MICROWAVE HOLOGRAPHY OF LARGE REFLECTOR ANTENNAS

by

Geoffrey Carlyle James, B.Sc. (Hons.)
Department of Physics

Submitted in fulfilment of the
requirements for the degree of
Doctor of Philosophy

University of Tasmania
Hobart
August, 1991

This thesis contains no material which has been accepted for the award of any other higher degree or graduate diploma in any tertiary institution. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person, except when due reference is made in the text.

Abstract:

MICROWAVE HOLOGRAPHY OF LARGE REFLECTOR ANTENNAS

Geoffrey C. James

Department of Physics, University of Tasmania,
G.P.O. Box 252C, Hobart 7001, Australia.

This thesis describes experiments in Microwave Holography using the 26 m radio telescope at the Mount Pleasant Observatory, and the development of data processing methods to overcome a major limitation of traditional holographic measurements.

Microwave Holography is a fast and effective technique for measuring the surface profile of reflector antennas. It requires measurement of the amplitude and phase of the antenna's far field radiation pattern. The Fourier transform relationship of the far field pattern to the aperture current function is used to estimate the aperture phase profile, which can be related directly to the surface profile. Using an unmodulated 12.7 GHz carrier signal transmitted by the AUSSAT-1 geostationary satellite, the 26 m diameter surface was measured to an accuracy of ± 53 μm with a resolution of 0.6 m. Using a beacon signal, weaker by 20 dB, gave a comparable accuracy of ± 64 μm. Each map took about 4 hours to record.

In practice only a small part of the complete radiation pattern can be measured, so information about high spatial frequencies in the aperture is lost. This causes detailed structure in the surface error map to be smoothed out. Most large reflectors are constructed with panels, and misaligned panels give discontinuities in surface error which cannot be resolved. This can lead to incorrect assessment of panel positions.

Significant improvement can be obtained by recognizing at the outset that the reflector surface is not smooth and continuous, but a collection of individual rigid panels. The Method of Successive Projections is an algorithm, with a simple geometrical interpretation, which allows information about the panel boundaries to be readily incorporated into the data reduction process. The algorithm is straight-forward to apply and is very flexible: aperture blockage effects do not disturb its operation, and panel distortions can be included in the analysis with minimal extra effort. Good results may be achieved while minimizing measurement time. This is an important consideration for high-performance antennas with a busy operating programme.

Acknowledgements

I am indebted to my supervisor, Dr Peter McCulloch, and Prof. Pip Hamilton. Their guidance and practical help was essential throughout my project, most especially in its early stages when the work was unfamiliar and 'independent research' was a difficult art. My work on Successive Projections was inspired by Dr Geoff Poulton, of the CSIRO Division of Radiophysics, who thereafter guided and criticized my work despite his own very demanding commitments. He will remain a model to me of a skillful, professional scientist.

My years with the Department of Physics would have seemed much longer but for the support and companionship of staff and students; unfortunately they are too numerous to name here. It was always a pleasure to work in such a good environment and to take part in several of the Department's activities. I was especially glad of the company of the astronomy students: Dr Andrew Klekociuk, Dr Marco Costa, Dr Stefan Dieters, Tas van Ommen, and Edward King.

Things must be made to work, and the technical staff gave their time generously and impressed me by their expertise, among them Barry Wilson, Gordon Wagner, Don Harding, John Smith, and Phil Button. Thankyou particularly to the Observatory Manager, Gordon Gowland, for his cheerful assistance.

When my own mathematical skills proved too inadequate I was happy to return to my lecturers in the Department of Mathematics. Many thanks to Prof. David Elliott, Dr David Paget, and Dr Glen McPherson.

I am very grateful to two external institutions. The CSIRO Division of Radiophysics supported my work materially during my several visits there, and this project was suggested after my time as a vacation scholar in 1986. It has been a pleasure to see and receive expert advice from staff whom I met then and afterwards: Dr Trevor Bird, Dr Graeme James, Dr Kelvin Wellington, Marc Calabretta, Robert Shields and Michael Kesteven in particular.

AUSSAT, Pty. Ltd. provided free of charge the carrier signal which made my experiments possible, and the Ground Station staff made considerable efforts on my behalf. In addition, AUSSAT funded my travel to conferences in Czechoslovakia and the USSR in 1990. My thanks to the Managing Director, W. G. Gosewinckel, to the Network Supervisor, Brian Bailey, to Chas Gregory, and to Anthony Belo for reliable satellite motion data often at short notice. This Australian company's support of scientific research was greatly appreciated and serves as an example to others.

I have been supported during my studies by the Commonwealth Government, through an Australian Postgraduate Research Award.

I have been lucky to have a wide circle of friends to relieve the hard work with lighter moments, making the years very memorable. My involvement with the Tasmania University Musical Society has been especially rewarding.

Most of all I thank my mother, father, and sister, who provided a happy home and endless encouragement, enabling me to complete my project. This thesis is dedicated to them.

Contents

Introduction

In 1937 Grote Reber built the first parabolic mirror designed for radio frequencies in Illinois, USA. He now lives 60 km from the Mount Pleasant Radio Astronomy Observatory in Tasmania. This thesis describes experiments in Microwave Holography at Mount Pleasant, and the development of data processing methods to overcome a major limitation of traditional holographic measurements.

Parabolic reflector antennas are almost universally used when a narrow directed beam is required for radio astronomy or microwave communication. The rising importance of millimetre-wave astronomy and space communication has placed increasing demands on the performance of these antennas, which is often limited by their reflector surface accuracy. Microwave Holography is a fast and effective technique for measuring the surface profile of reflector antennas. It requires measurement of the amplitude and phase of the antenna's far field radiation pattern. The Fourier transform relationship between the aperture current distribution and the radiation pattern can be used to estimate the complex aperture currents. Geometrical optics relates the phase of the aperture currents to deviations of the surface from a true paraboloid. Adjustment of the surface on the basis of holographic measurements can give significant improvements in antenna performance.

A limitation of Microwave Holography is that in practice only a small part of the complete radiation pattern can be measured. This is because measurement time must be minimized to avoid disrupting normal use of the antenna; there are also practical difficulties involved in measuring the very low-level sidelobes that occur far from the main beam of a large antenna. By measuring only in a region about the main beam, far field information corresponding to high spatial frequencies in the aperture is lost. Any detailed structure in the surface error map is smoothed out. Most large reflectors are constructed of panels, and misaligned panels give rise to discontinuities in surface error with high spatial frequencies. Low-resolution holographic

measurements can lead to incorrect assessment of panel positions in such cases.

Significant improvement can be obtained by recognizing at the outset that the reflector surface is not smooth and continuous, but a collection of individual rigid panels. The Method of Successive Projections is an algorithm, with a simple geometrical interpretation, which allows information about the panel boundaries to be readily incorporated into the data reduction process. The algorithm is straight-forward to apply and is very flexible: aperture blockage effects do not disturb its operation, and panel distortions can be included in the analysis with minimal extra effort. Good results may be achieved while minimizing measurement time. This is an important consideration for high-performance antennas with a busy operating programme.

The place of this thesis in relation to other work on reflector measurements is shown in chapter 1, which is a review of surface errors, their measurement, and their correction. Chapter 2 provides background material for the rest of the thesis, with discussions of antenna motion and panel deformation that contain significant new work. The remaining chapters contain primarily original work. Chapter 3 describes the holographic measurements at Mount Pleasant, which made use of the AUSSAT geostationary satellites. They were unusual in not requiring a saturated carrier for accurate results. Chapter 4 summarizes the systematic errors that can occur and describes a new way of analysing the random errors. Chapter 5 shows how Successive Projections can be used to fit panel positions to low-resolution holographic measurements, and chapter 6 investigates the algorithm's performance and other aspects of the panel-fitting problem.

The main findings of this thesis are given in the Conclusion. Any enquiries about equipment constructed for the holographic measurements or software written to process the data and apply the new algorithm should be directed to:

The Physics Department,
University of Tasmania,
G.P.O. Box 252C,
Hobart TAS 7001,
AUSTRALIA.

Chapter 1

Reflector Antenna Surfaces

1.1 Surface Tolerance Theory

Reflector antennas have a paraboloidal shape because this geometry reflects rays arriving parallel to the axis to a single focal point. Moreover, the ray length from a planar wavefront to the focus is a constant for all reflection points, as illustrated in figure 1.1. This means the radiation adds coherently at the focus. Equivalently, if radiation is emitted by a feed at the focus it will reflect to give a uniform phase in any aperture plane. The far field diffraction pattern from such an aperture has a narrow *main beam* and beyond this many *sidelobes*; an example can be seen in section 3.3.4. A good antenna concentrates as much energy as possible into the main beam, giving a high *gain*, and reduces sidelobe levels to below predetermined limits.

A small deformity of the reflector surface will change the ray lengths and so change the phase of the aperture field above the deformity. This is discussed in detail in section 2.1.2. Uniformity of phase is lost, and the effects of this on the antenna's gain and sidelobe levels are significant. In the manufacture of antennas, tolerances must be found for the surface accuracy if performance specifications are to be met.

1.1.1 Statistical Tolerance Theories

The first practically useful analysis of surface errors was a statistical one by Ruze [74]. Knowing the integral expression for the power pattern due to a given aperture field, he considered an aperture with phase errors whose statistics were known and derived the average power pattern for an ensemble of such apertures. The main features of his model for phase errors were:

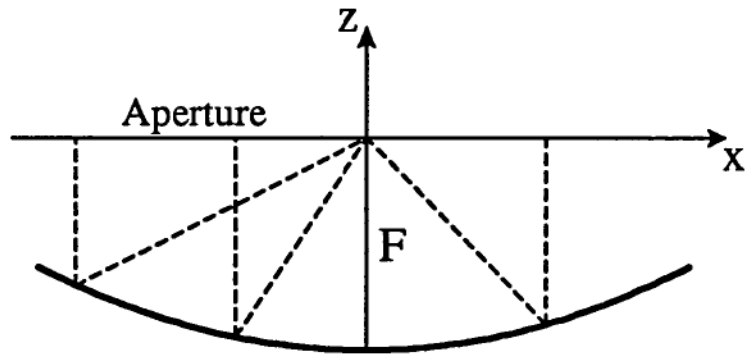


Figure 1.1: A cut in the x-z plane of a paraboloid of revolution with focal length F. The cut is described by $z = x^2/4F - F$. All the dashed ray paths to the aperture plane ($z = 0$) are the same length.

1. The aperture is subdivided into a large number of cylindrical 'hatboxes' or Gaussian 'hats' of phase distortion.

2. The hats' heights are small, and follow a Gaussian distribution with a uniform variance over the whole aperture.

3. The correlation distance of the errors, or the width of the hats or hatboxes, is very small compared to the antenna's diameter.

In the general case it is best to estimate the phase error variance with an illumination weighting, although for uniform variance it makes no difference. If the aperture field is given by the complex function $a(x, y)$ whose phase is $\phi(x, y)$, with the average phase being zero, the variance is estimated by

$$\overline{\delta^2} = \frac{\iint_{\text{ap}} |a(x,y)| \phi^2(x,y)\, dx\, dy}{\iint_{\text{ap}} |a(x,y)|\, dx\, dy}. \tag{1.1}$$

Ruze showed that the axial gain is reduced according to

$$G = G_0\, e^{-\overline{\delta^2}} = \eta \left(\frac{\pi D}{\lambda}\right)^2 e^{-\overline{\delta^2}} \tag{1.2}$$

where η is the aperture efficiency, D is the antenna diameter, and λ is the wavelength. This equation has been used by antenna engineers ever since. As a benchmark, rms surface errors of $\varepsilon = \lambda/15$ give a 3 dB reduction in gain; a surface should be designed to exceed this accuracy at its maximum working frequency. Every antenna has a maximum gain at some wavelength λ_{max}, beyond which surface errors cause the gain to decrease through their greater

effect on the aperture phase. For a shallow reflector the rms surface error ϵ produces a phase error variance

$$\overline{\delta^2} = \left(\frac{4\pi\epsilon}{\lambda}\right)^2$$

and the maximum gain occurs when

$$\lambda = \lambda_{\max} = 4\pi\epsilon. \tag{1.3}$$

This can be checked with gains measured by radio observations, and agreement with measured rms surface errors is usually good.

Ruze also identified a broad, scattered field which is added to the radiation pattern. As the correlation distance of the errors decreases, the scattered field is broadened, so that rough surfaces scatter the energy more diffusely. This is to be expected. However, Ruze's restriction to small correlation distances causes average sidelobe levels to be overestimated in many realistic situations. Vu [87] augments the original theory by analytically calculating the aperture correlation function for a tapered illumination. This allows a study of errors with large correlation distances. For example, Vu found that a correlation distance of $D/24$ gives average sidelobe levels 3 dB below those predicted by Ruze. Gain is also less affected when surface errors are very broad in scale.

The uniform variance restriction of Ruze's theory was partly lifted by Rahmat-Samii [62], who used an aperture divided into annular regions with different phase error variances. For some construction methods this would be more realistic than a uniform variance. Rahmat-Samii gave a closed expression for the average power pattern, while commenting that an average sidelobe level is not as useful a quantity as the probability of a sidelobe occurring within a specified range. Such information would help to decide manufacturing tolerances. The problem was tackled by Ling, et al. [41].

An entirely different statistical approach is due to Tripp [83]. Instead of a multiplicative phase error, he considered an additive perturbation of the aperture field which could occur at several random positions. Assuming that the scattering properties of the defect were known, he found a vector expression for the mean power pattern of an aperture randomly distributed with defects. This is a suitable model for an antenna subjected to a harsh environment for some time, where it will accumulate spot defects.

1.1.2 Origins of Surface Deformations

Whether or not a statistical approach to surface errors is best depends on what is known of the errors. For example, thermal distortion of a reflector support structure produces slowly changing errors with a large correlation distance. The effects of errors of this kind may be estimated statistically to help design a suitable structure. Once the reflector is built it may be possible to predict or measure thermal distortion, and then either to compensate for the gain variations or to correct the surface with actuators — 'active surfaces' are discussed in section 1.4.

A classification of errors by their correlation distance across the aperture is useful.

- Short correlation distance:
  - errors inherent in the surface construction such as ribs beneath a mesh surface.
  - panel setting errors for a reflector made of rigid panels.
  - profile errors in the panels themselves.
  - rapidly changing variations induced by the antenna drive system.
  - any spot deformities.
- Long correlation distance (flexure of the structure):
  - thermal distortion with ambient temperature or sunlight.
  - gravitational distortion as the antenna is tilted.
  - response to wind stresses.
- Errors which imply a different best-fit paraboloid. A linear aperture phase variation is removed simply by redefining the axis of the paraboloid. A quadratic variation, to first order, requires a refocusing by moving the feed axially. These are not errors of the surface at all.

At any time the total distortion of a reflector will contain many of these effects.

Thermal, gravitational, and wind distortion are likely to have very long correlation distances. Thus the near-in sidelobes are most affected, and a statistical error analysis is unsuitable for predicting the astigmatism, coma, and other features that are likely to occur. Rahmat-Samii [64] suggests methods

of interpolating known distortions and efficiently calculating the resulting far field pattern. Alternatively, Pontoppidan [58] shows how surface distortions with particular symmetries can be related directly to specific changes in the first few sidelobes of the pattern. He uses Zernike polynomials to represent a general distortion, where low-order modes correspond to long correlation distances. Such an analysis is useful because the reflector's backing structure determines which symmetries of distortion will occur, and action can be taken in the design stage to minimize the most serious.

The best example of a design to cater for surface distortions is the Effelsberg 100 m radio telescope, which deforms homologously under gravity as it is tilted. This means that while distortions of up to 6 cm occur at the edge of the reflector, they are such that a new paraboloid is made at every elevation. The inner 65 m of the reflector can be used at wavelengths as short as 1 cm. Careful design like this means that thermal and wind deformations, not gravity, limit the performance of well-designed modern antennas [69, sect. 5.6].

An existing antenna, then, has fixed and changing errors in its surface. Knowledge of these errors could be used to improve the manufacturing process if the antenna is one of a series, or if not it could be used to improve the antenna itself. It is very desirable to measure the surface errors of an antenna.

1.2 The Development of Microwave Holography

1.2.1 Before and After Scott and Ryle

It has long been known that there is a Fourier transform relationship between the aperture currents and the far field pattern of an antenna. The idea of finding imperfections in a radiating current distribution by measuring its complex far field pattern was first put into practice by Napier and Bates in 1973 [52]. They used acoustic antennas operating at a frequency of 34.5 kHz, and measured phase indirectly by recording a *hologram*: the sum of the test antenna's radiation with the radiation from a fixed reference antenna. In 1976, Bennett, et al. [6] used similar principles to perform microwave measurements of two parabolic reflector antennas, 3.0 m and 3.7 m in diameter. The test and reference antennas received radiation from a near-field source,

and holograms were formed by combining these signals. The derived aperture phase distribution was converted to a surface error profile.

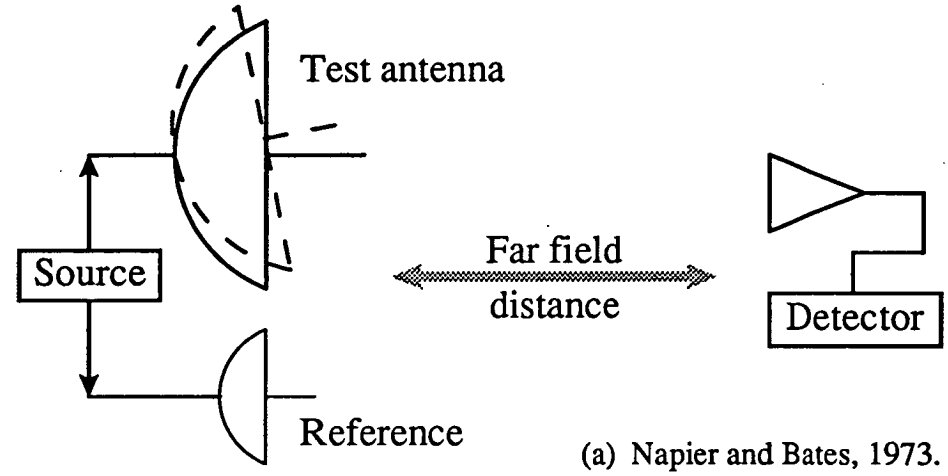
The following year, Scott and Ryle [76] showed how the complex receiving pattern of a radio telescope could be measured directly using a phase-switching receiver. Four 13 m antennas were measured using radiation at 15.4 GHz from a celestial source. Each test antenna was steered in a 17×17 grid of points about the on-source position while a reference antenna remained on-source, and the product of the resulting signals was formed by the receiver. The complex aperture distribution was found by a Fourier transform of these data, giving a grid of surface profile errors accurate to ± 0.1 mm. The measurement scheme is compared to the holographic technique of Napier and Bates in figure 1.2. The fundamental difference is that the radiation pattern's phase is determined directly by detecting a product instead of indirectly by detecting a sum. Although no hologram is formed, the Scott and Ryle method of reflector profile measurement has been called *Microwave Holography*.

Details of the implementation of Microwave Holography are left to the next two chapters. It has become a well-established method for surface metrology, and table 1.1 summarizes a variety of experiments performed over the past decade. A range of antenna sizes and frequencies can be seen. The time required for the measurements varied, but one night's observing was always sufficient and a few hours was typical: this is much faster than most of the alternative surface measurement methods described below.

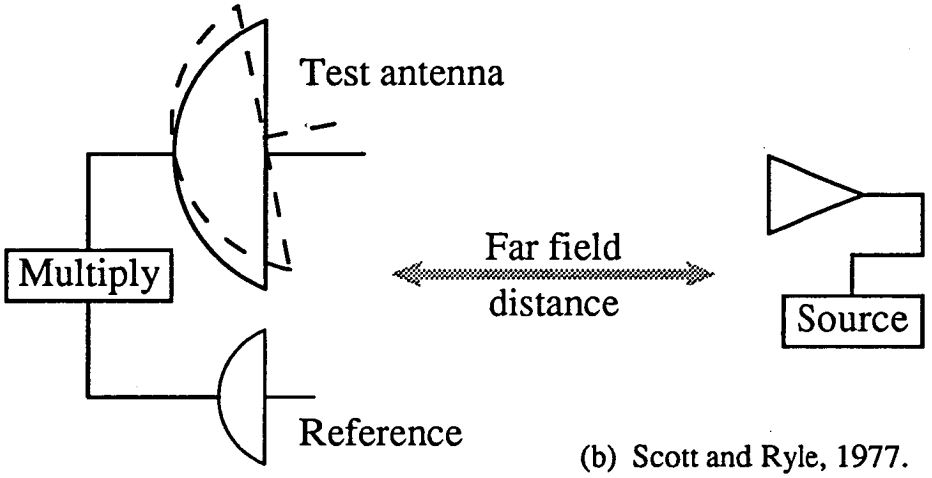
The accuracies σ in the table are the rms uncertainties in surface error for a single 'pixel' or resolution element. This was not given explicitly in some of the papers. In some it was deduced from the repeatability of independent surface maps, in which case it may reflect genuine changes in the surface shape rather than the intrinsic accuracy of the measurements. Despite this, the achieved accuracies are always in the range $0.001\lambda < \sigma < 0.01\lambda$. Thus the accuracy can be adapted to requirements by changing the frequency, provided that

- it is still possible to measure the far field phase, and
- the range of surface errors is not enough to cause 2π ambiguities in aperture phase.

When accuracy is limited by additive receiver noise giving an on-source signal-to-noise ratio $\mathcal{R}$, Scott and Ryle demonstrated that a map of dimen-



(a) Napier and Bates, 1973.



(b) Scott and Ryle, 1977.

Figure 1.2: Two methods of recording the complex far field pattern of an antenna: (a) formation of a hologram by summing radiation from the test antenna with a reference wave, and (b) measuring phase and magnitude by multiplying signals received by test and reference antennas for an incoming plane wave.

Ref.	Year	Diam.	Source	Grid	σ	λ/σ
[76]	1977	13 m	C 15.4 GHz	17×17	0.1 mm	195
[21]	1981	25 m	S 11.5 GHz	101×101	0.07 mm	369
[15]	1981	30 m	S 3.84 GHz	32×32	0.5 mm ?	156
[67]	1983	64 m	C 2.29 GHz	11×11	0.5 mm ?	262
[46]	1983	4.9 m	R 86.1 GHz	83×83	4 μm	870
[20]	1986	100 m	S 11.8 GHz	195×195	0.12 mm	212
[33]	1987	18 m	S 4.00 GHz	64×64	0.18 mm	416
[31]	1988	45 m	S 19.5 GHz	128×128	53 μm	290
[51]	1988	30 m	C 22.2 GHz	32×32	18 μm	764
[36]	1989	64 m	S 12.4 GHz	97×97	0.2 mm	121
$\star$	1991	26 m	S 12.7 GHz	53×53	53 μm	445

Table 1.1: A comparison of several Microwave Holography experiments from Scott and Ryle onwards ($\star$ is this project). The year is that of publication. The type of source used is indicated by a letter: C for a celestial source, S for a satellite-borne transmitter, and R for a ground range approximately $2D^2/\lambda$ in length.

sions $N \times N$ will have

$$\sigma \propto \frac{N}{\mathcal{R}}.$$

It is easier to achieve a high signal-to-noise ratio with a satellite source, and for this reason the maps of table 1.1 were more restricted in size when the source was celestial.

1.2.2 Phase Retrieval Holography

Millimetre-wave reflector antennas are playing an increasing role in radio astronomy, and they require very high surface accuracies. To be useful, Microwave Holography needs to give a measurement uncertainty smaller than the expected magnitude of surface profile errors. Frequencies of several tens of GHz or higher are required, and direct phase measurements are difficult.

The practicality of deducing the complex aperture current distribution from only the magnitude of the far field pattern was investigated by Morris [49]. He suggested using the Missel algorithm to recover the aperture phase from measurements of the in-focus and out-of-focus far field intensity. Subsequent tests at the IRAM 30 m diameter radio telescope at Pico Veleta [50] showed that the technique was effective, with the need for complicated receiving equipment obviated by sophisticated data processing. A disadvantage is that a much greater signal-to-noise ratio is required than for traditional holography. More recently, phase recovery from a single, out-of-focus intensity map has been investigated and validated [43]. Phase retrieval is an active area of research.

The measurements described in chapter 3 of this thesis were traditional, full-phase measurements. They were aimed at developing a simple but effective system which could be moved easily to other antennas. The algorithm for panel fitting in chapter 5 is very new but is related to the Missel algorithm, and others commonly used for phase retrieval. This connection is discussed as background to the Method of Successive Projections in section 5.1.3, and the solvability of the phase retrieval problem is discussed in section 6.2.3.

1.2.3 Related Techniques

There are some variations of the Microwave Holography method that can be used in special circumstances. When the three-dimensional structure of a reflector is important, Cook, et al. [14] show how particular planar cuts through the current distribution can be recovered. This is by consideration

of the three-dimensional Fourier transform of the current distribution, and the theory is described in section 2.3.2 where it becomes relevant for understanding the effects of antenna motion.

The focal-plane field distribution can also be related to the aperture currents by a Fourier transform [85]. This has been used to measure reflector deformities of the RATAN-600 radio telescope in the USSR [57]. The special equipment needed to probe focal-plane fields usually makes this technique more difficult than far field measurements. Another alternative is to deduce the aperture phase directly from near-field pattern measurements. Parini, et al. [54] determined the profile of a 1.2 m reflector by measuring only the near field phase. This was done by using 35 GHz electomagnetic waves and again using 40 kHz ultrasound. The size of the chamber that would be required makes near-field methods unsuitable for large reflectors.

A recent innovation is by Serabyn, et al. [77]. In the regime of submillimetre waves methods usually applied to optical systems can be copied. By using mirrors and a Mylar beam splitter, simultaneous on-axis and off-axis beams were produced for a 10.4 m diameter telescope operating in a band about 290 GHz. The mirrors were moveable, so the single detector received signals from the reference, on-source beam and the steerable, off-axis beam. This formed an interferometer which, in effect, examined the telescope's focal plane field. Serabyn, et al. measured the surface profile to an accuracy of ± 9 μm using this equipment and planetary radiation. The second beam replaced the second antenna usually required for interferometric measurements, and only one receiver was required, so the technique has a number of advantages.

1.3 Other Methods of Surface Measurement

Microwave Holography has been a viable technique for only one quarter of the time that large, parabolic reflectors have been built. There are several older methods for surface measurement which are still important, particularly during the construction of a reflector or immediately after its commissioning. They can be superior in some respects, such as the measurement of shadowed regions of the surface. Typically a large reflector will be adjusted to whatever tolerance traditional methods permit, and a holographic survey will then be performed, ideally at the shortest wavelength for which the remaining surface errors are determined unambiguously.

As mentioned in section 1.1.1, gain measurements can be used to give the rms surface error, but an actual map of the surface is needed if corrections are to be made. Template methods were the first to be used, but they are restricted to small antennas. By far the most common method for large reflectors has been triangulation or angle-range measurements to targets fixed on the surface. There are many variations on this theme. Immediately, there is an important difference with holographic surveys: errors indicated in a holographic map are averages over a region of the surface, whereas a target always indicates the position of a single point. If this point is over a panel adjustment screw this may be what is wanted; on the other hand, an average over the region between measured points is often a better measure of the effective surface error.

A 'theodolite-tape' measurement of the Effelsberg 100 m antenna is described by Greve [23]. Targets were positioned with a steel tape and observed by a theodolite at the vertex. The final survey gave the position of 6000 targets with a vertical accuracy of $\sigma \simeq 0.5$ mm, although σ varied with radius. The survey quality can be indicated by the ratio D/σ, so about 200 000 was achieved. This can be compared to the later holographic map of the same antenna [20] which gave $D/\sigma \simeq 830\,000$. Another theodolite-tape survey was performed by Kesteven et al. [35], where the accuracy achieved was $\sigma \simeq 0.11$ mm for a 22 m antenna of the Australia Telescope. This again gives $D/\sigma \simeq 200\,000$, which is a typical figure for a high-quality survey of this type. Both surveys had to be performed in the zenith position, for practical reasons and for the observers' comfort!

The method was developed further by Greve [24] in a measurement of the IRAM 30 m millimetre-wave radio telescope at Pico Veleta. The tape was supported by an arm and read using a transducer, after which a temperature correction was made. The theodolite was kept at a fixed angle for each ring of measurements, and a laser beam was focused through it onto target with a spherical surface to contact the reflector. Inside the target a position diode measured the offset of the laser beam, so human error could not enter the reading of data. The resulting accuracy was $\sigma \simeq 0.07$ mm or $D/\sigma \simeq 430\,000$. This was equalled by a holographic measurement with phase retrieval [50], and surpassed by a measurement with full phase [51] which gave $D/\sigma \simeq 1\,700\,000$. Remember that a direct comparison like this does not take into account that the holographic map was on a square grid with 486 points lying in the aperture, while the Greve survey had 600 points arranged in 11 rings

Antenna	Diameter	Traditional D/σ	Holographic D/σ
SRC Chilbolton	25 m	70 000	350 000
IRAM Pico Veleta	30 m	430 000	1 700 000
MPIfRA Effelsberg	100 m	200 000	830 000

Table 1.2: The 'quality' D/σ of holographic and traditional surface measurements of three antennas for which both are available. References are in the text.

on the aperture.

It is worthwhile to mention an earlier experiment by Slater [80] at the Chilbolton 25 m antenna for two reasons. Firstly, a specially constructed instrument ('Parabscan') measured the angles to targets from two points on the reflector axis, so that no range measurements were required. The survey of 512 targets had an accuracy $\sigma \simeq 0.36$ mm or $D/\sigma \simeq 70\,000$. Secondly, a laser ranging system was used afterwards to measure changes in the reflector's shape with elevation. This had a similar order of accuracy for a smaller number of points, and was much faster because differential measurements are always simpler than absolute ones. It is interesting that now, 20 years later, similar systems are being developed for application to active surfaces (below). Once again, a holographic measurement of the same antenna is available [21]: it was earlier than the others mentioned, and gave $D/\sigma \simeq 350\,000$. Table 1.2 summarizes the qualities of the holographic and traditional surveys of the three antennas measured in both ways. In these examples the holographic surveys are 4–5 times more accurate.

Accounts of theodolite-tape and similar surface measurements highlight several systematic effects which limit the accuracy achieved. Tape readings are affected by temperature changes, curves in the path, and positioning of the zero mark. Theodolite readings are affected by axis alignment, and several reference points must be measured repeatedly during the survey so that different rings or quadrants of data can be correctly combined. Often the greatest difficulty is the period of time during which consistent measurements must be made. The Effelsberg theodolite survey took 10 days, the Pico Veleta 6–10 nights, and the Chilbolton 26 nights for a team of 10 observers. Operator fatigue becomes a significant factor in these circumstances, as well as genuine changes in the surface shape due to environmental con-

ditions. Only the Australia Telescope measurements were performed in a period comparable to a typical holographic survey: about 4 hours for a 22 m antenna. This was possible because data logging was automated and the surface was already quite well set.

Completely different approaches to surface measurement are possible. An unconventional method is due to Payne, et al. [56]. A trolley with a 50 cm wheelbase was pulled along radial lines from the vertex: it had a depth sensor in the centre which allowed the curvature of the surface to be measured as a function of distance along the surface. By integration the slope and height functions could be found. The NRAO 11 m antenna at Kitt Peak was measured to an accuracy of $\sigma \simeq 0.09$ mm at the edge, or $D/\sigma \simeq 120\,000$. For larger antennas the accumulation of errors would be greater, and instead of a connected surface the trolley would have to traverse panel edges.

Photogrammetry is a method which can be applied to antennas of any size. Photographs are taken from several well-separated viewing positions after the placement of targets on the surface and other parts of the structure. By measuring from the photographs a true three-dimensional map is produced, with a high redundancy of information giving great precision. The antenna can be at any elevation without penalty, which is certainly not true of theodolite-type surveys. Fraser [17] describes photogrammetrical measurements of the DSS-15 34 m Cassegrain antenna near Goldstone. Both the main reflector and subreflector were measured using 9 and 6 camera stations respectively, and elevations of 6°, 45°, and 90° were used. For the main reflector an accuracy of $\sigma \simeq 0.17$ mm was achieved giving $D/\sigma \simeq 200\,000$, which is comparable to a good theodolite survey. The main disadvantage was the time necessary to digitize and measure the photographs to ± 0.4 μm: 18 hours, while the photographing time was only 6 hours.

This must be weighed against the value of things which only photogrammetry can do, particularly for Cassegrain antennas. Firstly, the subreflector can be independently measured to great precision. Secondly, by making sure that targets fixed beyond the edge of the subreflector are included in all photographs, the two reflectors can be described in a common reference frame. Other structural features, such as the ground and mount, can also be included. These data could be used to great benefit, so it is important to decide exactly what type of information is needed when choosing a survey method. If radio performance is the primary concern, Microwave Holography remains the most rapid and most precise of currently available methods.

1.4 Active Surfaces

The possibility of automatically adjusting the surface of a reflector is currently receiving serious attention. One application is being investigated by Clarricoats [11], who tested a 0.85 m mesh surface controlled by 52 actuators. This would be used on satellites for shaping the antenna beam to a desired coverage on Earth. For large reflectors an 'active surface' would have a different purpose: to compensate for deformations as they occur, allowing shorter wavelengths to be used.

This introduces a new set of surface measurement problems. Not all kinds of deformation can be predicted, so they must be measured accurately and quickly enough for effective compensation. The time scale may be faster than minutes, for example thermal reactions to the sun going behind a cloud. Fortunately, the measurements need only be differential provided that they are calibrated to a known surface shape determined by holography or other means. Although automatic adjustments were not considered, the laser ranging system mentioned above [80] followed this scheme. A microwave ranging system was tested by Findlay and Payne [16] at a frequency of 11.8 GHz, but, monitoring only four distances, it was intended for diagnosis rather than correction. These methods were accurate to approximately 0.4 mm and 0.15 mm.

Two large radio telescopes designed to have active surfaces are currently under construction: a 70 m antenna at Sufa, USSR and a 100 m offset antenna at Green Bank, USA. For the former, Shulga and Shamanin [78] described a microwave ranging system for operation at about 40 GHz; a single transponder has been used to determine path lengths to within ± 30 μm. A measurement time of some seconds and a surface setting time of some minutes are proposed. The Sufa antenna will be a Cassegrain design and the subreflector may be included in the microwave paths. However, problems with multiple reflections and atmospheric scintillations are anticipated. An important constraint is that the ranging system must not interfere with the radio astronomy receivers.

More information is available on the Green Bank antenna [55], which will be the largest fully steerable antenna in the World and will also be unique for its offset design. Completion is planned for 1995. It will have 1928 panels adjustable by 2255 actuators: because panel corners can be set accurately relative to each other, only one actuator is provided at each meeting of four

corners. The goal is compensation for all gravitational and thermal deformations and some vibrational modes. This should allow operation at frequencies above 40 GHz.

A laser ranging system is to be used; a prototype is accurate to ± 50 μm. Several lasers modulated at 1 GHz or similar will be mounted on the feed support arm, providing about 100 range measurements every second. These will be used to update a complex model of the structure once every second or faster, and surface corrections will be based on this model. Targets on the ground, the subreflector, and the feeds will be included in the measurements, allowing absolute pointing calculations to be made. The accuracy of surface corrections and the lack of aperture blockage may show other effects to be important to antenna performance, such as the size of gaps between panels. Microwave Holography will have a role: to provide an external reference standard for the laser measurements.

This section concludes with a measurement system for an active optical surface, such as could be implemented in the near term to compensate for atmospheric distortion in astronomical telescopes. It is interesting because it combines several principles whose application to radio reflectors has been discussed above. Sandler, et al. [75] show how in-focus and out-of-focus images of a reference star can be used to find the phase distortion across the telescope's aperture. An optical image is in fact the far field intensity pattern corresponding to the aperture phase function, so the recovery of phase from this pair of images is exactly analogous to the phase retrieval method of Morris [49]. The images are recorded with a single CCD camera, and a path length difference achieved with mirrors causes the defocus. Phase retrieval is performed by a parallel computational architecture called a 'neural network', modelled on the action of neurons in the human brain. The result is a set of low-order Zernike coefficients (c.f. [58]) describing the aperture phase, upon which surface corrections would be based. It is becoming more common for work in the optical and radio fields to be closely related; another example by Serabyn, et al. [77] was mentioned earlier.

Chapter 2

Microwave Holography Fundamentals

2.1 The Theoretical Basis

The "Fourier transform relationship between the aperture currents and the far field pattern" hides a number of subtleties and approximations in the case of a parabolic reflector antenna. It is well worth while to look at the origin of the expression and its application to Microwave Holography.

2.1.1 The Radiation Integral

Source currents which exist on the surface of a reflector antenna radiate into an infinite, homogeneous medium according to Maxwell's equations. Throughout this thesis the time dependence of the resulting harmonic fields is $\exp(-j\omega t)$. The solutions for $\mathbf{E}$ and $\mathbf{H}$, the electric and magnetic field intensities, can be written in terms of vector potentials $\mathbf{A}$ and $\mathbf{F}$:

$$\mathbf{E} = -\nabla \times \mathbf{F} + j\omega\mu\mathbf{A} - \frac{1}{j\omega\hat{\epsilon}}\nabla(\nabla \cdot \mathbf{A}) \tag{2.1}$$

$$\mathbf{H} = +\nabla \times \mathbf{A} + j\omega\hat{\epsilon}\mathbf{F} - \frac{1}{j\omega\mu}\nabla(\nabla \cdot \mathbf{F}). \tag{2.2}$$

Here μ and $\hat{\epsilon}$ are constitutive parameters of the medium. In terms of these vector potentials, Maxwell's equations become inhomogeneous Helmholtz equations:

$$\nabla^2\mathbf{A} + k^2\mathbf{A} = -\mathbf{J} \tag{2.3}$$

$$\nabla^2\mathbf{F} + k^2\mathbf{F} = -\mathbf{M}, \tag{2.4}$$

where $\mathbf{J}$ and $\mathbf{M}$ are electric and magnetic source currents and $k^2 = \omega^2\mu\hat{\epsilon}$.

These may be solved using a Green's function formulation. The coordinate system used is shown in figure 2.1. In terms of the Green's function

$$G(\mathbf{r},\mathbf{r}') = \frac{\exp(+jk|\mathbf{r}-\mathbf{r}'|)}{4\pi|\mathbf{r}-\mathbf{r}'|}$$

the solutions are

$$\mathbf{A}(\mathbf{r}) = \int_{V'} \mathbf{J}(\mathbf{r}')\, G(\mathbf{r},\mathbf{r}')\, dV' \tag{2.5}$$

$$\mathbf{F}(\mathbf{r}) = \int_{V'} \mathbf{M}(\mathbf{r}')\, G(\mathbf{r},\mathbf{r}')\, dV'. \tag{2.6}$$

Substituting these into the equations 2.1 and 2.2, and evaluating the derivatives of $G(\mathbf{r},\mathbf{r}')$, it follows that

$$\mathbf{E}(\mathbf{r}) = +jk\int_{V'} \left[\mathbf{M}(\mathbf{r}')\times\hat{\mathbf{R}} + \sqrt{\frac{\mu}{\hat{\epsilon}}}\left(\mathbf{J}(\mathbf{r}') - (\mathbf{J}(\mathbf{r}')\cdot\hat{\mathbf{R}})\hat{\mathbf{R}}\right)\right] G(\mathbf{r},\mathbf{r}')\, dV'$$

$$\mathbf{H}(\mathbf{r}) = -jk\int_{V'} \left[\mathbf{J}(\mathbf{r}')\times\hat{\mathbf{R}} - \sqrt{\frac{\mu}{\hat{\epsilon}}}\left(\mathbf{M}(\mathbf{r}') - (\mathbf{M}(\mathbf{r}')\cdot\hat{\mathbf{R}})\hat{\mathbf{R}}\right)\right] G(\mathbf{r},\mathbf{r}')\, dV'.$$

Terms of order $1/R^2$ and smaller have been neglected. These expressions, giving radiated fields in terms of the source current distributions, are often referred to as the *Kirchhoff-Huygens Diffraction Integrals*.

Some simplifications can be made for the radiation pattern in the far field of a perfectly conducting reflector. Firstly, $\mathbf{M} = 0$ since there can be no magnetic currents. Secondly, the *far zone approximation* is valid, whereby

$$\begin{array}{ll} R \simeq r & \text{for amplitude terms,} \\ R \simeq r - r'\cos\xi & \text{for phase terms.} \end{array}$$

These give

$$\mathbf{E}(\mathbf{r}) \simeq +jk\,\frac{e^{+jkr}}{4\pi r}\int_{V'} \sqrt{\frac{\mu}{\hat{\epsilon}}}\,[\mathbf{J}(\mathbf{r}') - (\mathbf{J}(\mathbf{r}')\cdot\hat{\mathbf{r}})\hat{\mathbf{r}}]\, e^{-jkr'\cos\xi}\, dV' \tag{2.7}$$

$$\mathbf{H}(\mathbf{r}) \simeq \sqrt{\frac{\hat{\epsilon}}{\mu}}\hat{\mathbf{r}}\times\mathbf{E}(\mathbf{r}). \tag{2.8}$$

Thus $\mathbf{E}$ and $\mathbf{H}$ are perpendicular to each other and to the direction of propagation in the far zone.

The use of these diffraction integrals for antenna analysis commenced with the radio scientists of the Second World War, whose work was collected by Silver [79]. They are introduced or derived in any textbook on antenna theory, for example [13, sect. 2.5, 3.6] or [3, sect. 3.2–6]. The summary above is based on the concise explanation of James [32, sect. 2.1].

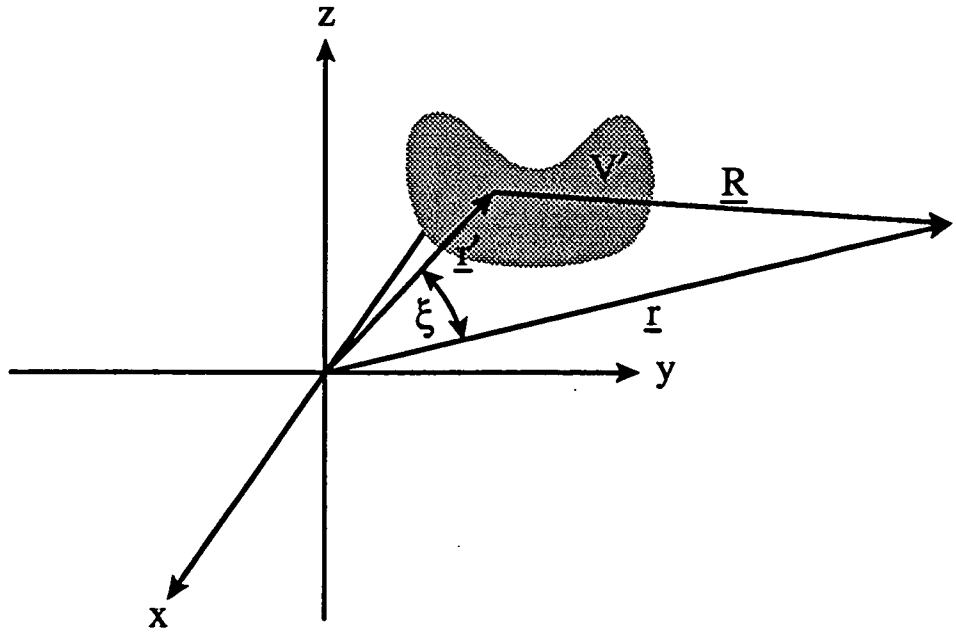


Figure 2.1: The coordinates $\mathbf{r}'$ for the integration of source currents in a volume V'. The field is computed at a point $\mathbf{r}$, or $\mathbf{R} = \mathbf{r} - \mathbf{r}'$ with respect to a source point.

The Aperture Integral Method

One way to express the field integral 2.7 in a tractable form is to transfer it to a planar surface by the principle of images. This is described neatly by Wood [84, p. 15]. Consider a thin conducting screen in the plane $z = 0$ where all sources are in the region $z < 0$, as shown in figure 2.2. Into the forward half-space $z > 0$ the electric current

$$\mathbf{J}_{\mathbf{ap}} = 2\hat{\mathbf{n}} \times \mathbf{H}(x, y, 0)$$

radiates a field $\mathbf{E}$ and $\mathbf{H}$ that is equivalent to the field that would exist with no screen. If the sources are generated by a reflector antenna, the imaginary screen is called the *aperture plane* of the antenna, and the currents are the *aperture currents*.

Geometrical Optics can be used to find the field radiated by the feed and reflected to the aperture plane. This involves the *local plane wave reflection* (LPWR) approximation that the incident radiation reflects as if it had met an infinite plane conductor; any currents on the rear of the reflector are ignored. Given this field, the aperture currents can be evaluated and equation 2.7

becomes

$$\mathbf{E}(\mathbf{r}) \simeq +jk\frac{e^{+jkr}}{4\pi r}\int_{A'}\sqrt{\frac{\mu}{\hat{\epsilon}}}\left[\mathbf{J}_{\mathrm{ap}}(\mathbf{r}') - (\mathbf{J}_{\mathrm{ap}}(\mathbf{r}')\cdot\hat{\mathbf{r}})\hat{\mathbf{r}}\right]e^{-jkr'\cos\xi}\,dA'. \qquad (2.9)$$

However, the geometry of figure 2.1 gives

$$r'\cos\xi = ux' + vy' + z'\cos\theta \text{ where } \begin{cases} u = \sin\theta\cos\phi \\ v = \sin\theta\sin\phi \end{cases}$$

with $z' = 0$ in the aperture plane. The vector operations can be taken outside the integral, giving

$$\mathbf{E}(\mathbf{r}) = \mathbf{T}(\mathbf{r}) - (\mathbf{T}(\mathbf{r})\cdot\hat{\mathbf{r}})\hat{\mathbf{r}} \qquad (2.10)$$

where

$$\mathbf{T}(\mathbf{r}) \simeq +jk\frac{e^{+jkr}}{4\pi r}\int_{A'}\sqrt{\frac{\mu}{\hat{\epsilon}}}\mathbf{J}_{\mathrm{ap}}(\mathbf{r}')e^{-j2\pi(ux'+vy')/\lambda}\,dA'. \qquad (2.11)$$

This is a Fourier transform. It is of a standard form if x' and y' are measured in wavelength units, eliminating the factor $1/\lambda$. These aperture variables transform to 'far field domain' variables u and v, the direction cosines of the radiation vector $\mathbf{r}$ to the x and y axes. The coordinate systems are shown in figure 2.2.

The Surface Integral Method

The image principle can also be applied to the reflector surface itself, giving a current

$$\mathbf{J}_{\mathrm{PO}} = 2\hat{\mathbf{n}} \times \mathbf{H}_{\mathrm{inc}}$$

where $\hat{\mathbf{n}}$ is the surface normal and $\mathbf{H}_{\mathrm{inc}}$ is the incident field from the feed. This is the *Physical Optics* (PO) approximation, which again assumes that each element of the surface behaves as an infinite plane conductor. The integral 2.7 can be applied directly to these currents. The LPWR and PO assumptions lead in general to slightly different radiated fields [84, p. 18]. However, if the aperture plane 'caps' the reflector, meeting its entire edge, the fields calculated by both methods are practically the same [88]. For offset geometries this cannot be the case, and PO followed by a surface integration should be used.

For a reflector surface integration the term $z'\cos\theta$ remains, so the integral 2.9 does not reduce to a Fourier transform. Rahmat-Samii and Galindo-Israel [66] show that evaluating this integral in aperture plane coordinates produces an infinite series of Fourier transforms. The first term is 2.11 with

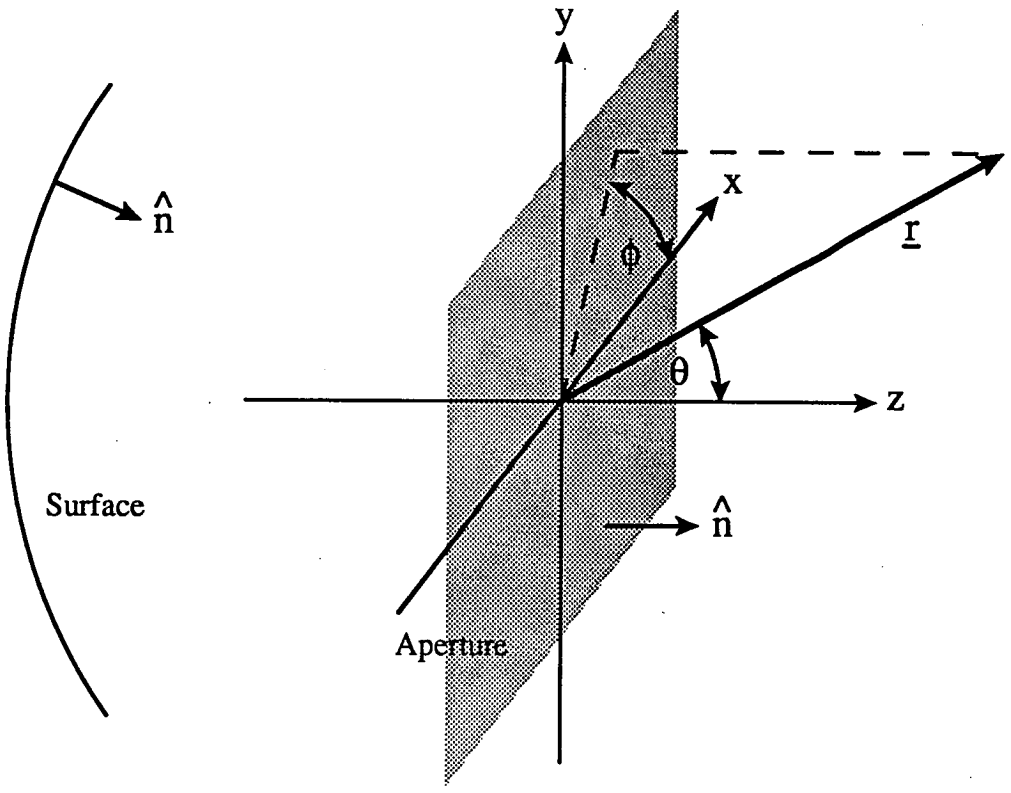


Figure 2.2: The aperture and far field coordinate systems, with origin placed at the focus. Primes are used when specifying a point in the aperture or on the surface.

$\mathbf{J}_{ap}$ replaced by a modified $\mathbf{J}_{PO}$, the modification being simply a projection of the surface currents onto the flat aperture with a phase shift matching the separation z'. Although the modified $\mathbf{J}_{PO}$ is treated like an aperture current it represents a different approximation, and the method remains a 'surface integration'.

Further terms of the series can be used if the first does not give an adequate approximation. The papers [48] and [66] show how the series expression is useful for offset geometries. It can also be used in cases where the depth of a reflector causes the second term to be significant. If the correct radiated field can be computed by the Fourier transform of a particular function $\mathbf{J}_1(\mathbf{r})$, a simple iterative process converges rapidly to $\mathbf{J}_2(\mathbf{r})$, which gives the correct field after computing two terms of the series of Fourier transforms. Then $\mathbf{J}_2(\mathbf{r})$ is a better estimate of the true aperture current function.

2.1.2 Application to Microwave Holography

By the principle of Reciprocity, discussed in [3, sect. 3.8] or [13, sect. 4.2], the roles of a transmitting and a receiving antenna can be exchanged and the voltages at their terminals will not be altered. The radiating and transmitting patterns of a test antenna, which can only be measured by using a second antenna, are thus always the same. This is why Microwave Holography determines the radiation pattern, even though it is done in receiving mode to allow a phase reference antenna to observe the same incident plane wave.

The amplitude and phase measurements are relative, so the spherical wave term $\exp(+jkr)/4\pi r$ in equation 2.11 is irrelevant. Assume for the present that the function $\mathbf{T}(\mathbf{r})$ is known from the measured radiation pattern. Then an inverse Fourier transform can be used to obtain either the aperture current $\mathbf{J}_{ap}(\mathbf{r}')$ or the projection of $\mathbf{J}_{PO}(\mathbf{r}')$ to the aperture depending upon the approximation used; these have identical phases. The aperture phase will be uniform for a perfect paraboloidal reflector fed from its focal point. Departures from uniformity can be related to errors in the reflector profile using the geometrical optics construction shown in figure 2.3, which is based on [63]. From the enlarged view,

$$\frac{\varepsilon}{\delta r} = \cos\eta \text{ and } \frac{\delta z}{\delta r} = \cos 2\eta.$$

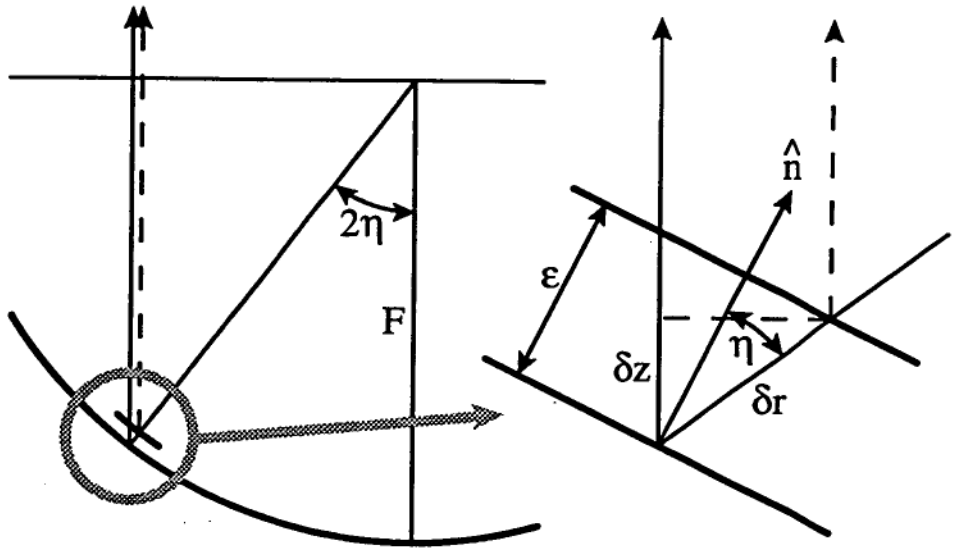


Figure 2.3: The effect of a small normal surface error ε on the ray path from the feed to the aperture plane.

The total path length change for the ray is thus

$$\delta r + \delta z = \frac{\varepsilon}{\cos\eta} + \frac{\varepsilon\cos 2\eta}{\cos\eta} = 2\,\varepsilon\cos\eta.$$

This causes a change in aperture phase ϕ obtained by multiplying by $4\pi/\lambda$. The cosine of η can be found as a function of the distance from the aperture centre and the focal length F, and this equation results:

$$\varepsilon = \phi\,\frac{\lambda}{4\pi}\sqrt{1+\frac{x'^2+y'^2}{4F^2}}. \tag{2.12}$$

This allows *normal surface errors* ε to be estimated from the recovered aperture phase, assuming they are small enough for the ray method to be accurate. The *axial surface errors* are the components of ε parallel to the reflector's axis, obtained by leaving out the square root factor. These serve as a direct measure of phase in units which are independent of frequency, which is often useful. Care must be taken with the sign of errors: because the time convention here is $\exp(-j\omega t)$ the rules are

- A *negative* phase error means the surface is too *high*;
- A *positive* phase error means the surface is too *low*.

Things which are Ignored

Only one output from the antenna is used for holography, hence only one polarization. Thus not all components of the vector $\mathbf{T}(\mathbf{r})$ are measured. Equation 2.10 simply states that the radiated field $\mathbf{E}(\mathbf{r})$ is the tangential component of $\mathbf{T}(\mathbf{r})$. This relationship is expressed in Cartesian coordinates by Ludwig [42], and for all three of his definitions of cross-polarization the following holds close to the main beam:

$$E_{\text{ref}} \simeq T_y \text{ and } E_{\text{cross}} \simeq T_x.$$

The Cartesian components of $\mathbf{T}$ transform independently to the components of $\mathbf{J}_{\text{ap}}$, so for example a signal with purely reference polarization gives only J_y aperture currents. Unless it is a Huygens source, however, a purely y-polarized feed produces some J_x currents due to the curvature of the reflector [3, pp. 617–9]. Thus, even for uniform illumination, J_y is graded across the aperture and would appear so in holography results.

In general, a satellite source will have an arbitrary but hopefully constant polarization with respect to the aperture coordinates, and the feed will be matched as well as it can be. Similar gradings will occur in the recovered aperture currents, but as long as a unique phase centre is presented by the feed all components of the currents have the same phase. Since this is all that matters for surface diagnosis, polarization effects are ignored in the experiments to be described. Feed support struts cause cross-polarization due to scattering, but also aperture blockage which has more serious consequences for holography — see section 4.1.1. Feed phase characteristics and polarization of celestial sources are mentioned in section 4.1.2.

Another effect which is ignored in this thesis is diffraction at the edge of the reflector. The surface integral truncates the Physical Optics currents in a non-physical way, and in reality currents are induced along the edge which alter the radiated far field in the side-lobe regions; they may dominate the field there. Bach [2] reviews scattering mechanisms neatly. Diffraction by curved edges can be treated effectively by the Geometrical Theory of Diffraction [32, sect. 6.6], and it would be interesting to investigate its importance in field regions where holographic measurements take place. After all, if some side-lobe field values are caused by edge currents rather than surface currents then to use the Fourier transform relationship is to interpret these values wrongly, and this is detrimental to the surface error map.

2.1.3 Dual Reflectors

Front-fed reflectors are paraboloidal so that a simple feed design gives a uniform phase in the aperture. This is not enough to ensure high antenna efficiency: a combination of near-uniform aperture illumination and low feed spillover is required. A different illumination profile is required if it is important to reduce side-lobe levels near the main beam. These aims can be achieved by careful feed design, a task which is made difficult by other constraints such as low cross-polarization.

The use of a dual reflector design, usually of the Cassegrain type, allows feed specifications to be relaxed. Even a standard Cassegrain geometry gives a higher efficiency because a more uniform illumination of the main reflector is permitted with lower spillover. Spillover is directed at the sky instead of the 'hot' ground. In addition, bulky receivers can be placed nearer to the feed than would be possible for a prime focus antenna.

By shaping the subreflector an ideal illumination profile may be approached. High efficiency, satisfactory side-lobe levels, and low feed VSWR may be achieved by adjusting reflector shapes instead of relying entirely on the feed characteristics. Originally, geometrical optics methods were used to find the subreflector shape which distributed feed power in the desired way, and then to find adjustments to the main reflector to compensate the resulting phase distortion. Better results can be obtained by taking into account diffraction at the subreflector, particularly when it is not large in terms of the wavelength. Illumination is still optimized by shaping the subreflector, but it is found that good performance can often be achieved without changing the main reflector: axial movement of the subreflector gives the necessary degree of freedom. A thorough review of dual reflector optimization methods may be found in [12].

Microwave Holography is not inhibited by the presence of a subreflector. The 'equivalent paraboloid', shown in figure 2.4, simply emphasizes that in the ideal case this is yet another aperture with uniform phase. The important change is in the interpretation of the data, since phase distortions represented in a holographic aperture map may be due to errors in the subreflector as well as the main reflector. Mayer, et al. [45] have used this idea to their advantage, designing from holographic measurements a new secondary reflector which compensates for errors in the main reflector. It was easier to fabricate the small reflector than to alter the large one.

Cassegrain antennas have some disadvantages: the main one for holo-

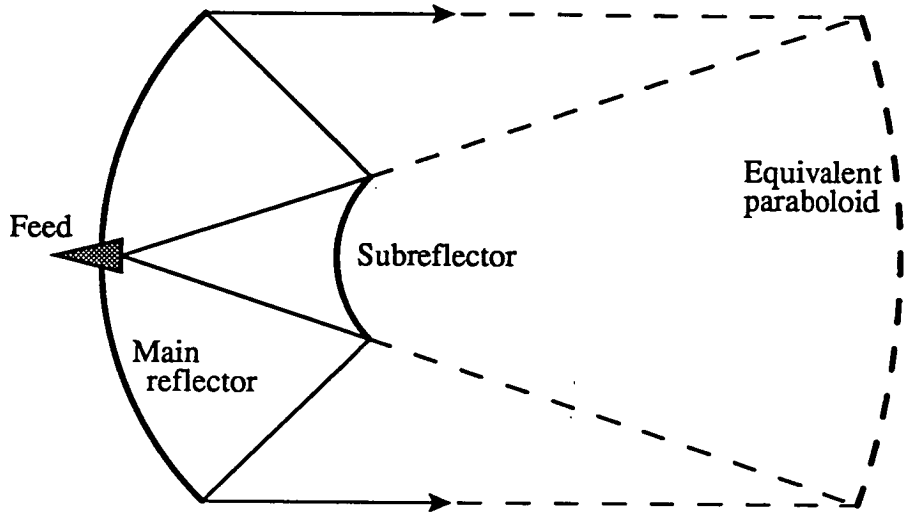


Figure 2.4: The equivalent prime-focus paraboloid for a Cassegrain antenna. Notice its long focal length.

graphic measurements is diffraction by the edge of the subreflector. Shaping allows the subreflector illumination to be more tapered, minimizing such diffraction, but effects have been noticed in holographic aperture maps all the same. Discussion of these effects is deferred until section 4.1.1.

2.2 Sampling and Truncation

The Fourier transform relationship between the aperture current function and the far field radiation pattern has been established. Holographic measurements are made in only a small solid angle in front of the antenna, otherwise the assumptions behind this relationship do not hold. In addition, the data do not form a continuous function but a discrete one at finite intervals of u and v. In this section the implications on the derived aperture current function are investigated.

For clarity, the arguments will be shown first for a one-dimensional aperture, and extended afterwards to two dimensions. Be aware that they are not mathematically rigorous because the existence of some infinite series is not proved. The notation introduced here will be used again in later chapters.

2.2.1 A One-Dimensional Aperture

Denote the aperture function by $a(x)$ and the resulting far field pattern by $A(u)$, where $u = \sin\theta$ measures the angle of the field point to the reflector axis. If all distances are expressed in wavelengths, these functions are related by a Fourier transform:

$$\begin{aligned} A(u) &= \int_{-\infty}^{+\infty} a(x) e^{-j2\pi ux}\, dx \\ a(x) &= \int_{-\infty}^{+\infty} A(u) e^{+j2\pi ux}\, du. \end{aligned} \tag{2.13}$$

In order to give the second integral infinite limits, $A(u)$ must be defined to be zero beyond the interval $[-1, +1]$, which represents directions in the forward hemisphere.

Let the sampling interval in u be δ_u, and define the discrete function $A_p(m)$ by

$$A_p(m) = A(m\delta_u) \text{ for all integral } m. \tag{2.14}$$

Approximating the Fourier integral for $a(x)$ with a sum gives the Fourier series pair

$$\begin{aligned} a_p(x) &= \sum_{m=-\infty}^{m=+\infty} A_p(m) e^{+j2\pi\delta_u mx}\, \delta_u \\ \delta_u A_p(m) &= \delta_u \int_{-\frac{1}{2\delta_u}}^{+\frac{1}{2\delta_u}} a_p(x) e^{-j2\pi\delta_u mx}\, dx, \end{aligned} \tag{2.15}$$

so the integral expressions for $A_p(m)$ and $A(m\delta_u)$ can be equated, after Roberts and Mullis [68, p. 114]:

$$\begin{aligned} \int_{-\frac{1}{2\delta_u}}^{+\frac{1}{2\delta_u}} a_p(x) e^{-j2\pi\delta_u mx}\, dx &= \int_{-\infty}^{+\infty} a(x) e^{-j2\pi\delta_u mx}\, dx \\ &= \sum_{k=-\infty}^{k=+\infty} \int_{\frac{k}{\delta_u}-\frac{1}{2\delta_u}}^{\frac{k}{\delta_u}+\frac{1}{2\delta_u}} a(x) e^{-j2\pi\delta_u mx}\, dx \\ &= \sum_{k=-\infty}^{k=+\infty} \int_{-\frac{1}{2\delta_u}}^{+\frac{1}{2\delta_u}} a(x+\tfrac{k}{\delta_u}) e^{-j2\pi\delta_u mx} e^{-j2\pi mk}\, dx \\ &= \int_{-\frac{1}{2\delta_u}}^{+\frac{1}{2\delta_u}} \left[\sum_{k=-\infty}^{k=+\infty} a(x+\tfrac{k}{\delta_u}) \right] e^{-j2\pi\delta_u mx}\, dx. \end{aligned}$$

This is true for all m, so the integrands must be equivalent, giving

$$\begin{aligned} a_p(x) &= \sum_{k=-\infty}^{k=+\infty} a(x+\tfrac{k}{\delta_u}) \\ &= \mathrm{III}(\delta_u x) * a(x). \end{aligned} \tag{2.16}$$

$\text{III}(t)$ is the sampling and replicating function of Bracewell [8], expressed using the impulse function by $\text{III}(t) = \sum_{k=-\infty}^{k=+\infty} \delta(t-k)$.

The effect of having a finite sampling interval is to replicate the aperture function at intervals in x of $1/\delta_u$. For this reason δ_u is chosen so that

$$\delta_u = \frac{\kappa}{D} \text{ with } 0 < \kappa \leq 1 \tag{2.17}$$

where D is the diameter of the antenna in wavelengths. Each replication of $a(x)$ is then separated from its neighbours because the aperture current $a(x)$ is zero outside the interval

$$\left[-\frac{D}{2}, +\frac{D}{2}\right] \subseteq \left[-\frac{1}{2\delta_u}, +\frac{1}{2\delta_u}\right].$$

Within these intervals the functions $a(x)$ and $a_p(x)$ are equal, so sampling has not affected the quality of the recovered aperture current.

The sampled field $A_p(m)$ is truncated to $\tilde{A}(m)$ by measuring only N_u samples, whence

$$\tilde{A}(m) = A_p(m)\text{II}\left(\frac{m}{N_u}\right) \text{ where } \text{II}(t) = \begin{cases} 1 & \text{if } |t| \leq \frac{1}{2} \\ 0 & \text{otherwise} \end{cases}. \tag{2.18}$$

The range of u covered by non-zero $\tilde{A}(m)$ is $W_u = N_u\delta_u$. Let $N_u = 2L_u + 1$ so that there are L_u far field points on either side of the main beam. By the convolution theorem for Fourier series,

$$\begin{aligned} \tilde{a}(x) &= \sum_{m=-\infty}^{m=+\infty} \tilde{A}(m) e^{+j2\pi\delta_u m x}\, \delta_u \\ &= \delta_u \int_{-\frac{1}{2\delta_u}}^{+\frac{1}{2\delta_u}} a_p(x') b(x-x')\, dx' \end{aligned}$$

where

$$\begin{aligned} b(x) &= \sum_{m=-\infty}^{m=+\infty} \text{II}\left(\tfrac{m}{N_u}\right) e^{+j2\pi\delta_u m x} \\ &= \sum_{m=-L_u}^{m=+L_u} \cos(2\pi\delta_u m x) + j \sum_{m=-L_u}^{m=+L_u} \sin(2\pi\delta_u m x) \\ &= \frac{\sin(\pi W_u x)}{\sin(\pi\delta_u x)}. \end{aligned} \tag{2.19}$$

This result is derived in appendix A; note that the sum of sine terms is zero. Inserting equation 2.16 and observing that $b(x)$ is periodic with period $1/\delta_u$,

it follows that

$$\begin{aligned}
\tilde{a}(x) &= \delta_u \int_{-\frac{1}{2\delta_u}}^{+\frac{1}{2\delta_u}} \sum_{k=-\infty}^{k=+\infty} a(x + \tfrac{k}{\delta_u}) b(x - x')\, dx' \\
&= \delta_u \sum_{k=-\infty}^{k=+\infty} \int_{\frac{k}{\delta_u} - \frac{1}{2\delta_u}}^{\frac{k}{\delta_u} + \frac{1}{2\delta_u}} a(x') b(x - x' + \tfrac{k}{\delta_u})\, dx' \\
&= \delta_u \int_{-\infty}^{+\infty} a(x') b(x - x')\, dx' \\
&= a(x) * \delta_u \frac{\sin(\pi W_u x)}{\sin(\pi \delta_u x)}. \qquad (2.20)
\end{aligned}$$

Thus the aperture current estimated from holographic data is the convolution of the true aperture current with $\delta_u b(x)$. A similar result was derived in [68, sect. 4.8]. The effects of this need some discussion.

2.2.2 The Recovered Aperture Current

The convolving function $\delta_u b(x)$ is most easily visualized when written

$$\delta_u b(x) = \delta_u \frac{\sin(\pi N_u \delta_u x)}{\sin(\pi \delta_u x)}.$$

It is graphed against $\delta_u x$ in figure 2.5 for $N_u = 21$. The curve is familiar as the amplitude response of a diffraction grating with N_u thin grooves. As N_u becomes large, the primary maxima get higher and the subsidiary maxima decrease more rapidly with $\delta_u x$: the function approaches $\mathrm{III}(\delta_u x)$ as expected.

The convolution for a particular offset is shown diagrammatically in figure 2.6. The primary maxima are responsible for replicating the aperture function at intervals of $1/\delta_u$. The presence of subsidiary maxima means that each $\tilde{a}(x)$ includes a contribution from $a(x)$ across the whole aperture: $\tilde{a}(x)$ is not zero beyond the physical aperture. This is termed *aliasing*, and it would be unacceptably damaging if $1/\delta_u < D$, causing the recovered current near the edge of the aperture to have an equal contribution from the opposite edge. Equation 2.17 ensures that this is not the case.

It is possible to control aliasing by varying κ and N_u. Here are some possibilities:

- *Decreasing κ with constant N_u* stretches the convolving function. This lessens aliasing by giving a greater distance between primary maxima, but also gives poorer resolution in the recovered aperture.

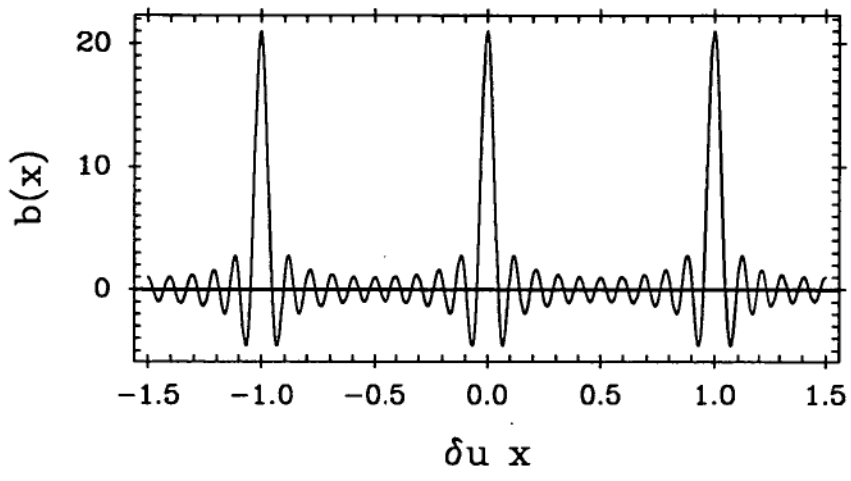


Figure 2.5: The convolving function $\sin(\pi W_u x)/\sin(\pi \delta_u x)$ graphed against $\delta_u x$ for $N_u = 21$.

- *Decreasing κ with constant W_u* lessens aliasing noticeably when the offset x approaches the edge of the aperture. The primary maxima are further apart, and the envelope of the subsidiary maxima reaches a lower level between them. The convolving function is only slightly changed near the primary maxima, since with respect to the peak value it is nearly independent of κ:

$$\frac{b(x)}{b(0)} = \frac{\sin(\pi W_u x)}{N_u \sin(\pi \delta_u x)} \simeq \frac{\sin(\pi W_u x)}{\pi W_u x} \text{ for } \delta_u x \ll 1.$$

- *Increasing N_u with constant κ* has been considered above. The convolving function approaches $\mathrm{III}(\delta_u x)$, so each aperture point is less dependent on the rest of the aperture.

