Synthesis Imaging in Radio Astronomy II
ASP Conference Series, Vol. 180, 1999
G. B. Taylor, C. L. Carilli, and R. A. Perley (eds.)

8. Deconvolution

Tim Cornwell
National Radio Astronomy Observatory, Socorro, NM 87801, U.S.A.

Robert Braun
NFRA, Dwingeloo, The Netherlands

Daniel S. Briggs
NCSA, Champaign-Urbana, IL 61801, U.S.A.

Abstract. This lecture describes how the visibility samples collected by an interferometric array can be used to produce a high quality image of the sky. In contrast to the linear methods of Lecture 7, these methods are all non-linear, and must create estimates of the visibility function at positions in the Fourier plane where it is not measured. The two most common algorithms used, CLEAN and MEM, are discussed in detail, with several variations and less common algorithms mentioned briefly. An example comparison between CLEAN and MEM is given on a simulated VLBA observation.

1. Deconvolution

As described in Lectures 1 and 2, an interferometric array provides samples of the complex visibility function of the source at various points in the (u, v) plane. Under various approximations, which are valid for a sufficiently small source in an otherwise blank region of sky (see Lecture 1, Sec. 4.2 and Lecture 2, Sec. 6), the visibility function $V(u, v)$ is related to the source intensity distribution $I(l, m)$ (multiplied by the primary beam of the array elements) by a two-dimensional Fourier transform:

$$V(u,v) = \int\int_S I(l,m) e^{-2\pi i(ul+vm)}\, dl\, dm\,, \tag{8–1}$$

where S denotes taking the integral over the whole sky, as in Equation 2–5.

Since only a finite number of noisy samples of the visibility function are measured in practice, $I(l, m)$ itself cannot be recovered directly. Either a model with a finite number of parameters, or some stable non-parametric approach, must be used to estimate $I(l, m)$. A convenient general purpose model $\widehat{I}$ of the source intensity that is capable of representing all the visibility data consists of a two-dimensional grid of δ-functions with strengths $\widehat{I}(p\Delta l, q\Delta m)$, where Δl and Δm are the separations of the grid elements in the two orthogonal sky coordinates. The visibility $\widehat{V}$ predicted by this model is given by

$$\widehat{V}(u,v) = \sum_{p=1}^{N_l}\sum_{q=1}^{N_m} \widehat{I}(p\Delta l, q\Delta m) e^{-2\pi i(pu\Delta l+qu\Delta m)}\,. \tag{8–2}$$

For simplicity we will henceforth denote the discrete form $\widehat{I}(p\Delta l, q\Delta m)$ by the notation $\widehat{I}_{p,q}$. Assuming reasonably uniform sampling of a region of the (u, v) plane, one can expect to estimate source features with widths ranging from $\mathcal{O}(1/\max(u, v))$ up to $\mathcal{O}(1/\min(u, v))$. The grid spacings, Δl and Δm, and

the number of pixels on each axis, N_l and N_m, must allow representation of all these scales. In terms of the range of (u,v) points sampled, the requirements are $\Delta l \leq \frac{1}{2u_{\max}}$, $\Delta m \leq \frac{1}{2v_{\max}}$, $N_l \Delta l \geq \frac{1}{u_{\min}}$, and $N_m \Delta m \geq \frac{1}{v_{\min}}$. This model has $N_l N_m$ free parameters, namely the cell flux densities $\widehat{I}_{p,q}$. The measurements constrain the model such that at the sampled (u,v) points

$$V(u_k, v_k) = \widehat{V(u_k, v_k)} + \epsilon(u_k, v_k)\,, \tag{8-3}$$

where $\epsilon(u_k, v_k)$ is a complex, normally distributed random error due to receiver noise, and k indexes the samples. At points in the (u,v) plane where no sample was taken, the transform of the model is free to take on any value. One can think of Equation 8–3 as a multiplicative relation

$$V(u,v) = W(u,v)(\widehat{V}(u,v) + \epsilon(u,v))\,, \tag{8-4}$$

where $W(u,v)$ is a weighted sampling function (see Lecture 7, Eq. 7–8) which is non-zero only for sampled points of the (u,v) plane,

$$W(u,v) = \sum_k W_k \delta(u - u_k, v - v_k)\,. \tag{8-5}$$

By the convolution theorem , this translates into a convolution relation in the image plane:

$$I^D_{p,q} = \sum_{p',q'} B_{p-p',q-q'} \widehat{I}_{p',q'} + E_{p,q}\,, \tag{8-6}$$

where

$$I^D_{p,q} = \sum_k W(u_k, v_k)\ \mathrm{Re}\left(V(u_k, v_k) e^{2\pi i (p u_k \Delta l + q v_k \Delta m)}\right) \tag{8-7}$$

and

$$B_{p,q} = \sum_k W(u_k, v_k)\ \mathrm{Re}\left(e^{2\pi i (p u_k \Delta l + q v_k \Delta m)}\right)\,. \tag{8-8}$$

