At the same time, the filter behavior at the low ($f \ll 1/a$) frequencies can be approximated by its Taylor expansion,

$$\Phi(f) \approx \frac{n}{6}(\pi a f)^2 + \mathcal{O}\big([\pi a f]^4\big). \tag{4.65}$$



(a) **(b)** **(c)**

Figure 4.1 Iterated filtering of $1/f^2$ noise via boxcar data windows. The window sizes are set using eq. (4.67) as (a) an equality, or (b) one half, and (c) one quarter that size. Top graphs show the effective filter pass functions $\Phi(f)$, initially (red solid), after two (blue dashed) and three (purple dotted) iterations, and the noise perfect whitening filter (thick black curves) for comparison. The filtering quickly approaches its asymptotic form after three iterations. The bottom graphs indicate the residual noise power levels as a result of the filter action. While (a) approaches the noise-whitening limit at the very low ($f \ll f_c$) frequencies, it leaves behind noise excesses due to the imperfect nature of the boxcar filter approximation. The excess can be trimmed by using smaller windows, at the expense of some overfiltering at the very low frequencies.



(a) **(b)** **(c)**

Figure 4.2 Iterated filtering of $1/f$ noise via boxcar data windows. Clearly, setting window sizes like in Fig. 4.1 for $1/f^2$ noise, leaves more residual excess in comparison. However, the smaller window sizes treat $1/f$ behavior acceptably, even though the power spectrum here falls slower than $1/f^2$. Somewhat unfortunately, the overfiltering at the very low frequencies is no longer avoidable.

In summary,

$$\Phi(f) \approx \begin{cases} \frac{n}{6}(\pi a f)^2 & f \ll \frac{1}{a} \\ \\ 1 & f \gtrsim \frac{1}{a}. \end{cases} \tag{4.66}$$

Comparing this to the optimal weak signal, or noise whitening, filters of Chapter 3, one can observe that estimating noise signals from consecutive blocks of data is approximately optimal for filtering Brownian ($1/f^2$) noise, especially once the higher frequency behavior is smoothed out after a few iterations. As such the approach is optimally suited for ground based astronomy, where atmospheric variations exhibit such a red noise spectrum (Bester et al. 1992). For a Brownian spectrum with $P(f) = (f_c/f)^2\sigma^2$ (here f_c is the '$1/f$ knee' where the spectral behavior blends into the white noise level with $P(f) \rightarrow \sigma^2$), the optimal window sizes a can be calculated as,

$$a \leq \sqrt{\frac{6}{n}}\,\frac{\sigma}{\pi f_c}, \tag{4.67}$$

for the appropriate noise whitening filters, or

$$a \leq \sqrt{\frac{6}{\lambda n}}\,\frac{\langle |S| \rangle}{\pi f_c}, \tag{4.68}$$

for optimal weak signal filters, in terms of the expected low frequency ($f \ll 1/a$) source signal amplitudes $\langle |S| \rangle$ and the adjustable filter parameter λ. By setting smaller window sizes a, more aggressive low frequency filtering is achieved. Similar filter properties can be produced with different numbers n of iterations, by adjusting a such that the product na^2 is kept invariant.

More generally, the filtering properties of noise estimators derived from contiguous data blocks with range a (analogous to the boxcar rejection windows) are useful for filtering a variety of $1/f^\alpha$ type noise. However, when $\alpha < 2$ i.e., the spectrum is not Brownian, these time-domain estimators tend to overfilter the signals at the low frequencies when compared to the appropriate optimal or noise whitening filters for nonBrownian $1/f$ noise types.

4.4.3 Additional Filtering

Filtering from data ranges, other than those analogous to boxcar windows, can provide similarly tuned filtering for different types of excess noise. For stationary noise we argued for using data ranges offset on a regular grid (reflecting that noise properties do not evolve in time). If the derived parameters are to be independent, these ranges need to be kept separated in the data. As a consequence of these two practical constraints, filtering for stationary noise via time domain estimators is typically associated with periodic data blocks. If blocks, like the one above, are offset at regular spacings of Na along the x direction, then muted–periodic rejection functions of the form,

$$\rho(f) \approx \frac{\sin(\pi a f)}{\pi a f} \times \left\{ \sum_m \delta(mN\pi f a) \right\} \tag{4.69}$$

are produced, i.e. the rejection function will be a periodic comb function with a characteristic envelope.

At times, such filter designs may prove too restrictive or simply not good enough. Then, the linear time-domain filtering provided by the described sequence of estimators has to yield to appropriate Fourier filters. While these are inevitably slower, they may remain feasible, especially when the filtering can be performed on small subsets of data (with $n \ll N_k$ elements), along certain coordinates only, typically requiring $\mathcal{O}(N_k \log n)$ operations for the full processing. However, note that notwithstanding whether filtering is performed via time-domain estimators or Fourier transforms, the effective evolution of the filtering properties (Section [4.4.1]) due to iterating in the analysis applies just the same.

Stationary noise, which uniformly present in the time domain, is not the only trouble affecting most data. When noise is localized to particular regions of the time stream unreliable measurements ought to be flagged (in an example of multiplicative time-series filter). In general, whenever a detector channel or an exposure of the detectors become suspect (based on objective criteria!), it is best to discard these altogether (e.g., by assigning zero weights, see Section 3.3.4).

4.5 Weight Estimation

Approximately correct weights can always be estimated from the covariance estimates as was demonstrated in Section 3.3.3. However, such an approach is expensive because the calculation of the autocorrelation requires $\mathcal{O}(N^2)$ operations when performed directly. Even with FFTs this only reduces to $\mathcal{O}(N \log_2 N)$ operations, and that is not even mentioning the inversion of the $N \times N$ full covariance matrix!

As such, some shortcutting in the weight calculation may be often desired or even unavoidable. Fortunately, we have shown that optimal filtering of weak signals tends to produce data with white, or nearly white noise residuals. This means, that the final noise weights of the data reduction have a tendency to approach simple $1/\sigma^2$ weights of independent noise, which can be estimated rather easily and requiring $\mathcal{O}(N)$ computations only.

Weights can be calculated at any time in the white noise assumption. However these will only be approximately proper weights once sufficient modeling (hence filtering!) of the data has already been achieved i.e., when the white noise assumption is nearly correct. As such, it is better to delay weight estimation in the pipeline till after the dominant signals have been modeled (even if only coarsely).

4.5.1 Weighting in Configuration Space

In the white noise limit, the expected level of the integrated residual variances in a subset $\mathcal{R}$ of the data is simply related to the underlying white noise RMS σ as,

$$\mathbf{E}\Big(\sum_{\mathcal{R}} |D - M|^2\Big) = \sigma^2 N_{df}(\mathcal{R}), \tag{4.70}$$

where $N_{df}(\mathcal{R})$, is the degrees of freedom in the subset $\mathcal{R}$ of the data given the extent of modeling already undertaken. The noise variances σ_k^2 are easily estimated from this as,

$$\hat{\sigma}_{\mathcal{R}}^2 = \frac{\sum_{\mathcal{R}} |D - M|^2}{N_{df}(\mathcal{R})} = \frac{\sum_{\mathcal{R}} |R|^2}{N_{df}(\mathcal{R})}. \tag{4.71}$$

If the noise of the time-series data $\sigma_{t,c}$ can be separated in detector and time only noise terms as $\sigma_{t,c} = \chi_t \cdot \sigma_c$, then the weights will separate accordingly as,

$$w_{t,c} = w_t \cdot w_c = \frac{1}{\chi_t^2} \cdot \frac{1}{\sigma_c^2}. \tag{4.72}$$

I.e., the time and detector weights are respectively given by (in a specific normalization) as,

$$w_t = \frac{1}{\chi_t^2}, \text{ and } w_c = \frac{1}{\sigma_c^2}. \tag{4.73}$$

Clearly, the same effective data weight $w_{t,c}$ can be produced by infinitely many combinations of w_t and w_c. To restrict oneself to a single pair of weights w_c and w_t from the family of possible pairings, consider the χ_t–σ_c pair for which

$$\left\langle \chi_t^2 \right\rangle = 1 \tag{4.74}$$

is satisfied inside a large subset $\mathcal{R}_0$ of measurements $D_{t,c}$. A restriction on $\mathcal{R}_0$ is that the χ_t–σ_c noise separation must remain a valid approximation in its full data range. In practice, this typically means a data block of a not too long contiguous integration (e.g., a $\lesssim$10 minute scan by the SHARC-2 array).

Under the above normalization, χ_t are dimensionless whereas σ_c carry the dimension of the signal and are meaningful as the characteristic RMS noise level for detector channel c over the relevant time span ($\mathcal{R}_0$) of the measurements. Then the set of interrelated noise estimators, restricted inside the subset $\mathcal{R}_0$, are obtained as,

$$\widehat{\sigma}_c^2 = \frac{1}{N_{df}(\mathcal{R}_c)} \sum_{t \in \mathcal{T}_c} |R/\widehat{\chi}_t|^2 \tag{4.75a}$$

$$\widehat{\chi}_{\mathcal{R}}^2 = \frac{1}{N_{df}(\mathcal{R})} \sum_{t, \forall c \in \mathcal{R}} |R/\widehat{\sigma}_c|^2 \tag{4.75b}$$

where $\mathcal{R}_c \subset \mathcal{R}_0$ is a subset of measurements $D_{t,c}$ for a selected detector channel c, and $\widehat{\chi}_t = \widehat{\chi}_{\mathcal{R}}$ for $t \in \mathcal{R}$. The subset of data used in determining $\widehat{\chi}_t$ and hence w_t i.e., $\mathcal{R} \subset \mathcal{R}_0$ should contain all samples $D_{t,c}$ which are measured in a time interval over which noise is assumed entirely stationary.

The two expressions (in eq. [4.75]) are interrelated in a nonlinear fashion. Thus, instead of aiming for algebraic solutions, the convergence to self-consistent set of χ_t and σ_c values can be achieved through iterations. The determination of separable weights, therefore, adds to the list of reasons for data analysis in an iterated environment. A practical starting condition is to assume all $\chi_t = 1$ (i.e., initial assumption of time-stationary noise) and begin with the estimation of detector weights w_c via $\widehat{\sigma}_c^2$.

4.5.2 Spectral Weight Estimation

One should note that intermediate weights in the weighting–filtering cycle are inconsequential, as long as convergence to the correct final values is granted. In fact, the sooner one can guess the final set of noise weights, the faster the weighting–filtering cycle can converge. Because final weights are calculable under the white noise assumption (provided an optimal level of filtering), these can be estimated early on form the white noise regions of the signal spectrum rather than the time-series signals.

This is possible, owing to Parseval's theorem (eq. [3.17]). Comparing it to eq. (4.75a) used for weight estimation, it is clear that,

$$\mathbf{E}\Big(\sum_{f \in \widetilde{\mathcal{R}}} |\widetilde{R}_f|^2 \Big) \propto \sigma^2 N_{df}(\mathcal{R}). \tag{4.76}$$

Under the appropriate normalization corresponding to the Fourier transform used for obtaining the spectral representation. When transforming white noise, all spectral components are white noise too, and thus,

$$\widehat{\sigma}^2 \propto \frac{N_f}{N_{df}(\mathcal{R}_0)} \big\langle |\widetilde{R}_f|^2 \big\rangle. \tag{4.77}$$

When noise is structured the spectral white noise level $\big\langle |R_f|^2 \big\rangle$ can be estimated from the appropriate spectral components of the white noise regions ($\widetilde{\mathcal{R}}_w \subset \widetilde{\mathcal{R}}$) of the spectrum as,

$$\left\langle |R_f|^2 \right\rangle \approx \frac{1}{N(\widetilde{\mathcal{R}}_w)} \sum_{f \in \widetilde{\mathcal{R}}_w} |R_f|^2. \tag{4.78}$$

Combining the above expressions thus provides,

$$\widehat{\sigma}^2 \propto \frac{N_f}{N_{df}(\mathcal{R}_0) N(\widetilde{\mathcal{R}}_w)} \sum_{f \in \widetilde{\mathcal{R}}_w} |R_f|^2, \tag{4.79}$$

under the same proportionality constant as eq. (4.76) above.