The resolution obtained in the aperture is determined by the width of the primary maxima. As in the field of optics, the resolution limit is specified by the peak-to-first-minimum distance of the intensity pattern, which is the peak-to-first-zero distance of the convolving function:

$$\delta_x = \frac{1}{W_u} = \frac{1}{N_u \delta_u}. \tag{2.21}$$

This is the Nyquist sampling interval determined by truncation of the far field pattern, and with it all spatial frequencies in the 'bandwidth' W_u can be represented. It is sufficient to evaluate the current $\tilde{a}(x)$ only for $x = p\delta_x$, where p is an integer in the range $-L_u, \ldots, +L_u$. Then the Fourier series

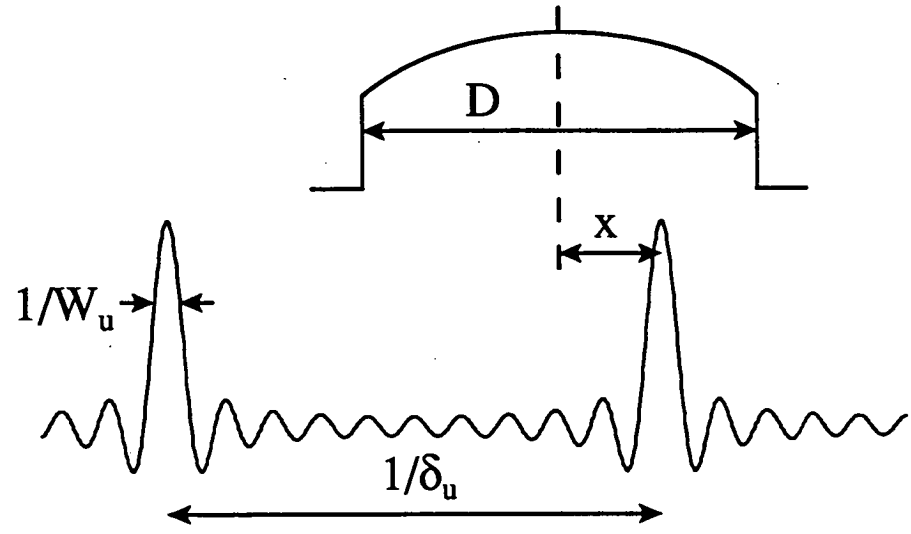


Figure 2.6: An example aperture function (amplitude only shown) and convolving function which when multiplied pointwise produce $\tilde{a}(x)$ at a particular offset x.

expression for $\tilde{a}(x)$ takes the form of an inverse discrete Fourier transform:

$$\begin{aligned}\tilde{a}(p\delta_x) &= \sum_{m=-\infty}^{m=+\infty} \tilde{A}(m) e^{+j2\pi\delta_u m\delta_x p}\,\delta_u \\ &= \sum_{m=-L_u}^{m=+L_u} \tilde{A}(m) e^{+j\frac{2\pi}{N_u}mp}\,\delta_u \\ &= \delta_u N_u a_{\mathrm{DFT}}(p). \end{aligned} \tag{2.22}$$

An inverse DFT is assumed to have a positive kernel exponent and include a scaling by $1/N_u$. The efficient FFT algorithm is available for its computation.

As an aside, imagine that the operations on the far field pattern had been performed in the opposite order, that is, the far field had been truncated first and then sampled. Let the truncated data be $A_t(u) = A(u)\mathrm{II}(u/W_u)$ and the convolution theorem for Fourier transforms gives

$$\begin{aligned} a_t(x) &= \int_{-\infty}^{+\infty} A_t(u) e^{+j2\pi ux}\,du \\ &= a(x) * \mathrm{sinc}\,(W_u x) \\ &= a(x) * \delta_u \frac{\sin(\pi W_u x)}{\pi\delta_u x}. \end{aligned}$$

The sampled data $\tilde{A}(m) = A_t(m\delta_u)$ for integral m, and as before the aperture

current is estimated by a Fourier series:

$$\begin{aligned}\tilde{a}(x) &= \sum_{m=-\infty}^{m=+\infty} \tilde{A}(m) e^{+j2\pi\delta_u m x}\, \delta_u \\ &= \mathrm{III}\,(\delta_u x) * a_t(x) \\ &= a(x) * \mathrm{III}\,(\delta_u x) * \delta_u \frac{\sin(\pi W_u x)}{\pi \delta_u x}. \end{aligned} \tag{2.23}$$

It can be shown (see appendix A) that the replication of the sinc function is the grating function $b(x)$ defined by equation 2.19, so this convolution is indeed the same as equation 2.20.

2.2.3 To Two Dimensions

These results can be generalized easily for the two-dimensional aperture shown in figure 2.2. Begin with the continuous far field function $A(u,v)$ and aperture function $a(x,y)$, where

$$\begin{aligned} u &= \sin\theta\cos\phi &&= \text{direction cosine to x-axis} \\ v &= \sin\theta\sin\phi &&= \text{direction cosine in y-axis} \end{aligned} \tag{2.24}$$

and x and y are measured in wavelengths. Then

$$\begin{aligned} A(u,v) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} a(x,y) e^{-j2\pi(ux+vy)}\, du\, dv \\ a(x,y) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} A(u,v) e^{+j2\pi(ux+vy)}\, du\, dv. \end{aligned} \tag{2.25}$$

Let the sampled far field be $A_p(m,n) = A(m\delta_u, n\delta_v)$ for integral m and n, and estimate the aperture current with

$$\begin{aligned} a_p(x,y) &= \sum_{m=-\infty}^{m=+\infty}\sum_{n=-\infty}^{n=+\infty} A_p(m,n) e^{+j2\pi(\delta_u m x + \delta_v n y)}\, \delta_u\, \delta_v \\ &= \sum_{k=-\infty}^{k=+\infty}\sum_{l=-\infty}^{l=+\infty} a(x+\tfrac{k}{\delta_u}, y+\tfrac{l}{\delta_v}) \\ &= \mathrm{III}\,(\delta_u x, \delta_v y) * a(x,y). \end{aligned}$$

Truncate $A_p(m,n)$ to N_u samples in u and N_v samples in v, defining $\tilde{A}(m,n) = A_p(m,n)\mathrm{II}\,(m/N_u, n/N_v)$. The corresponding aperture current estimate $\tilde{a}(x,y)$ forms a Fourier series pair with $\delta_u\delta_v\tilde{A}(m,n)$ as follows:

$$\begin{aligned} \tilde{a}(x,y) &= \sum_{m=-\infty}^{m=+\infty}\sum_{n=-\infty}^{n=+\infty} \tilde{A}(m,n) e^{+j2\pi(\delta_u m x + \delta_v n y)}\, \delta_u\, \delta_v \\ \delta_u\delta_v\tilde{A}(m,n) &= \delta_u\delta_v \int_{-\frac{1}{2\delta_u}}^{+\frac{1}{2\delta_u}}\int_{-\frac{1}{2\delta_v}}^{+\frac{1}{2\delta_v}} \tilde{a}(x) e^{-j2\pi(\delta_u m x + \delta_v n y)}\, dx\, dy. \end{aligned} \tag{2.26}$$

First Dimension	Second Dimension
$u = m\delta_u = m\kappa/D_x$	$v = n\delta_v = n\kappa/D_y$
$x = p\delta_x = p/W_u$	$y = q\delta_y = q/W_v$
$W_u = N_u\delta_u$	$W_v = N_v\delta_v$
$N_u = 2L_u + 1$	$N_v = 2L_v + 1$
$N_u\delta_u\delta_x = 1$	$N_v\delta_v\delta_y = 1$

Table 2.1: Sampling relationships for one and two dimensions. The aperture has dimensions D_x in x and D_y in y. All distances are measured in wavelengths.

The same working as before gives

$$\tilde{a}(x,y) = a(x,y) * \left[\delta_u \frac{\sin(\pi W_u x)}{\sin(\pi\delta_u x)}\, \delta_v \frac{\sin(\pi W_v y)}{\sin(\pi\delta_v y)}\right] \tag{2.27}$$

to match equation 2.20. The comments made about aliasing apply equally well in the two-dimensional case.

It is sufficient to evaluate the aperture current estimate $\tilde{a}(x,y)$ at discrete x and y, which turns the Fourier series into an inverse DFT:

$$\tilde{a}(p\delta_x, q\delta_y) = \delta_u\delta_v N_u N_v a_{\mathrm{DFT}}(p,q) \tag{2.28}$$

where

$$p = -L_u, \ldots, +L_u \text{ and } q = -L_v, \ldots, +L_v.$$

The sampling relationships are summarized in table 2.1. It is worth repeating the relation between the span of far field measurements and the resolution of the recovered aperture current function:

$$\delta_x = \frac{1}{W_u} \text{ and } \delta_y = \frac{1}{W_v}.$$

It seems that the only way to improve the aperture resolution is to take a greater number of far field measurements. In chapter 5 another method will be introduced.

A useful equation is derived from Parseval's relation for Fourier series. The average power in the aperture is given by

$$\delta_u\delta_v \int_{-\frac{1}{2\delta_u}}^{+\frac{1}{2\delta_u}} \int_{-\frac{1}{2\delta_v}}^{+\frac{1}{2\delta_v}} |\tilde{a}(x,y)|^2 \, dx\, dy \;=\; \sum_{m=-\infty}^{m=+\infty} \sum_{n=-\infty}^{n=+\infty} \left|\delta_u\delta_v \tilde{A}(m,n)\right|^2$$

which can be written

$$\|\tilde{a}\|^2 = \delta_u{}^2 \delta_v{}^2 \|\tilde{A}\|^2. \tag{2.29}$$

The symbol $\|\cdot\|^2$ has two meanings in this equation, but it is not ambiguous. When applied to a periodic function it refers to the average integral over one period of the squared absolute value. When applied to a discrete function it refers to the sum of the squared absolute value over the entire range.

2.3 The Antenna Geometry

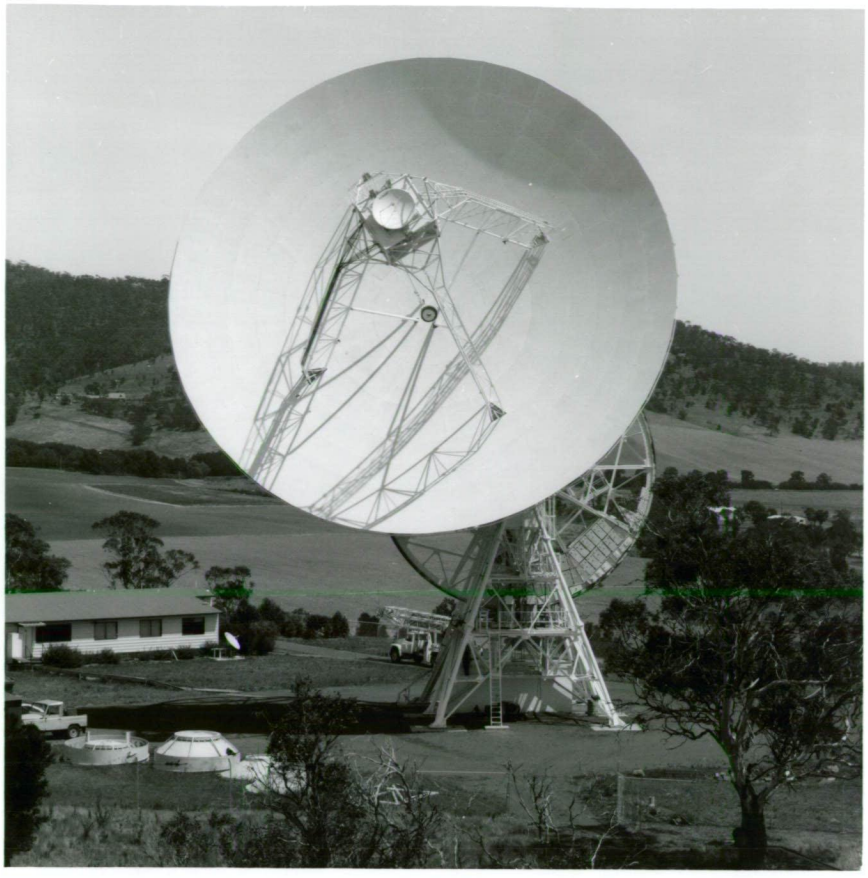
2.3.1 The Mount and (u, v) Coordinates

Now that the theoretical basis of holographic measurement is understood, it is time to look at some properties of real antennas. This thesis is about large, steerable reflectors, and in particular the Mount Pleasant 26 m radio telescope where the measurements were performed. This antenna, illustrated in figure 2.7, is unusual because it has an XYEW mount. The reflector is tilted to the East or West about the 'Y-axis'. The inclination of this upper axis to the North or South is determined by tilting about the 'X-axis' which runs East-West (hence the designation of the mount). A similar scheme is the XYNS mount, which has the lower axis running North-South instead.

Most modern antennas have an 'azimuth/elevation' mount, here designated AZEL, where the dish tilts up or down on an elevation axis which itself is rotated about a vertical azimuth axis. This mount is the simplest to build, and has the advantage that its axes correspond to conventional polar horizon coordinates. The geometry of an antenna has an important bearing on the way holographic data are collected, so it is impossible to omit a discussion of the AZEL mount. In fact, all three mounts mentioned have features in common, and they can be treated by the same analysis.

Adopt a rectangular horizon coordinate system that is right-handed and has the x-axis pointing to the East. In this system the axis angles for each mount are shown in figure 2.8. When the axis angles are zero the antenna

Figure 2.7: *Overleaf* shows the 26 m radio telescope at the Mount Pleasant Observatory, and the 2.1 m reference antenna. When the former photograph was taken the reference antenna was mounted on top of the feed cabin; later it was mounted at the north-eastern foot of the radio telescope so that the source remained in its main beam during large holographic maps.



points up for XY mounts or due North for the AZEL mount. In other positions these equations are found for the unit pointing vector $\hat{\mathbf{a}} = (x, y, z)$ for the antenna:

$$\left.\begin{aligned} x &= \sin Y \\ y &= \cos Y \sin X \\ z &= \cos Y \cos X \end{aligned}\right\} \text{XYEW} \tag{2.30}$$

$$\left.\begin{aligned} x &= \cos Y \sin X \\ y &= \sin Y \\ z &= \cos Y \cos X \end{aligned}\right\} \text{XYNS} \tag{2.31}$$

$$\left.\begin{aligned} x &= \cos E \sin A \\ y &= \cos E \cos A \\ z &= \sin E \end{aligned}\right\} \text{AZEL} \tag{2.32}$$

The similarity of these equations is clear. It arises because each mount consists of two perpendicular axes, with a separation which is usually zero in the AZEL case. If either XY mount is tipped over so its X-axis points into the ground, it becomes an AZEL mount. An equatorial mount also has this property, although it will not be considered here: the use of computers for rapid coordinate conversions has removed the main reason for adopting this mount, and few are built nowadays.

The far field position (u, v, w) is referred to a coordinate system fixed to the aperture, while the position of a source, whether celestial or man-made, is expressed in horizon coordinates for antenna pointing. If the antenna is pointed to the vicinity of a source at horizon position (x, y, z), this vector must be expressed in aperture coordinates to find the corresponding far field position. Conversely, to measure the received field at a far field position (u, v, w) antenna axis angles must be found such that this vector converts to the source position (x, y, z) in horizon coordinates.

Provided that the source can be assumed infinitely removed, the transformation is a three-dimensional rotation of the coordinate system. The aperture frame used in section 2.1 will be oriented so that it matches the horizon frame when the antenna points upwards, with the x-axis running East; the two frames are illustrated in figure 2.9. As the antenna tilts the aperture origin moves, but the amount is negligible compared to the 37850 km range of AUSSAT-1. Three-dimensional rotations can be composed of rotations about the coordinate axes using the Euler angle method. By re-labelling the axes when necessary, the transformation for any mount can be expressed in

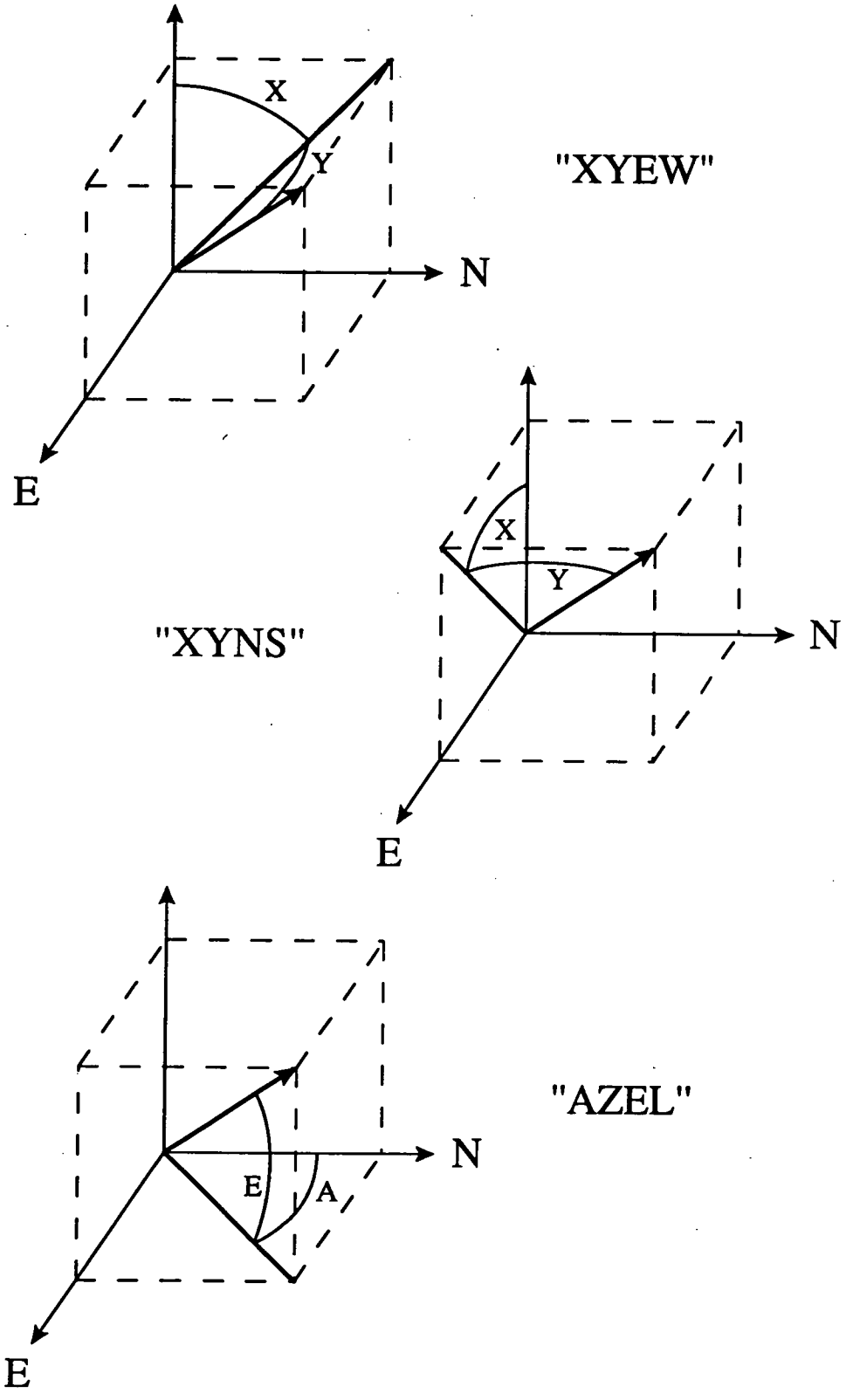


Figure 2.8: The positive sense of the axis angles of three types of antenna mount.

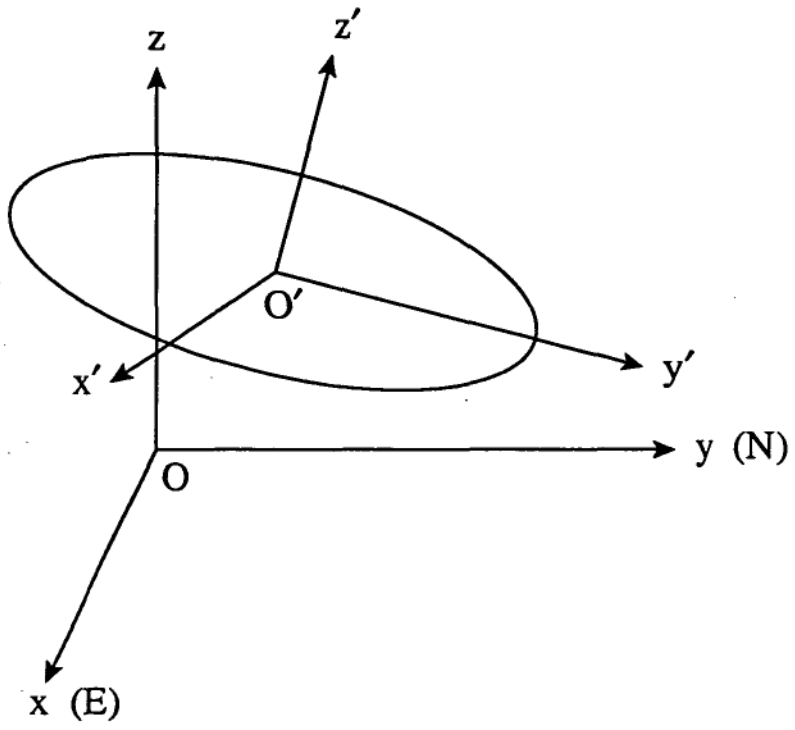


Figure 2.9: The horizon and aperture frames, related by a three-dimensional rotation and a translation.

a standard form. This connects the mount geometries in a nice way, and is very helpful when designing software.

The Goldstein convention [22, sect. 4.4] for Euler angles will be followed here. The original, horizon coordinate system is rotated anticlockwise about the z-axis by an angle ϕ. This new system is rotated anticlockwise about its x-axis by an angle θ. Finally, the second new system is rotated anticlockwise about its z-axis by an angle ψ to give the transformed, aperture coordinate system. Any three-dimensional rotation can be represented by unique Euler angles ϕ, θ, and ψ, and in matrix notation the aperture vector pointing to a source with horizon coordinates (x, y, z) can be written

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{pmatrix} \cdot \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \tag{2.33}$$

There is no need to write the product matrix in full; it will be denoted by

$\mathbf{P}(\phi, \theta, \psi)$. If the aperture system is indicated by a prime, then

$$\mathbf{r}' = \mathbf{P}(\phi, \theta, \psi) \cdot \mathbf{r}. \tag{2.34}$$

If the horizon source vector $\mathbf{r}$ is a unit vector, so is $\mathbf{r}'$ and its first two components (u, v) are the Fourier transform variables for the far field domain.

The Euler angles defined above do not always correspond to rotations about the antenna's axes; this can be achieved by re-labelling the coordinate axes. No change is necessary for the AZEL mount, but for the XY mounts the horizon vector must be 'rolled up' one position:

$$\mathbf{r} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \overset{\text{roll up}}{\longrightarrow} \begin{pmatrix} y \\ z \\ x \end{pmatrix} = \uparrow \mathbf{r} \tag{2.35}$$

where the $\uparrow$ prefix indicates a vector rolled up by one row. This is a rotation of the frame by 120° about the vector $(1, 1, 1)$, and it can be transformed to a similarly rotated aperture frame, giving $\uparrow \mathbf{r}'$ instead of $\mathbf{r}'$. Now the Euler angles for the transformation can be given in terms of the antenna's axes' angles for the XYEW mount:

$$\uparrow \mathbf{r}' = \mathbf{P}(-X, Y, 0) \cdot \uparrow \mathbf{r} \tag{2.36}$$

and the XYNS mount:

$$\uparrow \mathbf{r}' = \mathbf{P}(-Y, X, 0) \cdot \uparrow \mathbf{r} \tag{2.37}$$

and the AZEL mount:

$$\mathbf{r}' = \mathbf{P}(-A, E - \pi/2, 0) \cdot \mathbf{r}. \tag{2.38}$$

By reversing the 'roll' it is easy to obtain (u, v, w) from $\uparrow \mathbf{r}'$.

The far field position can be found from the source and antenna positions using these equations. In order to take observations, however, the antenna must be steered to give a predetermined far field position, requiring the equations to be solved for X and Y or A and E. Because the third Euler angle ψ is zero for all the mounts, only ϕ and θ need be found from the two rotationally related vectors $\mathbf{r}$ and $\mathbf{r}'$. Appendix B shows how this can be done.

How necessary are these computations? After all, the aperture frame is oriented so that turning on one axis or the other corresponds closely to changing u or v. The far field maps are small in extent, so a square grid

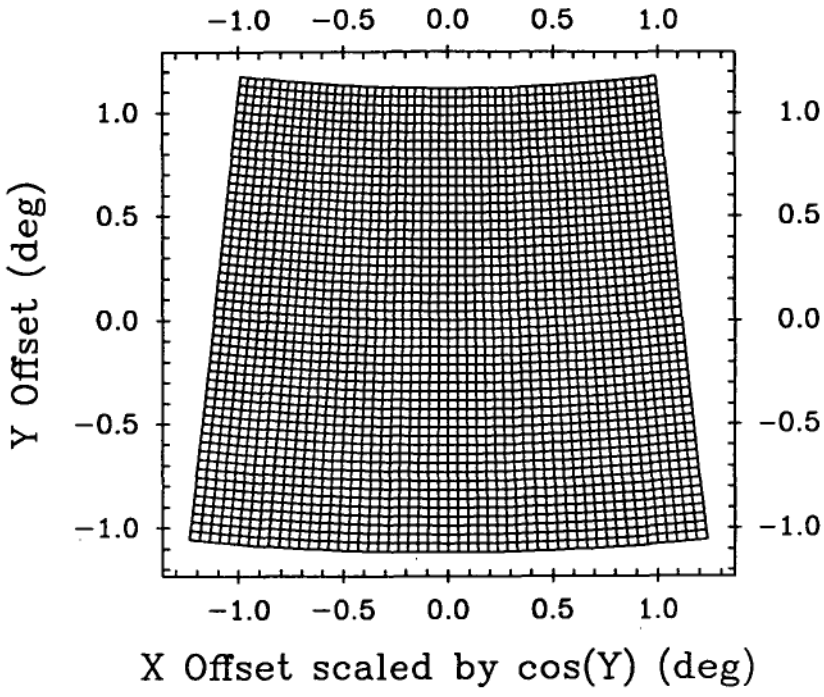


Figure 2.10: A square grid in (u, v) as it appears in (X, Y) for a XYEW mount. Positions are shown relative to the grid centre, which is 10° from the eastern pole. The grid size is 53 × 53 points and the intervals $\delta_x = \delta_y = 0.00075$.

in (X, Y) or (A, E) might be close enough to a square grid in (u, v). This depends on the proximity of the mount's pole, where the antenna's direction is independent of the angle of its lower axis. The XYEW mount has poles on the horizon due East and due West, so turning the X-axis has no effect when pointing in these directions. The XYNS mount has poles to the North and South, and the AZEL mount has a pole at the zenith. (This is why the Mount Pleasant antenna has an XY mount: it was originally designed for tracking satellites, where a pole at the zenith is a serious liability.) Close to a pole, antenna motion maps onto the sky in a very non-linear fashion. Figure 2.10 shows lines of constant u and v as they appear when converted to X and Y angles for an XYEW antenna. The original grid was square, but any attempt to fit a square (X, Y) grid will result in serious errors. This is an extreme case only 10° from the eastern pole, but the effect is significant even when observing AUSSAT-1, 76° from the pole.

The problem of obtaining a square (u, v) grid for holographic measurements has been addressed by Rahmat-Samii [65] for an AZEL mount. The method described above can be applied to any mount consisting of two per-

pendicular axes with no extra effort. In combination with the equations for translation of the origin in the next section, the rotation matrix formulation allows a complete transformation from ground-based to antenna-based coordinates.

2.3.2 Antenna Motion and the Phase Reference

If the phase reference antenna is placed on the ground, care must be taken to refer the far field phase to an origin fixed to the main antenna. The origin is best placed somewhere on the reflector's axis to maintain symmetry; its position along the axis has a significance which will be discussed soon. Up to now the far field phase has not been defined precisely. It is the phase of the antenna's radiation with respect to a spherical wave emanating from the origin. In terms of reception of incoming planar wavefronts, it is the phase of the antenna's output voltage with respect to the phase of the wavefront seen at the origin.

Figure 2.11 is a diagram of a general antenna mount with a 'lower' (X or azimuth) axis and an 'upper' (Y or elevation) axis. Let the perpendicular separation of the axes be d. Choose an origin a distance l along the reflector axis from its intersection with the mount's upper axis. It is assumed that these two do cross, and at the point closest to the lower axis; this is nearly always the case, and the simplification is not a fundamental one. The important points in the diagram are the centre of the lower axis V, the centre of the upper axis W, the aperture frame origin O, and the location R of the phase reference antenna. V is stationary and is known as the 'VLBI' point of the antenna. Finally, denote the unit vector along the reflector axis by $\hat{\mathbf{a}}$, and the unit vector towards the source by $\hat{\mathbf{s}}$.

The phase error due to the separation of R from O is of course the path difference $\mathbf{R} \cdot \hat{\mathbf{s}}$ shown in the diagram, or in radians

$$\phi_t = +k\,\mathbf{R} \cdot \hat{\mathbf{s}} \tag{2.39}$$

where $k = 2\pi/\lambda$ and $\mathbf{R}$ is the vector from O to R. This phase will have a component due to motion of the source and a component due to motion of the antenna as it covers the (u, v) grid. It is useful to separate these because on-source calibration measurements can show the source motion component independently. This is discussed in section 3.2.1.

The position of R must be specified somehow, so it is assumed that $\mathbf{R}_V$

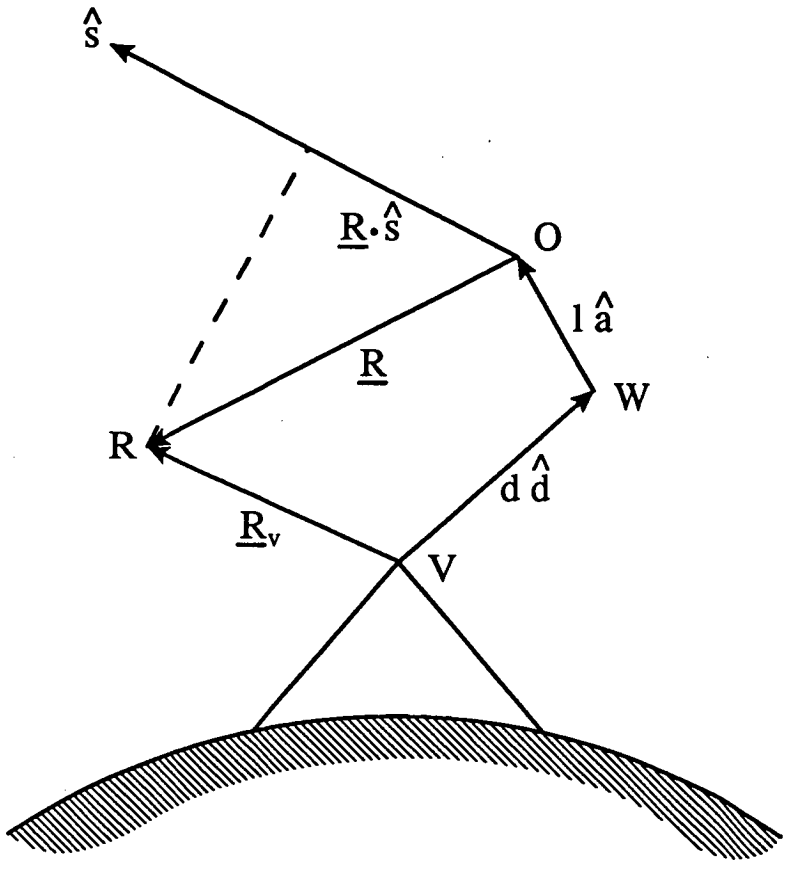


Figure 2.11: A general antenna mount with two orthogonal rotation axes whose closest points V and W are separated by d. The aperture frame origin O is a distance l along the reflector axis from W, and the phase reference antenna is located at R. Unit vectors in the source and axis directions are $\hat{\mathbf{s}}$ and $\hat{\mathbf{a}}$.

is known, the vector from V to R. Then

$$\mathbf{R} = \mathbf{R}_V - d\,\hat{\mathbf{d}} - l\,\hat{\mathbf{a}}. \tag{2.40}$$

Here $\hat{\mathbf{d}}$ is the unit vector from V to W, which is easily shown to be

$$\hat{\mathbf{d}} = \begin{pmatrix} 0 \\ \sin X \\ \cos X \end{pmatrix}_{\text{XYEW}} = \begin{pmatrix} \sin X \\ 0 \\ \cos X \end{pmatrix}_{\text{XYNS}} = \begin{pmatrix} \sin A \\ \cos A \\ 0 \end{pmatrix}_{\text{AZEL}}. \tag{2.41}$$

The components of $\hat{\mathbf{a}}$ are given in equations 2.30, 2.31, and 2.32. Thus

$$\phi_t = +k\,(\mathbf{R}_V - d\,\hat{\mathbf{d}} - l\,\hat{\mathbf{a}}) \cdot \hat{\mathbf{s}} \tag{2.42}$$

can be computed. The phase ϕ_s due to source motion can be isolated from the phase ϕ_a due to antenna motion by using the on-source position as a reference, where $\hat{\mathbf{d}} = \hat{\mathbf{d}}_0$ and $\hat{\mathbf{a}} = \hat{\mathbf{s}}$. Thus

$$\phi_s = +k\,(\mathbf{R}_V - d\,\hat{\mathbf{d}}_0 - l\,\hat{\mathbf{s}}) \cdot \hat{\mathbf{s}} \tag{2.43}$$

giving

$$\phi_a = \phi_t - \phi_s = kd\,(\hat{\mathbf{d}}_0 - \hat{\mathbf{d}}) \cdot \hat{\mathbf{s}} + kl\,(\hat{\mathbf{s}} - \hat{\mathbf{a}}) \cdot \hat{\mathbf{s}}$$

or

$$\phi_a = \phi_a|_{l=0} + kl(1 - \cos\theta). \tag{2.44}$$

The far field phase w.r.t. O is related to the measured phase by

$$\phi_{\text{true}} = \phi_{\text{measured}} - \phi_t = \phi_{\text{measured}} - \phi_s - \phi_a. \tag{2.45}$$

The linear dependence of ϕ_a on the origin position l is not a surprise, and it has no effect on the aperture currents obtained by Fourier transform. However, the $-kl\cos\theta$ term modifies the far field phase in a way which does slightly affect the recovered currents, and it is not obvious why moving the 'aperture plane' containing the origin should make a difference. The total phase modification due to l is approximately quadratic in θ, since

$$kl(1 - \cos\theta) \simeq \frac{kl\theta^2}{2} \text{ for small } \theta.$$

It was originally identified in this context by Bennett, et al. [6].

The significance of this is that the extent of the map in the far field domain is giving some resolution in the z direction, normal to the aperture. In fact, the reflector surface of the antenna is a three-dimensional current

distribution that has a three-dimensional Fourier transform. The far field at a particular frequency is a cut through this domain that is the surface of a sphere, radius k, centred at the origin (the 'Ewald' sphere, a term borrowed from crystallography). Usually, holographic measurements occupy only a small region of this surface around the main-beam direction. The ability of wide-angle measurements to 'see around' the source distribution is made manifest by improved z resolution.

It can be demonstrated that the far field vector potential is the weighted sum of the two-dimensional Fourier transforms of each constant-z slice of the current distribution, and the weighting function is a phase term that is constant for $z = 0$: the aperture plane transform is not distorted. The equation can be abbreviated like this:

$$\text{far field} \simeq \frac{e^{+jkr}}{4\pi r} \int_z (\text{2-D F.T. of slice at } z) e^{-jkz\cos\theta}\, dz$$

where the integral covers the extent of the source currents. Thus the phase term $-kl\cos\theta$ produced by axial shifts of the origin has a clear physical significance: it is to edit the far field measurements so that the aperture plane $z = 0$ is the most faithfully recovered slice of the source current distribution.

This argument is made clearer by considering the two-dimensional inverse Fourier transform of the equation above:

$$\text{I.F.T. of far field} \simeq \frac{e^{+jkr}}{4\pi r} \int_z (\text{source slice at } z) * C_z(x,y)\, dz.$$

Each slice of the source currents is convolved with a function $C_z(x,y)$ before it is added to the others, forming what has until now been called the 'recovered aperture field'. $C_0(x,y)$ for $z = 0$ is a sinc function representing the horizontal aperture resolution δ_x and δ_y from section 2.2. The aperture plane itself is convolved with $C_0(x,y)$, while other planes are convolved with a broadened function that smears their features. The smoothing effect of this function increased rapidly with $|z|$ when the far field domain covers a large part of the Ewald sphere, leaving the aperture plane less and less perturbed.

The reader is referred to the paper by Cook, et al. [14], which describes these ideas in more detail and reports wide-angle antenna measurements that allowed surface features to be detected beneath feed support legs.

Scott and Ryle [76] explain the same phenomenon by saying that the aperture plane phase faithfully represents the surface errors provided that the distance of the surface from the plane is small compared to the Fresnel distance appropriate to the scale size of the errors. The radiation from a

very small surface feature will begin to evolve towards its far field pattern before reaching the aperture plane, so such a feature will be ill-determined by holograhic data. In section 4.1.2 it is shown that, for the experiments at Mount Pleasant, the uncertainties in a surface error map due to these propagation effects are smaller than the random errors in the map.

2.4 Panel Position Estimation

Most large reflector antennas do not have a continuous surface, but one made from panels that have been constructed as segments of the underlying paraboloid. The end result of Microwave Holography is an aperture phase function estimated on a grid of aperture points, from which corrections must be made to the panel positions if it is desired to improve the surface profile. It is very awkward to make corrections based on a contour map of surface errors deduced from the aperture phases, or any other direct form of display. Panel positions must be estimated somehow from the sparse aperture phase function, and then the heights of corners and other adjustment points must be calculated.

The first assumption necessary to do this is that the profile errors in each panel are smaller than the errors due to poor panel setting. Clearly, if this is not the case there is little use in adjusting the panels' positions. There are usually many panels on a reflector: figure 2.12 shows the panel layout for the Mount Pleasant 26 m antenna, which has 252 individual panels, although several are shadowed by the feed support structure as indicated by the grey area. When the panels are so much smaller than the reflector they can be engineered accurately.

The other assumption is that panel positions can be fitted to aperture phase instead of the errors normal to the surface. This is easier because the curved profile of a panel produces a flat aperture phase profile if it is the correct shape. Tilts and distortions of the panel are assumed to produce corresponding deviations of phase in the aperture region directly above the panel. This would be true were it not for the propagation effects discussed in the last section. Panel edges will not generally match and are separated by a gap, causing a very small-scale feature in the source current distribution which will become confused while propagating to the chosen aperture plane. However, the effect will be negligible over most of a panel's area and can be ignored in practice (section 4.1.2).

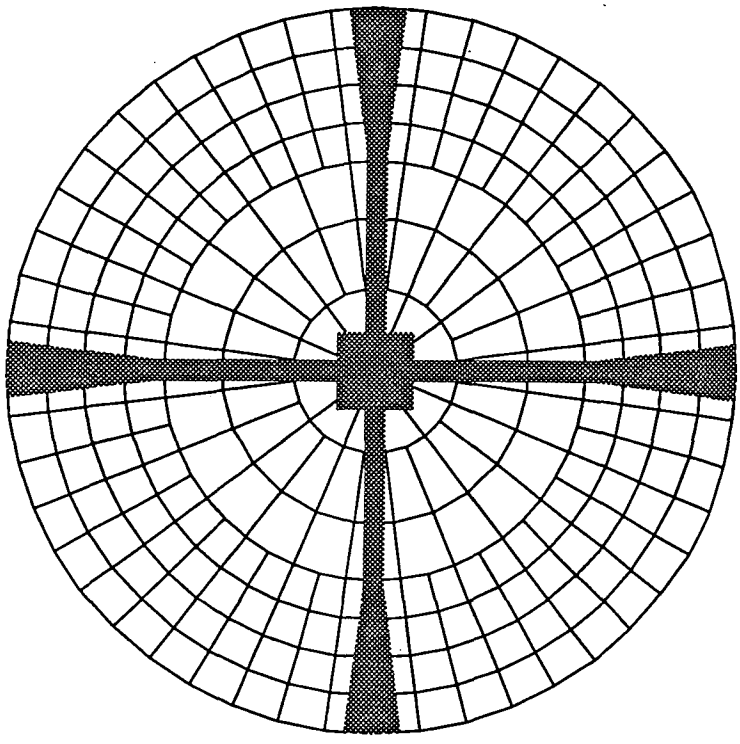


Figure 2.12: Panel layout for the Mount Pleasant 26 m antenna. Shadowing by the feed cabin and support legs is shown in grey.

Much more important is the limited spatial resolution of a map, δ_x and δ_y. The estimated aperture phase at any point is not the true phase at that point but an average of all phases in the neighbourhood, determined by the convolving function in section 2.2.2. Points on either side of a panel edge are significantly affected by the neighbouring panel: a point right on the edge represents neither panel meaningfully. With time and sensitivity constraints, a holography map will often have quite a small number of points lying on each panel [7]: they may all be within one resolution cell of the edge. In this chapter panel fitting proceeds regardless: the problem is addressed fully in chapter 5.

2.4.1 Fitting Rigid Panels by Least Squares

The aperture phase above a rigid panel follows a plane, and it is convenient to express the equation to the plane in the panel-based coordinate system shown in figure 2.13. To convert an aperture point (x, y) to panel coordinates (x_p, y_p) requires a rotation and a translation:

$$\begin{pmatrix} x_p + r_0 \\ y_p \end{pmatrix} = \begin{pmatrix} \cos\theta_0 & \sin\theta_0 \\ -\sin\theta_0 & \cos\theta_0 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} \tag{2.46}$$

or

$$\begin{aligned} x_p &= r\cos(\theta - \theta_0) - r_0 \\ y_p &= r\sin(\theta - \theta_0). \end{aligned} \tag{2.47}$$

Here (r, θ) is the polar form of (x, y), and (r_0, θ_0) is the geometrical centre of the trapezoidal panel. The equation to describe a plane is

$$\phi = \beta \frac{x_p}{h/2} + \alpha \frac{y_p}{w/2} + \ell, \tag{2.48}$$

so that β represents a panel tilt towards or away from the centre of the reflector, while α represents a tilt at right-angles to this. The coordinates are normalized by the panel's height and average width, so the radian difference across the panel is β or α.

The algorithm in appendix C for fitting a least-squares plane can be used to estimate the parameters β, α, and ℓ from a set of aperture phases ϕ_i at known points in the panel. The uncertainty of these estimates is not important. Of more practical significance is the uncertainty of subsequent estimates of the phase at panel corners, that is, the standard error σ_c in the estimate ϕ_c obtained by equation 2.48 for (x_c, y_c) at a panel corner or another

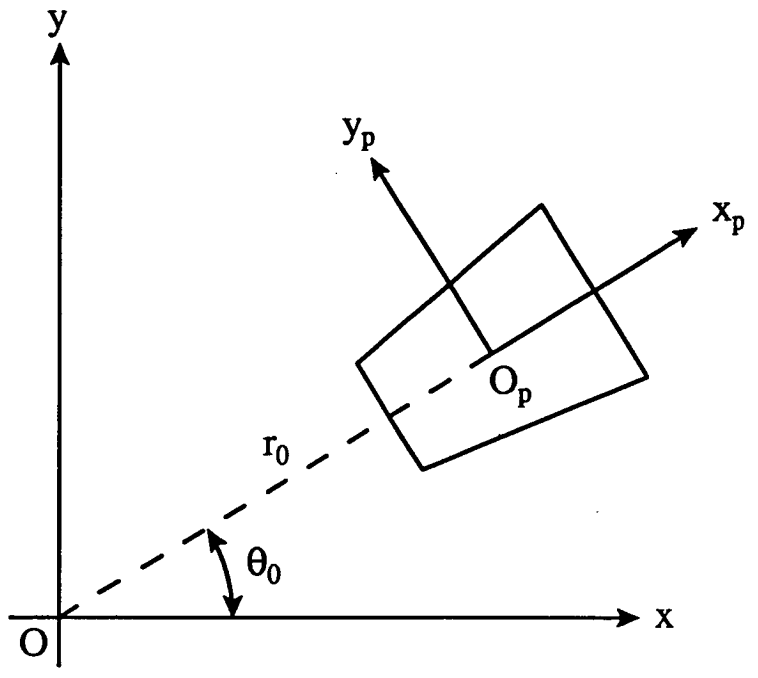


Figure 2.13: The panel-based coordinate system (x_p, y_p) and its relation to the aperture coordinate system (x, y).

adjustment point. These figures, converted to normal surface errors, are the data used to adjust the panels.

This is a textbook example of multivariate linear regression, for example see Guttman, et al. [26]. In the notation of appendix C, the following quantity must be computed:

$$\begin{aligned} A &= \frac{1}{S} + \frac{S_{vv}}{\Delta}(x_c - S_x)^2 + \frac{S_{uu}}{\Delta}(y_c - S_y)^2 \\ &\quad - \frac{2S_{uv}}{\Delta}(x_c - S_x)(y_c - S_y) \end{aligned}$$

where $\Delta = S_{uu}S_{vv} - S_{uv}^2$ is the matrix determinant. Then, if the standard errors of the data ϕ_i are all equal to σ, the standard error of ϕ_c is given by

$$\sigma_c = \sigma\sqrt{A}.$$

The errors in ϕ_i vary over the reflector, but each panel covers a small area so they are approximately constant over a panel.

It is interesting to consider a simple, rectangular panel over which the phase is sampled on a regular grid as shown in figure 2.14. The n^2 points are at integral spacings, and the lower left point is $(1, 1)$. The lower left panel corner is shown as $(1/2, 1/2)$, but depending on the registration of the grid

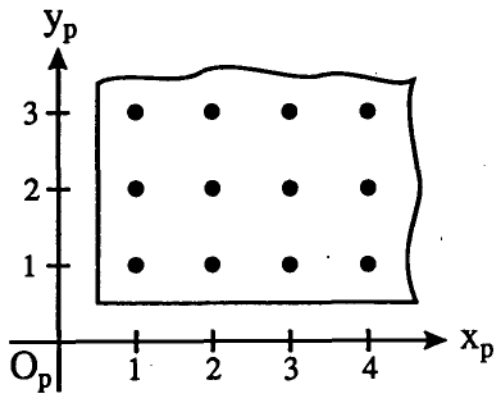


Figure 2.14: A rectangular panel sampled on a regular grid of n by n points at integral spacings.

n	$n^2A_{(0,0)}$	$n^2A_{(1/2,1/2)}$	$n^2A_{(1,1)}$
2	19.0	9.00	3.00
3	13.0	7.75	4.00
4	11.0	7.40	4.60
5	10.0	7.25	5.00
6	9.40	7.17	5.29
7	9.00	7.13	5.50
8	8.71	7.10	5.67

Table 2.2: Values of n^2A for several n and three positions of the panel's lower left corner.

with the panel it could as well be $(0,0)$ or $(1,1)$. Does this make a large difference to the uncertainty σ_c? The quantity A for the three cases can be shown to be

$$\begin{aligned} A_{(0,0)} &= \frac{1}{n^2}\left(1+6\frac{n+1}{n-1}\right) \\ A_{(1/2,1/2)} &= \frac{1}{n^2}\left(1+\frac{6n^2}{n^2-1}\right) \\ A_{(1,1)} &= \frac{1}{n^2}\left(1+6\frac{n-1}{n+1}\right) \end{aligned}$$

and in table 2.2 n^2A is given for several n. It can be seen that

$$\sigma_c = n\sqrt{A}\frac{\sigma}{n} \longrightarrow \sqrt{7}\frac{\sigma}{n} \text{ as } n \longrightarrow \infty. \tag{2.49}$$

A similar law with $n = \sqrt{P_p}$ can be expected to hold for trapezoidal panels with a less regular distribution of P_p samples. This is not the same as the intuitive result that when 3 independent parameters are estimated from P_p samples then $\sigma_c \simeq \sigma/\sqrt{P_p - 3}$. When 4 corners are allowed to vary independently, a reasonable estimate of σ_c is $\sqrt{4/3}$ times the result for 3 corners, although this is not justified here. Certainly, it is expected that $\sigma_c \propto \sigma/\sqrt{P_p}$ for large P_p.

2.4.2 Panel Deformations

Many modern large antennas, particularly for radio astronomy, are expected to operate at millimetre wavelengths. Panel position estimates which assume undistorted panels are often inadequate at this level of precision. At the very least, a formulation which allows the four corners of a trapezoidal panel to move independently is desirable. To do this, some assumptions are needed about the way a panel distorts.

Von Hoerner [86] performed an investigation of three common panel constructions, and showed that a simple model for panel distortions was adequate for all. The rms deviation from the model was less than two percent of the corner movement causing the twist. His relationship for the change in panel profile $z(x_p, y_p)$ due to a twist was

$$z(x_p, y_p) \propto x_p y_p$$

using a panel-based rectangular coordinate system where the origin is stationary. He developed this into an equation for the panel profile after a

general twist, tilt and vertical lift, and gave a least-squares fitting procedure for the equation parameters given measurements of the panel height.

A limitation of the method he described is that it assumes a grid of sample points symmetrical about the y_p axis, which for Microwave Holography data on a rectangular aperture grid will almost never be true. Here, a different equation is given which can be fitted to data on a general grid. In addition, the four free parameters of the equation are directly related to the phase errors at the four panel corners. The derivation begins with the corner errors, and proceeds by making physical assumptions about the distortions that are consistent with von Hoerner's.

A diagram of a single trapezoidal panel is shown in figure 2.15. Let it have height h and widths a and b, and let $w = (a+b)/2$ and $d = (b-a)/2$, the average width and half width difference. The corners are numbered, and the aperture phase at each corner is defined as follows:

$$\begin{aligned}
\phi(x_b, y_1) = \phi_1 \quad &\text{at corner 1 :} \quad (x_b, y_1) = (+h/2, -b/2) \\
\phi(x_b, y_2) = \phi_2 \quad &\text{at corner 2 :} \quad (x_b, y_2) = (+h/2, +b/2) \\
\phi(x_a, y_3) = \phi_3 \quad &\text{at corner 3 :} \quad (x_a, y_3) = (-h/2, -a/2) \\
\phi(x_a, y_4) = \phi_4 \quad &\text{at corner 4 :} \quad (x_a, y_4) = (-h/2, +a/2).
\end{aligned}$$

The phase $\phi(x_p, y_p)$ is desired at any point in the panel. Define (x_a, y_a) and (x_b, y_b) to be the points where the radial line from the reflector centre through (x_p, y_p) meets the trapezium. This means for $a \neq b$

$$\frac{y_b - y_a}{b-a} = \frac{y_a}{a} = \frac{y_b}{b} \tag{2.50}$$

and so

$$y_a + y_b = \frac{w}{d}(y_b - y_a). \tag{2.51}$$

Interpolating between these intercepts by similar triangles,

$$\frac{y_b - y_p}{x_b - x_p} = \frac{y_p - y_a}{x_p - x_a}$$

which rearranges using equation 2.51 to give

$$y_p = \frac{x_p}{h}(y_b - y_a) + \frac{1}{2}(y_a + y_b) = \left(\frac{x_p}{h} + \frac{w}{2d}\right)(y_b - y_a). \tag{2.52}$$

Assume that the phase ϕ remains linear along lines of constant x_p when the panel is twisted. Thus the panel is rigid along lines parallel to sides a

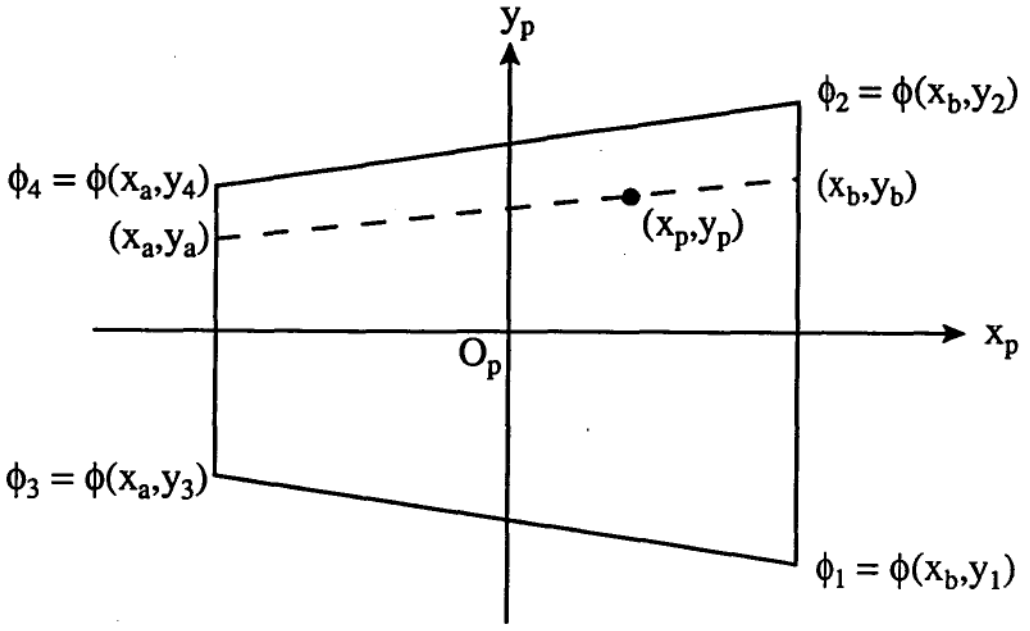


Figure 2.15: A trapezoidal panel and the construction used to find the aperture phase function when corners 1–4 have arbitrary heights.

and b. The phases at the radial intercepts can be found by similar triangles above these sides:

$$\frac{\phi_4 - \phi_a}{y_4 - y_a} = \frac{\phi_a - \phi_3}{y_a - y_3}$$

and

$$\frac{\phi_2 - \phi_b}{y_2 - y_b} = \frac{\phi_b - \phi_1}{y_b - y_1}.$$

These simplify to the pair of equations

$$\begin{aligned} \phi_a &= \frac{y_a}{a}(\phi_4 - \phi_3) + \frac{1}{2}(\phi_3 + \phi_4) \\ \phi_b &= \frac{y_b}{b}(\phi_2 - \phi_1) + \frac{1}{2}(\phi_1 + \phi_2) \end{aligned}$$

and so, using equation 2.50,

$$\begin{aligned} \phi_b - \phi_a &= \frac{y_b - y_a}{2d}(\phi_2 - \phi_1 - \phi_4 + \phi_3) + \frac{1}{2}(\phi_1 + \phi_2 - \phi_3 - \phi_4) \\ \phi_a + \phi_b &= \frac{y_b - y_a}{2d}(\phi_2 - \phi_1 + \phi_4 - \phi_3) + \frac{1}{2}(\phi_1 + \phi_2 + \phi_3 + \phi_4). \end{aligned}$$

At this stage the distortion parameters are defined:

$$4\tilde{\tau} = -\phi_1 + \phi_2 + \phi_3 - \phi_4 \quad (2.53)$$

$$4\beta = +\phi_1 + \phi_2 - \phi_3 - \phi_4 \quad (2.54)$$

$$4\tilde{\alpha} = -\phi_1 + \phi_2 - \phi_3 + \phi_4 \quad (2.55)$$

$$4\ell = +\phi_1 + \phi_2 + \phi_3 + \phi_4. \quad (2.56)$$

It is intuitive that these parameters will be sensitive to a twist, radial tilt, tangential tilt, and vertical lift respectively. The symbols are chosen to agree with von Hoerner's usage. Thus,

$$\phi_b - \phi_a = \frac{2\tilde{\tau}}{d}(y_b - y_a) + 2\beta$$
$$\phi_a + \phi_b = \frac{2\tilde{\alpha}}{d}(y_b - y_a) + 2\ell.$$

Now assume that ϕ is linear along radial lines such as that through (x_p, y_p). This restricts the panel shape enough to define the equation for ϕ at any point. Using similar triangles above the radial line,

$$\frac{\phi_b - \phi}{x_b - x_p} = \frac{\phi - \phi_a}{x_p - x_a}$$

which gives

$$\begin{aligned}
\phi &= \frac{x_p}{h}(\phi_b - \phi_a) + \frac{1}{2}(\phi_a + \phi_b) \\
&= \frac{x_p}{h}\left[\frac{2\tilde{\tau}}{d}(y_b - y_a) + 2\beta\right] + \frac{1}{2}\left[\frac{2\tilde{\alpha}}{d}(y_b - y_a) + 2\ell\right] \\
&= \tilde{\tau}\,\frac{x_p}{h/2}\,\frac{y_p}{w/2 + x_p d/h} + \beta\,\frac{x_p}{h/2} + \tilde{\alpha}\,\frac{y_p}{w/2 + x_p d/h} + \ell \qquad (2.57)
\end{aligned}$$

after substituting for $(y_b - y_a)$ using equation 2.52. This should be compared with equation 2.48. For a rectangular panel, $d = 0$ and it is an obvious generalization including von Hoerner's cross-term $\tilde{\tau} x_p y_p$ suitably normalized. For a non-rectangular panel, the coordinate y_p is modified to depend slightly on x_p (d/h is small). A least-squares fitting procedure is given in appendix C which estimates $\tilde{\tau}$, β, $\tilde{\alpha}$, and ℓ for data on an arbitrary set of coordinates.

There is an advantage of von Hoerner's equation over 2.57: his distortion parameters are 'decoupled', while here $\tilde{\alpha}$ contains a twist and β contains a vertical shift whenever $a \neq b$. For example, set $\phi_1 = \phi_3 = -\delta$ and $\phi_2 = \phi_4 = +\delta$, giving $\tilde{\alpha} = \delta$ by equation 2.55 and all other parameters zero. If $a \neq b$ these sides clearly have different slopes and so the panel has a twist. A similar interaction of β and ℓ is not of concern, but it may be that twisting is to be suppressed in some situations, which cannot be done by simply setting $\tilde{\tau} = 0$ and using the same equation.

A common-sense modification to equations 2.53 and 2.55 is to define

$$\frac{4\tau}{w/2} = -\frac{\phi_1}{y_1} + \frac{\phi_2}{y_2} + \frac{\phi_3}{y_3} - \frac{\phi_4}{y_4} \qquad (2.58)$$
$$\frac{4\alpha}{w/2} = -\frac{\phi_1}{y_1} + \frac{\phi_2}{y_2} - \frac{\phi_3}{y_3} + \frac{\phi_4}{y_4}. \qquad (2.59)$$

Then there is truly no twist when $\tau = 0$. It can be shown that

$$\left.\begin{aligned} \tilde{\tau} &= \tau + \frac{d}{w}\alpha \\ \tilde{\alpha} &= \alpha + \frac{d}{w}\tau \end{aligned}\right\} \text{ or } \left\{\begin{aligned} \frac{ab}{w^2}\tau &= \tilde{\tau} - \frac{d}{w}\tilde{\alpha} \\ \frac{ab}{w^2}\alpha &= \tilde{\alpha} - \frac{d}{w}\tilde{\tau} \end{aligned}\right. \tag{2.60}$$

Thus τ and α can be calculated easily once a least-squares fit has provided $\tilde{\tau}$ and $\tilde{\alpha}$. Moreover, substitution into equation 2.57 shows that the equation for undeformed panels 2.48 is recovered exactly when $\tau = 0$; in fact

$$\phi = \tau \frac{x_p + hd/2w}{h/2 + x_p d/w} \frac{y_p}{w/2} + \beta \frac{x_p}{h/2} + \alpha \frac{y_p}{w/2} + \ell. \tag{2.61}$$

The formulation for twisted panels is completely consistent with the simpler case, it has 'decoupled' twist and tilt parameters, and all its parameters are logically related to the corner phase errors.

Chapter 3

Holography Experiments at Mount Pleasant

Several holographic surveys of the Mount Pleasant 26 m antenna were made during the course of this research. This chapter describes the last and most accurate survey, undertaken in April 1991 using both carrier and beacon signals from the AUSSAT-1 satellite. An illumination-weighted accuracy of ± 53 μm of normal surface error was obtained with the carrier signal. Data taking and reduction procedures had developed sufficiently by this stage to form a model for future experiments. The results were used to adjust the position of most reflector panels, and detailed information on large-scale structural deformations was obtained.

3.1 Experimental Equipment

3.1.1 Design Philosophy

The holographic measurement system used at Mount Pleasant was developed with portability in mind. At the time of writing a compact unit based on this experience is under construction. A major factor in the design has been the availability of monochromatic signals from the AUSSAT satellites; there are three satellites currently in geostationary Earth orbit.

The down-link band is 12.25–75 GHz, the 'Ku-band', whose wavelength of about 24 mm is suitable for unambiguous holographic measurements of reflectors intended for cm wavelengths. The advantages of using these signals are the potential for very high signal-to-noise ratios as the bandwidth is narrowed, the low expense of front-end units designed specifically for the 'Ku-

band', and the feasibility of direct phase measurements at these frequencies. A significant disadvantage is that the antenna can only be measured at the azimuth and elevation of the satellite. It is not possible to study deformations of the surface due to the changing effects of gravity at different antenna attitudes.

The alternative is to use cosmic point sources of high intensity. These can be either thermal sources observed at a frequency where they are unresolved by the antenna, or strong maser sources which are discovered from time to time [5, 51]. If several are available the antenna can be measured in a variety of positions, although each position will be an average because a map takes several hours to record. This would give an indication of the magnitude and nature of gravitational deformations. Unfortunately, a radio system with a wide bandwidth is needed to achieve sufficient signal-to-noise with a cosmic source, whereas a narrow bandwidth is desirable for monochromatic satellite-borne sources.

It was decided that the simplicity of a narrow-band system, and the availability and portability of front-end units, outweighed the restriction on source positions. In addition, any number of commercial antennas are available for collecting Ku-band down-link signals, ideal for a phase reference antenna. A 2.1 m diameter *Andrew* model was purchased.

3.1.2 Front-End and Portability

Originally, the same model low-noise block down-converter (LNB) was used for both antennas: a *MASPRO LNB-SCAU-500* with a noise temperature of 180 K. Later, after an unfortunate accident involving the Sun, the 26 m antenna was fitted with a *MASPRO LNC-SCF-778* having a noise temperature of 130 K. These commercial units had built-in local oscillators of frequency 11.30 GHz which were free-running, and could not be locked to a time standard for phase stability. The units were adapted by allowing for an external l.o. signal to be fed to the mixer. Close to each LNB was placed a *MITEQ* phase-locked oscillator which could be tuned over the range 11.00–50 GHz. It was used only in the range 11.29–39 GHz, never far from the design frequency of the mixer, and no significant changes in signal level were noticed as the oscillator was adjusted. The ability to change this frequency allowed more flexibility in later intermediate-frequency (i.f.) stages.

The combined LNB and oscillator weighs just a few kilograms, including a feed horn. It should not be difficult to mount this in any antenna which

has the facility to change feeds; in some cases an existing feed horn may be preferred. The other hardware which must be improvised at each new site is a solid mount for the 2 m reference antenna: once it is pointing at the satellite it does not need to be moved again. At Mount Pleasant the mounting bracket rotated about the top of a $3\frac{1}{2}$" diameter vertical pipe, welded to the foot of the main antenna. Providing a similar pipe should not be a problem elsewhere.

The portable i.f. and sampling system, when construction is complete, will operate independently at any site except for two things. Firstly, a frequency reference is needed for locking all the local oscillators in the system so that signal phase is preserved. Secondly, there must be a means of scanning the antenna in either axis with accurately known limits and rates, or else with an accurate record of the positions passed as a function of time. This facility will generally be available for large antennas, but an interface with the control system may require some special software.

3.1.3 Using AUSSAT Satellites for Holography

There are three satellites in geostationary orbit at the following longitudes:

AUSSAT-1	160°E
AUSSAT-2	156°E
AUSSAT-3	164°E.

Each satellite has several 'beams' for illuminating different parts of Australia and the South-West Pacific. To each beam several transponders can be connected: there are eleven 12 W transponders for the lower part of the Ku-band, and four 30 W transponders for the high frequency channels. The beam appropriate for Tasmania is the 'SE' beam, which has a vertical polarization of the E-vector. The 12 W and 30 W transponders provide about 42 dBW and 46 dBW equivalent isotropic radiated power at Hobart.

Phase measurements could not be made with a modulated down-link signal, so AUSSAT Pty. Ltd. supported this research by providing a full-strength carrier with no modulation on several occasions. The 30 W 12725 MHz transponder on AUSSAT-1 was used. This enabled very high-quality surface maps to be made before the surface was adjusted.

In addition to the down-link channels there are some unmodulated beacons, 20 dB less powerful, used for other purposes. Each satellite has a pair of beacons at frequencies between 12748 MHz and 12750 MHz. At both

12373 MHz and 12565 MHz AUSSAT-1 has two beacons separated by about 100 kHz. The strongest is at the lower frequency, and this beacon was used several times for holography measurements, giving maps whose quality compared well with the carrier maps.

AUSSAT-1 appears from Hobart at an azimuth of $18.15 \pm 0.04°$ and an elevation of $39.01 \pm 0.02°$, with a range of 37850 ± 10 km. The periodic variation in position is due to an imperfect orbit. It must be taken into account because the 26 m antenna's beamwidth at this frequency is 0.05°. The Ground Station staff were able to send position predictions covering each period of observation: these were used by the observing software at Mount Pleasant. Predictions spaced at hourly intervals could be interpolated to an accuracy of $\pm 0.001°$, which is of the same order as the antenna's pointing accuracy.

3.1.4 Down-Conversion, Amplification, and Filtering

The i.f. system used at Mount Pleasant had three stages of down-conversion to a final frequency of 10.7 MHz. It is shown in figure 3.1. The frequencies indicated are for the 12725 MHz carrier. The upper sideband was used at each conversion, so the 10.7 MHz phase had the same sense as the original phase.

It was essential to use phase-locked local oscillators throughout for two reasons. Firstly, the signal frequency had to be stable within the 30 kHz pass-band, which was ensured by locking to an accurate frequency standard. Secondly, the phase relationship between the signals in channels A and R had to be preserved. This meant locking the two front-end *MITEQ* oscillators to the same reference signal, provided by a frequency synthesizer itself locked to the frequency standard. The oscillators' phase loops multiplied by 108, so the reference signal of 105.5 MHz had to be carried to the two front-end units with a high degree of phase stability. It was found that sufficient stability was obtained without a feed-back system.

Different gains were required for the carrier signals and the weaker beacon signals. Only the latter required 20 dB amplifiers to follow the 410 MHz band-pass filters. The variable 1 dB step attenuators were used to adjust the final output levels to roughly 1 volt maximum, avoiding sampler overflow while giving ample dynamic range.

At 10.7 MHz the i.f. signals were filtered to a bandwidth of 30 kHz. These 0.28% band-pass filters were cheap, commercial units made for f.m.

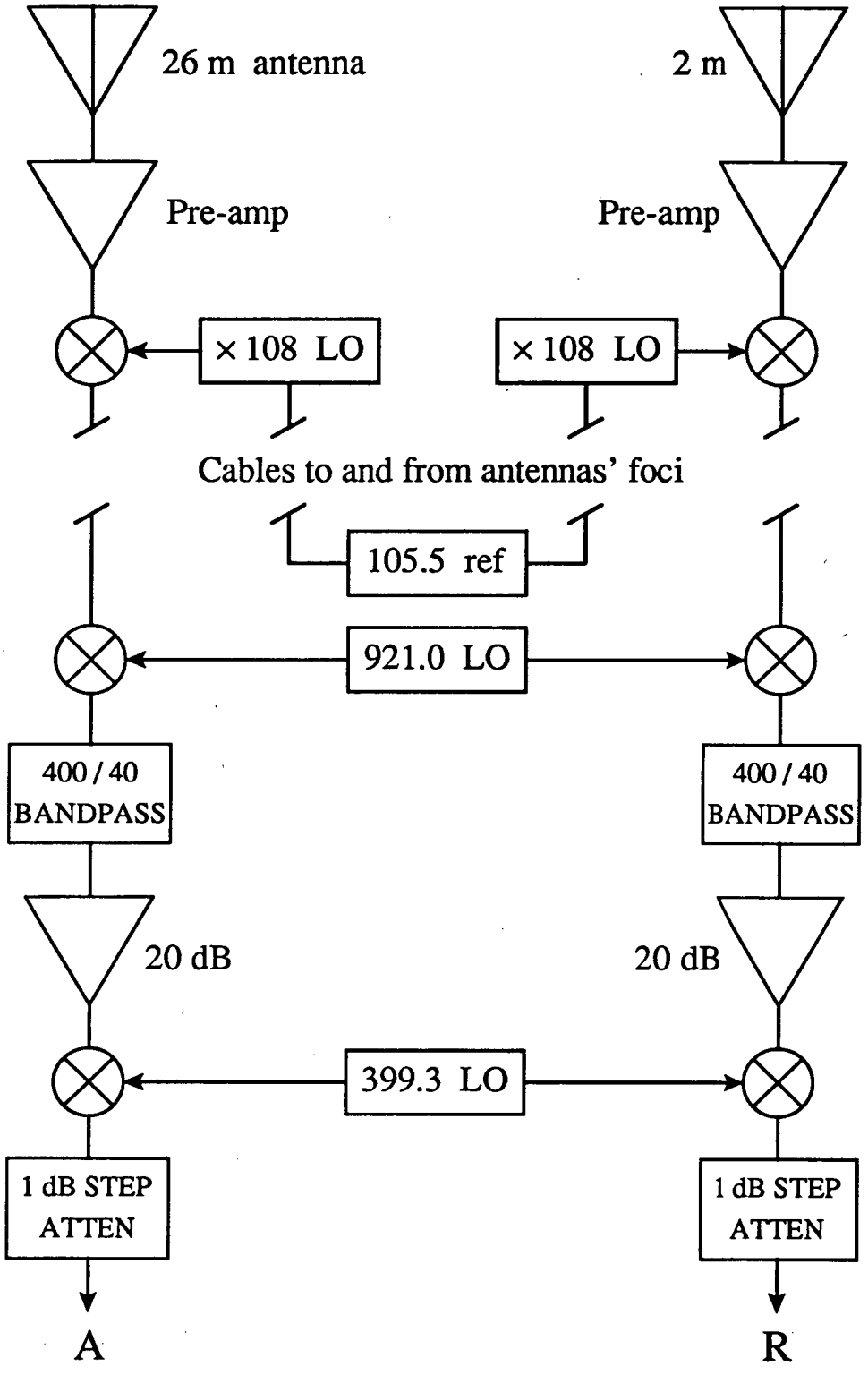


Figure 3.1: The front-end and frequency-conversion system used for holographic measurements at Mount Pleasant. Frequencies are indicated for a carrier signal at 12725 MHz, although the 20 dB amplifiers were necessary only for the weaker beacon signal.

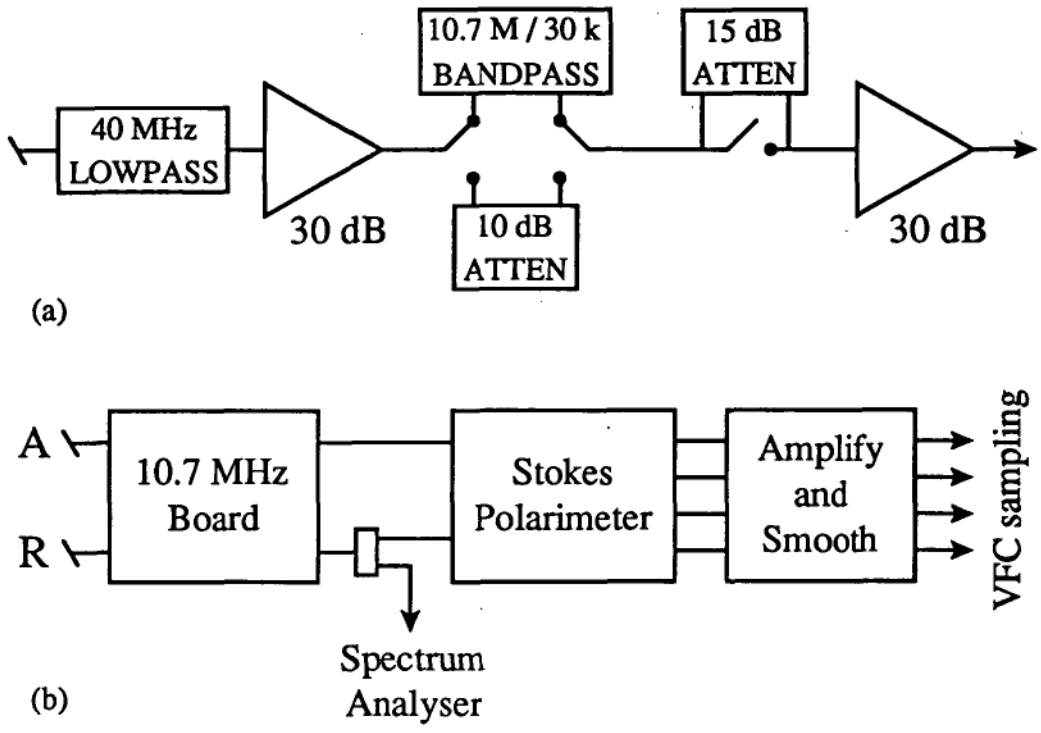


Figure 3.2: The filtering and detecting system: (a) one channel of the 10.7 MHz board, and (b) the polarimeter and sampling arrangements.

radio. Figure 3.2(a) shows one channel of the board incorporating the filters. The pass-band attenuation of about 10 dB was mimicked by attenuators so that the signals could be switched past the filters with no change in output level. This was convenient for locating the beacons and checking that there was no interference nearby. In addition, an extra 15 dB of attenuation could be switched by computer into the 26 m channel when close to the on-source position. This enabled a much greater gain to be used without causing saturation and sampler overflow, giving increased sensitivity in the far-out sidelobe regions of the antenna pattern.

