MAP MAKING WITH NON-PHASE STABLE INTERFEROMETERS

P. N. Wilkinson
University of Manchester, Jodrell Bank, U.K.
A. C. S. Readhead
California Institute of Technology, Pasadena, U.S.A.

1. INTRODUCTION AND BACKGROUND

The ability to make reliable maps of radio sources at any available frequency and on all feasible angular scales is obviously a fundamental goal of observational radio astronomy. Unfortunately, until very recently, our map-making capability has been extremely restricted, limited in fact to those physical baselines and frequencies where phase-stable interferometers could be operated. However it is now becoming increasingly clear that phase stability is not an essential prerequisite for reliable mapping. The basic point is that whereas in principle there are an infinite number of brightness distributions which could give rise to the observed amplitudes of a set of Fourier components whose phases are unknown, in practice the *a priori* information that the required distribution is real and positive severely constrains the range of possibilities. Both Bates and his co-workers (e.g. Bates and Napier 1974) and Ross *et al*. (1978) have pointed out that the brightness distributions we are seeking are examples of 'entire' functions. In particular it is known that the real and imaginary parts of such functions are not independent (see references in Ross *et al*.) and that by studying the positions of their complex zeroes one may well be able to deduce the brightness distribution from knowing only the modulus of its Fourier transform. Unfortunately such a rigorous approach appears to be rather difficult to implement in practice. However the results of Napier and Bates (1974) have confirmed that in two dimensions reliable structure determinations can be made without phase information.

Based on this knowledge one can have some confidence in the various pragmatic approaches to the problem which have been developed. Of course the best known method of analysing fringe amplitude data is that of model fitting. In our opinion this method has obtained an unduly bad reputation for producing ambiguous results when in fact the ambiguity really arises because of the paucity of the observational data. When more extensive data have been collected, for example by Purcell on 3C147 (see Wilkinson *et al*.1977) a careful model fitting analysis assuming Gaussian components produced what has subsequently

C. van Schooneveld (ed.), Image Formation from Coherence Functions in Astronomy, 83-91.

https://doi.org/10.1017/S0252921100074765 Published online by Cambridge University Press

been confirmed to be a good representation of the source structure. In fact the basic precepts of model fitting are sensible. If, for example, gaussian components are used they are: (i) assume that the sky is basically smooth and non-negative (i.e. blank) (ii) assume that the source is basically smooth (Gaussians are infinitely differentiable) (iii) use the minimum number of components to fit the data within the noise (i.e. extract the least amount of information while remaining consistent with the data). These are just the ideas underlying the maximum entropy method! The major problem with model-fitting - like all non-linear approaches - is that it can take extravagant amounts of computer time to arrive at an acceptable solution for all but the simplest sources. However for simple sources it remains a convenient way of parametrising the visibility data. Baldwin and Warner (1978) have demonstrated another approach which seems to work well for complex sources - at least if they contain several well separated compact regions or one dominant compact feature. This method may run into difficulties if the emission is smoothly distributed however and Baldwin and Warner give little indication of the computing effort required to reach an acceptable solution. The method applied by Conway and Stannard (1975) to map the 3C273 jet also depends critically on the presence of a strong compact feature in the source.

In many cases of practical interest some phase information is always available no matter how badly disturbed are the individual phases in an interferometer network. The 'closure' phase (Jennison 1958; Rogers et al. 1974) is a linear combination of observed phases around closed loops of baselines and contains no systematic errors. The only restriction its use implies is that the fringe amplitude must always be above the noise on all baselines in the loop. For N telescopes there are at most (N-1)(N-2)/2 independent closure phase relationships; one may in fact include all baselines with fewer than this but this is not desirable. Each closure phase tells us as much about the shape of the source as does each fringe amplitude and thus it is clearly logical to include this phase information in the reconstruction. By this means one expects to reduce the residual uncertainties in amplitude only reconstructions. Chief among these uncertainties is the $\pm 180^{\circ}$ position angle ambiguity and this is resolved by the closure phase. Still however the absolute position of the source on the sky is lost. Three methods of using this information have been published (Fort and Yee 1976; Wittels et al. 1977; Readhead and Wilkinson 1978) and the results confirm that with the extra, and extremely strong, constraints on the visibility phase imposed by the closure phase one may anticipate being able to reconstruct arbitrary brightness distributions with high reliability.