This method is primarily useful for obtaining better detector weights w_c under stationary noise, by weighted Fourier transforming the residuals $R_{t,c} \in \mathcal{R}_0$ along the time direction only, i.e., using the temporal spectral estimate of the detector time streams for the detector channel weight estimate. In this way eq (4.75a) above can be effectively replaced by eq. (4.79) to expedite the convergence in the weighting–filtering cycle of the analysis.

4.5.3 Robust Weight Estimation

As an alternative of spectral weighting, which requires Fourier transforms, typically along the time direction of the data, one may use robust weight estimation, by deriving weight based on the median, not the weighted mean, square deviation. For a normalized probability distribution $p(x)$ of a random variable x, the median square deviation value x^2_{med} is provided by the condition $p(x^2) = 1/2$. For the Gaussian distribution (eq. [3.35]) this is found numerically to be at,

$$x^2_{\text{med}} \approx 0.454937\, \sigma^2_x. \tag{4.80}$$

By inverting the above expression, we can estimate the variance of zero-mean, quasi-Gaussian distributions in a robust way, using the median square deviation for the estimate. This is primarily useful for noise distributions that are mostly Gaussian, but exhibit non-Gaussian tails at higher deviations (e.g., from localized structures). More generally median weighting can be used when excess noise corrupts only a fraction of the time-stream data i.e., for detector weighting (w_c) under short bursts of transient noise.

4.5.4 Fractional Degrees of Freedom

The counting of the degrees of freedom for any particular subset of the data, used in the weight estimation (see above), requires some careful thought. When the accounting is not done correctly, some of the data will be over- or under-weighted in some systematic fashion. Even small initial deviations from fair weights can grow with iterations, providing unrealistic solutions where only a fraction of the data carry nonzero weights, while others becoming zero-weighted. For this reason, the accurate measuring of the degrees of freedom associated to any subset of the data is absolutely critical.

For linear estimators, we can write the incremental model solution as a linear combination of the observed point estimators,

$$\delta \widehat{A}_i = \sum_{\mathcal{R}_i} f_i \, \delta \widehat{A} = \sum_{\mathcal{R}_i} f_i \, (R/g_i), \tag{4.81}$$

where $\delta \widehat{A}_k = R_k / G_k$ is a point estimator of A for a single datum, providing a point estimator of A_i. Comparing the above to the expression for the maximum-likelihood estimator (eq. [4.18]), one can immediately identify the coefficients as,

$$f_{i,k} = \frac{w_k |g_{i,k}|^2}{\sum_{\mathcal{R}_i} w |g_i|^2}. \tag{4.82}$$

The coefficients $f_{i,k}$, are readily interpreted as the fractional contribution of the measurement D_k to the derivation of the model parameter A_i, which as a result obey the condition,

$$\sum_{\mathcal{R}_i} f_i = 1. \tag{4.83}$$

This *fractional dependence* of the parameter A_i on the measurement D_k is useful for calculating the effective number of parameters i.e., the lost degrees of freedom, for any subset $\mathcal{R}$ of the data (not only the subset $\mathcal{R}_i$ used for deriving A_i), simply by summing $f_{i,k}$ within the set defined by $\mathcal{R}$, as $N_M(\mathcal{R}) = \sum_{\mathcal{R}} f_k^i$. In this way, the total degrees of freedom remaining on an arbitrary data subset $\mathcal{R}$ after the estimation of model parameters A_i is consequently calculated as,

$$N_{df}(\mathcal{R}) = N(\mathcal{R}) - \sum_{\mathcal{R}} \sum_i f_i, \tag{4.84}$$

where $N(\mathcal{R})$ is the number of measurement with nonzero weights inside the data subset $\mathcal{R}$. While seemingly innocent, it must be reemphasized that the accurate accounting of the degrees of freedom in this way is ultimately crucial for the stability of the weighting-filtering cycle of any data reduction approach.

4.6 Gains and Nonlinear Detector Response

Thus far we have made the implicit assumption of linear detector response i.e., that the observed signals D are produced by the the linear (matrix) transformation $\mathbf{D} = \mathcal{M} \cdot \mathbf{A}$. However, most measurement apparatus has nonlinearities in their response to underlying signals. Often, the deviations from linearity are small enough to be negligible, whereas at times these may be significant. Deviations from linearity become critical for weak signal analysis when the effective variation of gains $\delta g = \partial \mathcal{M} / \partial A_i A_i$ exceed the the gain accuracy requirement (eq. [3.62]). Fortunately, the estimators may be tweaked to account for the nonlinear detector response.

When detector response is not linear, gains are typically loading dependent (apart from the typical variation along measurement coordinates) i.e., can be expressed as a function of the 'noiseless' underlying total power signal $\overline{D}_k$ via the functions $G_{i,k}(\overline{D}_k)$ replacing the matrix elements $\mathcal{M}_{i,k}$ in the mapping of parameters A_i which are written as,

$$D_k = \sum_i G_{i,k}(\overline{D}_k) A_i. \tag{4.85}$$

Since $\overline{D}_k$ are essentially the measured data D_k, this conveniently means that gains are well determined as long as the functionals $G(D)$ are known accurately, such that the estimation of A_i can proceed without hindrance. However, owing to the nonzero $\partial G_{i,k} / \partial A_i$ terms, the χ^2 minimizing condition $\partial \chi^2 / \partial A_i$ is slightly modified, compared to the maximum-likelihood estimators of Section 4.3.1, leading to a an appropriately altered eq. (4.18) for expressing the the incremental maximum-likelihood estimators as,

$$\delta A_i^{(n)} = \frac{\sum_{\mathcal{R}_i} w \, g_i^*(D) R}{\sum_{\mathcal{R}_i} w \, |g_i(D)| \, |G_i(D)|}. \tag{4.86}$$

Here, $g_{i,k}(D_k) = \partial D_k / \partial A_i$ is, by definition, the momentary small-signal gain affecting the way small incremental changes to A_i are mapped into the time stream. After a simple manipulation, this small-signal gain is found to be,

$$g_{i,k}(D_k) = \frac{G_{i,k}(D_k)}{1 - G'_{i,k}(D_k) A_i}, \tag{4.87}$$

in terms of $G'_{i,k} = \partial G_{i,k} \partial D$. In this way, the $g_{i,k}$ assume knowledge of A_i for estimating A_i. This circularity can be approached by iterated solutions where $g_{i,k}^{(0)} = G_{i,k}$ is assumed initially to obtain a first estimate of A_i and prior A_i values are then used for calculating increasingly accurate $g_{i,k}$ values according to eq. (4.87). Similarly to eq. (4.21), the uncertainty of $\widehat{A}_i$ is provided for nonlinear response by,

$$\widehat{\sigma}_i^2 = \frac{1}{\sum_{\mathcal{R}_i} w \, |g_i(D)| \, |G_i(D)|}, \tag{4.88}$$

and the fractional dependence of $\widehat{A}_i$ on the measurement D_k is calculable, in simplified form using the above result as,

$$f_{i,k} = w_k \, |g_{i,k}(D_k)| \, |G^*_{i,k}(D_k)| \, \widehat{\sigma}_i^2, \tag{4.89}$$

which is useful for estimating the degrees of freedom for arbitrary data subsets according to eq. (4.84). For robust estimators, the point estimates $\widehat{A}_{i,k}$ of A_i are simply expressed from eq. (4.86) as,

$$\widehat{A}_{i,k} = \frac{R_k}{G_{i,k}(D_k)}. \tag{4.90}$$

With these slight modifications, all prior discussion holds true for the case of nonlinear detector responses.

4.6.1 Gain Estimation

In Chapter 3 we demonstrated the critical role of accurate gain knowledge for effective filtering of dominant noise. At times, gains can be determined accurately (according to the criterion of eq. [3.62]) by calibration procedures which are external to the scientific measurement. This is more likely the case when gains are stable and detector responses are sufficiently linear. However, complex nonlinearities or variations to the effective gains from measurement to measurement (e.g., slight changes in the bath temperature for bolometers, or changes due to the wear and tear of the apparatus) can make such external gain characterizations either inaccurate or impractical. Therefore, it is sometimes desirable that the relevant gains are estimated from the data themselves.

Suppose, that the functional form of the gains are known to the desired accuracy (at least in the signal range of the data), but not the various scalings constants g_μ each of which apply to the mapping of parameters inside the corresponding domain $\mathcal{D}_\mu$ i.e.,

$$G_{i,k}(D) = g_\mu \, \gamma_{i,k}(D_k) \quad \text{for } A_i \in \mathcal{D}_\mu, \tag{4.91}$$

for $k \in \mathcal{R}_j$, with $\gamma_{i,k}(D)$ well known but the coefficients g_μ not determined accurately. The functions $\gamma_{i,k}(D)$ are measure the relative coupling of the parameter A_i to the different measurements D_k, and may be derived from theoretical considerations, or else sufficiently characterized by an external calibration procedure. Then, accurate gain knowledge requires the determination the coefficients g_μ to the desired precision.

It is possible to derive a maximum-likelihood estimate from the condition $\partial\chi^2/\partial g_\mu = 0$ yielding incremental solutions for the gain coefficients as,

$$\delta\widehat{g}_\mu = \frac{\sum_{\mathcal{R}_\mu} w \, \Upsilon^*_\mu \, R}{\sum_{\mathcal{R}_\mu} w \, |\Upsilon_\mu|^2}, \tag{4.92}$$

with corresponding gain uncertainties of

$$\widehat{\sigma}^2_\mu = \frac{1}{\sum_{\mathcal{R}_\mu} w\, |\Upsilon_\mu|^2}, \tag{4.93}$$

and the fractional dependence of the gain constants on the measurements, used for the correct estimation of the degrees of freedom on arbitrary subsets of the data as,

$$f_{\mu,k} = \widehat{\sigma}^2_\mu w_k \, |\Upsilon_{\mu,k}|^2. \tag{4.94}$$

Each of the above are expressed using,

$$\Upsilon_{\mu,k} = \sum_{\mathcal{D}_\mu} \gamma_{i,k}(D_k) \widehat{A}_{i,k}, \tag{4.95}$$

where Υ_k is the unscaled compound template signal from the parameters A_i inside the gain coefficient domain $\mathcal{D}_\mu$ for which $\widehat{g}_\mu$ is the best-fit scaling given the measurements D_k. The data ranges $\mathcal{R}_\mu$ are defined in the usual way to indicate the subset of data affected by the gain constant g_μ, and is thus the union of the ranges $\mathcal{R}_i$ for those parameters $A_i \in \mathcal{D}_\mu$.

Clearly, when both the gain coefficients g_μ and parameters A_i in $\mathcal{D}_\mu$ are fitted to the data, all combinations that yield the same $g \mu A_i$ product provide equally good solutions i.e., gain fitting does not provide absolute gain constants only relative ones. As such, it is advisable to enforce a practical normalization of the derived gain coefficients, such that the estimated parameters A_i are physically meaningful. This is especially true for the source signals since the objective of the experiment is the faithful determination of these. Thus, it is recommended that gain constants are renormalized after every round of gain fitting to provide,

$$\langle \widehat{g}_k \rangle = G_0, \tag{4.96}$$

where G_0 is the absolute average gain of the source signals, which is to be determined by an appropriate calibration measurement (e.g., by observing source of known 'brightness') or otherwise (e.g by incorporating ancillary data like a measurement of the optical depth τ for ground-based submillimeter astronomy). As such, it is clear that when the initial signal parameters $A^{(0)}{}_i$ are zero valued, a practical starting point for data reduction without prior assumptions, some initial gain constants need to be assumed. These initial gains ought to be subject to the same normalization condition as above. Therefore, unless further information is available, a practical starting point could be where all initial gain coefficients are equal with $g^{(0)}_\mu = G_0$, and the parameters necessary for the determination of the g_μ values are estimated before the estimation of g_μ themselves.

Given the imperfect modeling (i.e., also filtering) of data that result from inaccurate gains (Section 3.5.2), it is clear that one should estimate gains for the dominant model components as soon as those are determined and not proceed to the modeling of the weaker signals until the gain solutions achieve the desired precision (cf. Section 3.5.2.1). In this way, embedded sub-cycles of gain estimation, with the corresponding re-derivation of the dependent parameters, may be necessary within the iterative pipeline that is established otherwise.