3.1.5 Measurement of Phase

The method of phase measurement is the core of the signal processing system, after such efforts have been made to ensure that phase is preserved through the i.f. stages. A differential phase must be measured between the A and R signals. As the system was developed attempts were made to measure the phase directly, so that it could be displayed and recorded in degrees as data were taken. These methods were not sensitive enough, and simple

phase shift and multiplication was more effective, producing sine and cosine outputs from which the phase was deduced later. Here is a description of each method.

A Vector Voltmeter

The Hewlett-Packard 8405A Vector Voltmeter displays the relative phase of a signal with respect to any reference that it can lock onto. A voltage proportional to the phase is available. Unfortunately it was not possible to lock onto the weak beacons received by the reference antenna: at this early stage the i.f. system had much less gain than the design above. In any case, it was intended eventually to build a compact unit which did not rely on a bulky piece of equipment.

A Rectangular Phase Detector

A phase detector circuit based on a design from the Hewlett-Packard 8754A Network Analyzer was built. It worked by amplifying and limiting two 1 MHz i.f. signals, producing square waves of the same phase. The rising and falling edges of these waves triggered digital circuitry which produced a new waveform, having a mark/space ratio proportional to the phase lag or lead of one square wave behind or ahead of the other. The area under this waveform was averaged to give a voltage proportional to the phase difference. This circuit gave an adequate phase measurement for several sidelobes beyond the main beam, but again it was not sensitive enough for a map of any useful size.

A Polarimeter

It is a common technique to multiply two signals in phase and quadrature, then apply a low-pass filter, to obtain outputs proportional to the cosine and sine of the signals' phase difference. This is how the U and V outputs of a Stokes polarimeter are produced for polarization measurements in radio astronomy. Indeed, after some tests to prove the feasibility of the method, an existing Stokes polarimeter was modified to accept the holography i.f. signals, as seen in figure 3.2(b). As a bonus, the I_x and I_y outputs gave the total power in the main and reference channels. I_x was not used to measure the far field power, however, because it was proportional to the square of the 26 m antenna output amplitude, while the observable $\mathcal{A}$ obtained from U

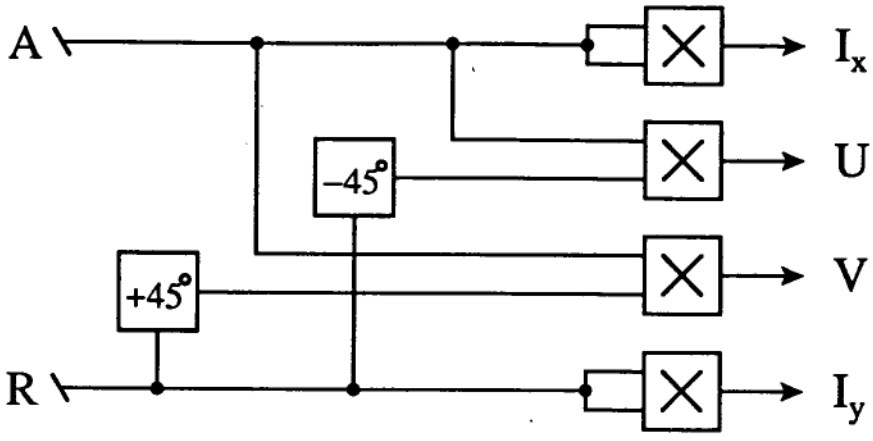


Figure 3.3: The Stokes polarimeter used for phase determination.

and V was directly propotional to the amplitude — see equation 3.1 below. Thus $\mathcal{A}$ is much more sensitive to low-level sidelobe structure than I_x.

Figure 3.3 is a diagram of the polarimeter. Phase shifts were implemented with simple RC filters, perfectly adequate for this narrow-band application. Conventional quadrature multiplication would have required a 90° phase shift achieved by using the filter in its stop-band. To avoid the consequent large attenuation, the phase difference was achieved with matching low-pass and high-pass filters operating close to their common turnover frequecy. With exact component values there would be a 45° phase lag in the low-pass case and a 45° phase advance in the high-pass case. Because the components' values were only approximately correct, the total phase difference was less than 90°; this was taken into account in the calibrations described below. The circuit was designed as part of a 50 Ω system, and the attenuation expected for each filter was 5 dB.

The four outputs I_x, U, V, and I_y were sampled by integration using voltage-frequency converters. Prior to this they were smoothed with a time constant somewhat larger than the sampling interval of 25 ms. Many successive samples were averaged at each point of the grid, giving effective integration times of up to 2.5 s.

Calibrating the Polarimeter

To quantify the operation of the polarimeter, consider the model for the cross-products shown in figure 3.4. Phase shifts of $-\alpha$ and $+\beta$ are applied to the reference signal $r(t)$, after which the products are formed with $a(t)$.

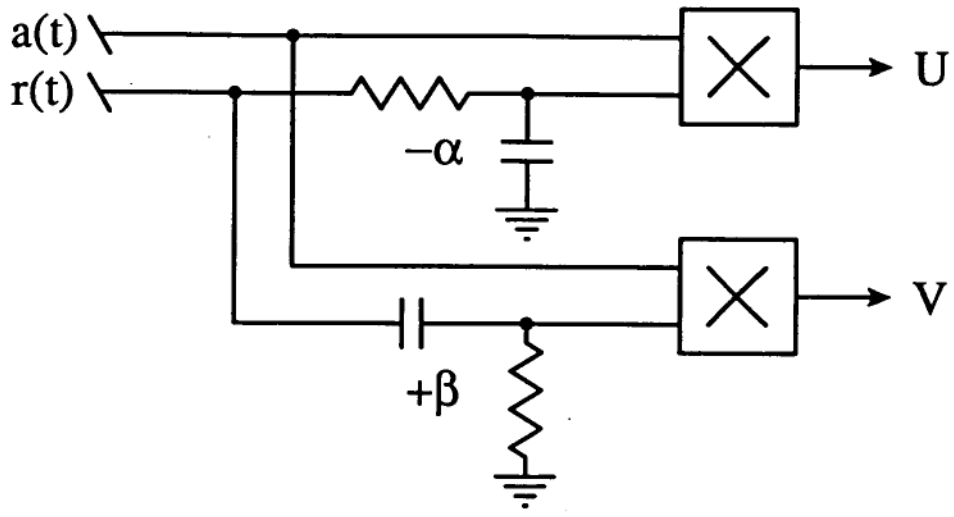


Figure 3.4: A model for the cross-product part of the polarimeter.

The voltages U and V result after smoothing to remove all high frequency components of the products.

With a monochromatic source and suitable filtering of unwanted noise, the input waveforms can be assumed to be sinusoidal with a phase difference ϕ:

$$\begin{aligned} a(t) &= A\cos(\omega t + \phi) \\ r(t) &= R\cos(\omega t). \end{aligned}$$

If the combined effect of each multiplier and preceeding filter is to give a gain G_α in one case and G_β in the other, the smoothed outputs are

$$\begin{aligned} U &= \tfrac{1}{2}G_\alpha AR\cos(\phi + \alpha) \\ V &= \tfrac{1}{2}G_\beta AR\cos(\phi - \beta). \end{aligned}$$

It is as useful to measure $\phi - \beta$ as ϕ, since it is the relative phase at different grid points that is important. A short manipulation gives

$$\tfrac{1}{2}G_\beta AR\sin(\phi - \beta) = \frac{V\cos(\alpha + \beta) - U(G_\beta/G_\alpha)}{\sin(\alpha + \beta)}$$

which requires for its computation only that the gain ratio G_β/G_α and the total phase shift $\alpha + \beta$ are known. The 'observable' $\mathcal{A}$ obtained at each grid point has this for its imaginary part and V for its real part, giving

$$\mathcal{A} = \tfrac{1}{2}G_\beta AR\exp j(\phi - \beta). \tag{3.1}$$

The gain ratio and phase shift were measured using this property: if $a(t)$ and $r(t)$ are two synthesized waves of very similar frequency, their relative

phase will change at a constant rate. Then the outputs U and V will vary sinusoidally in time, with a phase difference $\alpha + \beta$. Under these conditions U and V were recorded at precise time intervals for several cycles. Sinusoids were fitted to these data using the least-squares method described in appendix C, and the gain ratio and phase shift were deduced from the amplitude and phase of the fits. The 10.7 MHz board was included in the setup, so the value of the A channel attenuator could be measured by switching it in and observing the change in amplitude. The values obtained for these parameters were:

$$\begin{aligned} \alpha + \beta &= 75.9 \pm 0.9° \\ G_\beta / G_\alpha &= 0.821 \pm 0.012 \\ \text{sw. att.} &= 15.29 \pm 0.13 \text{ dB}. \end{aligned}$$

3.2 Data Collection and Reduction

When the antenna is steered about the source the quantity $\mathcal{A}$ varies according to the far field pattern, and if there were no other variations the normalized antenna pattern would be $\mathcal{A}(u,v)/\mathcal{A}(0,0)$. This section discusses some systematic effects which occur during data taking, and the way they are compensated in the data reduction.

3.2.1 Data Collection

Microwave Holography data should ideally be collected at a rectangular grid of source positions (u,v) with respect to the antenna aperture, as discussed in detail in section 2.3. The practicality of doing this depends upon the dynamical properties of the antenna steering system. Many systems for radio telescope antennas have a high damping constant to avoid unwanted oscillations about the source position, which means that motion between two nearby sources is very slow. A large map takes far too long to collect in these circumstances.

Instead, a map can be built from many parallel scans in one axis direction or the other, with samples taken 'on the fly'. To minimize sample position errors it is best to scan using the axis attached to the reflector rather than the fixed axis: that is, Y for an XY mount or elevation for an AZEL mount. Referring to figure 2.10 it is seen that these scans are straight lines in (u,v) because motion is towards or away from a pole, while scans in the other direc-

tion would be curved. Samples taken while moving are smeared over a region surrounding the nominal sample position, but this has a completely recoverable effect on the aperture currents. If integration occurs over a fraction f of the sample interval, the measured quantity is not $A(u)$ but

$$A'(u) = A(u) * \frac{1}{f\delta_u} \mathrm{II}\left(\frac{u}{f\delta_u}\right),$$

using a one-dimensional notation. In the transform domain, this replaces $a(x)$ by

$$a'(x) = a(x)\frac{\sin(\pi f \delta_u x)}{\pi f \delta_u x}. \tag{3.2}$$

The phase is unchanged, and the amplitude can be corrected by dividing by this sinc function along the direction of scanning.

Interrupt-controlled software is needed to take samples at the correct times, and there are two possible sources for interrupt signals. Either the antenna is driven at an accurate rate and samples are triggered at uniform intervals of time, or else the antenna control system produces a triggering signal when the antenna passes preset angular positions. The first system is preferable for these reasons:

1. When the control system allows, the antenna can be scanned on both axes simultaneously with the same time interrupts, allowing straight paths in (u, v) to be followed more closely near a pole;

2. The control system will have a finite cycle time in its servo loop which limits the accuracy of the interrupts it produces.

In this case it is important that the antenna is steered accurately. A wise measure would be to read its position on every sample interrupt. It was to avoid this complication that control system interrupts were used at Mount Pleasant, where the 0.1 s cycle time is adequate.

From the considerations in chapter 6 a map size of 53×53 samples was chosen for the 26 m antenna measurements. The sampling interval was $\delta_u = \delta_v = 0.00075$ or 0.043°, resulting in aperture resolutions of 0.593 m at 12725 MHz and 0.610 m at 12373 MHz — just over 5 sample points per panel on average. With integration times of up to 2.5 s at a scanning rate of 0.75° per minute, smearing was by as much as 73%. The desired (u, v) end-points of each scan were converted to (X_1, Y_1) and (X_2, Y_2) as discussed in section 2.3.1, and the scan was performed in Y from Y_1 to Y_2 while holding X constant at $(X_1 + X_2)/2$. The maximum deviation of the (u, v) points

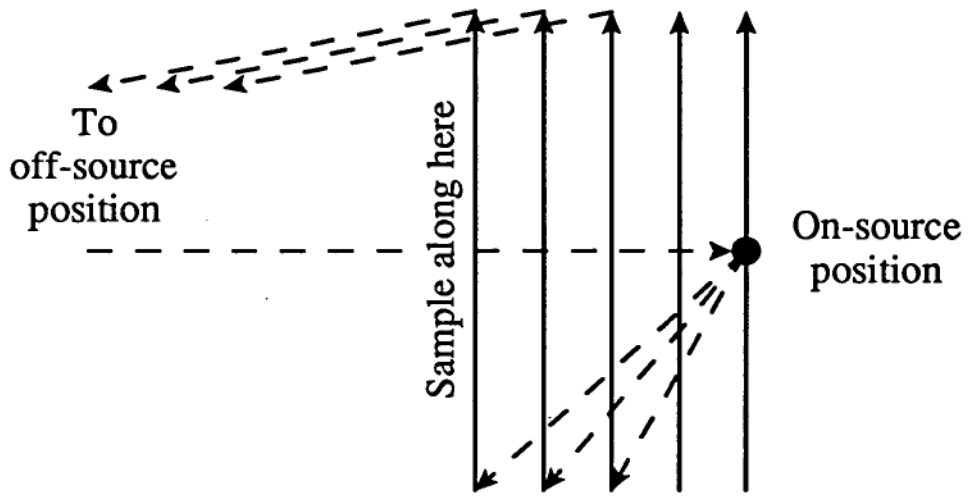


Figure 3.5: The antenna scanning pattern, including on-source and off-source calibration points. Sampling occurs along the solid arrows.

traversed in this way from a rectangular grid was 6.8×10^{-6} or 0.9% of the sample interval. This is not significant: the resolution of the antenna's pointing system is 2.9" or 1.9% of the sample interval.

Variations in instrumental gain and phase and satellite signal strength can appear as false features in the map unless they are measured and removed. To monitor them an on-source calibration sample was taken before every scan, as well as a sample some degrees off-source to monitor the zero levels. The scanning pattern including calibration points is shown in figure 3.5. The satellite position was recomputed prior to the on-source samples using the data provided by AUSSAT; this ensured that the satellite was centred in the main beam and that the scan end-point calculations were up to date. To aid in interpolating the satellite position and calibration signal applicable to each sample point, the following times were recorded:

$$J_{\mathrm{sat}} = \text{Julian Date of satellite position prediction} \quad (3.3)$$

$$J_{\mathrm{cal}} = \text{Julian Date of on-source calibration point} \quad (3.4)$$

$$J_{\mathrm{scn}} = \text{Julian Date of first sample in the scan.} \quad (3.5)$$

Julian Date was the most convenient form to keep the time in: it is continuous across changes in minute, hour, day, month, and year.

All of this information relating to each scan — the calibration samples, the satellite position, the times, the scan end-points — was stored in a small header in front of the scan data. This was separate to the header which

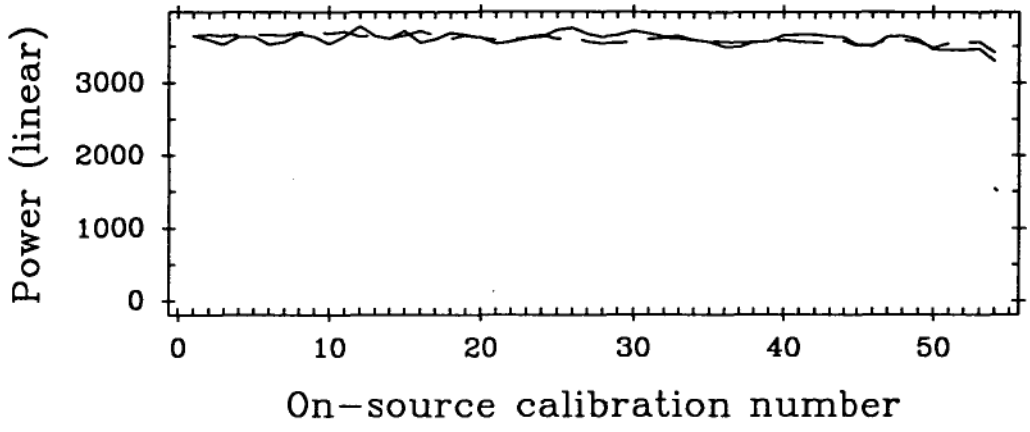


Figure 3.6: Signal power estimated from on-source $|\mathcal{A}|$ (solid) and I_x (dashed).

contained characteristics of the whole map, such as antenna name, frequency, interrupt method, and the number and length of scans. Although the format of the headers will always be arbitrary, the information they must contain is common to all experiments. For this reason the header format used at Mount Pleasant is given in appendix D. It is suitable for describing data taken with control system or time interrupts.

Before proceeding with any data analysis, the calibration samples were checked for noisy or discontinuous behaviour. One useful test was a graph of the 26 m power I_x and the magnitude of $\mathcal{A}$ derived from U and V for the on-source samples. Because $I_x \propto A^2$ and by equation 3.1 $|\mathcal{A}| \propto AR$, both are proportional to the signal power that determines amplitudes A and R. If they differed a change of gain of the 26 m receiver relative to the 2 m was indicated. As figure 3.6 shows for a typical experiment, no significant change occurred.

It is also interesting to graph the phase of $\mathcal{A}$. A variation in the on-source phase is expected because the main and reference antennas form a two element interferometer which can resolve motion of the satellite. This phase term is given by equation 2.43 where the predicted satellite position is $\hat{\mathbf{s}}$. The actual and expected phase variations are shown in figure 3.7 for two separate experiments. In the upper graph the trend is predicted quite well; in the lower graph there are large deviations which indicate another systematic effect. This is likely to be due to the lengthy cables which are required to reach the antennas' foci from the control room: about 80 m for

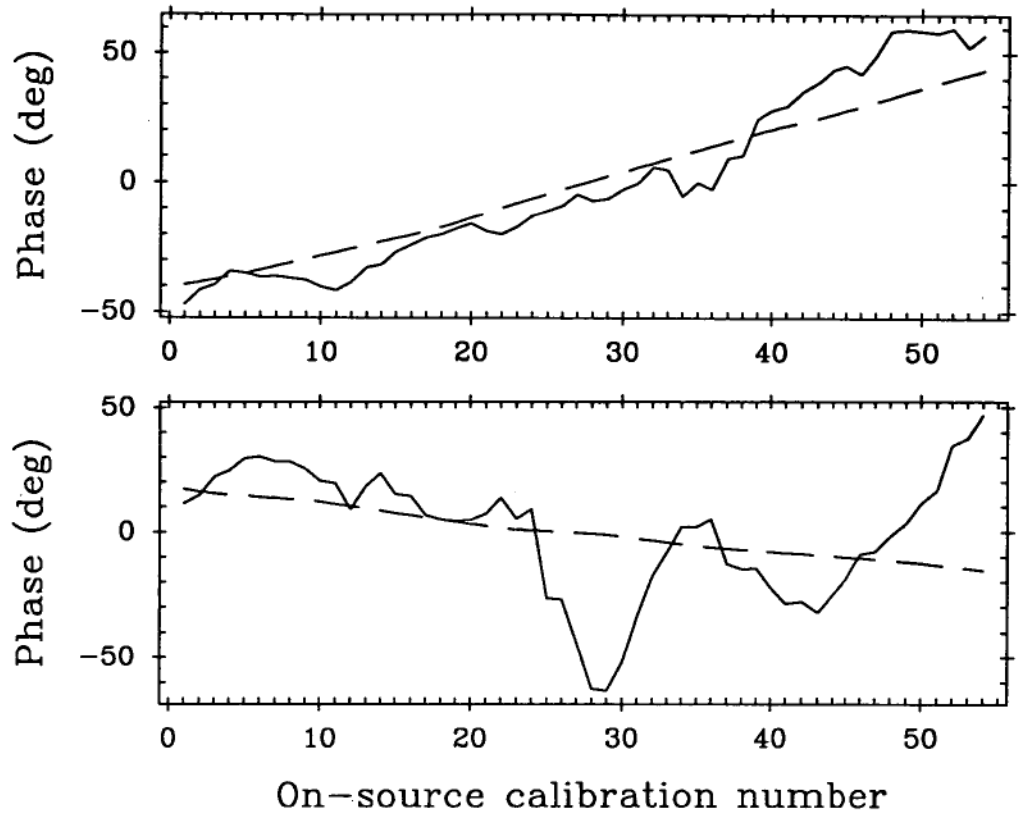


Figure 3.7: On-source calibration phase (solid) and predicted interferometer phase due to satellite motion (dashed) for two experiments.

the 26 m antenna. Changes in their temperature and position will affect their electrical length. By interpolating the on-source phase between calibration samples these variations can be compensated at each sample position, leaving a residual phase error of less than $\pm 10°$.

3.2.2 Data Reduction

The raw data for a $N \times N$ map consists of the following information:

$$N \text{ scans having } \begin{cases} N \text{ data samples } U \text{ and } V \\ + \text{ antenna positions at each end} \\ + \text{ satellite positions at each end} \\ + \text{ calibration samples at each end} \\ + \text{ times when each quantity is correct.} \end{cases}$$

The data reduction procedure to produce a map of aperture currents is best described in point form. The description assumes that an XY antenna has

been measured with a satellite source but the methods are universal.

1. Use a linear interpolation between the scan end-points to find the antenna position (X, Y) for each sample.

2. Similarly, interpolate using J_{sat} and J_{scn} to find the corresponding satellite position $\hat{\mathbf{s}}$.

3. Compute (u, v) for each sample using equation 2.36 with $\mathbf{r} = \hat{\mathbf{s}}$.

4. Compute $\mathcal{A}$ from U and V using equation 3.1 (remember from section 3.1.5 that $\mathcal{A}$ is more sensitive than I_x to low-level structure).

5. Interpolate linearly using J_{cal} and J_{scn} to estimate the on-source signal at the time of each sample, and deduce $\mathcal{A}_0$ at the grid centre.

6. Compensate for on-source amplitude and phase variations (which incorporate satellite motion) by finding

$$A(u, v) = \mathcal{A}/\mathcal{A}_0.$$

7. Find the phase correction for antenna motion ϕ_a using equation 2.44, and correct

$$A(u, v) \longrightarrow A(u, v) \exp(-j\phi_a).$$

8. Compile the scans into a full map, interpreting them as rows or columns according to the scanning coordinate and reversing them if necessary. Compile the (u, v) positions similarly.

9. Perform a least squares fit for the origin and sampling intervals of closest rectangular (u, v) grid. No re-gridding is necessary if the error is smaller than the likely pointing errors.

10. Perform a two-dimensional fast Fourier transform with a positive kernel exponent to signify an $\exp(-j\omega t)$ time convention. The aperture sampling intervals are obtained from the fitted (u, v) sampling intervals by table 2.1. The map is now $a(x, y)$.

11. Minimize phase wrapping as much as possible by adding a constant offset to the aperture phase $\arg a(x, y)$.

12. Fit a least squares plane to the aperture phase over unshadowed points (x, y), and subtract it.

13. Correct the amplitude for sample smearing by applying equation 3.2 in the direction corresponding to scanning.

Steps 11 and 12 require some explanation. Phase 'wrapping' occurs when the range of phases in the map exceeds 2π. Reducing phases to the same period then causes sharp jumps between $-\pi$ and $+\pi$. The number of discontinuities between neighbouring points can be minimized by shifting the phase of the whole map — a uniform phase shift does not affect the focusing properties of the reflector. If wrapping still occurs in an unshadowed part of the aperture, the surface errors range over more than half a wavelength and the experiment should be repeated at a lower frequency. In shadowed regions, diffraction can cause steep phase gradients and some wrapping is often unavoidable. To the unshadowed points, however, a least squares plane can be fitted whose gradient corresponds to a constant pointing error in the (u, v) domain; that is, the (u, v) grid centre did not agree with the on-source position. By subtracting the plane in step 12 the effect of this error is removed from the map.

Changing the position of the feed also causes characteristic variations in aperture phase. Neglecting pointing errors, the phase map represents deviations of the surface from the paraboloid determined by the position of the feed during the measurements. If the phase is zero uniformly across the aperture, the surface conforms to the unique paraboloid that has its focus at this position. If it is not, adjustments to the surface based on the phase will force conformity. On the other hand, lateral and axial motion of the feed may be used to change the target paraboloid and reduce the number of surface adjustments required. In this case the three feed position parameters should be included in the least squares fit for the three plane parameters, giving six parameters in all. The result defines a 'best fit' paraboloid for the existing surface: it is no coincidence that a general paraboloid has six free parameters.

Godwin, et al. [19] have discussed optimisation of the feed position in this way. Often the feed mounting does not allow three degrees of freedom of movement, in which case the least squares fit should take advantage of whatever freedom exists, the remainder of the phase errors being removed by surface adjustments. This was done, for example, for the Effelsberg 100 m antenna survey [20] where there was only one direction for lateral feed adjustment. A Cassegrain or Gregorian antenna has a subreflector with its own positioning errors, and if the phase errors characteristic of these are known

Name	Date	Time (AEST)	MHz	Integ.	Smear
MB	25/4	0810–0840	12725	0.5 s	36.4%
C2	26/4	0020–0405	12725	1.0 s	29.1%
C3	26/4	0423–0811	12725	2.5 s	72.7%
B1	1/5	0011–0416	12373	2.5 s	72.7%
B2	1/5	0421–0828	12373	2.5 s	72.7%
C4	2/5	0004–0447	12725	2.5 s	72.7%
C5	2/5	0453–0848	12725	2.5 s	77.6%

Table 3.1: A summary of the holographic maps of the 26 m antenna. Names are given for easy reference. The size of all maps was 53×53 except for the first of 15×15, and sampling was $\delta_u = \delta_v = 0.00075$. This gave $\kappa = 0.825$ at 12725 MHz (carrier) and $\kappa = 0.802$ at 12373 MHz (beacon).

more parameters can be introduced to the fit. Otherwise, subreflector deformations and position errors will be embodied in corrections to the main reflector.

Another issue which can be important for high-precision surfaces is the shape of the panels themselves. Panels are fabricated to agree with a particular paraboloid, and systematic aperture phase errors are caused if the panels are adjusted to form a paraboloid with a different focal length. The result is a 'scalloping' of the phase map [27].

3.3 Results and Surface Adjustment

3.3.1 Surface Error Maps

The characteristics of the maps taken in April and May 1991 are summarized in table 3.1. As the times indicate, most of the data were collected in the small hours of the morning when the temperature of the structure was most stable; these times coincided with the availability of carrier signals from AUSSAT. The Autumn weather was kind, and most nights were cold and very still. Only the final morning was unsettled, with gusty 50 kph winds causing such large pointing inaccuracies that the map C4 is useless and the C5 map is unreliable.

On 29 and 30 April the reflector panels were adjusted according to the

results of maps C2 and C3. There was thus one morning before the adjustments during which two contiguous, high-quality maps were taken, and another such morning after the adjustments. It was to allow this that the scan speed was chosen to keep the time required for the maps at about 4 hours. Having a 'mate' for each map taken under almost identical conditions was very helpful for separating errors intrinsic to the holographic technique from day-to-day changes in the reflector. Once good repeatability was established for these pairs, they were averaged vectorially to improve the signal-to-noise by a factor $\sqrt{2}$. The vector average of maps C2 and C3, denoted by AV1, is shown as a contour plot of normal surface errors in figure 3.8. The average of maps B1 and B2, denoted by AV2, is shown in figure 3.9. The contour interval for each is 0.5 mm: on this scale most of the features to be seen are not artifacts of noise but genuine.

The aperture current amplitude depends upon the illumination pattern of the feed and the shadow of the feed support structure. Figures 3.10 and 3.11 show the illumination functions according to AV1 and AV2. The edge taper is approximately 12 dB and there is shadowing of both the incoming plane wave and the reflected spherical wave. Several other sharp dips in the illumination are caused by particularly poor panels: for example, the access hatch lid is on the southern edge of the western leg. Steep and irregular contours of aperture phase can usually be identified with a poor panel or group of panels. When light strikes at a low angle the bumpy surface of these panels can be seen with the eye.

3.3.2 Consistency and Surface RMS

Root mean square (rms) phase errors and differences are commonly used to compare and examine aperture phase estimates; beforehand it is wise to think carefully about the weighting to use. If $a(x,y)$ is the aperture current and $\phi(x,y)$ is its phase, then equation 4.20 gives that $\sigma_{\phi(x,y)} \propto 1/|a(x,y)|$. It is appropriate to weight the least squares fit by $1/{\sigma_\phi}^2 \propto |a(x,y)|^2$ when estimating panel positions: see appendix C. The same weight is useful when comparing two independent estimates of $\phi(x,y)$ for the following reason. The discrepancy at each aperture point is $\Delta = \phi_1 - \phi_2$, but σ_Δ is not constant over the aperture because of an illumination dependence $\sigma_\phi = k/|a|$. This is discussed in section 4.2.4, where $k = \text{rms}\,|\tilde{e}|/\sqrt{2}$. Now $\sigma_{|a|\phi} = |a|\sigma_\phi = k$

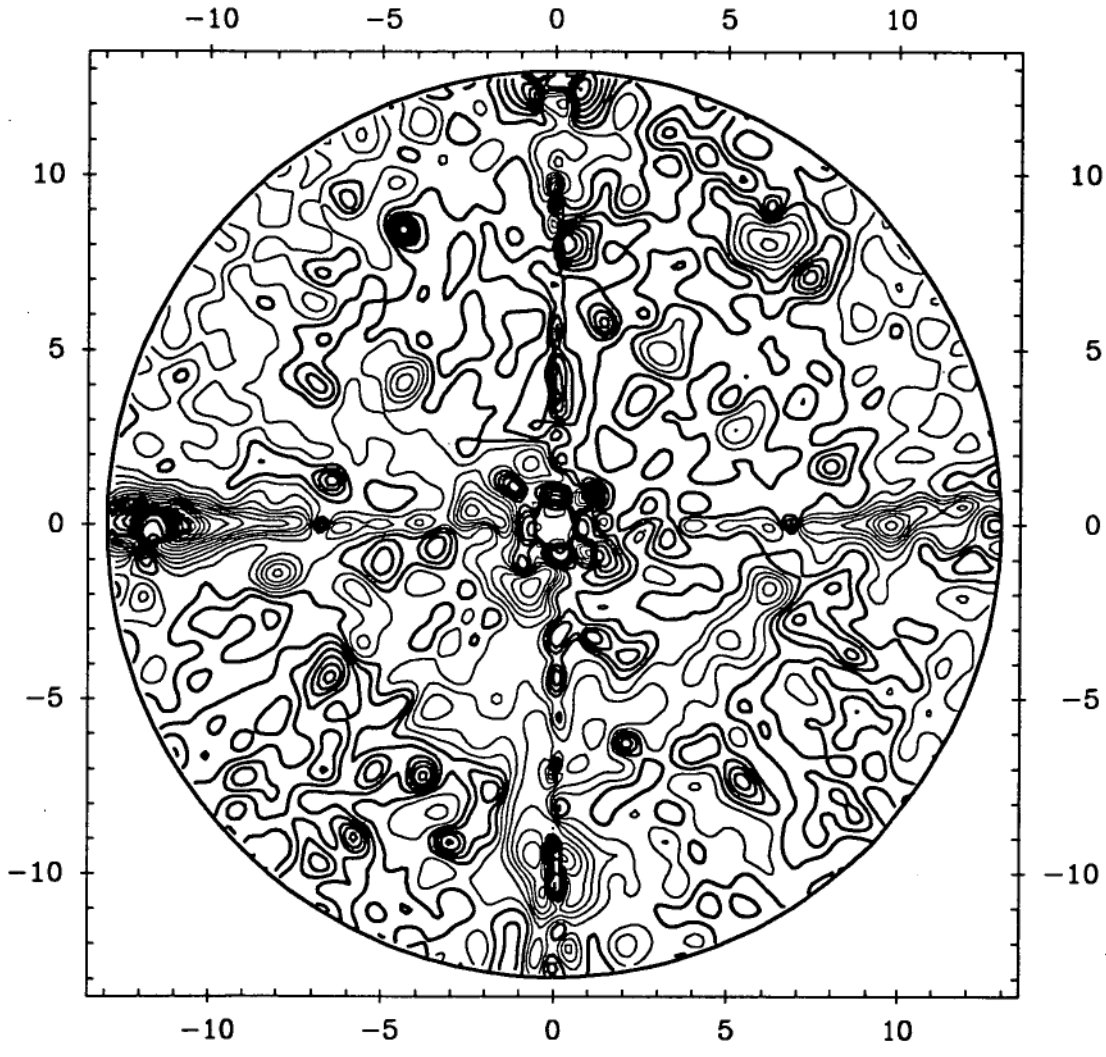


Figure 3.8: Map AV1, the vector average of maps C2 and C3, as a contour plot with contour interval 0.5 mm. Negative contours are thinner. The scales are in metres.

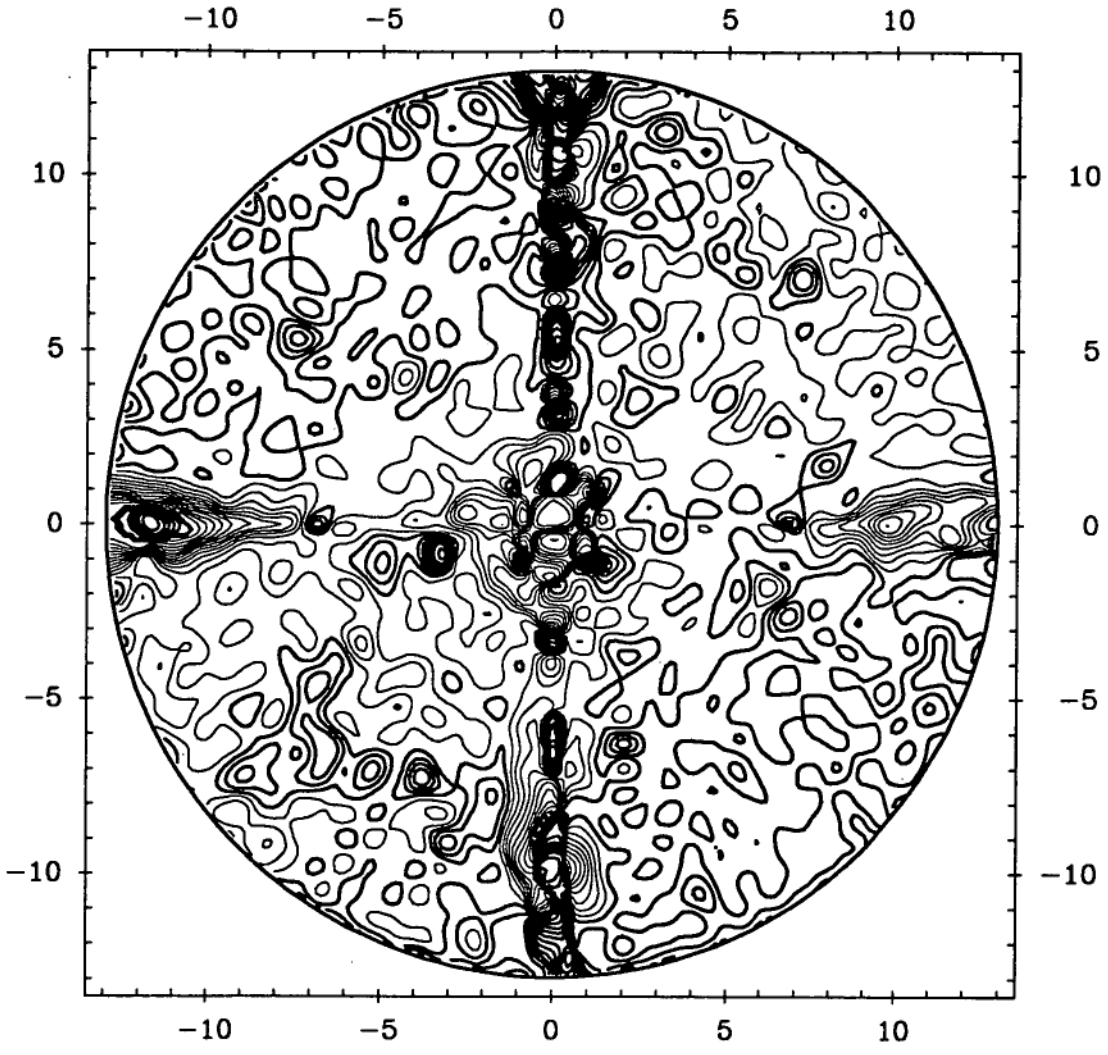


Figure 3.9: Map AV2, the vector average of maps B1 and B2, as a contour plot with contour interval 0.5 mm. Negative contours are thinner. The scales are in metres.

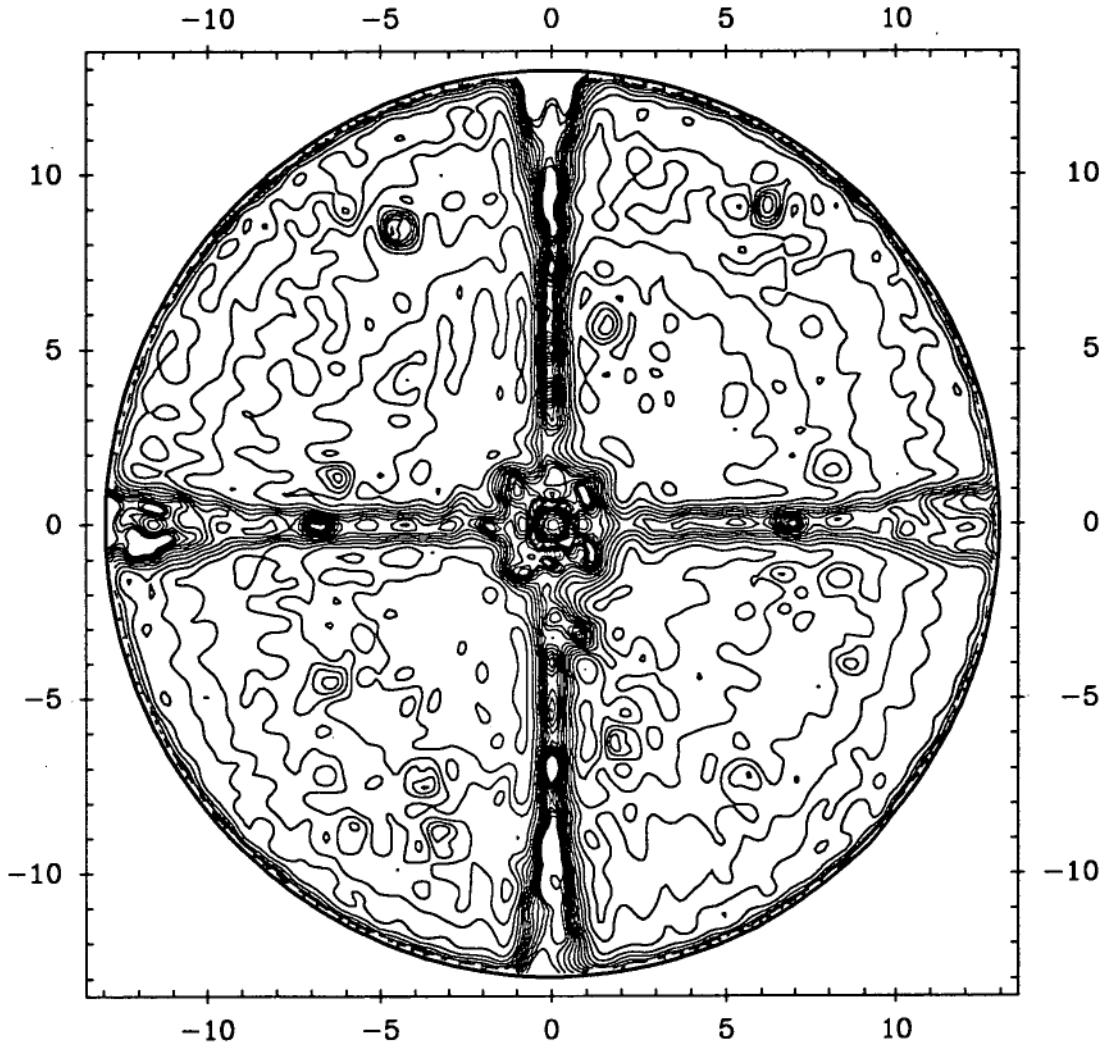


Figure 3.10: The aperture illumination according to AV1 in 1.5 dB contours. The scales are in metres.

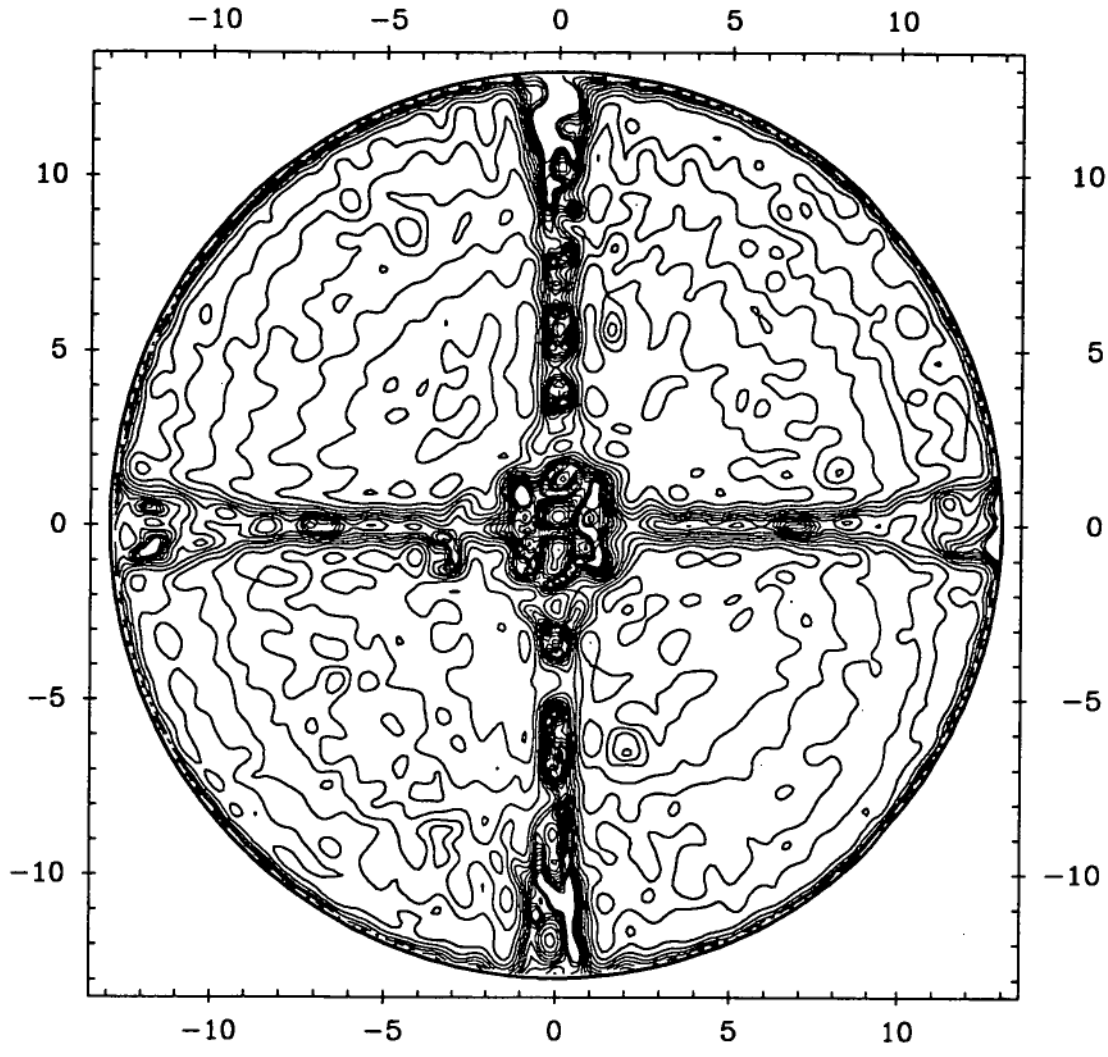


Figure 3.11: The aperture illumination according to AV2 in 1.5 dB contours. The scales are in metres.

which is constant over the aperture, as is

$$\sigma_{|a|\Delta} = \sigma_{|a|\phi_1 - |a|\phi_2} = \sqrt{2}\,\sigma_{|a|\phi} = \sqrt{2}\,k.$$

This can be estimated from two maps of n aperture phase estimates:

$$\sqrt{2}\,k = \sigma_{|a|\Delta} \simeq \sqrt{\frac{1}{n}\sum_{\text{ap}} |a|^2(\phi_1 - \phi_2)^2}.$$

Thus an $|a|^2$ weighting is necessary to find the constant of proportionality. The formation of an rms phase discrepancy with this weight gives a direct indication of the data quality; the unweighted rms is also useful to highlight the effect of the aperture taper.

It is entirely different to compare an estimated function $\phi(x, y)$ with zero, that is to find the rms phase error over the aperture. The purpose of such a statistic is to show the extent of deviation of the surface from a paraboloid, and to indicate the effect on the antenna's performance. The tolerance theory discussed in section 1.1.1 shows that both the gain and sidelobe levels depend on

$$\overline{\delta^2} = \frac{\iint_{\text{ap}} |a(x,y)|\phi^2(x,y)\,dx\,dy}{\iint_{\text{ap}} |a(x,y)|\,dx\,dy}.$$

Antenna gain and sensitivity calibrations using astronomical sources will estimate this statistic. Thus, if a single statistic is desired to embody a complete surface measurement it is best to weight the aperture phases by $|a|$. To summarize for the discrete case, when comparing independent data the *rms phase difference* is

$$\text{rms}\,\Delta = \sqrt{\frac{\sum_{\text{ap}} |a_i|^2(\phi_{1i} - \phi_{2i})^2}{\sum_{\text{ap}} |a_i|^2}}, \tag{3.6}$$

but when studying a single aperture function the *rms phase error* is

$$\sqrt{\overline{\delta^2}} = \sqrt{\frac{\sum_{\text{ap}} |a_i|\phi_i{}^2}{\sum_{\text{ap}} |a_i|}}. \tag{3.7}$$

Prior to their averaging contiguous maps were compared by forming the rms phase difference, masking points which were shadowed by the feed structure. The comparisons are shown in table 3.2 for each ring of panels. Ring 1 is the vertex region and is not shown, and ring 2, while included, is mostly shadowed and does not typify the accuracy of the phases. The other rings show the expected increase in phase error with radius. The overall rms phase

Ring	Outer Radius	Panels	C2 and C3	B1 and B2
2	2.92 m	12	2.37°	2.86°
3	5.43 m	24	1.17°	1.70°
4	7.47 m	24	1.58°	2.14°
5	8.87 m	48	1.69°	2.03°
6	10.24 m	48	1.90°	2.14°
7	11.58 m	48	2.29°	2.45°
8	12.95 m	48	3.07°	3.51°
2–8	12.95 m	252	1.82°	2.20°

Table 3.2: Ring-by-ring rms phase differences in degrees between contiguous maps.

difference of two maps expected to be identical gives the 'repeatability' of the measurements. Thus, C2 and C3 agree to $\pm 1.82°$ while B1 and B2 agree to $\pm 2.20°$, or if no amplitude weighting is used $\pm 2.36°$ and $\pm 3.01°$.

These numbers are frequency-dependent, but can be converted to axially resolved surface repeatability by multiplying by $\lambda/4\pi$. Alternatively the phase maps can be converted to normally resolved surface errors by equation 2.12 and then compared, which gives a more practical measure. It is found that C2 and C3 agree to ± 0.075 mm and B1 and B2 agree to ± 0.091 mm. Remember too that these are difference statistics, and the uncertainties in each map, if equal, would be a factor $\sqrt{2}$ smaller again. This gives ± 0.053 mm and ± 0.064 mm, or as a fraction of the wavelength 0.22% and 0.26%, an excellent result.

Both the phase and surface error results show that the carrier maps are more precise. The difference is not great, yet the signal power is 20 dB greater for the carrier than for the beacon. If precision was determined primarily by receiver noise a much greater difference would have occurred. This indicates that other sources of error must have been limiting the precision. Chapter 4 discusses this in detail.

Now that the reliability of the maps has been tested, the rms phase error representing the quality of the surface can be found by equation 3.7. The average maps AV1 and AV2 are used for this purpose, and the improvement due the surface adjustment can be measured by comparing them. Table 3.3 shows the rms deviation from the ideal paraboloid in phase and normal surface er-

RMS of:	Weighted?	AV1	AV2	Improvement
phase	yes	20.7°	17.8°	14.0%
error	yes	0.735 mm	0.689 mm	6.3%
phase	no	21.0°	19.6°	6.7%
error	no	0.740 mm	0.718 mm	3.0%

Table 3.3: Root mean square deviation of aperture phase and normal surface error from the ideal paraboloid for maps AV1 and AV2.

ror, with and without amplitude weighting. The improvement is smaller without amplitude weighting, indicating that the adjustments were more effective for the inner part of the reflector. The improvement is also smaller for normally resolved errors, indicating the same thing: normal errors increase with radius with respect to axial errors, accentuating defects of the outer part of the reflector.

For the reasons given above, the amplitude-weighted phase error is the best measure of reflector performance, so an improvement of 14.0% appears to have been achieved. According to the Ruze equation 1.2, the reduction in gain at Ku-band has changed from -0.57 dB to -0.42 dB, and the frequency at which the gain peaks is about 40 GHz.

3.3.3 The Surface Adjustments

This section looks at the surface adjustments in detail, and compares them with the change in the surface profile measured by holography. Reflector panel adjustments were made on the basis of the average map AV1. The Successive Projections method, described in chapter 5, was applied for 10 iterations on an expanded grid of 241×241 points. This improved the effective aperture resolution, allowing panel positions to be estimated that were much less affected by neighbouring panels. From the tilts and vertical shift of each panel, four corner adjustments were found which would make the panel conform to the best-fit paraboloid. This was under the assumption that the panel was itself undistorted: at this stage Successive Projections had not been thoroughly tested with a panel twist parameter.

There are 252 panels making up the reflector surface: see figure 2.12. All panels which

- were not affected by aperture blockage, and
- required more than 0.5 mm adjustment at at least one corner

were adjusted at all corners, a total of 191 panels. The panels in the outer four rings (numbers 5–8) do indeed have an adjustment screw at each corner with a 28 t.p.i. thread. The recommended corner adjustments were converted to screw turns and applied to the nearest quarter-turn or 0.23 mm. The panels in rings 2–4 have adjustment screws on some edges, and these were turned according to a mental interpolation between the corners.

A significant number of panels could not be adjusted properly because of a mechanical constraint. Either a screw was at the limit of its travel or occasionally a panel came hard against its mounting brackets. These problems arose because the panel mounts were not fixed accurately onto the girders of the backing structure. The mounts can be moved to allow the panels sufficient travel, but not by a precise distance and not without removing the panel in many cases. With limited time available it was decided to leave the problem until some future date, after which another holographic survey can be done. In total, 53 panels were affected.

The adjustments actually applied to the panels were recorded while on the surface. Panels positions can be derived from the map AV2 exactly as from AV1, and the difference between the corner positions so derived can be compared directly with the corner adjustments. This is an end-to-end test of the measurement and adjustment cycle. The most effective way to display the comparison is by a correlation plot of measured adjustment against actual adjustment. Each corner of each panel produces a single point on this plot, which is shown in figure 3.12. Since holography results are not reliable for panels affected by shadowing, these panels are excluded from the plot; to avoid clutter at the centre, so are panels which weren't adjusted at all.

There is a clear positive correlation with unit slope: the panels were moved in the correct direction. Unexpectedly, however, there is a scatter of roughly ± 1 mm about the line '$y = x$'. The repeatability of maps indicates a precision an order of magnitude better than this for individual aperture phases, and with an average of 5 points per panel a similar precision, perhaps slightly worse, is expected for the phase at panel corners. How can this be?

Performing the Successive Projections iterations with panel twists enabled did not change the plot significantly. The discrepancy is unlikely to be due to panel distortion. The difference in source frequencies for AV1 and AV2 is

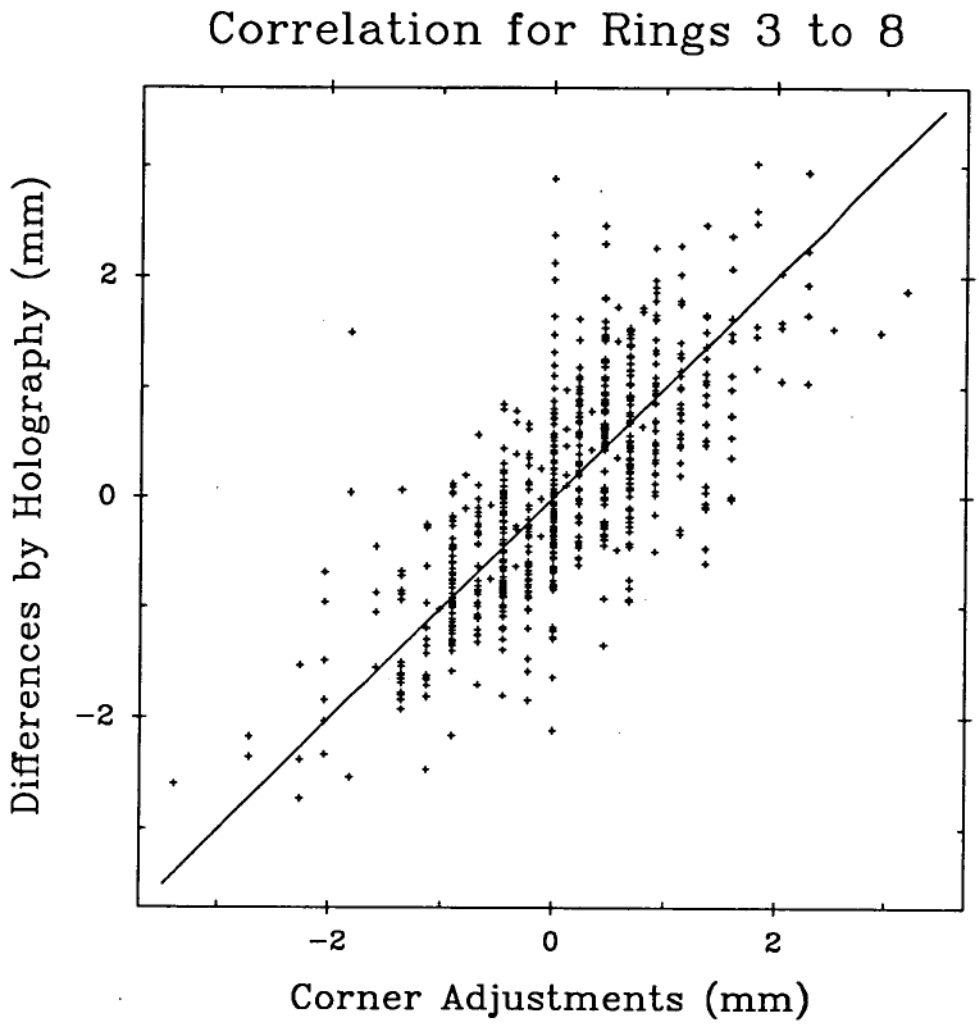


Figure 3.12: A correlation plot of measured changes in panel corner heights against actual adjustments applied. Each point represents a single corner of a panel. Panels affected by shadowing and panels not adjusted are not included. The line '$y = x$' is shown.

relatively small and should not cause a large random error. It is unfortunate that maps C4 and, to a lesser extent, C5 were disturbed by wind, preventing a reliable test of the effect of frequency. There were no other differences in measurement methods which would affect the comparisons without also affecting the repeatability of contiguous maps. The remaining likely possibility is that between 26 April and 1 May there were changes to the surface profile in addition to the panel adjustments. Step 12 of the data reduction process in the previous section guarantees that such changes would not give a surplus of points above or below the $y = x$ line, consistently with figure 3.12.

This is not so for the correlation over a part of the aperture. The aperture was divided into 12 segments: 8 octants of the annulus of the outer rings 5–8 and 4 quadrants of the annulus of rings 3–4. In each of these a separate correlation plot of measured change against actual adjustment was performed. They appear in figures 3.13, 3.14, and 3.15. The $y = x$ line is shown in all plots: in most the points tend to lie above or below the line. In these segments this represents a systematic raising or lowering of the surface not accounted for by the adjustments. The average shifts are given on a diagram of the reflector in figure 3.16, showing how systematic the effect is. The surface changed shape in an astigmatic way, opposite in sign on opposite sides of the dish, with an overall lowering of the inner region.

Underneath the surface backing structure is a square of girders which is attached to the Y-wheel. The orientation of the square, which is shown, may explain the symmetry of the change in shape. It has three probable causes:

1. Mechanical hysteresis after moving the antenna and returning to 39° elevation for AUSSAT-1;

2. Hysteresis after the structure experienced several days of temperature change;

3. An actual difference in temperature on the two mornings of measurement.

It is not possible to separate these factors without further measurements. Nevertheless, it is very useful to know that changes in the surface profile of roughly 0.5 mm amplitude can be expected. For example, further surface adjustments will be of limited value unless the structure can be stiffened.

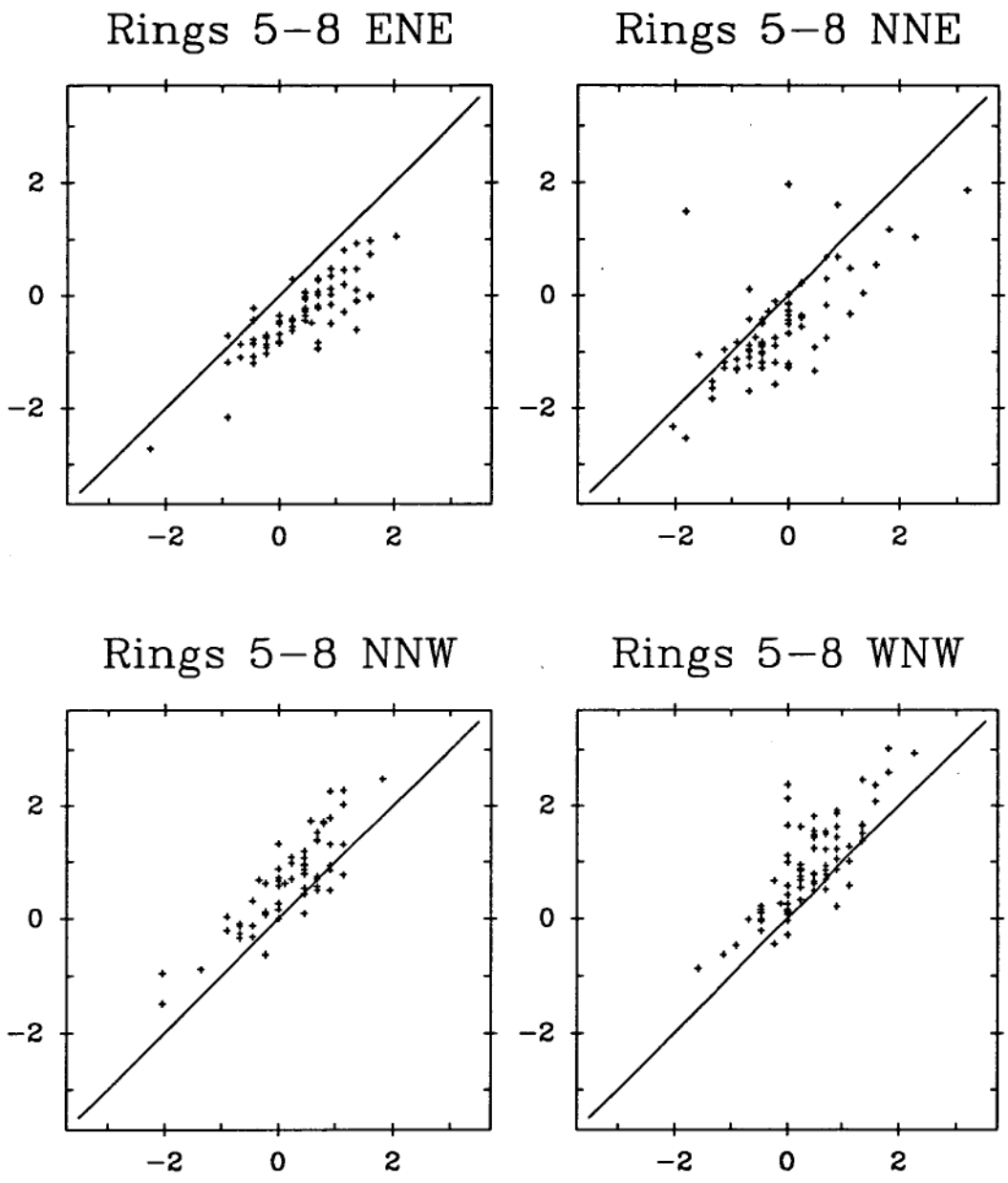


Figure 3.13: Correlation plots of measured change against surface adjustment for the 4 northern octants of rings 5–8, each containing 20 panels. The scales are in mm.

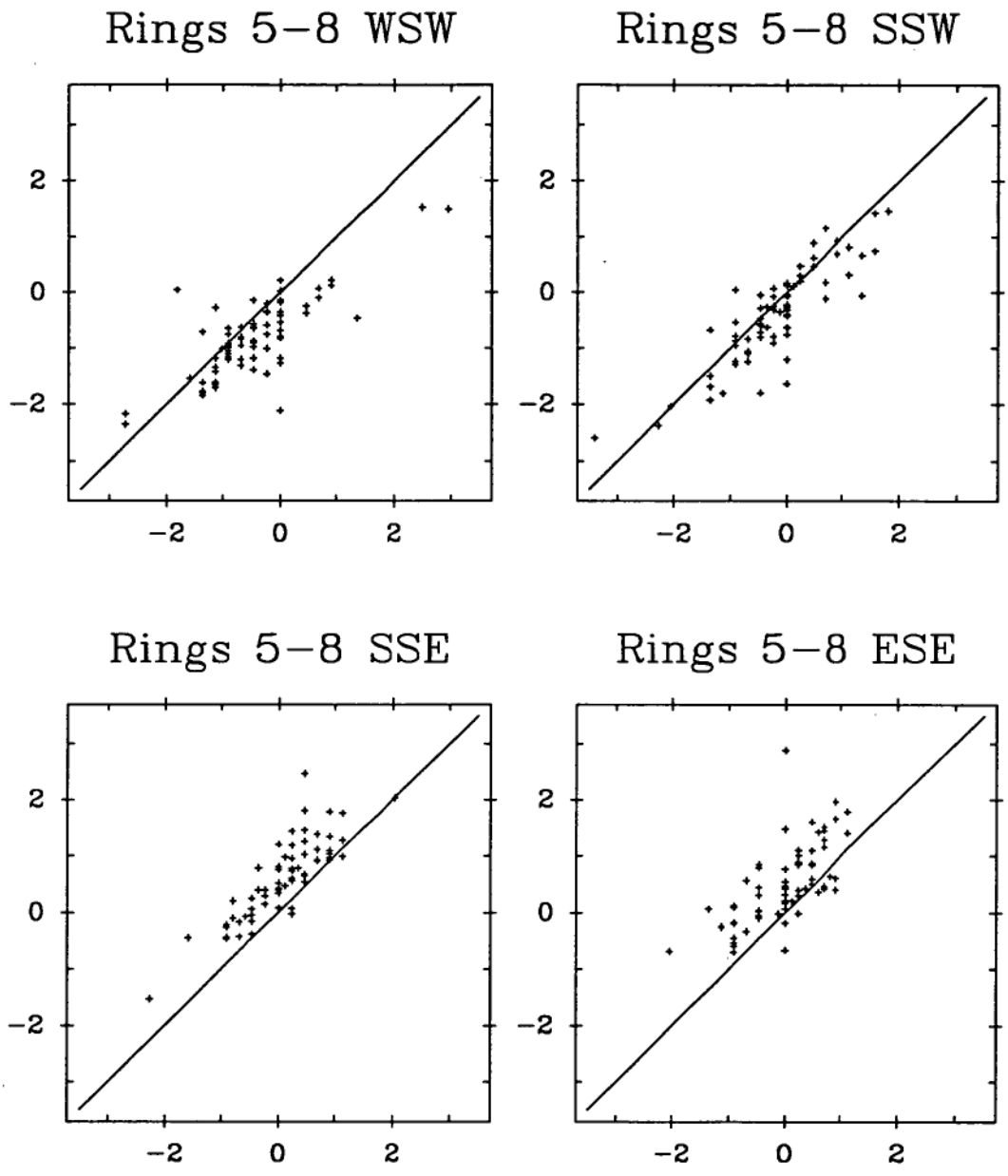


Figure 3.14: Correlation plots of measured change against surface adjustment for the 4 southern octants of rings 5–8, each containing 20 panels. The scales are in mm.

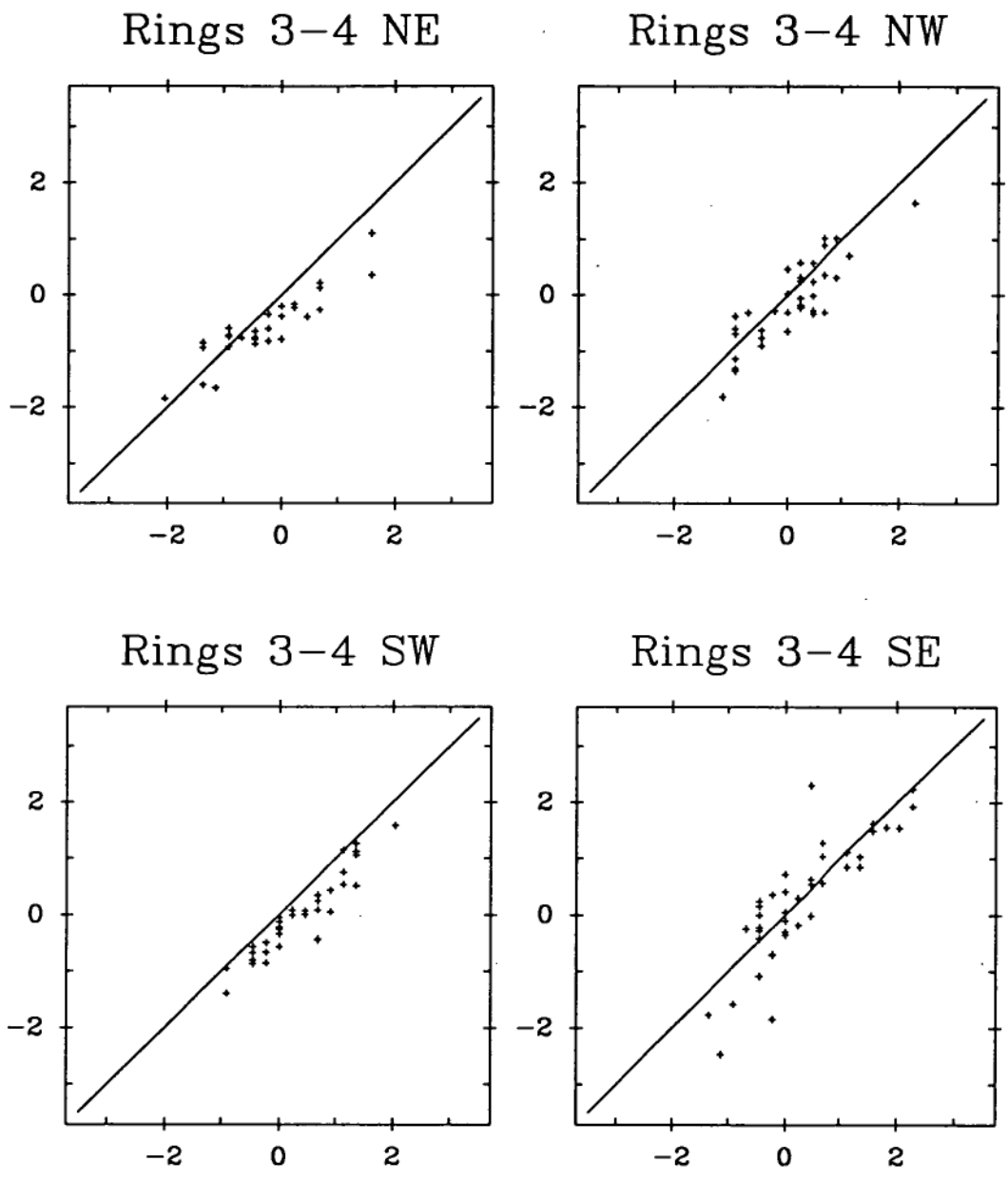


Figure 3.15: Correlation plots of measured change against surface adjustment for the 4 quadrants of rings 3–4, each containing 10 panels. The scales are in mm.

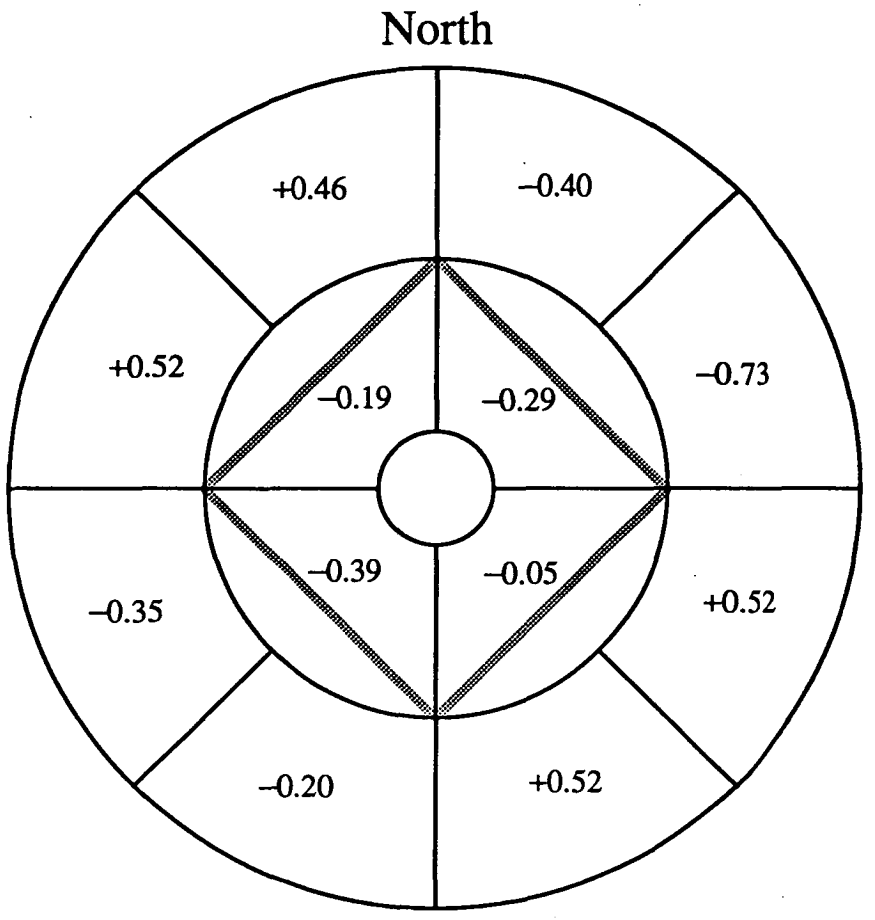


Figure 3.16: The average discrepancy between measured changes in the surface and changes made by adjusting panels. The view is down on the antenna in the zenith position, with North at the top of the page. Thick grey lines mark the square of girders supporting the reflector. The units are mm.

3.3.4 Antenna Properties Deduced

An advantage of the holographic method is that it provides information about the antenna's aperture illumination and far field radiation pattern in addition to the surface profile. Of course, this information is of limited practical use because it is specific to the feed and frequency used for the measurements. The far field pattern can be estimated at any frequency, if an aperture illumination function is provided, by scaling the measured aperture phase function appropriately and performing a Fourier transform.