$E_{p,q}$ in Equation 8–6 is the noise image obtained by replacing V in Equation 8–7 by $\epsilon(u_k, v_k)$. Note that the $B_{p,q}$ given by Equation 8–8 is the point spread function (beam) that is synthesized after all weighting has been applied (and after gridding and grid correction if an FFT was used; to keep the notation concise, we will not signify this gridding and grid correction explicitly). The Hermitian nature of the visibility has been used in this rearrangement.

Equation 8–4 represents the constraint that the model $\widehat{I}_{p,q}$, when convolved with the point spread function $B_{p,q}$ (also known as the *dirty beam*) corresponding to the sampled and weighted (u,v) coverage, should yield $I^D_{p,q}$ (known as the *dirty image*).

The weighting function $W(u,v)$ can be chosen to favor certain aspects of the data. For example, setting $W(u_k, v_k)$ to the reciprocal of the variance of the error in $V(u_k, v_k)$ will optimize the signal-to-noise ratio in the final image, whereas setting it to the reciprocal of some approximation to the local density of samples will minimize the sidelobe level. Robust weighting is a hybrid approach which attempts to achieve a good balance between these criteria, (see Lecture 7).

We shall now examine the possible solutions of the convolution equation.

1.1. The "principal solution" and "invisible distributions"

Let us now consider whether the convolution equation has a unique solution. Clearly if some of the spatial frequencies allowed in the model are not present in the data, then changing the amplitudes of the corresponding sinusoids in I will have no effect on the fit to the data. In effect, the dirty beam filters out these spatial frequencies. Let Z be an intensity distribution containing only these unmeasured spatial frequencies. Then $B * Z = 0$. Hence, if I is a solution of the convolution equation, so too is $I + \alpha Z$ where α is any number. Thus, as usual, the existence of homogeneous solutions implies the general non-uniqueness of any solution in the absence of boundary conditions. An important point to note is that Equation 8–6 cannot be solved by linear methods, such as $I' = A \times D$ where A is some matrix, since the homogeneous solutions Z will also be absent from I'. Thus, conventional deconvolution procedures such as inverse filtering, Wiener filtering, etc. (e.g., Andrews & Hunt 1977) will not work: a nonlinear procedure is required.

Interferometrists call the homogeneous solutions "invisible distributions" (Bracewell & Roberts 1954) or "ghosts". The solution having zero amplitude in all the unsampled spatial frequencies is usually called the "principal" solution. Invisible distributions arise from two causes: firstly, the (u, v) coverage extends only up to finite spatial frequencies, so that the invisible distributions correspond to finer detail than can be resolved; secondly, holes may exist in the (u, v) coverage.

The problem of image construction thus can be reduced to that of choosing plausible invisible distributions to be merged with the principal solution. The shortcomings of the principal solution must be considered before tackling this problem.

1.2. Problems with the principal solution

If the data are obtained on a regular grid then the principal solution can be computed very easily: one must simply choose the weighting function in Equation 8–7 so that the bias in weight due to the vagaries of sampling are corrected. For each grid point the visibility samples are summed with appropriate weights, and the total weight normalized to unity. In such circumstances, known as uniform weighting, the principal solution is thus equal to the dirty image and is given by the convolution of the true brightness distribution with the dirty beam. For most synthesis arrays currently in use, the dirty beam has sidelobes in the range 1% to 10%. Sidelobes represent an unavoidable confusion over the true distribution of any emission in the dirty image, which can be resolved only either by making further observations or by introducing *a priori* information such as the limits in extent of the source. For example, consider uniformly weighted observations of a point source: the dirty image is just the dirty beam centered on the point source position. Without *a priori* information we cannot tell whether the source is a point or is shaped like the dirty beam. Of course we know that Stokes parameter I must be positive and that usually radio sources do not resemble dirty beams (in particular they do not have sidelobe patterns extending to infinity) and so we could use this information as an extra clue. One further unsatisfactory aspect of the principal solution, besides its implausibility, is that

it changes (sometimes drastically) as more visibility data are added. A better estimator would possess greater stability.

A priori information is thus the key; in the rest of this lecture we consider two algorithms which use different constraints on the invisible distributions to derive solutions to the convolution equation. These algorithms, CLEAN and the Maximum Entropy Method (MEM), are still the predominant ones used for deconvolution of radio synthesis images.

