

APERTURE SYNTHESIS WITH A NON-REGULAR DISTRIBUTION OF INTERFEROMETER BASELINES

J.A. HÖGBOM

Stockholm Observatory, Sweden

In high-resolution radio interferometry it is often impossible for practical reasons to arrange for the measured baselines to be regularly distributed. The standard Fourier inversion methods may then produce maps which are seriously confused by the effects of the prominent and extended sidelobe patterns of the corresponding synthesized beam. Some methods which have been proposed for avoiding these difficulties are discussed. In particular, the procedure CLEAN is described in some detail. This has been successfully applied to measurements taken with several different radio telescopes and appears to be the best method available at the time of writing.

1. INTRODUCTION

Aperture synthesis measurements are usually made at a set of interferometer spacings and orientations that form a regular pattern in the baseline (u, v) diagram. Such a regular coverage has many practical advantages both in connection with the formal synthesis calculations and the later astronomical interpretation of the synthesis map. However, there are occasions when irregularities in the baseline coverage cannot be avoided. Interference or malfunctioning of some part of the equipment can make it necessary to reject certain portions of the measurements and this will leave gaps in an otherwise regularly covered u, v plane. The gaps give rise to undesirable sidelobes in the synthesized beam, making the synthesis map difficult or impossible to interpret. Similar problems arise when the u, v plane has been covered by a coarse grid of measurements as this will give rise to prominent grating responses in the synthesized beam pattern.

It may, in certain cases, be impossible (or impractical) to arrange for the interferometer measurements to fall on a regular grid in the baseline (u, v) diagram. This is the case for measurements taken with instruments such as the Caltech and Green Bank interferometers and, in general, for measurements that involve large interferometer spacings. Interferometers can be operated over spacings up to the full diameter of the Earth and from space vehicles and one shall ultimately want to use such measurements in a systematic way to synthesize very high-resolution maps of small diameter sources. Occultations of radio sources by the Moon give rise to similar problems. A few occultations of the same source (or one occultation measured at several observatories) will deliver a number of strip scans at a non-regular set of position angles. These strip scans are equivalent to a u, v coverage along radii at these same position angles.

Knowledge of what can and what cannot be achieved with a non-regular coverage of the u, v plane is obviously very important in connection with the design of future high-resolution synthesis instruments. Already existing instruments can also be used more efficiently if the formal requirement of a particular regular baseline coverage can be dropped.

Problems involved in the reduction of non-regularly spaced interferometer measurements for which amplitude and phase information is available are discussed in this paper. The aim shall be to produce a map that is equivalent – in so far as this is possible – to that which we would have obtained by the normal synthesis calculations if the measurements had in fact covered the relevant region of the u, v plane in a uniform manner. One might object that such a procedure cannot exist. According to a theorem in Fourier analysis, this region contains just as many independent measurements as there are independent directions in the observed field as seen with the synthesized beam. A smaller number of measurements cannot be sufficient to specify the map completely. The confusion caused by sidelobes and grating responses should therefore be regarded as an unavoidable expression of the fact that one cannot solve for more unknowns than there are equations. However,

significant deflections from zero intensity will usually occupy only a small fraction of a high-resolution synthesis map. This becomes obvious when one compares the brightness temperature sensitivity of a high-resolution radio telescope with typical values of the sky brightness temperature and present knowledge about radio source statistics. A typical high-resolution map can usually be described by saying that it is *empty* but for certain exceptions which can be specified in a table. Clearly, the number of measurements must be at least as large as the number of items in this table, but this will usually be a fraction of the total number of independent synthesized beam directions in the observed field.

2. CONSEQUENCES OF A NON-REGULAR DISTRIBUTION OF THE MEASURED BASELINES

The theory of aperture synthesis does not in itself require that the u, v plane be covered according to a regular pattern, but it turns out that regular patterns will in general produce maps that are relatively easy to interpret. In its basic form, the theory states that the synthesis map obtained by taking the Fourier transform of the measurements $W(u, v)$ is proportional to the convolution of the true brightness distribution T_b (enveloped by the primary beam P of the antenna combination) with the synthesized beam G . If all the baselines can be reduced to one plane, the u, v plane, which is at right angles to the center of the observed field (we shall discuss this assumption later), then this Fourier relation takes the following form:

$$\{P(\ell, m) T_b(\ell, m)\} * G(\ell, m) \propto \iint_{-\infty}^{+\infty} W(u, v) g(u, v) \exp[i 2\pi(u\ell + vm)] du dv \quad (1)$$

ℓ and m are coordinates in the sky (directional cosines with respect to the u and v axes respectively), $i = \sqrt{-1}$, and $g(u, v)$ is the function by which we choose to weight the (complex) interferometer measurements $W(u, v)$ in the transform calculations. u and v are expressed in wavelengths. The synthesized beam equals the suitably normalized Fourier transform of this weight function:

$$G(\ell, m) = \text{const} \iint_{-\infty}^{+\infty} g(u, v) \exp[i 2\pi(u\ell + vm)] du dv \quad (2)$$