A way of obtaining relative position information and of ensuring that at least the relative phases of the Fourier components are correct is to calibrate the interferometer phase with a nearby point source (e.g. Peckham 1973). Clearly then a map can be made using conventional techniques for the analysis of synthesis data. The success of this method depends entirely on the presence of a sufficiently strong,

https://doi.org/10.1017/S0252921100074765 Published online by Cambridge University Press

compact, reference object close enough to the source of interest and this angular distance is a function of interferometer baseline and observing frequency. Phase-referencing has been successfully employed at 408 and 1660 MHz on the MK IA-MK III and MK IA-Defford baselines (24 km and 127 km) using reference sources of ≳100 mJy. If the switching period is ≲10 mins relative phase deviations ≲0.2 rad can be achieved on arbitrary sources. At 408 MHz the reference source is usually in the same beam whereas at 1660 MHz angular throws of up to $\sim 2^{\circ}$ have been used. At lower and higher frequencies, especially on longer baselines, little work has yet been done but it is to be expected in the near future. However if phase-referencing does prove to be difficult, at metre and short centimetre wavelengths on VLBI baselines, the closure phase will always be available and we shall now discuss our method of using it.

2. A PRACTICAL RECONSTRUCTION METHOD USING CLOSURE PHASE

Readhead and Wilkinson (1978) have demonstrated a linear approach to the use of closure phase which can best be understood with an example. From a four telescope six baseline interferometer network one can obtain three independent closure phase relations, C, Thus:

$$C_{123} = \phi_1 + \phi_2 - \phi_3$$

$$C_{245} = \phi_2 + \phi_4 - \phi_5$$

$$C_{346} = \phi_3 + \phi_4 - \phi_6$$

where the ϕ_i are the six unknown visibility phases. If we have *a priori* knowledge which enables us to establish three of the phases independently then the other three can be calculated from these closure phase relations. This *a priori* information exists in the form of the fringe amplitudes for, as we have discussed above, much can be said about the visibility phases from the amplitudes alone. In practice simple models have proved quite adequate for providing such starting phases. The flow diagram for the iterative procedure we have adopted is shown in Fig. 1. Note that after the initial use of the closure phases the method uses exactly the same software as is contained in any standard CLEAN (Högbom 1974) package.

Fig. 2 shows the method at work on a simple source and enables us to understand why convergence to a stable solution takes place. The source in the top left hand corner is a point double with 2:1 flux ratio as would be mapped with a four telescope, phase-stable, network using CLEAN. Below it we show the first iteration obtained using the three closure phases and assuming that three of the visibility phases are zero. Obviously as all the phases are incorrect the resulting map contains spurious features and the essence of the method lies in its use of CLEAN to remove these errors. The fundamental point is that if we subtracted sources over a large area of sky we should, on Fourier