The aperture illumination function for the Ku-band feed has already been shown. Far field cuts in the u and v directions are shown in figure 3.17 for AV1 and AV2. The highest sidelobe is respectively 16 dB and 13 dB below the main beam. The main beam and inner sidelobes are graphically illustrated in two dimensions in figures 3.18 and 3.19. All far field data has been interpolated for display by 'padding' with zeros in the aperture domain, which increases the far field resolution without adding any higher-frequency information. The small map MB, which was collected on a dense grid and could be graphed without interpolation, matched well with AV1. Thus the interpolation method provides a faithful representation of the pattern.

The far field has a four-fold symmetry due to the quadrupod feed support structure. A marked asymmetry in AV1 has been improved in AV2: this could be due to the surface adjustments, but the different source frequencies of the maps and the changing astigmatism both may have contributed. The same far field region for the single map C5, less affected by windy conditions than C4, was very similar to AV2. This indicates that the small change in frequency was not an important factor.

Several standard antenna properties are related to integrals of the normalized antenna power pattern $P_n(\theta, \phi)$, where the maximum P_n is unity. These are discussed in any text on antennas or radio astronomy, for example see Kraus [37, chap. 2] or Rohlfs [69, sect. 4.5]. $P_n(\theta, \phi)$ is directly obtainable from holographic measurements over a limited range of angles θ from the main beam, and so the integrals can be estimated. For example, the *beam solid angle*:

$$\Omega_A = \iint_{4\pi} P_n(\theta, \phi)\, d\Omega$$

and the *main beam solid angle*:

$$\Omega_M = \iint_{\text{main beam}} P_n(\theta, \phi)\, d\Omega.$$

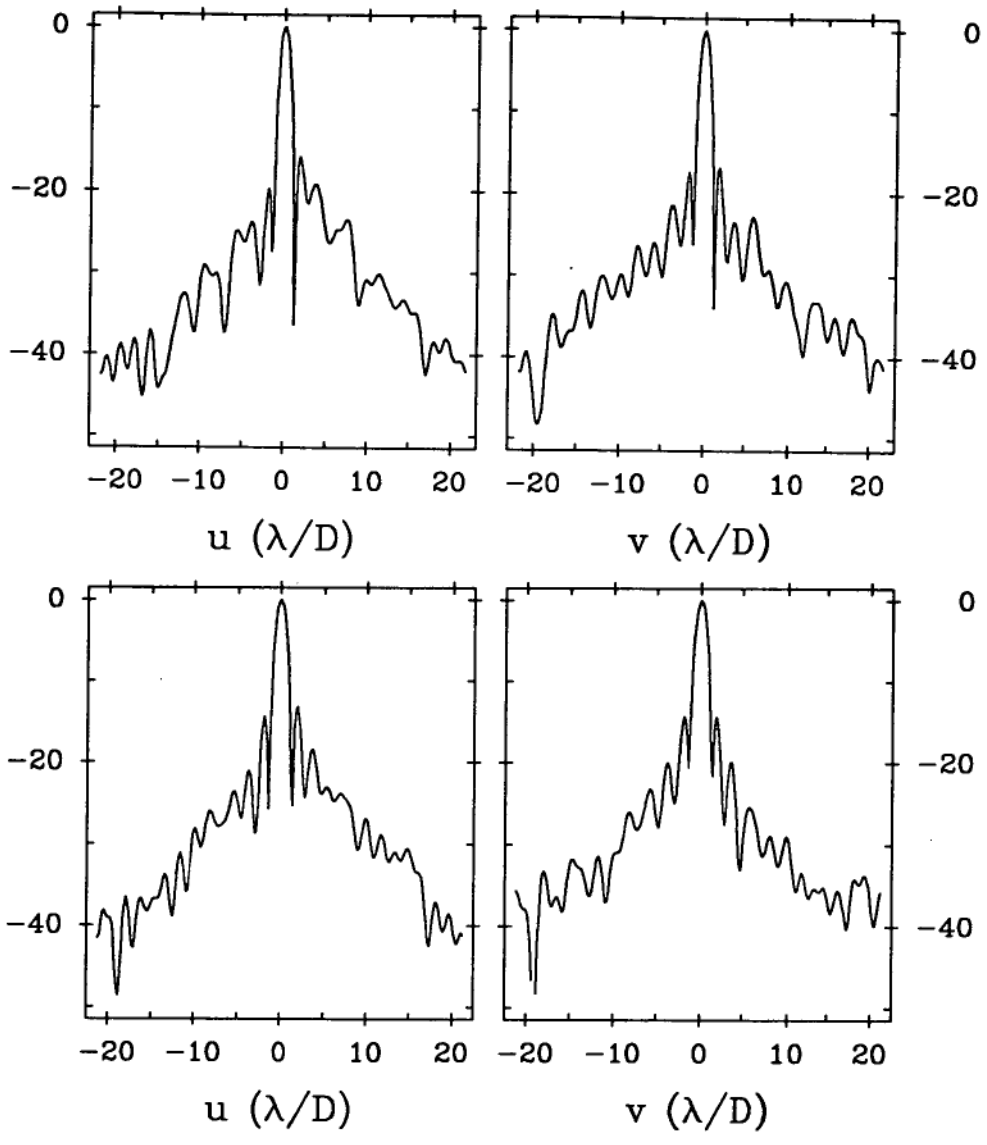


Figure 3.17: Far field cuts across the centre of the map AV1 (above) and AV2 (below). The vertical scales are in dB of power, the horizontal in units of $\lambda/D \simeq 0.053°$.

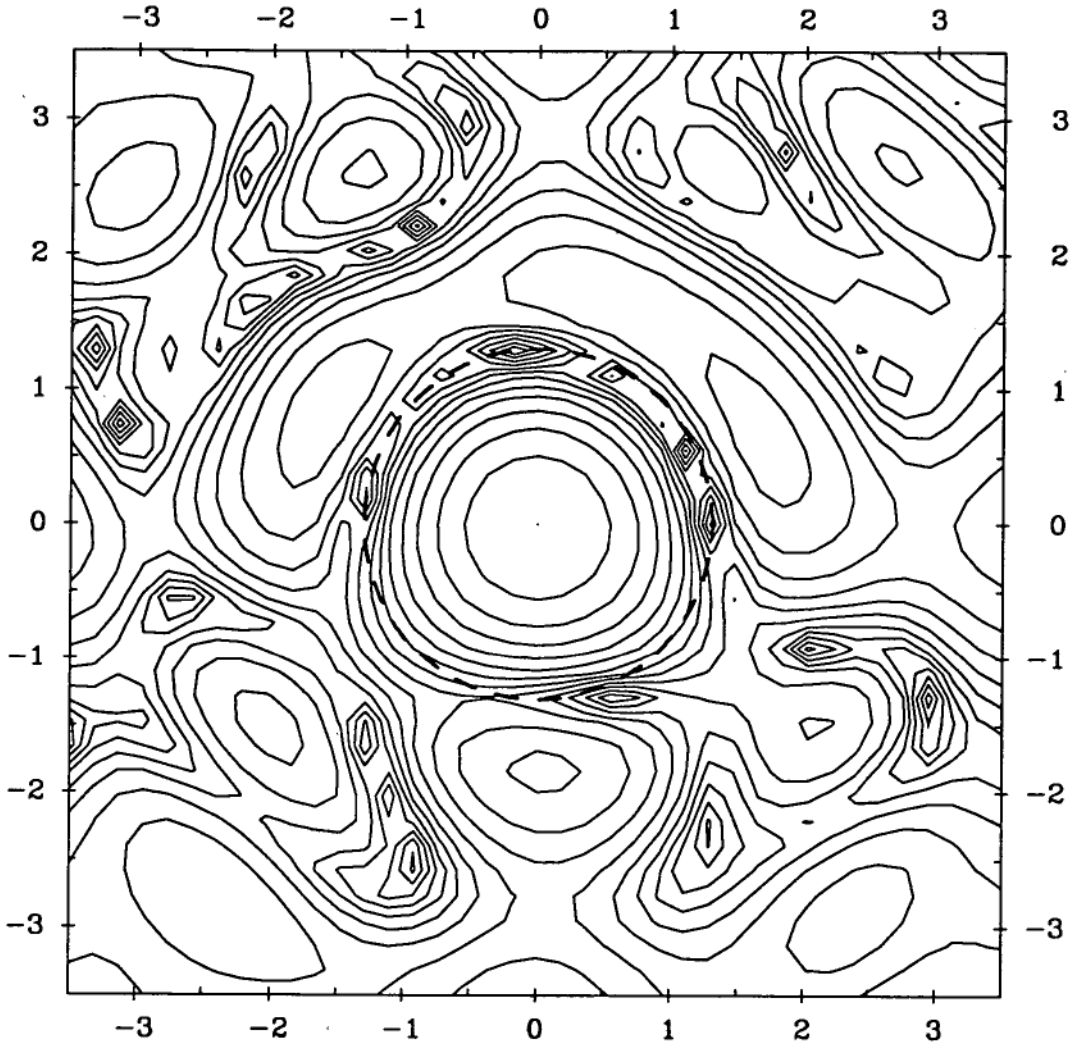


Figure 3.18: The main beam and inner sidelobes from map AV1. The contour interval is 3 dB. The dotted line is an arbitrary boundary for the 'main beam' used for estimating Ω_M. The scales are in units of $\lambda/D \simeq 0.053°$.

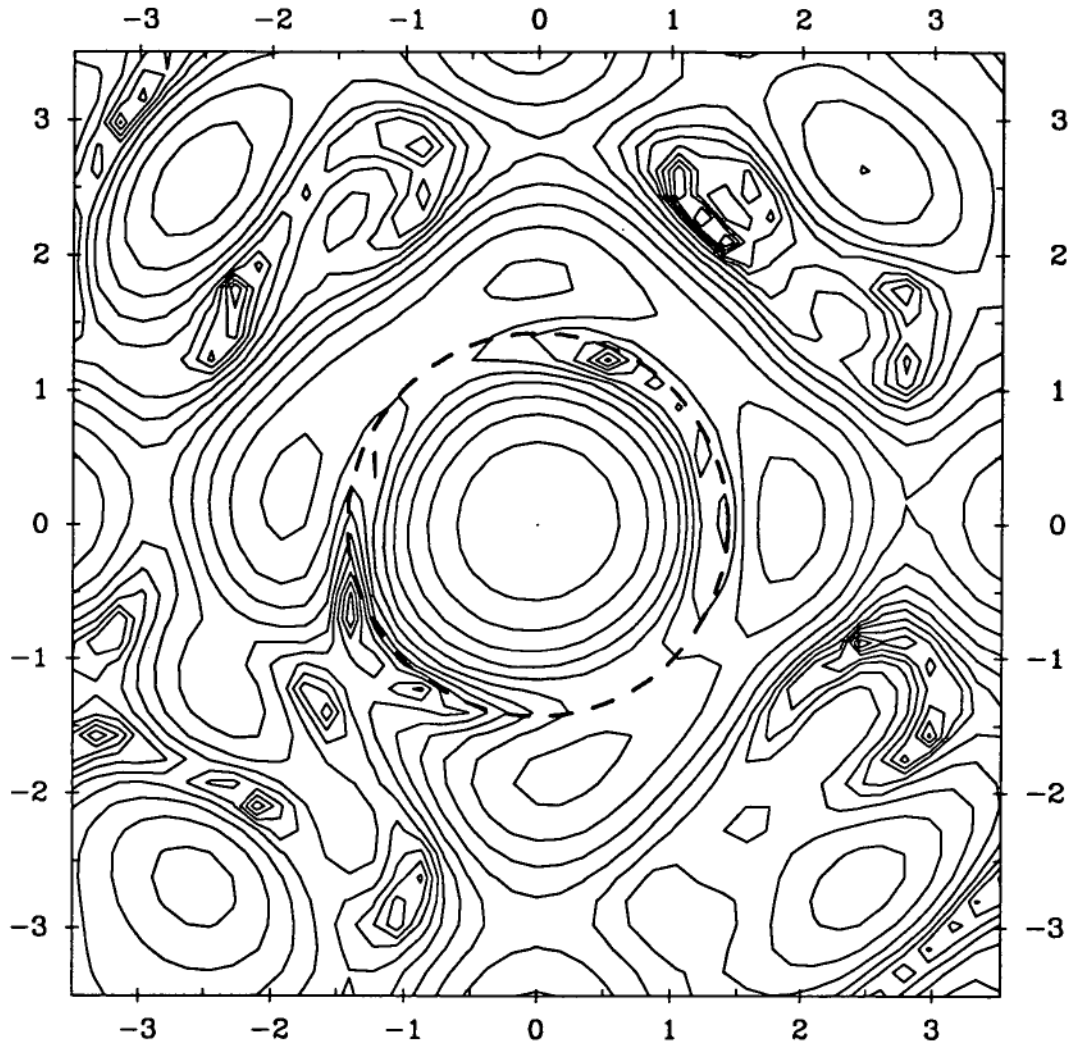


Figure 3.19: The main beam and inner sidelobes from map AV2. The contour interval is 3 dB. The dotted line is an arbitrary boundary for the 'main beam' used for estimating Ω_M. The scales are in units of $\lambda/D \simeq 0.053°$.

Quantity	Map AV1	Map AV2	Map C5
$\mathcal{D}$	69.8 dB	68.7 dB	69.5 dB
A_e	424 m^2	345 m^2	389 m^2
η_M	0.755	0.761	0.757
η_A	0.804	0.654	0.738

Table 3.4: Upper limits for some antenna properties deduced from the integrals Ω_A and Ω_M for the maps indicated.

These are both in steradian measure. Clearly Ω_A will be underestimated, perhaps severely, when all but 0.00158 sr of the pattern is assumed zero for the map size used here. Also, the value of Ω_M will depend on a somewhat arbitrary boundary to the main beam being chosen.

Upper limits for the directivity $\mathcal{D}$, the effective aperture A_e, the beam efficiency η_M, and the aperture efficiency η_A can be found using these relations:

$$\mathcal{D} = \frac{4\pi}{\Omega_A} \quad A_e = \frac{\lambda^2}{\Omega_A}$$

$$\eta_M = \frac{\Omega_M}{\Omega_A} \quad \eta_A = \frac{A_e}{A_g}.$$

A_g is the geometrical area of the antenna's aperture. In table 3.4 these quantities are shown as estimated from the maps AV1, AV2, and C5. Their variability prevents a statement about the effect of the adjustments or the astigmatism, but supplemented by other antenna calibrations these data may be useful. Note the usage here is that of Rohlfs: A_e is the A_{em} of Kraus, who then defines an effective aperture which takes account of the electrical losses in the antenna.

Chapter 4

Sources of Error

The Fourier transform of the measured radiation pattern can never represent the reflector surface profile exactly. Some discrepancies are caused by processes which can potentially be measured and their effects corrected, though this is not always done. These are systematic errors, and the major ones are discussed in section 4.1. Other discrepancies are caused by random processes which differ unpredictably for each map. These random errors are discussed in section 4.2, where a model by which they can be categorized is proposed and the magnitude of the effect of each category on the recovered aperture is calculated. In section 4.3 this model is shown to describe adequately random errors in the Mount Pleasant holography experiments.

4.1 Systematic Errors

The most obvious disruption to the aperture maps in the previous chapter was the shadow of the feed cabin and its support struts. Aperture blockage will be discussed first, followed by several other effects. The paper by Morris, et al. [51] will be cited several times: it has an especially thorough error analysis for holography with a cosmic source.

4.1.1 Aperture Blockage

Any reflector not of an offset design is inevitably blocked to some extent by a feed or subreflector and its support structure. To account fully for the effects of blockages on an antenna's electrical properties, mutual coupling, multiple scattering, and multiple shadowing all need to be considered. Typical of comments in the literature [73] is that this "is an electromagnetic problem

which defies exact solution within the foreseeable generation of computer technology." A major object of further work in Microwave Holography must be to quantify the main effects of aperture blockage on the measurements, and then to compensate for them.

In this thesis blockage has been accounted for by assuming simply that there is no radiation from regions of the reflector which are in the geometrical shadow of the feed or support struts. This is called the *null field hypothesis*. It is a reasonable approach when the blocking elements are several wavelengths across. It means that all information about shadowed reflector panels is lost, and partly shadowed panels may not be reliably determined. However, fully illuminated panels, for which correct setting is more important, are only minimally affected if at all because blockage effects are localized about the geometrical shadow. The use of the null field hypothesis and the localization of blockage effects will be discussed now.

Consider the next stage of sophistication, the *induced field ratio hypothesis*. In the terminology of radiation rather than reception, this assumes that currents are induced on a conducting strut by rays emerging from its shadow on the reflector surface, as if the strut was infinitely long and the rays represented free-space plane waves. Rusch, et al. [71, 70] have considered radiation from such currents, defining the Induced Field Ratio (IFR) of this field to the hypothetical field from the geometrical shadow alone. Thus IFR $= 0$ if the blockage has no effect at all, and IFR $= -1$ corresponds to the null field hypothesis. In [72] the radiation is shown to occur in cones about each strut which in the far field give large circles intersecting at the main beam. In the small region where holographic measurements are made the arcs appear as straight lines, and the Mount Pleasant data shows ridges of increased field intensity along the u and v axes in accordance with this, as can be seen in figure 4.1.

The support struts above the 26 m reflector are composed of three parallel, cylindrical girders several inches in diameter with numerous smaller poles between forming a triangular lattice. This is similar to many other large antennas. With a wavelength $\lambda \simeq 24$ mm almost every pole is wider than 2λ, and most are much wider. The graphs in [71] and [70] show that in these circumstances the IFR will be quite close to -1 for the components of the strut. Certainly, if the entire strut is treated approximately as a single girder the null field hypothesis will hold with good accuracy. Tests show that the null field hypothesis also predicts ridges of far field intensity along the u and

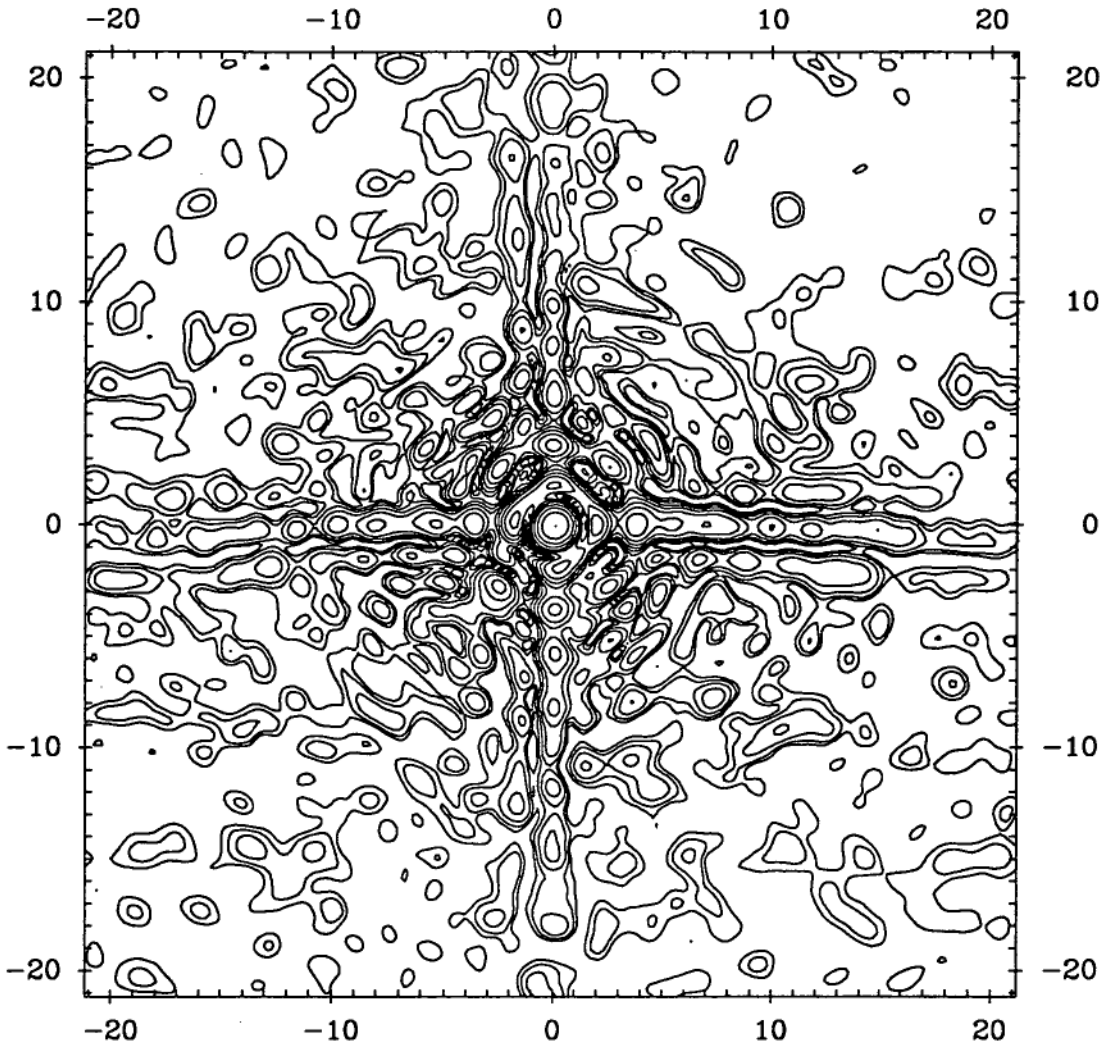


Figure 4.1: The entire measured far field intensity map AV2, shown in contours at 3 dB intervals down to a level 42 dB below the main beam. The intensity is higher in ridges along the u and v axes. The scales are in units of $\lambda/D \simeq 0.053°$.

v axes in the region of interest for holography. All this suggests that strut currents caused by reflected rays may not be important.

The analysis of [72] ignored a second type of shadowing: the rays from the feed which meet the struts before reaching the reflector surface. The two types of shadowing can be identified in the maps 3.10 and 3.11 of the illumination of the 26 m reflector: the four struts meet the reflector at about 8 m from the centre, and at these points the strut shadows begin to diverge from a rectangular shape. In the terminology of radiation again, the spherical wave from the feed is diffracted by the struts to induce currents in the triangular shadow region out to the edge of the reflector. Radiation from these currents was ignored because in many far field regions currents due to the reflected plane wave are dominant. A full treatment would use geometrical optics methods [32, chap. 6] to investigate the field incident on the reflector, or perhaps a simple model such as [73, appendix] would be useful.

It may not be unreasonable, then, to use the null field hypothesis to predict the radiation from a blocked aperture, at least in the main beam region. Conversely, the aperture currents recovered by Microwave Holography from blocked aperture radiation may exhibit something like a null field in the geometrical shadow. The 26 m aperture illumination maps do show a sharp drop in the rectangular, plane wave shadows, subject to the 0.6 m aperture resolution. This is particularly true of the northern and southern struts, which carry large bundles of cables to the focus: notice how much deeper the shadows are beneath these. On the other hand, the triangular, spherical wave shadows of the eastern and western struts do contain some structure. It is not always sufficient to treat the strut as a simple, opaque girder.

To see whether diffraction patterns are likely to be detrimental to aperture phase recovery well outside the geometrical shadow, consider a simple case. When the antenna receives a plane wave, the diffraction pattern on the reflector caused by a strut will be much the same as that of any body with the same shape. This is because the loss of part of the wavefront is the most important occurence, and the presence or lack of radiating currents does not change the pattern significantly in the region of the shadow. Modelling a strut by an infinitely long conducting strip of the same width, the diffraction pattern can be found by assuming each edge behaves like the edge of an infinite plane: the struts are 0.73 m wide or 31λ at 12725 MHz, so the approximation is good. Equations for the field from each edge are given in [32, pp. 117–8], and figures 4.2(a) and 4.3(a) show the resulting patterns

when the strut is 2 m and 9 m above the reflector surface. The focal length is 10.97 m so this is a realistic range of heights.

The recovered currents would not look like this. They must first be convolved with the smoothing function corresponding to the aperture resolution, $\delta_x = 0.59$ m for the carrier maps in chapter 3. The results, after applying the convolution by discrete Fourier transform, are shown in figure 4.2(b) and 4.3(b). The amplitude has a wave extending well beyond the shadow, but at both heights the phase variation has mostly disappeared by the shadow's edge. If this behaviour is typical, holographic surface maps will not be drastically affected by aperture blockage except in and very near the geometrical shadow. A proper understanding of blockage effects would nevertheless be very welcome.

Subreflector Diffraction

One blocking body which may be amenable to analytical study is the subreflector of a dual reflector antenna. Its shape is well defined, although allowance would have to be made for adjustment of its position. Because the subreflector does not produce a well-defined beam, it is more efficiently analyzed using the Geometrical Theory of Diffraction [39] than by laborious Physical Optics integrals. A complication is that subreflectors are often shaped to differ from the standard Cassegrain design, as discussed in section 2.1, and they may have a 'skirt'. This can mean that the geometrical shadow of the subreflector made by the feed wave lies inside the edge of the main reflector. It has been found [30, 34] that the dominant feature of subreflector diffraction is a set of ring-shaped ripples in amplitude and phase near the edge of the aperture.

Attempts up to now to compensate for these features have been unconvincing. Instead, Ishiguro, et al. [29, 30, 31] did further holography using a feed at the prime focus. They showed, by the use of absorbers, that other important scatterers were structural members around the feed horn and the circular vertex plate (which was raised above the reflector surface). Another approach was taken by Kesteven [34]. The antenna was taken to be ideal at its commissioning following a theodolite survey, and holographic 'reference' maps were made which included the effects of subreflector and strut diffraction. Following adjustments to the surface at a later date another holographic map was made, and when the maps were subtracted the diffraction effects cancelled quite well, leaving residual surface errors. When diffraction effects

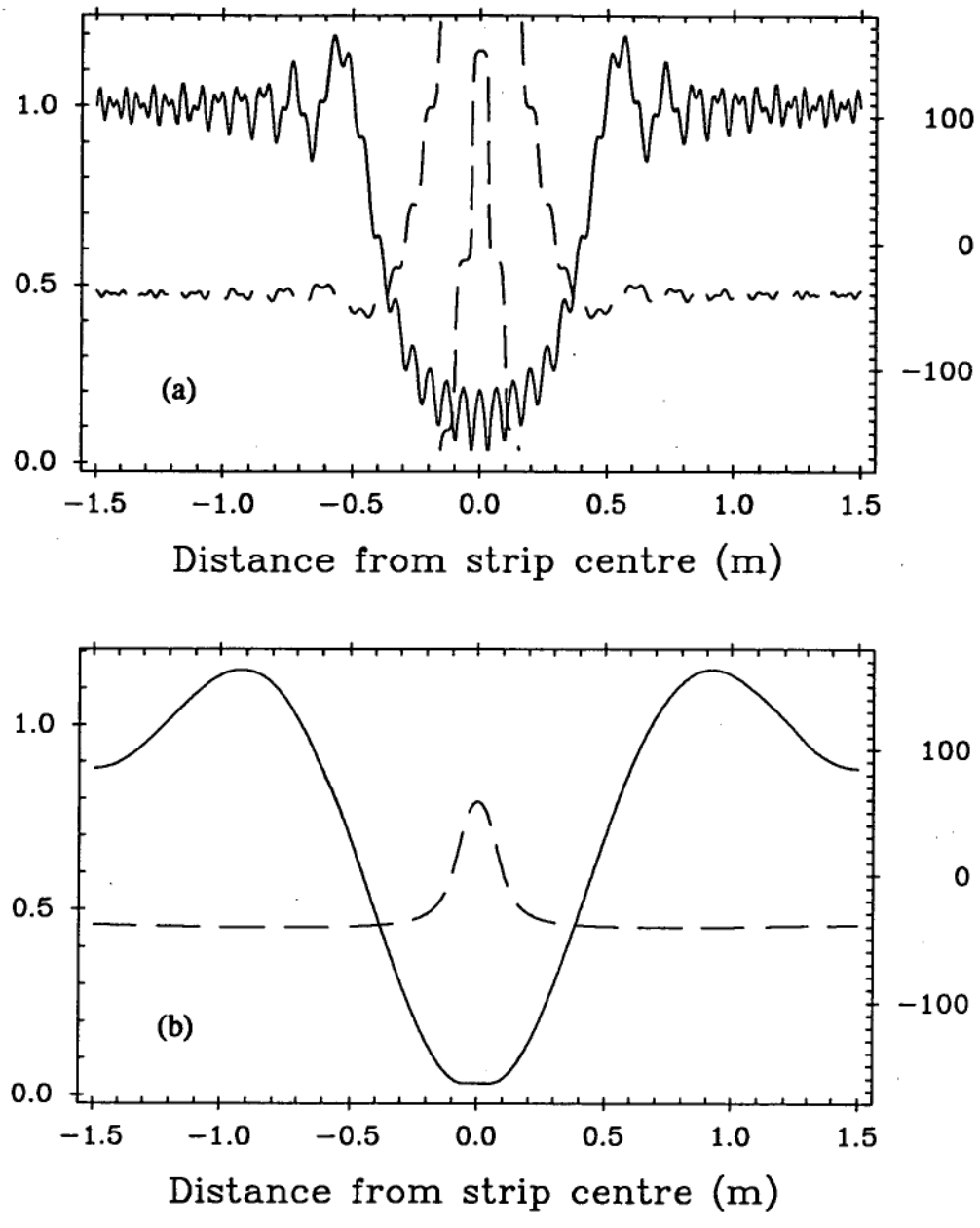


Figure 4.2: The diffraction pattern (a) of a plane wave 2 m beyond its normal incidence onto an infinitely long, conducting strip of width 0.73 m. The wavelength is 23.6 mm. In (b) the pattern has been convolved with the smoothing function corresponding to an aperture resolution $\delta_x = 0.59$ m. The solid line is the amplitude relative to the incident wave (left-hand scale); the dashed line is the phase (right-hand scale in degrees).

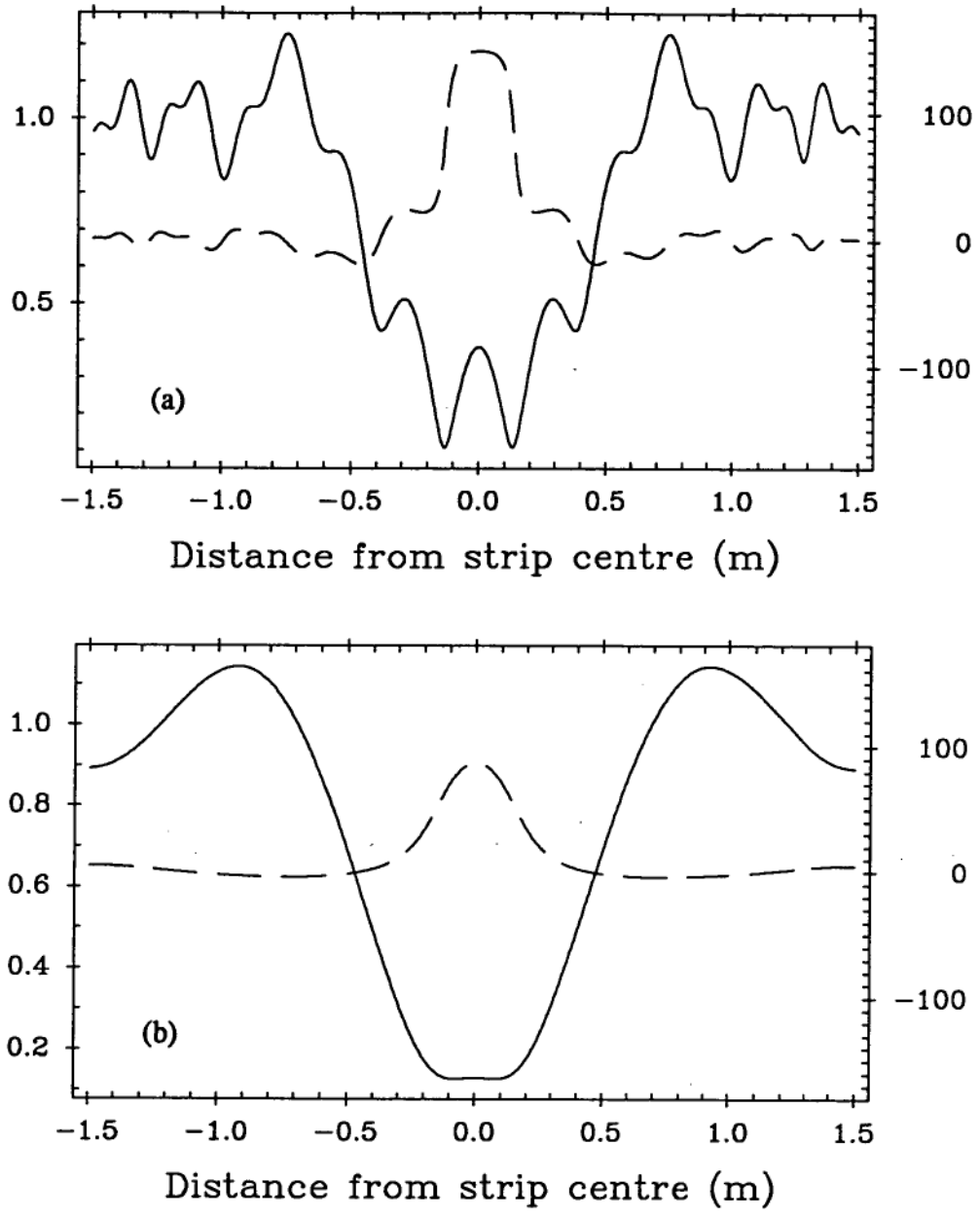


Figure 4.3: The diffraction pattern (a) of a plane wave 9 m beyond its normal incidence onto an infinitely long, conducting strip of width 0.73 m. The wavelength is 23.6 mm. In (b) the pattern has been convolved with the smoothing function corresponding to an aperture resolution $\delta_x = 0.59$ m. The solid line is the amplitude relative to the incident wave (left-hand scale); the dashed line is the phase (right-hand scale in degrees).

are so repeatable they may be characterized by direct measurement rather that computation.

4.1.2 Other Systematic Errors

Radiation Pattern Approximations

Two other electromagnetic processes occuring in a reflector antenna are diffraction by currents at its edge and propagation from the surface of a deep paraboloid to the aperture plane. As discussed in section 2.1, these do not enter the simple Fourier transform relationship used for holography, although a correction for the latter was suggested. They were ignored in the analysis of the Mount Pleasant data.

The Reconstruction Plane

Related to the depth of the reflector is the ambiguity in defining a single 'aperture plane' for the recovered currents. It was shown in section 2.3.2 that surface features above and below the chosen aperture plane are confused by a convolution before being represented in that plane. This is because detailed features on the wavefront are lost while propagating to the plane. The recovery plane used in the Mount Pleasant analysis was halfway between the levels of the reflector vertex and rim, or 5.55 m from the point W.

To estimate the error introduced by using only one plane, the analysis was repeated for a plane through the vertex and again for a plane through the rim, 3.82 m higher. The weighted rms aperture phase difference between the resulting maps was formed, giving for the four high-quality maps:

Map	Weighted rms Δ
C2	0.833°
C3	0.826°
B1	0.841°
B2	0.833°

The average rms difference of 0.83° should be compared with the repeatability data in table 3.2, for which a single recovery plane was used. Although a significant effect, propagation errors are not dominant.

The Feed Phase Response

Just as the feed illumination pattern dominated the aperture current amplitude maps, the feed phase pattern must be present in the aperture phase maps. It is unlikely that any feed presents a unique phase centre over the entire angle of illumination. Seen from the feed, the edge of the 26 m reflector is wide at 61° from the axis, and it is possible that a phase variation over the outer parts of the reflector caused a larger surface measurement error than the random errors shown in table 3.2. Some workers [51, 30] have measured the feed pattern in a separate experiment prior to holographic measurements, but this was not done at Mount Pleasant.

Polarization of the Source

Polarization has been mentioned in section 2.1, and usually it does not affect holographic measurements. For celestial sources, however, it can be important because the orientation of any source polarization will change with respect to the antenna during the course of observations. The result is a spurious astigmatism which can be minimized by repeating the measurements with the feed rotated by 90°, as described in [51].

Pointing Errors, Satellite Motion, and Calibrations

A constant pointing offset is not detrimental the the surface map, it simply causes a linear phase gradient which is removed in processing (section 3.2.2). Systematic errors from the ideal sampling grid have been mentioned already in section 2.3.1, and they can be minimized by judicious antenna scanning. Random pointing errors caused by wind and the drive system are studied in the next section.

Satellite motion can be accurately compensated using position predictions as well as on-source calibration measurements: see sections 2.3.2 and 3.2.1. The same measurements also give information about changes in signal path length, which may occur in the cabling due to temperature chages and cable motion. In addition, they compensate for slow changes in the signal level. The taking of on-source calibration samples is the single most effective technique for removing systematic errors.

Gravitational Deformations

Any large antenna is a massive structure, and the surface and feed support will change in shape under gravity as the antenna moves to different attitudes. Modern antennas are designed according to principles of homology, so that the shape remains a paraboloid with a varying focal length and axis orientation. The Mount Pleasant antenna was not designed this way, and its deformations are especially difficult to study because it has an XY mount. Any part of the reflector's edge may be closest to the ground so deformations depend on both axis angles, unlike an AZEL mount where they should be independent of azimuth.

This is not a problem of inaccuracies in a single holographic map, but a problem of interpretation of the map in the light of some theory of deformation. Maps for several antenna attitudes are necessary for a proper study. Because the deformations will be on large spatial scales they may be represented by low-resolution holographic maps, so a set of small maps would be sufficient and measurement time need not be excessive. Gain curve measurements also may give insight into deformations if they have a symmetry matching that of the antenna's mount.

There may, however, be small changes in the shape of the surface and feed support during the taking of a single map. These may result in changes in the focal length, for which the existing phase drift compensation will perform some remedy [51].

Cross-Coupling in the I.F. System

If the identical i.f. systems for the main and reference antennas are not perfectly isolated, a spurious output will occur in each channel proportional to the power in the other channel. In fact, it is suspected that this appeared as the dominant 'random' error in the Mount Pleasant measurements, even though it is not strictly random in origin. A discussion is deferred until section 4.3.2.

4.2 Random Errors

4.2.1 Origins and Categories

Random error or 'noise' cannot be predicted no matter how much is known about the system. It must be treated statistically. The dominant source of

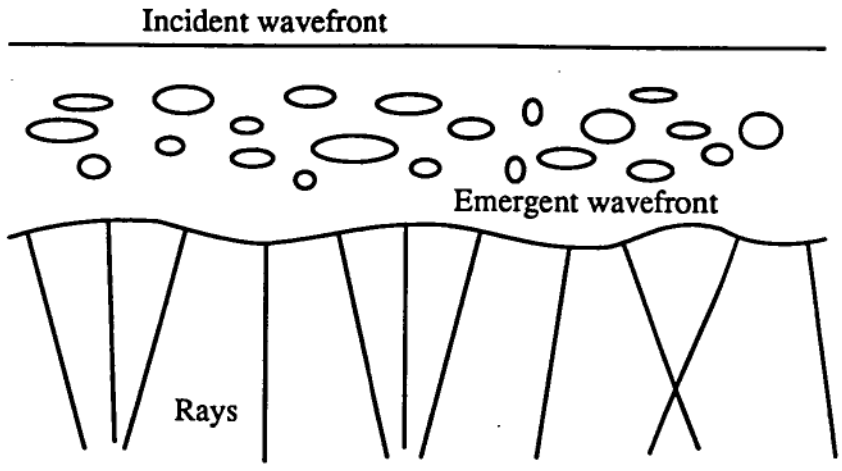


Figure 4.4: A disturbed wavefront resulting from a medium with irregular refractive properties, and the corresponding rays which converge or diverge (after figure 13.15 of Thompson, et al.).

noise in radio astronomy is the receiver, which produces a spurious output with a constant power per unit bandwidth. It is equivalent to the noise power available from a resistor kept at the 'noise temperature' of the receiver T_R, and great efforts are made to cool receivers and so minimize this power. For Microwave Holography using celestial sources receiver noise can limit the dynamic range of the measurements. A satellite, as will be seen, gives such a high antenna temperature that receiver noise can be less important than other sources of error.

Propagation through the neutral and ionized atmosphere affects the incoming signal in several ways. The refractivity of water vapour is much greater in the radio region than in the optical region, causing absorption and propagation delay. Variations in water vapour density rather than in temperature are the dominant cause of phase fluctuations across an initially plane wavefront. The rays corresponding to a disturbed wavefront will converge or diverge as illustrated in figure 4.4, and amplitude scintillation will result in addition to that caused by changes in absorption. Continuing the ray model, the angle of arrival will also fluctuate as the ray path varies. In the ionosphere the electron density causes absorption and refraction, although most of the associated effects scale with frequency as ν^{-2} and are small above several GHz or so. Propagation effects in the atmosphere are discussed extensively by Thompson, et al. [82, chap. 13] in the context of VLBI.

For a small baseline interferometer the rms phase difference due to the

neutral atmosphere increases with separation to a maximum at a distance of a few kilometers. Scale sizes in ionization irregularities are predominantly a few kilometers or less. Baselines used for holography are very small on these scales, so the signal at each antenna passes through almost identical pieces of atmosphere, and phase fluctuations are expected to be small. Angle-of-arrival variations may be more important; however, at 22 GHz they were estimated by Morris, et al. [51] to be of the order of 1" rms, which is less than the 2.9" pointing resolution of the 26 m antenna.

In order to quantify their effect on the recovered aperture currents, it is helpful to categorize these and other random errors according to their effect on the measured complex far field. The main distinction is between additive and multiplicative errors. Additive errors have the same expected magnitude at all points of the far field map. A multiplicative error such as a change in signal power has a smaller absolute magnitude when it occurs in a sidelobe, because the antenna output changes less than it would if the signal were received in the main beam. Thus an additive error should be much more detrimental to the aperture recovery than a multiplicative error which has the same expected magnitude when on-source.

Denote the true far field pattern by $A(u,v)$, and let a particular measured field be $A'(u,v)$ on a discrete grid of (u,v) points. Random errors will produce a different $A'(u,v)$ for every measurement. Four categories of error will be modelled as follows:

1. Additive errors. Sources are the receiver noise temperature and the quantization noise. At Mount Pleasant, at least 40 polarimeter U and V samples were averaged at each grid point, while the on-source level $|A(0,0)|$ was about 4000 counts. Thus the quantization noise was an insignificant 104 dB down or less. Receiver noise adds an independent error to each component of $A(u,v)$, so denote the error by $\varepsilon = \varepsilon_r + j\varepsilon_i$ and

$$A'(u,v) = A(u,v) + \varepsilon. \tag{4.1}$$

2. Amplitude errors. The satellite transponder's output power can be maintained by AUSSAT at a very steady level, so any substantial fluctuations in $|A(u,v)|$ above the receiver noise can be attributed to atmospheric scintillation. They are multiplicative in nature, so let ρ be a small real number and

$$A'(u,v) = A(u,v) + \rho A(u,v). \tag{4.2}$$

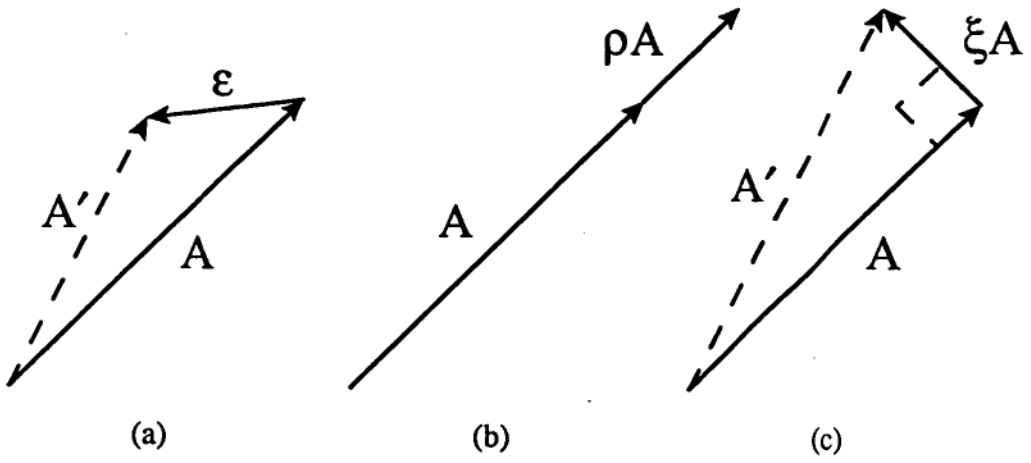


Figure 4.5: Representations in the complex plane of (a) a complex additive error, (b) an amplitude error, and (c) a phase error. A' is the measured value of the true far field A at some point.

3. Phase errors. Those due to the atmosphere are expected to be small, but there may also be non-systematic changes in the signal path length following the antennas (particularly in cables). A phase error is always multiplicative, and if its magnitude ξ is small

$$A'(u,v) = A(u,v)e^{j\xi} \simeq A(u,v) + j\xi A(u,v). \tag{4.3}$$

4. Pointing errors. The servo error of the drive system is one source, and under windy conditions this can be large. The angle-of-arrival fluctuations described above are also, in effect, pointing errors. If the error is μ in the u direction and ν in the v direction a first order Taylor approximation can be used to give

$$A'(u,v) = A(u+\mu, v+\nu) \simeq A(u,v) + \mu \left.\frac{\partial A}{\partial u}\right|_{(u,v)} + \nu \left.\frac{\partial A}{\partial v}\right|_{(u,v)}. \tag{4.4}$$

Representations in the complex plane of error types 1–3 appear in figure 4.5.

In this model the quantities ε_r, ε_i, ρ, ξ, μ, and ν all have zero-mean Gaussian statistics. The errors are statistically independent from point to point and from map to map. This may not be a realistic assumption in some cases; for example, the pointing servo error would be highly correlated between successive grid points. Nevertheless, the assumption makes finding the level of error in the aperture domain a tractable problem, and the results are useful for comparative studies as long as some caution is used. In the rest

of this section, the effect of each category of error on the recovered aperture phase is deduced, and the results are tested by a numerical simulation. This is far better than using a simulation to derive empirical formulae. Ultimately, in section 4.3, the data will tell whether this method is useful.

4.2.2 Propagation of Errors to the Aperture

Errors propagate to the estimated aperture current function through the discrete Fourier transform. When the errors are additive as in equation 4.1 it is usual to use noise power considerations to estimate aperture errors. For other types of error, however, this is not possible and a different analysis must be used.

Using the notation of section 2.2, consider a set of far field measurements at discrete points $u = m\delta_u$ and $v = n\delta_v$. The index ranges are $m = -L_u, \ldots, +L_u$ and $n = -L_v, \ldots, +L_v$, and the sampled and truncated far field map $\tilde{A}(m,n)$ is zero beyond this range. For a particular experiment, let the measured map differ from the true field by $\tilde{E}(m,n)$, giving a change of $\tilde{e}(x,y)$ in the recovered aperture currents $\tilde{a}(x,y)$:

$$\begin{aligned} \tilde{A}'(m,n) &= \tilde{A}(m,n) + \tilde{E}(m,n) \\ \tilde{a}'(x,y) &= \tilde{a}(x,y) + \tilde{e}(x,y). \end{aligned}$$

By the Fourier series expression for $\tilde{a}(x,y)$ in equation 2.26,

$$\tilde{e}(x,y) = \sum_{m=-\infty}^{m=+\infty} \sum_{n=-\infty}^{n=+\infty} \tilde{E}(m,n) e^{+j2\pi(\delta_u m x + \delta_v n y)}\, \delta_u\, \delta_v.$$

The autocorrelation theorem for Fourier series gives

$$|\tilde{e}(x,y)|^2 = {\delta_u}^2 {\delta_v}^2 \sum_{m'=-\infty}^{m'=+\infty} \sum_{n'=-\infty}^{n'=+\infty} \mathcal{C}\{\tilde{E}\}(m',n') e^{+j2\pi(\delta_u m' x + \delta_v n' y)} \tag{4.5}$$

where

$$\mathcal{C}\{\tilde{E}\}(m',n') = \sum_{m=-\infty}^{m=+\infty} \sum_{n=-\infty}^{n=+\infty} \tilde{E}^*(m,n) \tilde{E}(m+m', n+n'). \tag{4.6}$$

This is the autocorrelation function of the error at offset (m',n').

Each experiment will have a different error function $\tilde{E}(m,n)$. Consider the expected value of the aperture current error over all members of a large ensemble of experiments. Following, for example, MacDonald [44] each aperture point (x,y) has:

$$\left\langle |\tilde{e}(x,y)|^2 \right\rangle = {\delta_u}^2 {\delta_v}^2 \sum_{m'=-\infty}^{m'=+\infty} \sum_{n'=-\infty}^{n'=+\infty} \left\langle \mathcal{C}\{\tilde{E}\}(m',n') \right\rangle e^{+j2\pi(\delta_u m' x + \delta_v n' y)}.$$

Now the assumption that errors at different far field points are independent implies that the autocorrelation has expectation value zero for all but zero offsets:

$$\left\langle \mathcal{C}\{\tilde{E}\}(m',n') \right\rangle = 0 \text{ except when } m' = n' = 0.$$

This applies even when the error involves $A(u,v)$ in equations 4.2 to 4.4. The sum reduces to a single term, and

$$\left\langle |\tilde{e}(x,y)|^2 \right\rangle = {\delta_u}^2 {\delta_v}^2 \left\langle \mathcal{C}\{\tilde{E}\}(0,0) \right\rangle$$

or, taking the square root,

$$\operatorname{rms} |\tilde{e}| = \delta_u \delta_v \sqrt{\left\langle \mathcal{C}\{\tilde{E}\}(0,0) \right\rangle}.$$

The (x,y) dependence has been dropped: aperture current errors are constant across the aperture if the far field errors are uncorrelated between different points.

Additive and Multiplicative Errors

By equation 4.1 additive complex errors have the form $\tilde{E}(m,n) = \varepsilon$, so using equation 4.6

$$\left\langle \mathcal{C}\{\tilde{E}\}(0,0) \right\rangle = \left\langle \sum_{m=-\infty}^{m=+\infty} \sum_{n=-\infty}^{n=+\infty} |\varepsilon|^2 \right\rangle = N_u N_v \left\langle |\varepsilon|^2 \right\rangle$$

so

$$\operatorname{rms} |\tilde{e}| = \delta_u \delta_v \sqrt{N_u N_v} \operatorname{rms} |\varepsilon|. \tag{4.7}$$

Amplitude errors have the form $\tilde{E}(m,n) = \rho \tilde{A}(m,n)$, and now the sum becomes the norm of $\tilde{A}$:

$$\begin{aligned} \left\langle \mathcal{C}\{\tilde{E}\}(0,0) \right\rangle &= \left\langle \sum_{m=-\infty}^{m=+\infty} \sum_{n=-\infty}^{n=+\infty} |\rho \tilde{A}(m,n)|^2 \right\rangle \\ &= \left\langle \rho^2 \right\rangle \sum_{m=-\infty}^{m=+\infty} \sum_{n=-\infty}^{n=+\infty} |\tilde{A}(m,n)|^2 \\ &= \left\langle \rho^2 \right\rangle \|\tilde{A}\|^2. \end{aligned}$$

Taking the square root and using equation 2.29,

$$\operatorname{rms} |\tilde{e}| = \|\tilde{a}\| \operatorname{rms} \rho. \tag{4.8}$$

Phase errors, with $\tilde{E}(m,n) \simeq j\xi \tilde{A}(m,n)$, can be treated in exactly the same way, giving

$$\operatorname{rms} |\tilde{e}| \simeq \|\tilde{a}\| \operatorname{rms} \xi. \tag{4.9}$$

Pointing Errors

Pointing errors require a bit more work. By equation 4.4,

$$\tilde{E}(m,n) \simeq \mu \left.\frac{\widetilde{\partial A}}{\partial u}\right|_{(m,n)} + \nu \left.\frac{\widetilde{\partial A}}{\partial v}\right|_{(m,n)}$$

where the tilde over the derivative functions indicates their evaluation at discrete points. Thus

$$\begin{aligned}
\left\langle \mathcal{C}\{\tilde{E}\}(0,0)\right\rangle &\simeq \left\langle \mathcal{C}\left\{\mu\frac{\widetilde{\partial A}}{\partial u} + \nu\frac{\widetilde{\partial A}}{\partial v}\right\}(0,0)\right\rangle \\
&\simeq \left\langle \sum_{m=-\infty}^{m=+\infty}\sum_{n=-\infty}^{n=+\infty}\left|\mu \left.\frac{\widetilde{\partial A}}{\partial u}\right|_{(m,n)} + \nu \left.\frac{\widetilde{\partial A}}{\partial v}\right|_{(m,n)}\right|^2 \right\rangle \\
&= \left\langle \mu^2\right\rangle \left\|\frac{\widetilde{\partial A}}{\partial u}\right\|^2 + \left\langle \nu^2\right\rangle \left\|\frac{\widetilde{\partial A}}{\partial v}\right\|^2 . \qquad (4.10)
\end{aligned}$$

Here the cross terms have been eliminated using the independence of μ and ν, which means $\langle \mu^* \nu\rangle = \langle \nu^* \mu\rangle = 0$.

The Fourier transform relationship between $a(x,y)$ and $A(u,v)$, equation 2.25, will now be exploited. By the Derivative Theorem,

$$\left.\frac{\widetilde{\partial A}}{\partial u}\right|_{(m,n)} = \left.\frac{\partial A}{\partial u}\right|_{(m\delta_u, n\delta_v)} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}[-j2\pi x a(x,y)]e^{-j2\pi(\delta_u m x + \delta_v n y)}\,dx\,dy,$$

but $a(x,y) = 0$ beyond the aperture, so write

$$\delta_u\delta_v \left.\frac{\widetilde{\partial A}}{\partial u}\right|_{(m,n)} = \delta_u\delta_v \int_{-\frac{1}{2\delta_u}}^{+\frac{1}{2\delta_u}}\int_{-\frac{1}{2\delta_v}}^{+\frac{1}{2\delta_v}}[-j2\pi x a(x,y)]e^{-j2\pi(\delta_u m x + \delta_v n y)}\,dx\,dy.$$

The right hand side is a Fourier series integral, and the Parseval relation gives

$$\delta_u{}^2\delta_v{}^2 \left\|\frac{\widetilde{\partial A}}{\partial u}\right\|^2 = 4\pi^2\|xa\|^2.$$

Similarly,

$$\delta_u{}^2\delta_v{}^2 \left\|\frac{\widetilde{\partial A}}{\partial v}\right\|^2 = 4\pi^2\|ya\|^2.$$

Substituting these expressions into equation 4.10 gives the final result:

$$\operatorname{rms}|\tilde{e}| \simeq 2\pi\sqrt{\|xa\|^2 \operatorname{rms}^2\mu + \|ya\|^2 \operatorname{rms}^2\nu}. \qquad (4.11)$$

As for additive and multiplicative errors, the rms aperture current error due to pointing errors is expressed in terms of integrals of the aperture current. Note that in this case the integrals $\|xa\|$ and $\|ya\|$ will be strongly dependent on the aperture illumination function, decreasing rapidly as the edge taper increases.

4.2.3 The Signal-to-Noise Ratio

The level of random errors or noise is usually specified by the signal-to-noise ratio in the on-source position, denoted by $\mathcal{R}$. If there are several types of noise present their effects on the recovered aperture differ, so more than one signal-to-noise ratio must be specified. First, a strict definition of $\mathcal{R}$ is needed. Let the instantaneous error E add to the on-source response without noise A to produce a measured response A'. The signal-to-noise ratio will be defined by

$$\mathcal{R} = \frac{|A|}{\text{rms}\,|E|} = \frac{|A|}{\text{rms}\,|A' - A|}. \tag{4.12}$$

For additive and multiplicative random errors $\mathcal{R}$ can be derived from the rms values of quantities $|\varepsilon|$, ρ, and ξ. Consider equations 4.1 to 4.3:

1. Additive errors: $E = \varepsilon \Rightarrow \text{rms}\,E = \text{rms}\,|\varepsilon|$.
2. Amplitude errors: $E = \rho A \Rightarrow \text{rms}\,E = |A|\text{rms}\,\rho$.
3. Phase errors: $E \simeq j\xi A \Rightarrow \text{rms}\,E \simeq |A|\text{rms}\,\xi$.

The $\mathcal{R}$ due to each error in isolation, denoted by $\mathcal{R}_n$, $\mathcal{R}_a$, and $\mathcal{R}_p$ respectively, is thus

$$\mathcal{R}_n = \frac{|A|}{\text{rms}\,|\varepsilon|} \quad \mathcal{R}_a = \frac{1}{\text{rms}\,\rho} \quad \mathcal{R}_p = \frac{1}{\text{rms}\,\xi}\ . \tag{4.13}$$

In general all these errors will be present to a degree. The total error is given by

$$E \simeq \varepsilon + \rho A + j\xi A$$

from which it can be shown that

$$\frac{1}{\mathcal{R}_t^2} = \frac{1}{\mathcal{R}_n{}^2} + \frac{1}{\mathcal{R}_a{}^2} + \frac{1}{\mathcal{R}_p{}^2}. \tag{4.14}$$

This overall signal-to-noise ratio is not a very useful quantity because it cannot be related to the aperture current errors. A more sensible definition for $\mathcal{R}_t$ is the additive $\mathcal{R}_n$ which would be required to account for all aperture errors $\text{rms}\,|\tilde{e}|$ if $\mathcal{R}_a = \mathcal{R}_p = f_u = f_v = 0$. This will be followed up in section 4.3.3.

Before repeating the equations for $\text{rms}\,|\tilde{e}|$ in terms of these ratios, it is convenient to express pointing errors in a relative fashion too. Define the fractional pointing errors

$$f_u = \frac{\text{rms}\,\mu}{\delta_u} \text{ and } f_v = \frac{\text{rms}\,\nu}{\delta_v}. \tag{4.15}$$

Now the equations in section 4.2.3 can be written as follows:

$$\textit{Additive}: \quad \text{rms}\,|\tilde{e}| = \frac{\delta_u \delta_v \sqrt{N_u N_v}\,|A(0,0)|}{\mathcal{R}_n} \tag{4.16}$$

$$\textit{Amplitude}: \quad \text{rms}\,|\tilde{e}| = \frac{\|\tilde{a}\|}{\mathcal{R}_a} \tag{4.17}$$

$$\textit{Phase}: \quad \text{rms}\,|\tilde{e}| \simeq \frac{\|\tilde{a}\|}{\mathcal{R}_p} \tag{4.18}$$

$$\textit{Pointing}: \quad \text{rms}\,|\tilde{e}| \simeq 2\pi\sqrt{\|xa\|^2 \delta_u{}^2 f_u{}^2 + \|ya\|^2 \delta_v{}^2 f_v{}^2}. \tag{4.19}$$

4.2.4 Aperture Phase Errors

For practical purposes, the uncertainty in aperture phase is of interest. This is determined at each aperture point by the phase between $\tilde{a}$ and $\tilde{a}'$, which will vary with the magnitude and phase of the complex error $\tilde{e}$ as shown in figure 4.6. If the error were always perpendicular to $\tilde{a}$ the rms deviation of aperture phase would clearly be $\sigma_\phi = \text{rms}\,|\tilde{e}|/|\tilde{a}|$, using a small angle approximation. If instead the phase $\arg \tilde{e}$ is distributed uniformly in $[-\pi, +\pi]$ this is reduced by a factor $\sqrt{2}$, and so

$$\sigma_\phi = \frac{\text{rms}\,|\tilde{e}|}{\sqrt{2}\,|\tilde{a}|}. \tag{4.20}$$

In section 4.3 it is shown that a uniform phase distribution for $\tilde{e}$ can be assumed. Thus equations 4.16 to 4.19 can be used directly to find uncertainties in recovered aperture phase and hence in surface error. Section 2.4 shows how σ_ϕ affects the estimation of panel positions. While rms $\tilde{e}$ is constant over the aperture for random far field errors, σ_ϕ increases towards the aperture edge for a tapered illumination. This effect can be seen in table 3.2.

It will be helpful to find the average phase uncertainty over a region of the aperture. For example, this can be compared to the rms phase difference between independent maps in section 3.3.2. How should the current amplitude $|\tilde{a}(x,y)|$ be averaged to form the denominator of equation 4.20? A suitable method comes from equation 3.6 for the rms phase difference and the argument preceeding it. For continuous functions this time, the equation can be written

$$\text{rms}\,\Delta = \sqrt{\frac{\iint_{\text{ap}} |\tilde{a}(x,y)|^2 \Delta^2(x,y)\,dx\,dy}{\iint_{\text{ap}} |\tilde{a}(x,y)|^2\,dx\,dy}} = \sigma_{|\tilde{a}|\Delta} \sqrt{\frac{\iint_{\text{ap}} dx\,dy}{\iint_{\text{ap}} |\tilde{a}(x,y)|^2\,dx\,dy}}$$

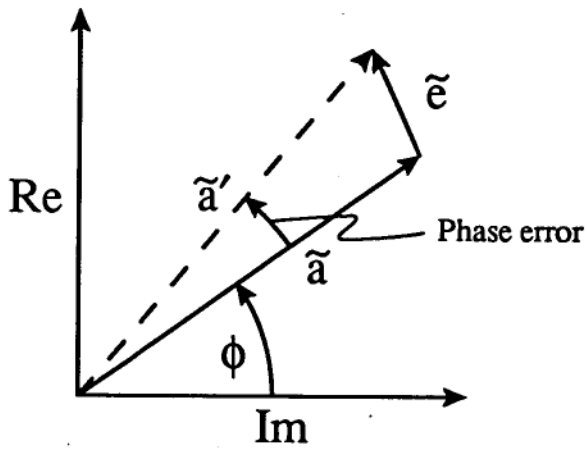


Figure 4.6: The uncertainty in aperture phase σ_ϕ is determined by the magnitude and direction of the complex error $\tilde{e}$.

where $\sigma_{|\tilde{a}|\Delta} = \sqrt{2}\,k$ with $\sigma_\phi = k/|\tilde{a}|$. But by equation 4.20 $k = \text{rms}\,|\tilde{e}|/\sqrt{2}$, and the final factor above is $1/\text{rms}\,|\tilde{a}|$. Thus

$$\text{rms}\,\Delta = \frac{\text{rms}\,|\tilde{e}|}{\text{rms}\,|\tilde{a}|}$$

can be compared directly with the results of section 3.3.2. To represent the uncertainty in aperture phase rather than phase difference, define

$$\overline{\sigma_\phi} = \frac{\text{rms}\,|\tilde{e}|}{\sqrt{2}\,\text{rms}\,|\tilde{a}|}. \tag{4.21}$$

This is the appropriate analogue of equation 4.20. The average can be taken over all or part of the aperture as desired.

The Uniform Illumination Case

To illustrate the way in which the equations for $\text{rms}\,|\tilde{e}|$ and σ_ϕ can be used, consider a circular aperture of radius r, uniformly illuminated with current a:

$$\tilde{a}(x,y) = a\,e^{j\phi(x,y)} \text{ for } x^2 + y^2 \leq r^2.$$

Assume that $\tilde{a}(x,y) \simeq a(x,y)$ (ignore the convolution) and that the phase profile $\phi(x,y)$ is always close to zero. Then the on-source field and the

aperture norms are

$$\begin{aligned} A(0,0) &\simeq \iint_{\text{ap}} a\,dx\,dy &&= \pi r^2 a \\ \|\tilde{a}\|^2 &= \delta_u\delta_v \iint_{\text{ap}} a^2\,dx\,dy &&= \delta_u\delta_v \pi r^2 a^2 \\ \|xa\|^2 &\simeq \delta_u\delta_v \iint_{\text{ap}} x^2a^2\,dx\,dy &&= \tfrac{1}{4}\delta_u\delta_v \pi r^4 a^2 \\ \|ya\|^2 &\simeq \delta_u\delta_v \iint_{\text{ap}} y^2a^2\,dx\,dy &&= \tfrac{1}{4}\delta_u\delta_v \pi r^4 a^2. \end{aligned}$$

For a circular aperture $\delta_u = \delta_v = \kappa/2r$. Making this substitution and using equations 4.16 through 4.20, the phase uncertainty due to each error type is

$$\textit{Additive}: \quad \sigma_\phi = \frac{\pi\kappa^2\sqrt{N_u N_v}}{4\sqrt{2}\,\mathcal{R}_n} \tag{4.22}$$

$$\textit{Amplitude}: \quad \sigma_\phi = \frac{\sqrt{\pi}\,\kappa}{2\sqrt{2}\,\mathcal{R}_a} \tag{4.23}$$

$$\textit{Phase}: \quad \sigma_\phi = \frac{\sqrt{\pi}\,\kappa}{2\sqrt{2}\,\mathcal{R}_p} \tag{4.24}$$

$$\textit{Pointing}: \quad \sigma_\phi = \frac{\pi\sqrt{\pi}\,\kappa^2 f}{4}. \tag{4.25}$$

Pointing errors have been assumed to be equal in u and v.

The first equation is the classical Scott and Ryle result [76]. It shows that the pointwise phase uncertainty increases for larger map sizes. This is because wide-angle far field measurements, while giving a higher density of points in the aperture, are increasingly dominated by the constant level of additive noise. On the other hand, multiplicative random errors have an effect which is independent of the map size. The same is true for pointing errors, but remember that σ_ϕ will be much smaller for a realistic, tapered illumination because of the nature of $\|xa\|$ and $\|ya\|$. Further on, in table 4.2, these uniform illumination estimates are compared with a full evaluation of the integrals for the 26 m antenna.

4.2.5 Numerical Evaluation and Simulation

Using a Discrete Fourier Transform

The aperture current estimate $\tilde{a}(x,y)$ is evaluated at discrete points $x = p\delta_x$ and $y = q\delta_y$ using an inverse discrete Fourier transform, forming $a_{\text{DFT}}(p,q)$. For an error analysis it is necessary to have equations for rms $|e_{\text{DFT}}|$. By

equation 2.28, where the inverse DFT includes a normalization by $1/N_u N_v$,

$$\tilde{a}(x,y) = \delta_u \delta_v N_u N_v \, a_{\mathrm{DFT}}(p,q)$$

Clearly

$$\mathrm{rms}\,|\tilde{e}(x,y)| = \delta_u \delta_v N_u N_v \,\mathrm{rms}\,|e_{\mathrm{DFT}}(p,q)|.$$

The norm of $\tilde{a}$ can be estimated by approximating the integral by sums over discrete x and y:

$$\begin{aligned}
\|\tilde{a}\|^2 &= \delta_u \delta_v \int_{-\frac{1}{2\delta_u}}^{+\frac{1}{2\delta_u}} \int_{-\frac{1}{2\delta_v}}^{+\frac{1}{2\delta_v}} |\tilde{a}(x,y)|^2 \, dx \, dy \\
&\simeq \delta_u \delta_v \sum_{p=-L_u}^{p=+L_u} \sum_{q=-L_v}^{q=+L_v} |\tilde{a}(p\delta_x, q\delta_y)|^2 \, \delta_x \, \delta_y \\
&= {\delta_u}^3 {\delta_v}^3 {N_u}^2 {N_v}^2 \delta_x \delta_y \sum_{p=-L_u}^{p=+L_u} \sum_{q=-L_v}^{q=+L_v} |a_{\mathrm{DFT}}(p,q)|^2 \\
&= {\delta_u}^2 {\delta_v}^2 N_u N_v \|a_{\mathrm{DFT}}\|^2.
\end{aligned}$$

The terms were simplified using $N_u \delta_u \delta_x = N_v \delta_v \delta_y = 1$. To proceed further an assumption is required: that for the purposes of noise calculation $a(x,y) \simeq \tilde{a}(x,y)$, ignoring the convolution discussed in section 2.2.2. Unless the recovered currents themselves are badly corrupted, this has a negligible effect on the estimated error. Then

$$\begin{aligned}
\|xa\|^2 \simeq \|x\tilde{a}\|^2 \simeq {\delta_u}^2 {\delta_v}^2 N_u N_v {\delta_x}^2 \|pa_{\mathrm{DFT}}\|^2 \\
\|ya\|^2 \simeq \|y\tilde{a}\|^2 \simeq {\delta_u}^2 {\delta_v}^2 N_u N_v {\delta_y}^2 \|qa_{\mathrm{DFT}}\|^2.
\end{aligned}$$

Substitution into equations 4.16 to 4.19 yields

$$\textit{Additive}: \quad \mathrm{rms}\,|e_{\mathrm{DFT}}| \simeq \frac{|A(0,0)|}{\sqrt{N_u N_v}\,\mathcal{R}_n} \tag{4.26}$$

$$\textit{Amplitude}: \quad \mathrm{rms}\,|e_{\mathrm{DFT}}| \simeq \frac{\|a_{\mathrm{DFT}}\|}{\sqrt{N_u N_v}\,\mathcal{R}_a} \tag{4.27}$$

$$\textit{Phase}: \quad \mathrm{rms}\,|e_{\mathrm{DFT}}| \simeq \frac{\|a_{\mathrm{DFT}}\|}{\sqrt{N_u N_v}\,\mathcal{R}_p} \tag{4.28}$$

$$\textit{Pointing}: \quad \mathrm{rms}\,|e_{\mathrm{DFT}}| \simeq 2\pi \sqrt{\frac{\|pa_{\mathrm{DFT}}\|^2 {f_u}^2}{{N_u}^3 N_v} + \frac{\|qa_{\mathrm{DFT}}\|^2 {f_v}^2}{N_u {N_v}^3}}. \tag{4.29}$$

The expressions for aperture phase uncertainty translate directly to the discrete case:

$$\sigma_\phi = \frac{\mathrm{rms}\,|e_{\mathrm{DFT}}|}{\sqrt{2}\,|a_{\mathrm{DFT}}|} \tag{4.30}$$

and

$$\overline{\sigma_\phi} = \frac{\text{rms}\,|e_{\text{DFT}}|}{\sqrt{2}\,\text{rms}\,|a_{\text{DFT}}|}, \tag{4.31}$$

where

$$\text{rms}\,|a_{\text{DFT}}| = \sqrt{\frac{\sum_{\text{ap}} |a_{\text{DFT}}(p,q)|^2}{\sum_{\text{ap}} 1}}.$$

4.2.6 Testing by Simulation

The usefulness of equations 4.26 to 4.29 depends on the applicability of the noise models and the practicality of isolating each signal-to-noise ratio. Their correctness can be tested by numerical simulation.

Given a simulated complex far field on a grid of map points, it is simple to use computer-generated random numbers to add errors of the additive, amplitude, and phase types. To include a pointing error it is necessary to sample the simulated field at a much closer spacing than the Nyquist interval λ/D. This can be done by Fourier interpolation. Far field samples at points adjacent to the correct sample positions are chosen to represent pointing errors in discrete steps, the direction and number of steps being selected randomly. This scheme leads to a program with the following outline:

1. Produce an aperture function in an array of complex numbers.
2. Enlarge the array with zeros to many times its size.
3. Compute the far field pattern by a forward DFT: the 'padded' array will cause samples to be interpolated for pointing error simulation.
4. Generate a set of random Gaussian deviates, and use these to add a chosen type of error to the far field pattern.
5. Compute the 'recovered' aperture function by an inverse DFT, and compare this with the original function.
6. Repeat the last two steps many times to accumulate the rms error at each aperture point.
7. Display the error statistics and compare with those predicted by the appropriate equation.