In order to calculate the Fourier integral in Eq. (1), the product $W \cdot g$ must be known for all baselines (u, v) . However, the correlation function $W(u, v)$ is only available for the finite number of baselines at which it has been measured. If there is no other information available, then it is reasonable from the point of view of information theory to use the completely unknown values of W elsewhere with the weight $g=0$. The weight function $g(u, v)$ will then be very irregular (finite at all measured points and zero over the rest of the u, v plane) and, as a consequence of this, the synthesized beam will be accompanied by an extended pattern of undesirable sidelobes (figure 1 and 2). A synthesis map calculated directly according to Eq. (1) can be so disturbed by the effects of these sidelobes that it becomes essentially useless (figure 3a). I shall use the terms “dirty beam” and “dirty map” to describe this particular synthesized beam and synthesis map respectively. If we choose to give the same weight $g=1$ to all the measured points, then the dirty map is identical with the so called principal solution.

The principal solution dirty map is one member of an infinite family of solutions that all agree with the available measurements. Other solutions can be produced by allocating any finite values to the correlation function W at the not measured baselines (u, v) and then treating these allocated values as if they were additional real measurements. The principal solution dirty map has the unique property, if no other information is available, of being the solution with the lowest probable rms deviation from the true sky brightness distribution. The absence of other information here means that, as far as we know, the true brightness distribution could be any random function of the direction in the sky.

We are, however, discussing a different case. The true brightness temperature will be below the sensitivity level over most of the sky and only a fraction of the observed field is expected to contain significant departures

from zero intensity. This in itself is equivalent to a great amount of information which could be added to the information produced by the interferometer measurements. The problem is how to incorporate this *a priori* information into the procedure by which we calculate the sky brightness distribution over the observed field.

3. VARIOUS WAYS OF INCORPORATING A PRIORI KNOWLEDGE ABOUT THE SKY BRIGHTNESS DISTRIBUTION

The undesirable sidelobe effects are caused by gaps in the u, v coverage; consequently it would seem natural to bridge these gaps by means of some simple interpolation procedure using the surrounding measured points. However, this can only improve the situation if neighbouring points are correlated and if we know the exact nature of this correlation. For instance, if the position of the source is known, then we can perform the interpolation in such a way that the picture of this source is less, or not at all, disturbed by its own sidelobes. The common method of dividing the u, v plane into cells and letting all the measurements within each individual cell be represented by their vector average placed at the center of the cell, can in this way improve the map of a source situated at the instrumental phase center. Improved results can be obtained with more sophisticated interpolation schemes (see e.g. Burns and Yao 1970). The disturbances caused by the sidelobes from all other sources in the field will, however, remain or even become more pronounced. A simple interpolation scheme can indeed improve the situation but only in certain special cases. The gaussian convolution applied to the Westerbork measurements (Brouw 1971, van Someren Gréve 1973) and the corresponding procedures used at other observatories are examples of interpolation procedures. Their great merit is that they transform the measurements to the rectangular grid of points that is required for the Fast Fourier transform algorithm. They are not solutions to the problem of how to eliminate the adverse effects of gaps in the u, v coverage. Grating rings and other sidelobe disturbances remain and some extra confusion may be introduced by the aliasing caused by the imposed regular grid structure.

This first method used by radio astronomers for reducing non-regularly spaced data was the method of *model fitting*. If, for instance, we have measured three numbers: the total flux and the amplitude and phase at one $E-W$ interferometer spacing, then we can solve for the characteristics of a three parameter source model. If the source is a uniform circular disc – that is the model – then we can calculate three parameters, e.g. the source diameter, the disc brightness temperature and the right ascension position coordinate. The trouble is that the source may not at all look like a circular disc and the problem of choosing suitable models and solving the resulting set of equations becomes increasingly difficult for larger numbers of measurements and thus potential model parameters. In practice, model fitting sometimes deteriorates to a mixture of wishful thinking and root mean square adjustments.

The maximum entropy method described by Ables (1973) is based on a mathematical principle that determines which solution we should choose among the infinite family of possible solutions all agreeing with the available interferometer measurements. This principle states that the selected solution should be that which contains the least amount of information (the “maximum entropy”) according to a definition of information analogous to that which enters the concept of entropy in thermodynamics. In short, the minimum information solution is also that which is likely to contain the least amount of false information. This principle, coupled with the requirement that the map be compatible with the available interferometer measurements, defines a set of equations which can in principle be solved for any distribution of measured baselines (u, v). However, the equations for irregularly spaced measurements in two dimensions appear to require excessive computing efforts and it is not yet clear how useful the method will be for the problems discussed in this paper.