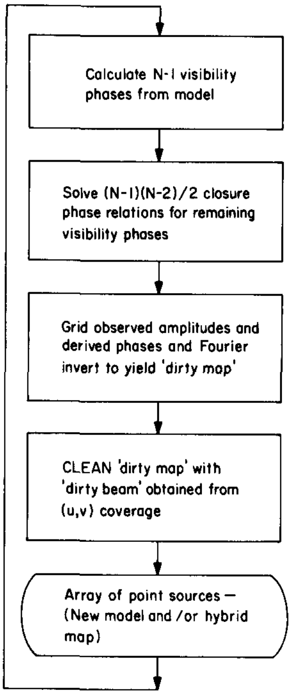

Fig. 1. Flow diagram for a linear reconstruction method using closure phase

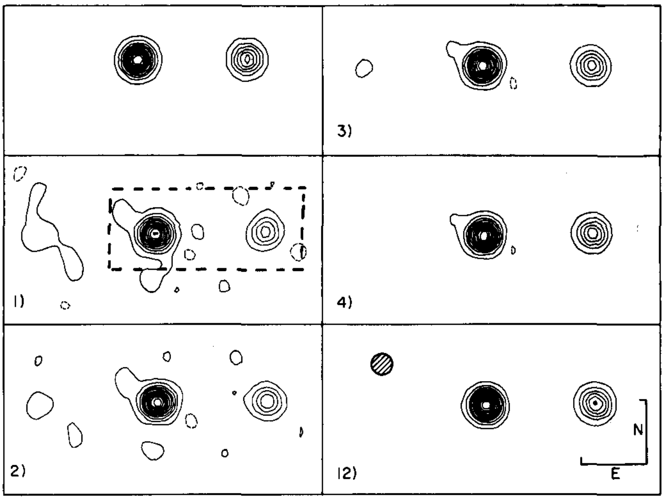

Fig. 2. Illustration of method on 2:1 point double source

https://doi.org/10.1017/S0252921100074765 Published online by Cambridge University Press

transformation, reproduce the input phases exactly since CLEAN is a linear process; in this case no convergence would occur. However in line with the original philosophy behind CLEAN we assume that most of the sky is empty and restrict our search to a 'window' around where we suspect that the source lies. We thereby systematically reject spurious components and the map therefore tends towards the true distribution as the iterations proceed. In Fig. 2 the CLEAN window is the whole area of each map and the iteration number is indicated in the bottom left hand corner. The effect of the window size on the rejection and hence the speed of convergence can be judged by noting that when we only subtracted sources within the area inside the dashed line convergence was twice as fast. More visibility data is obviously helpful and with a five telescope network and the smaller window convergence was three times as fast.

Thus for a simple source no initial subjective judgement at all is required to arrive at the correct structure. For more complicated sources it is sensible to try and start with a plausible model to reduce the number of iterations required to arrive at the correct solution. In order to convince ourselves of the efficacy of our approach on non-trivial sources we performed 'blind' tests on simulated amplitude and closure phase data to reproduce as closely as possible the real situation. We regard such tests as vital for any brightness reconstruction method where the observer can interact to aid convergence. The results of two such tests are shown in Figs. 3 and 4. In each case a) is the test source (not convolved with the restoring beam) b) is the CLEAN solution from full amplitude and phase data from a four station network and c) is the solution obtained from amplitude and closure phase. In Fig. 4(d) we show the marked improvement obtained by adding a fifth station to the array.

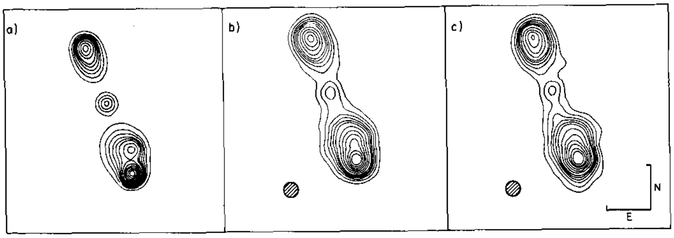

Fig. 3. 'Blind' test of method

To help the method to converge in these more complicated examples we have used three other ways of reducing the spurious features; (i) return only positive point sources for use in the initial phase calculation (ii) in the first few iterations, when many weak spurious features exist, do not clean very deeply - increase the depth as the

https://doi.org/10.1017/S0252921100074765 Published online by Cambridge University Press