For simpler programming and display, the simulation can be performed in one dimension. Equations for $\text{rms}\,|e_{\text{DFT}}|$ are obtained by setting $N_v = 1$ in

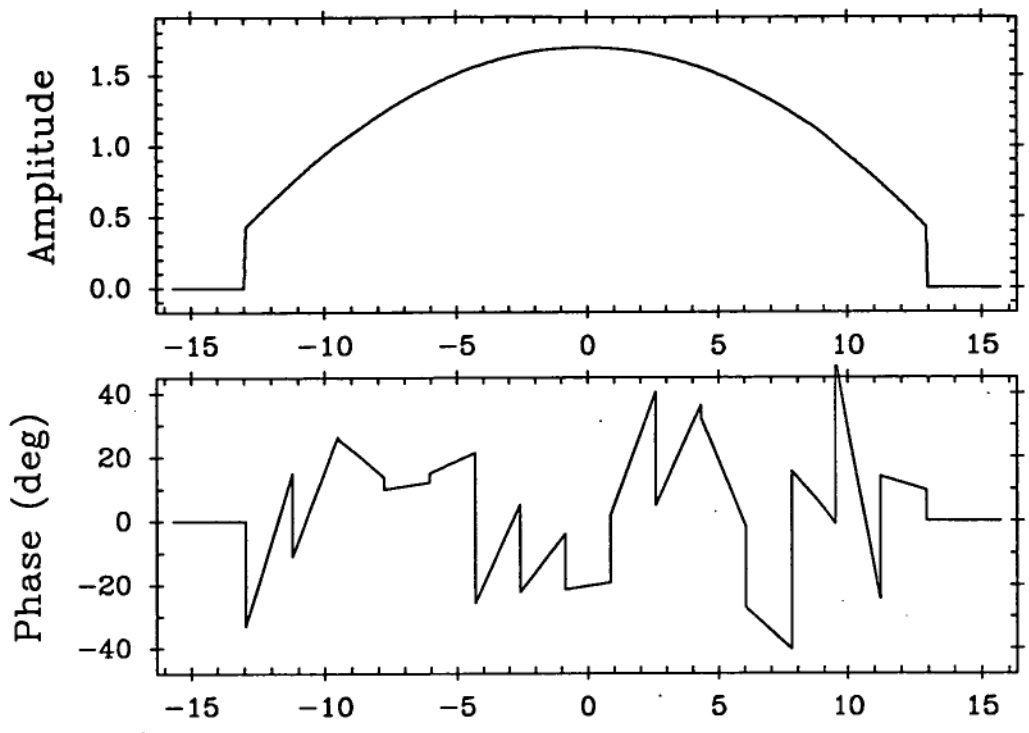


Figure 4.7: A simulated one-dimensional aperture current function. The illumination taper is −12 dB and the weighted rms phase is 18.9°. The horizontal scale is in metres.

the two dimensional equations. This is just as effective a test as a full 2D simulation.

The numerical experiment had these parameters, chosen to match the carrier maps described in chapter 3:

wavelength:	$\lambda = 23.6$ mm
diameter:	$D_x = 25.9$ m
sampling:	$\delta_u = 7.50 \times 10^{-4}$ giving $\kappa = 0.825$
data size:	$N_u = 53$ giving $\delta_x = 0.593$ m.

An aperture current function was produced with an arbitrary phase profile in the form of 15 linear segments to represent panels. The aperture illumination was a 'parabola on a box' of the form

$$|\tilde{a}(x)| = 1 - c\left(\frac{2x}{D_x}\right)^2 \text{ where } c = 1 - 10^{T/20}. \tag{4.32}$$

The edge taper T was set to −12 dB, and the weighted rms phase was 18.9°. The amplitude and phase profiles are shown in figure 4.7.

Each category of noise was investigated separately, with equal signal-to-noise ratios $\mathcal{R}_n = \mathcal{R}_a = \mathcal{R}_p = 100$ and $f_u = 0.048$ for pointing errors ($f_v = 0$ of course). Steps 4 and 5 were repeated 1000 times to accumulate statistics on $e_{\text{DFT}}(p)$ at each aperture point $x = p\delta_x$. Figure 4.8 shows the resulting aperture phase uncertainty σ_ϕ as a function of fractional radius $2x/D_x$. The predicted uncertainty is graphed, and matches the simulation well. To demonstrate the validity of equation 4.30 for σ_ϕ, the rms phase of $e_{\text{DFT}}(p)$ is shown in figure 4.9. If $\arg e_{\text{DFT}}(p)$ does indeed follow a uniform distribution in $[-\pi, +\pi]$ its rms would be $\pi/\sqrt{3}$ or 104°. This level is graphed, and again the agreement is good.

It is important that the predictions are reliable for any experiment. The wavelength, diameter, and sampling parameters were changed one-by-one and the simulation repeated, and the predictions remained of a similar quality. Equations 4.16 to 4.19, and their discrete analogues 4.26 to 4.29, are thus reliable predictors for random errors in Microwave Holography or any experiment where a function is estimated by measurements of its Fourier transform.

4.3 Measuring the Random Errors

In this section it is shown how the signal-to-noise ratios and pointing errors of the previous section were estimated by monitoring the on-source signal. An assumption required is that the random errors in $A'(u, v)$ do not depend on the antenna's position except as indicated by the defining equations of the error categories, 4.1 to 4.4. That is, all effects of scanning the antenna must be accounted for as systematic errors. This is not entirely possible: for example, pointing errors while driving at a constant rate are difficult to measure, and cable motion may be important to the phase. In the end, however, comparisons with the repeatability of the Mount Pleasant maps justify this approach by showing that the important errors are adequately predicted.

The on-source signal was monitored by sampling it for unbroken periods of several minutes. Holography samples were spaced unevenly in time: they were almost contiguous while scanning but between scans were long gaps. It is assumed that in either case the random processes behind the quantities ε_r, ε_i, ρ, ξ, μ, and ν produced no correlation: that is, their spectra were 'white'. Thus the on-source data had to be examined for drifts and correlated

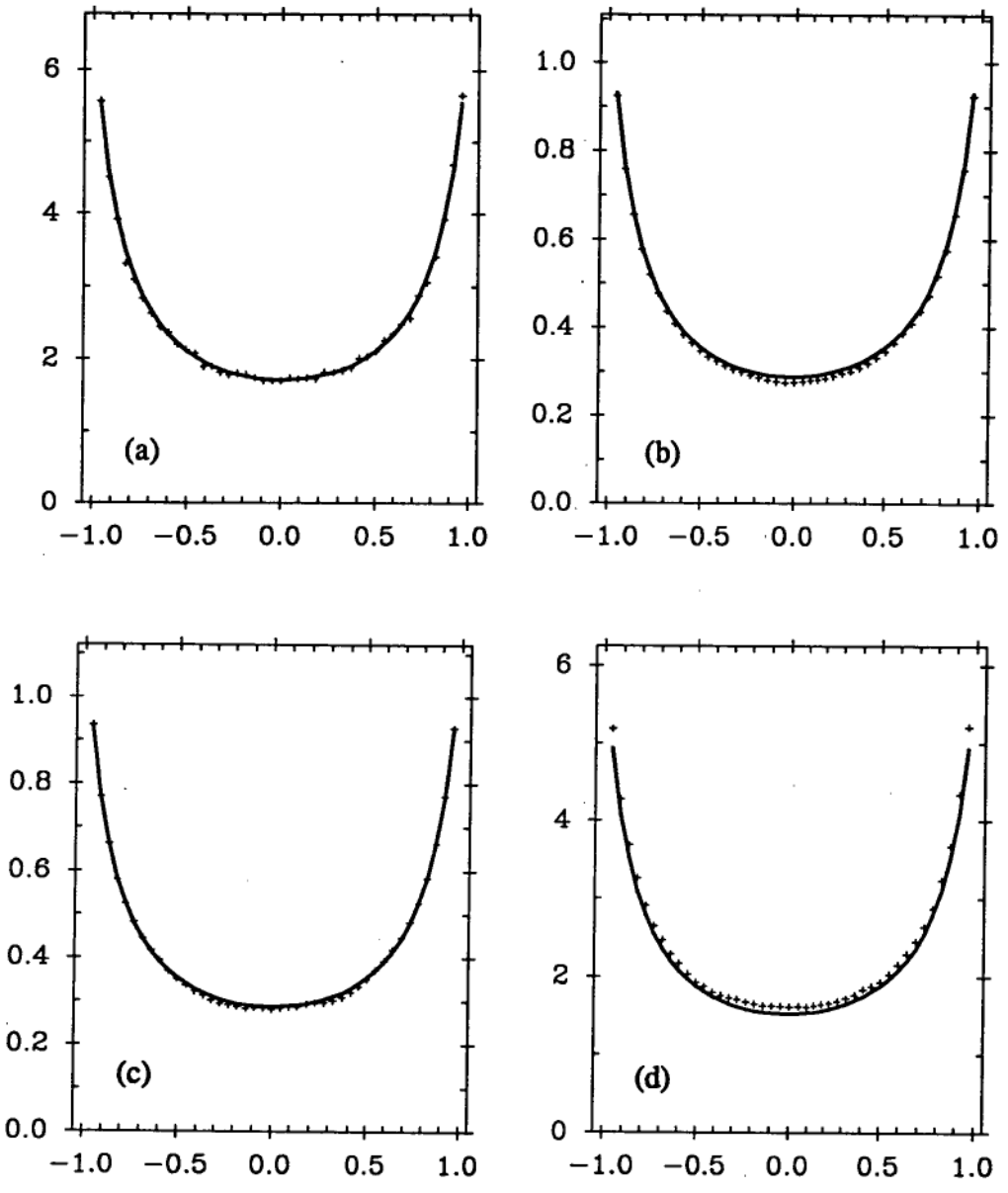


Figure 4.8: A numerical investigation of the effects of four categories of random error: (a) additive, (b) amplitude, (c) phase, and (d) pointing. The aperture phase uncertainty is graphed in degrees as a function of fractional radius $2x/D_x$. The predicted uncertainty is graphed for comparison.

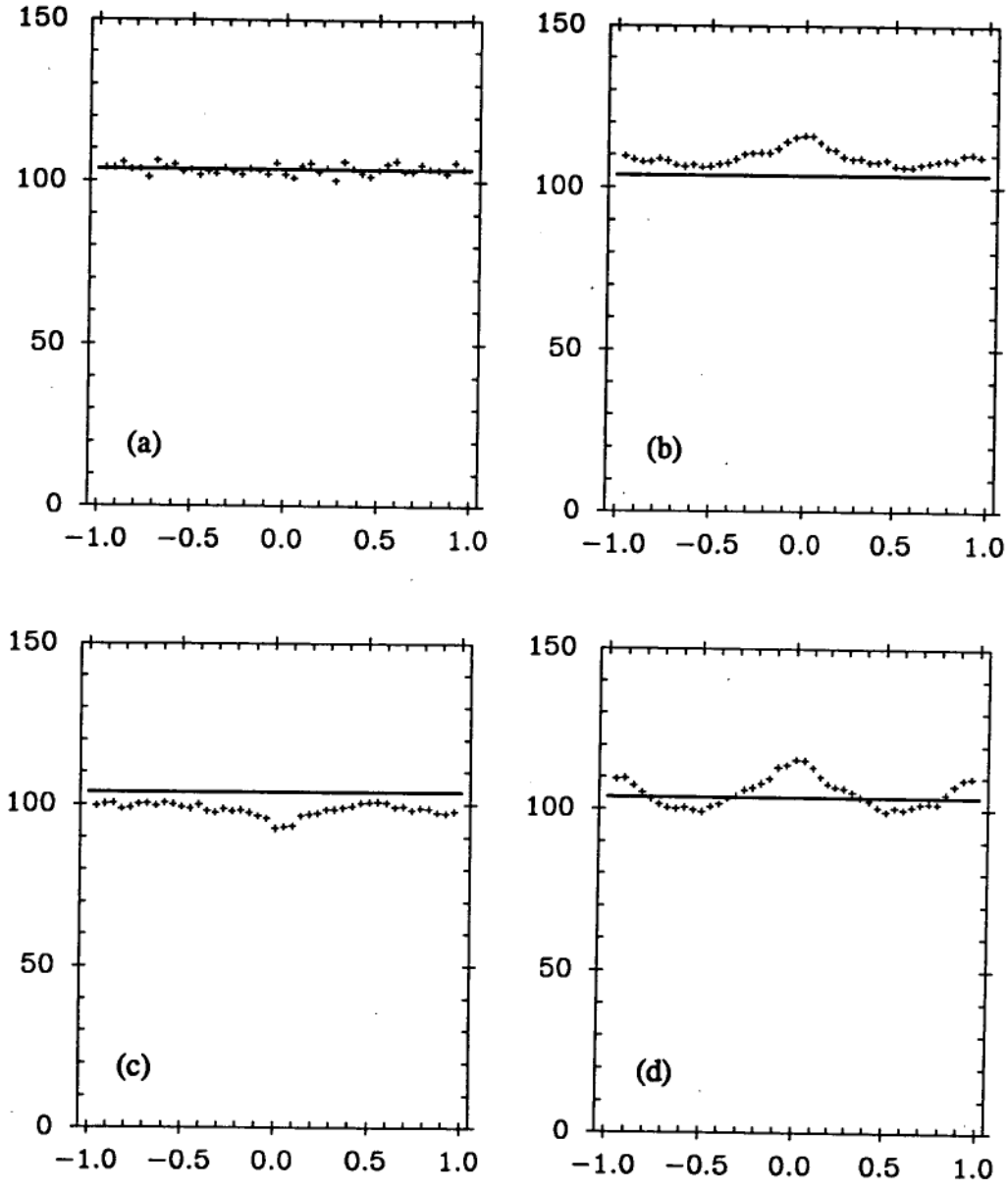


Figure 4.9: The rms phase of the error phasor $e_{\mathrm{DFT}}(p)$ for the numerical simulations in the previous figure, graphed in degrees as a function of fractional radius $2x/D_x$. The level of 104° expected for a uniform phase distribution is graphed for comparison.

variations, and these explained and removed before estimating noise levels.

4.3.1 The Data and Pointing Errors

Several data runs were taken, each of 1024 samples at 1 s intervals, or 17 min in length. Since scans making up a map began roughly 4.5 min apart, the runs included frequencies low enough to be well-resolved by the inter-scan calibration points. Most maps had a 2.5 s integration time for each grid position, so the runs included all frequencies that weren't smoothed out by this time constant. Each of the 1024 samples had an integration time of 0.75 s: this would have governed the envelope of the noise power spectrum. However, it was not a critical parameter because, as will be seen, fluctuation levels were not principally determined by the integration time.

Three runs were recorded while observing the 12725 MHz carrier: one directly on-source, one slightly off at 2 dB down from the beam maximum, and one 3° off-source. This is the position used for zero level calibrations between scans. The intention was that the steep slope at the sides of the main beam would make the output at 2 dB down very sensitive to pointing errors, giving an estimate of their magnitude. Multiplicative errors would be most important in the on-source data, while the off-source data, consisting only of the additive errors remaining when $A(u,v) \simeq 0$, would measure the receiver noise. A similar set of three runs was recorded for the 12373 MHz beacon, but only the carrier data will be displayed here.

Figure 4.10 shows the cosine and sine outputs ($\mathrm{Re}\,\mathcal{A}$ and $\mathrm{Im}\,\mathcal{A}$) calculated from U and V as a function of time for the on-source run. The corresponding amplitude and phase are also shown. The cosine and sine are dominated by a slow phase drift, but the amplitude has a clear periodic feature of roughly 220 s period. This was mirrored exactly by the I_x (26 m power) output. The rms deviation of the amplitude is 80 counts on an average level of 4390 counts, or 1.8%.

The 2 dB down run is shown in the same way in figure 4.11. It shows a much stronger periodicity in amplitude: the rms is 219 counts on an average of 3473 counts, or 6.3%. This suggests that the cause is a cyclic pointing error in both runs, which is reasonable behaviour for a servo-controlled drive system. A considerable range of periods would occur, so the the shorter 175 s period in the second run is not surprising. Nearly the entire amplitude rms is due to the cycle, so, in conjunction with the slope of the beam at the point where data were taken, it gives the pointing error. The slope was estimated

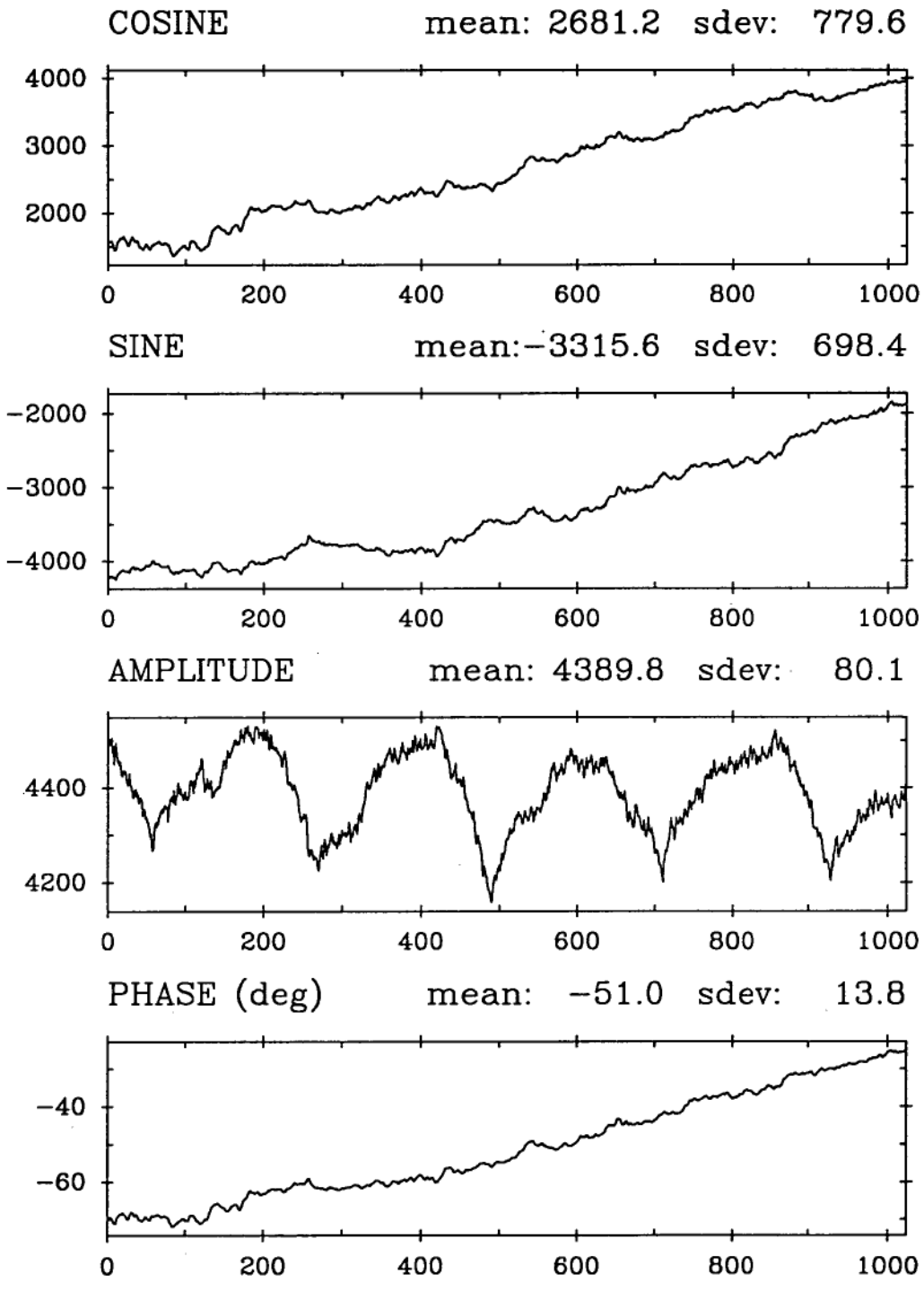


Figure 4.10: Cosine output, sine output, amplitude, and phase as a function of time for the 1024 s on-source data run at 12725 MHz. Units for the vertical axes are counts and degrees.

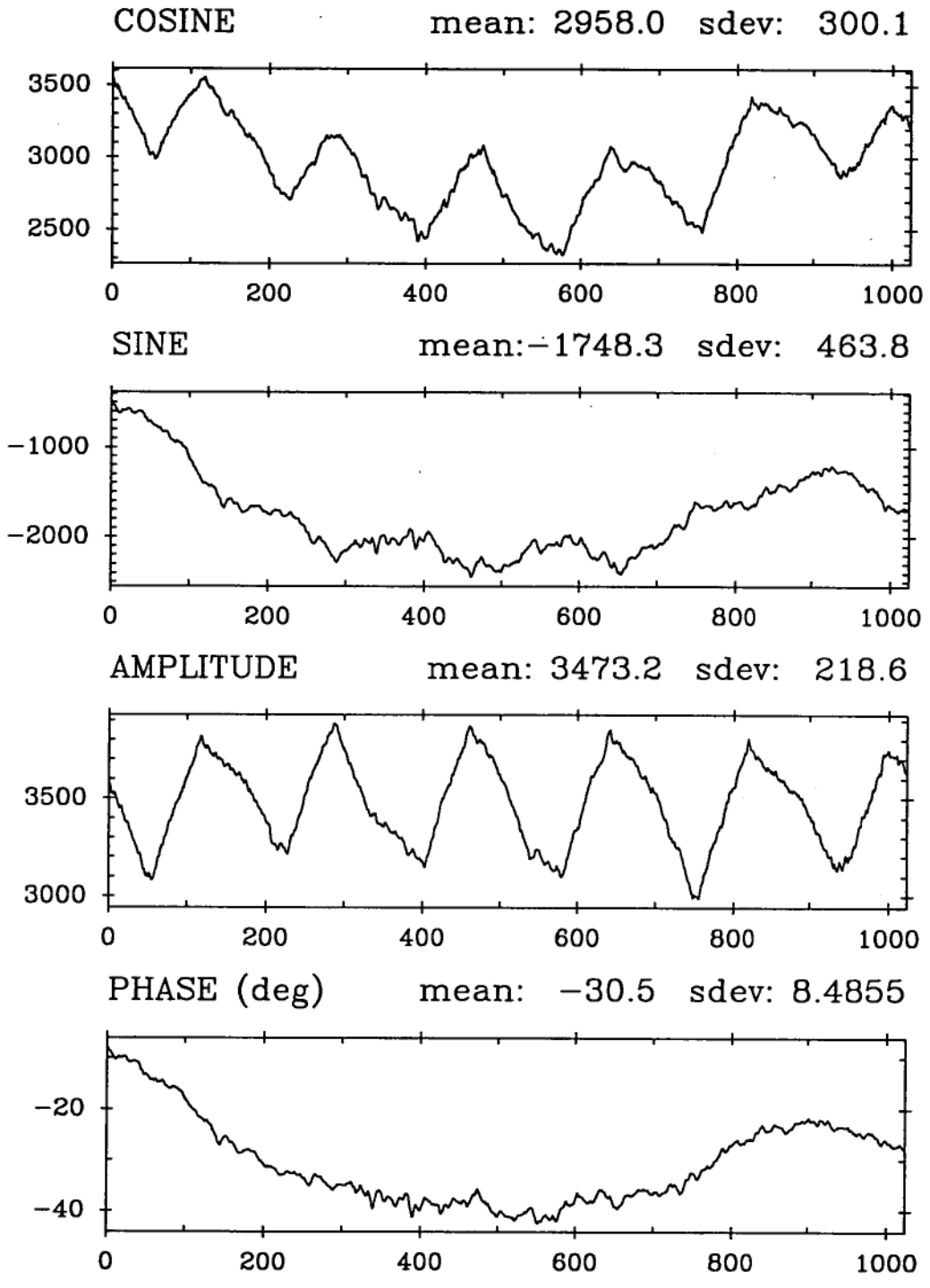


Figure 4.11: Cosine output, sine output, amplitude, and phase as a function of time for the data run taken 2 dB down from the peak of the main beam at 12725 MHz. Units for the vertical axes are counts and degrees.

at 12725 MHz and 12373 MHz from the beam maps in figures 3.18 and 3.19. Summarizing the results at these frequencies:

	12725 MHz	12373 MHz
Amplitude variation	-2.05 ± 0.55 dB	-2.02 ± 0.45 dB
Slope of beam	$0.0055 \pm 0.0005°$/dB	$0.0064 \pm 0.0006°$/dB
Pointing error	$\pm 0.0030°$	$\pm 0.0029°$

The pointing error was similar for both 2 dB down runs. If it is assumed that the same pointing error will exist at right-angles to the gradient, then

$$\text{rms}\,\mu = \text{rms}\,\nu \simeq 5.1 \times 10^{-5} \Rightarrow f_u = f_v \simeq 0.069.$$

These can be used to predict aperture domain errors with two fairly strong suppositions: that the servo system introduces similar errors when the antenna is in motion, and that the correlation of pointing errors at successive grid points can be ignored. Clearly the results should be treated with caution, but at least they provide a handle on the magnitude of the effect. They will be given later.

The problem now remains of separating the pointing error effect from the on-source amplitude variations. A clue is had to the nature of the low frequency drift or 'signal' as distinct from the 'noise' by plotting the power spectral density (PSD) on a logarithmic power scale. The result is shown separately for the amplitude and phase of the on-source data in figure 4.12. It is legitimate to consider these independently because fluctuations in amplitude and phase are supposed to arise from different atmospheric and instrumental causes. It can be seen that at very low frequencies the almost-linear noise spectra disappear beneath a sharply rising feature. This is assumed to contain pointing errors and other drifts which are resolved by inter-scan calibration points. In order to find the noise rms in each spectrum, a filter must be designed to remove this feature.

'Wiener' filtering is a suitable technique when approximate models are available for the signal and noise components of a spectrum [61, sect. 12.6]. If $|S(f)|^2$ is the signal PSD and $|N(f)|^2$ is the noise PSD, the Wiener filter is given by

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

This can be multiplied by the spectrum to isolate the signal. It does not require accurate representations of the noise and signal spectra. It simply provides a sensible interpolation between the extreme cases $\Phi(f) = 0$ when

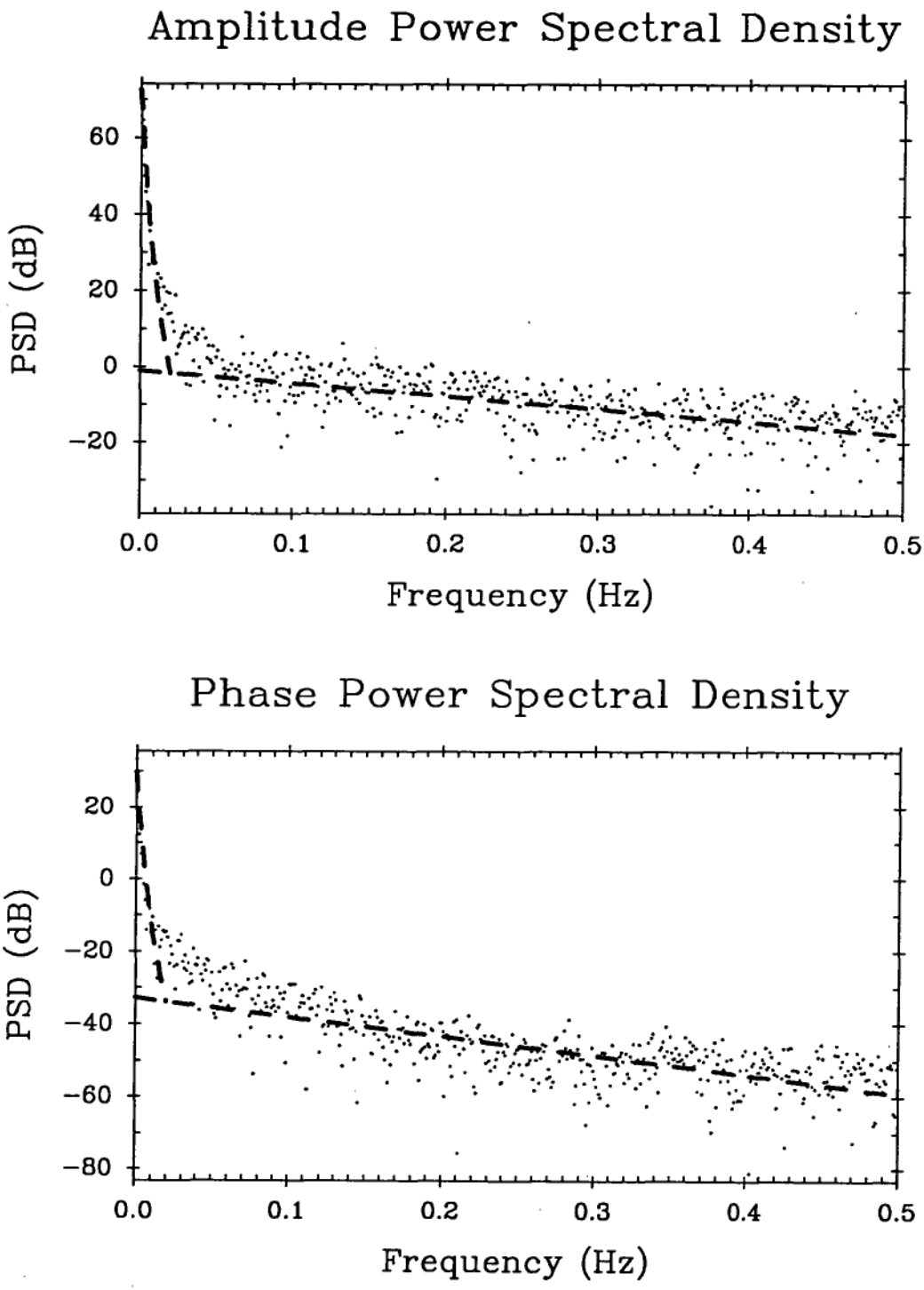


Figure 4.12: Power spectral density of the on-source amplitude and phase data. The frequency resolution is 0.98 mHz, determined by the data length of 1024 s. The 'signal + noise' and 'noise' models are shown as dotted lines.

noise dominates and $\Phi(f) = 1$ when the signal dominates: one which minimizes the rms difference of the filter output from the uncorrupted signal. A conservative cut-off frequency helps to prevent noise being mistaken as signal, which would cause the noise rms to be under-estimated. From figure 4.12 a cut-off of 19.5 mHz (20 bins) was decided upon, and an hyperbola fitted to the signal spectrum to meet the linear noise spectrum at this frequency. These models are shown on the plots. The robustness of the technique was tested by increasing the cut-off frequency to 29.3 mHz (30 bins). At most, this decreased the noise rms by 25%, usually much less; this is an acceptable level of accuracy for a comparative study. On the other hand, a cut-off of 9.8 mHz (10 bins) failed to remove the periodic pointing drifts.

Figure 4.13 shows the remaining 'drift' components after applying this filter to the amplitude and phase spectra. Subtracting these from the original data gives the 'noise' components which are also shown. Separate rms deviations σ can be found for each graph. The noise components are free from pointing effects, and σ_n is probably dominated by amplitude and phase multiplicative errors. Before this is assumed the off-source data must be examined for the level of additive noise.

4.3.2 Additive and Multiplicative Errors

Slow drifts should be removed from the off-source data as well, because zero levels are checked before each scan of a holography map. Two cut-off frequencies were tried, one on either side of that used for the on-source data, and the rms noise deviations compared to see if there was a substantial change. It is appropriate to treat the cosine and sine outputs separately for additive noise, since the U and V outputs should be independent for a complex correlator subjected to thermal noise [82, p. 161]. The rms deviations in count units for the off-source runs were:

	Cut-off 9.8 mHz	Cut-off 29.3 mHz
Cosine, 12725 MHz	4.30	4.00
Sine, 12725 MHz	3.50	3.29
Cosine, 12373 MHz	4.44	3.76
Sine, 12373 MHz	4.25	3.89

These values are reasonably steady, so a consensus of 4.0 counts was chosen:

$$\sigma_r = \text{rms}\,\varepsilon_r \simeq 4.0$$
$$\sigma_i = \text{rms}\,\varepsilon_i \simeq 4.0$$

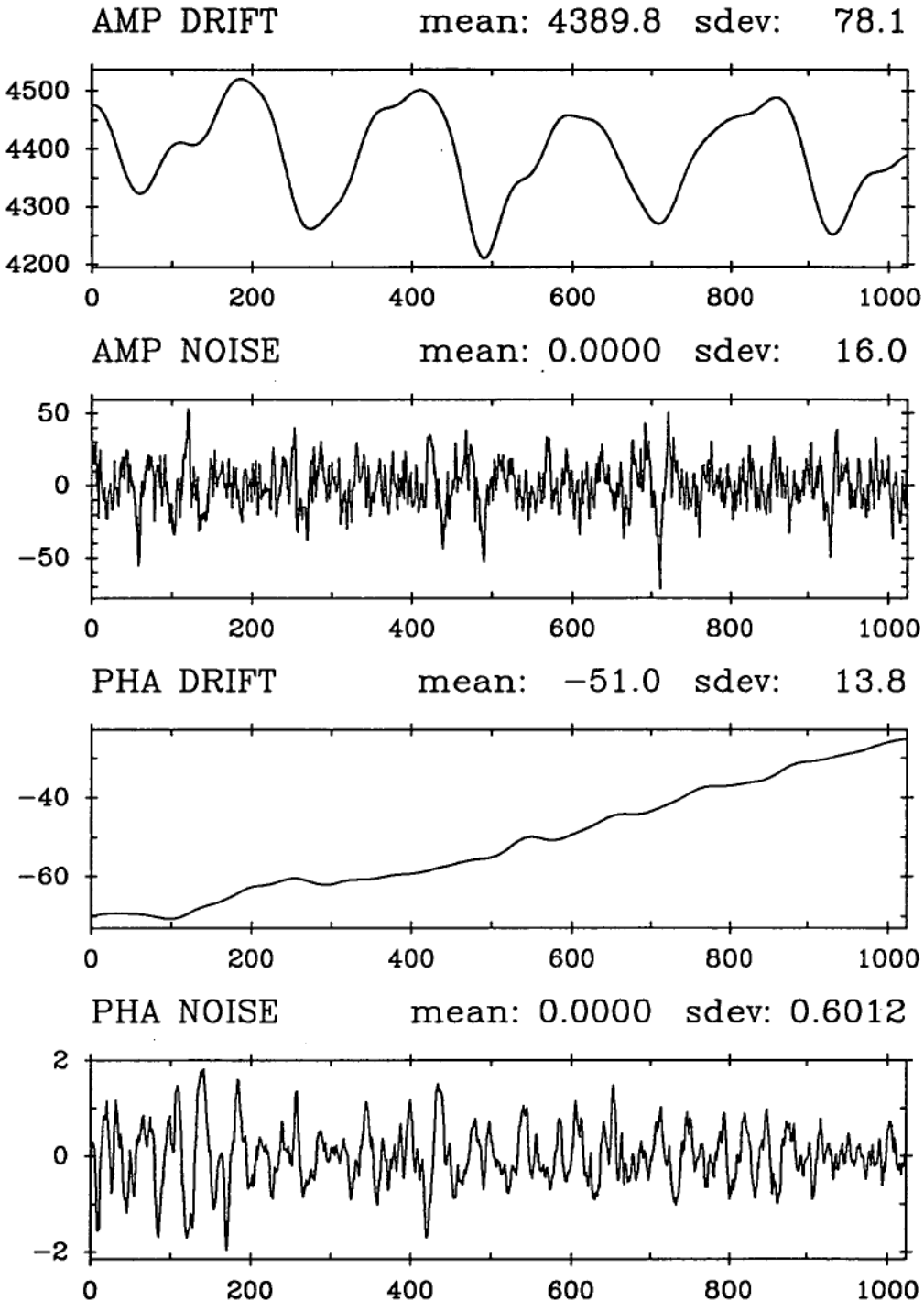


Figure 4.13: Amplitude drift and noise, then phase drift and noise for the 12725 MHz on-source data run obtained by Wiener filtering the PSD with a cut-off frequency of 19.5 mHz. Units for the vertical axes are counts and degrees.

where $\varepsilon = \varepsilon_r + j\varepsilon_i$ is the complex error (equation 4.1). Figure 4.14 shows the drift and noise components at 12725 MHz for the sharper filter.

A question arises: with both carrier and beacon signals amplified to give similar output levels, why was the receiver noise not much greater in proportion for the weaker beacon signal? Another strange thing is that the cosine and sine outputs can be seen to have a very high, negative correlation. This is not the correct behaviour for receiver noise. In fact, the receiver noise is expected to be at a lower level by the following argument: even the beacon signal would give an antenna temperature T_A much greater than the system temperature T_S. The minimum detectable signal for a correlator with 30 kHz pre-detection bandwidth followed by a 0.75 s integration is

$$\Delta T_A = \frac{\sqrt{2}\,T_S}{\sqrt{30 \times 10^3 \times 0.75}} = 0.0094\,T_S \ll 0.0094\,T_A.$$

Now $T_A \propto 3415$ counts for the on-source beacon signal, giving a noise level of $\Delta T_A \ll 32$ counts. Thus it is likely that the receiver noise was well below the 4.0 count level obtained above.

The correlation of the outputs suggests that cross-coupling in the i.f. chain might be the cause of the signal: even when off-source the reference channel remains at full strength. The isolation of the 10.7 MHz board in figure 3.2(a) was measured by applying a signal to one input and observing the ratio of the direct output to the output of the other channel. It was found that:

$$\begin{aligned} \text{Isolation } A \longrightarrow R: &\quad 55 \text{ dB} \\ \text{Isolation } R \longrightarrow A: &\quad 47 \text{ dB}. \end{aligned}$$

All on-source and off-source data runs were recorded with the 15 dB A attenuator switched in; measurement gave the gain difference of the two channels as 13.5 dB in this case. To give similar on-source outputs the R signal was thus 13.5 dB down from the A signal, and the level of cross-power in the A channel was 60.5 dB down. For on-source levels of roughly 4000 counts this corresponds to 3.8 counts, which is close to the observed off-source rms and would be the same for carrier and beacon measurements.

There are difficulties with interpreting a cross-coupling effect as a random error. A simple model for non-reactive cross-coupling has

$$\begin{aligned} a(t) &= A\cos(\omega t + \phi) + X_R R\cos(\omega t) \\ r(t) &= R\cos(\omega t) + X_A A\cos(\omega t + \phi) \end{aligned}$$

for the signals reaching the polarimeter, instead of the simpler equations in section 3.1.5. Here $X_A = -55$ dB and $X_R = -47$ dB. Neglecting terms of

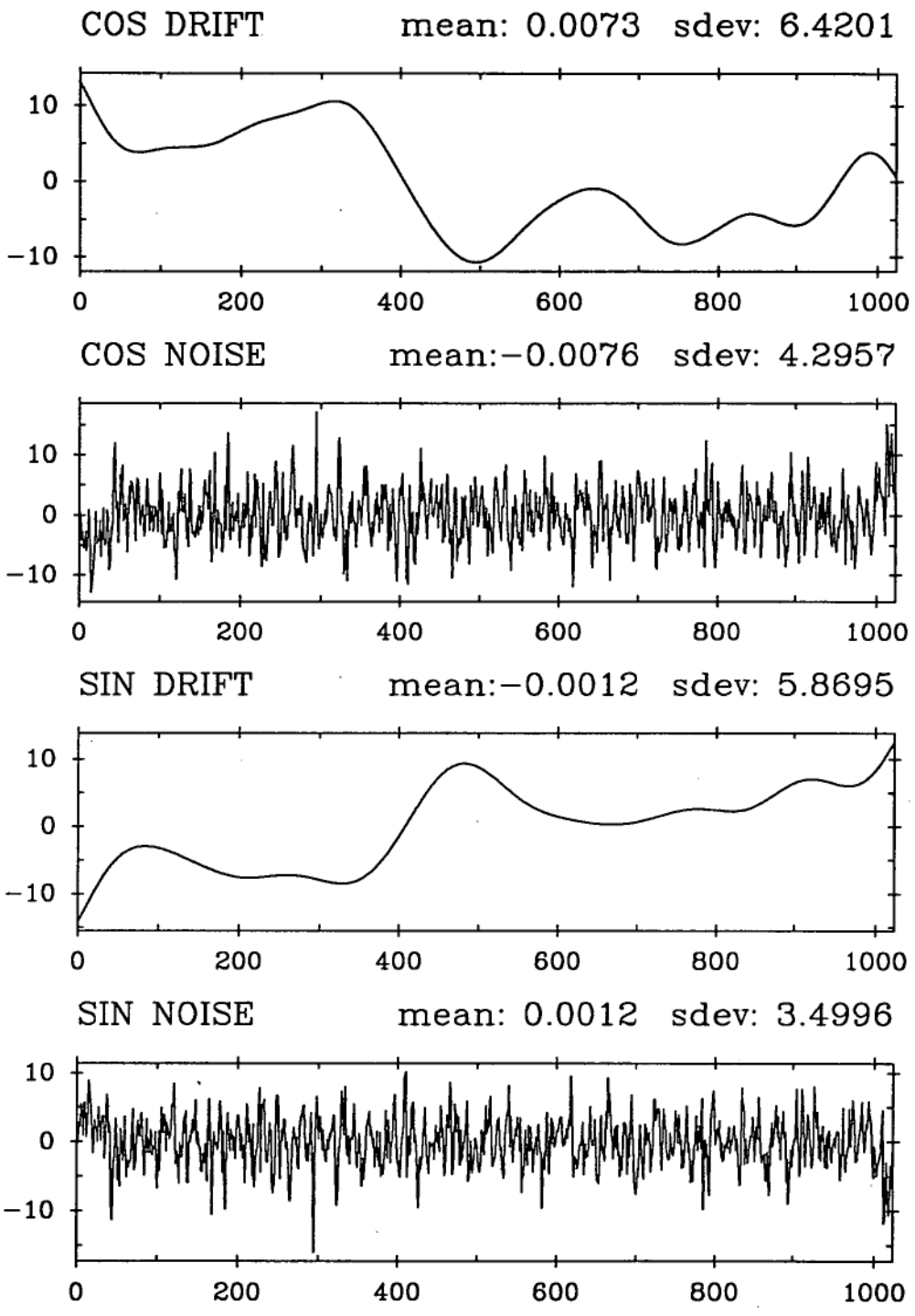


Figure 4.14: Cosine drift and noise, then sine drift and noise for the 12725 MHz off-source data run obtained by Wiener filtering the PSD with a cut-off frequency of 9.8 mHz. The vertical scales are in counts.

order $X_A X_R$ and setting $G_\alpha = G_\beta = 1$, the smoothed polarimeter outputs are

$$U' = U + \tfrac{1}{2} X_R R^2 \cos\alpha + \tfrac{1}{2} X_A A^2 \cos\alpha$$
$$V' = V + \tfrac{1}{2} X_R R^2 \cos\beta + \tfrac{1}{2} X_A A^2 \cos\beta$$

where U and V are the original outputs. In the sidelobe regions A^2 gets small rapidly and a constant vector offset of magnitude $\frac{1}{2} X_R R^2$ is left, which by the argument above is about 3.8 counts. Figure 4.14 displays fluctuations of this magnitude, not a constant offset. Some more investigation is required, although the correlation of the outputs remains a strong argument in favour of cross-coupling as the major contributor.

Now the additive noise can be separated from the on-source amplitude and phase noise. Let the on-source response be A' with phase ϕ'. The effect of $\mathrm{rms}\,\varepsilon_r = \mathrm{rms}\,\varepsilon_i = \sigma$ on $|A'|$ and ϕ' can be shown, in the strong signal case [82, pp. 259–62], to be

$$\left.\begin{aligned} \sigma_{|A'|} &\simeq \sigma \\ \sigma_{\phi'} &\simeq \sigma / |A'| \end{aligned}\right\} \text{ where } \sigma \simeq 4.0 \text{ and } |A'| \simeq 4390.$$

The 12725 MHz data is used as a example. This additive component can be subtracted quadratically from the rms values indicated in figure 4.10. Thus the multiplicative amplitude error was

$$\sigma_{|A'|} \simeq \sqrt{16.0^2 - 4.0^2} = 15.5 \text{ counts}$$

and the multiplicative phase error was

$$\sigma_{\phi'} \simeq \sqrt{0.60^2 - \left(\tfrac{180 \times 4.0}{\pi \times 4390}\right)^2} = 0.60°.$$

To find the signal-to-noise ratios defined in section 4.2.3, note that $|\varepsilon|$ follows a Rayleigh distribution for which

$$\mathrm{rms}\,|\varepsilon| = \sqrt{\langle |\varepsilon|^2 \rangle} = \sqrt{2}\,\sigma.$$

Incidentally, this is different to

$$\sigma_{|\varepsilon|} = \sqrt{\langle |\varepsilon|^2 \rangle - \langle |\varepsilon| \rangle^2} = \sqrt{2 - \pi/2}\,\sigma.$$

Thus, by equation 4.13,

$$\mathcal{R}_n = \frac{|A|}{\mathrm{rms}\,|\varepsilon|} \simeq \frac{4390}{\sqrt{2} \times 4.0} = 776.$$

Quantity	Carrier	Beacon
$\mathcal{R}_n$	776	604
$\mathcal{R}_a$	283	148
$\mathcal{R}_p$	96	62
$f_u = f_v$	0.069	0.069

Table 4.1: Signal-to-noise ratios and fractional pointing errors estimated from on-source and off-source data runs, for both the carrier signal and the beacon signal. All quantities are dimensionless.

Given $A' = A + \rho A$ and $\phi' = \phi + \xi$ corresponding to equations 4.2 and 4.3, the amplitude and phase rms deviations can be evaluated:

$$\sigma_{|A'|} = \sqrt{\langle |A'|^2 \rangle - \langle |A'| \rangle^2} = |A| \mathrm{rms}\, \rho$$
$$\sigma_{\phi'} = \sqrt{\langle \phi'^2 \rangle - \langle \phi' \rangle^2} = \mathrm{rms}\, \xi.$$

Thus, again by equation 4.13,

$$\mathcal{R}_a = \frac{1}{\mathrm{rms}\, \rho} \simeq \frac{4390}{15.5} = 283$$
$$\mathcal{R}_p = \frac{1}{\mathrm{rms}\, \xi} \simeq \frac{180}{\pi \times 0.60} = 96.$$

Similar calculations can be done for the 12373 MHz data. A summary of the results is in table 4.1.

4.3.3 Estimated and Actual Aperture Errors

The estimates of $\mathcal{R}_n$, $\mathcal{R}_a$, $\mathcal{R}_p$, f_u, and f_v can be inserted straight-away into equations 4.26 to 4.29, and the average aperture phase uncertainty can be found with equation 4.31. Using equations 4.22 to 4.25 the phase uncertainties assuming a uniform illumination can also be found, and it is interesting to compare these. Table 4.2 shows the phase uncertainty according to both methods for each error category and for each signal source. The additive noise dominates all other sources, if the random error model is correct, even though its signal-to-noise ratio is very much the greatest. As expected, the effect of pointing errors is over-estimated by a uniform illumination analysis. Notice that the effect of additive noise is also over-estimated in this case: the Scott and Ryle equation neglects the taper both in finding $\mathrm{rms}\,|e_{\mathrm{DFT}}|$ and in

$\overline{\sigma_\phi}$ from:	Carrier Map AV1	Beacon Map AV2
Additive	1.27° (1.48°)	1.40° (1.80°)
Amplitude	0.10° (0.10°)	0.19° (0.19°)
Phase	0.30° (0.31°)	0.45° (0.46°)
Pointing	0.07° (3.72°)	0.07° (3.52°)
Total	1.31°	1.48°

Table 4.2: The average aperture phase uncertainty arising from each category of error according to the models developed in this chapter. The necessary aperture integrals were computed for maps AV1 and AV2, and the corresponding error parameters were taken from the previous table. In brackets are the aperture phase uncertainties predicted by the uniform illumination equations.

weighting σ_ϕ appropriately, and the omissions do not cancel out. The effect of multiplicative errors, on the other hand, is well predicted by the uniform illumination equations.

The total phase uncertainty is the square root of the sum of the square of each contributing uncertainty. The 'total' signal-to-noise ratio may be found from this as suggested in section 4.2.3: $\mathcal{R}_t$ is the additive snr that would be sufficient to account for the total aperture error. It is 752 or 57.5 dB for the carrier map and 570 or 55.1 dB for the beacon map.

The phase uncertainty in each ring of panels can also be found, as it was in section 3.3.2 from the repeatability of contiguous maps. These can be compared directly after dividing the repeatability figures by $\sqrt{2}$ because they arise from a difference of phases. In table 4.3 this comparison is done for maps AV1 and AV2 (compare with table 3.2). It is very convincing, but remember that this confirms only the additive noise predictions, since they dominated all others. Ring number 2 was affected by shadowing. For other rings the variation of $\overline{\sigma_\phi}$ with radius is well modelled by equation 4.31, particularly for the carrier data. This has the important consequence that the repeatability of contiguous maps appears to have been limited by random errors, not by systematic effects. The beacon data shows more deviation from the model, with predictions a little low in rings 3–5 and a little high in rings 6–8. In the light of the arguments leading to figure 3.16 a slight distortion of the entire reflector might be responsible for this.

Ring Number	Carrier Map AV1		Beacon Map AV2	
	$\overline{\sigma_\phi}$	rms $\Delta/\sqrt{2}$	$\overline{\sigma_\phi}$	rms $\Delta/\sqrt{2}$
2	1.12°	1.68°	1.28°	2.02°
3	0.96°	0.83°	1.06°	1.20°
4	1.04°	1.12°	1.15°	1.51°
5	1.25°	1.20°	1.40°	1.44°
6	1.46°	1.34°	1.69°	1.51°
7	1.67°	1.62°	1.98°	1.73°
8	2.21°	2.17°	2.72°	2.48°
2–8	1.31°	1.29°	1.48°	1.56°

Table 4.3: The predicted aperture phase uncertainty $\overline{\sigma_\phi}$ in each ring is compared with the phase repeatability rms Δ for contiguous maps in section 3.3.2.

The approach to random errors shown in this chapter has correctly predicted their effects in the aperture domain. The on-source and off-source signals were examined and multiplicative random errors, while dominating the on-source fluctuations, were confirmed to have little effect in comparison to the constant additive error level. Rather than receiver noise, it is suspected that cross-coupling in the i.f. system is responsible for the additive noise. The mechanism by which this occurs needs more investigation, but in the meantime it would be interesting to re-design the circuit board before further holographic measurements are made. If this is presently the limitation on achievable map accuracy then the lack of substantial improvement by use of a strong carrier signal is explained.

Chapter 5

Panel Fitting by Successive Projections

The remainder of this thesis is devoted to the problem of panel fitting from sparsely sampled Microwave Holography data. This chapter develops an algorithm based upon the Method of Successive Projections (SP) onto convex sets. In section 5.1 the problem is outlined and a history of other SP applications is given. Section 5.2 discusses the 'mechanics' of the algorithm, and section 5.3 shows how it was applied to the Mount Pleasant data.

5.1 The Method of Successive Projections

5.1.1 The Resolution Problem

A limitation of Microwave Holography is that in practice only a small part of the complete radiation pattern can be measured. It was seen in section 2.2.2 that as a result the recovered aperture current function is convolved with a function of resolution width (δ_x, δ_y) in wavelengths where

$$\delta_x = \frac{1}{W_u} \text{ and } \delta_y = \frac{1}{W_v}.$$

The spatial frequencies represented by the far field samples are limited to a 'bandwidth' of (W_u, W_v) about the zero frequency, main beam position. Thus any aperture features with high spatial frequencies are not recovered. Physically, this means that such features scatter power to far field regions which are not measured.

Practical holographic measurements of large reflector antennas are aimed at adjusting surface panels. The difficulty created by limited spatial resolu-

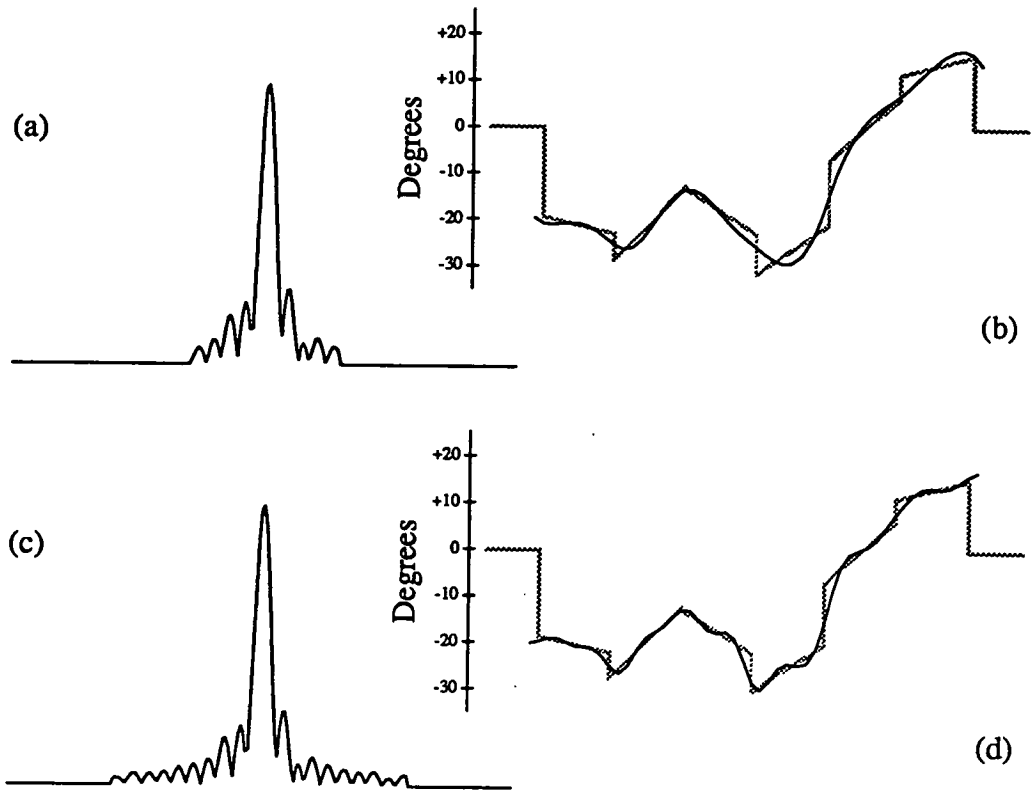


Figure 5.1: A one-dimensional simulation showing the effect of truncation of the far field, plotted on a linear amplitude scale in (a) and (c). The corresponding recovered aperture phase functions are plotted in degrees in (b) and (d). Far field phase and aperture amplitude are not shown.

tion was alluded to in section 2.4, although panel fitting proceeded anyway. Aperture samples which are within a resolution cell of a panel edge are significantly affected by the neighbouring panel. Many holographic maps have very few points per panel — the number of points per panel P_p is an average quantity since panels have different shapes and orientations to the sample grid — so the problem is serious. A one-dimensional illustration of this is shown in figure 5.1. An aperture with 6 'panels' was simulated by 6 segments of linear phase, uniformly illuminated. Sampling out to the third sidelobe shown in (a) gave the recovered aperture phase in (b), against which is plotted the original phase. Estimated panel positions will be very poor. In (c) twice the extent of far field has been sampled, and the recovered aperture phase (d) follows the original more closely.

A typical panel layout for a real, two-dimensional aperture is that of the Mount Pleasant 26 m antenna. As can be seen in figure 2.12 the panels are

trapezoidal and have edges lying on radial lines and polygonal lines about the reflector's centre. Discontinuities and gaps along these lines constitute aperture features of high (spatial) frequency, and the aperture maps in chapter 3 contain only slight indications that these features exist. This is understandable, because a resolution cell is about 0.6×0.6 m while typical panel dimensions are 1.0–1.5 m. Yet it is known precisely where the panel edges lie, so there should be no need to recover their locations from the data. Can this knowledge possibly be incorporated into the data analysis to improve the estimated panel tilts and offsets?

It can indeed, though first the edge locations must be described in a precise mathematical way. This is done by saying they are boundaries of regions of linearly varying aperture phase. As discussed in section 2.4, profile errors in individual panels must be small beside panel setting errors if panel corner adjustments are to have any benefit. It is assumed now that the panels are undistorted, so the aperture phase will follow a plane given by equation 2.48 above each panel. The range of possible aperture current functions is then restricted to a set of those fulfilling this criterion. Simultaneously, the functions are restricted by holographic measurements to a set consistent with the measured far field pattern. The intersection of these sets, if it exists, contains functions that agree with all existing knowledge of the aperture: both the far field measurements and the locations of panel edges.

It is most convenient to describe the second set in the far field domain. Precisely, define the sets as follows:

> *The set* S_1 contains all radiation patterns that agree with the measured pattern, to within a constant factor. Beyond the region of measurement there is no restriction on the pattern.
>
> *The set* S_2 contains all aperture current functions with linearly varying phase above each panel and zero magnitude past the reflector edge.

It is possible to speak of the intersection of these sets despite their being defined in opposite Fourier domains, because there is a one-to-one correspondence between elements in each domain. The Fourier transform also preserves distance: it is an isomorphic linear operator. The domains are thus indistinguishable except for the physical meaning of their elements. Another interpretation is to imagine transforming every element of S_1 to the aperture domain. Then the sets can be compared directly as when they were first

introduced. Of course this is impossible because both sets contain an infinite number of elements, so it is better simply to transform a particular function to the appropriate domain when it needs to be checked against one or other of the sets.

These arguments will be formalized later. The important result is that the problem of panel fitting has been reduced to the problem of finding the intersection of two sets in the 'space' of aperture current functions. There are a number of techniques available for finding the intersection of sets. Here the Method of Successive Projections will be examined. It is a technique which is conceptually simple and has been applied successfully to other problems of electromagnetics. It lies behind several popular algorithms for phase retrieval and image enhancement.

5.1.2 Successive Projections

Consider the two sets in the two-dimensional plane illustrated in figure 5.2(a): each set is the interior and edge of a closed boundary curve. Beginning at any point in the plane, the sets' intersection can be approached by dropping a perpendicular onto each set alternately as illustrated. If the sets do not intersect, as shown in 5.2(b), the same process will approach their nearest points. The sequence of points obtained in either case is guaranteed to converge provided that the sets are *convex*, that is, any point lying between two elements of a set is also an element. The same method can be used for any number of convex sets, and figure 5.2(c) shows an example of three intersecting sets and a sequence of perpendiculars.

The point at which a perpendicular from an exterior point meets the set boundary is called the *projection* onto the set; the projection from an interior point is the point itself. Clearly the projection is the element of the set nearest to the original point. More formally, consider a *closed* set S, a closed set being one that contains all its limit points. Given a norm $\|\cdot\|$, the projection y from a point x onto the set S satisfies

$$y \in S \text{ and } \|y - x\| = \min_{y' \in S} \|y' - x\|. \tag{5.1}$$

If the set is convex this point is unique for each x, otherwise there may be several or even infinitely many projections. For example, the set $\{x \mid \|x - x_c\| \geq 1\}$ has infinitely many projections from the point x_c, as shown in figure 5.3.

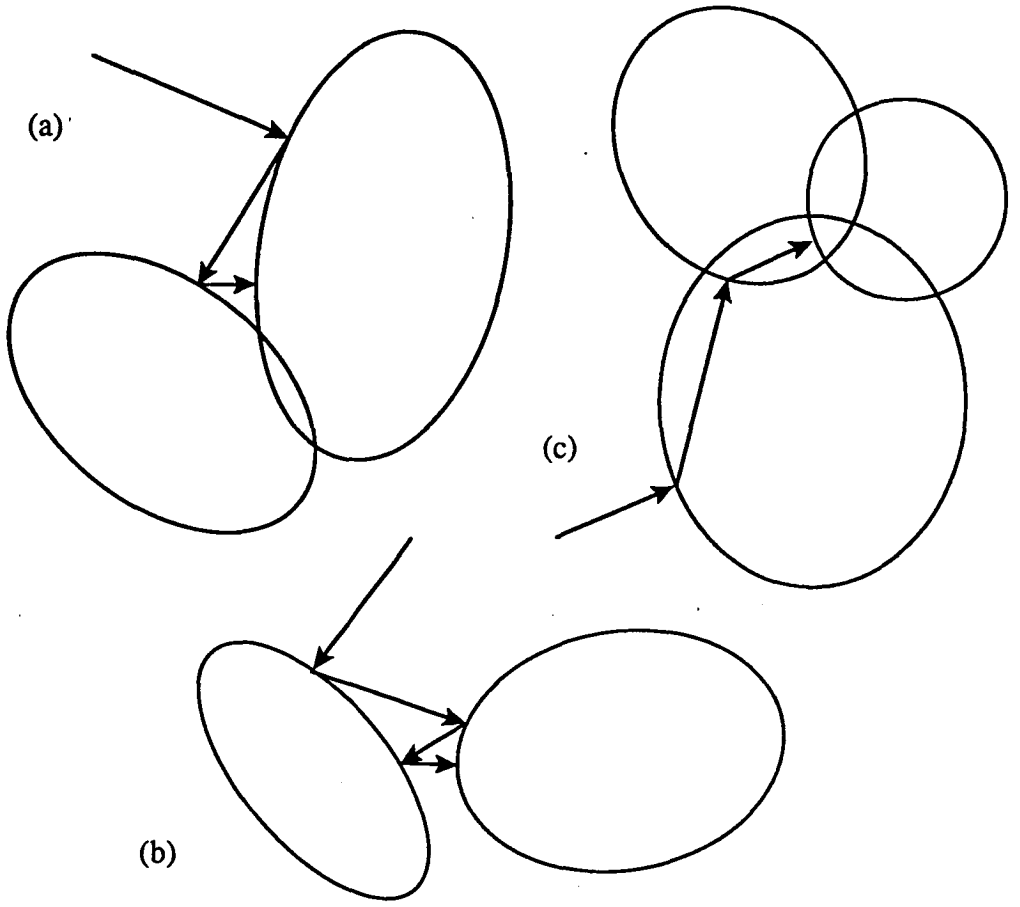


Figure 5.2: Dropping perpendiculars onto convex sets in the two-dimensional plane: (a) two intersecting sets, (b) two disjoint sets, and (c) three intersecting sets.

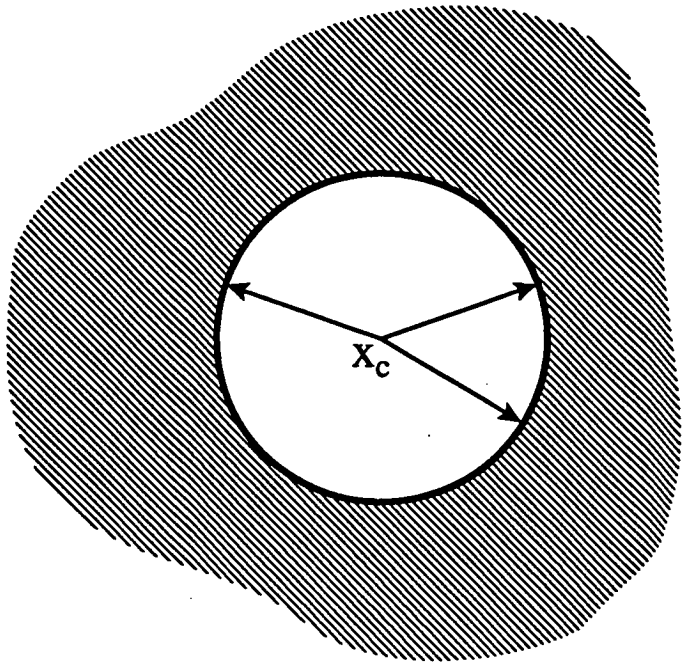


Figure 5.3: An example of a set with infinitely many projections from a particular point x_c.

The schemes shown in figure 5.2 can thus be called Methods of Successive Projections for finding a common point of convex sets (or the nearest point of two disjoint sets). A group of Russian mathematicians [25] described the method rigourously in 1967 for convex sets in a general Hilbert space, and showed that it could be applied to practical problems. They considered schemes for ordering the projections onto more than two sets: cyclic ordering, choosing the remotest set each time, or approximately choosing the remotest set. Their main contribution was to demonstrate *strong* convergence of the sequence of projections for infinite-dimensional Hilbert spaces, provided that one of several alternative conditions on the sets was satisfied. This means the sequence is convergent in the norm: if the starting point is x_0 and the nth projection is x_n, then there is a limit point x for which

$$\lim_{n\to\infty} \|x_n - x\| = 0. \tag{5.2}$$

Strong convergence is always guaranteed for closed, convex sets in a finite dimensional Hilbert space.

Under the right conditions, convergence is geometric for intersecting convex sets. This can be shown easily for the situation in figure 5.2(a), as demonstrated by figure 5.4 which magnifies the region of intersection. It is

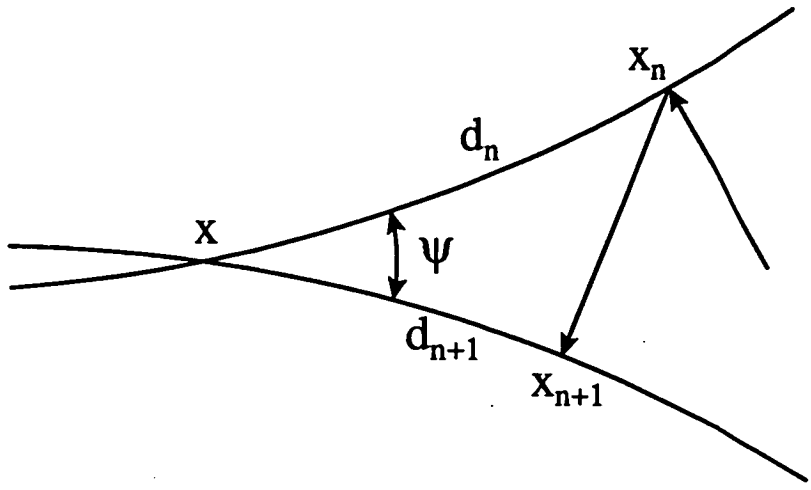


Figure 5.4: The region of intersection of two convex sets in the two-dimensional plane. Two projection points x_n and x_{n+1} are shown, and their distances from the limit point x are d_n and d_{n+1}. Because the set boundaries are almost straight on this scale, $d_{n+1}/d_n \simeq \cos\psi$ or $d_{n+1} \simeq d_n \cos\psi$. Thus convergence is geometric.

also true, in the two-dimensional plane at least, for the disjoint convex sets in 5.2(b): the distance of the projection x_n from the closest point of approach on the same set decreases geometrically. This is demonstrated in figure 5.5. Observe that in the intersecting case the length of projections $\|x_{n+1} - x_n\|$ also decreases geometrically. This is a useful quantity for monitoring the convergence in an experiment, where naturally the limit point is not known.

The technique as discussed so far can already be applied to useful problems. A Successive Projections formulation was used by Abo-Taleb and Fahmy [1] in the design of FIR digital filters. They used a large number of sets, each set representing the constraint on the amplitude response at a particular frequency. A filter can be specified by a vector of real coefficients $\mathbf{x} = (x_1, \ldots, x_N)$ by expressing the frequency response $H(f)$ as a combination of 'basis' functions

$$H(f) = \mathbf{x} \cdot \Phi(f) = \sum_{j=1}^{j=N} x_j \Phi_j(f). \tag{5.3}$$

Abo-Taleb and Fahmy dealt with two-dimensional filters, so the notation is simplified here. Let the desired amplitude response $G(f)$ be specified at a set of frequencies $f_1, \ldots, f_M$, and allow a tolerance $\pm E$ at each frequency. Then a suitable filter is one for which $|G(f_i) - H(f_i)| \leq E$ at each f_i, or in

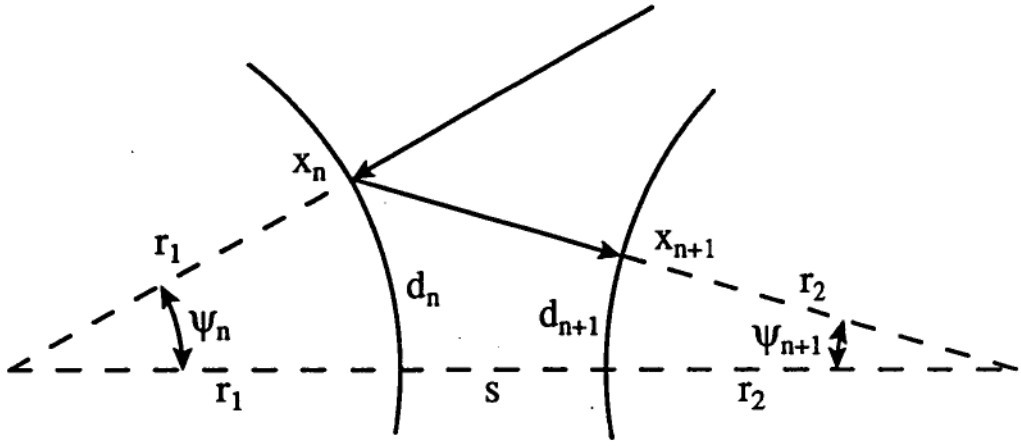


Figure 5.5: Two disjoint sets with local radii of curvature r_1 and r_2 at the points of closest approach, which are separated by a distance s. Other symbols are as in the previous figure, except ψ_i is now the angle to projection x_i. Let $D = r_1 + s + r_2$ and the sine rule gives $r_1/\sin\psi_{n+1} = D/\sin(\psi_n + \psi_{n+1})$. After a small angle approximation, $D/r_1 \simeq 1 + \psi_n/\psi_{n+1}$, but $d_n \simeq r_1\psi_n$ and $d_{n+1} \simeq r_2\psi_{n+1}$, so $d_{n+1} \simeq d_n r_2/(r_2 + s)$. Similarly $d_{n+2} \simeq d_{n+1}r_1/(r_1 + s)$, so convergence is geometric with an alternating factor.

other words $\mathbf{x} \in Q_i$ for each i, where

$$Q_i = \{\mathbf{x} \mid |G(f_i) - \mathbf{x} \cdot \mathbf{\Phi}(f_i)| \leq E\}. \tag{5.4}$$

The problem has become that of locating a vector in the intersection of M sets Q_i which exist in the space of real N-vectors. In this way it is very similar to the pictorial examples in figure 5.2, which have $N = 2$. It is shown in [1] that the sets are closed and convex, and that projections onto them can be calculated easily.

5.1.3 Non-Convex Sets

Many existing signal processing algorithms make use of Successive Projections methodology, although their originators did not couch them in these terms. A classic example is the Gerchberg-Saxton algorithm for phase retrieval [18]. This estimates the phase of a function whose amplitude only is known in both Fourier domains, using an iterative scheme:

1. Begin with a random phase function.

2. Impose the known 'time' domain amplitude onto the phase.

3. Apply a Fourier transform.

4. Impose the known 'frequency' domain amplitude onto the phase.

5. Apply an inverse Fourier transform.

6. Repeat from item 2 until the phase is stable.

Consider the following subset of the space of complex functions:

$$S_a = \{f(x) \mid |f(x)| \text{ matches a known amplitude function}\}. \qquad (5.5)$$

Items 2 and 4 are simply projections onto two sets of this kind, because no matter what 'distance' measure is used the nearest element of S_a to a function will match the specified amplitude and have unchanged phase. Thus the Gerchberg-Saxton algorithm is a Successive Projections application with one set in each of two isomorphic domains, just like the sets S_1 and S_2 in section 5.1.1. If preferred, the transform and its inverse can be thought of as part of the frequency domain projection, but then the algorithm loses its pleasing symmetry.

A closely related algorithm is due to Missel [47]. Known in this case is the amplitude of two images of the same object related by a focus change. The projection operation is the same as above, but one defocused image is related to its mate by a Fourier transform to the diffraction plane, a phase change corresponding to the defocus, and a transform back to the image plane. The whole is equivalent to a convolution. This algorithm has found practical use for phase-less antenna measurements [49]. Phase retrieval will be discussed further in section 6.2.3: whether or not a unique phase solution exists has an important bearing on the problem.

The advantages of regarding these existing algorithms as applications of a general, well-founded technique are two-fold. Firstly, it gives insight into their operation where otherwise they seem to be based upon educated intuition. Secondly, it opens the way to a rigorous treatment of convergence and robustness, which previously have been investigated using numerical tests. This may be a profitable area for future research.

There is one immediate difficulty for the phase retrieval problem: the set S_a is not convex. This is why users of the Gerchberg-Saxton or Missel algorithms have had to take great care to avoid misleading behaviour. A major advance was made by Levi and Stark [40] when they showed that a kind of convergence can be ensured for two non-convex sets by optimizing

the *relaxation parameters* of the projections. There is no similar result for three or more sets. If $P(x_n)$ is the projection of x_n onto some set, the next point in the sequence can be chosen to fall short of or to overshoot $P(x_n)$:

$$x_{n+1} = x_n + \lambda_n(P(x_n) - x_n) \text{ where } 0 < \lambda_n \leq 2. \tag{5.6}$$

Levi and Stark showed how to find for each projection the range of relaxation parameters λ_n which give 'convergence' according to their criterion: the sum of distances to the two sets decreases monotonically. In fact, $\lambda_n = 1$ is always within this range so $x_{n+1} = P(x_n)$ will always reduce the summed set distance. Other values of λ_n are usually chosen to accelerate convergence.

Monotonic reduction of the summed set distance does not imply strong convergence to a limit point as would be the case for convex sets. This allows the existence of 'traps' as illustrated in figure 5.6(a), and these can halt convergence. Even with convex sets 'tunnels' can occur as in figure 5.6(b), giving extremely slow convergence. There may be ways to escape traps and tunnels; at least it is possible to detect them when they occur.

Poulton [59] has applied SP to antenna power pattern synthesis using a formulation similar to that of Abo-Taleb and Fahmy above. The method had to be extended to the space of complex N-vectors to allow for the complex radiation patterns of antennas. The pattern constraints in this case gave non-convex sets, so the application demonstrated that convergence can be obtained even when there are many non-convex sets. The lesson to be learned is that projections are likely to converge if the problem is solvable. Asking impossible things of the method, such as phase retrieval when a unique solution is not guaranteed, is likely to give unreliable behaviour.