An obvious property of real brightness distributions is that they are exclusively positive. Schell (1965) and Biraud (1969) point out that the Fourier transform of a positive function must itself be an autocorrelation function and that this property can be used to compute finite values for spacings which are greater than those contained in the antenna system and so increase the angular resolution. The authors compute these

values according to different criteria, but the main aim is to achieve a physically acceptable (i.e. non-negative) solution which involves a minimum number of not measured spacings. It should be possible to use these methods also to fill in gaps in the baseline coverage. A different procedure has been described by Högbom (1969). The “dirty beam” described earlier contains negative as well as positive sidelobes and the dirty map will then also in general contain regions of negative intensities. The real brightness distribution, however, must be zero or positive so we can reduce the differences between the map and the true sky by replacing all negative values by zeroes. The Fourier transform of the map will then no longer agree with the known values of W at the measured baselines. The agreement is restored by adding to the map Fourier components that make up for these differences. This will produce new areas of negative intensities which are treated in the same way and the process is repeated until the changes in the map from one iteration to the next become insignificant. The method works well in many cases and will even display certain features on the map at an increased resolution. The unspecified beam shape can, however, cause trouble during the iterations and also make it difficult to judge the reality of details on the final map.

4. MAIN PRINCIPLES OF THE PROCEDURE CLEAN

The procedure that appears to work best at the time of writing has become known as “CLEAN”. It is an iterative procedure that operates in the map plane and which uses the known shape of the dirty beam to distinguish between real structure and sidelobe disturbances on the dirty map. Let us assume that the map in figure 2 has in fact been obtained by a direct Fourier transform of a set of ideal noiseless interferometer measurements of a particular piece of sky. The map certainly has a very “dirty” and unsatisfactory appearance but, if we find that it is identical in every individual detail to the expected dirty beam, then the interpretation becomes clear: there is a point source at the position of the maximum deflection on the map. We can formally confirm this conclusion by subtracting out a full dirty beam pattern centered at this position and with the corresponding amplitude: there should then be nothing left. If we now return to this empty map a clean beam pattern – i.e. the ideal main beam of a similar shape but without the disturbing sidelobes – at this position and with the same amplitude, the result will be the same as that which we would have obtained with normal synthesis calculations if the relevant region in the u,v plane had been completely covered with measurements.

The single point source is a trivial case but it illustrates the main principle of the procedure CLEAN. Observe that the interpretation of the dirty map in terms of a point source in an otherwise empty piece of sky is not based on synthesis theory. Other information was used to reject an infinite number of other possible interpretations as being too unlikely to be taken seriously. The real sky may, for instance, look exactly like the complex structure that was interpreted as sidelobes and be empty at the position of the main maximum; this latter would then in fact be a systematic pile-up of sidelobes from the many weaker but real details in the field. In rotational synthesis (supersynthesis) we could in a similar way argue that one of the grating rings is a real feature while the other rings *and* the central maximum are grating responses caused by this ring structure in the sky. Such interpretations are rejected partly because the structures themselves look artificial, but mainly because the real sky brightness distribution should have no special relation to the particular sidelobe pattern produced by the available set of interferometer baselines. A significant correlation between the dirty map (DM) and the dirty beam (DB) will most likely be due to a source or source component at the position of the maximum correlation. The correlation function is given by the convolution $DM * DB$. The Fourier transform equations (1) and (2) can be written symbolically:

$$\begin{aligned} DM &\stackrel{\text{FT}}{\longleftarrow} W \cdot g \\ DB &\stackrel{\text{FT}}{\longleftarrow} g \end{aligned} \quad (3)$$

and it follows from the convolution theorem that

$$\text{DM} * \text{DB} \stackrel{\text{FT}}{\leftarrow} W \cdot g^2 \quad (4)$$

If we have chosen the weights g to be unity at all measured points and zero everywhere else, then $W \cdot g^2$ will be identical to $W \cdot g$. Thus $\text{DM} * \text{DB}$ must also be identical to DM and there is no need to perform the convolution: the dirty map is itself the correlation function that we want. The maximum deflection from zero on the dirty map is also the place at which it is most strongly correlated with the full dirty beam pattern. There may be reasons for using different weights g for different measurements (e.g. matching the main maximum of the DB to the shape of the CB). The DM will then not be equal to $\text{DM} * \text{DB}$ but to $\text{DM}' * \text{DB}'$ where the primes indicate the map and the beam that would have been obtained with the weights $g' = \sqrt{g}$. This will only affect the procedure CLEAN in so far that the measurements, as expected, will exert their influence on the solution with different weights g . The procedure CLEAN can now be described in terms of a few simple steps.