The fitting of gains (via coefficients) also presents an important subtlety. As prior parameter models were estimated from the old set of gains, the derivation of new gains implicitly alter the shape of these models too, which no longer represent the 'best' estimates of the data following gain estimation. This is not necessarily a problem in an iterated reduction scheme, where newer and improved parameter models are derived with the successive iterations. However, with the effective change gain dependent models, one has to pay attention to always recalculate residuals $R = I - M$ to reflect the change of gains. A glance at eq. (4.91) reveals that this can be done by adjusting the residuals as

$$\delta R_k = -\sum_{\mu} \delta\widehat{g}_\mu \sum_{\mathcal{D}_\mu} \gamma_{i,k}(D)\, A_i, \tag{4.97}$$

after the incremental estimates of the renormalized gain coefficients $\delta\widehat{g}_\mu$ are derived.

Last but not least, one ought to mention that the fitting of gain constants renders the data reduction problem fundamentally nonlinear in nature, as both components of the products $g_\mu A_i$ are fitted (thus it is now a quadratic problem in effect). As such, gain estimation comes with the usual caveats of nonlinear modeling of data. These can mean undesired convergence to local minima, or narrow 'valleys' in parameter space which are not easily explored etc. To obtain the optimal solutions despite these difficulties, the user of the data reduction has to be careful, and at times creative, in identifying a closely ideal configuration of the data-reduction pipeline.

4.7 Comparison to Alternative Methods

By now we have demonstrated that time-series analysis through a pipeline of estimators, can be effective at reducing data optimally in the presence of $1/f$-type noise (esp. Brownian noise). However, to do justice to other established models of data reduction it is fair to provide a comparison between those and the sequential estimator method discussed here. We shall use the the methods of matrix inversion through normal equations and that of Singular Value Decomposition (SVD) as the standards of comparison.

4.7.1 Direct Matrix Inversion

A common approach to data analysis is to directly invert the mapping of parameters under noise (eq. [4.5]), expressing the parameter vector $\mathbf{A}$ from the data $\mathbf{D}$ after ignoring the noise term, via the normal equations (Press et al. 1986) as,

$$\widehat{\mathbf{A}} = [\mathcal{M}^T \mathcal{M}]^{-1} \mathcal{M}^T \cdot \mathbf{D}. \tag{4.98}$$

Unfortunately, that is often easier said than done. One problem is that the set of parameters that intuitively arise, e.g., from physical considerations, in eq. (4.1) can be degenerate, in the sense that different combinations of model parameters may produce identical signal signatures. When such degeneracies exist, the matrix $\mathcal{M}^T \mathcal{M}$ is singular and, therefore, not directly invertible. Although the degeneracies involved maybe removed by explicit incorporation of appropriate constraints (which reduce the effective number of fitted parameters), such constraints can be tedious or nontrivial to formulate, especially when complex observing strategies are used in producing the data set. Also, the sheer size of the matrix operations involved, most generally requiring N_k^2 operations (although some shortcutting may be possible at times), can make this approach unattractive for large data sets. The required calculation can be significantly less intensive when the matrices mainly contain zero components (e.g., can be reduced as block diagonal).

Aside from the practicalities of inverting large and complex matrices, the method raises additional concerns, even when otherwise feasible. Compare this approach to the general framework of weak signal data reduction presented in Chapter 3. To incorporate noise weights, the matrix multiplication can be appropriately converted to noise weighted sums. These essentially reproduce the χ^2 minimizing normal equations (Press et al. 1986). What is less clear is how the matrix inversion relates to thee 'optimal' filtering of the data. Filtering, may be performed externally, (e.g., by prewhitening) but arises also from the constraints which are used to remove the degeneracies and make the matrix inversion possible. As these constraints themselves can be complex, so will be the filtering properties that they procure.

4.7.2 Singular Value Decomposition

Singular Value Decomposition (SVD) is a more widely applied method of data reduction, directly suitable for degenerate data models. SVDs identify a single (approximate) solution to the matrix equation (eq. [4.5]) from a the family of possible (approximate) solutions, by picking one which offers a minimal χ^2 deviation (or another extremum). As such SVDs invariably yield solutions to the χ^2 minimizing problem, although these may not be entirely unique. In fact, linear combinations of 'solution vectors' with identical eigenvalues provide equally good solutions themselves (for a better, more detailed discussion of SVDs refer to Press et al. 1986). This should be expected when identical signal structures can be produced via different degenerate parameter sets, linear combinations of which have identical signal signatures also. In other words, from a space of equally plausible solutions, each with identical χ^2 values, SVD identifies but a single one, leaving one wondering about the propriety of this choice.

It should be plainly obvious that no mathematical approach, hence not the SVD either, should be able to magically resolve the inherent degeneracies of the problem. While in case of the direct matrix inversion, via the normal equations, the filtering properties are determined explicitly by constraints supplied, no such external constraining is needed for the SVD approach. Therefore, the filtering properties of SVD are more difficult to determine, and may even be arbitrary, since the solutions are not always exactly unique. It is only via further physical constraining that this value can become fully meaningful.

SVD approaches, therefore, do not get around the complexity of the filtering problem presented by the direct matrix inversion approach. At the same time, SVDs are computationally even more expensive.

Last but not least, neither the SVD nor the direct matrix inversion is directly suited for solving nonlinear parametrization. Although this may be overcome by an iterated scheme with converging estimates, the costly approximate solutions provided at every turn are not that different from the estimates that result from the inexpensive linear pipeline approach.

4.8 Conclusions

We demonstrated that closely optimal analysis of data with $1/f$-type noise (especially Brownian noise) can be achieved through a sequence of statistical estimators which are derived for every model parameter, one at a time. The type of estimators used can be chosen to provide the desired statistical properties. The modeling of noise signals is an equivalent approach to the spectral filtering discussed in Chapter 3. When optimally configured, the method easily achieves $1/f^2$ noise whitening, and converges to properly weighted spectra when a set of intertransforming observing patterns allow for a 'stacking' of these. Optimal pipeline configurations are achieved through:

- Careful choice of noise parametrization, with respective data ranges. $1/f$ noise is best parametrized from regularly spaced, nonoverlapping, blocks of time-series data.
- Optimal ordering of the parameter estimators.
- Identifying an optimal number of iterations for each of the various signal component models. A handful of iterations is usually adequate for reaching optimal solutions.
- Estimation of quasi-proper statistical noise weights, with pivotal importance assigned to the proper accounting of the degrees of freedom for arbitrary data subsets.
- Estimation of gain constants to the prescribed accuracy, and recalculating the residuals when gains are adjusted.

- Additional Filtering via FFT or otherwise, if desired, and the discarding of unreliable data.

Iterating is necessary partially to reach a self-consistent weights and filters (see Chapter 3), but also to mold effective filtering properties, arrive at accurate gain parameters or deal with equivalent nonlinearities, and to identify and flag unreliable data. The initial filtering properties provided by the relevant noise estimators evolve through the iterations. As the source model is re-derived multiple times, the degenerate noise signals gradually make their way back into the source model. Thus, excessive iterating is generally counterproductive.

The weighting (including flagging) and gain estimation steps can be inserted in between the derivation of signal component parameter vectors. Granted, some insertion points are better justified than others. Typically, one should delay the estimation of weaker signal components until a accurate set of gains and weight are derived from the modeling of more dominant signals.

In its most optimal form, the method's computation costs grows linearly with the data size, placing it at a great advantage of alternative approaches, such as matrix inversion via normal equations or Singular Value Decomposition, or even simple Fourier transform-based analysis. At the same time its filtering properties are well-characterized and easily tracked, is sharp contrast to the matrix methods which leave these obscured under their relative mathematical complexity.

The faint galaxies of Chapter 2, and a variety of SHARC-2 papers whose scientific results also benefited from the *CRUSH* data reduction, provide proof of the effectiveness of this approach. The method also provides a viable form of data reduction for the very large data sets expected from future bolometer arrays, and the analysis weak signals under $1/f$ noise interference in general.

Bibliography

Audley, M., et al. 2004, preprint (astro-ph/0410439)

Bester, M., Danchi, W. C.. Degiacomi C. G., Greenhill, L. J., & Townes, C. H. 1992, ApJ, 392, 357

Dowell, C. D., et al. 2003, SPIE Conf. Proc. Vol. 4855, Millimeter and Submillimeter Detectors for Astronomy, ed. T. G. Phillips, & J. Zmuidzinas, 73

Dupac, X., & Giard, M. 2002, MNRAS, 330, 497

Emerson, D. T., & Graeve, R. 1988, A&A, 190,353

Emerson, D. T. 1995, ASPC, 75, 309

Ferreira, P. G., Jaffe A. H. 2000, MNRAS, 312, 89

Gibb, A. G., Scott, D., Jenness, T., Economou, F., Kelly, B. D., & Holland, W. S. 2005, ASPC, 247, 585

Jenness, T., Economou, F., Scott, D., Kelly, B. D., & Holland, W. S. 2004, ASPC, 314, 428

Johnstone, D., Wilson, C. D., Moriarty-Schieven, G., Giannakopoulou-Creighton, J., & Gregersen, E. 2000, ApJS, 131, 505

Narayan, R., & Nityananda, R. 1986, ARA&A, 24, 127

Poutanen, T., et al. 2006, A&A, 449, 1311

Press W. H., Flannery B. P., & Teukolsky, S. A. 1986, Numerical Recipes. The Art of Scientific Computing (Cambridge: University Press)

Tegmark, M. 1997a, PhRvD, 55, 5895

Tegmark, M. 1997b, PhRvD, 56, 4514

Appendix A

Note on the Effective Number of Beams Inside an Aperture

The effective number of beams $N_{\rm beam}$ inside some aperture $a(\mathbf{x})$ can be counted by the number of independent Fourier components necessary to parametrize the aperture under some limiting beam resolution. Thus, the underlying flux distribution $f(\mathbf{x})$ is convolved by the beam response $\beta(\mathbf{x})$, normalized such that $\int d\mathbf{x}\,\beta(\mathbf{x}) = 1$, and windowed by the aperture $a(\mathbf{x})$ with the image inside the aperture expressed as $f_a(\mathbf{x}) = f(\mathbf{x}) \otimes \beta(\mathbf{x}) \cdot a(\mathbf{x})$ i.e., $F_a(\omega_x) = F(\omega_x) \cdot B(\omega_x) \otimes A(\omega_x)$ in the Fourier domain, under tranformation with $e^{-i\omega\cdot\mathbf{x}}$. Hence, for white noise, $\langle |F(\omega_x)| \rangle = B(\omega_x) \otimes A(\omega_x)$, where the spectral resolution is given by $A(\omega_x)$. The number of independent spectral noise components is then expressed as,

$$N_{\rm comp} = N_{\rm beam} = \frac{\int d\omega_x\,|A(\omega_x) \otimes B(\omega_x)|^2}{\int d\omega_x\,|A(\omega_x)|^2} \tag{A.1}$$

For quasi-Gaussian beams and pillbox apertures the product $A \otimes B$ will also be nearly Gaussian, owing to the central limit theorem, with a widened spread $\sigma^2_{A\otimes B} \approx \sigma^2_A + \sigma^2_B$, under the implicit Gaussian approximation to $A(\omega_x)$ with an equivalent σ_A spread. Therefore, the above can be approximated as,

$$N_{\rm beam} \approx \frac{\sigma^2_A + \sigma^2_B}{\sigma^2_A} \approx 1 + \frac{\sigma^2_a}{\sigma^2_\beta} \tag{A.2}$$

Recalling the uncertainty principle for Gaussian Fourier transforms $\sigma_a(\mathbf{x})\sigma_A(\omega_x) = 1$, and observing that the variances relate to the areal integrals as $A = 2\pi\sigma^2$ for both pillbox apertures and Gaussian shapes, the number of independent beams are expressed, in terms of the areas as,

$$N_{\rm beam} \approx 1 + \frac{A}{A_{\rm beam}} \tag{A.3}$$

The above result is intuitively justified, as at least a single beam (or spectral component) is always necessary to describe the image in an aperture, no matter how small, and for large apertures one should expect the number of beams to grow linearly with the aperture as $N \sim A/A_{\rm beam}$. The results will approximately remain valid for generic apertures as long as the deviations from a pillbox shape are not too extreme i.e., for apertures that are contiguous, not too elongated, and for beams that are reasonably approximated by Gaussian shapes.