Poulton [60] has also investigated the use of projection methods for data reduction in Fresnel zone antenna measurements. Yet another SP application in disguise is the Papoulis image enhancement algorithm [53] which can be applied when a function is known to have a band-limited spectrum. This too has been used in antenna measurements [63]. Successive Projections is clearly a technique with a wide range of existing and potential applications. There are only two prerequisites for its use:

- The problem must be represented as the intersection of several sets;
- Projections onto the sets must be readily calculable.

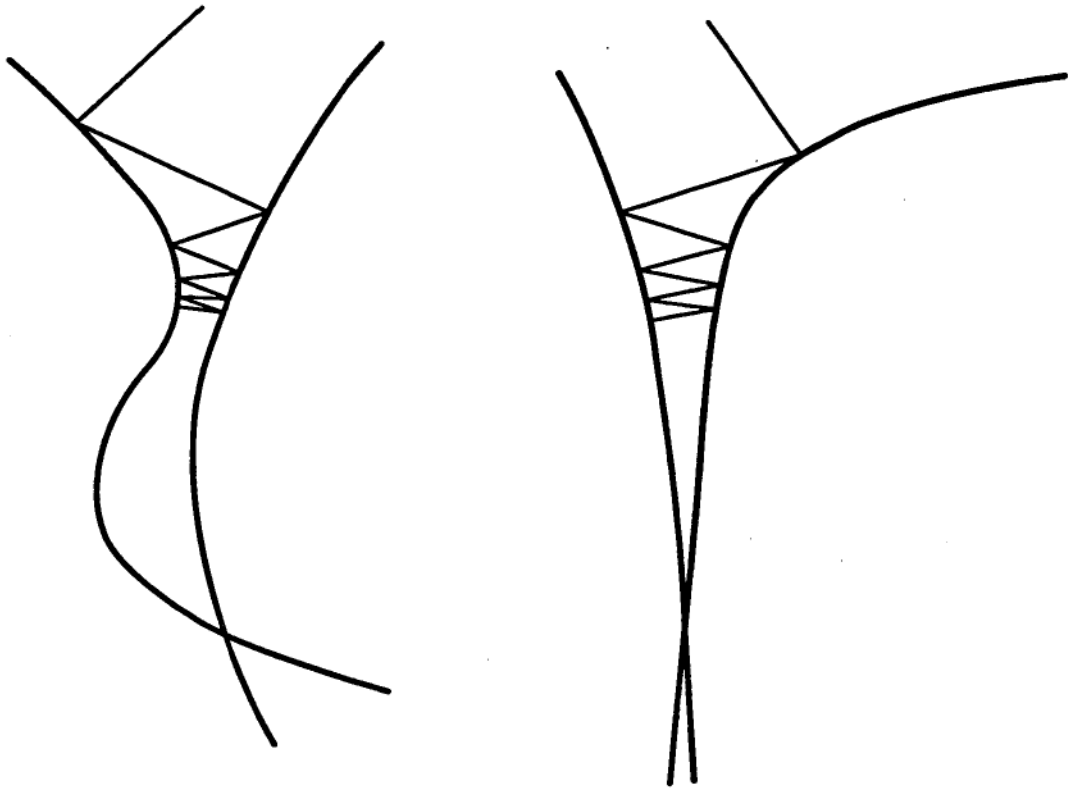


Figure 5.6: Pathologies of Successive Projections with two sets: (a) a 'trap' and (b) a 'tunnel' (after Levi and Stark).

5.2 Theory for Panel Fitting

The application of Successive Projections to the panel fitting problem will now be considered in detail. One-dimensional far field patterns and aperture current functions will be used. This gives a less cluttered notation, and all the results can be generalized easily to two dimensions.

5.2.1 The Sets and their Hilbert Space

The Intersection

Return to the set S_1 defined in section 5.1.1: it does not constrain the far field pattern beyond the region of measurement. This is why there is 'room' in S_1 to accommodate or come close to a part of S_2. If a pattern in S_1 is represented by the complex function $A(u)$, then it is fixed by the measurements (to within a constant factor) in the region $|u| \leq W_u/2$, but for larger u the terms can vary as necessary to 'square up' the corresponding aperture phase over each panel. The effect of applying SP to the sets is to estimate these terms in a way consistent with the known layout of panels in the aperture. The result is similar to that seen in figure 5.1, where doubling the measurement span gave a more faithful representation of the panels.

Figure 5.7 shows how two sets may either overlap, or just touch, or merely come close at some points. Sets of functions in an abstract space are more difficult to imagine than ovals drawn on a page, but the significance of these three cases is the same. If the sets S_1 and S_2 have a considerable overlap then there are many functions which satisfy the criteria of both sets, and without further information one is as good as any other. It is expected that SP sequences, for example figure 5.2(a), starting from different points x_0 would approach different limit points on the boundary of $S_1 \cap S_2$, and in this way an idea of the extent of the intersection might be formed. A similar behaviour might result when the sets just touch, because convergence is likely to be very slow due to the 'tunnel' effect of figure 5.6(b), giving a variety of endpoints after a finite number of iterations.

In the disjoint case SP approaches the point of closest approach of the two sets, representing the best estimate that can be obtained with conflicting information. The 'intersecting' region of sets S_1 and S_2 may actually look something like figure 5.8, because it is found below that S_1 is convex but S_2 has a complicated boundary. If so, whether or not they really intersect

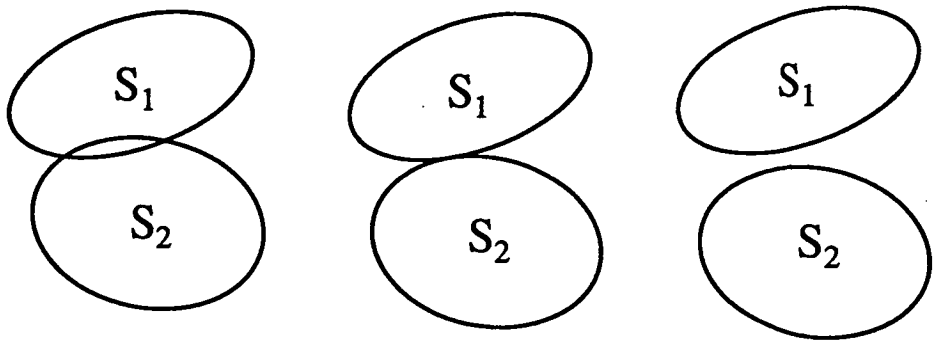


Figure 5.7: Three possibilities for the meeting of two sets.

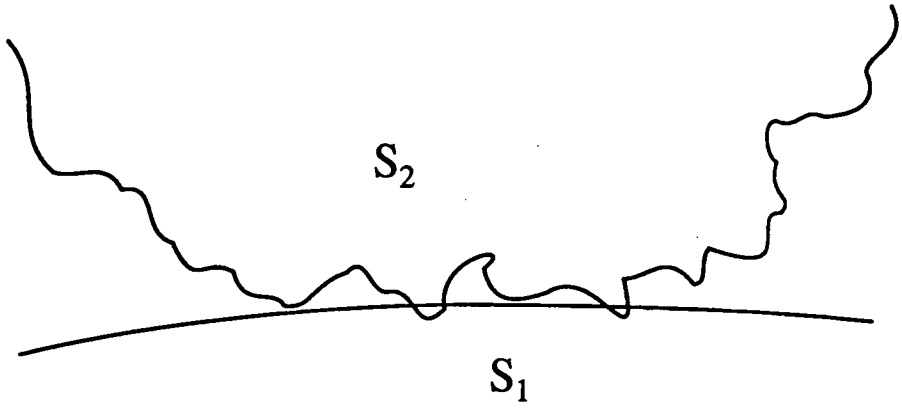


Figure 5.8: A possible 'intersection' of a convex and a non-convex set.

is immaterial. Successive Projections will approach some point near this region in a finite number of iterations, and this will be good enough provided that the region is acceptably small. This can be judged by comparison with the uncertainty in panel positions expected due to measurement errors, and ultimately with the accuracy of the panel adjusting mechanism. In section 6.1 these issues are investigated by numerical experimentation.

A decision has to be made on the span of the radiation pattern to attempt to recover. It was not mentioned in the definition of S_1 that the pattern is restricted to the forward hemisphere, $-1 < u < +1$, by the assumptions behind the aperture integration — see section 2.1.2. In practice it is wasteful to let all the unmeasured far field in this region be determined by SP. All that is necessary is that a sufficient span is determined beyond W_u to give accurate panel fits by the least squares method in appendix C. A larger span would give a smaller resolution cell in the aperture and thus sharper panel boundaries, but the fitted panels would be the same because no new infor-

mation would be available to the algorithm. The presence of measurement errors would make this a doubly irrelevant exercise. A span W'_u of 4 to 5 times W_u was found adequate in this study, the rest of the far field pattern remaining undetermined.

Which Hilbert Space?

Consider now the function spaces in which the sets S_1 and S_2 exist. A radiation pattern spanning W'_u is in the *complex space* $L^2[-W'_u/2, +W'_u/2]$, which is a Hilbert space as required by Gubin, et al. [25]. The properties of Hilbert spaces are discussed in any good textbook on functional analysis, for example the one by Kreyszig [38]. The transform appropriate for such a radiation pattern is a Fourier series if a periodic extension is assumed, so the corresponding aperture current function is in the Hilbert space l^2 of infinite complex sequences. This is inconvenient, because a different norm must be used in each domain.

A better approach is to deal directly with the discrete far field and aperture current functions which are measured and derived in Microwave Holography experiments. This means the notation is immediately applicable to the data, and both domains are in the same Hilbert space. Breaking with section 2.2, the discrete functions will be denoted by

$$\textit{Far field pattern}: \quad A = A(u) \text{ where } u = m\delta_u \tag{5.7}$$

$$\textit{Aperture currents}: \quad a = a(x) \text{ where } x = p\delta_x. \tag{5.8}$$

The integers m and p range over the N'_u values $-L'_u, \ldots, +L'_u$ where $N'_u\delta_u = W'_u$. It is understood that the convolution of equation 2.20 has taken place, and that the functions are defined only at these discrete points: they are really complex N'_u-vectors. The advantages of this notation over "(A_i)" and "(a_i)" are that the vectors' physical meanings are emphasized and a clutter of subscripts is avoided. The disadvantage is that for the rest of this chapter the reader must remember that u and x are indices, not continuous arguments.

Thus both sets of functions exist in the *unitary space* $\mathbf{C}^n$ of ordered n-tuples of complex numbers. It is a Hilbert space, and the norms are

$$\|A\| = \sqrt{\sum_u A(u)A^*(u)} \text{ and } \|a\| = \sqrt{\sum_x a(x)a^*(x)}. \tag{5.9}$$

This is nice because minimizing the distance of an estimate from the intersection $S_1 \cap S_2$ amounts to a conventional least squares optimization. A discrete

Fourier transform relates every vector a to a unique vector A, and Parseval's relation gives

$$\|A\| = \sqrt{N'_u}\|a\| \text{ and so } \|A_1 - A_2\| = \sqrt{N'_u}\|a_1 - a_2\|. \tag{5.10}$$

This is the distance-preserving property that was mentioned earlier.

Defining the Sets

To make a notation which is suitable for one or two dimensions the indices u and x will be divided into sets. Let $u \in M$ span the range of measured far field values, and let $u \in R$ span the rest of the range W'_u. In one dimension

$$\begin{aligned} M &= \{u \mid u = m\delta_u \text{ for } m = -L_u, \ldots, +L_u\} \\ R &= \{u \mid u = m\delta_u \text{ for } m = -L'_u, \ldots, -L_u - 1, +L_u + 1, \ldots, +L'_u\}. \end{aligned}$$

It is easy to define similar sets for a pair of indices (u, v). In the aperture domain, some indices x will lie outside the physical aperture: call this set X_0. The physical aperture is divided up into N_p panels, and the set X_k contains all indices x lying in panel number $k = 1, \ldots, N_p$. Let the phase $\phi(x) = \arg a(x)$ so that $a(x) = |a(x)| \exp(j\phi(x))$. A *panel function* is a phase function for which

$$\phi(x) = a_k x + b_k \text{ for } x \in X_k \tag{5.11}$$

where a_k and b_k are real numbers generally differing for each k. An example is in figure 5.1. In two dimensions the equation of a plane must be used.

Let the far field measurements be $A_m(u)$ for $u \in M$. The sets are formally defined as follows:

$$S_1 =' \{A \mid A(u) = \mu A_m(u) \text{ for } u \in M \text{ and real } \mu\} \tag{5.12}$$

$$S_2 = \{a \mid \phi(x) \text{ is a panel function and } a(x) = 0 \text{ for } x \in X_0\}. \tag{5.13}$$

The first set can easily be shown to be convex. Consider two elements A_1 and A_2, and let $0 < t < 1$ for some real t:

$$u \in M \Rightarrow tA_1(u) + (1-t)A_2(u) = t\mu_1 A_m(u) + (1-t)\mu_2 A_m(u) = \mu_3 A_m(u)$$

where μ_3 is also a real number. Functions lying between A_1 and A_2 satisfy the criterion for membership of S_1, so the set is convex.

The second set can just as easily be shown not to be convex. A linear combination of two elements a_1 and a_2 is a vector addition at each index x, as

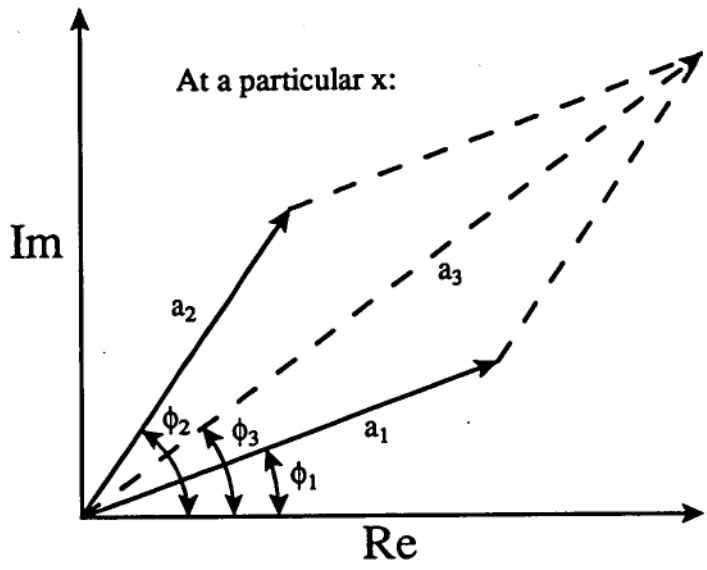


Figure 5.9: The vector sum of $a_1(x)$ and $a_2(x)$, giving a resultant phase $\phi_3(x)$.

illustrated in figure 5.9. The resultant angle $\phi_3(x)$ depends on the magnitudes $t|a_1(x)|$ and $(1-t)|a_2(x)|$, so a linear variation from one x to the next will not generally be preserved. Notice, however, that if the angular separation $|\phi_1(x)-\phi_2(x)|$ is small the angle $\phi_3(x)$ is less dependent on the magnitudes. It is this property that allows convergence to be attained. It is very difficult to imagine the 'shape' of the boundary of S_2 because of the non-linear behaviour of phase under addition. For example, the mid-point of two elments of S_2 with equal amplitudes $r(x)$ is

$$\tfrac{1}{2}r(x)e^{j\phi_1(x)} + \tfrac{1}{2}r(x)e^{j\phi_2(x)} = r(x)\cos\left(\tfrac{\phi_1(x)-\phi_2(x)}{2}\right)e^{j\left(\frac{\phi_1(x)+\phi_2(x)}{2}\right)}.$$

This is an element of S_2 because of the linear phase combination, but either side of the mid-point this will not be the case. Perhaps figure 5.8 is a good representation of the boundary!

5.2.2 Deriving the Projections

A constraint of constant total power was added to the projections for convenience in monitoring the progress of the algorithm. This is equivalent to projecting onto $S_1 \cap S_p$ and $S_2 \cap S_p$, where S_p is the set of complex N_u'-vectors with a predetermined power $\|\mathbf{x}\|^2 = P$. Just as in figure 5.3, S_p is not convex but there exists a unique projection onto it from every complex vector except zero: simply scale the vector by a suitable real factor. No difficulties are

expected, and indeed power constraint was found to make no practical difference to the convergence behaviour. The conditions which made it essential in [59] do not exist here.

The Projection onto S_1

From an initial far field pattern A_0 a projection $A \in S_1$ is to be found with the same total power, in other words $\|A - A_0\|$ is to be minimized with the constraint $\|A\|^2 = \|A_0\|^2$. Define the power of the initial point and the power contained in the measurements A_m as follows:

$$P = \sum_{u \in M \cup R} |A_0(u)|^2 \text{ and } P_M = \sum_{u \in M} |A_m(u)|^2. \tag{5.14}$$

Let $A(u) = x(u) + jy(u)$ and similarly for A_0 and A_m. The boundary constraint is then

$$x(u) - \mu x_m(u) = 0 \text{ and } y(u) - \mu y_m(u) = 0 \text{ for } u \in M$$

and the power constraint is

$$\sum_{u \in M \cup R} |A(u)|^2 - P = 0.$$

Introduce Lagrangian multipliers $2\alpha_u, 2\beta_u$ for each of the former and λ for the latter. The problem is solved by minimizing the function

$$\begin{aligned} J &= \sum_{u \in M \cup R} |A(u) - A_0(u)|^2 \\ &+ \sum_{u \in M} 2\alpha_u [x(u) - \mu x_m(u)] + \sum_{u \in M} 2\beta_u [y(u) - \mu y_m(u)] \\ &+ \lambda \left[\sum_{u \in M \cup R} |A(u)|^2 - P \right]. \end{aligned}$$

Taking partial derivatives,

$$\frac{\partial J}{\partial x(u)} = 2x(u) - 2x_0(u) + 2\alpha_u + 2\lambda x(u) \tag{5.15}$$

$$\frac{\partial J}{\partial y(u)} = 2y(u) - 2y_0(u) + 2\beta_u + 2\lambda y(u) \tag{5.16}$$

where the α_u and β_u terms are dropped when $u \in R$. Now for μ:

$$\frac{\partial J}{\partial \mu} = -\sum_{u \in M} [2\alpha_u x_m(u) + 2\beta_u y_m(u)]. \tag{5.17}$$

If these derivatives are set to zero, the first two give for $u \in R$

$$x(u) = \frac{x_0(u)}{1+\lambda} \text{ and } y(u) = \frac{y_0(u)}{1+\lambda}.$$

Thus let $\rho = 1/(1+\lambda)$ and the projection can be written

$$A(u) = \begin{cases} \mu A_m(u), & u \in M \\ \rho A_0(u), & u \in R \end{cases} . \tag{5.18}$$

This shows that, with suitable scaling, the projection is obtained by simply inserting the measurements into the region M of the present estimate of the far field pattern.

For $u \in M$ equations 5.15 and 5.16 give

$$\alpha_u = x_0(u) - (1+\lambda)x(u) = x_0(u) - \frac{\mu}{\rho}x_m(u)$$

$$\beta_u = y_0(u) - (1+\lambda)y(u) = y_0(u) - \frac{\mu}{\rho}y_m(u),$$

and these can be substituted into equation 5.17 and the whole equated with zero:

$$\sum_{u\in M}\left[x_0(u)x_m(u) - \frac{\mu}{\rho}{x_m}^2(u) + y_0(u)y_m(u) - \frac{\mu}{\rho}{y_m}^2(u)\right] = 0$$

or

$$\text{Re}\sum_{u\in M} A_0(u){A_m}^*(u) = \frac{\mu}{\rho}P_M.$$

Obtain another equation by substituting equation 5.18 into the power constraint:

$$\sum_{u\in M}\mu^2\left[{x_m}^2(u) + {y_m}^2(u)\right] + \sum_{u\in R}\rho^2\left[{x_0}^2(u) + {y_0}^2(u)\right] = P$$

or

$$\rho^2\sum_{u\in R}|A_0(u)|^2 = P - \mu^2 P_M.$$

In addition to the definitions 5.14, define:

$$S_R = \sum_{u\in R}|A_0(u)|^2 \text{ and } S_M = \text{Re}\sum_{u\in M} A_0(u){A_m}^*(u). \tag{5.19}$$

Then

$$\rho S_M = \mu P_M \text{ and } \rho^2 S_R = P - \mu^2 P_M$$

which can be solved to give

$$\mu = \left(\frac{P_M{}^2 S_R}{P S_M{}^2} + \frac{P_M}{P}\right)^{-\frac{1}{2}} \text{ and } \rho = \left(\frac{S_R}{P} + \frac{S_M{}^2}{P P_M}\right)^{-\frac{1}{2}}. \tag{5.20}$$

These are the parameters necessary to calculate the projection 5.18.

It is also useful to know the distance from the point A_0 to its projection A, which is the distance from A_0 to S_1. Thus,

$$\begin{aligned} \|A - A_0\|^2 &= \sum_{u \in M} |A_0(u) - \mu A_m(u)|^2 + \sum_{u \in R} |A_0(u) - \rho A_0(u)|^2 \\ &= (P - S_R) - 2\mu S_M + \mu^2 P_M + (1-\rho)^2 S_R. \end{aligned}$$

From the identities above 5.20 can be proved the further identity

$$\mu S_M + \rho S_R = \frac{P}{\rho},$$

which allows this expression to be reduced to the attractive result

$$\|A - A_0\|^2 = 2P\left(1 - \frac{1}{\rho}\right). \tag{5.21}$$

The Projection onto S_2

Analogously to the projection onto S_1, a point $a \in S_2$ is to be found as close as possible to an initial aperture current function a_0. Thus $\|a - a_0\|$ is to be minimized with the constraint $\|a\|^2 = \|a_0\|^2$. Polar representation of complex numbers is better here, so let $a(x) = r(x)\exp(j\phi(x))$ and $a_0(x) = r_0(x)\exp(j\phi_0(x))$. The situation at each index x is illustrated in figure 5.10, so by the cosine rule

$$\|a - a_0\|^2 = \sum_x d^2(x) = \sum_x \left[r^2(x) + {r_0}^2(x) - 2r(x)r_0(x)\cos(\phi(x) - \phi_0(x))\right].$$

The definition of S_2 allows the boundary constraint to be incorporated directly into the function

$$\begin{aligned} J &= \sum_{k=1}^{N_p} \sum_{x \in X_k} \left[r^2(x) + {r_0}^2(x) - 2r(x)r_0(x)\cos(a_k x + b_k - \phi_0(x))\right] \\ &+ \lambda\left[\sum_{k=1}^{N_p} \sum_{x \in X_k} |a(x)|^2 - P\right]. \end{aligned}$$

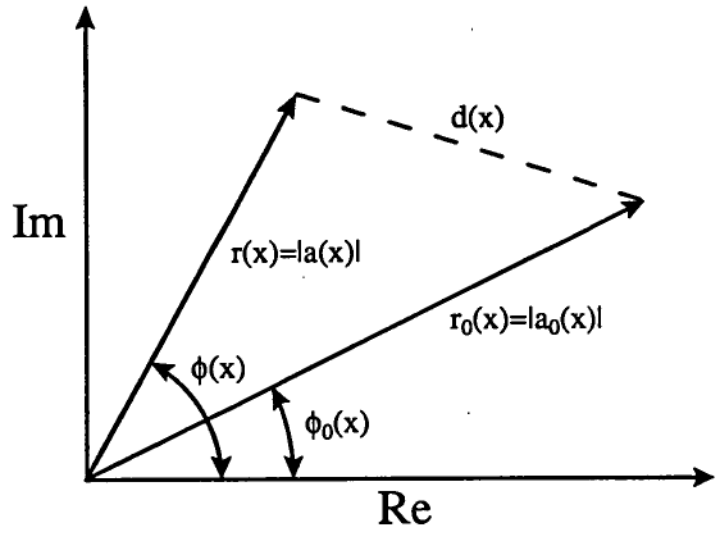


Figure 5.10: At each aperture position x the original phasor $a_0(x)$ is changed to a phasor of the projected aperture $a(x)$. The distance between them is $d(x) = |a(x) - a_0(x)|$ as shown.

A power constraint is included as before, and because $r(x) = 0$ for $x \in X_0$ there is no need to include these indices in J. Taking partial derivatives,

$$\begin{aligned}
\frac{\partial J}{\partial r(x)} &= 2r(x) - 2r_0(x)\cos(a_k x + b_k - \phi_0(x)) + 2\lambda r(x) \text{ unless } x \in X_0 \\
\frac{\partial J}{\partial a_k} &= \sum_{x \in X_k} 2xr(x)r_0(x)\sin(a_k x + b_k - \phi_0(x)) \text{ for } k = 1, \ldots, N_p \\
\frac{\partial J}{\partial b_k} &= \sum_{x \in X_k} 2r(x)r_0(x)\sin(a_k x + b_k - \phi_0(x)) \text{ for } k = 1, \ldots, N_p.
\end{aligned}$$

The first derivative, equated to zero, gives

$$r(x) = \frac{1}{1+\lambda} r_0(x)\cos(\phi(x) - \phi_0(x)). \tag{5.22}$$

This is the expected result: in figure 5.10 the length $d(x)$ is at first perpendicular to the phasor $a(x)$, and then the entire vector a is scaled by $1/(1+\lambda)$ to restore the correct power. Substituting this into the other derivatives, also equated to zero, gives for $k = 1, \ldots, N_p$

$$\sum_{x \in X_k} x{r_0}^2(x)\sin 2(a_k x + b_k - \phi_0(x)) = 0 \tag{5.23}$$

$$\sum_{x \in X_k} {r_0}^2(x)\sin 2(a_k x + b_k - \phi_0(x)) = 0. \tag{5.24}$$

The identity $2\sin\theta\cos\theta = \sin 2\theta$ was used. This author has not been able to solve this set of $2N_p$ equations for a_k and b_k, but neither has it been necessary. If convergence is being achieved the change in phase $|\phi(x) - \phi_0(x)|$ should be small at each x, and the small angle approximation can be used to obtain these equations:

$$a_k \sum_{x \in X_k} {r_0}^2(x)x^2 + b_k \sum_{x \in X_k} {r_0}^2(x)x = \sum_{x \in X_k} {r_0}^2(x)x\phi_0(x) \qquad (5.25)$$

$$a_k \sum_{x \in X_k} {r_0}^2(x)x + b_k \sum_{x \in X_k} {r_0}^2(x) = \sum_{x \in X_k} {r_0}^2(x)\phi_0(x). \qquad (5.26)$$

In this form they are recognizable as the normal equations for a least squares, straight line fit to the ordered pairs $(x, \phi_0(x))$, with a weight of ${r_0}^2(x)$ at each point. A separate fit is performed over each panel. If the projection were to be devised intuitively an identical method would probably be chosen; it is good to know that it can be justified.

When the phase difference between the initial point and the best-fit panels is large these equations are not strictly true. In practice the phases stabilized rapidly in the first couple of iterations except sometimes in the shadow region of aperture blockages. Because S_2 does not describe a blocked aperture the shadowed aperture points were not included in the projection calculations. Even so, something must be done with the shadowed points once the fits have been performed. The scheme shown in figure 5.11 was adopted: as long as $|\phi(x) - \phi_0(x)| \leq \pi/2$ equation 5.22 was taken as correct, but for obtuse angles $r(x)$ was forced to zero. It would not make physical sense for $a(x)$ to have a phase opposite to the fitted $\phi(x) = a_k x + b_k$.

5.2.3 The Algorithm, and a 1D Example

The Successive Projections algorithm for panel fitting is shown in figure 5.12. Each iteration consists of a Fourier transform to the far field domain, a projection onto S_1, an inverse Fourier transform to the aperture domain, and a projection onto S_2. This loop can be entered at different points. The starting point can be the measured far field pattern, augmented to a cover a wider angle from the main beam by surrounding the data with arbitrary field values. The effect of the iterations is to estimate the field at these extra positions consistently with the panel information provided. Alternatively, the starting point can be an aperture current function with arbitrary amplitude and phase profiles, finely sampled so that the corresponding far field pattern

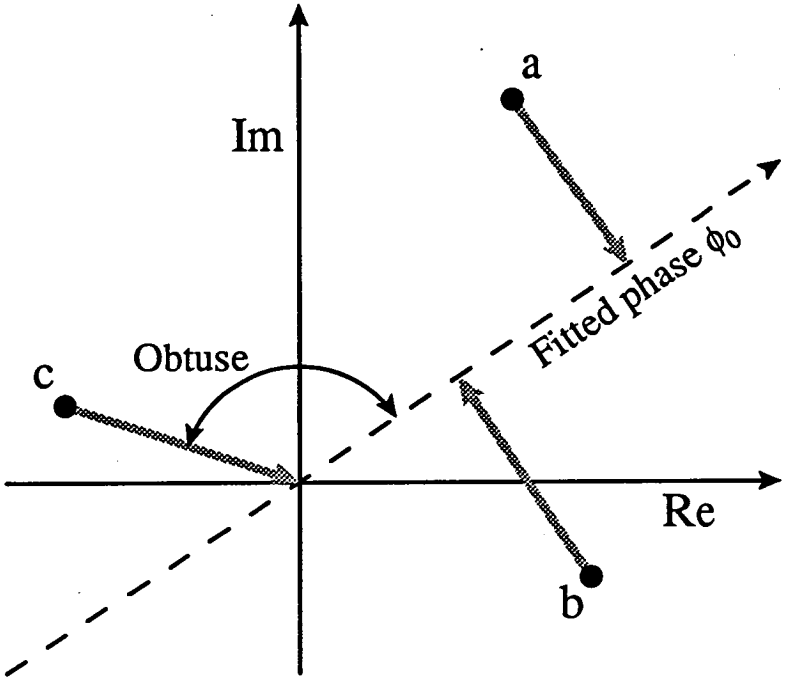


Figure 5.11: At each aperture point, the complex current is projected to its component along the fitted phase ϕ_0 as shown for a and b here. If the phase change is obtuse, as for c, this operation is not sensible, and the current is set to zero.

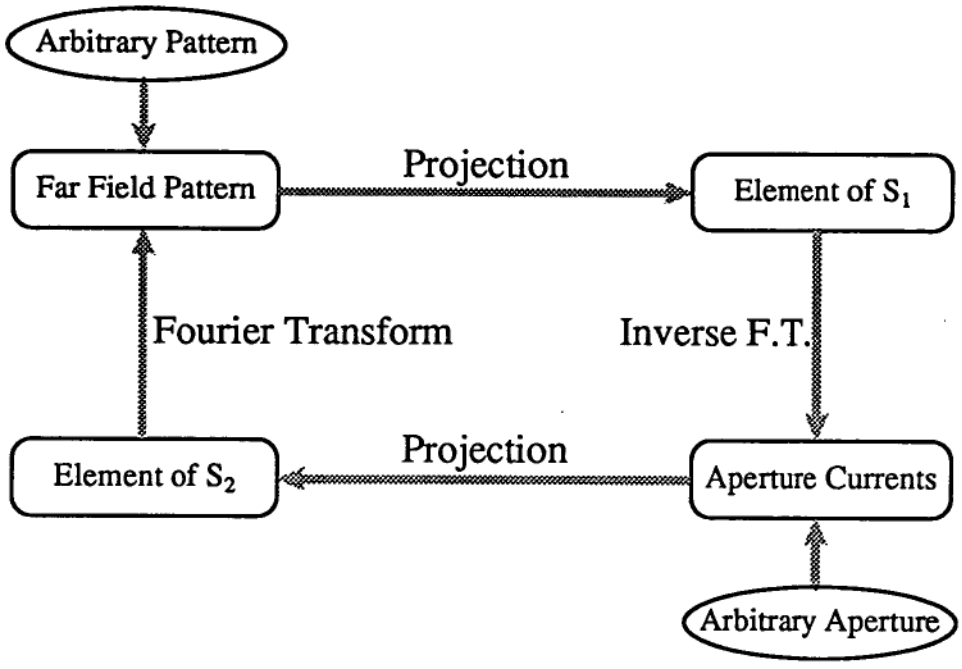


Figure 5.12: The Successive Projections algorithm: one 'iteration' is one cycle through this loop, requiring two Fourier transforms and two projections.

covers a wide angle $W_u' = 1/\delta_x'$. A frequently used starting point is the measured far field pattern augmented with zeros to a wider angular coverage. In the aperture domain this gives a smooth interpolation of the aperture function returned by traditional processing, where only an inverse Fourier transform is used.

As mentioned above, a span W_u' of 4 to 5 times the original W_u is usually quite adequate. What happens in the two domains is pictured in figure 5.13 for $W_u' = 3W_u$. The extra far field samples create aperture samples lying between the original 'tall arrows'. The increased number of points in each panel results in a better estimated profile.

The effectiveness of the method is demonstrated by a one-dimensional example. The aperture simulated for figure 5.1 is reproduced in figure 5.14(a), overlaid by the inverse Fourier transform of $N_u = 15$ far field samples. Sampling parameters were such that an average of $P_p = 2.0$ aperture samples lay on each panel. In figure 5.14(b) the far field samples have been augmented with zeros until $N_u' = 71$, and SP applied for just 3 iterations. The correct panel positions have been almost entirely reclaimed.

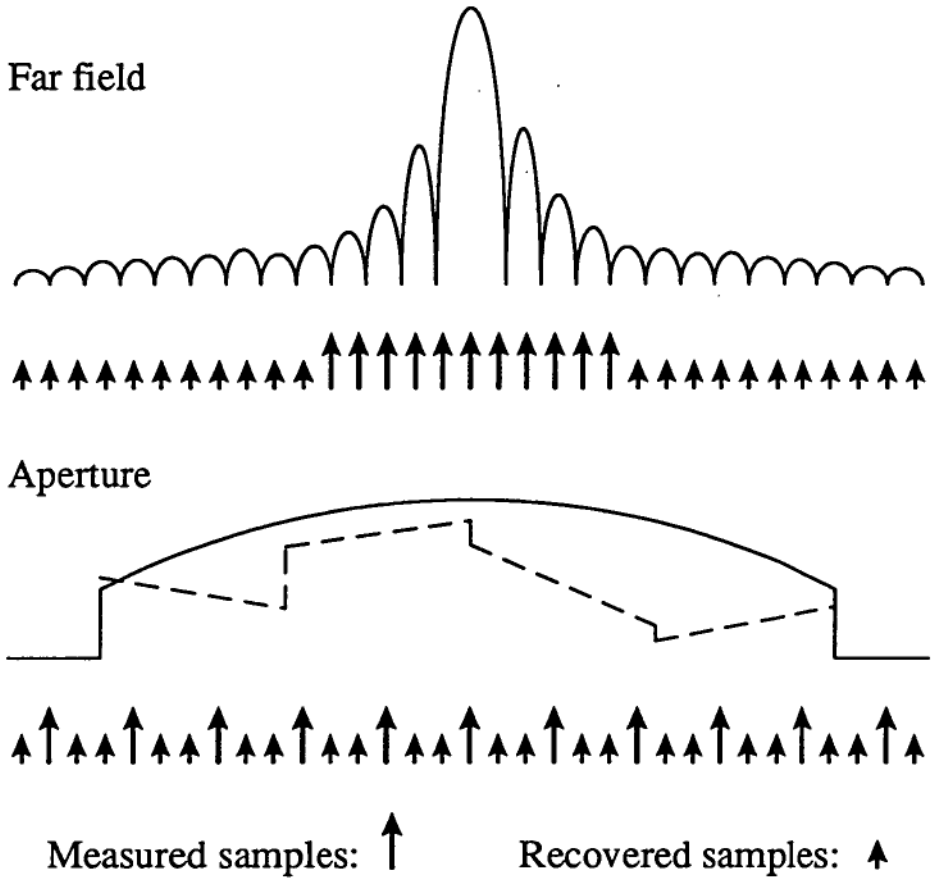


Figure 5.13: A picture showing the significance of the additional far field samples recovered using Successive Projectons.

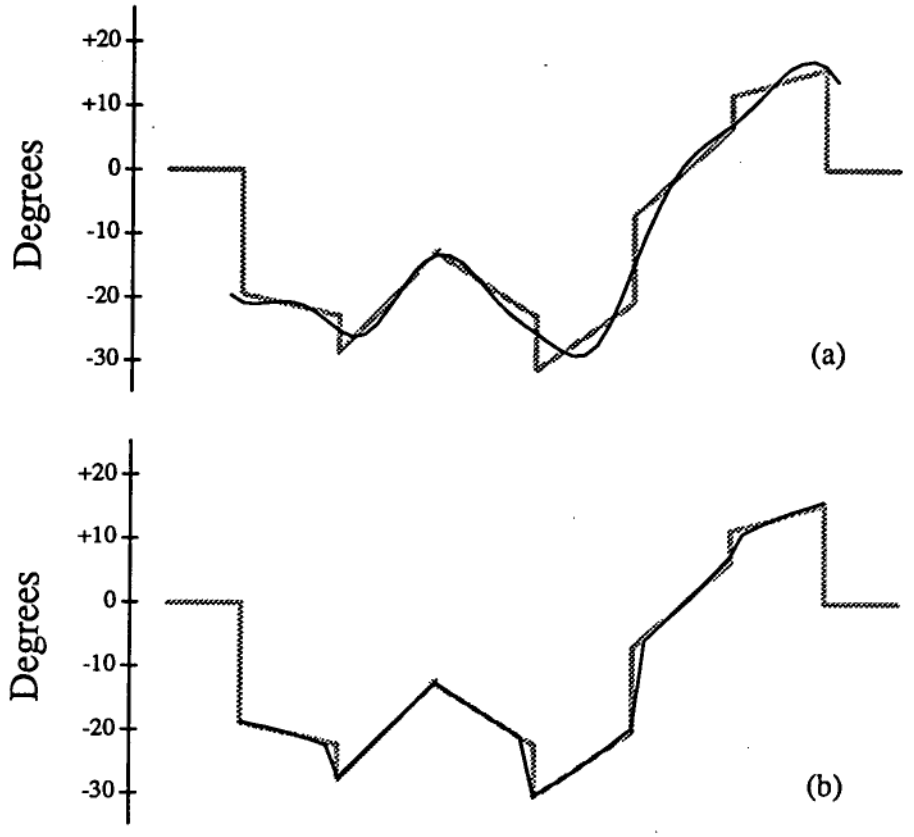


Figure 5.14: A simulated one-dimensional aperture current function (a), and the result of 3 Successive Projections iterations (b).

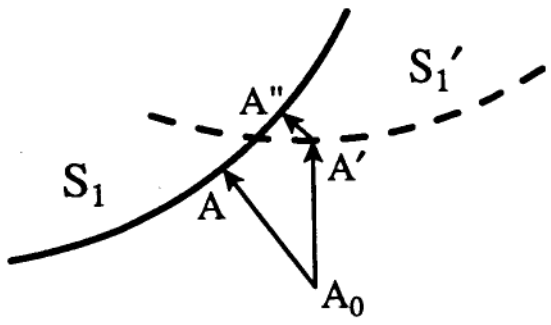


Figure 5.15: From some point A_0 the projection onto S_1 is A. With measurement errors the projection onto S_1' is A'. The distance of A' from the correct set S_1 is the distance to its projection onto S_1, labelled A''.

5.2.4 The Effect of Measurement Errors

If due to measurement errors the correct far field pattern $A_m(u)$ is replaced by a slightly different $A'_m(u)$, then an incorrect set S_1' is used for Successive Projections in place of S_1. The effect on projections is shown in figure 5.15. Since there is no reason to prefer A over A'', the distance $\|A' - A''\|$ is a measure of the error in the projection caused by the measurement errors. It is the separation of the boundaries of S_1 and S_1'. This distance will now be found.

Let the projections be

$$A'(u) = \begin{cases} \mu' A'_m(u), & u \in M \\ \rho' A_0(u), & u \in R \end{cases} \quad \text{and } A''(u) = \begin{cases} \mu'' A_m(u), & u \in M \\ \rho'' A'(u), & u \in R \end{cases} .$$

By equation 5.21 the squared distance from the first projection to the second is

$$\|A' - A''\|^2 = 2P\left(1 - \frac{1}{\rho''}\right).$$

If real and imaginary components are written as in section 5.2.2, and if the errors are given by

$$A'_m(u) - A_m(u) = E(u) = p(u) + jq(u)$$

then these sums can be found:

$$\begin{aligned} S'_R &= \sum_{u \in R} \left[x_0^2(u) + y_0^2(u)\right], \\ S''_R &= \sum_{u \in R} \left[x'^2(u) + y'^2(u)\right] \end{aligned}$$

$$
\begin{aligned}
&= \sum_{u\in R}\left[(\rho' x_0(u))^2 + (\rho' y_0(u))^2\right] \\
&= \rho'^2 S'_R \\
&= P - \mu' P'_M \text{ by previous identities,} \\
S''_M &= \sum_{u\in M}\left[x'(u)x_m(u) + y'(u)y_m(u)\right] \\
&= \sum_{u\in M}\left[\mu' x'_m(u)x_m(u) + \mu' y'_m(u)y_m(u)\right] \\
&= \mu'(P_M + S_E), \text{ where} \\
S_E &= \sum_{u\in M}\left[p(u)x_m(u) + q(u)y_m(u)\right].
\end{aligned}
$$

Equation 5.20 can be used to find ρ''. Firstly, without loss of generality, assume that the power in the measurements is the same with and without errors: $P'_M = P_M$. This can always be arranged by scaling A_m or A'_m. Then

$$
\begin{aligned}
\frac{1}{\rho''^2} &= \frac{S''_R}{P} + \frac{{S''_M}^2}{PP'_M} \\
&= \frac{P - \mu'^2 P_M}{P} + \frac{\mu'^2(P_M + S_E)^2}{PP_M} \\
&= 1 + \mu'^2\left(\frac{2S_E}{P} + \frac{{S_E}^2}{PP_M}\right)
\end{aligned}
$$

If the errors are small compared to the measurements, $S_E \ll P$ and the second term can be ignored beside the first. The number μ' is the scaling applied to the measurements when projecting onto S'_1, and as long as P_M is similar to P it will be close to unity. Otherwise, an unrealistic portion of the total power is going to the recovered, distant sidelobes. Using $\sqrt{1+x} \simeq 1 + x/2$ gives

$$
\frac{1}{\rho''} \simeq 1 + \frac{\mu'^2 S_E}{P},
$$

and so

$$
\|A' - A''\|^2 = 2P\left(1 - \frac{1}{\rho''}\right) \simeq -2\mu'^2 S_E.
$$

It only remains to prove that this is positive. The constraint $P'_M = P_M$ is enough to determine the sum S_E. In appendix E it is shown that $2S_E = -\|E\|^2$, so if $\mu' \simeq 1$ the distance from any projection onto S'_1 to the correct set S_1 is

$$
\|A' - A''\| \simeq \|E\| = \sqrt{\sum_{u\in M} |A'_m(u) - A_m(u)|^2}. \tag{5.27}
$$

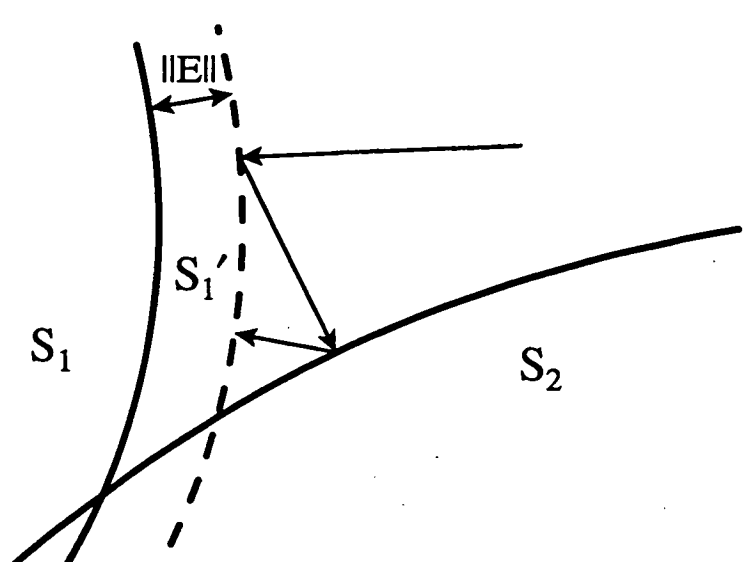


Figure 5.16: A representation of the statement, "noise power is conserved under Successive Projections." The projections onto S_1', which has been affected by errors $E(u)$, are never farther than $\|E\|$ from the boundary of the error-free S_1.

Thus SP iterations will never wander farther than $\|E\|$ from S_1, as illustrated simplistically in figure 5.16. By analogy with the traditional Fourier transform estimation of aperture currents, it can be said that the noise power is conserved under Successive Projections.

5.3 Application to Mount Pleasant Data

5.3.1 Parameters and Results

The success of one-dimensional simulations, and the encouraging results of the tests in section 6.1, indicated that SP might be applied to experimental data with confidence. Both the carrier signal and the beacon signal maps of the 26 m antenna were processed in this way before estimating the panel corner adjustments described in section 3.3. Before displaying the results some details will be given.

The measured far field maps were on a grid of 53×53 samples at intervals $\delta_u = \delta_v = 0.00075$. This grid was extended to 241×241 samples for SP, this number being chosen mainly for consistency with earlier two-dimensional

tests. Thus the far field span was increased by a factor

$$\frac{W_u'}{W_u} = \frac{W_v'}{W_v} = 4.55.$$

In the aperture domain this gave a corresponding decrease in the sampling interval, and hence an increase in the average number of points per panel. The parameters at each frequency were:

	12725 MHz		12373 MHz	
	Old	New	Old	New
Map Width on Sky	2.3°	10.4°	2.3°	10.4°
Aperture δ_x	0.593 m	0.130 m	0.610 m	0.134 m
Points per Panel	5.35	110.9	5.10	104.7

The large number of points lying on an average panel ensured that smoothing near panel edges had a minimal effect on fitted panel positions. Blockage shadowing was taken into account in these figures. Even the smallest, partly obscured panels contained over 40 points in the augmented maps.

Those aperture points in the geometrical shadow of the feed housing or its support legs were ignored when projecting onto S_2. The steep phase gradients caused by diffraction then had little effect: see figures 4.2 and 4.3. Phases which differed greatly from the fitted planes resulted in a small or zero projected amplitude at that point, as shown in figure 5.11. This meant that shadowed regions were effectively zeroed prior to a Fourier transform to the far field domain, in accordance with the null field hypothesis of section 4.1.1.

Ten SP iterations had a dramatic effect on the aperture maps. The starting point was the interpolated aperture obtained by extending the far field with zeros. In figure 5.17 one quadrant of the interpolated 12725 MHz surface error map has been enlarged and plotted with a fine 0.2 mm contour interval. In figure 5.18 the same quadrant is shown after the iterations. Panel structure which did not appear in the original maps has been restored convincingly. There are some panels, apart from those near shadow regions, which contain distinct 'hills' or 'valleys', two in particular in the fifth ring from the edge. Inspection of the surface confirms that these panels are misshapen.

In section 3.3.3 a systematic change in the surface shape apart from the panel adjustments was discovered. This prevented an end-to-end test of Successive Projections, because it was not possible to compare directly the 'before' and 'after' positions of an individual panel. Section 6.1 provides alternative convincing evidence that the algorithm recovers panel positions reliably.

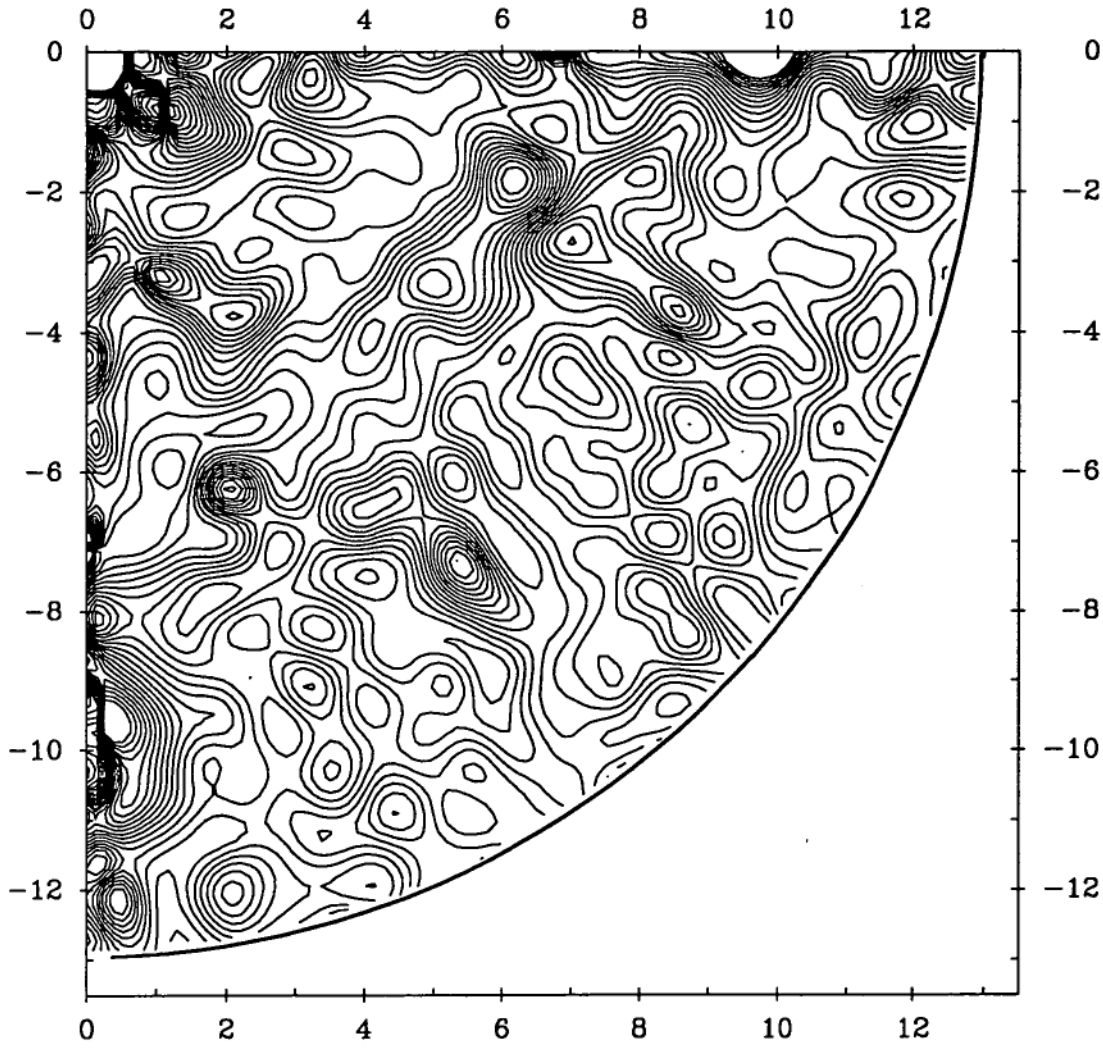


Figure 5.17: One quadrant of the 12725 MHz surface error map before any SP iterations are applied. The contour interval is 0.2 mm, and the scales are in metres.

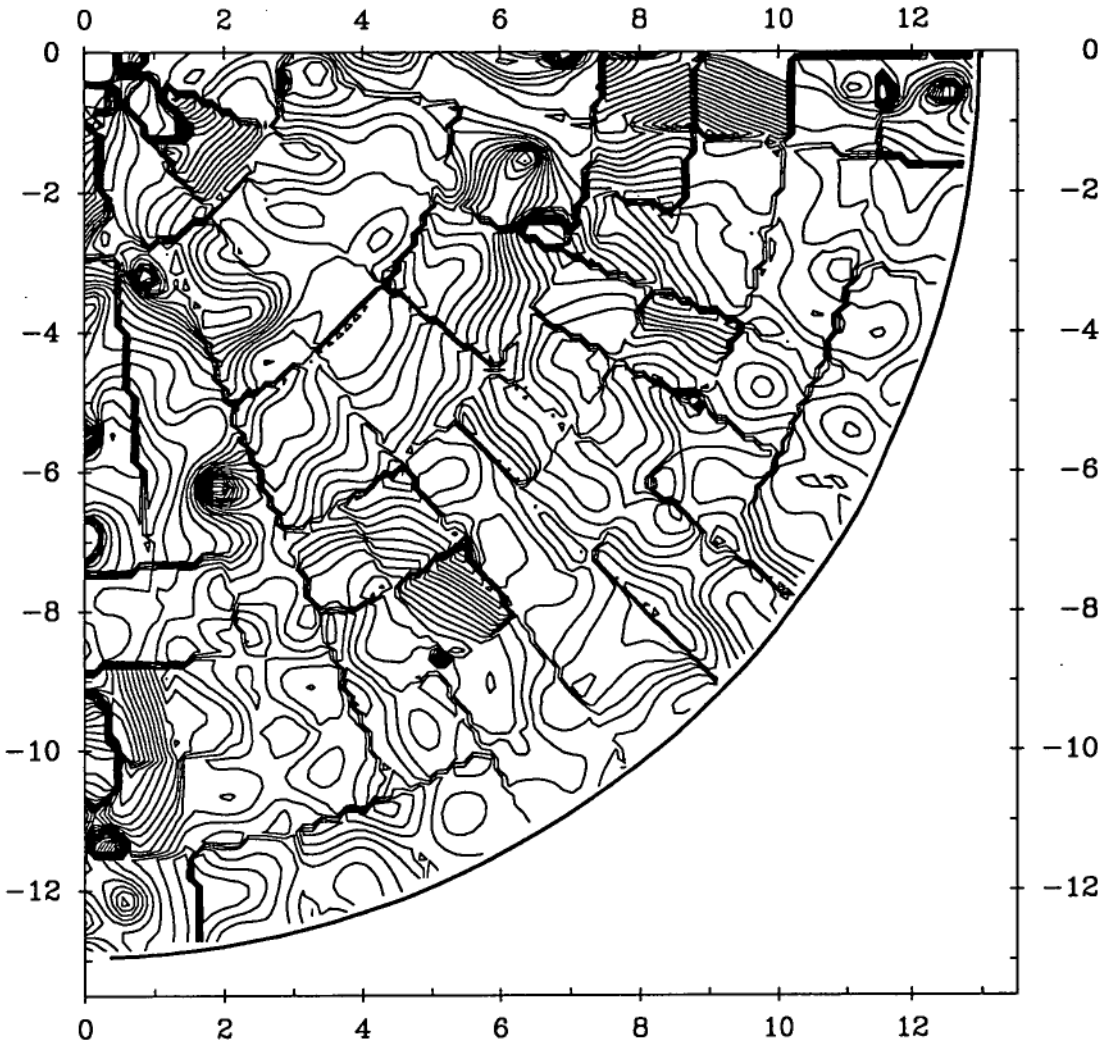


Figure 5.18: One quadrant of the 12725 MHz surface error map after 10 SP iterations have been applied. The contour interval is 0.2 mm, and the scales are in metres.

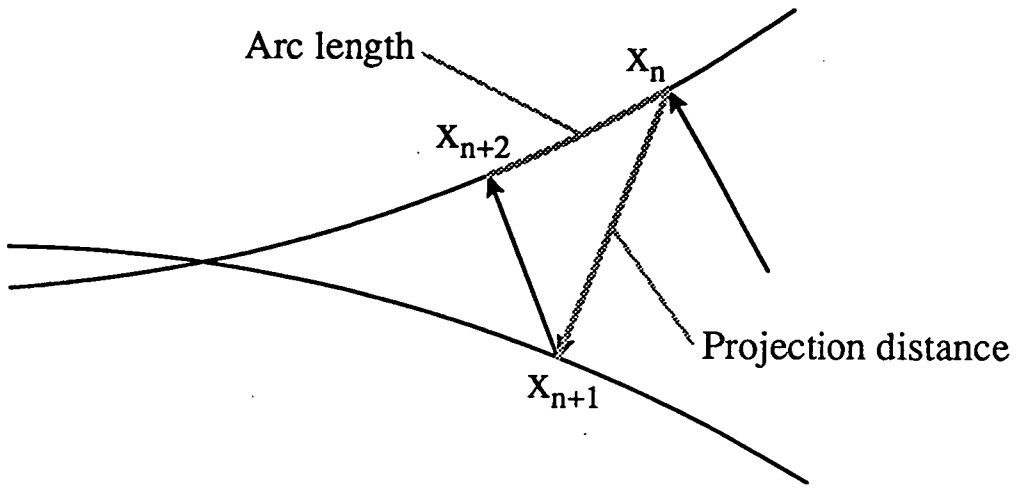


Figure 5.19: Projection lengths and arc lengths for two sets in the plane.

5.3.2 Looking at the Convergence

It is possible to get an idea of the progress of the algorithm by looking at the lengths of the projections. If the sets were convex a geometric rate of convergence would be expected, both for the distance to the limit point and the distance from one projection to the next, as shown in section 5.1.2. The latter can be calculated readily. Another quantity which may be useful is the separation of a projection from the previous one onto the same set; it is shown in figure 5.19. In the two-dimensional plane this distance would approximate the arc length along the set's boundary.

Table 5.1 shows these quantities and demonstrates the good behaviour of SP for the average map made at 12725 MHz. Dimensions in the far field domain have been divided by $\sqrt{N'_u N'_v} = 241$ so they can be compared with dimensions in the aperture domain, as dictated by equation 5.10. Because the starting point was the interpolated aperture, the first operation was a projection onto S_2. Thus the sequence of projection distances is $33.3, 28.6, 26.3, \ldots$ and of arc lengths is $10.55, 7.20, 6.37, \ldots$. If not quite geometric, they certainly decrease monotonically and in a very regular way.

It is interesting that the arc lengths begin at about 1/3 of the projection lengths, but decrease to about 1/7 after 10 iterations. This shows that most progress is made at the beginning, as expected from the one-dimensional example which gave a good result after only 3 iterations. It may also show that the projections entered a tunnel: very small arc lengths are a symptom of tunnel behaviour for sets in the plane like those in figure 5.6. If convergence

Iteration Number	Projection Distance Onto S_2	Onto S_1	'Arc Length' On S_2	On S_1
1	33.3	28.6		
2	26.3	25.3	10.55	7.20
3	24.4	23.9	6.37	5.12
4	23.3	22.9	4.98	4.35
5	22.5	22.2	4.26	3.90
6	21.9	21.6	3.82	3.57
7	21.3	21.0	3.52	3.34
8	20.8	20.6	3.28	3.14
9	20.3	20.1	3.09	2.98
10	19.9	19.7	2.94	2.85

Table 5.1: Convergence of Successive Projections iterations for the carrier data extended to a 241 × 241 grid. The units are arbitrary but can be compared with the square root of the total aperture domain power, or 208.8.

was hindered in this way, an investigation of acceleration techniques might be worthwhile. On the other hand, panel positions were quite stable enough after a small number of iterations, so acceleration was not a priority for this study.

Convergence from a range of different starting points was not investigated using the data. This was because noise, sampling parameters, and aperture blockage may all have caused vagaries of the convergence path and could not be studied independently. It was more effective to try different starting points using simulated data, where at least the limit point was known in advance. This experiment is described in section 6.1.

5.3.3 Efficiency and Flexibility

The time taken for each SP iteration was considerable, and was dominated by the two 241 × 241 point discrete Fourier transforms. The data array is large, requiring 908 blocks of 512 bytes for its storage in single precision. CPU times for a single iteration were compared on three different *Digital VAX* computers with these results:

Machine	Time (m:s)
VAX 11/750	35:19
MicroVAX 3500	4:10
VAXstation 3100	1:31

No effort was made to improve the efficiency of the Fourier transform, for example by choosing a power of two for the array size. An optimal size for the recovered region of the far field could be investigated. However, for this study all that mattered was that the size was sufficient to estimate panels well. On modern machines the algorithm still executes in a small fraction of the time required to record a map (about 4 hours), let alone to adjust the panels afterwards, and this is really the important issue for practical use.

At this point it is appropriate to consider another technique for fitting panels to Microwave Holography data, one of a kind that has been used with success by other workers [28]. This is a gradient search technique in the space of panel parameters — tilts and vertical shifts — to minimize the difference between derived and measured far field patterns. The contribution to the total field from each panel may be calculated analytically if it is approximated that each panel is uniformly illuminated; this is a reasonable approximation for a large reflector with many panels. The tilts and shift of the panel enter the expression as parameters, so using the symbols from section 2.4.1

$$A_k(u,v) = F_k(u,v;\beta_k,\alpha_k,\ell_k,r_k) \text{ for } k = 1,\ldots,N_p.$$

Here r_k is the illumination amplitude on the panel k.

The total field is the sum of these contributions, and can thus be expressed explicitly as a function of $4N_p$ parameters:

$$A(u,v) = \sum_{k=1}^{N_p} F_k(u,v;\beta_k,\alpha_k,\ell_k,r_k).$$

(An expression like this will be discussed more fully in section 6.2.) Derivatives can be calculated, and a gradient search technique such as the conjugate gradient method can be used to minimize the error

$$J = \sum_{\text{meas}} |A(u,v) - A_m(u,v)|^2.$$

The steps of this method will be much more rapid than SP projections because no discrete Fourier transforms are necessary. However there will be many more steps; for the Mount Pleasant 26 m antenna 1008 steps would be

Iteration Number	Projection Distance Onto S_2	Onto S_1	'Arc Length' On S_2	On S_1
1	31.7	26.9		
2	24.4	23.4	10.61	7.20
3	22.4	21.8	6.30	5.03
4	21.3	20.9	4.86	4.21
5	20.4	20.1	4.13	3.75
6	19.8	19.5	3.67	3.41
7	19.2	18.9	3.36	3.17
8	18.7	18.4	3.12	2.98
9	18.2	18.0	2.94	2.83
10	17.8	17.6	2.80	2.70

Table 5.2: Convergence of Successive Projections iterations for the carrier data, this time with twisting of panels permitted.

only one for each unknown parameter, and typically many times this number would be required in a search.

This is certainly a viable alternative method, but it lacks the flexibility of the SP formulation. Here is an example: alter the definition of the set S_2 so that twisted panels are permitted. The necessary equations are given in section 2.4.2, and appendix C gives a least squares fitting method which can be incorporated into the projection onto S_2. The additional computation is minimal, and when the altered algorithm was applied to the 26 m data no change in execution speed was seen. The projection lengths and arc lengths are given in table 5.2, and convergence is even slightly more rapid than in table 5.1. Perhaps this is because the new S_2 represents the real surface more closely.

There can be no similar extension to the gradient search method: it is the linear variation of phase which enables the Fourier integral of a tilted panel to be evaluated. Without it, the integral cannot be expressed in terms of simple functions. This is one way in which the flexibility of Successive Projections is of advantage for practical problems.

Chapter 6

Behaviour of the Algorithm

Now that the Successive Projections algorithm for panel fitting has been described and applied, its usefulness and reliability must be demonstrated. Section 6.1 makes use of numerical experiments to study convergence, firstly for a particular example and then for an ensemble of convergence paths to highlight generic features. Section 6.2 takes a different approach by studying the solvability of a restricted form of the panel fitting problem that can be analysed rigorously. Both sections lead to the same conclusions on the significance of some experimental parameters, and an optimal design for Microwave Holography experiments is indicated. Section 6.3 is an excursion to show an interesting extension of a different Successive Projections algorithm.

6.1 Convergence Behaviour

6.1.1 Is It Always Monotonic?

When applied to the Mount Pleasant data, Successive Projections converged with a smoothly and monotonically decreasing projection distance and 'arc length'. The limit point was not known, of course, so the distance of the projections from the limit could not be monitored. Was it true that

$$\lim_{n\to\infty} \|A_n - A\| = 0 \text{ and } \lim_{n\to\infty} \|a_n - a\| = 0,$$

as claimed for convex sets in section 5.1.2? In a simulated experiment the correct aperture current function is known, so the distance to it from each projection can be monitored. Ideally it would be the limit point of the sequence of projections, but in reality the limit point might be anywhere on

the boundary of a possibly large intersection $S_1 \cap S_2$. The distance to the original aperture is a measure of progress towards its complete recovery.

Another advantage of a simulated experiment is that its parameters can be changed easily. The most important ones are the map size $N_u \times N_v$ and the sampling intervals δ_u and δ_v. These affect the measured data and hence the set S_1, so the parameters control the shape of the sets' intersection. This is how the information content of the data has a bearing on the convergence of SP. An efficient experiment is one where good convergence is obtained with the least possible data.

A one-dimensional aperture was simulated, and its radiation pattern calculated by a discrete Fourier transform. The aperture had the same dimensions as the Mount Pleasant antenna: a width of 26 m with 15 panels across. It was given a 12 dB illumination taper with the profile of equation 4.32. Like the 12725 MHz measurements, the radiation pattern was sampled at an interval $\delta_u = 0.00075$, so $\lambda = 23.6$ mm and $\kappa = 0.825$. The number of samples N_u was varied, and because the one-dimensional SP iterations were very fast a large number, 1000, could be applied in each case. The sampled far field pattern was, as usual, extended with zeros to $N'_u \simeq 241$ to give a starting point. A significant parameter is P_p, the average number of points per panel in the aperture map obtained by a DFT of the sampled far field alone. This is the traditional holographic processing method. The sample interval in the aperture is given by $\delta_x = 1/W_u = 1/N_u\delta_u$, from which can be derived

$$P_p = \frac{\kappa N_u}{N_p} \tag{6.1}$$

where N_p is the number of panels across the physical aperture. These values of N_u were used:

N_u	31	33	35	45
P_p	1.70	1.81	1.92	2.47

It is worthwhile to plot three quantities as a function of iteration number: the rms phase difference of the aperture estimate a_n from the original aperture a, the distance $\|a_n - a\|$, and the length of the projection onto S_1. The former is the most important measure for holography, while the other two are concerned more with the geometry of the sets. All three are plotted for the above values of N_u in figures 6.1 to 6.4. These plots use logarithmic scales so that geometric convergence would give straight lines; the equations

are

$$\begin{aligned}
\text{rms}\,\Delta\ (\text{dB}) &= 20\log_{10}\left(\frac{\sum_x |a_n(x)|^2 |\phi_n(x) - \phi(x)|^2}{\sum_x |a_n(x)|^2}\right) \\
&\quad - \text{ditto for } n = 0 \\
\delta a\ (\text{dB}) &= 20\log_{10}\left(\frac{\|a_n - a\|}{\|a_n\|}\right) \\
\text{Length}\ (\text{dB}) &= 20\log_{10}\left(\frac{d_P(a_n)}{\|a_n\|}\right).
\end{aligned}$$

where $d_P(a_n)$ is the distance to a_n from the previous projection. Thus rms Δ is compared with the original rms phase difference, and the distances are compared with the norm of a_n, or the square root of its total power.

The change in behaviour as $P_p \uparrow 2$ is dramatic. The quality of the recovered phase improves until at $P_p = 1.92$ the rms is 1/8 of the original phase error. By $P_p = 2.47$ no further improvement has occured, except that the limit is reached a little faster. This is significant because it shows that SP is giving no more than it receives. Two points are needed on a one-dimensional 'panel' to determine its tilt, so in this sense far field data which gives P_p significantly less than 2 does not contain complete tilt information. It would be worrying if such data were magically imbued with this knowledge just by applying SP, whose purpose is to make use of existing information on panel boundaries.

The development of a minimum in the graph of δa is surprising when the projection length $d_P(a_n)$ decreases so steadily in every case. This does not indicate that convergence is failing, however, because as mentioned above the original aperture is not necessarily the limit point of the sequence. As usual, it is difficult to imagine the shape of $S_1 \cap S_2$, but the qualitative change in convergence about $N_u = 35$ must be due to the contraction of S_1 as its membership criterion becomes more strict. That is, more functions are excluded from membership when the pattern is specified at a larger number of points. Even if the reason is unclear, the way the minimum and subsequent rise occurs can be seen by graphing the aperture function. Figure 6.5 shows the aperture amplitude and phase at the minimum in the graph for $N_u = 45$, which is after 5 iterations, and the same after 1000 iterations. Although the phase is noticeably better the amplitude has developed distortions near large discontinuities in phase. The set S_2 is very strict about phase and very loose about amplitude, so this is reasonable.

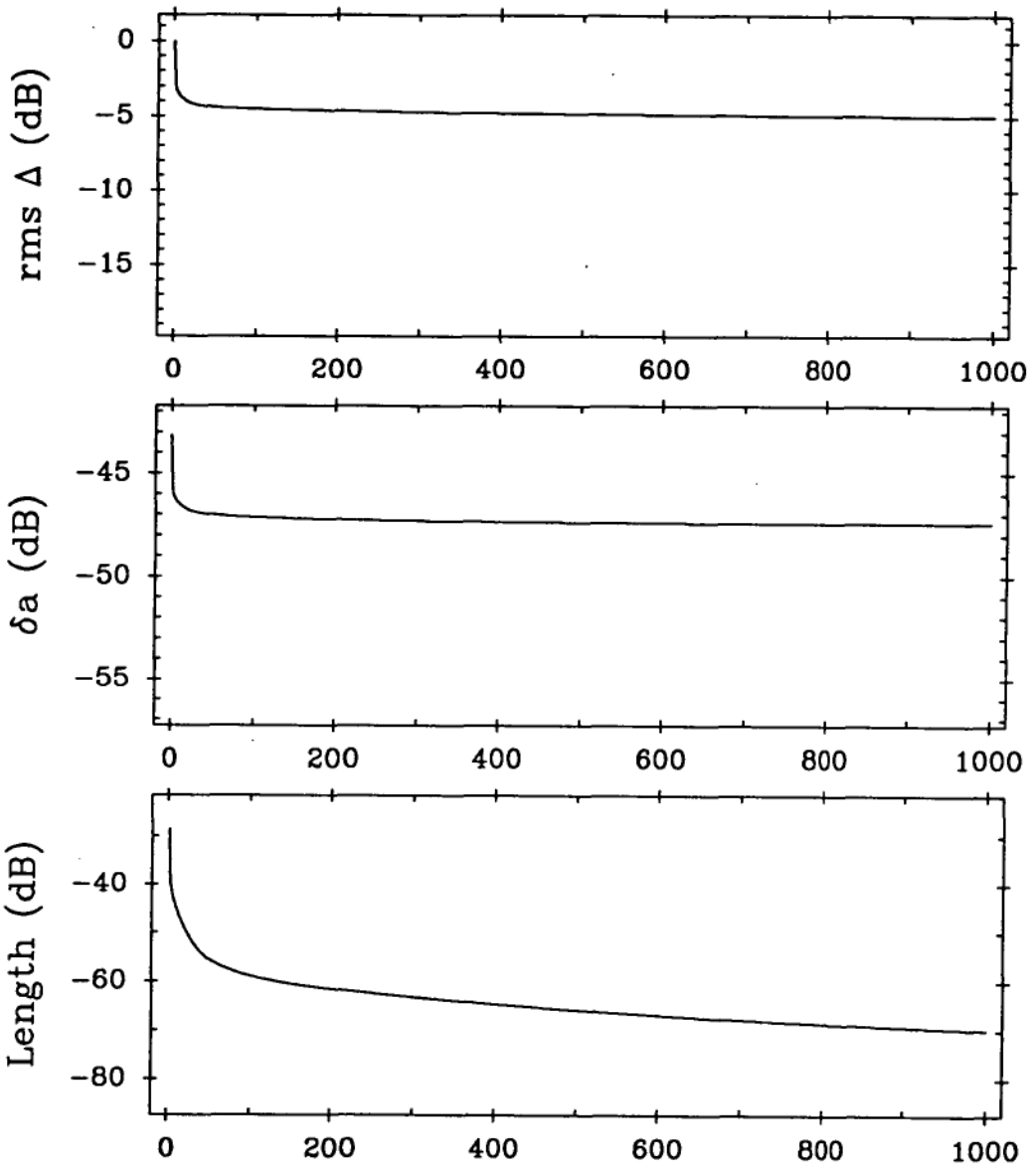


Figure 6.1: The convergence of the rms phase difference from the original aperture, the Hilbert space distance to the original aperture, and the length of the projection onto S_1. They are plotted on logarithmic scales against the SP iteration number. The number of far field samples was $N_u = 31$.