- I. Compute the DM and the DB by standard Fourier inversion methods. The weights g should be chosen in such a way that the main lobe of the DB becomes a good fit to the selected CB.
- II. Subtract over the whole map a DB pattern which is centered at the point at which the DM has its maximum absolute value $|I_o|$ and which is normalized to γI_o at the beam center. The fraction γ will be called the *loop gain*.
- III. Repeat step II, each time replacing the DM by the remaining map from the previous iteration. Stop the iterations when the current value of I_o is no longer significant in view of the general noise level on the map.
- IV. Return to the final remaining map all those components that were removed in step II, but do this in the form of clean beams with the appropriate positions and amplitudes.

The individual iteration will represent a correct interpretation of the situation if the value I_o did in fact contain a contribution with an amplitude of at least γI_o from some real feature. Clearly, for optimum safety, one should choose an infinitesimally small value for the loop gain γ . The process could then be visualized as one in which a ceiling to the map absolute value was lowered while continuously shaving off whatever protrudes through this ceiling along with the associated patterns of dirty sidelobes. However, if a certain deflection is significant in the sense that it indicates the presence of some real feature, then it should be safe to assume that this feature makes a substantial contribution to this deflection. One could, of course, make a full probability analysis in each individual case but in practice it is found that there is little further improvement when the loop gain is reduced below $\gamma = 1/2$. Even $\gamma = 1$ will give perfectly good results in the simplest cases. Figure 3 shows a dirty map produced from measurements with the Green Bank interferometer and the end product of using the CLEAN procedure with 1, 2 and 6 iterations and a loop gain of $\gamma = 1$. The last map could be described as empty but for certain exceptions which can be specified in a table containing less than 30 numbers. This is to be compared with the 100 measured quantities, i.e. the amplitudes and the phases at 50 interferometer baselines (figure 1). Polarization measurements can be handled with a complex version of CLEAN or by processing the Q and U maps separately; a survey of the structure and polarization of 78 radio sources has been made using this technique (Högbom and Carlsson 1973).

The CLEAN iterations can be programmed in an efficient manner. In a first step we go through the whole array looking for the element with the largest absolute value $|I_o|$. At the same time we compute the sum of all absolute values encountered. Dividing this sum by the total number of elements we get the average absolute value. The decision as to whether the deflection I_o is significant can now be based on the criterion that $|I_o|$ exceeds this average value by some specified factor of the order of 4. This full survey of the map deflections may not be necessary at every iteration. It can be enough to make note of other regions containing high absolute values and then to restrict the search for the maximum deflection to these regions during the next few iterations. The subtraction of a dirty beam pattern component over the whole map with a specified loop gain γ will require one multiplication per array element. However, the loop gain need not be a fixed quantity. It should be

less than about $1/2$ and also should not be too small because this will mean an unnecessarily large number of iterations. If we allow the loop gain to vary within a factor of two between, say, $0.25 < \gamma \leq 0.5$, then the multiplications can be replaced by subtractions. Let the dirty map and dirty beam be normalized so that their maxima are 4 and 1 respectively. Now, a direct subtraction in the first iteration corresponds to $\gamma = 0.25$. The maxima in the following iterations will be lower and hence the loop gain correspondingly higher. When a maximum has an absolute value of less than 2 which would imply $\gamma > 0.5$, we double the map – at the expense of one addition per element – before performing the subtraction which will then again correspond to a loop gain $\gamma < 0.5$. Returning the clean beams after the last iteration is no problem because they only affect small areas of the map.

5. LIMITATIONS TO THE PROCEDURE CLEAN

A non-linear procedure such as CLEAN is of doubtful value in practice unless the user can judge whether or not a final cleaned map is reliable. Clearly, CLEAN cannot deliver a correct map if the source is in fact so complex that it cannot be described adequately with less parameters than there are measurements. In such a case one finds after a few iterations that the next maximum deviation I_0 will no longer be significantly greater than the general confusion level on the map (i.e. indicate a more than 50% probability that it is partly due to some real feature at this position). The iterations will therefore be interrupted at an early stage when there is still a great amount of sidelobe confusion noise over the map. After returning the clean patterns of the first few iterations, this noise remains and will mask the detailed structure of the source in much the same way as would an excessive amount of receiver noise. Clearly, noisy data will have the same effect: CLEAN will only pass through a few iterations but this should at least give some improvement. In both cases it may be possible to continue the iterations if the remaining map and the dirty beam are first smoothed to a lower resolution. This is especially useful in bringing out weak extended distributions that would otherwise be lost in the noise. Purely random data will result in no or perhaps a single iteration (depending upon how the 50% criterion has been formulated). At the other extreme – complete u,v coverage and zero noise – CLEAN may pass through any number of iterations, but the clean beam is then identical to the “dirty beam” and the net effect is, as it should be, equal to zero.