Appendix B

CRUSH: A Practical Guide

B.1 *CRUSH* for SHARC-2

The aim of this appendix is to provide a practical guide of incorporating the data reduction principles of *CRUSH* (Chapters 3 and 4) into an effective implementation. This is done through the example of the SHARC-2 reduction.

SHARC-2 has 384 semiconductor bolometer detectors arranged in a 32×12 format with closely packed $4.85''$ square pixels.[1] The AC-biased, demodulated and amplified, detector signals are sampled at every 36 ms to provide the time-stream data. Thus the data cube is 3-dimensional, with x and y coordinates describing the detector columns and rows respectively, and a time axis t. Data is taken normally in 2–10 minute integrations.[2] Multiple scans on a single target are generally isolated in time, and can be distributed over days, even years.

Scans. Because data is generally not a contiguous block of sample, or because noise properties may be slowly varying over long integrations, division into smaller chunks is often practical for creating smaller data blocks for which the stationary noise assumption holds (see Section 3.3.3). The *CRUSH* analysis approach is mainly optimal for the analysis of data with stationary, and especially $1/f$-type noise (Section 3.3.3.1). For SHARC-2, the noise properties of individual scans (2–10 minutes typically) are reasonably described as stationary.

The full data set can be a collection of (possibly noncontiguous) scans (see fig. B.1). Accordingly, models (noise components, gain and weight estimates, detector flags etc.) ought to be calculated *separately* for each scan. The source model is the exception to that, since the same source signals are present in all scans that observed the same target. Thus a single source model is used with a collection of scans.

---

[1] The effective pixels are slightly rectangular with $4.89'' \times 4.82''$ on the Cassegrain mount and $4.49'' \times 4.74''$ on the Nasmyth mount. For consistency, all fluxes in *CRUSH* are scaled to the effective Cassegrain values when referred to the input.

[2] Shorter integrations of 2-3 minutes are used for pointing and calibration, while 10 minute scans are typically used for science targets.

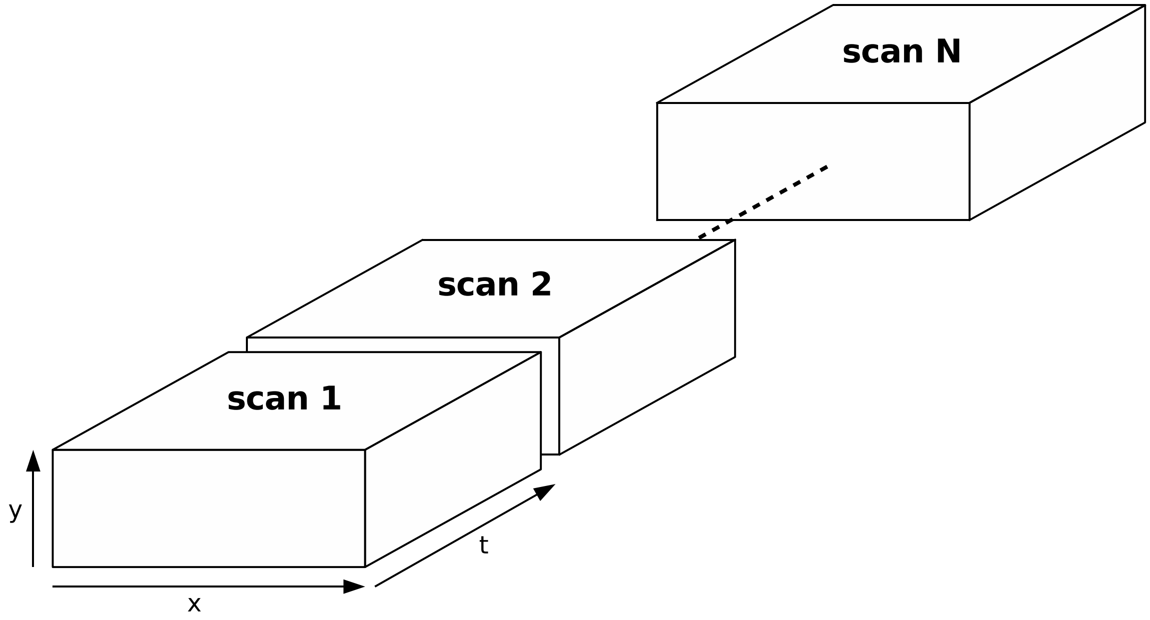


Figure B.1 The full data cube typically consists of a collection of scans, which may have been taken at different times. Even when data is taken in a single long contiguous block, a division into smaller, stationary blocks may be practical. SHARC-2 scans are typically $2 - 10$ minutes long. Analysis is performed for each scan separately, except for the source model which is estimated from the full data cube.

The separation of modeling into scans (or other practical stationary subdivisions) is useful from a computational perspective also. Each scan may be processed separately, in parallel, for the most part, providing a way to process even extremely large datasets with a cluster of computers. These computers need to work in unison only for the appropriate derivation of the source model, whereas they may process data independently otherwise.

Slicing. The constraints of nonoverlapping parameter ranges for the independence of the parameter estimates (Section 4.4), and the contiguous data blocks (i.e., boxcar windows) that are effective for the treatment of $1/f$-type noise signals via *CRUSH* (Section 4.4.2), in practice mean that each scan (or analogous contiguous stationary blocks of data) of $N_x \times N_y \times N_t$ samples[3] is sliced into smaller cubes of sizes $n_x \times n_y \times n_t$ samples for the estimation of stationary (i.e., uniform) noise components. The slices provide characteristic regions $\mathcal{R}_m$ (Section 4.4) for the various component signals m (Section 4.2) in that scan. The incremental parameters that are derived from the $\mathcal{R}_{m,j}$ are simply low-resolution representations of the scan residuals at the time of the estimation. A schematic illustration of the data slicing used for the various SHARC-2 models is shown in Figure B.3.

The combination of ranges $\mathcal{R}_i$ and corresponding gains $g_{i,tc}$ for a parameter A_i provide its

---

[3] for SHARC-2 $N_x = 32$ and $N_y = 12$.

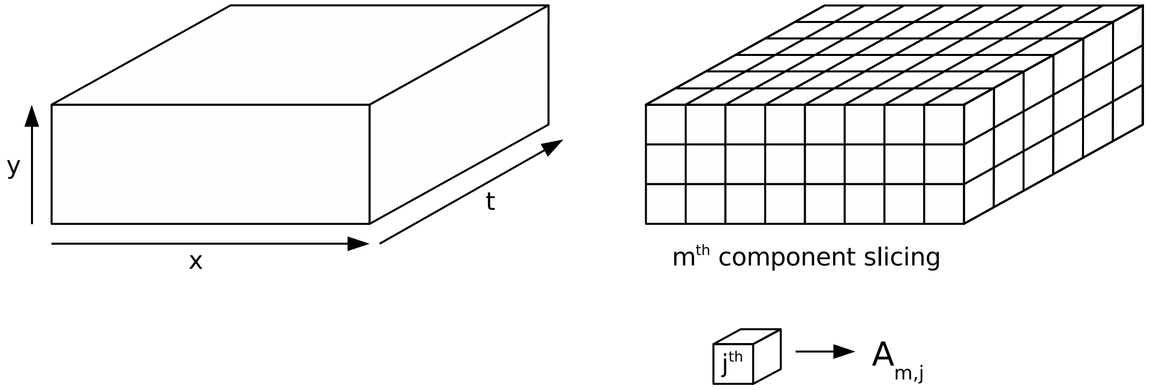


Figure B.2 Slicing of scan data for deriving parameter estimates. The practical constraints on how to parametrize noise and other signal components (Section 4.2) such that these are independent (Section 4.4) for a given component signal K_m, and are effective for $1/f$-type noise filtering (Section 4.4.2), is fulfilled if data is sliced into smaller cubes. Statistical estimates for the jth cube provide a signal parameter $A_{m,j}$, which is the effective value of K_m within that region. Thus, parametrization of $1/f$-type noise signals happens via low-resolution snapshots of the residuals.

mapping $\mathcal{M}$ into the time stream (Section 4.3.1). Thus, each model parameter affects only the slice of data from which it derives. The slicing of data determines the $1/f$-filtering properties of the resulting estimates (see Section 4.4.2).

Apart from noise signals, similar slicing is practical for other models also. Detector gains and weights, DC levels, and time weights can be estimated from similar cubical slices (see Sections 4.6 and 4.5. Because each scan comes with a separate set of models (except source), the slicing affects similarly all scans of a data set.

A summary of the SHARC-2 models is provided in Table B.1, indicating the typical slicing (i.e., parameter ranges $\mathcal{R}$) used for each, as well as the relevant gains and other properties. The various model components shown in the table, are discussed in further detail further below.

Modeling. The jth parameter of the mth signal component K_m (see Section 4.2) i.e., $A_{m,j}$, is derived from the jth block with shape and size $\mathcal{R}_m$ (i.e., $\mathcal{R}_m, j$) via any viable statistical estimator of choice (see Section 4.3 for possible statistical estimators). Accordingly, the estimation process of all models $\mathbf{A}_m$ is fully described by its mapping $\mathcal{M}_m$, or equivalently by the slicing into cubical regions $\mathcal{R}_{m,j}$ and corresponding gains $g_{m,j,tc}$

Ordering. The ordering of component models is an important aspect of configuring effective pipelines. The following considerations need to be taken into account for determining an optimal pipeline order.

1. Modeling of bright signals generally precedes that of the fainter ones (see Section 3.5.2.1 and also 4.5).

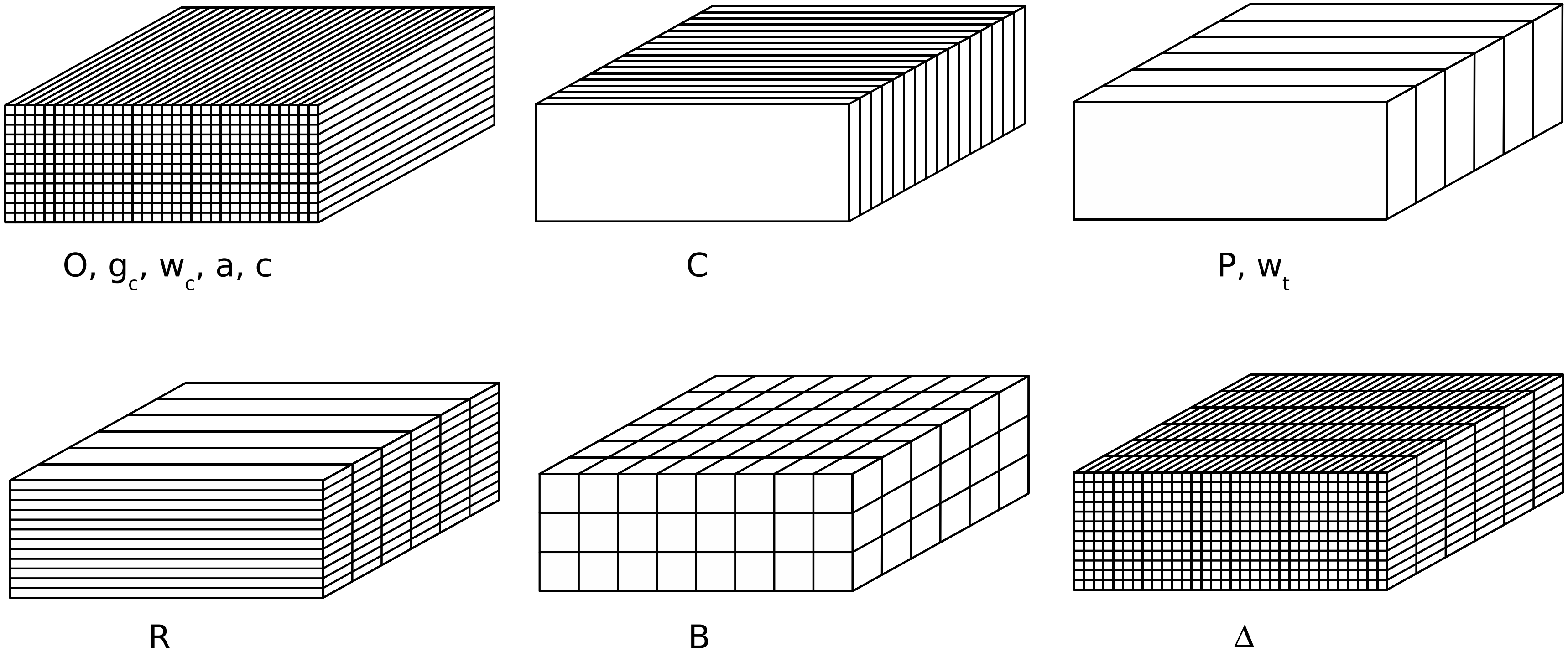


Figure B.3 Slicing Example for SHARC-2 scan data. Schematic examples of the different data slicings which are used in the SHARC-2 reduction. The symbols correspond to pipeline elements, which are summarized in Table tab:SHARC2-pipeline. The slicing determines the effectiveness of $1/f$-type filters, while the ordering of estimator vectors in the pipeline (see Section 4.4 on filtering in CRUSH) determines which act as filters on the source model (or one another) and which do not.