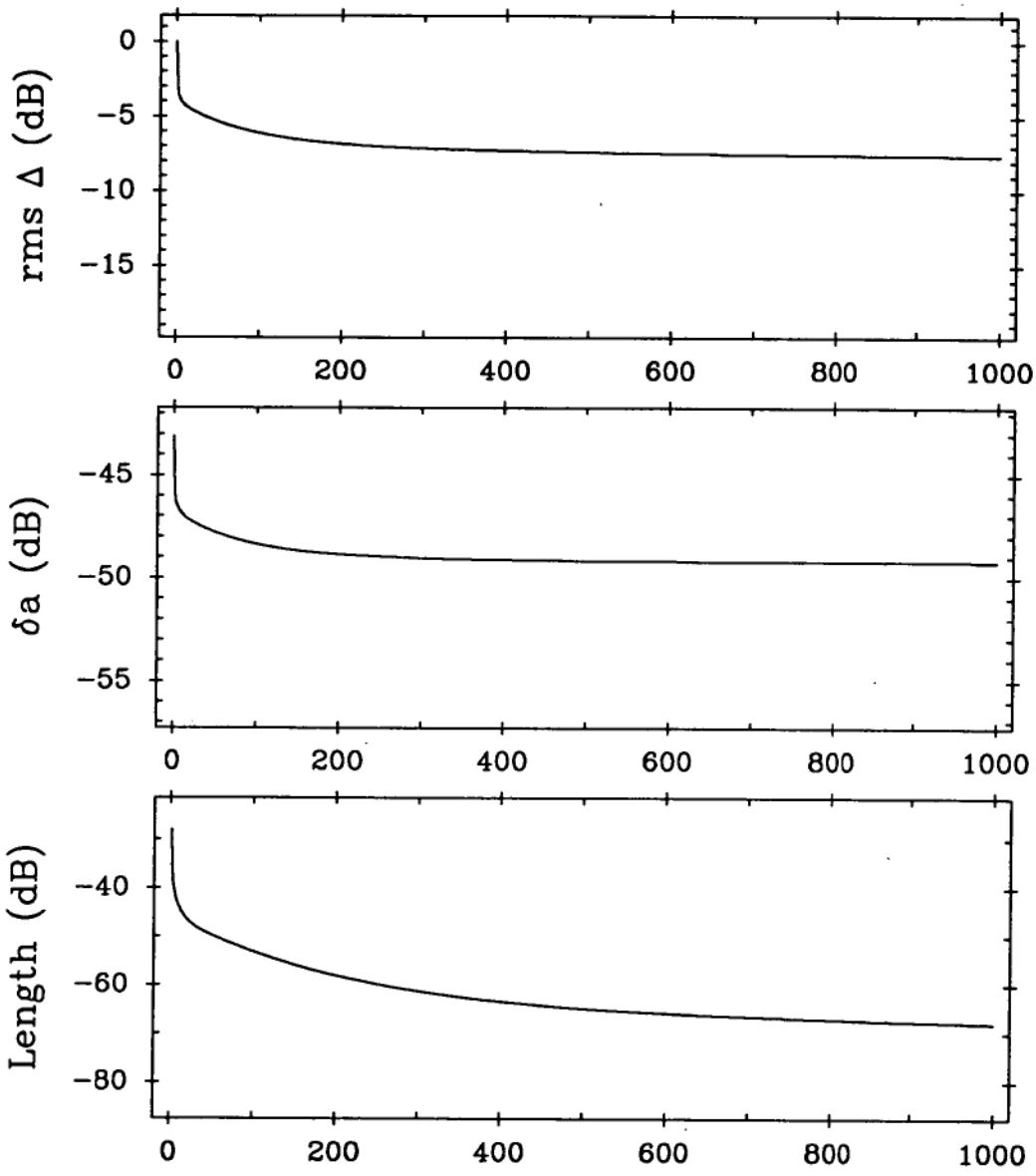


Figure 6.2: Convergence graphs for $N_u = 33$.

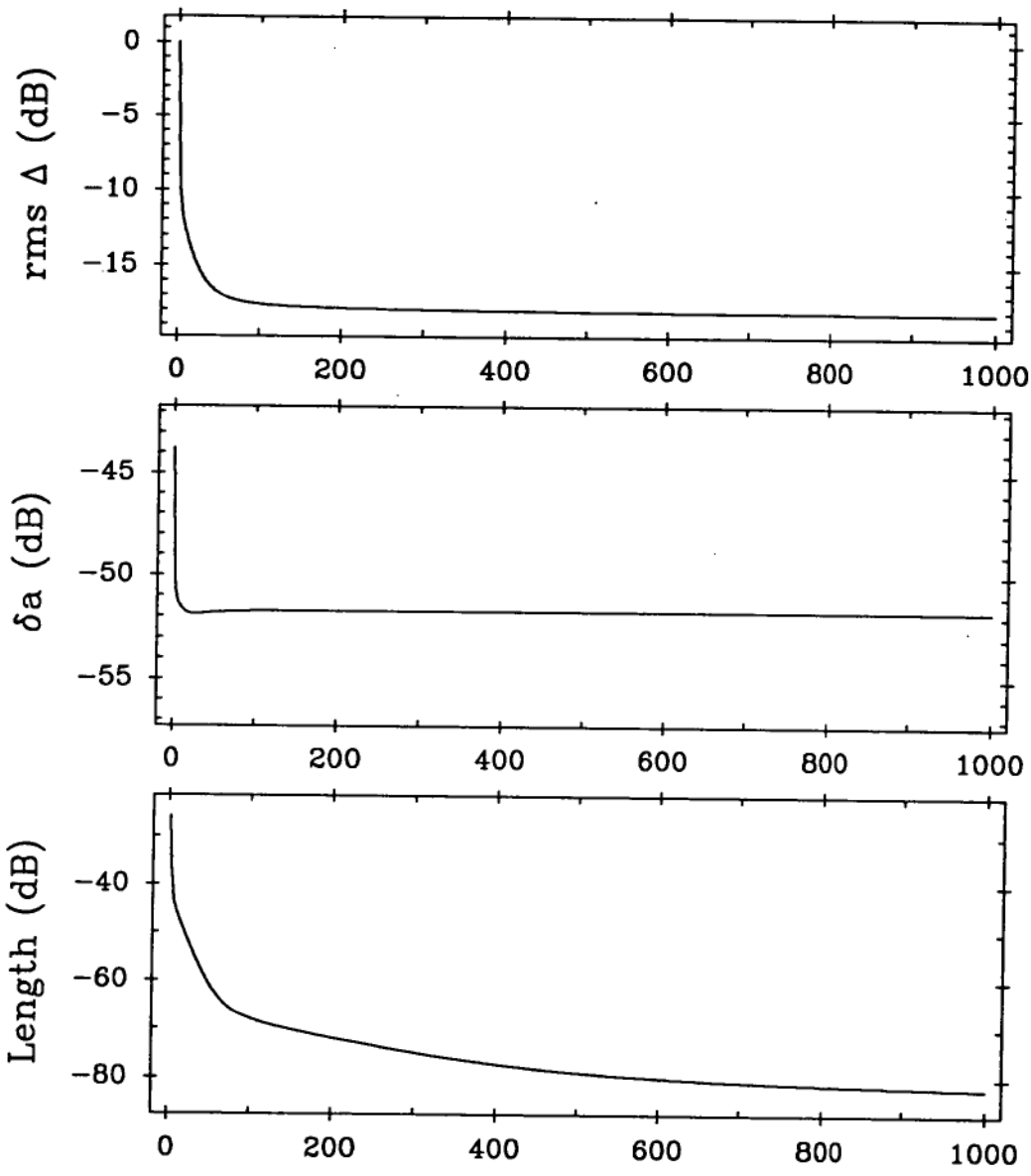


Figure 6.3: Convergence graphs for $N_u = 35$.

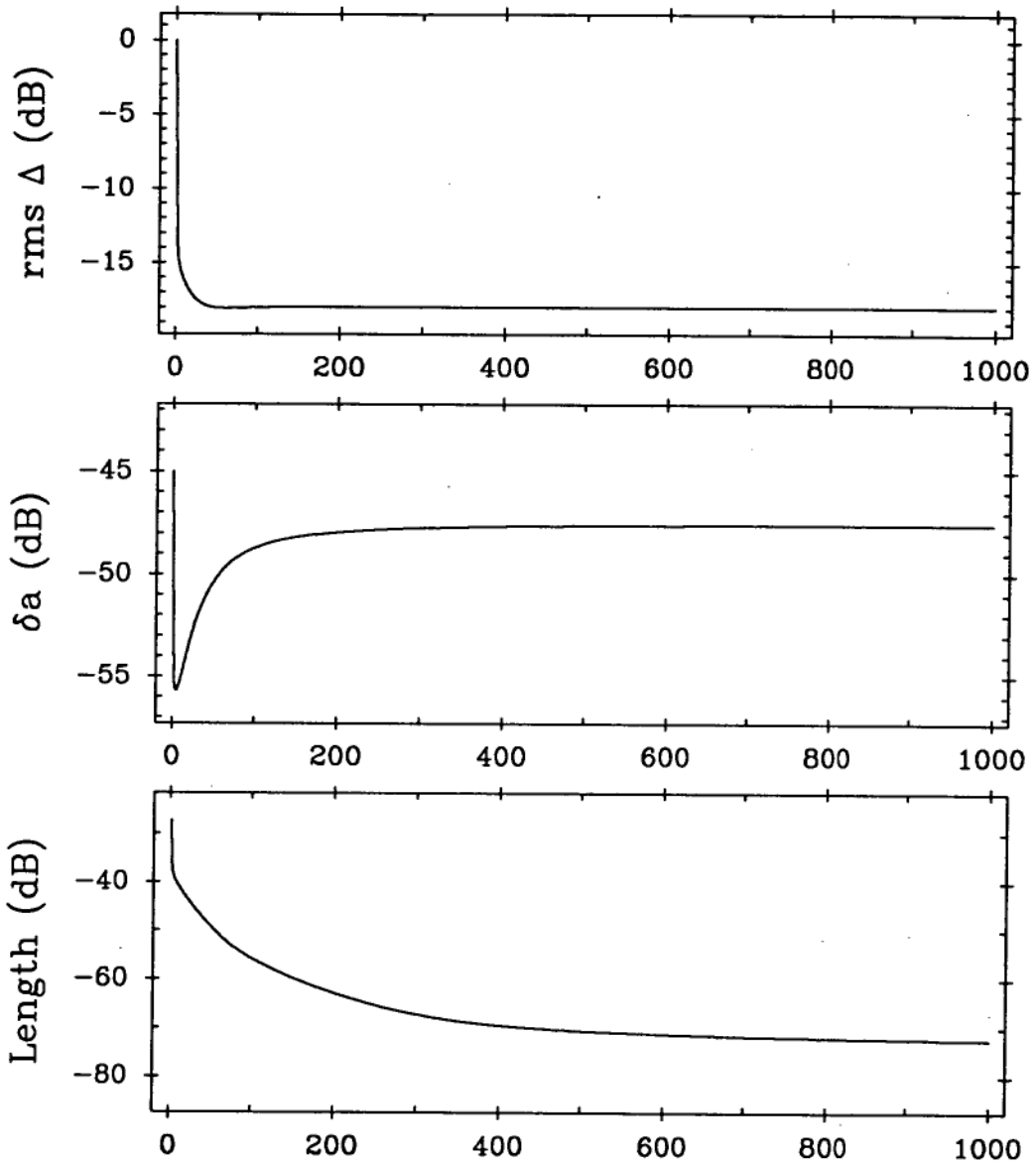


Figure 6.4: Convergence graphs for $N_u = 45$.

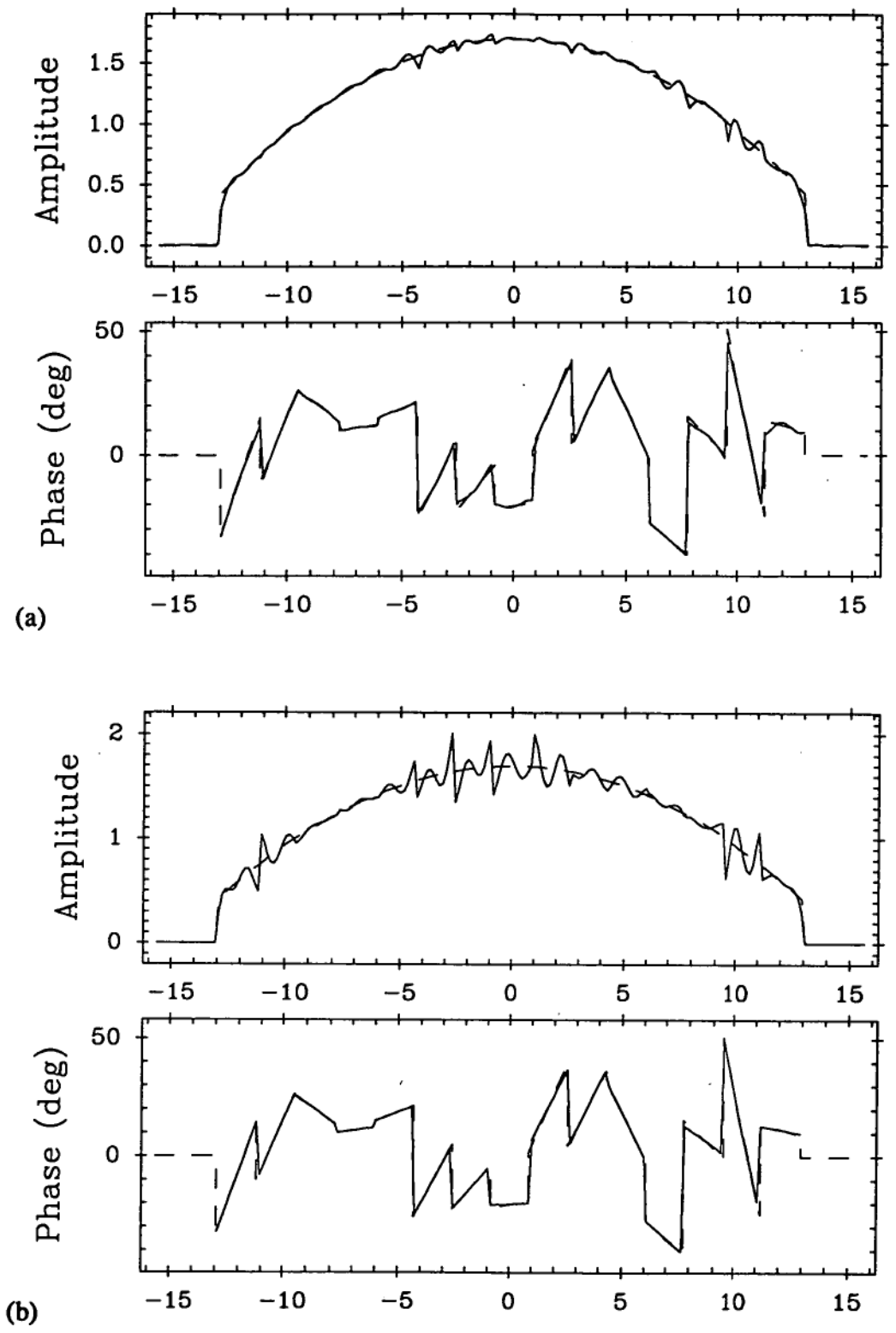


Figure 6.5: The aperture amplitude and phase for $N_u = 45$ after (a) 5 SP iterations and (b) 1000 SP iterations, with the original, simulated aperture shown in dashes. The horizontal scales are in metres.

A study of graphs of the aperture estimates also helps in understanding the behaviour of rms Δ. The aperture amplitude and phase after 100 iterations is shown for $N_u = 33$ and for $N_u = 35$ in figure 6.6 (a) and (b). In (a) most panels have been incorrectly estimated, while in (b) nearly all panels have been recovered well. A graph for $N_u = 37$, or $P_p = 2.03$, would show the 4^{th} and 5^{th} panels also recovered well.

6.1.2 Changing the Starting Point

It is important that consistent results are obtained with different choices of the starting point of Successive Projections iterations. Changing the starting point is a way of probing the shape of $S_1 \cap S_2$, because the position of the limit point on the boundary is likely to change too. A wide spread of results would indicate that the intersection was not small enough to specify the correct aperture function properly. Increasing the number of far field samples is a way of contracting S_1 and, perhaps, the intersection too. Does the number of points per panel have the same importance in this situation that it had in the section above?

A numerical experiment was carried out using the same simulated aperture as above, with 15 panels and a 12 dB illumination taper. There were slight changes: $\lambda = 24.2$ mm and $\kappa = 0.801$. Three values $N_u = 31$, 41, and 61 were tried, but the extended span was $N_u' = 241$ in each case, allowing the same starting points to be used. An ensemble of 24 different aperture functions to use as starting points was concocted, some with uniform phase, some with random phase, some with illumination taper, some without, and some with an unusual illumination function. Two examples are illustrated in figure 6.7.

Twenty SP iterations were applied beginning from each of these points, but always using far field data for the same simulated aperture. The 20 different convergence paths resulted in an ensemble of end-points, and ideally all would have estimated the aperture with some success. The number of far field samples N_u was changed, and computer-generated Gaussian noise was sometimes added to them, mimicking a series of holography 'experiments'; each one produced a new ensemble of end-points. The scheme is outlined in figure 6.8. The rms phase difference rms Δ was used to measure the success of each convergence path in an ensemble, this time using linear units. This is a more practical measure than the Hilbert space distance $\|a_{20} - a\|$ because good phase recovery is the aim of Microwave Holography. It will be used

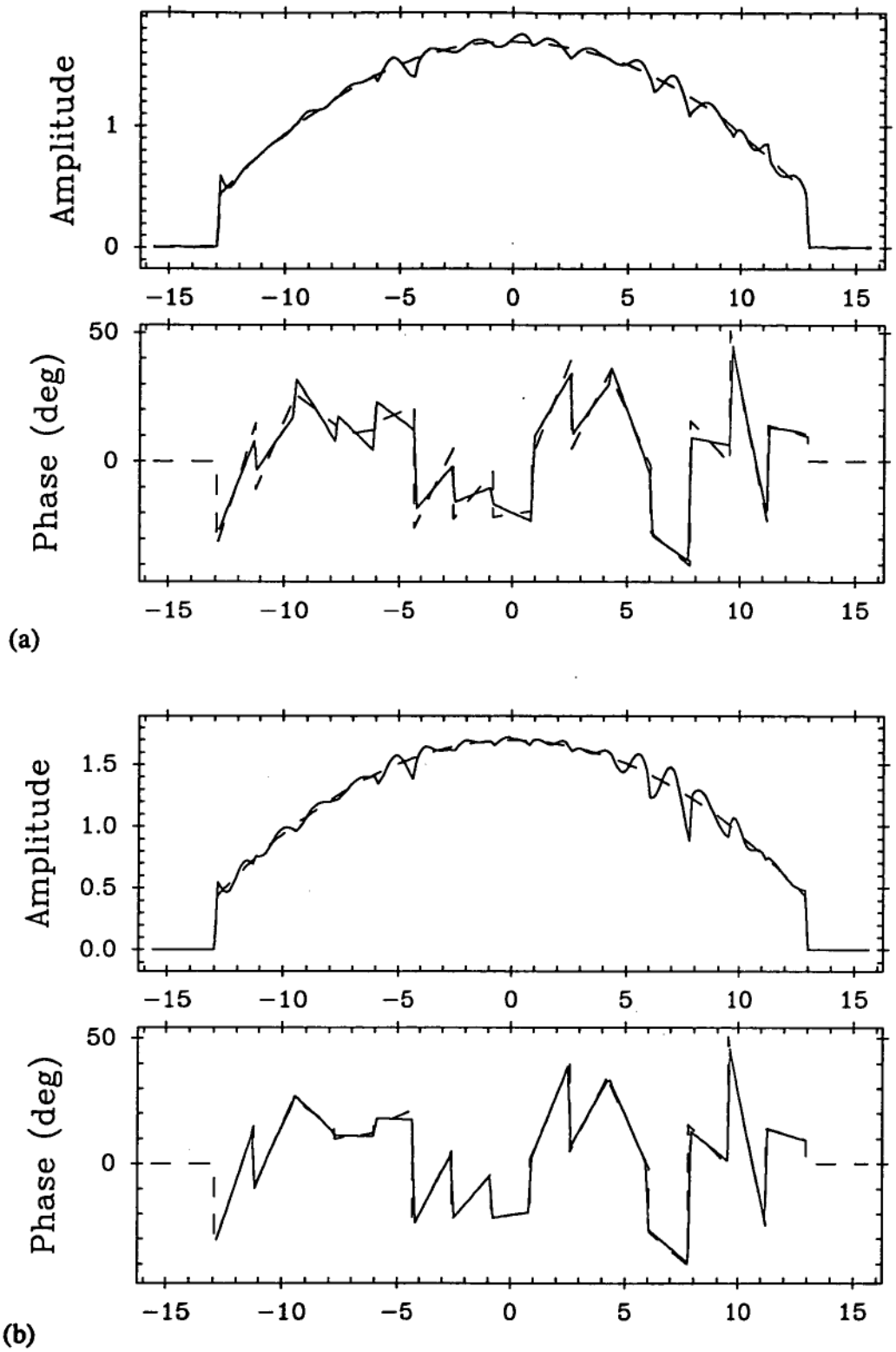
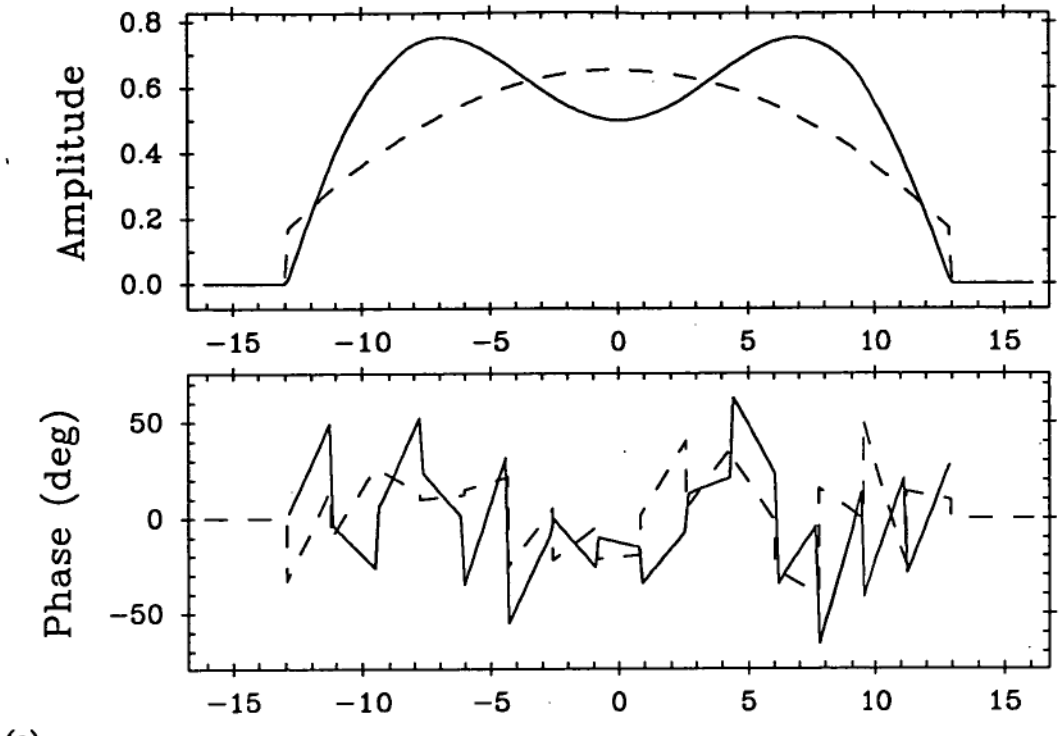
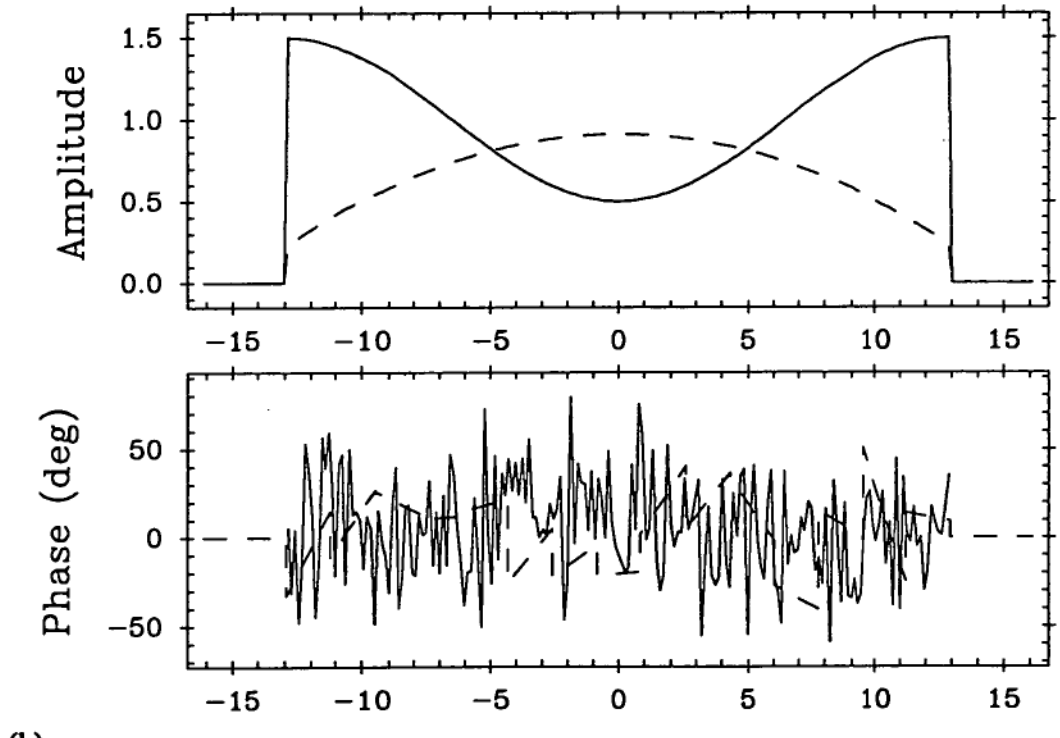


Figure 6.6: The aperture amplitude and phase after 100 SP iterations for (a) $N_u = 33$ and (b) $N_u = 35$. In the latter, nearly all panels have been recovered well. The horizontal scales are in metres.



(a)



(b)

Figure 6.7: Two of the 24 different aperture functions used as starting points for SP. The amplitude and phase of the original aperture is shown dashed. The horizontal scales are in metres.

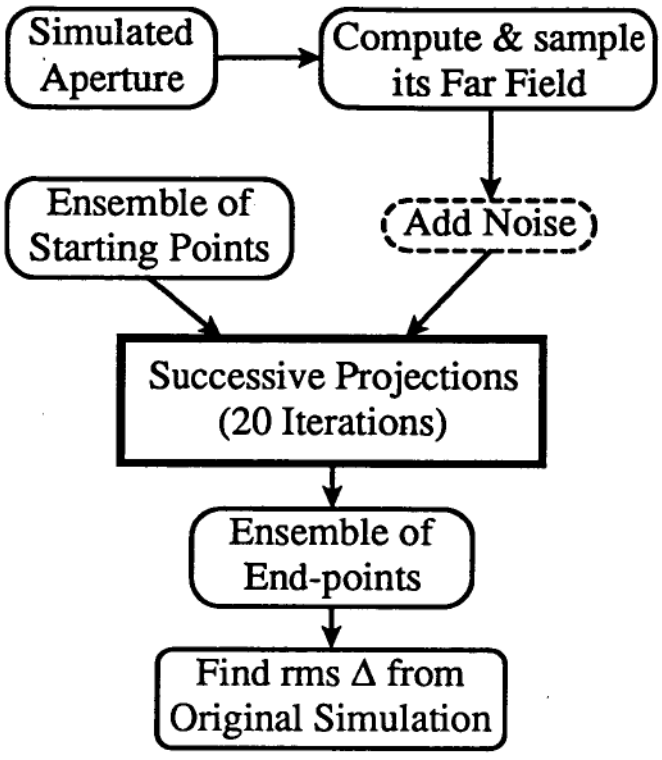


Figure 6.8: A numerical test of the effect of changing the starting point for SP. Various holography 'experiments' were simulated by changing the far field sampling and noise level, and their success was guaged by seeing how well the ensemble of end-points clustered about the original aperture function.

frequently in this chapter.

Figure 6.9 shows the effect on ensemble convergence of N_u. The ensemble range of rms Δ is shown by a box from the minimum to the maximum for $N_u = 31$, 41, and 61. Within the boxes the distribution was fairly uniform. A box of a different shade shows the range of rms Δ for the 24 starting points: from 19° to 46°. The main feature is that $N_u = 41$ gave a much smaller and more consistent value than $N_u = 31$, while taking 61 samples instead of 41 made only a slight improvement beyond this. This ties well with the observations above because $P_p = 1.66$, 2.19, and 3.26 respectively for these values of N_u, indicating that $P_p \simeq 2$ is the important turning point. To illustrate the result that would be obtained in each case without the benefit of SP, panels were fitted to the direct Fourier transform of the N_u samples and the rms Δ of the result shown as a dot on the same diagram. It is worth noting that this diagram and the following ones looked much the same for

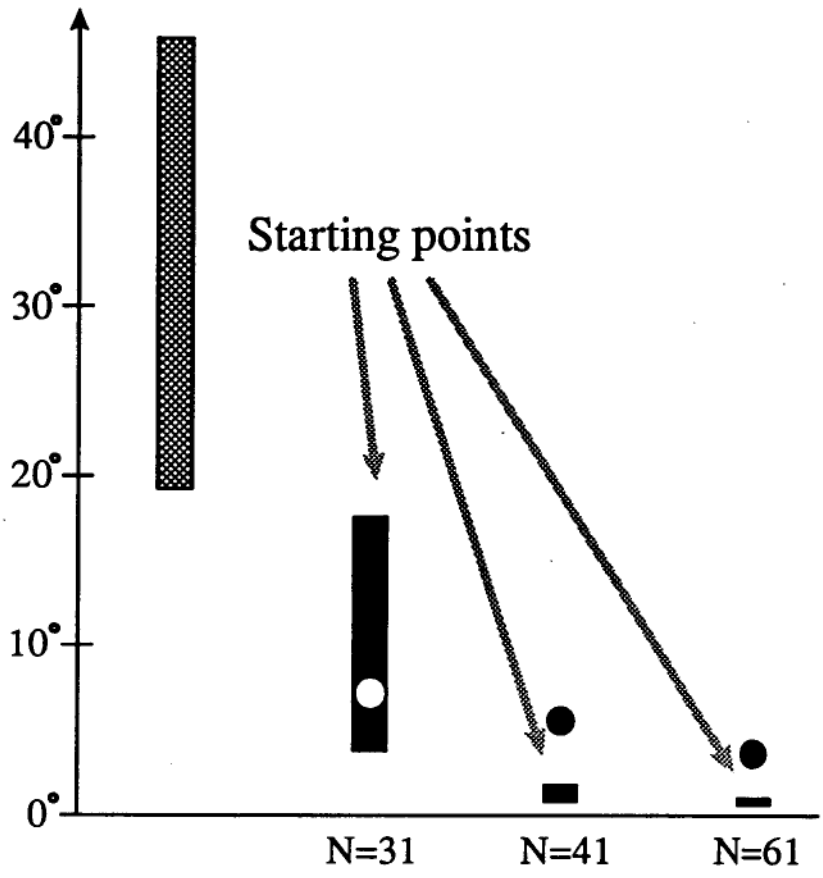


Figure 6.9: The effect of 20 SP iterations on an ensemble of 24 different starting points for a simulated one-dimensional aperture. Each box shows the range of resulting rms Δ obtained for a different N_u. The left-most box shows the range for the starting points. The dots show the rms Δ obtained without the use of Successive Projections.

10, 20, or 100 SP iterations.

Adding noise to the data at the 50 dB level gave a result entirely consistent with the results of section 4.2 for $\mathcal{R}_n = 50$ dB. Using the uniform illumination equation 4.22 to estimate the expected aperture phase error, it is found that $\sigma_\phi \simeq 0.36°$–$0.50°$ for this range of N_u. In figure 6.10 the three boxes are seen to have risen by roughly this amount, illustrating the assertion made in section 5.2.4 that noise power is preserved by SP.

So far the sampling interval δ_u has been kept constant such that $\kappa = \delta_u D_x/\lambda = 0.801$. One reason for having $\kappa < 1$ is that aliasing effects are reduced. Some iterative transform algorithms have performed better when κ is reduced (see section 6.2) so it is important to see if this is true for

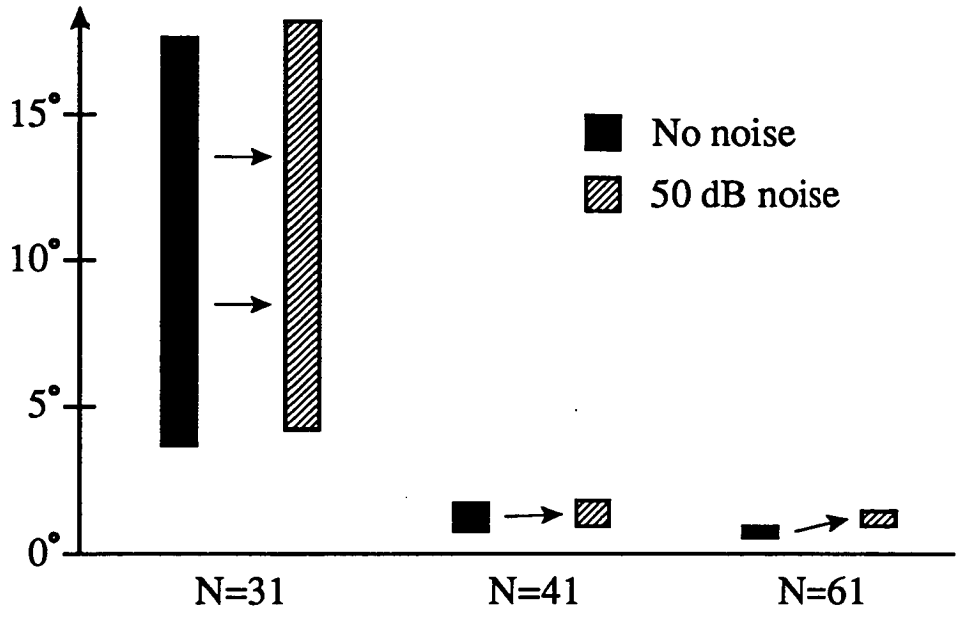


Figure 6.10: The dark boxes are as for the previous figure. The adjacent ones show the effect of adding Gaussian noise at the 50 dB level.

panel fitting with SP. In order to separate such an effect from changes in P_p the number of samples must be increased when κ is decreased, according to equation 6.1. Table 6.1 gives the sampling parameters used for all the simulated experiments discussed here. In figure 6.11 it is seen that reducing κ to 0.606 had little effect on the convergence of the ensemble for $P_p = 1.66$. However, reducing κ to 0.670 and then to 0.538 had a noticeably bad effect for $P_p = 2.19$! This is very interesting and suggests that it is best to sample at close to the Nyquist interval, as long as aliasing problems are not incurred. Since this means fewer samples are required for the same far field span W_u, measurement time will be reduced.

6.1.3 A Two-Dimensional Numerical Experiment

A *Digital VAXstation 3100* computer became available after the completion of the one-dimensional ensemble studies. This made it feasible to do a similar experiment with a full two-dimensional simulated aperture, using data grids of the size used for the Mount Pleasant experiments. The effect of noise and sampling interval were not investigated again, but the choice of a suitable far field map size was of great interest with the April 1991 experiments about to begin. The opportunity was also taken to apply a 'null field' blockage to

κ	N_u	P_p
0.801	31	1.66
0.606	41	1.66
0.801	41	2.19
0.670	49	2.19
0.538	61	2.19
0.801	61	3.26

Table 6.1: Sampling parameters used in the one-dimensional ensemble experiments.

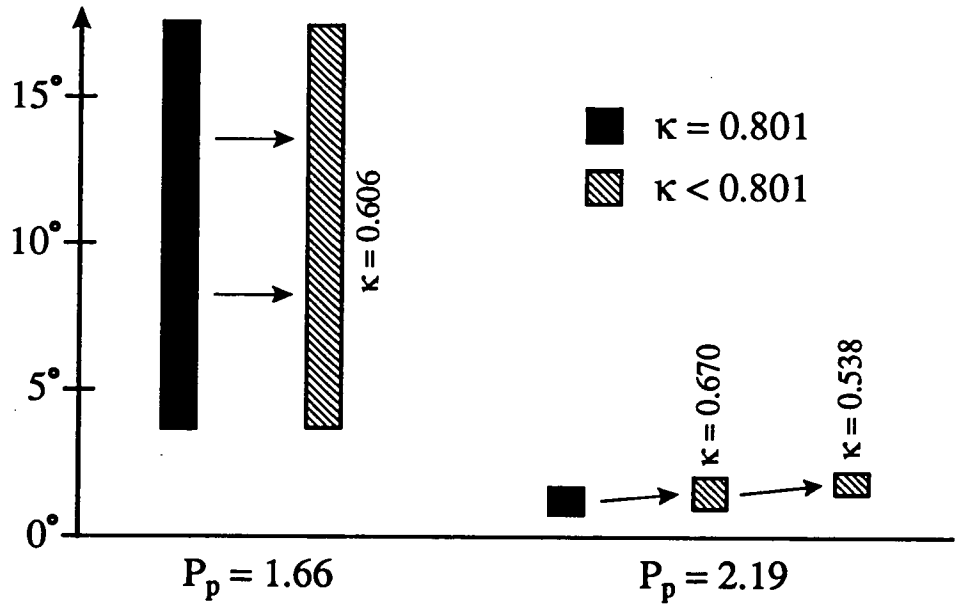


Figure 6.11: The dark boxes are as for the previous two figures, with $\kappa = 0.801$. The others show the effect of reducing κ to 0.606 for $N_u = 31$, and to 0.670 and then 0.538 for $N_u = 41$, maintaining P_p at a constant value.

the aperture and test its effect on Successive Projections.

An aperture with randomly oriented panels laid out as in figure 2.12 was given a 12 dB illumination taper, and its far field radiation pattern calculated by a discrete Fourier transform both with and without zeroing of the shadowed parts of the aperture. Various sizes of map were sampled from these patterns, each defining a different set S_1. Their sizes and corresponding numbers of aperture points per panel, for $\lambda = 24.2$ mm and $\kappa = 0.801$, were as follows:

$N_u \times N_v$	P_p (blocked)	P_p (clear)
41 × 41	3.05	3.37
45 × 45	3.60	3.98
49 × 49	4.27	4.75
53 × 53	5.10	5.62
57 × 57	5.91	6.48

The aperture blockage lowered the average number of useable points per panel, as can be seen. Three points are required to determine the tilts and vertical shift of a two-dimensional panel. Because the panels have a range of sizes the average P_p should be considerably larger than 3 to ensure that the smaller panels are sufficiently covered.

Eighteen arbitrary apertures were also simulated, similarly to the one-dimensional ones, to be starting points for SP iterations. They had an rms Δ from the first aperture of between 27° and 41°. With a grid extended to 241 × 241 points, each blocked and unblocked map was processed by 10 iterations of SP beginning at all 18 starting points. Thus a total of 1800 iterations were required, which was quite a significant processing and data management task. As in the one-dimensional case, the ensembles of end-points were compared with the original aperture using rms Δ. The range and average of rms Δ over the ensembles is shown in table 6.2, and the values for a blocked original aperture are very similar to those for a clear aperture. This confirms that Successive Projections is not adversely affected by a null field blockage when it is handled as discussed in section 5.3.1.

The same data are shown graphically in figure 6.12, except that the clear aperture cases have been omitted. For comparison, the rms Δ resulting when panels are fitted without using SP is shown for each map size. This graph can be compared with figure 6.9. The trend to lower rms Δ for larger maps is smoother than for the one-dimensional case because the panels have such a variety of sizes. The ensemble average is always near the bottom of the range:

Far Field	Blocked aperture			Clear aperture		
Map Size	Min.	Ave.	Max.	Min.	Ave.	Max.
41 × 41	6.0	7.1	11.4	6.0	7.6	11.6
45 × 45	4.2	5.2	8.9	4.4	5.6	8.9
49 × 49	3.2	4.0	6.2	3.1	4.0	6.1
53 × 53	2.6	3.1	4.6	2.3	2.9	4.0
57 × 57	2.2	2.6	3.8	1.8	2.2	2.9

Table 6.2: The range and average of rms Δ over each ensemble of SP endpoints for the two-dimensional experiment. The units are degrees of aperture phase.

some more erratic starting points tended to give an unusually high rms Δ. The benefit of using SP is clear. Reading across the graph, for example, a 45×45 map usuall gave a more accurate aperture phase estimate than did traditional analysis applied to a 57×57 grid. This is a 38% reduction in the number of far field samples, which would lead to an equivalent saving in observing time.

On the basis of this graph a map size of 53×53 was chosen for the April 1991 experiments. It was a safe compromise between measurement time and quality of recovered panel positions, although smaller maps would have been almost as effective. The observations in this section imply an optimal design for future Microwave Holography experiments. It is not advantageous to sample more closely than the Nyquist interval ($\kappa = 1$) except as required to eliminate aliasing errors, and furthermore Successive Projections permits a smaller extent of the radiation pattern to be measured to achieve a given accuracy of aperture recovery. This means the available observing time is better used by taking several holographic maps of minimum size instead of fewer, larger maps. Among the benefits are:

1. The effective signal-to-noise ratio, which dictates the achievable accuracy of estimated panel positions, increases in proportion to the number of maps;

2. The experiment is more immune to environmental or instrumental factors which may disrupt a single map;

3. Data are collected closer to the main beam, where the field is deter-

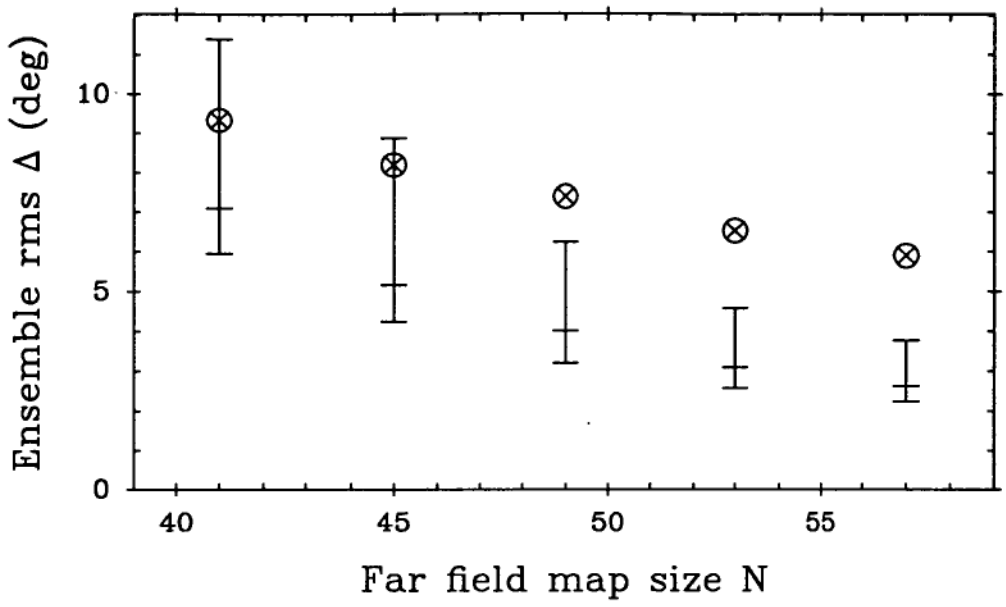


Figure 6.12: A graph of the ensemble average and range of rms Δ as a function of map dimension $N = N_u = N_v$: the average is the cross-piece in the centre of each bar. The rms Δ when panels are fitted without SP is shown as a cross in a circle.

mined more accurately by Fourier transform.

6.2 Analytical Solution of the Linear Problem

Empirical investigations tell something of the behaviour of the Successive Projections algorithm with different experimental parameters, but it would be better to have a more rigorous description. This section looks at the possibility of using well-known analytical techniques to solve a one-dimensional, linear model of a Microwave Holography experiment. This contrasts to the general problem which requires iterative solution, but it only applies when the distortions of the surface are small.

It is hoped that results from the linear model can be extended to the general problem. The justification is that the basis of the experiment is the same in both cases; only the numbers are changed. Thus, restrictions obtained on experimental parameters may be fundamental to the experiment, rather than introduced by the processing method. Several benefits arise from the linear

analysis. Since it is independent of the measured data, the feasibility of an experiment can be tested before it is performed. The results indicate which derived quantities are likely to be sensitive to measurement errors. The model can in principle be extended to a two-dimensional aperture.

6.2.1 Making a Linear Approximation

The aim is to calculate the illumination, offset, and tilt of each panel of the antenna directly from the radiated far field. To express the far field in terms of properties of individual panels, look at the field radiated by each panel. The total far field is the sum of these. Using the subscript p to denote a panel number in the range $1, \ldots, N_p$, write equation 2.13 in the form

$$A(u) = \sum_{p=1}^{N_p} A_p(u) \text{ where } A_p(u) = \int_{-\infty}^{+\infty} a_p(x) e^{-j2\pi ux}\, dx.$$

In order to calculate this Fourier transform analytically, assume that the illumination is constant over each panel. Let the panel be bounded by x_{p-1} and x_p, and let the aperture phase vary linearly between these edges from ϕ_p to ϕ'_p. The situation is shown in fiture 6.13. Then the following quantities can be defined:

$$\begin{aligned}
r_p &= |a_p(x)| \text{ for } x \in [x_{p-1}, x_p) \\
g_p &= \frac{x_p + x_{p-1}}{2} \text{ is the panel centre} \\
h_p &= x_p - x_{p-1} \text{ is the panel size} \\
\ell_p &= \frac{\phi'_p + \phi_p}{2} \text{ is due to a panel offset} \\
\beta_p &= \frac{\phi'_p - \phi_p}{2} \text{ is due to a panel tilt.}
\end{aligned}$$

The aperture current for each panel can now be written as

$$a_p(x) = r_p \,\mathrm{II}\left(\tfrac{x-g_p}{h_p}\right) \exp j\left(2\beta_p \tfrac{x-g_p}{h_p} + \ell_p\right). \tag{6.2}$$

The Fourier transform of $\mathrm{II}(x)$ is $\mathrm{sinc}(u)$. Using the similarity and shift theorems, the transform of the magnitude of $a_p(x)$ can be found easily. The phase gradient represented by β_p becomes a shift in the far field domain: as the panel is tilted its field follows. The phase offset ℓ_p remains unchanged under Fourier transform. These facts give

$$A_p(u) = r_p e^{+j\ell_p} e^{-j2\pi u g_p} h_p \mathrm{sinc}\left(h_p u - \tfrac{\beta_p}{\pi}\right). \tag{6.3}$$

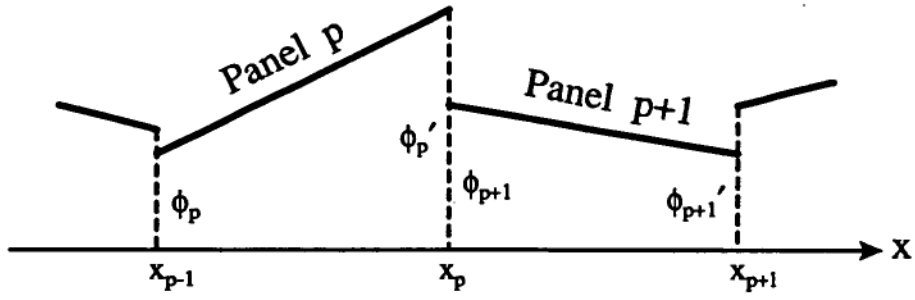


Figure 6.13: The panel geometry of a one-dimensional aperture.

To obtain an equation linear in ℓ_p, assume that it is small enough to write

$$e^{+j\ell_p} = \cos\ell_p + j\sin\ell_p \simeq 1 + j\ell_p.$$

Also assume that β_p is small enough to allow the sinc function to be approximated by its first order Taylor expansion about $h_p u$, giving

$$h_p \mathrm{sinc}\left(h_p u - \tfrac{\beta_p}{\pi}\right) \simeq h_p \mathrm{sinc}\,(h_p u) + \frac{\beta_p}{\pi u}\left[\mathrm{sinc}\,(h_p u) - \cos(\pi h_p u)\right]. \qquad (6.4)$$

In the special case of $u = 0$ this expression must be set simply to h_p; a quadratic term would be needed to improve on this.

Equation 6.3 can now be multiplied out and the real and imaginary parts separated. A final assumption, that the cross-term $\ell_p\beta_p$ can be neglected, gives the following expressions for the field due to a single panel:

$$\begin{aligned}
\mathrm{Re}\,A_p(u) &\simeq h_p\mathrm{sinc}\,(h_p u)\cos(2\pi u g_p)\cdot r_p \\
&+ h_p\mathrm{sinc}\,(h_p u)\sin(2\pi u g_p)\cdot r_p\ell_p \\
&+ \tfrac{1}{\pi u}\left[\mathrm{sinc}\,(h_p u) - \cos(\pi h_p u)\right]\cos(2\pi u g_p)\cdot r_p\beta_p \qquad (6.5) \\
\mathrm{Im}\,A_p(u) &\simeq -h_p\mathrm{sinc}\,(h_p u)\sin(2\pi u g_p)\cdot r_p \\
&+ h_p\mathrm{sinc}\,(h_p u)\cos(2\pi u g_p)\cdot r_p\ell_p \\
&+ -\tfrac{1}{\pi u}\left[\mathrm{sinc}\,(h_p u) - \cos(\pi h_p u)\right]\sin(2\pi u g_p)\cdot r_p\beta_p. \qquad (6.6)
\end{aligned}$$

These equations are linear in the quantities $(r_p, r_p\ell_p, r_p\beta_p)$, which have become the unknowns to be calculated for each p. The components of the total far field are linear in all of the $3N_p$ unknowns:

$$\mathrm{Re}\,A(u) \simeq \sum_{p=1}^{N_p}\mathrm{Re}\,A_p(u) \text{ and } \mathrm{Im}\,A(u) \simeq \sum_{p=1}^{N_p}\mathrm{Im}\,A_p(u).$$

Experimental measurements give $\mathrm{Re}\,A(u)$ and $\mathrm{Im}\,A(u)$ for N_u values $u = u_1, \ldots, u_{N_u}$, resulting in an over-determined set of equations whose solution

is a general linear least-squares problem. A design matrix can be constructed by placing the unknowns and the data into vectors. For brevity, denote the six coefficients in equations 6.5 and 6.6 by $a_p(u), b_p(u), \ldots, f_p(u)$. Then the problem is summarized by this matrix equation:

$$\begin{pmatrix} a_1(u_1) & b_1(u_1) & c_1(u_1) & a_2(u_1) & \cdots \\ d_1(u_1) & e_1(u_1) & f_1(u_1) & d_2(u_1) & \\ a_1(u_2) & b_1(u_2) & c_1(u_2) & a_2(u_2) & \\ d_1(u_2) & e_1(u_2) & f_1(u_2) & d_2(u_2) & \\ a_1(u_3) & b_1(u_3) & c_1(u_3) & a_2(u_3) & \cdots \\ \vdots & & & \vdots & \end{pmatrix} \cdot \begin{pmatrix} r_1 \\ r_1\ell_1 \\ r_1\beta_1 \\ r_2 \\ r_2\ell_2 \\ \vdots \end{pmatrix} \simeq \begin{pmatrix} \mathrm{Re}\, A(u_1) \\ \mathrm{Im}\, A(u_1) \\ \mathrm{Re}\, A(u_2) \\ \mathrm{Im}\, A(u_2) \\ \mathrm{Re}\, A(u_3) \\ \vdots \end{pmatrix}. \tag{6.7}$$

This can be written symbolically as

$$\begin{array}{ccccc} \mathbf{A} & \cdot & \mathbf{x} & \simeq & \mathbf{b} \\ 2N \times 3N_p & & 3N_p \times 1 & & 2N \times 1 \end{array}$$

where clearly $2N > 3N_p$ for an over-determined problem.

The constraints imposed by these assumptions would appear to be quite rigid, but a visual check shows that the approximation is good for a reasonable range of values. An aperture with 6 randomly displaced panels was simulated, the maximum ℓ_p being 0.48 rad and the maximum β_p 0.21 rad. Its far field was estimated in three ways:

(a) summing of the panel fields given by equation 6.3, with the illumination on each panel replaced by its average amplitude;

(b) multiplying the design matrix by the parameter vector shown in equation 6.7;

(c) performing a discrete Fourier transform of the currents evaluated on a closely sampled set of aperture points.

The results are illustrated in figure 6.14: the 'diameter' was 26 m, the wavelength 24.2 mm, and the edge taper −12 dB. Over the first few sidelobes, at least, the field is reproduced quite well by the linear model. Most of the phase discrepancy is due to the averaging of each panel's illumination, as expected for only 6 panels. The matrix problem was solved using Fourier transform result for a right-hand vector, giving panel parameters r_p, ℓ_p, and β_p within a few percent of their correct values.

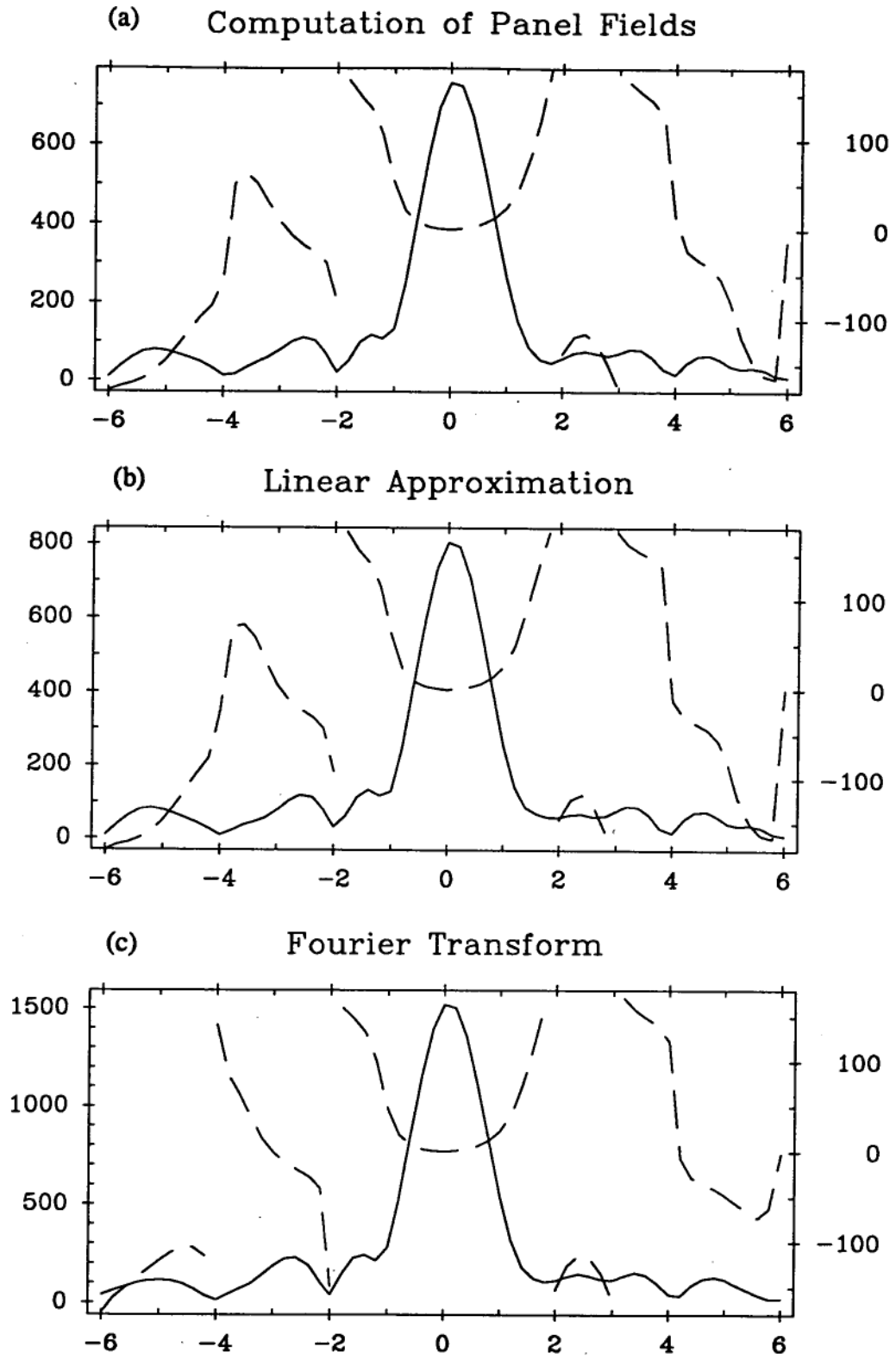


Figure 6.14: Three ways of computing the field of an aperture with 6 panels and a −12 dB taper; (a), (b), and (c) are described in the text. The amplitude (solid line, left-hand scale) and the phase (dashed line, right-hand scale in degrees) are shown. The horizontal scales are in units of $\lambda/D \simeq 0.053°$.

This shows that the computations were correct. Remember, however, that it is unimportant whether or not the linear model fails for most realistic situations. The question is whether the experimental design allows the problem to be solved at all.

6.2.2 Behaviour of the Design Matrix

In the words of Press, et al. [61, sect. 2.9], Singular Value Decomposition (SVD) is the method of choice for solving linear least squares problems. From the present point of view, the great advantage of SVD is its ability to diagnose where problems will arise when the design matrix is ill-conditioned in some way. It gives a detailed account of the effect on the solution in these cases. It cannot fail: the decomposition can always be done, no matter how singular the matrix.

Using SVD, any matrix $\mathbf{A}$, having at least as many rows as columns, can be expressed as a product of three matrices:

$$\mathbf{A} = \mathbf{U} \cdot \mathbf{W} \cdot \mathbf{V}^{\mathrm{T}}. \tag{6.8}$$

Here $\mathbf{U}$ is a column-orthonormal matrix with the same shape as the original matrix, $\mathbf{W}$ is a diagonal matrix, and $\mathbf{V}$ is a square orthonormal matrix. Both of the latter have the same number of columns as $\mathbf{A}$, so there is one diagonal element w_i for each unknown in the equations. These are the *singular values* of the matrix. Because the inverse of $\mathbf{A}$ is constructed by the product

$$\mathbf{A}^{-1} = \mathbf{V} \cdot \mathrm{diag}\,(1/w_i) \cdot \mathbf{U}^{\mathrm{T}},$$

the solution will run into difficulties if some of the w_i are zero or very small compared to the others. Each small w_i represents a degree of freedom in the unknown vector which will not be recovered properly.

The possibility of finding a solution to the linear panel fitting problem is of more interest than the solution itself. The singular values of the design matrix show whether the problem can be solved, and how robust the solution is likely to be to measurement errors. Observing the singular values as experimental parameters are changed gives insight into the information content of data that would result from these experiments. If the design matrix behaves badly, there is something wrong with the experiment which is independent of the data collected.

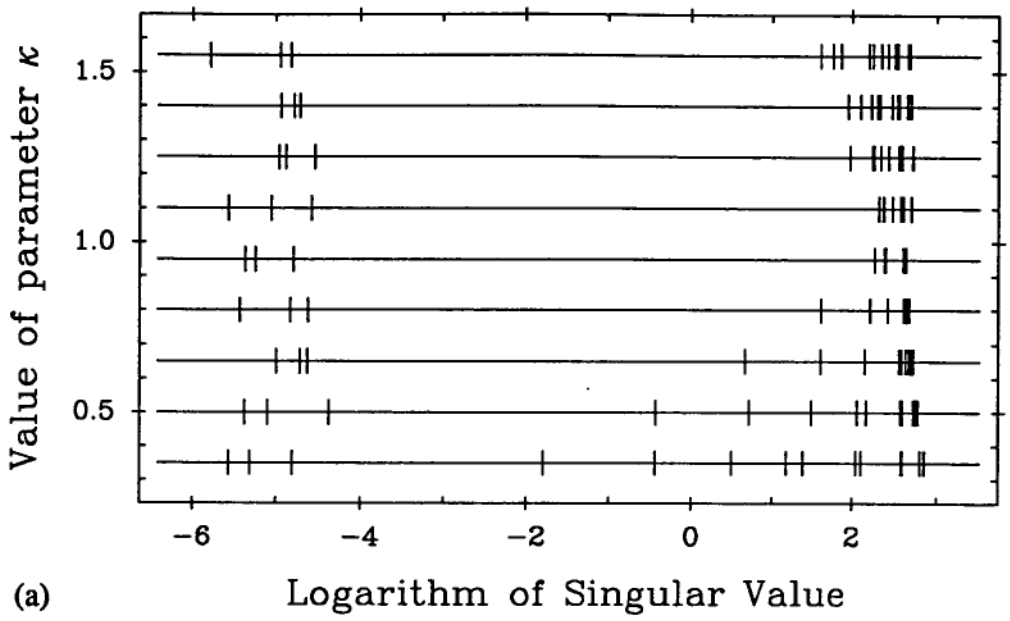
The parameters of interest are the sampling factor κ and the number of samples N_u. Related to these is the average number of independent sample

points on each panel in the aperture domain, given by equation 6.1. The problem is specified completely by the sampling parameters, which determine a set of u_i values, and the panel layout, which provides the remaining terms of equations 6.5 and 6.6. A design matrix can then be formed.

A panel layout with 6 equally-sized panels, $h_1 = h_2 = \ldots = h_6$, was investigated using SVD. There were 18 unknowns, or columns of matrix **A**. In figure 6.15 the 18 singular values of **A** are plotted on a logarithmic scale for several different values of κ. In (a) $N_u = 9$, and there are just as many equations as unknowns, $2N = 3N_p$. Clearly this is not enough: for all κ there are three exceedingly small singular values, so the equations are not independent even for Nyquist sampling. This may be because the coefficients in equations 6.5 and 6.6 have so many related terms. In (b) $N_u = 15$, and it can be seen that for κ around or just below unity the w_i are all similar, indicating a robust solution, while elsewhere there can be a range over several orders of magnitude. This is expected from Fourier sampling theory: $\kappa = 1$ gives Nyquist sampling. If it is increased the edges of the aperture are dominated by aliasing, while if it is reduced the samples will lose their independence, causing degenerate equations. Note that $P_p = 2$ for $\kappa = 0.8$, and it is below this value that trouble occurs.

The columns of the matrix **V** corresponding to singular values of zero form an orthonormal basis for the null space of **A**. These are the linear combinations of unknowns to which the data will be completely insensitive. Less extremely, columns of **V** corresponding to unusually small singular values are combinations that will be ill-determined by the data. It is instructive to look at columns for some of the singular values in figure 6.15(b). Table 6.3 shows the columns for the smallest w_i in several experiments; recall the order of terms in vector **x** of equation 6.7. Firstly, $N = 15, \kappa = 0.5$ gives a linear combination of primarily $r_p\beta_p$ terms, showing that panel tilts are the most difficult quantities to recover when the going gets tough. Secondly, $N = 15, \kappa = 1.55$ gives a mixture of $r_p\ell_p$ and $r_p\beta_p$ terms, but notice that they are concentrated at the outer panels. Undersampling causes precisely these quantities to be contaminated by aliasing with the neighbouring image.

A different way of displaying the design matrix behaviour is to fix the average number of points per panel, P_p, which is expected to have a large bearing on the success of tilt recovery. Figure 6.16 gives singular value plots for (a) $P_p = 1.5$ and (b) $P_p = 2$. The extremely low values for large κ are due to the small number of equations: the top line in each diagram corresponds



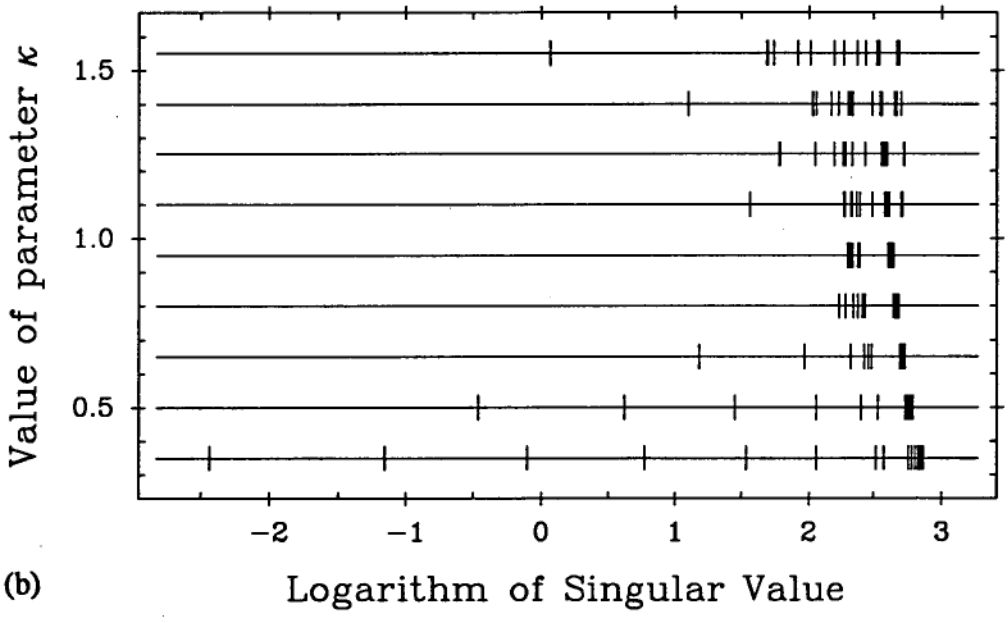


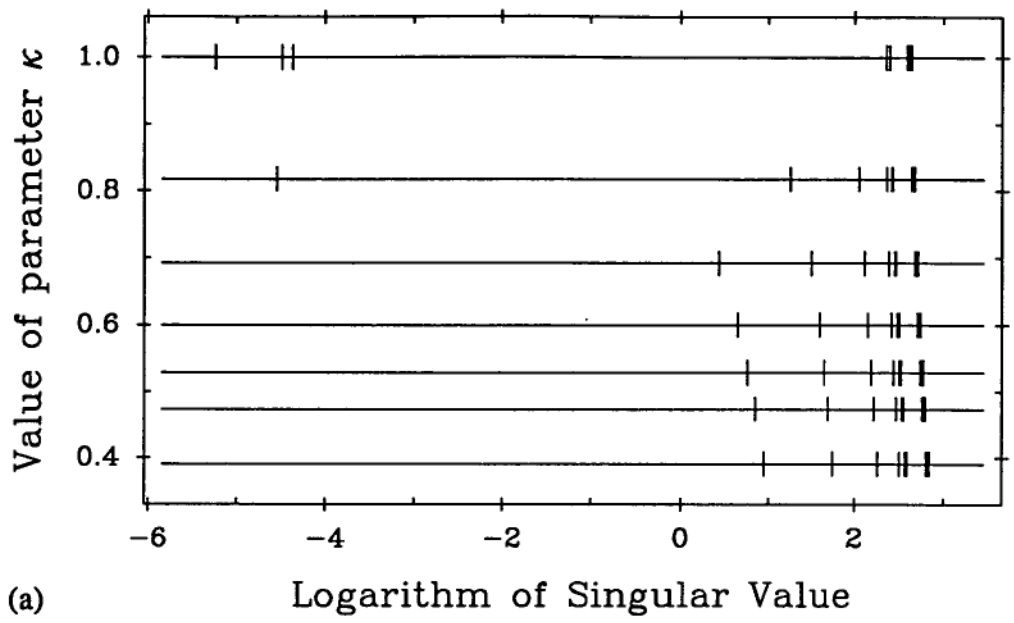
Figure 6.15: Singular values of the design matrix **A** for a several hypothetical one-dimensional Microwave Holography experiments on a 6-panel reflector. Each horizontal line represents a different experiment, and holds 18 singular values on a logarithmic scale. The sampling ratio κ is varied while $N_u = 9$ for (a) and 15 for (b).

Unknown	$N = 15$ $\kappa = 0.5$	$N = 15$ $\kappa = 1.55$	$P_p = 1.5$ $\kappa = 0.692$	$N = 31$ $\kappa = 0.8$
r_1	0.000	0.000	0.000	
$r_1\ell_1$	0.021	−0.305	−0.032	−0.408
$r_1\beta_1$	0.049	−0.352	−0.101	0.000
r_2	0.000	0.000	0.000	
$r_2\ell_2$	0.060	−0.398	−0.049	−0.408
$r_2\beta_2$	0.300	0.353	−0.360	0.000
r_3	0.000	0.000	0.000	
$r_3\ell_3$	0.039	0.001	−0.024	−0.408
$r_3\beta_3$	0.634	−0.001	−0.597	0.000
r_4	0.000	0.000	0.000	
$r_4\ell_4$	−0.039	−0.001	0.024	−0.408
$r_4\beta_4$	0.634	−0.001	−0.597	0.000
r_5	0.000	0.000	0.000	
$r_5\ell_5$	−0.060	0.398	0.049	−0.408
$r_5\beta_5$	0.300	0.353	−0.360	0.000
r_6	0.000	0.000	0.000	
$r_6\ell_6$	−0.021	0.305	0.032	−0.408
$r_6\beta_6$	0.049	−0.352	−0.101	0.000

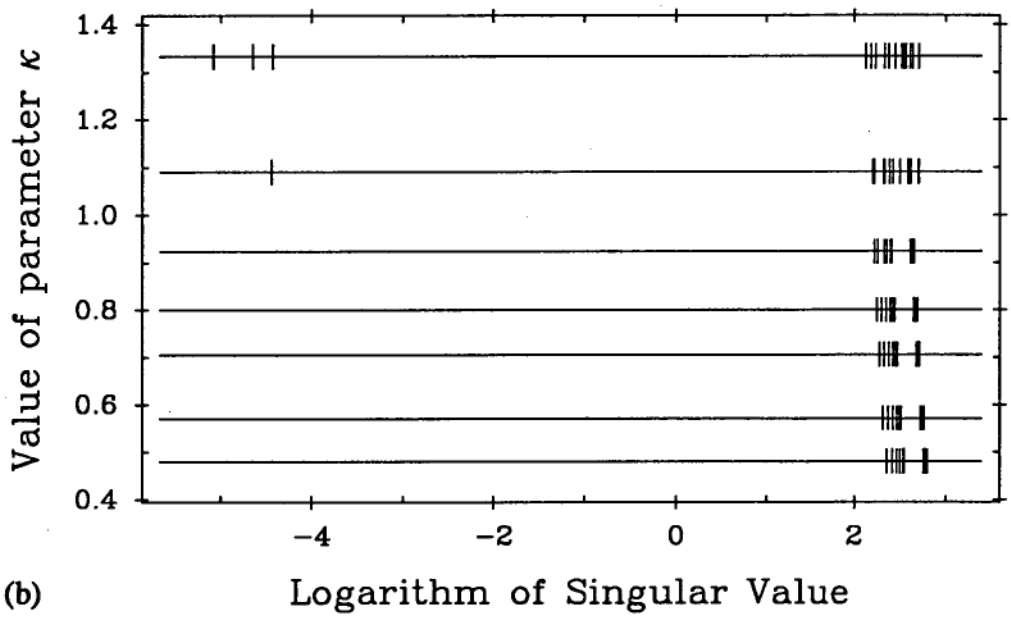
Table 6.3: Columns of matrix $\mathbf{V}$ corresponding to the smallest singular value of the design matrix of selected experiments.

to $N = 9$. In other regions the w_i are very well behaved for $P_p = 2$ but trickle downwards for $P_p = 1.5$. The third column in table 6.3 is for the lowest w_i when $P_p = 1.5, \kappa = 0.692$. It is another combination of $r_p\beta_p$ terms, and since r_p is always very well recovered it is knowledge of panel tilts which will suffer. For $P_p = 2$ there is no evidence that decreasing κ is beneficial, which agrees with the findings of section 6.1.

The SVD analysis agrees with intuition and experimental experience, but does it show what conditions *must* be fulfilled by an experiment? This is a strong statement; it means that other techniques for determining aperture quantities from measured data, such as Successive Projections, will be subject to the same restrictions. If so, SVD can be used as a tool to select optimal parameters for the experiment. To make this approach more convincing it is



(a)



(b)

Figure 6.16: Singular values as in the previous figure but for fixed P_p, so that N_u varies with κ. In (a) $P_p = 1.5$ and in (b) $P_p = 2$.

worthwhile to consider a different problem.

6.2.3 The Phase Retrieval Problem

Phase retrieval holography is a successful surface measurement technique which is particularly useful at high frequencies when direct phase recording is difficult. Typically, two amplitude maps of the radiation pattern, one or both with a defocus applied, are processed using an iterative Fourier transform algorithm as discussed in section 5.1.3. It has recently been shown by McCormack, et al. [43] that the complex aperture current function can be recovered with some success from a single map of the power pattern of the antenna. If the power can be expressed linearly in terms of the aperture quantities (r_p, ℓ_p, β_p), then SVD can be applied to this problem as well.