Figure 4 shows cleaned maps of the source 3C 130 prepared from measurements at 200, 100, 50 and 30 baselines taken with the Westerbork synthesis telescope. The first map is in itself a good illustration of how the CLEAN procedure can help towards an efficient use of a synthesis telescope. The minimum observing time for two-dimensional synthesis with this instrument using the standard inversion methods is 12 h. The 200 baselines (10 h angles, each with 20 spacings) required only $3\frac{1}{3}$ h of instrument time.

The map based on 100 baselines is very similar. The 50 baseline map is seriously disturbed but still correct in its main features. 30 baselines is obviously not enough, and the map is dominated by a prominent confusion noise.

The general run of the zero-level contour in figure 4a shows that the map is free from the usual depression caused by the missing zero spacing in the interferometer measurements. As far as the procedure CLEAN is concerned, this depression is just one part of the dirty beam sidelobe pattern which it attempts to eliminate. The procedure cannot detect large extended distributions that are completely resolved at all the measured baselines, but it will remove the depression caused by those source components that it detects during the iterations.

An important limitation to the use of CLEAN in its present form is related to the fact that the actual distribution of baselines will normally be three-dimensional, each baseline having some vector component w wavelengths parallel to the direction of the observed piece of sky and hence at right angles to the u,v plane. The measurements W and the weights g will be functions of all three variables (u,v,w) and the Fourier

integrals in Eqs. (1) and (2) should strictly be expressed in their three-dimensional form. The exponential factor in these equations then becomes:

$$\exp [i 2\pi (u\ell + vm + wn)] \quad (5)$$

The three directional cosines ℓ , m and n are related by

$$\ell^2 + m^2 + n^2 = 1 \quad (6)$$

Write:

$$\begin{aligned} wn &= w (1 - \ell^2 - m^2)^{1/2} \\ &\simeq w - 1/2 w \rho^2 \quad (\rho^2 = \ell^2 + m^2) \end{aligned} \quad (7)$$

then the exponential factor (5) can be written:

$$\exp [i 2\pi w] \cdot \exp [i 2\pi (u\ell + vm - 1/2 w \rho^2)] \quad (8)$$

The first factor can be formally incorporated in the weight function $g(u, v, w)$; it is the analytical expression that corresponds to reducing the phases at the measured baselines to their projections on the u, v plane. The term $1/2 w \rho^2$ in the argument expresses the fact that this projection looks different as seen from directions away from the center of the field (i.e. for $\rho \neq 0$). The two-dimensional formulae (1) and (2) will be correct if this term is always $\ll 1$: The effect of having larger values within the observed field of sky is to make the sidelobe structure of the synthesized beam noticeably different for sources situated in different parts of the synthesis map. Consequently, cleaning with a fixed dirty beam shape will have a positive influence on the map only when all operations are restricted to an area centered at the field (phase) center and with a radius ρ_{\max} given by, approximately

$$w_{\max} \cdot \rho_{\max}^2 \approx 1 \quad (9)$$

Thus, if $w_{\max} \cdot \lambda = 10$ km and $\lambda = 6$ cm, then CLEAN should not be used to correct for disturbances caused by sources outside a circle of radius $\rho_{\max} \simeq 0.00245$ radians, or about 8 arcmin. There are ways by which one can take this variable synthesized beam pattern into account but they will be at the expense of a considerably increased computing complexity. However, two effects tend to make this limitation less important. First, the interferometer measurements will usually be made while the antennas follow the observed field for some time. The u, v plane will therefore be measured along a set of lines (or paths) rather than at a set of discrete points and this will make the far sidelobe level approximately inversely proportional to the square root of the distance from the beam center. A finite bandwidth will cause a similar reduction of the distant sidelobe level. A strict statistical study of these effects in the case of randomly distributed baselines has not been made, but it seems that sources outside the area $\rho \leq \rho_{\max}$ should in general have a very reduced disturbing effect on the cleaned central part of the map.

REFERENCES

- Ables, J.: 1973, *Astron. Astrophys. Suppl.* (this issue).
 Biraud, Y.: 1969, *Astron. Astrophys.* **1**, 124.
 Brouw, W.N.: 1971, *Thesis*, University of Leiden.
 Burns, W.R. and Yao, S.S.: 1970, *Astron. Astrophys.* **6**, 481.
 Högbom, J.A.: 1969, *Rep. URSI 16th Gen. Assembly*, Ottawa.
 Högbom, J.A. and Carlsson, I.: 1973, *Astron. Astrophys.* (in preparation).
 Schell, A.C.: 1965, *Radio and Electronic Engineer* 21.
 Someren Greve, H. van: 1973, *Astron. Astrophys. Suppl.* (this issue).