2. Accurate gain knowledge is typically needed for a sufficient removal of bright signals. Thus, it is recommended that faint signal modeling is delayed until the desired gain convergence is achieved (see Section 3.5.1).

3. Priorly estimated components impose filtering on the later component estimates (see Section 4.4.1 and also 4.2).

4. It may not be necessary to model the very faint signals when these are insignificant in comparison to the source signals. Exception can be made if cleaner residuals are desired nonetheless (see Section 4.4.1.2).

To satisfy constraints, some pipeline elements may be delayed until later iterations before estimation is attempted. Consider, as an example a typical pipeline (fig. B.5) that was used for a $\simeq$20 Jy point source in good weather ($\tau \approx 1$). For models and conditioning pipeline elements (e.g., consistency tests) that are associated with faint signals, an upper limit to χ^2 (see Section B.4) may be used as a condition before solutions are to be attempted.

Iterations. Once an optimal pipeline order has been established, one must determine how long to keep iterating. Here, the considerations are:

1. Achieve sufficient convergence on gains (Section 3.5), weights (Section 3.4.1) and data flagging.

2. Reach a desired level of source filtering via the models whose estimation has preceded the first source estimate (see Section 4.4.1.2). Most filtering is provided at the first source model generation. Subsequent iterations may produce more optimal filtering, although at more relaxed levels the further one iterates.

Generally, only a handful of iterations are necessary to reach optimal solutions. For SHARC-2 data, 5–12 iterations, with 3–5 source model generations, are typically used.

B.2 SHARC-2 Pipeline Elements

A summary of the SHARC-2 pipeline elements (except source) is shown in Table B.1. A detailed discussion of each of the component follows below.

B.2.1 Signal Models

Pixel Offsets. The DC component of the (total power) detector signals. In SHARC-2 this consists of a part subtracted in hardware via DACs, and a residual which needs to be 'measured' in the

Table B.1. Summary of SHARC-2 Pipeline Elements

		characteristic $\mathcal{R}$							
Component Name	Symbol	n_x	n_y	n_t	gain	typ. flux	α	ext. int.	condition
Pixel Offsets	O	1	1	N_t	1	100 Jy	$\cdots$	none	
Correlated Noise	C	32	12	1	$G(D_k)$	10 Jy	2	$\gtrsim 2'$	
Pixel Weights	w_c	1	1	N_t	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\widehat{\chi}_r^2 < 2$
De-spiking	σ	1	1	1	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\widehat{\chi}_r^2 < 2$
Pixel Gains	g_c	1	1	N_t	$\cdots$	$\cdots$	$\cdots$	$\cdots$	C modeled.
Time Weights	w_t	32	12	10-100	$\cdots$	$\cdots$	$\cdots$	$\cdots$	
Sky Polynomials (order n)	$P(n)$	32	12	1-10	$g_c\overline{\Gamma}_c$	0.1 Jy	1	$\gtrsim 1'$	
Row Bias Drifts	R	32	1	10-30	$g_c\overline{\Gamma}_c$	0.02 Jy	1	directional	
Detector Drifts	Δ	1	1	300-1000	1	<0.01 Jy	1	$n_t \Delta_t v$	
Spectral Test	Σ	1	1	$n_f = 1, 2, 4, ...N_f$	$\cdots$	...	$\cdots$	$\cdots$	$\widehat{\chi}_r^2 < 2$
Temporal Test	T	1	1	$2, 4, 8...n_t(\Delta)$	$\cdots$	...	$\cdots$	$\cdots$	$\widehat{\chi}_r^2 < 1.2$
Block Correlations	B	4-8	4-6	1-10	$g_c\overline{\Gamma}_c$	<0.01 Jy	?	$n_x \times n_y$	
Acceleration Response	a	1	1	N_t	$g_c\overline{\Gamma}_c$	<0.01 Jy	$\cdots$	bias?	$\widehat{\chi}_r^2 < 1.2$
Chop Residual	c	1	1	N_t	$g_c\overline{\Gamma}_c$	<0.01 Jy	$\cdots$	bias?	$\widehat{\chi}_r^2 < 1.2$

Note. — Summary of SHARC-2 Models. The various pipeline elements (except for the source model) that are used in the reduction of SHARC-2 data are listed. Each of these are estimated for each scan separately in the data set. Nonsignal model names (e.g., weighting, gain estimation and consistency checks) appear italicized. The cubical blocks of $n_x \times n_y \times n_t$ provide the typical characteristic data ranges $\mathcal{R}$ used for the model parameters in the SHARC-2 reduction. These together with the signal gains provide the mapping $\mathcal{M}$ of the modeled signal parameters into the time stream. Also shown are the typical equivalent $350\,\mu$m fluxes of the modeled signals per SHARC-2 pixel, and the power index α for the characteristic $1/f^{\alpha}$ type noise power spectra. The last two columns indicate whether the models impose filtering of the source when precede that in the pipeline, and the typical conditions used for evaluating whether the particular modeling should be attempted in a given iteration.

reduction software. The hardware subtracted part with $\simeq$10 mV is equivalent to $\simeq$$10^4$ Jy signals, while the residual DC signal is typically on the order of 100 Jy per detector pixel. The model O thus is a single DC value per pixel for every scan.

The DC offsets are a function of the loading power P_{load} and the bias voltage V_{bias}. Thus,

$$O_c = f_c(P_{load}, V_{bias}), \tag{B.1}$$

with a different appropriate functions f_c for every detector channel c. The DC levels can be used to provide a direct measure of the detector loading when f_c are invertible under the bias used. It is thus possible to provide direct line-of-sight measures of the atmospheric optical depth τ, once f_c are adequately characterized (Section B.3.1). *CRUSH* has this capability for SHARC-2 $350\,\mu$m operation in good weather, i.e. $\tau \lesssim 1$ under the standard detector biasing of operation.

Correlated Noise. Most of the atmospheric variation is correlated among the full SHARC-2 detector array, and exhibits a $1/f^2$ Brownian spectrum. Therefore, the correlated component C is modeled for each 36 ms SHARC-2 across all the detectors. Since the signal is dominant, with typical RMS of 10 mJy min$^{-1/2}$, it is important to use accurate loading dependent signal gains $G(D_k)$.

Sky Polynomials. One way for modeling higher order atmospheric structures (cf. fig. B.3) is to fit 2-dimensional polynomials P every so often to the entire array. Typically 1st or 2nd order polynomials are useful. When the modeling of spacial-temporal atmospheric structures precedes the source, it will cause the source model to be filtered as far as these polynomials are able to describe brief exposures of the source in the time stream.

Row Bias Drifts. The detectors of SHARC-2 are biased at a constant voltage, with all detectors of a row supplied with the same bias. Because the detector bias is related to the DC level, small bias fluctuations can result in signal artifacts. With DC offsets typically on the 10^4 Jy equivalent flux scale, bias fluctuation on the order of 10 ppm can cause electronic noise signals on the 100 mJy level. These exhibit a canonical $1/f$, or pink noise, spectrum typical to electronic noise. Modeling row bias drifts R requires determining one bias offset per detector row at regular intervals. When the modeling of these bias drifts precedes the modeling of the source, total filtering of the source structures in the direction perpendicular to the rows results. However, because this is an anisotropic noise source, the filtering can be avoided entirely by rotation of the field or the detectors relative to one another (see Sections 3.6.6.1 and 4.4.1.3), in a single coherent reduction of the dataset.

Detector Drifts. Unfortunately, the SHARC-2 detectors themselves are not stable over time scales of $\gtrsim$30 s. For reaching the sensitivity limit of the instrument, it may be necessary to model (i.e., filter) the $1/f$ component. This is implemented by estimating a drift value Δ every so often ($\lesssim$30 s) to each pixel. When preceding the source modeling, this will filter all source structures that are larger than the distance swept by the scanning detectors in the time scale of the estimates. Note, that consequently, sensitive scan patterns should loop around within this limiting time scale.

Block Correlations. An alternative to polynomial fitting, and truer to the *CRUSH* approach, higher order atmospheric structures are modeled trough low resolution estimates of the scan data (cf B in fig. B.3). Individual parameter estimates are derived for a block of adjacent detectors (e.g., 8×6 pixels are typically used), averaged over some small number of 36 ms samples. Because the modeling has to be performed on very short timescales to be effective, the detector movement is typically negligible and the source filtering is mainly imposed by the estimator block size in the slicing (see Section 4.4.2) when modeling precedes that of the source. For 8 pixel blocks, spacial components (e.g., Fourier components) larger than about 15 pixels (i.e., $\gtrsim 1'$) will end up completely filtered.

Acceleration Response. The SHARC-2 time stream is suspected to contain signals related to telescope acceleration. One reason maybe that semiconductor bolometers are susceptible to microphonics. More significantly perhaps, the surface of large submillimeter telescopes (e.g., the CSO) could wobble when accelerated in pursuing a scanning pattern. With the wobbling the illumination $d\Omega_c$ (see below in *Pixel Gains* of the detectors is expected to change systematically. Assuming the change is anisotropic i.e., dependent on the magnitude but not the direction of telescope acceleration a, the illumination $d\Omega_c$ of detector c can be approximated locally by a quadratic form as,

$$d\Omega_c(a) = d\Omega_{0,c} - \omega_c a^2, \tag{B.2}$$

around the optimal static values $d\Omega 0, c$, which are known either from measurements or theoretical calculations (e.g., optical modeling). Given that changes in the illumination modulate the optical loading P_{load} of the detectors, which is otherwise nearly constant throughout the typical SHARC-2 scan, such variations result in effective component signals (see Section 4.2) as,

$$K_{tc} \approx \omega_c' \, a_t^2. \tag{B.3}$$

Thus, the coefficients ω_c' can be estimated (Section 4.3) for each detector from its full time-series signal. When these estimates precede the source modeling some interference with the source structures is expected but may be difficult to characterize. However, because acceleration changes the detector illumination one-sidedly, some sort of a bias may result in the source model, when that follows the acceleration-related terms in the pipeline.

Chop Residuals. If a chopping secondary is used instead of, or in addition to, a scanning primary, similar changes in the illumination (as above) are to be expected. Thus, for a well-tuned and symmetric secondary mirror, such that the static position is optimal, the illumination is a function of the deviation $d\theta$ from the optimal position and thus,

$$d\Omega_c(d\theta) = d\Omega_{0,c} - \kappa_c d\theta^2 \tag{B.4}$$

leading to time-series (component) signal artifacts from chopping which are,

$$K_{tc} \approx \kappa_c' \, d\theta_t^2 \tag{B.5}$$

Thus, the constant κ_c' can be estimated in for every detector using all samples from a scan. Similar biases may result as for the *Acceleration Response* above.

When the optics is not symmetric it may be necessary to include a linear term too, leading to signal artifacts of

$$K_{tc} \approx \lambda_c \, d\theta_t + \kappa_c' \, d\theta_t^2 \tag{B.6}$$

This is the more applicable case for SHARC-2, owing to the astigmatic secondary mirror of the CSO. In fact, for small chopper excursions, the quadratic term is assumed to be negligible in comparison to the linear one. Either or both relevant coefficients can be estimated (Section 4.3) for each detector for the duration of a scan.