Taking the earlier equations 6.5 and 6.6 as a starting point, rewrite them with the abbreviated coefficients used for the design matrix:

$$\begin{aligned} \mathrm{Re}\, A_p(u) &\simeq a_p(u) r_p + b_p(u) r_p \ell_p + c_p(u) r_p \beta_p \\ \mathrm{Im}\, A_p(u) &\simeq d_p(u) r_p + e_p(u) r_p \ell_p + f_p(u) r_p \beta_p. \end{aligned}$$

Summing over p to produce the total field maintains linearity, but squaring the sum to find the power does not. To make any progress the additional assumption of uniform illumination is needed, so set $r_p = 1$ for all p. Further cross terms $\ell_p \beta_q$ must be neglected in addition to those for $p = q$ neglected previously, but as before the analysis is useful despite the constraint placed on panel displacements. Dropping the u dependence for clarity,

$$\begin{aligned} \mathrm{Re}^2 A &= \left(\sum_{p=1}^{N_p} \mathrm{Re}\, A_p \right)^2 \\ &\simeq \sum_{p=1}^{N_p} (a_p + b_p \ell_p + c_p \beta_p) \sum_{q=1}^{N_p} (a_q + b_q \ell_q + c_q \beta_q) \\ &\simeq \sum_{p=1}^{N_p} \sum_{q=1}^{N_p} (a_p a_q + a_p b_q \ell_q + a_p c_q \beta_q + b_p \ell_p a_q + c_p \beta_p a_q) \\ &= \sum_{p=1}^{N_p} \sum_{q=1}^{N_p} (a_p a_q + 2 a_q b_p \ell_p + 2 a_q c_p \beta_p). \end{aligned}$$

Similarly for the imaginary part, so that after rearrangement

$$\mathrm{Re}^2 A - \sum_{p,q=1}^{N_p} a_p a_q \simeq 2 \sum_{p=1}^{N_p} \left[\ell_p \cdot b_p \sum_{q=1}^{N_p} a_q + \beta_p \cdot c_p \sum_{q=1}^{N_p} a_q \right]$$

$$\mathrm{Im}^2 A - \sum_{p,q=1}^{N_p} d_p d_q \;\simeq\; 2\sum_{p=1}^{N_p}\left[\ell_p \cdot e_p \sum_{q=1}^{N_p} d_q + \beta_p \cdot f_p \sum_{q=1}^{N_p} d_q\right].$$

Adding these,

$$|A|^2 - \sum_{p,q=1}^{N_p}(a_p a_q + d_p d_q) \simeq$$
$$2\sum_{p=1}^{N_p}\left[\ell_p\left(b_p\sum_{q=1}^{N_p} a_q + e_p\sum_{q=1}^{N_p} d_q\right) + \beta_p\left(c_p\sum_{q=1}^{N_p} a_q + f_p\sum_{q=1}^{N_p} d_q\right)\right]. \qquad (6.9)$$

The quantity on the right hand side is a linear function of ℓ_p and β_p, and all coefficients can be calculated from the previous design matrix **A**. Numerical tests show that equation 6.9 gives the same result for $|A(u)|^2$ as the matrix multiplication 6.7, as long as ℓ_p and β_p are sufficiently small — they must be much smaller than the previous model required.

The solvability of the problem is determined solely by the coefficients of ℓ_p and β_p, which form a second design matrix. This is much smaller than **A**, having $2N_p$ columns or unknowns, and N_u rows because each far field sample now consists of a single quantity $|A(u)|^2$. A singular value decomposition of this matrix, using the same example aperture as before, shows that it is extraordinarily ill-conditioned. Singular values are plotted with $P_p = 3.5$ and various κ in figure 6.17(a). Despite the large P_p, exactly half of the singular values are vanishingly small in most cases ($\kappa = 1$ is unusual), so it seems that half the information has been lost. This situation can be compared with discarding the imaginary part of each complex data point, in other words removing every second row of **A**. If this is done, the singular values of the reduced matrix behave in just the same way.

One thing has been forgotten, however. All phase retrieval experiments make use of measurements taken either with a Fresnel zone source or with an axial defocus applied to the antenna. In the first case a spherical wavefront will meet the aperture instead of a planar one, while in the second case only a curved wavefront will arrive coherently at the focus. Either way the result is approximated by a quadratic phase variation across the aperture. McCormack, et al. found that this variation was necessary to remove the 'twin-image' ambiguity of the aperture function, which otherwise causes retrieval algorithms to confuse a function with its conjugate reflection. Another justification is that the dynamic range needed to record the map is reduced, giving less stringent noise requirements. Indeed, the very low singular values

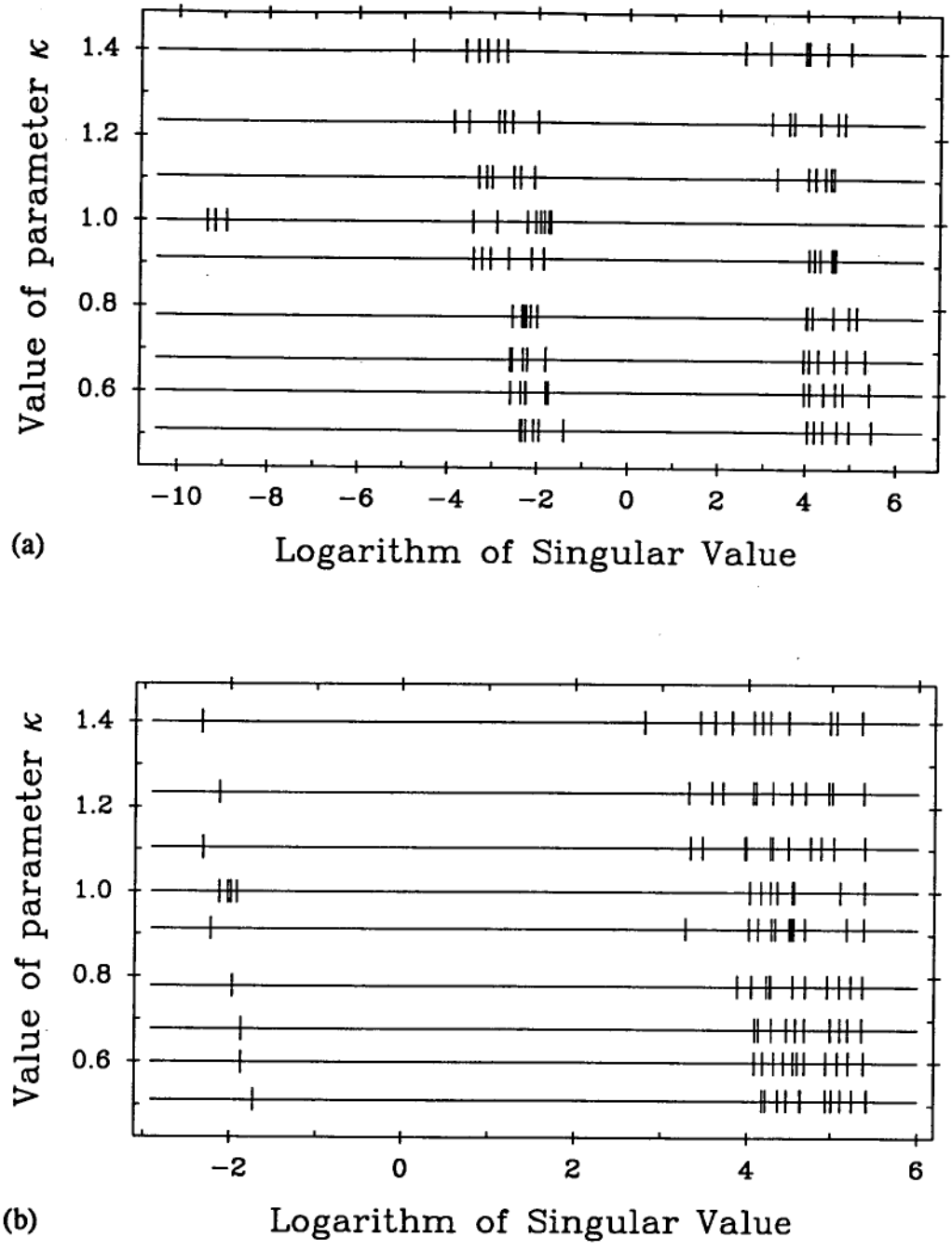


Figure 6.17: Singular values for a series of phase-less experiments with $P_p = 3.5$ and various κ. In (b) there is a quadratic phase term in the aperture reaching 180° at the edge, while in (a) there is none.

really indicate the extreme sensitivity to noise of the corresponding combinations of panel parameters.

A quadratic term can be incorporated into the analysis above by modifying equation 6.2 as follows:

$$a_p(x) \longrightarrow a_p(x)e^{j\gamma x^2}.$$

Analogously to the approximation that the illumination is constant over each panel, the same must be assumed for the quadratic phase term if an analytic solution is to be found. The new equations for the field due to a single panel are

$$\begin{aligned}
\mathrm{Re}\, A_p(u) &\simeq h_p \mathrm{sinc}\,(h_p u)\cos(\gamma g_p^2 - 2\pi u g_p)\cdot r_p \\
&+ -h_p \mathrm{sinc}\,(h_p u)\sin(\gamma g_p^2 - 2\pi u g_p)\cdot r_p \ell_p \\
&+ \tfrac{1}{\pi u}\left[\mathrm{sinc}\,(h_p u) - \cos(\pi h_p u)\right]\cos(\gamma g_p^2 - 2\pi u g_p)\cdot r_p \beta_p \\
\mathrm{Im}\, A_p(u) &\simeq h_p \mathrm{sinc}\,(h_p u)\sin(\gamma g_p^2 - 2\pi u g_p)\cdot r_p \\
&+ h_p \mathrm{sinc}\,(h_p u)\cos(\gamma g_p^2 - 2\pi u g_p)\cdot r_p \ell_p \\
&+ \tfrac{1}{\pi u}\left[\mathrm{sinc}\,(h_p u) - \cos(\pi h_p u)\right]\sin(\gamma g_p^2 - 2\pi u g_p)\cdot r_p \beta_p.
\end{aligned}$$

When these changes are made to the coefficients of $\mathbf{A}$, they automatically carry to the phase retrieval matrix because one is derived from the other. With a quadratic phase term γ chosen to give a 180° change at the edge of the dish, SVD analysis of the phase retrieval matrix gave the singular values plotted in figure 6.17(b). Except for γ this plot was otherwise generated just as (a), but it has only one tiny singular value for each experiment. Again, $\kappa = 1$ is unusual. The corresponding column of $\mathbf{V}$ was always as shown in the final column of table 6.3. It represents a uniform phase shift across the aperture that is unknowable: of course this is of no concern at all, and the problem can be effectively solved.

This result should be compared with the findings of other workers in phase retrieval. Taylor [81] reviews the theory behind one-dimensional problems of the kind modelled here. He concludes that the existence of a unique solution for the phase is not assured, and some prior knowledge of the situation must be provided. He summarized a variety of phase retrieval algorithms, which were successful or not depending on the particular function and initial estimate used. In the present case a uniform illumination had to be assumed to make the problem tractable, and the aperture function was known to be localized to the physical aperture. These facts constituted prior knowledge of the solution, and were probably responsible for the success of the analysis.

Of course the real experiment being modelled has two dimensions, not one, and this is a fundamental change. Bates [4] concludes that there is an effectively unique solution to the two-dimensional phase retrieval problem when the image is 'localized' in the sense which he defines. It is unique apart from three 'trivial characteristics' of the image, here the aperture function: a uniform phase shift, an origin shift, or a conjugate reflection. As mentioned already, McCormack, et al. [43, sect. 5] did consider conjugate reflection important and showed that a quadratic phase shift helped to avoid the ambiguity. They, too, concluded that the two-dimensional problem could be solved except for contrived, non-physical cases while the one-dimensional problem, in general, could not. Both authors made use of the theoretical work of Bruck and Sodin [10], who showed that the prime-ness of multi-dimensional polynomials lies behind this difference.

Even with this assurance of solvability the necessity of using every bit of a priori information is continually stressed, particularly when the data are contaminated by measurement noise. McCormack, et al. use an 'aperture domain management' scheme which incorporates an estimate of the aperture illumination function that can co-exist with noise outside the aperture support. Another way of providing additional information is to reduce the data sampling interval to less than the Nyquist interval, that is $\kappa < 1$. Bates explains that this information is not redundant because the square of the amplitude is the Fourier transform of the auto-correlation of the image, which has twice the extent of the image itself and so requires $\kappa = 0.5$ for its adequate representation. Supporting this, users of many iterative transform algorithms for phase retrieval have found that reducing the sampling interval can improve performance greatly. Thus it may be significant that the singular values in figure 6.17(b) are noticeably closer together for low κ, yet misbehave for $\kappa = 1$. No such effect occurs in figure 6.16(b), agreeing with the finding in section 6.1 that reducing κ is of no advantage for the full-phase Successive Projections algorithm.

Analyses of restrictive, linear models of a standard holography experiment and a phase-less experiment have shown behaviour with great physical significance. The restrictions have not changed the possibility or otherwise of obtaining a good result with a given set of experimental parameters. The models are thus useful for studying properties of experiments more rigorously than is possible by empirical methods.

6.3 An Alternative Formulation

The speed of the Successive Projections algorithm for panel fitting, as it has been described so far, is limited by the two large discrete Fourier transforms which are performed in every iteration. In the previous section, equation 6.3 gave the one-dimensional far field radiation from a single panel in an explicit form. The radiation from an aperture composed of such panels can be computed rapidly from the panel parameters — illumination, tilt, and vertical shift — without using a DFT.

This opens the possibility of a new SP formulation of the panel fitting problem, similar to that used by Poulton [59] for antenna power pattern synthesis. Each measured far field value $A_m(u_i)$ determines a set of panel parameters that give a far field agreeing at this point. It will be seen that equation 6.3, with a Taylor approximation, is of a form that allows the projection onto this set to be calculated. There is one set for each u_i, giving an algorithm with many sets, hence many projections in each iteration. Of interest is

- the effectiveness of this algorithm,
- its speed compared to 'old' SP, and
- its extension to two dimensions.

These are investigated in the following sections.

6.3.1 One Dimension

An essential part of Poulton's method was to form an expression which was linear for a set of complex unknowns. This can be done for the field due to a set of panels

$$A(u) = \sum_{p=1}^{N_p} A_p(u)$$

if the first order Taylor expansion 6.4 is used in equation 6.3. The result is

$$\begin{aligned} A_p(u) &= r_p e^{+j\ell_p} e^{-j2\pi u g_p} h_p \mathrm{sinc}\left(h_p u - \tfrac{\beta_p}{\pi}\right) \\ &\simeq r_p e^{+j\ell_p} e^{-j2\pi u g_p} \left[h_p \mathrm{sinc}\,(h_p u) + \frac{\beta_p}{\pi u}(\mathrm{sinc}\,(h_p u) - \cos(\pi h_p u))\right] \\ &= a_p F_p(u) + b_p G_p(u), \end{aligned} \tag{6.10}$$

where

$$\begin{aligned} a_p &= r_p e^{+j\ell_p} \\ b_p &= r_p \beta_p e^{+j\ell_p} \\ F_p(u) &= h_p \mathrm{sinc}\,(h_p u) e^{-j2\pi u g_p} \\ G_p(u) &= \frac{1}{\pi u}\left[\mathrm{sinc}\,(h_p u) - \cos(\pi h_p u)\right] e^{-j2\pi u g_p}. \end{aligned}$$

This is equivalent to equations 6.5 and 6.6 except that no small angle approximation has been made for ℓ_p. If the complex numbers a_p and b_p can be found (the same symbols have a different meaning in section 6.2) then the panel parameters r_p, ℓ_p, and β_p can be derived from them.

The entire field is

$$A(u) = \sum_{p=1}^{N_p} \left[a_p F_p(u) + b_p G_p(u)\right] \tag{6.11}$$

which will be written as

$$A(u) = \sum_{j=1}^{2N_p} x_j \Phi_j(u) = \mathbf{x} \cdot \mathbf{\Phi}(u). \tag{6.12}$$

These vectors are simply concatenations,

$$\begin{aligned} \mathbf{x} &= (a_1, b_1, a_2, b_2, \ldots, a_{N_p}, b_{N_p}) \\ \mathbf{\Phi}(u) &= (F_1(u), G_1(u), \ldots, F_{N_p}(u), G_{N_p}(u)), \end{aligned}$$

to put the expression into a standard form. Compare with equation 5.3 where $\mathbf{x}$ and $\mathbf{\Phi}(f)$ are real vectors. The problem of panel fitting to Microwave Holography data $A_m(u_i)$ for $i = 1, \ldots, N_u$ can now be stated as follows: find a vector $\mathbf{x}$ such that

$$|A_m(u_i) - \mathbf{x} \cdot \mathbf{\Phi}(u_i)| \leq E \text{ for each } i = 1, \ldots, N_u$$

where E is the permitted tolerance about a measured field value. Such a vector can be found by the Method of Successive Projections if these constraints are made to define sets:

$$Q_i = \{\mathbf{x} \mid |A_m(u_i) - \mathbf{x} \cdot \mathbf{\Phi}(u_i)| \leq E\}. \tag{6.13}$$

The intersection of all N_u sets will contain suitable vectors $\mathbf{x}$, plus a host of others which do not represent apertures. This is because a_p and b_p defined above have equal phase ℓ_p, so $\mathbf{x}$ must consist of pairs of elements with equal

phase. Once again, the flexibility of SP allows a simple solution: define one more set

$$R = \{\mathbf{x} \mid \arg x_{j-1} = \arg x_j \text{ for even } j\} \tag{6.14}$$

and project onto R in addition to the Q_i. The projections from any $\mathbf{x}_0$ onto Q_i and R can be derived using differential techniques; the derivations will not be repeated here. A single equation gives the projection onto Q_i:

$$\mathbf{x} = \mathbf{x}_0 - \left(1 - \frac{E}{|\mathbf{\Phi} \cdot \mathbf{x}_0 - A_m|}\right) \frac{\mathbf{\Phi} \cdot \mathbf{x}_0 - A_m}{\mathbf{\Phi}^* \cdot \mathbf{\Phi}} \mathbf{\Phi}^*. \tag{6.15}$$

The arguments u_i have been left out for clarity. The projection onto R is best described in steps to be applied to each pair of elements a'_p and b'_p of $\mathbf{x}$:

1. Calculate $C = \mathrm{Re}^2 a'_p - \mathrm{Im}^2 a'_p + \mathrm{Re}^2 b'_p - \mathrm{Im}^2 b'_p$;
2. Calculate $S = 2\mathrm{Re}\, a'_p \,\mathrm{Im}\, a'_p + 2\mathrm{Re}\, b'_p \,\mathrm{Im}\, b'_p$;
3. Find the projection phase from $2\phi = \arctan(S/C)$;
4. The projected elements are the components of a'_p and b'_p in this direction,
$$a_p = \mathrm{Re}\left(a'_p e^{-j\phi}\right) e^{+j\phi} \text{ and } b_p = \mathrm{Re}\left(b'_p e^{-j\phi}\right) e^{+j\phi}.$$

Encouraging results were obtained when these equations were implemented for the simulated one-dimensional aperture used in previous sections. Firstly, decisions had to be made on two aspects of the algorithm: the order of projections onto the large number of sets, and the tolerances E to allow about each measurement A_m. A lengthy study could be made, but a simple course was to watch the convergence of the rms aperture phase difference for a variety of example experiments. A tolerance $E = 0$ usually gave the best performance and was adopted, although further tests of the effects of measurement noise are needed. Three schemes for ordering the projections were tried:

- project onto Q_i in alternation with R for increasing u_i;
- project onto all Q_i for increasing u_i followed by a single projection onto R;
- project onto Q_i for the on-source sample $u_i = 0$, then the two Q_i one sample interval away, and so on outwards, followed by a single projection onto R.

Iteration	New SP rms Δ	Old SP rms Δ
0	19.63°	19.63°
1	3.70°	5.99°
2	1.10°	2.51°
3	1.23°	1.60°
4	1.23°	1.31°
5	1.24°	1.20°
6	1.24°	1.14°
7	1.24°	1.11°
8	1.24°	1.08°
9	1.24°	1.07°
10	1.24°	1.05°

Table 6.4: Convergence of the new, many-set SP method beginning at a uniformly illuminated, zero phase aperture. The rms phase difference from the phase of an original, one-dimensional aperture with 15 panels is shown. Convergence of the old, two-set SP method is shown for comparison.

Each pass through all the sets constituted one iteration, or repeating unit, of the algorithm. The first scheme did not perform as well as the other two, but these were quite similar: sometimes one was better, sometimes the other. The third one was adopted because of its greater symmetry.

The 15 panel aperture simulated for section 6.1 was used again to test the new algorithm. It was 26 m across, with $\lambda = 23.6$ mm and $\kappa = 0.825$. It had a 12 dB edge taper, so the approximation of a uniform illumination on each panel could be tested. Convergence was observed in the usual way, by the decrease of rms Δ, and a starting vector $\mathbf{x}_0$ was chosen to represent a uniformly illuminated, zero phase aperture. Table 6.4 shows rms Δ for 10 iterations of the new SP method together with 10 iterations of the old, two-set SP method, for which the same data and starting point were used. The number of far field samples taken was $N_u = 41$.

A noticeably better result was obtained with the old method, yet it is interesting that the new method converged rapidly to a steady state. Although not shown, the projection lengths were also constant beyond the sixth iteration. More surprising is that the same steady state arose for any randomly

N_u	P_p	rms Δ
31	1.71	—
33	1.82	—
35	1.93	$\star$
37	2.04	0.81°
39	2.14	1.25°
41	2.25	1.24°
43	2.36	0.97°
45	2.47	0.93°
47	2.58	1.00°
49	2.69	1.05°
51	2.80	1.12°
61	3.35	1.05°

Table 6.5: The steady states obtained with the new SP formulation for several N_u. For $N_u = 35$ ($\star$) a variety of end-points were obtained as the starting vector was changed, with rms Δ in the range 1°–4°. For smaller N_u no sensible limits were obtained.

chosen starting vector. A possible explanation is that the sets did not intersect and a repeating limit cycle was reached, the same sequence of projections recurring for each iteration. The minimum just before reaching the steady state, here at iteration 2, was also typical behaviour.

The steady states obtained for different spans of far field data were compared by testing an ensemble of random starting vectors for several N_u. As can be seen in table 6.5, a common steady state occurred for the whole ensemble only when $P_p \geq 2$, confirming again the importance of this parameter. The limit of rms Δ behaved in an irregular way as N_u was increased, and there seems to be no benefit in taking a very large set of measurements. This also agrees with previous experience. To show that useful recovered apertures were obtained, the tenth iterated vector for $N_u = 41$ was converted to panel parameters r_p, ℓ_p, and β_p and graphed in figure 6.18. The stepped illumination function follows the original as well as can be expected, and the panel tilts are well recovered except in regions of low illumination.

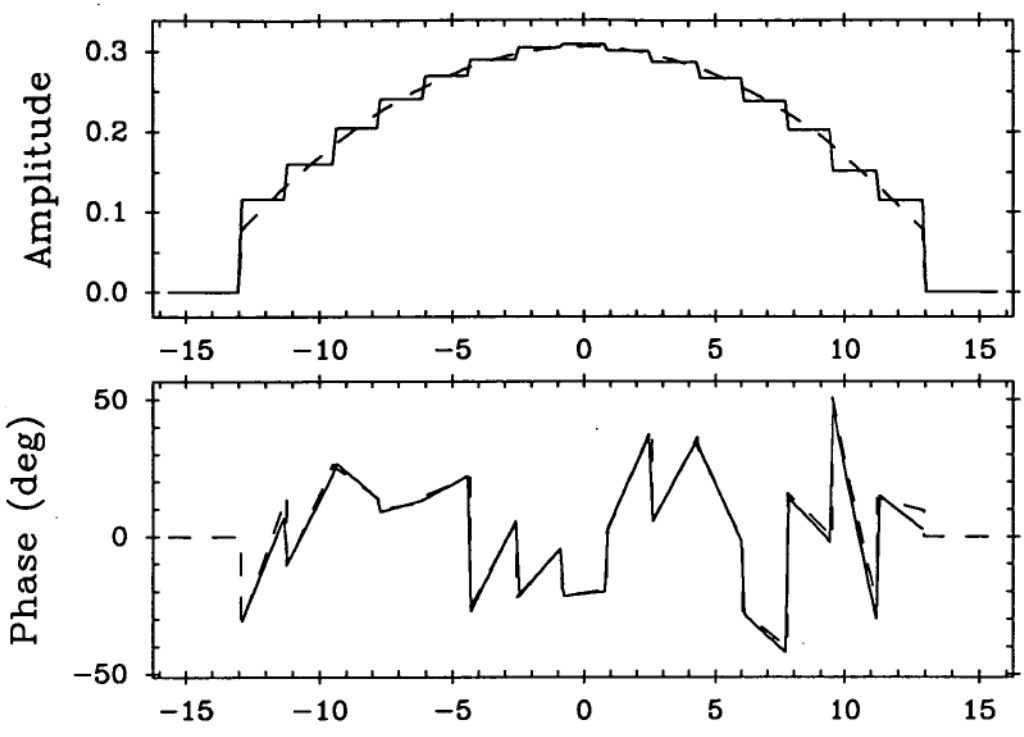


Figure 6.18: The recovered aperture function after 10 iterations of the new SP method. The number of far field samples was $N_u = 41$.

6.3.2 Two Dimensions

The one-dimensional tests above are encouraging. Most importantly, the algorithm was extremely fast because the vectors $\Phi(u_i)$ in equation 6.15 could be evaluated once and used for every iteration thereafter. Can this formulation be extended successfully to a two-dimensional aperture with a realistic layout of panels? As will be seen shortly, there were difficulties which prevented effective use of the method in its present form.

A linear expression for the total field $A(u, v)$ is obtained in the same way as before: by summing explicit expressions for the fields of N_p panels. The field of a trapezoidal panel with constant illumination r_p, vertical shift ℓ_p, and tilt parameters α_p and β_p can be calculated using standard Fourier transform operations. The result will be given without derivation here, using the panel geometry given in section 2.4.2. For a general panel centred at azimuth θ_0 and radius r_0 some rotational transformations are needed, so define

$$\begin{pmatrix} u' \\ v' \end{pmatrix} = \begin{pmatrix} \cos\theta_0 & \sin\theta_0 \\ -\sin\theta_0 & \cos\theta_0 \end{pmatrix} \cdot \begin{pmatrix} u \\ v \end{pmatrix}.$$

Then the field of the panel is

$$A_p(u,v) = r_p e^{+j\ell_p} e^{-j2\pi u' r_0} I\left(u' - \frac{\beta_p}{\pi h_p}, v' - \frac{\alpha_p}{\pi w_p}\right). \tag{6.16}$$

The integral $I(u,v)$ is the field of the panel if it were at the origin:

$$I(u,v) = \frac{jh_p}{2\pi v}\left[\frac{\sin(\pi P h_p)}{\pi P h_p} e^{-j\pi v w_p} - \frac{\sin(\pi Q h_p)}{\pi Q h_p} e^{+j\pi v w_p}\right] \tag{6.17}$$

where

$$P = u + \frac{v d_p}{h_p} \text{ and } Q = u - \frac{v d_p}{h_p}.$$

When the panel is rectangular, $d_p = 0$ and $I(u,v)$ reduces to the expected double sinc function,

$$I(u,v) = h_p \frac{\sin(\pi u h_p)}{\pi u h_p} w_p \frac{\sin(\pi v w_p)}{\pi v w_p}.$$

A first order Taylor expansion allows equation 6.16 to be linearized:

$$A_p(u,v) = r_p e^{+j\ell_p} e^{-j2\pi u' r_0}\left[I(u',v') - \frac{\beta_p}{\pi h_p} I_u(u',v') - \frac{\alpha_p}{\pi w_p} I_v(u',v')\right] \tag{6.18}$$

where I_u and I_v are partial derivatives of I which will not be reproduced here.

This is the analogue of equation 6.10. Define complex numbers a_p, b_p, and c_p to contain all the unknown panel parameters:

$$a_p = r_p e^{+j\ell_p} \quad b_p = \beta_p a_p \quad c_p = \alpha_p a_p.$$

Then functions can be defined in an obvious way to give

$$A_p(u,v) = a_p F_p(u,v) + b_p G_p(u,v) + c_p H_p(u,v)$$

which makes the parallel between the one- and two-dimensional cases very clear. The formulation of the panel fitting problem proceeds in an identical fashion from this point, except that the vectors $\mathbf{x}$ and $\Phi(u,v)$ are concatenations of triplets of terms instead of pairs:

$$\begin{aligned} \mathbf{x} &= (a_1, b_1, c_1, a_2, b_2, c_2, \ldots) \\ \Phi(u,v) &= (F_1(u,v), G_1(u,v), H_1(u,v), \ldots). \end{aligned}$$

The same sets Q_i are used, and the projections onto them are still found by equation 6.15. R is now the set of vectors made of equi-phase triplets, and the projection onto it is done by a simple extension of the previous method.

It is interesting that the problems turn out to be the same. Unfortunately, the first order Taylor expansion does not adequately represent the far field pattern of a two-dimensional panel. The field computed by equation 6.18 for a single panel of the Mount Pleasant 26 m antenna is shown in figure 6.19. It is plotted over the far field region that would be covered by a 53×53 holographic map with $\kappa = 0.825$. The sum of such fields for all 252 panels should reproduce the field of the whole antenna. In (a) the panel is not tilted, while in (b) tilts amounting to aperture phase gradients of 45° radially and laterally have been applied. That is, $\beta = \alpha = 45°$. Considerable distortion has occurred, so attempting to determine these tilts from holographic measurements would give unreliable results if linearity was assumed.

There is another, practical difficulty. Each $\Phi(u, v)$ has $3N_p$ complex elements, and there are $N_u N_v$ separate field points (u_i, v_i) for which this vector must be pre-calculated and stored. This constitutes a matrix of $3N_p N_u N_v$ elements, or 2.12×10^6 for the current example. The working space required by the algorithm is thus 33K blocks of 512 bytes, and the access time for such a large stored object will be considerable on most computers, not to mention the inconvenience of storing it. The computational time for $N_u N_v$ projections is also significant: if the norms $\Phi^* \cdot \Phi$ are pre-calculated $\sim 6N_p$ complex multiplications occur in each projection, or $\sim 6N_p N_u N_v$ in each iteration. This is still a smaller operation than the two $N'_u \times N'_v$ Fourier transforms required for an iteration of the old kind, so a speed advantage remains a possibility.

These problems — the poor Taylor approximation and the size of the workspace — make the new SP algorithm unwieldy in its present form. It is an interesting extension of previous work, however, and further research may show that a many-set formulation can be practical.

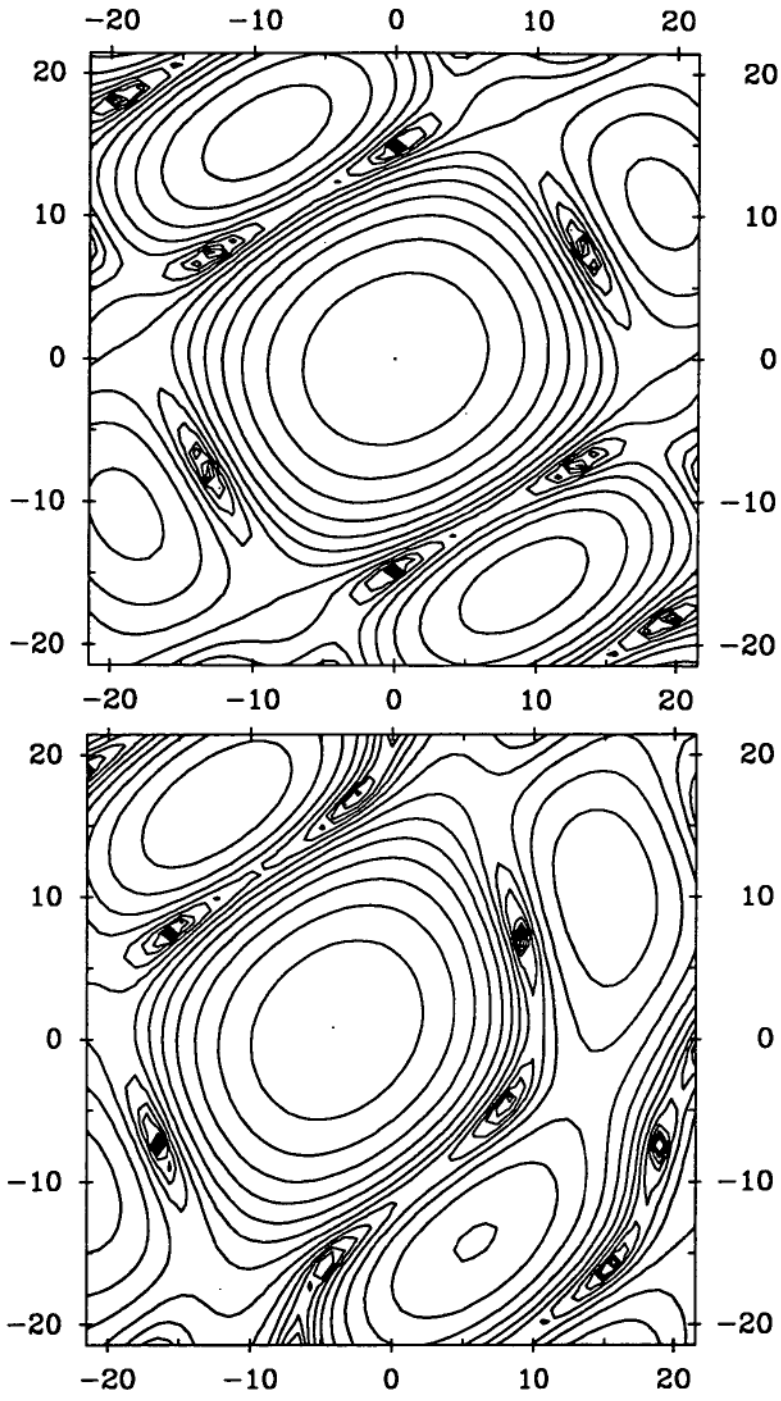


Figure 6.19: The field from a single panel of the 26 m antenna, computed for uniform illumination using a first order Taylor expansion of the explicit equation. Above, the panel is not tilted, while below there are tilts $\beta = \alpha = 45°$. The scale units are λ/D, which is somewhat less than the beamwidth of the whole antenna.

Conclusion

There have been two aspects to this project. Chapters 2 to 4 describe Microwave Holography experiments at the Mount Pleasant 26 m antenna. Chapters 5 and 6 describe the Successive Projections algorithm for panel fitting, which overcomes a major short-coming of traditional holographic measurements: limited aperture resolution.

As well as being essential for the Mount Pleasant experiments, the thorough treatment of antenna mounts and panel distortions in chapter 2 is useful for experiments with any antenna. The phase corrections necessary to compensate for antenna motion highlight the physical significance of the 'aperture plane'. The panel distortion model can be applied with a general grid of sample points, and consists of a conventional description of panel tilts summed with an independent 'twist' term.

Simple and effective hardware has been developed for holographic measurements using a signal from the AUSSAT-1 satellite. The front-end receiver and 2 m reference antenna are portable for measurements at other sites. Both carrier and beacon signals were used in the Ku-band, 12.25–75 GHz, and accuracies of ± 53 μm and ± 64 μm respectively were obtained for the estimated normal surface error at each aperture point. Panel adjustments based on these data resulted in an improved surface rms for the Mount Pleasant antenna, and the symmetry and size of systematic surface distortions were identified.

Systematic and random errors in Microwave Holography are discussed in chapter 4, and for random errors a model is proposed where additive, amplitude only, phase only, and pointing errors are treated separately. The effect of each category on the recovered aperture currents is derived, and the signal-to-noise ratios in the model can be estimated from the on-source signal. The model agrees very well with the repeatability of the holographic aperture maps. It seems likely that cross-coupling in the i.f. system was the dominant source of error.

Chapter 5 introduces the Method of Successive Projections, an iterative algorithm which is seen to describe the operation of several existing optimization algorithms. It can be applied to any problem that can be represented as the location of the intersection of several sets, provided that the projections onto the sets are readily calculable. Its application to the problem of fitting panel tilts and offsets to Microwave Holography data is considered in detail, and good convergence was obtained for the Mount Pleasant data. The processing time on a modern computer is a fraction of the 4 hours required to record the maps, and the results are expressed in a practically useful way as the adjustments needed at each panel corner.

The new algorithm is examined in chapter 6, first by studying the regularity of convergence with simulated data, then by finding generic features in an ensemble of convergence paths. Optimal experimental parameters are suggested that allow high-quality panel fitting to the smallest possible sets of data. Sampling of the far field should be as close to the Nyquist interval as aliasing allows, and the maps are sufficiently large if they give at least 3 aperture points on every panel. Several small maps make better use of the available observing time than a few large maps, as long as the Successive Projections algorithm is used to find panel positions. Otherwise, limited aperture resolution prevents accurate panel fitting.

This claim is supported by a rigorous analysis of the linear problem that arises when panel tilts and offsets are small. This analysis can be extended to include phase-less measurements, giving behaviour agreeing with the experience of other workers and thus confirming the validity of the linear approximation. Another Successive Projections formulation of the panel fitting problem has many sets instead of just two. It is too unwieldy in its current form, but refinements may be possible to make it practical.

There are several other areas where further research would be profitable. The panel fitting algorithm might be improved by the use of convergence acceleration techniques. The accuracy with which panel twists and higher-order distortions are recovered needs investigation. An end-to-end test of the algorithm would be possible by adjusting panels on a rigid reflector, where panel setting errors truly dominate other systematic distortions. Then 'before' and 'after' panel positions could be compared as in chapter 3. A mm-wave antenna would be a suitable 'workshop' for this test, and the panel-fitting tool which Successive Projections provides will be most useful for these extremely accurate structures.

A more sophisticated way of incorporating aperture blockage effects may be possible. The effects of subreflector blockage and diffraction have never been thoroughly treated. These are the major problems remaining for Microwave Holography, and it is possible that Successive Projections will find application including prior knowledge of diffraction effects into the analysis. The Method of Successive Projections itself is useful in a much wider context, as can be judged from the range of applications mentioned. It may allow a rigorous treatment of many existing signal processing algorithms which currently are used but not well understood.

Bibliography

[1] A. Abo-Taleb and M. M. Fahmy, "Design of FIR two-dimensional digital filters by successive projections," *IEEE Trans. Circuits Syst.*, vol. CAS-31, no. 9, pp. 801–5, 1984.

[2] H. Bach, "A Note on Antennas: Definitions and Methods," *IEEE Trans. Antennas Propagat.*, vol. AP-35, no. 3, pp. 325–7, 1987.

[3] C. A. Balanis, *Antenna Theory: Analysis and Design*, John Wiley & Sons, 1982.

[4] R. H. T. Bates, "Fourier Phase Problems are Uniquely Solvable in More than One Dimension," *Optik*, vol. 61, no. 3, pp. 247–62, 1982.

[5] W. Batrla, H. E. Matthews, K. M. Menten, and C. M. Walmsley, "Detection of strong methanol masers towards galactic H II regions," *Nature*, vol. 326, pp. 49–51, 1987.

[6] J. C. Bennett, A. P. Anderson, P. A. McInnes, and A. J. T. Whitaker, "Microwave Holographic Metrology of Large Reflector Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-24, no. 13, pp. 295–303, 1976.

[7] J. C. Bennett and M. P. Godwin, "Necessary Criteria for the Diagnosis of Panel Misalignments in Large Reflector Antennas by Microwave Metrology," *Electron. Lett.*, vol. 13, no. 16, pp. 463–5, 1977.

[8] R. N. Bracewell, *The Fourier Transform and its Applications (second edition)*, chap. 10, McGraw-Hill, 1978.

[9] R. N. Bracewell and A. R. Thompson, "The Main Beam and Ringlobes of an East-West Rotation-Synthesis Array," *Astrophys. J.*, vol. 182, pp. 77–94, 1973.

[10] Yu. M. Bruck and L. G. Sodin, "On the ambiguity of the image reconstruction problem," *Optics Commun.*, vol. 30, pp. 304–8, 1979.

[11] P. J. B. Clarricoats, "Reconfigurable Reflector Antennas for Spacecraft Applications," presented at the *XXIII General Assembly of the URSI*, Prague, Czechoslovakia, August–September 1990.

[12] P. J. B. Clarricoats and G. T. Poulton, "High-Efficiency Microwave Reflector Antennas — A Review," *Proc. IEEE*, vol. 65, no. 10, pp. 1470–1504, 1977.

[13] R. E. Colin and F. J. Zucker, *Antenna Theory*, McGraw-Hill Book Company, 1969.

[14] G. G. Cook, A. P. Anderson, A. J. T. Whitaker, and J. C. Bennett, "High Resolution Three-Dimensional Microwave Imaging of Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-37, no. 6, pp. 768–79, 1989.

[15] D. J. Edwards, "Determination of Large Reflector Profiles by Microwave Wavefront Reconstruction," *Post Off. Electr. Eng. J. (GB)*, vol. 74, pp. 106–12, 1981.

[16] J. W. Findlay and J. M. Payne, "An Instrument for Measuring Deformations in Large Structures," *IEEE Trans. Instrum. Meas.*, vol. IM-23, no. 3, pp. 221–6, 1974.

[17] C. S. Fraser, "Microwave Antenna Measurement," *Photogramm. Eng. Remote Sensing*, vol. 52, no. 10, pp. 1627–35, 1986.

[18] R. W. Gerchberg and W. O. Saxton, "A Practical Algorithm for the Determination of Phase from Image and Diffraction Plane Pictures," *Optik*, vol. 35, no. 2, pp. 237–46, 1972.

[19] M. P. Godwin, A. P. Anderson, and J. C. Bennett, "Optimisation of feed position and improved profile mapping of a reflector antenna from microwave holographic measurements," *Electron. Lett.*, vol. 14, no. 5, pp. 134–6, 1978.

[20] M. P. Godwin, E. P. Schoessow, and B. H. Grahl, "Improvement of the Effelsberg 100 meter telescope based on holographic reflector surface measurement," *Astron. Astrophys.*, vol. 167, pp. 390–4, 1986.

[21] M. P. Godwin, A. J. T. Whitaker, J. C. Bennett, and A. P. Anderson, "Microwave diagnostics of the Chilbolton 25 m antenna using the OTS satellite," *Inst. Elec. Eng. Conf. Pub.*, vol. 195, pp. 232–6, 1981.

[22] H. Goldstein, *Classical Mechanics (second edition)*, Addison-Wesley, 1980.

[23] A. Greve, "Metrology of the Effelsberg 100 Meter Radio Reflector," *Zt. f. Vermessungswesen*, 106 Jahrg., Heft 6, pp. 308–15, 1981.

[24] A. Greve, "Reflector Surface Measurements of the IRAM 30-m Radio Telescope," *Int. J. Infrared Millim. Waves*, vol. 7, no. 1, pp. 121–35, 1986.

[25] L. G. Gubin, B. T. Polyak, and E. V. Raik, "The method of projections for finding the common point of convex sets," *USSR Comput. Math. Math. Phys.*, vol. 7, no. 6, pp. 1–24, 1967.

[26] I. Guttman, S. S. Wilks, and J. S. Hunter, *Introductory Engineering Statistics*, sect. 15.7, John Wiley & Sons, 1971.

[27] R. E. Hills, "Phase-recovery holography in the Fresnel region," presented at the *International Workshop on Microwave Holography*, Nizhnij Arkhyz, USSR, September 1990.

[28] R. E. Hills, private discussions.

[29] M. Ishiguro, S. S. Hayashi, K. Morita, T. Takahashi, N. Kaifu, M. Masuda, K. Ohgi, and H. Kondo, "Development of a High-resolution Surface Measuring System Using Radio Holography," *Nobeyama Radio Observatory*, Technical Report No. 10, 1987. Reprinted from *Tech. Report, Inst. Electronics, Information and Communications Engineers, Japan*, vol. A.P87-28, pp. 67–74, 1987. Both are in Japanese.

[30] M. Ishiguro, K. Morita, S.S. Hayashi, M. Masuda, T. Ebisui, and S. Betsudan, "Influence of Aperture Illumination on the Surface Measurement Error Using Radio Holographic Metrology," *Nobeyama Radio Observatory*, Technical Report No. 11, 1987. Reprinted as above.

[31] M. Ishiguro, K. Morita, S. Hayashi, T. Masuda, T. Hirukoi, and S. Betsudan, "Measurement of dish accuracy of 45-metre radio telescope using radio wave holography," *Technical Reports, Mitsubishi Electric Corporation*, vol. 62, no. 5, pp. 69–74, 1988. In Japanese.

[32] G. L. James, *Geometrical Theory of Diffraction for Electromagnetic Waves (third edition revised)*, Peter Peregrinus Ltd., 1986.

[33] A. P. Kalcina, R. D. Hawke, and G. T. Neil, "Optimising Satellite Earth Station Antennas by Microwave Holography," *J. Electr. Electron. Eng. (Aust.)*, vol. 7, no. 4, pp. 267–70, 1987.

[34] M. J. Kesteven, "A Holographic Survey of Sydney-1A, The OTC 18 m Antenna at Oxford Falls," *Contract Report No. AE8905-F1, RPP 3391*, CSIRO Division of Radiophysics, Australia, 1989.

[35] M. J. Kesteven, B. F. Parsons, and D. E. Yabsley, "Antenna Reflector Metrology: The Australia Telescope Experience," *IEEE Trans. Antennas Propagat.*, vol. AP-36, no. 10, pp. 1481–4, 1988.

[36] M. Kesteven, K. Wellington, M. Calabretta, D. Cooke, L. Loew, B. Parsons, R. Shields, and D. Yabsley, "A Report on the 1988/89 Campaign to Assess and Readjust the Surface of the Parkes Radio Telescope," CSIRO Division of Radiophysics, Australia, 1989.

[37] J. D. Kraus, *Antennas (second edition)*, McGraw-Hill, 1988.

[38] E. Kreyszig, *Introductory Functional Analysis with Applications*, John Wiley & Sons, 1978.

[39] S. W. Lee, P. Cramer, K. Woo, and Y. Rahmat-Samii, "Diffraction by an Arbitrary Subreflector: GTD Solution," *IEEE Trans. Antennas Propagat.*, vol. AP-27, no. 3, pp. 305–16, 1979.

[40] A. Levi and H. Stark, "Image restoration by the method of generalised projections with application to restoration from magnitude," *J. Opt. Soc. Am. A*, vol. 1, no. 9, pp. 932–43, 1984.

[41] H. Ling, Y. T. Lo, Y. Rahmat-Samii, "Reflector Sidelobe Degradation Due to Random Surface Errors," *IEEE Trans. Antennas Propagat.*, vol. AP-34, no. 2, pp. 164–72, 1986.

[42] A. C. Ludwig, "The Definition of Cross Polarization," *IEEE Trans. Antennas Propagat.*, vol. AP-21, pp. 116–9, 1973.

[43] J. E. McCormack, G. Junkin, and A. P. Anderson, "Microwave metrology of reflector antennas from a single amplitude," *IEE Proceedings*, vol. 137, pt. H, no. 5, pp. 276–84, 1990.

[44] D. K. C. MacDonald, *Noise and fluctuations*, John Wiley & Sons, 1962.

[45] C. E. Mayer, J. H. Davis, and H. D. Foltz, "Texas 5-m Antenna Aperture Efficiency Doubles from 230–300 GHz with Error Compensating Secondary," *IEEE Trans. Antennas Propagat.*, vol. AP-39, no. 3, pp. 309–17, 1991.

[46] C. E. Mayer, J. H. Davis, W. L. Peters, and W. J. Vogel, "A Holographic Surface Measurement of the Texas 4.9-m Antenna at 86 GHz," *IEEE Trans. Instrum. Meas.*, vol. IM-32, no. 1, pp. 102–9, 1983.

[47] D. L. Missel, "An examination of an iterative method for the solution of the phase problem in optics and electron optics: I & II," *J. Phys. D: Appl. Phys.*, vol. 6, pp. 2200–25, 1973.

[48] R. Mittra, Y. Rahmat-Samii, V. Galindo-Israel, and R. Norman, "An Efficient Technique for the Computation of Vector Secondary Patterns of Offset Paraboloid Reflectors," *IEEE Trans. Antennas Propagat.*, vol. AP-27, no. 3, pp. 294–304, 1979.

[49] D. Morris, "Phase Retrieval in the Radio Holography of Reflector Antennas and Radio Telescopes," *IEEE Trans. Antennas Propagat.*, vol. AP-33, no. 7, pp. 749–55, 1985.

[50] D. Morris, H. Hein, H. Steppe, and J. W. M. Baars, "Phase retrieval radio holography in the Fresnel region: tests on the 30 m telescope at 86 GHz," *IEE Proceedings*, vol. 135, pt. H, no. 1, pp. 61–4, 1988.

[51] D. Morris, J. W. M. Baars, H. Hein, H. Steppe, C. Thum, and R. Wohlleben, "Radio-holographic reflector measurement of the 30-m millimeter radio telescope at 22 GHz with a cosmic signal source," *Astron. Astrophys.*, vol. 203, pp. 399–406, 1988.

[52] P. J. Napier and R. H. T. Bates, "Antenna — aperture distributions from holographic type of radiation-pattern measurement," *Proc. Inst. Elec. Eng.*, vol. 120, no. 1, pp. 30–4, 1973.

[53] A. Papoulis, "A New Algorithm in Spectral Analysis and Band-Limited Extrapolation," *IEEE Trans. Circuits Syst.*, vol. CAS-22, no. 9, pp. 735–42, 1975.

[54] C. G. Parini, A. K. K. Lau, and P. J. B. Clarricoats, "Phase-only reflector antenna metrology," *IEE Proceedings*, vol. 136, pt. H, no. 4, pp. 343–9, 1989.

[55] J. M. Payne, "An Active Surface for the Green Bank Telescope," presented at the *XXIII General Assembly of the URSI*, Prague, Czechoslovakia, August–September 1990.

[56] J. M. Payne, J. M. Hollis, and J. W. Findlay, "New method of measuring the shape of precise antenna reflectors," *Rev. Sci. Instrum.*, vol. 47, no. 1, pp. 50–5, 1976.

[57] G. A. Pinchuk and A. A. Stotskii, "Use of radioholography to test the RATAN-600 radio telescope," *Astrofiz. Issled. Izv. Spets. Astrofiz. Obs.*, vol. 16, pp. 135–45, 1982.

[58] K. Pontoppidan, "Reflector Antennas — An Overview of Surface Distortion Effects," *Proc. JINA '86*, pp. 82–93, Nice, 1986.

[59] G. T. Poulton, "Antenna power pattern synthesis using method of successive projections," *Electron. Lett.*, vol. 22, no. 20, pp. 1042–3, 1986.

[60] G. T. Poulton, "Applications for Projection Methods in Antenna Design and Measurement," presented at the *Second Australian Symposium on Antennas*, CSIRO Division of Radiophysics, Epping, Australia, February 1989.

[61] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes — The Art of Scientific Computing*, Cambridge University Press, 1986.

[62] Y. Rahmat-Samii, "An Efficient Computational Method for Characterizing the Effects of Random Surface Errors on the Average Power Pattern of Reflectors," *IEEE Trans. Antennas Propagat.*, vol. AP-31, no. 1, pp. 92–8, 1983.

[63] Y. Rahmat-Samii, "Surface diagnosis of large reflector antennas using microwave holographic metrology: An iterative approach," *Radio Science*, vol. 19, no. 5, pp. 1205–17, 1984.

[64] Y. Rahmat-Samii, "Effects of deterministic surface distortions on reflector antenna performance," *Ann. Télécommun.*, vol. 40, nos. 7–8, pp. 350–60, 1985.

[65] Y. Rahmat-Samii, "Microwave Holography of Large Reflector Antennas — Simulation Algorithms," *IEEE Trans. Antennas Propagat.*, vol. AP-33, no. 11, pp. 1194–1203, 1985.

[66] Y. Rahmat-Samii and V. Galindo-Israel, "Shaped Reflector Antenna Analysis using the Jacobi-Bessel Series," *IEEE Trans. Antennas Propagat.*, vol. AP-28, no. 4, pp. 425–35, 1980.

[67] Y. Rahmat-Samii, S. Gulkis, G. S. Levy, B. L. Seidel, L. E. Young, M. J. Batty, and D. L. Jauncey, "Microwave Holographic Surface Measurement of the Tidbinbilla 64-m Antenna," *Proc. Astron. Soc. Aust.*, vol. 5, no. 2, pp. 270–2, 1983.

[68] R. A. Roberts and C. T. Mullis, *Digital Signal Processing*, Addison-Wesley, 1987.

[69] K. Rohlfs, *Tools of Radio Astronomy*, Springer-Verlag, 1986.

[70] W. V. T. Rusch, "Forward Scattering from Cylinders of Triangular Cross Section," *IEEE Trans. Antennas Propagat.*, vol. AP-26, no. 6, pp. 849–50, 1978.

[71] W. V. T. Rusch, J. Appel-Hansen, C. A. Klein, and R. Mittra, "Forward Scattering from Square Cylinders in the Resonance Region with Application to Aperture Blockage," *IEEE Trans. Antennas Propagat.*, vol. AP-24, no. 2, pp. 182–9, 1976.

[72] W. V. T. Rusch, O. Sørensen, and J. W. M. Baars, "Radiation Cones from Feed-Support Struts of Symmetric Paraboloidal Reflectors," *IEEE Trans. Antennas Propagat.*, vol. AP-30, no. 4, pp. 786–90, 1982.

[73] W. V. T. Rusch, L. R. Welch, and G. E. Mires, "Observation-Point-Dependent Blocking Shadows on a Reflector Antenna," *IEEE Trans. Antennas Propagat.*, vol. AP-37, no. 6, pp. 690–7, 1989.

[74] J. Ruze, "Antenna Tolerance Theory — A Review," *Proc. IEEE*, vol. 54, no. 4, pp. 633–40, 1966.

[75] D. G. Sandler, T. K. Barrett, D. A. Palmer, R. Q. Fugate, and W. J. Wild, "Use of a neural network to control an adaptive optics system for an astronomical telescope," *Nature*, vol. 351, pp. 300–2, 1991.

[76] P. F. Scott and M. Ryle, "A rapid method for measuring the figure of a radio telescope reflector," *Mon. Not. R. Astr. Soc.*, vol. 178, pp. 539–45, 1977.

[77] E. Serabyn, T. G. Phillips, and C. R. Masson, "Surface figure measurements of radio telescopes with a shearing interferometer," *Appl. Opt.*, vol. 30, no. 10, pp. 1227–41, 1991.

[78] V. Shulga and Yu. Shamanin, "The equipment for phase measurements of 70-m radio telescope in Sufa," presented at the *International Workshop on Microwave Holography*, Nizhnij Arkhyz, USSR, September 1990.

[79] S. Silver (ed.), *Microwave Antenna Theory and Design*, McGraw-Hill Book Company, 1949.

[80] R. H. Slater, "Metrology and radio performance of reflector of Chilbolton aerial," *Proc. Inst. Elec. Eng.*, vol. 118, no. 12, pp. 1691–7, 1971.

[81] L. S. Taylor, "The Phase Retrieval Problem," *IEEE Trans. Antennas Propagat.*, vol. AP-29, no. 2, pp. 386–91, 1981.

[82] A. R. Thompson, J. M. Moran, and G. W. Swenson, *Interferometry and Synthesis in Radio Astronomy*, John Wiley & Sons, 1986.

[83] V. K. Tripp, "A New Approach to the Analysis of Random Errors in Aperture Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-32, no. 8, pp. 857–63, 1984.

[84] P. J. Wood, *Reflector antenna analysis and design*, Peter Peregrinus, 1980.

[85] S. von Hoerner, "Telescope Surface Measurement with Two Feeds," *IEEE Trans. Antennas Propagat.*, vol. AP-26, no. 6, pp. 857–60, 1978.

[86] S. von Hoerner, "Internal Twist and Least-Squares Adjustment of Four-Cornered Surface Plates for Reflector Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-29, no. 16, pp. 953–8, 1981.

[87] T. B. Vu, "Influence of correlation interval and illumination taper in antenna tolerance theory," *Proc. Inst. Elec. Eng.*, vol. 116, no. 2, pp. 195–202, 1969.

[88] A. D. Yaghjian, "Equivalence of Surface Current and Aperture Field Integrations for Reflector Antennas," *IEEE Trans. Antennas Propagat.*, vol. AP-32, no. 12, pp. 1355–8, 1984.

Appendices

A The Convolving Function

In section 2.2.1 it is necessary to sum a finite cosine series:

$$\sum_{m=-L}^{m=+L} \cos my \text{ where } y = 2\pi\delta_u x.$$

This can be done by starting with the standard result

$$\sum_{m=0}^{n-1} \cos my = \frac{\sin\frac{ny}{2}\cos\frac{(n-1)y}{2}}{\sin\frac{y}{2}},$$

which can be proved by summing a geometrical series with first term 1 and ratio e^{jy}. Now, putting $n-1=L$,

$$\sum_{m=0}^{m=L} \cos my = \frac{\sin\frac{(L+1)y}{2}\cos\frac{Ly}{2}}{\sin\frac{y}{2}}.$$

Thus

$$\begin{aligned}
\sum_{m=-L}^{m=+L} \cos my &= 2\sum_{m=0}^{m=L} \cos my - 1 \\
&= \frac{2\sin\frac{(L+1)y}{2}\cos\frac{Ly}{2} - \sin\frac{y}{2}}{\sin\frac{y}{2}} \\
&= \frac{2\sin\frac{Ly}{2}\cos\frac{y}{2}\cos\frac{Ly}{2} + 2\cos\frac{Ly}{2}\sin\frac{y}{2}\cos\frac{Ly}{2} - \sin\frac{y}{2}}{\sin\frac{y}{2}} \\
&= \frac{\left[2\sin\frac{Ly}{2}\cos\frac{Ly}{2}\right]\cos\frac{y}{2} + \left[2\cos^2\frac{Ly}{2} - 1\right]\sin\frac{y}{2}}{\sin\frac{y}{2}} \\
&= \frac{\sin\frac{2Ly}{2}\cos\frac{y}{2} + \cos\frac{2Ly}{2}\sin\frac{y}{2}}{\sin\frac{y}{2}} \\
&= \frac{\sin\left(\frac{2Ly}{2} + \frac{y}{2}\right)}{\sin\frac{y}{2}} \\
&= \frac{\sin\frac{Ny}{2}}{\sin\frac{y}{2}}
\end{aligned}$$

where $N = 2L + 1$, and this result is used in section 2.2.1.

Given this result, the replication of the sinc function can be shown to equal the same function provided that this Fourier series is accepted:

$$\sum_{k=-\infty}^{k=+\infty} N\mathrm{sinc}\, N(\theta + k) = 1 + 2\sum_{k=1}^{k=L} \cos(2\pi k\theta).$$

This result is quoted without proof in [9, p. 86]. Written another way,

$$\begin{aligned} \mathrm{III}(\theta) * N\mathrm{sinc}\,(N\theta) &= 2\sum_{k=0}^{k=L} \cos(2\pi k\theta) - 1 \\ &= \frac{\sin(\pi N\theta)}{\sin(\pi\theta)}. \end{aligned}$$

B Euler Angles ϕ and θ

In section 2.3 it is necessary to deduce the Euler angles ϕ and θ of the rotation matrix $\mathbf{P}(\phi, \theta, 0)$, given vectors $\mathbf{r}$ in the original frame and $\mathbf{r}'$ in the rotated frame. The transformation with $\psi = 0$ may be written in full as

$$\begin{aligned} x' &= & x\cos\phi & + & y\sin\phi & & \\ y' &= - & x\sin\phi\cos\theta & + & y\cos\phi\cos\theta & + & z\sin\theta \\ z' &= & x\sin\phi\sin\theta & - & y\cos\phi\sin\theta & + & z\cos\theta. \end{aligned}$$

Making the substitution

$$\left.\begin{aligned} x &= a\cos\alpha \\ y &= a\sin\alpha \end{aligned}\right\} \Rightarrow a = \sqrt{x^2 + y^2} \text{ and } \tan\alpha = y/x$$

gives the equations

$$\begin{aligned} x' &= & a\cos(\alpha - \phi) & & & \\ y' &= & a\sin(\alpha - \phi)\cos\theta & + & z\sin\theta \\ z' &= - & a\sin(\alpha - \phi)\sin\theta & + & z\cos\theta. \end{aligned}$$

The second and third equations are a two-dimensional rotation by angle θ which may be inverted to give

$$\begin{aligned} a\sin(\alpha - \phi) &= y'\cos\theta & - & z'\sin\theta \\ z &= y'\sin\theta & - & z'\cos\theta. \end{aligned}$$

A second substitution

$$\left.\begin{aligned} y' &= b\cos\beta \\ z' &= b\sin\beta \end{aligned}\right\} \Rightarrow b = \sqrt{y'^2 + z'^2} \text{ and } \tan\beta = z'/y'$$

gives

$$\begin{aligned} a\sin(\alpha-\phi) &= b\cos(\beta+\theta) \\ z &= b\sin(\beta+\theta). \end{aligned}$$

From these equations it is simple to show that

$$|b\cos(\beta+\theta)| = t \text{ and } |a\sin(\alpha-\phi)| = t$$

where, because $|\mathbf{r}| = |\mathbf{r}'|$ for vectors related by a rotation,

$$t = \sqrt{a^2 - x'^2} = \sqrt{b^2 - z^2} = \sqrt{|\mathbf{r}|^2 - x'^2 - z^2}.$$

Thus the angles ϕ and θ can be found from these relations:

$$\begin{aligned} \sin(\alpha-\phi) &= \pm t/a \\ \cos(\alpha-\phi) &= x'/a \\ \sin(\beta+\theta) &= z/b \\ \cos(\beta+\theta) &= \pm t/b \end{aligned}$$

where either both upper or both lower signs must be chosen. When t is imaginary, that is $|\mathbf{r}|^2 - x'^2 - z^2 < 0$, there are no solutions. This behaviour is understood by considering the locus of solutions (ϕ, θ, ψ) when all three Euler angles are allowed to vary. The locus is traced out as the rotated frame is turned around the vector $\mathbf{r}'$. It will cut the $\psi = 0$ plane either twice or not at all, with two degenerate solutions when $t = 0$. The other special cases to consider are $a = 0$ and $b = 0$, for which it is easy to show that ϕ and θ respectively are undetermined. The other angle is found as follows:

$$a = 0 \;\Rightarrow\; x = y = 0 \;\Rightarrow\; \begin{cases} \sin\theta = y'/z \\ \cos\theta = z'/z \end{cases}$$

$$b = 0 \;\Rightarrow\; y' = z' = 0 \;\Rightarrow\; \begin{cases} \sin\phi = y/x' \\ \cos\phi = x/x' \end{cases}.$$

These computations make a simple computer algorithm.

C Fitting by Least Squares

Following are outlines of least squares fitting methods referred to but not described in the text. In section 3.2 a plane is fitted to the recovered aperture phase function and subtracted so that a phase ramp caused by a pointing offset does not affect the result. In section 3.1 the polarimeter is calibrated

by fitting a sinusoid to the variation of the U and V outputs with time. In section 2.4 deformations of a reflector panel are modelled by four parameters which must be fitted from the aperture phase information, either as a separate operation or later as part of the Successive Projections algorithm.

It is not necessary to estimate the errors of the fitted parameters in these applications, because the repeatability of independent experiments is always used as an indication of the level of error from all sources. Only the equations for the parameter estimates themselves are given below.

C.1 Fitting a Plane

Given a set of points (x_i, y_i) in the two-dimensional plane, a function

$$\phi = ax + by + c$$

is to be fitted to measurements ϕ_i at each point, minimizing the chi-square merit function

$$\chi^2 = \sum_i w_i \left(\phi_i - ax_i - by_i - c\right)^2 .$$

Here i ranges over the data, and w_i is the weight for each point which is often chosen to be $1/{\sigma_i}^2$ where σ_i is the uncertainty associated with ϕ_i. It is assumed that the errors in the ϕ_i are much greater than the errors in x_i or y_i.

This algorithm is an extension of the usual method for fitting data to a straight line, for example see Press, et al. [61, sect. 14.2]. Form the sums

$$S = \sum_i w_i \quad S_x = \frac{1}{S}\sum_i w_i x_i \quad S_y = \frac{1}{S}\sum_i w_i y_i \quad S_\phi = \frac{1}{S}\sum_i w_i \phi_i$$

and define

$$u_i = x_i - S_x \text{ and } v_i = y_i - S_y .$$

Then form the sums

$$S_{uu} = \sum_i w_i u_i u_i \quad S_{uv} = \sum_i w_i u_i v_i \quad S_{vv} = \sum_i w_i v_i v_i$$
$$S_{u\phi} = \sum_i w_i u_i \phi_i \quad S_{v\phi} = \sum_i w_i v_i \phi_i .$$

The coefficients of x and y are found by solving the matrix equation

$$\begin{pmatrix} S_{uu} & S_{uv} \\ S_{uv} & S_{vv} \end{pmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} S_{u\phi} \\ S_{v\phi} \end{pmatrix},$$

and finally c is found by the equation

$$c = S_\phi - aS_x - bS_y .$$

C.2 Fitting a Sinusoid

Now a general sinusoid

$$\phi = \Phi \sin(\omega t + \delta) + \Phi_0$$

must be fitted to measurements ϕ_i for ordinates t_i. Recast the problem by defining

$$x = \sin \omega t \text{ and } y = \cos \omega t$$

and let

$$\left.\begin{array}{rcl} a &=& \Phi \cos \delta \\ b &=& \Phi \sin \delta \\ c &=& \Phi_0 \end{array}\right\} \Rightarrow \Phi = \sqrt{a^2 + b^2} \text{ and } \tan \delta = b/a.$$

Then

$$\phi = ax + by + c$$

which is the equation for a plane, and the previous algorithm can be applied without change.

C.3 Fitting a Twisted Panel

The equation that models a twisted panel is

$$\phi = (\tau y + \beta)x + (\alpha y + \ell) = \tau xy + \beta x + \alpha y + \ell$$

and for a least squares fit the function to be minimized is

$$\chi^2 = \sum_i w_i (\phi_i - \tau x_i y_i - \beta x_i - \alpha y_i - \ell)^2 .$$

The independent variables x and y are related to x_p and y_p of section 2.4 by

$$x = \frac{x_p}{h/2} \text{ and } y = \frac{y_p}{w/2 - x_p d/h}$$

where h, w, and d are constants of the panel. When twisting of the panel is suppressed, $\tau = 0$ and the plane fitting method above can be used to find the other parameters.

When all four parameters must be determined, form the sums S, S_x, S_y, S_ϕ, S_{uu}, S_{uv}, S_{vv}, $S_{u\phi}$, and $S_{v\phi}$ exactly as for the planar fit. In addition, form the sums

$$S_{uvu} = \sum_i w_i u_i v_i u_i \quad S_{uvv} = \sum_i w_i u_i v_i v_i$$

$$S_{uv\phi} = \sum_i w_i u_i v_i \phi_i \quad S_{uuvv} = \sum_i w_i u_i{}^2 v_i{}^2 .$$

Then calculate the following elements, whose names indicate that they are part of a 3-D matrix equation:

$$\begin{aligned} a_{11} &= S_{uuvv} + S_{uvu}S_y + S_{uvv}S_x - {S_{uv}}^2/S \\ a_{21} &= S_{uvu} + S_{uu}S_y + S_{uv}S_x \\ a_{31} &= S_{uvv} + S_{uv}S_y + S_{vv}S_x \\ v_1 &= S_{uv\phi} - S_{uv}S_\phi. \end{aligned}$$

Find the tilt parameters β and α by solving the matrix equation

$$\begin{pmatrix} a_{11}S_{uu} - a_{21}S_{uvu} & a_{11}S_{uv} - a_{21}S_{uvv} \\ a_{11}S_{uv} - a_{31}S_{uvu} & a_{11}S_{vv} - a_{31}S_{uvv} \end{pmatrix} \cdot \begin{pmatrix} \beta \\ \alpha \end{pmatrix} = \begin{pmatrix} a_{11}S_{u\phi} - a_{21}v_1 \\ a_{11}S_{v\phi} - a_{31}v_1 \end{pmatrix}.$$

The twist parameter τ and the constant offset ℓ are given by

$$\begin{aligned} \tau &= \frac{v_1 - \beta S_{uvu} - \alpha S_{uvv}}{a_{11}} \\ \ell &= S_\phi - \tau\,(S_x S_y + S_{uv}/S) - \beta S_x - \alpha S_y. \end{aligned}$$

D Header Information for Scans

Here is the header used at Mount Pleasant to record information about each scan (see section 3.2). The language is FORTRAN, and declarations are in the order of storage. A single sample consisted of a `real` value from each polarimeter channel. This defined a 'record' size of 16 bytes, and it was convenient to maintain this division for the header: each line declares 16 bytes.

```
double precision  jdcal, jdscn       ! for previous scan
double precision  jdsat, scanno      ! for this scan
double precision  sleft, sright      ! source position
double precision  xstart, ystart     ! scan start position
double precision  xstop, ystop       ! scan stop position
double precision  xrate, yrate       ! antenna axis rates
double precision  xmark, ymark       ! interrupt spacing
real              cal_offsource(4)   ! Ix, U, V, and Iy
real              cal_onsource(4)    ! Ix, U, V, and Iy
```

A celestial source position would not be specified in azimuth and elevation, so the notation (`sleft,sright`) allows for any coordinate system. Antenna positions may refer to an XY mount or an AZEL mount. The method of

obtaining times prevented `jdcal` and `jdscn` from being available until after the scan, so they were stored in the following header. The final calibration samples occupied a separate header which followed the last scan, and `jdcal` for the last scan was appended to this.

The quantities which should appear in the map header are shown below. Others can be included, but these are essential. Some explanation: if a `prediction_file` is not available, or if the source is celestial, a fixed position can be given. In either case the coordinate system used is indicated by `source_coord`, and a constant offset of (`left_offset`, `right_offset`) can be applied. If sample interrupts are triggered at constant intervals in time, `sample_gap_time` is this interval. Otherwise, control system interrupts are used, and one of `xmark` and `ymark` above is the constant angular interval between sample interrupts. Clearly, only one axis can give interrupts even if both are in motion, so one of `xmark` and `ymark` should be set to zero. The number of central grid points having the A channel attenuator switched in is indicated by `attenuate_extent`.

```
character*50      dataset_name
character*50      antenna_name
character*50      source_name
double precision  frequency
character*50      prediction_file    ! from AUSSAT data
double precision  source_left        ! if no predictions
double precision  source_right       ! if no predictions
integer           source_coord
double precision  left_offset
double precision  right_offset
integer           interrupt_mode     ! time or system
double precision  integration_time
double precision  sample_gap_time
integer           number_scans
integer           length_scans
integer           attenuate_extent   ! from map centre
```

E Proof that $2S_E = -\|E\|^2$

In section 5.2.4 the constraint $P'_M = P_M$ was said to be enough to determine the sum S_E. This will now be proved. Using the notation of that section,

$$\begin{aligned}
S_E &= \sum_{u\in M} [p(u)x_m(u) + q(u)y_m(u)] \\
&= \sum_{u\in M} [(x'_m - x_m)x_m + (y'_m - y_m)y_m] \text{ dropping the index} \\
&= \sum_{u\in M} \left[x'_m x_m - {x_m}^2 + y'_m y_m - {y_m}^2\right] \\
&= \sum_{u\in M} \left[x'_m x_m - {x'_m}^2 + y'_m y_m - {y'_m}^2\right] \text{ because } P'_M = P_M \\
&= \sum_{u\in M} [(x_m - x'_m)x'_m + (y_m - y'_m)y'_m].
\end{aligned}$$

Add the second and the last lines of this equality:

$$\begin{aligned}
2S_E &= \sum_{u\in M} [(x'_m - x_m)x_m + (y'_m - y_m)y_m] \\
&\quad + \sum_{u\in M} [(x_m - x'_m)x'_m + (y_m - y'_m)y'_m] \\
&= \sum_{u\in M} [(x'_m - x_m)(x_m - x'_m) + (y'_m - y_m)(y_m - y'_m)] \\
&= -\sum_{u\in M} \left[p^2(u) + q^2(u)\right] \\
&= -\|E\|^2.
\end{aligned}$$

This is the desired result.