J.A. Högbom

Stockholm Observatory
 S - 13300 Saltsjöbaden, Sweden

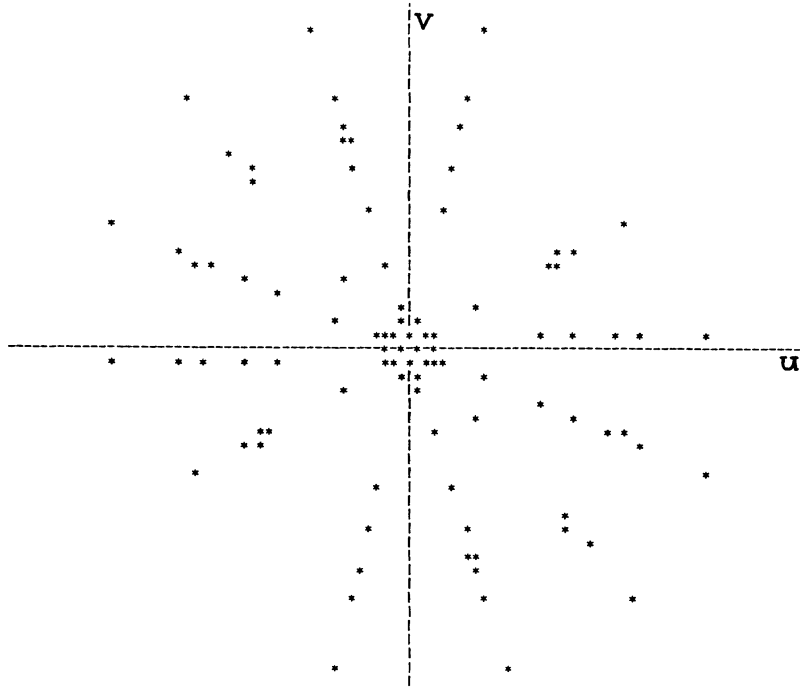


Figure 1 Example of a baseline coverage obtained with the Green Bank Interferometer. The measured points have been mirrored through the origin to give a better impression of the overall structure.