B.2.2 Supplementary Models

Pixel Gains. The full loading dependent SHARC-2 detector gains for extended input signals (e.g., atmospheric loading, the Moon, or some extended GMC) can be expressed in terms of the illumination of the detector $d\Omega_c$ i.e., the integrated solid angle covered by the detector or its beam, the loading functional form $\Gamma(D_{tc})$, and the scaling constant g_c, as

$$G_c(D_{tc}) = g_c \, d\Omega_c \, \Gamma_c(D_{tc}) \tag{B.7}$$

With the strong correlated atmospheric background affecting all detectors simultaneously, the estimation of the scaling constant g_c becomes possible from the correlated parameters C_i, via the gain coupling $\gamma_{i,tc}$,

$$\gamma_{i,tc} = \delta_{i,t} \, d\Omega_c \, \Gamma_c(D_{tc}) \tag{B.8}$$

substituted in eqs. (4.92)-(4.95). The delta function simply associates a correlated parameter C_i with a single exposure at time sample t, which is its natural connection (see *Correlated Noise* above). Section 4.6.1 offers more detail on gain estimation.

When the illumination of detectors is approximately uniform ($d\Omega_c \approx d\Omega_{c'}$), and the loading variations during the scans are small ($\delta\Gamma_c(D_k)/\overline{\Gamma}_c \ll 1$), then the actual signal gains are crudely approximated using the mean loading term $\overline{\Gamma}_c$ and the scaling constants g_c as

$$G \approx g_c \overline{\Gamma}_c. \tag{B.9}$$

For SHARC-2 $350\,\mu$m data this approximation is typically accurate to 2–3 significant figures under the typical atmospheric loading (emissivity) variations of <1% in a 10 minute scan. Thus, the approximate gains can be used for the adequate subtraction of the weaker ($\lesssim$1 Jy) signals, even when recovering the faintest ($\simeq$10 mJy) targets. Because the electronic gains for all detectors channels are effectively the same (within the 2–3 digit accuracy) all gains can be refered to the input, and thus the use of $g_c\overline{\Gamma}_c$ are generally adequate for all weak signals (see see Section 3.5.2).

When the loading dependence $G(D_{tc})$ is well characterized, it provides an absolute gain calibration. The fitting of gain constants is useful only for relative fine-tuning of the detector gains (see Section 4.6.1). *CRUSH* provides the built-in gain calibration for SHARC-2 $350\,\mu$m data based on the DC offsets (which typically trump all other signals i.e., $D_{tc} \approx O_c$). However, the final calibration of the source image has further gain dependence on the atmospheric optical depth τ and the beam efficiency η (see Section B.3). Thus, without accurate characterization of these extra terms, the built-in gain calibration is no substitute for frequent calibrating against objects with known $350\,\mu$m fluxes.

Pixel Weights. Pixel weights are used for characterizing the final noise properties of the residuals under the time variant vs. detector noise separation assumption (see Section 4.5.1). Because effective reductions result in filtering that is approximately noise whitening, the weights can be calculated under the white noise assumption, with transient properties absorbed into the time weight terms (see below). A few iterations over the more dominant signals are typically necessary to arrive at a consistent set of weights and signal models (cf. weighting–filtering cycle in Section 3.4.1). The estimation of weights should generally wait until the dominant signals are removed from the weighting domains, whether temporal or spectral (see Section 4.5), so that the white noise assumption is approximately correct.

Time Weights. Time weights are a complement to detector weights (above), and are meant to deal with transient noise features in the data steam. Thus similar considerations apply. Time weighting is described in Section 4.5.1.

B.2.3 Consistency Checks & Flagging

A critical aspect in the reduction of weak signals is to be able to identify data (e.g., detector channels) which are unreliable. In the SHARC-2 reduction a number of tests serve this purpose. These tests are inserted into the pipeline and provide flagging (Section 3.3.4) of questionable detectors or time samples. Tests are used both for flagging and unflagging of suspected detector channels, depending whether they pass or fail the test at every round. Thus the iterations may evaluate models over different sets of detectors. It is generally best to arrive to a stable set of active detectors and data points before proceeding to the critical faint source signals (esp. in deep mode reductions).

Residual Spikes. The simplest test is to search for far outliers in the data which are inconsistent with the measured white noise properties (see weighting in Section 4.5) of the signals. For example, under the Gaussian noise $5\,\sigma$ outliers are expected only in every couple of million data points only. Farther outliers should be even more rare. Thus, a relatively more frequent occurrence of residual farther-lying data may indicate a detector with unusual behavior. Accordingly, the first few extreme outliers are flagged point-for-point in the SHARC-2 reduction, but once these exceed a critical user-defined fraction of the detector data, the entire detector channel is flagged.

Temporal Test. Similar to the search for spikes, but instead of looking at individual data points the test is performed on averages derived from data blocks of varying sizes. Performed for each detector separately, the these blocks involve varying numbers of time samples for a given detector channel. Thus, the test is effective for identifying transient features (of all sizes) in the detector time-streams. As the *CRUSH* reduction filters low frequency components heavily (see filtering, Section 4.4.2) it makes sense to try window sizes between 2 samples and the window size used for the detector drifts, because these will filter low frequency noise on the shorter times scales and all spacial scales. The data blocks in this test may overlap to preference. Detectors which fail the test are flagged (see Section 3.3.4).

Spectral Test. Very much like the *Temporal Test* but operating on the spectral domain obtained by the Fourier transforms of the detector signals along the time direction. The spectral window sizes can vary between 1 frequency channel to $N_f/2$.

B.3 The Source Model

As mentioned the source model alone ties a set of scans together, and is thus derived from the entire data cube (with all scans). Otherwise all other modeling and testing is performed separately for each scan. The mapping of the time series (tc indeces) into map pixels (with xy indeces) is via,

$$\mathcal{M}_{tc}^{xy} = \delta_{tc}^{xy}\, g_{tc}. \tag{B.10}$$

The Kronecker delta expresses that data points are uniquely associated with map pixels (for simplicity!). With coinciding data and map indeces, the data is mapped with an appropriate gain g_{tc} into the source map $\mathbf{S}_{xy}$. This gain is related to the atmospheric gains $G(D_{tc})$ (see *Correlated Noise*) but with the extra terms for atmospheric extinction (via the time variable optical depth τ_t) and for detector beam efficiency η_c,

$$g_{tc} = \eta_c\, e^{-\tau_t}\, G(D_{tc}). \tag{B.11}$$

Just like $G(D_{tc})$ can be replaced simply by $g_c\overline{\Gamma}_c$ for fain signals, the same can be done when the source signals are small relative to the variation of sky (C) in magnitude, leading to faint source gains of,

$$g_{tc} = \eta_c\, e^{-\tau_t}\, g_c\overline{\Gamma}_c. \tag{B.12}$$

With the mapping (data ranges via the delta functions, and gains) defined, the source estimation proceeds as usual, using a statistical estimator of choice, such as a maximum-likelihood estimate (Section 4.3.1), maximum-entropy estimate (Section 4.3.2) or robust median estimate (Section 4.3.3).

However, since the sensitive and accurate recovery of the source signal is the objective of the experiment, it plays a more special role in the pipeline than the other signal (and noise) components. Specifically, some manipulation of the source image may be desirable beyond the pure statistical estimation, to make it more “physical” to the observer. Some of the useful postprocessing operations are:

- Leveling, since the constant source term is irrecoverable due to $1/f$-type noise interference.
- Noise re-estimation, because uncertainties are calculated (according to eq. [4.21]) with weights in the white noise assumption, which may not characterize the resulting image.
- Smoothing and filtering to improve appearance and/or detection significance of the images.
- Masking, to concentrate on the recovery of selected image regions only.
- Clipping, especially to discard insensitive parts of the image.

Some of these operations are presented in further detail below. Note, that calibration of the source fluxes is strongly dependent on the gain terms provided by the optical depth τ and the beam efficiency in the field over the detector η_c, which are not determined from the data otherwise. Of the two terms, the estimation of τ may be possible from the data, either entirely, or differentially around an otherwise determined mean value $\overline{\tau}$ (Section B.3.1). The beam efficiency, on the other hand, needs to be measured externally (e.g., by frequent observation of calibrator sources with known $350\,\mu$m fluxes!). The variations of the focus quality, affecting the main beam efficiency of the telescope, is likely the strongest contributor to the blind calibration uncertainty of SHARC-2 images. With the SHARC-2 focus models after April 2005, blind calibrations of ≃15%(!) seems possible, provided a reasonable guess at τ is supplied.

Leveling. Strictly speaking images recover not absolute fluxes but rather a relative distribution of those. Flux measurements are possible via comparison of different image region relative to one another. Specifically, by identifying a part of the image as “blank sky” source fluxes can be deduced by comparing flux levels relative to that patch. However, provided that most SHARC-2 images contain mostly blank sky, its level can be estimated reasonably via a robust estimate, such as a weighted median of the image pixel fluxes (Section 4.3.3). This way, the resulting image fluxes are generally accurate measures of the absolute source brightness. *CRUSH* levels the derived images in this way for SHARC-2 data. However, this convenient built-in leveling feature should not serve as a substitute to sound judgment by the astronomer.

Noise re-estimation. Noise estimates on the image pixels are calculated statistically from the time-stream noise properties (eq. [4.21]), by assuming that the residual noise is well characterized as Gaussian. As such, these are only guiding numbers for images which may have more complex noise structures. When the image noise is uniform in nature, and variations are caused only because of the varying coverage of the different image pixels, the white noise estimates may be scaled by an appropriate constant to provide reliable image noise estimates. The scaling constant can be determined, as an example, from the median χ^2 deviation (see Section 4.5.3). *CRUSH* offers this possibility for SHARC-2 images.

If noise adjustments are done scan-by-scan (on the partial maps yielded by every scan), based on a median or average χ^2_n for scan n, it makes sense to adjust the weights used in constructing the source image as,

$$w_{tc} \longrightarrow \frac{w_c\, w_t}{\left\langle \chi^2_n \right\rangle}. \tag{B.13}$$

Smoothing and Filtering Smoothing is a convolution by a beam-shape $B(\mathbf{x})$

$$S'(\mathbf{x}) = S(\mathbf{x}) \otimes \beta(\mathbf{x}), \tag{B.14}$$

where $\beta(\mathbf{x})$ is the integral normalized beam,

$$\beta(\mathbf{x}) = \frac{B(\mathbf{x})}{\int d\mathbf{x}'\, B(\mathbf{x}')}. \tag{B.15}$$

The normalization is intended to preserve integrated fluxes.

Since $B(\mathbf{x}')$ represents the response to a point source at an offset of $\mathbf{x}'$, $B(\mathbf{x}')$, it is to be treated as a gain term. As such, the weighted convolution becomes,

$$S'(\mathbf{x}_i) = \frac{\sum_j w(\mathbf{x}_i - \mathbf{x}_j)\beta(\mathbf{x}_j)S(\mathbf{x}_i - \mathbf{x_j})}{\sum_j w(\mathbf{x}_i - \mathbf{x}_j)\beta(\mathbf{x}_j)}, \tag{B.16}$$

according to to weighted averages replacing arithmetic ones (see Section 3.3). Interestingly, under the integral normalization of eq. (B.15), the normalized beam response $\beta(\mathbf{x}')$ behaves more like a weight than as a gain term, as the particular form of the denominator reveals. In this light, the error on the smoothed image is,

$$\sigma_S^2(\mathbf{x}_i) = \frac{1}{\sum_j w(\mathbf{x}_i - \mathbf{x}_j)\beta(\mathbf{x}_j)}. \tag{B.17}$$

The exact same results are obtained if one fits beam features of shape $\beta(\mathbf{x}')$ to the unsmoothed source map $S(\mathbf{x})$ by minimizing the weighted χ^2 model deviation i.e., producing maximum-likelihood estimates (see Section 4.3.1 on maximum-likelihood estimates). Therefore, images $S'(\mathbf{x})$, smoothed by an effective beam $\beta(\mathbf{x}')$, represent the most probable distribution of sources which are of underlying shape $\beta\mathbf{x}'$ and, therefore, smoothing readily provides a way for point source extraction for images characterized by noise that is independent from map pixel to map pixel (and less strictly map beam to map beam).