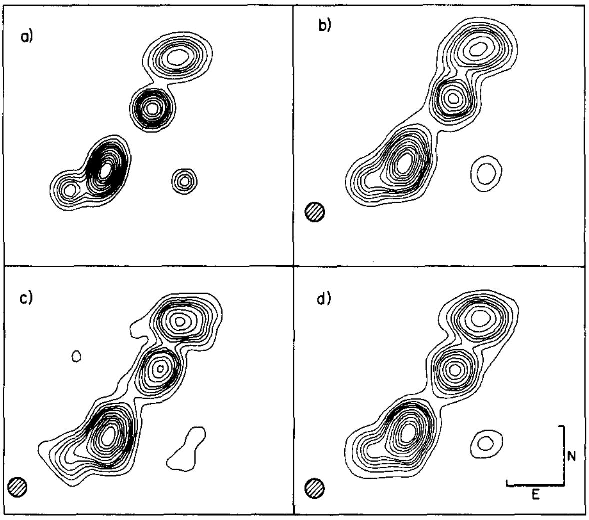

Fig. 4. 'Blind' test of method

iterations proceed (iii) starting from a stable map which is suspected to be close to the true distribution derive other maps by changing the N-1 baselines on which the phase is calculated in the initial step. If the first map is the true distribution then these subsequent maps will all be the same to within the noise. If it contains minor errors then the other maps will be different and the error regions tend to move to new locations in each map. Thus a better map for use as the input to the next iteration can be obtained by adding these trial maps together, for the correct areas in them add coherently and are thereby enhanced with respect to the spurious ones.

It is important to give some idea of the speed of this process. In the examples shown in Figs. 3 and 4 (array size about 30 x 30) each iteration (one job) took typically 1 minute CPU time on an IBM 370/158; this involved cleaning and returning $\lesssim$200 point sources per iteration. On average we allowed ourselves ∿20 jobs, trying various different approaches, before comparing our preferred solution with the test source.

A particularly neat application of this method arises when the source contains a bright compact region which can be used as a phase reference. In this case convergence is extremely rapid (e.g. Wilkinson et al. 1977). Fig. 5 shows a test on simulated four station data on Purcell's model of 3C147. Here again a) is the model b) is the CLEAN solution and c) is the solution using closure phase. Two things are noteworthy: first the good agreement between the full and the closure phase solutions and second the fact that one can clearly reproduce most of the features in such an extended source out to ∿20 beam diameters. This test gives us confidence that our actual maps of 3C147 (Wilkinson

https://doi.org/10.1017/S0252921100074765 Published online by Cambridge University Press

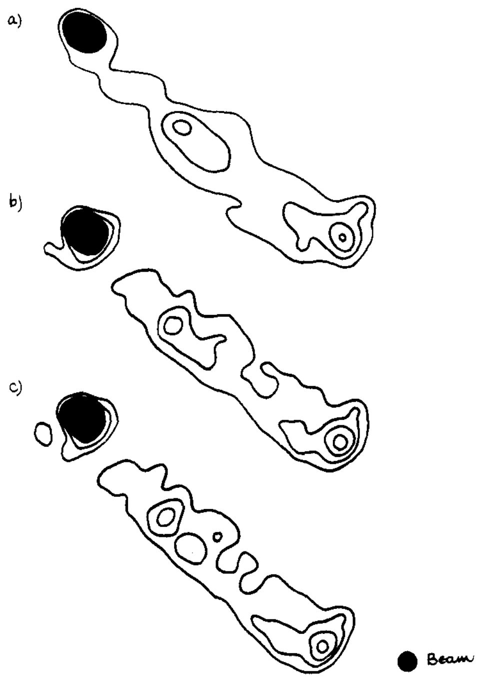

Fig. 5. Test on Purcell's model of 3C147

et al.1977) accurately represent the compact features in the source.

3. CONCLUSIONS

The obvious way to achieve a degree of phase stability in an interferometer network is to employ phase-referencing techniques. Not only is positional information collected but the ability to make long coherent integrations means that weak sources can be studied. However even if this is not possible it is clear that one can still obtain data which are sufficient for determining the structures of the stronger sources. The closure phase is particularly easy and error-free data to collect and indeed there may well be cases where individual relative phases can only be determined to say $\pm$ 30° r.m.s. while the more restricted but more accurate closure phase (say $\pm$ 5° r.m.s. on a strong source) can lead to a more reliable map.

ACKNOWLEDGEMENTS

We thank M.H. Cohen for his enthusiastic support and D.H. Rogstad for many discussions regarding his CLEAN package. PNW gratefully acknowledges the receipt of the Royal Society Weis Fellowship. ACSR acknowledges support via NSF grant no. AST 77-00247.