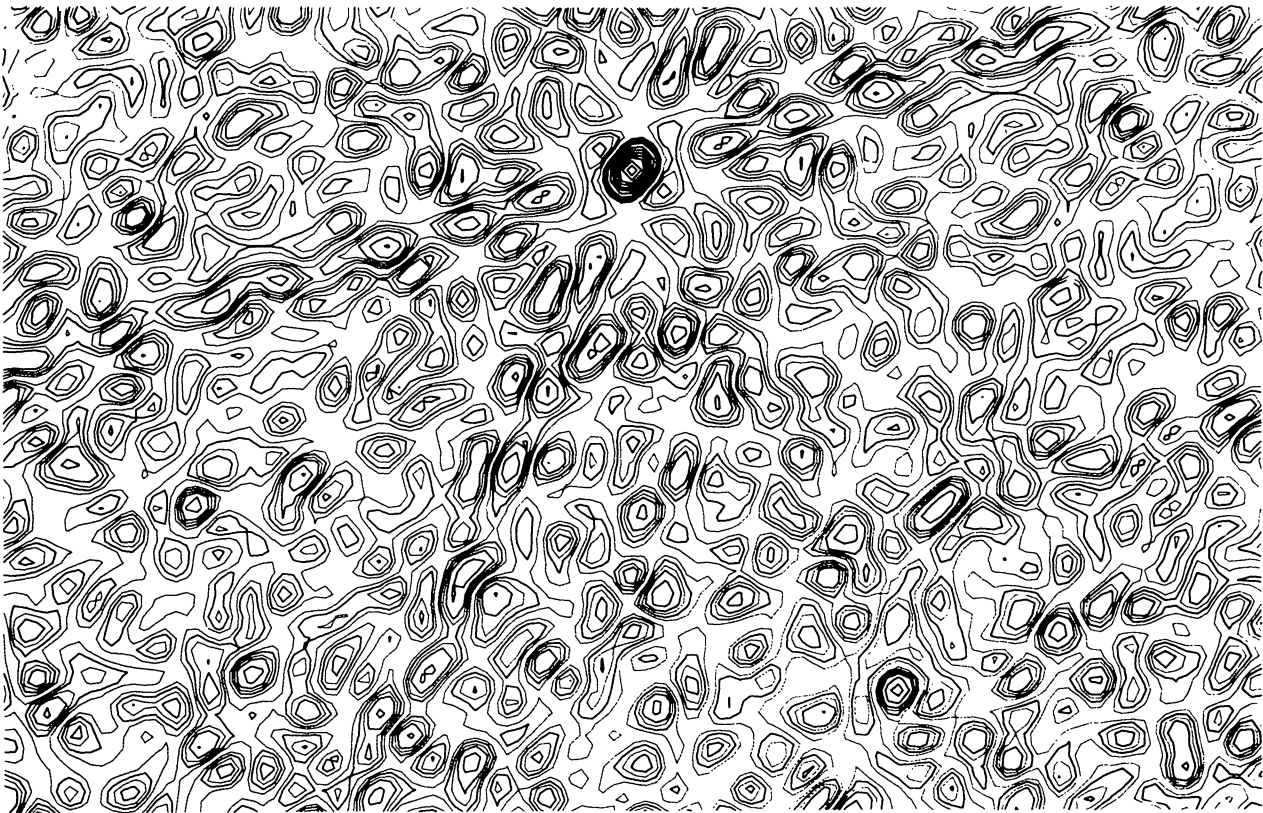


Figure 2 The “dirty beam” corresponding to the u,v coverage shown in figure 1. Contours are drawn at 5, 10, 15, 20, 30 etc. % of the beam maximum (top center). No distinction has been made in the figure between positive and negative contours; half of the sidelobes are in fact negative.

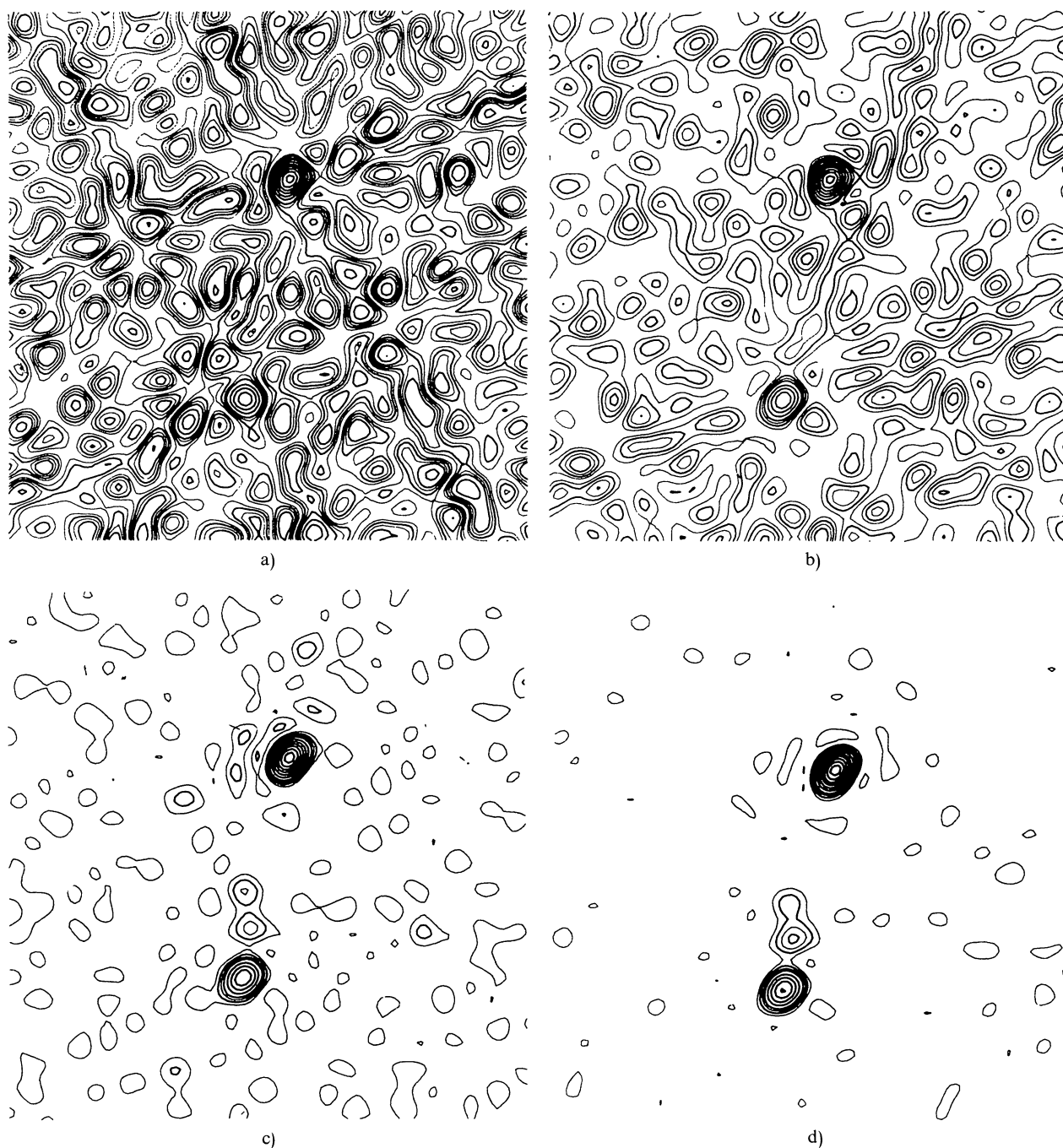


Figure 3 Illustrating the effect of the CLEAN procedure on measurements at 2695 MHz of the radio source 3C 244.1 taken with the u,v coverage shown in figure 1. Contours are drawn at the same intensity ratios as in figure 2. a) the “dirty map” b) cleaned map after one iteration with the loop gain $\gamma=1$ and subsequent return of the clean beam, c) same, but after two iterations and the return of the two clean beams. The north preceding component is extended and there are some weaker components present, d) 6 iterations. The further improvement here is due to the cleaning of the sidelobes from the less intense features that remain after the two first iterations.