2. The CLEAN Algorithm

The CLEAN algorithm, which was devised by J. Högbom (1974), provides one solution to the convolution equation by representing the radio source by a number of point sources in an otherwise empty field of view. A simple iterative approach is employed to find the positions and strengths of these point sources. The final deconvolved image, usually known as the CLEAN image, is the sum of these point components convolved with a CLEAN beam, usually Gaussian, to de-emphasize the higher spatial frequencies which are usually spuriously extrapolated.

We now describe some of the currently available CLEAN algorithms, including two variants of the Högbom algorithm which are better suited to large images.

2.1. The Högbom algorithm

This algorithm proceeds as follows:

1. Find the strength and position of the peak (i.e., of the point brightest in absolute intensity) in the dirty image, $I^D_{p,q}$. If desired, one may search for peaks only in specified areas of the image, called *CLEAN windows.*

2. Subtract from the dirty image, at the position of the peak, the dirty beam B multiplied by the peak strength and a damping factor γ (≤ 1, usually termed the *loop gain*).

3. Record the position and magnitude of the point source subtracted in a model.

4. Go to (1) unless any remaining peak is below some user-specified level. The remainder of the dirty image is now termed the residuals.

5. Convolve the accumulated point source model $\widehat{I}_{p,q}$ with an idealized CLEAN beam (usually an elliptical Gaussian fitted to the central lobe of the dirty beam).

6. Add the residuals of the dirty image to the CLEAN image formed in (5).

The last stage is not always performed but can often provide useful diagnostic information, for example about the noise on the map, residual sidelobes, "bowls" near the center of the image (Sec. 3.3 below), etc.

2.2. The Clark algorithm

Clark (1980) has developed an FFT-based CLEAN algorithm. A large part of the work in CLEAN is involved in shifting and scaling the dirty beam; since this is essentially a convolution it may, in some circumstances, be more efficiently performed via two-dimensional FFTs. Clark's algorithm does this, finding approximate positions and strengths of the components via CLEAN using only a small patch of the dirty beam.

In detail, the Clark algorithm has two cycles, the major and minor cycles. The *minor cycle* proceeds as follows:

1. A beam patch (a segment of the discrete representation of the beam) is selected to include the highest exterior sidelobe.

2. Points are selected from the dirty image if they have an intensity, as a fraction of the image peak, greater than the highest exterior sidelobe of the beam.

3. A list-based Högbom CLEAN is performed using the beam patch and the selected points of the dirty image. The stopping criterion for the CLEAN is roughly such that any remaining points would not be selected in step (2).

The algorithm then proceeds to a *major cycle* in which the point source model found in the minor cycle is transformed via an FFT, multiplied by the weighted sampling function that is the inverse transform of the beam, transformed back and subtracted from the dirty image. Any errors introduced in a minor cycle because of the beam patch approximation are, to some extent, corrected in subsequent minor cycles.

2.3. The Cotton–Schwab algorithm

Cotton & Schwab (as described in Schwab 1984) have developed a variant of the Clark algorithm in which the major cycle subtraction of CLEAN components is performed on the *ungridded* visibility data. Aliasing noise and gridding errors can thus be removed provided that the inverse Fourier transform of the CLEAN components to each (u, v) sample has sufficient accuracy. Two routes are used for the inverse transform: for small numbers of CLEAN components, a 'direct Fourier transform' is performed and so the accuracy is limited by the precision of the arithmetic. In the other extreme of a large number of CLEAN components, an FFT is more efficient but inevitably some errors are introduced in interpolating from the grid to each (u, v) sample. Currently, high order Lagrangian interpolation is used.

The other considerable advantage of the Cotton–Schwab algorithm, besides gridding correction, is its ability to image and CLEAN many separate but proximate fields simultaneously. In the minor cycle each field is CLEANed independently, but in the major cycles, CLEAN components from all fields are removed. In calculating the residual image for each field, the full phase equation, including the w-term, can be used. Thus, the algorithm can correct what is commonly called the "non-coplanar baselines" distortion of images (see Lectures 2, 7 and 19).

The Cotton–Schwab algorithm is often faster than the Clark CLEAN, the major exception occurring for data sets with a large number of visibility samples, where gridding over and over again becomes prohibitively expensive. The Cotton–Schwab algorithm also allows CLEANing with smaller guard bands around the region of interest, hence with smaller image sizes.

This algorithm is implemented in NRAO's Astronomical Image Processing System (AIPS) as the classic task `MX` and the modern version, `IMAGR`.

2.4. Other related algorithms

Several algorithms have been invented with the aim of correcting some deficiencies of CLEAN.