REFERENCES

Baldwin, J. E. and Warner, P. J.: 1978, Mon.Not.R.astr.Soc., 182, 411.
Fort, D. N. and Yee, H. K. C.: 1976, Astron.Astrophys., 50, 19.
Högbom, J. A.: 1974, Astron.Astrophys.Suppl., 15, 417.
Jennison, R.C.: 1958, Mon.Roy.astr.Soc., 118, 276.
Napier, P.J. and Bates, R. H. T.: 1974, Astron.Astrophys.Suppl., 15, 427.
Peckham, R. J.: 1973, Mon.Not.Roy.astr.Soc., 165, 25.
Readhead, A. C. S. and Wilkinson, P. N.: 1978, Astrophys.J., 223, 25.
Rogers, A. E. E., et al.: 1974, Astrophys.J., 193, 293.
Ross, G., et al.: 1978, Proc.Roy.Soc., 360, 25.
Wilkinson, P.N. et al.: 1977, Nature, 269, 764.
Wittels, J. J., et al.: 1976, Astron.J., 81, 933.

DISCUSSION

Comment T.A. CLARK
I concur with Wilkinson's comments that model-fitting has been given "bad press". When the model fitting is approached with statistical procedures, either as described by Cotton or below, they can be considered quite similar to maximum likelihood estimation.
An additional comment on phase closure analysis. A slightly different approach taken by L.K. Hutton and me was to form a two dimensional "bed-of-nails" delta-function representation of the source. By direct Fourier Transform, the entire source was simultaneously adjusted in a weighted non-linear least squares analysis to minimize the differences between the actual amplitude and closure phase data and that predicted by the model. An efficient gradient-search least-squares algorithm made this quite computationally efficient. Initial constraints consisted of a maximum window size determined by the shortest baselines and/or a priori source knowledge, source grid spacing determined by maximum resolution, and a positive flux constraint on each grid point. A starting point that was found necessary was to begin with one point non zero to "nail down" the final map. Generally the final map "grew" with the strongest component at the position of this "nail". The final delta-function map was convolved with a gaussian beam similar to the point source response determined from the U-V coverage.

After the presentation by Wilkinson and Readhead a general discussion developed on the subjects of closure phase and positivity. Several of the previous speakers answered questions and comments.

Comment G. POOLEY
Several people have mentioned using a constraint that the data are positive; in measuring polarization data you cannot use this constraint.

Question B.C. CLARK
In the case where one is approaching the signal-to-noise limit, one presumably wishes to prevent, or at least discourage, the phases measured on the baselines where the amplitudes are low, and signal-to-noise ratios bad, from entering into the closure phases, and hence affecting the solved phases on other baselines. Is there a systematic procedure for handling this?
Reply J.E. BALDWIN
Our analysis of 5C7 does not bear on the closure phase problem but we did indeed find it was essential to remove amplitude data where it fell below 2σ. Otherwise, this noisy data combined with the phase of the trial conspired to put extra flux density into the trial source in the next hybrid map. The 2σ limit was an arbitrary decision and there may be some more proper procedure.

Comment by D.B. SHAFFER to J.P. HAMAKER
If you are willing to give up all positional information, phase closure will give you the correct shape map (but not its position).
Reply J.P. HAMAKER
Let me answer your remark in terms of an array which produces not only closure phase observations but also contains enough redundancy to solve for all instrumental phase errors. What we can obtain then are one-dimensional strip scans of perfect shape but unknown position, as you suggest. To align these in a two-dimensional map, however, one does need the relative position of these scans, if one wants to avoid the radial streaks which I showed before. These relative positions are not provided by the closure phase or redundancy information and one must make an educated guess of one sort of another. Part of my short paper (J.P. Hamaker: "Kneading", this volume) has to do with precisely this point. I shall try to think your remark over and presently make some additional comments.

Comment R. SRAMEK
When discussing closure phase, it is best to think of the phase of a single baseline interferometer to consist of 1) a baseline dependent part ϕ_{12} resulting from e.g. source structure, correlator problems, etc. and 2) an antenna dependent part $\phi_1 - \phi_2$ which is due to e.g. front-end problems, the atmosphere, etc. The closure phase is independent of the antenna dependent phase, no matter what the antenna geometry, VLBI or connected antennas, or the origin of ϕ_1 and ϕ_2.