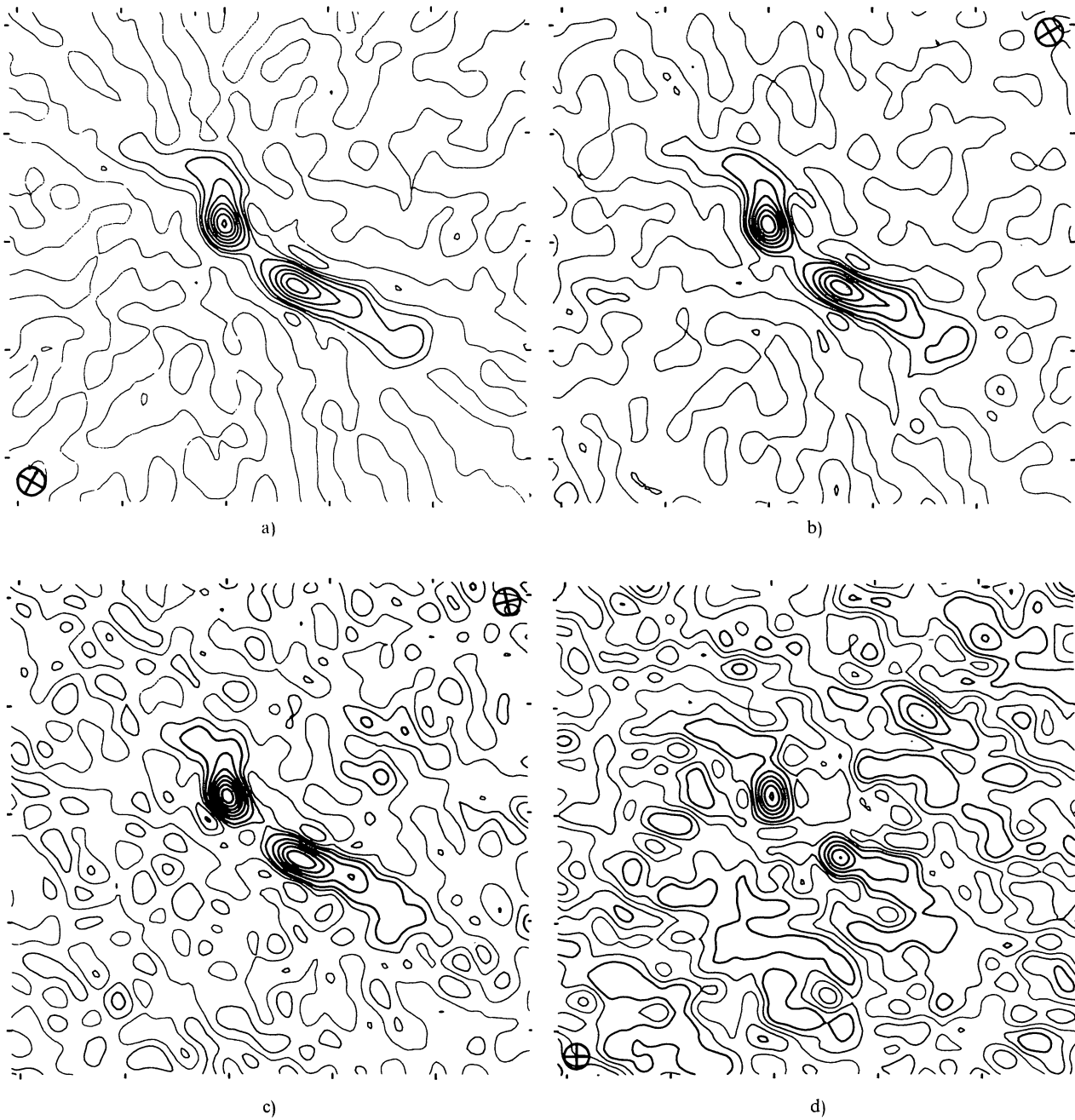


Figure 4 Illustrates the effects of CLEAN when the number of measurements used for the calculations is reduced successively until it is no longer sufficient to define the source structure uniquely. The cleaned maps have been prepared from 1415 MHz measurements with the Westerbork synthesis telescope taken at a) 200, b) 100, c) 50 and d) 30 baselines. Contours are plotted at 0, 1/2, 1, 2, 3, etc. units; no distinction is made in the figure between positive and negative contours.