Smoothing by Gaussians that are not beam-sized can be used more generally for filtering unwanted high frequency image noise. Varying beam sizes produce quasi-optimal filtering (see Section 3.4.3.2 on optimal weak signal filters) of the images. The wider the Gaussian shape used for smoothing the heavier is the filtering of the higher spatial frequencies. Apart from granting more certain fluxes, smoothing can greatly improve the appearance of SHARC-2 images (see fig. B.4).

Smoothing can also be used for high-pass spatial filtering of images (i.e., discarding extended structures), by subtracting a smoothed image from the original. For example, if one wishes to filter away spatial structures $\gtrsim 1'$ from SHARC-2 maps, it can be achieved by subtracting an image smoothed by $1'$ from the original. *CRUSH* offers both smoothing and high-pass spatial filtering of SHARC-2 images.

Clipping Clipping of underexposed, or otherwise noisy, image pixels may be useful both for improving image appearance, and to discard image pixels that carry unreliable flux measurements. Flux clipping (e.g., the clipping of negative fluxes, or clipping at arbitrary flux levels) can also be useful to provide images which are physically more justified (even if statistically not at all!). Clipping of the negative image fluxes is additionally useful for producing the positive images required for maximum-likelihood estimation (see Section 4.3.2).

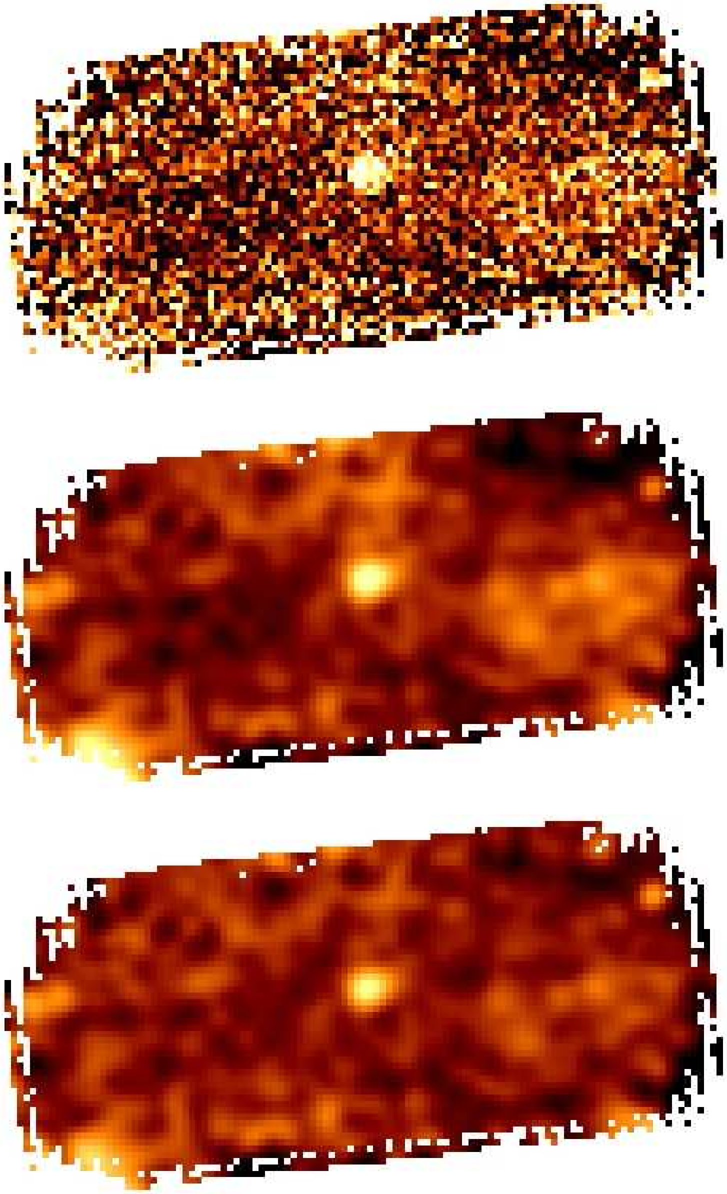


Figure B.4 Smoothing and Filtering of Images (3C345 from scan 10715). The top image is the maximum-likelihood source map, on a 1.6″ grid, smoothed by a 6″ FWHM Gaussian (middle), and filtered additionally for spatial scales over 30″ (bottom). A small ($\simeq 1'$ diameter) circular dip, resulting from the filtering, may be picked out by the careful eye.

B.3.1 Optical Depth

The measurement of the optical depth is critical for the flux calibration of SHARC-2 images. It was shown that DC offsets are loading dependent. The loading power can be broken down into an optical loading component and all else (external) as,

$$P_{load} = P_{optical} + P_{ext}, \tag{B.18}$$

where the optical loading is dependent on the detector illumination $d\Omega_c$ and the apparent sky brightness, which can be approximated by the relevant part of the gray- or blackbody spectrum $B_\nu(T)$ falling inside the detector bands (via a temperature T and an emissivity index β), and the optical thickness τ as,

$$P_{optical} \propto d\Omega_c \left(1 - e^{-\tau}\right) \int_{\text{in band}} \nu^\beta \, B_\nu(T) \, d\nu. \tag{B.19}$$

The integrated emission is considerably simplified in the Rayleigh-Jeans regime, ($h\nu \ll kT$), leading to

$$P_{optical} \propto d\Omega_c \left(1 - e^{-\tau}\right) T \tag{B.20}$$

in that limit. Under the assumption that the optical depth comes from atmospheric layers close to the ground (as expected for water vapor), the temperature T can be approximated by the ambient value T_{amb}. Thus, if P_{excess} is characterized, the DC offsets can be used to provide a direct measure of τ. Comparing to eq. B.1,

$$1 - e^{-\tau} \propto f_c^{-1}(O_c, V_{bias}), \tag{B.21}$$

for individual detectors. Better measures of the optical depth result from combining the measurements of all useful detector pixels into an appropriately formulated weighted average (see Section 3.3). The constants of proportionality come from an external calibration process, e.g., via comparison to independent measurements of the optical depth. This has been done for SHARC-2, and the τ values derived from the DC levels are in generally good agreement with the SHARC-2 in-band optical depth measures inferred from the 225 GHz radiometer- or the $350\,\mu$m tipper data.

Differential Optical Depth Variations Even if external measures of the optical depth are available and utilized, the optical depth may vary significantly over the duration of integrations (see atmospheric variations). Thus, a differential momentary adjustment is possible around the mean τ value determined otherwise.

Consider the breakdown of τ into its mean value $\overline{\tau}$ over the period of the scan, and the differential variation $delta\tau$ around it,

$$\tau_t = \overline{\tau} + \delta\tau_t. \tag{B.22}$$

Then,

$$\delta\hat{\varepsilon}_t = \left\langle G_c(D_{ct})\,\frac{\partial\varepsilon_t}{\partial O_c}\right\rangle C_t, \tag{B.23}$$

assuming that C_t is a zero-mean signal over the period of a scan. Here, the emissivity ε of the atmosphere is,

$$\varepsilon = 1 - e^{-\tau} \tag{B.24}$$

A solid measure of the differential emissivity variation is obtained by the appropriate weighted averaging (Section 3.3) of the values provided by the individual detectors.

B.4 χ^2 Estimation

When weighting is appropriately performed using only the white-noise dominated regions of the data (e.g., by spectral weighting of stationary noise), then a true estimate of the chi^2 becomes possible. The arithmetic definition of the reduced chi^2 i.e., $\chi_r^2 N_{df} = \sum_k |R/\sigma|^2$, can be translated to weighted data (see Section 3.3),

$$\widehat{\chi}_r^2 = \frac{N_k}{N_{df}}\,\frac{\sum_k w|R|^2}{\sum_k w} \tag{B.25}$$

Deviations from the expected $\mathbf{E}(\chi_r^2) = 1$ are indicative of the level of excess noise remaining unfiltered in the data stream. Optimally configured reduction pipelines should produce $\widehat{\chi}_r^2 \approx 1$ i.e., residuals that exhibit uniform white noise (see Section 4.4.1). Larger $\widehat{\chi}_r^2$ values are indicative of insufficient noise filtering while too low $\widehat{\chi}_r^2$ values reveal overfiltering as compared to optimally filtered data. Accordingly, appropriate adjustments should be made to model parameters (e.g., time- and spacial resolutions) and pipeline configurations (e.g., ordering, iterations and specific model activations). See Section 4.4 for a detailed discussion of the filtering properties of iterated reduction pipelines.

The $\widehat{\chi}_r^2$ measure may be additionally useful for providing an objective criterion for evaluating whether certain pipeline elements are ready for estimation, or should be delayed until later iterations. See Table B.1 for typical conditions used in the reduction of SHARC-2 data.

B.5 SHARC-2 Example Pipelines

Example 1: Default Pipeline. The following pipeline is typically used for the default reduction of SHARC-2 data. Each iteration consists of the sequence (cf. 4.40),

$$O\;C\;w_c\;\sigma\;g_c\;T\;w_t\;c\;P\;R\;\mathbf{S}\;\Sigma\;\Delta\;B\;a. \tag{B.26}$$

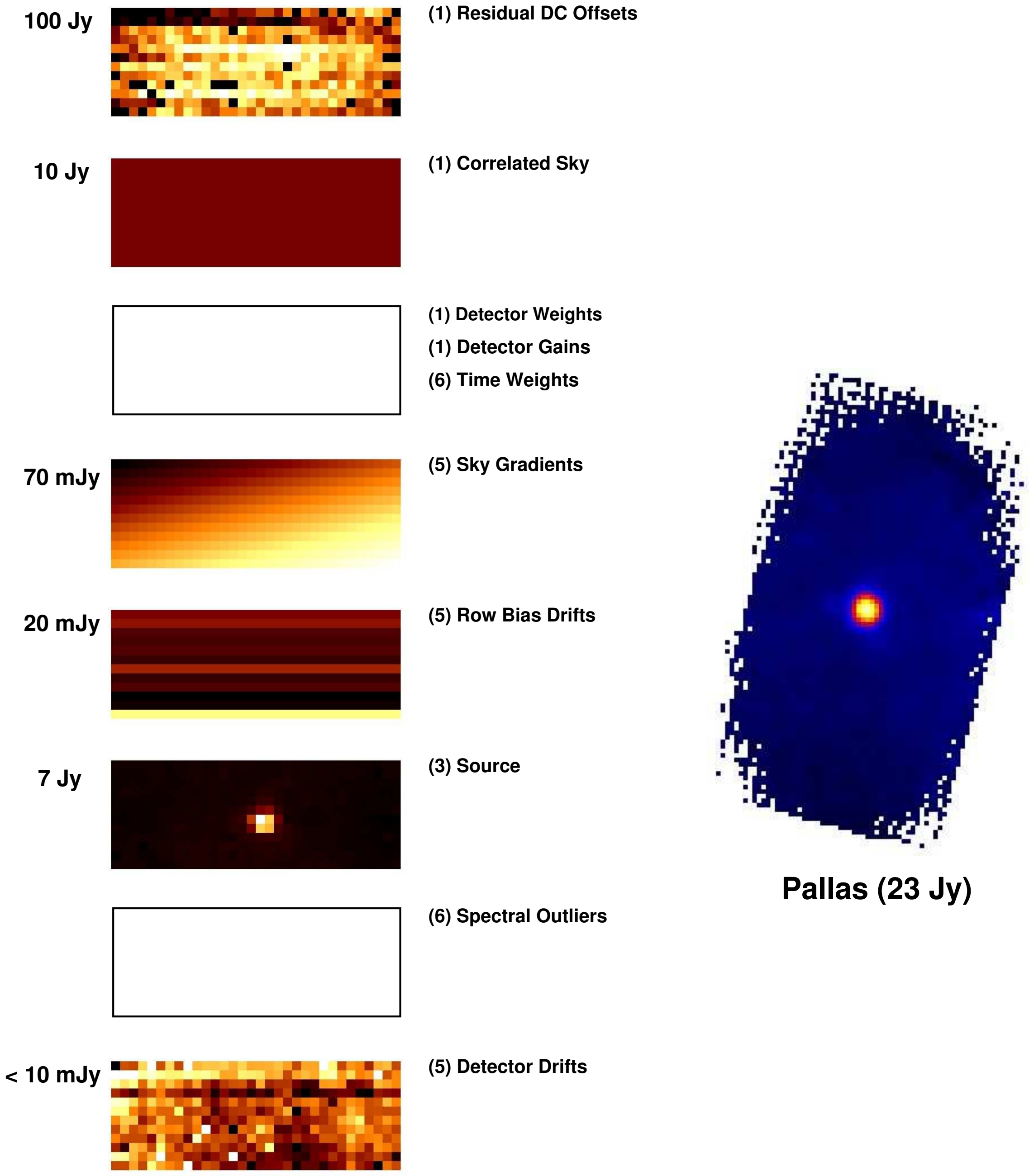


Figure B.5 Modeling example of a SHARC-2 scan (Pallas). The series of images on the left show the signal models projected back onto the SHARC-2 array for a single 36 ms SHARC-2 exposure of Pallas (frame 256 of scan 22813), resulting from a typical reduction (Example 1). The images are not shown on the same color scale. Instead, their typical equivalent RMS fluxes are indicated to the left. For the source flux the apparent peak flux is shown for comparison, as observed with atmospheric extinction. The iteration in which the models are first derived are shown in brackets. In this case no real filtering of the source is imposed by the other models, which are delayed in the pipeline until after the source model extraction.

The source **S** is inserted based on a preliminary assessment of its brightness. Every element of the sequence may be skipped over in the particular iteration. What elements are activated and when may be guided by conditions, such as a maximum $\hat{\chi}^2$ requirement. Thus, for the particular reduction of Pallas (scan 22813), a 23 Jy point source, the following model sequence was used.

Iteration 1-2: $O\ C\ w_c\ g_c$
Iteration 3-4: $O\ C\ w_c\ g_c\ \mathbf{S}$
Iteration 5: $O\ C\ w_c\ g_c\ P(1)\ R\ \mathbf{S}\ \Sigma\ \Delta$
Iteration 6-8: $O\ C\ w_c\ g_c\ w_t\ P(1)\ R\ \mathbf{S}\ \Sigma\ \Delta$
Final: **S**

As filtering of the source (**S**) is provided by those signals models only which precede it in the unfolded pipeline (Section 4.4.1.2), in this particular case only O and C impose filtering of the source map, leaving most of its extended structures unperturbed. Particularly, the spatially more sensitive models P and R have been delayed until 2 source generations have been obtained, such that these do not filter spacial source structures. A faithful, although possibly too noisy, a source reconstruction is achieved thus. The spacial sky structure was modeled through a 1st order polynomial $P(1)$ i.e., as gradients across the array. The reduction resulted in a final $\hat{\chi}^2$ of 1.385. The source image is smoothed by a Gaussian with 4" FWHM.

Example 2: Compact Sources. The same sequence is used as above, but this time allowing for spacial filtering of the source image. Specifically, P and R now precede the first source estimate, leading to the unfolded pipeline as,

Iteration 1-2: $O\ C\ w_c\ g_c$
Iteration 3-4: $O\ C\ w_c\ g_c\ P(1)\ R\ \mathbf{S}$
Iteration 5: $O\ C\ w_c\ g_c\ T\ w_t\ P(1)\ R\ \mathbf{S}\ \Sigma\ a\ \Delta$
Iteration 6-8: $O\ C\ w_c\ \sigma\ g_c\ T\ w_t\ P(1)\ R\ \mathbf{S}\ \Sigma\ a\ \Delta$
Final: **S**

This time, a final $\hat{\chi}^2$ of 1.081 was reached. The spacial filtering of the source is apparent, but point source peak fluxes can be correctly estimated from scaling the final image by $Q = 1.20$ (see Section 3.4.4 on corrections). The source image is smoothed by a 6" FWHM Gaussian.

Example 3: Deep Fields. At times, the remapped noise from the time-stream is still too high for the scientific results to show. Then, even more aggressive filtering may be necessary. The following pipeline is used typically for deep field point sources (in this case a 10 minute scan in the Lockman Hole):

Iteration 1-2: $O\ C\ w_c\ g_c$

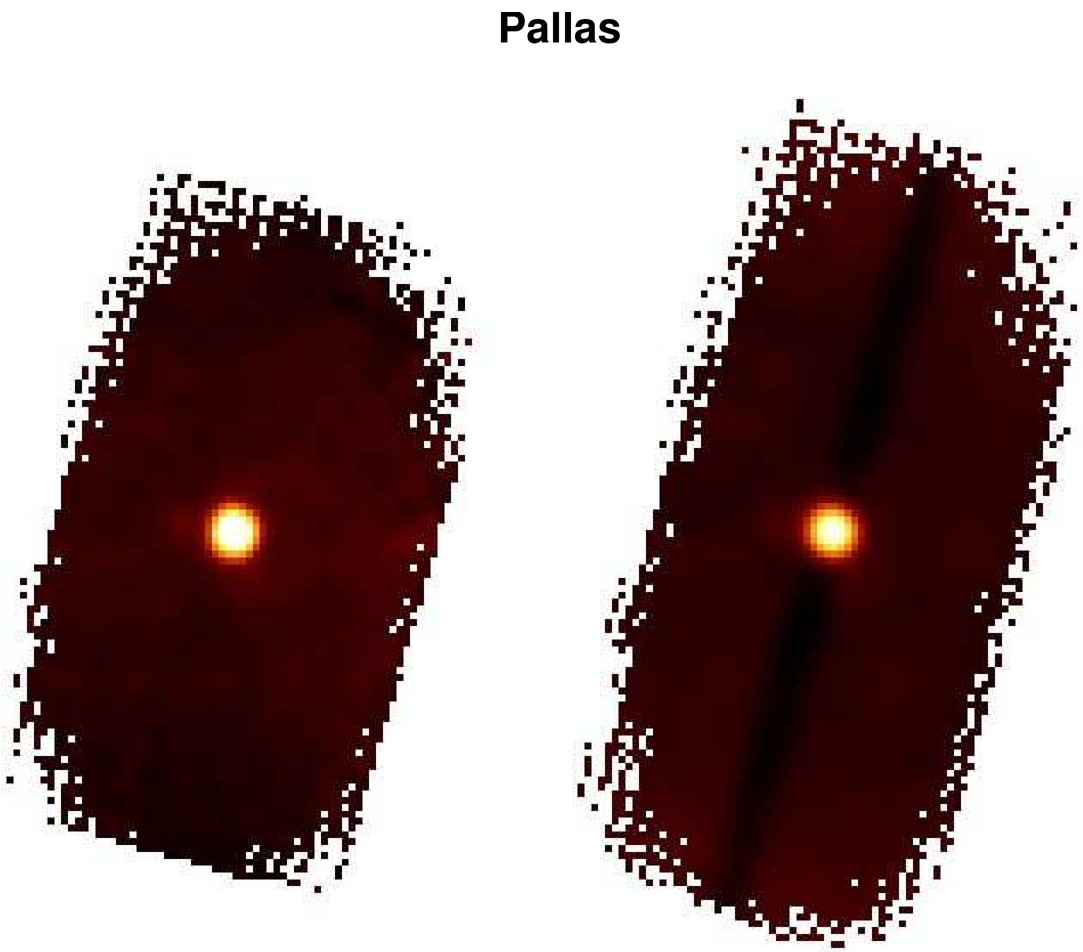
Iteration 3: $O\ C\ w_c\ g_c\ P(2)\ R\ \Delta\ \Sigma$

Iteration 4: $O\ C\ w_c\ g_c\ T\ P(2)\ w_t\ R\ \Delta\ \Sigma\ a$

Iteration 5-9: $O\ C\ w_c\ \sigma\ g_c\ T\ P(2)\ w_t\ R\ \Delta\ \Sigma\ a$

Iteration 10-12: $O\ C\ w_c\ \sigma\ g_c\ T\ P(2)\ w_t\ R\ \Delta\ \Sigma\ a\ \mathbf{S}$

Here a a second order polynomial $P(2)$ is used for the sky structure, and all models present filtering of the source, which is solved only in the last three iterations to minimize filter evolution. Still additional filtering is possible via B a higher resolution sky model than $P(2)$, or by using a higher order polynomial. The final $\hat{\chi}^2$ was 1.050 in this case. The filtered peak fluxes are recovered for point sources by scaling the final image with $Q = 1.29$ (see Section 3.4.4 on corrections). The source image is smoothed by the beam (8.5″ FWHM Gaussian), and thus represent maximum-likelihood beam fits to the data.



Example 1 **Example 2**

Figure B.6 Example reductions of the asteroid Pallas (a 23 Jy point source from scan 22813). The heavier filtering via the modeling of the row bias drifts preceding the source is apparent in the dark 'trench' around the source in the direction of the SHARC-2 rows. Such filtering is effortlessly avoided by *CRUSH* when significant field rotation is present in the coherently reduced data set (see Section 4.4.1.3). The differing map sizes owe to different sets of pixels being flagged in the two example reductions.

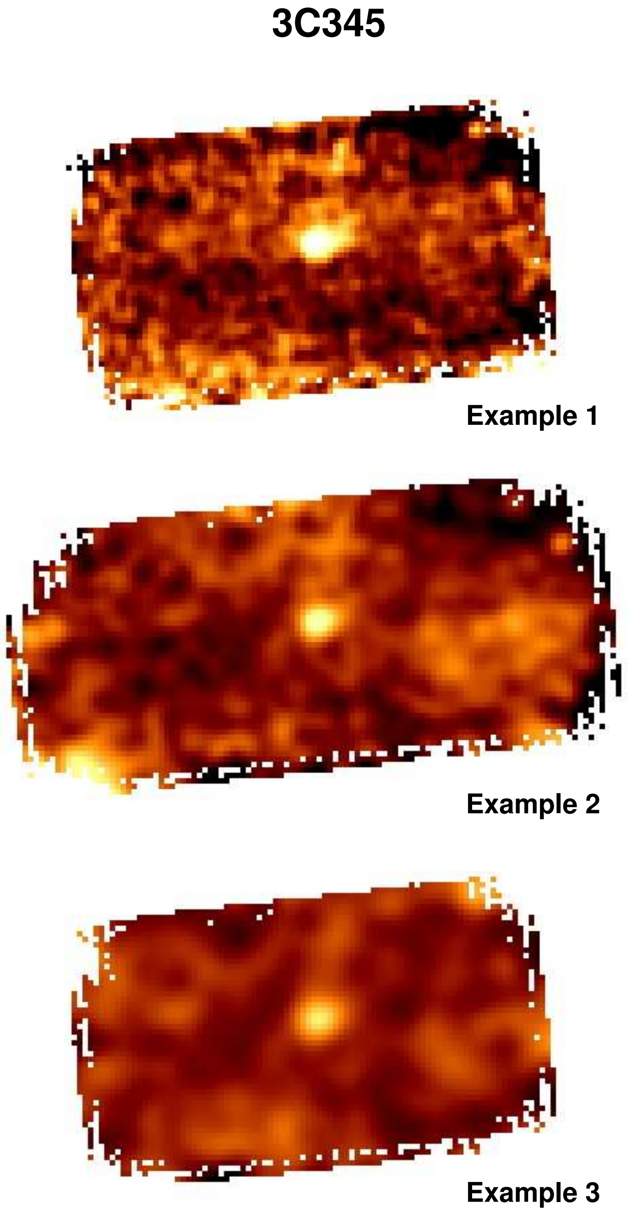


Figure B.7 Example reductions of 3C345 ($\simeq 0.8$ Jy point source, from scan 10715). As more filtering is imposed by the various signal models preceding the source, the flatter backgrounds are achieved. At the same time, the with heavier smoothing applied (from top to bottom), the peak fluxes decrease, although the integrated flux measures stay the same. The differing map sizes owe to different sets of pixels being flagged in the example reductions. The deep reduction (Example 3, bottom) has 4 columns on each side of the array clipped explicitly for producing cleaner images.