Steer, Dewdney & Ito (1984) developed a variant of the Clark algorithm in which the minor cycle is replaced by a step of simply taking all points above a sidelobe-dependent threshold, scaling them and then subtracting normally in the major cycle. The saving in time can be considerable compared to CLEAN, but the radio astronomy community has relatively little experience with this variant of the algorithm. For some sources it can suppress the well known striping of extended emission, but for high precision deconvolution of moderately compact objects, it does not appear superior to the basic CLEAN.

Segalovitz & Frieden (1978) proposed an *ad hoc* modification of the *dirty* beam to enhance the smoothness of the resulting CLEAN image. Cornwell (1983) justified a similar prescription as forcing the minimization of the image power (i.e., the sum of the squares of the pixel values) and thus pushing down the extrapolated visibility function. Both approaches seem again to ameliorate partially the striping instability to which CLEAN is susceptible but possibly at cost in the overall stability of the algorithm.

Keel (1988) extended the domain of CLEAN to conventional optical imaging with the 'σ-CLEAN', where instead of searching for the maximum in the dirty image residuals, one searches for the peak in signal-to-noise at each pixel. This change is necessary due to the different character of the noise involved. In spite of working quite adequately, this has not proved popular in optical work and other restoration algorithms are generally used instead.

2.5. Practical Details and Problems of CLEAN Usage

Theoretical understanding of CLEAN is relatively poor even though the original algorithm is about 25 years old. Schwarz (1978, 1979) has analyzed the Högbom CLEAN algorithm in some detail. He notes that in the noise-free case the least-squares minimization of the difference between observed and model visibility, which CLEAN performs, produces a unique answer if the number of cells in the model is not greater than the number of independent visibility measurements contributing to the dirty image and beam (*cf.* Eqs. 8–7 and 8–8), counting real and imaginary parts separately. This rule is unaffected by the distribution of (u, v) sample points so that, in principle, super-resolution is possible if enough data points are available. In practice, however, the introduction of noise and the use of the FFT algorithm to calculate the dirty image and beam corrupts our knowledge of the derivatives of the visibility function upon which super-resolution is based. (As shown in Chapter 5 of Briggs (1995), CLEAN is in fact particularly *bad* at the visibility extrapolation involved in super-resolution and

is not recommended for the purpose.) Even if the FFT is not used, the presence of noise means that independence of the data must be redefined. Schwarz has in fact produced a noise analysis of the least-squares approach but it involves the inversion of a matrix of side $N_l N_m$ and so is totally impractical for typical image sizes; furthermore, we are really interested in CLEAN, not the more limited least-squares method, since CLEAN will still produce a unique answer in circumstances where the least-squares method is guaranteed to fail. To date no one has succeeded in producing a noise analysis of CLEAN itself. The existence of instabilities in CLEAN, which will be discussed later, makes such an analysis highly desirable.

Schwarz also proves three conditions for the convergence of CLEAN:

1. The beam must be symmetric.
2. The beam must be positive definite or positive semi-definite. Thus the eigenvalues must be non-negative.
3. The dirty image must be in the *range* of the dirty beam. Roughly speaking, there must be no spatial frequencies present in the dirty image which are not also present in the dirty beam.

All three of these conditions are obeyed in principle for the dirty image and beam calculated by Equations 8–7 and 8–8 if the weighting function is nowhere negative. In practice, however, numerical errors, and the gridding and grid-correction process may cause violation of these conditions. The CLEAN algorithm will therefore diverge eventually. CLEANing close to the edge of a dirty image computed by an FFT is particularly risky. Even the most simple case of a three pixel image has been demonstrated by Tan (1986) to be potentially chaotic. Still, in real cases the algorithm seems to work well.

Marsh & Richardson (1987) showed that for the case of an isolated point source, the CLEAN algorithm approximately minimizes the sum of the pixel values in the component model, subject to the constraint that these are positive. That is, it returns the minimum flux solution consistent with the data. By comparison with empirical results of deconvolvers which explicitly minimize this criterion, clearly this is only an approximation even in simple cases—it is not obvious how far this insight should be trusted.

Thus most of our understanding of CLEAN still comes from a combination of guessing how to apply intuition and Schwarz's analysis to real cases, and much practical experience on real and test data. In the rest of this section we will attempt to summarize the current lore concerning how the algorithm should be used, and how it can fail.

2.6. The use of boxes

The region of the image which is searched for the peak can be limited to those areas (known as the CLEAN *windows* or *boxes*) within which emission is known or guessed to be present. These boxes effectively restrict the number of degrees of freedom available in the fitting of the data. Schwarz's work (and common sense) tells us that the number of such degrees of freedom should be minimized but that the CLEAN window should include all real emission in the image. For a simple source in an otherwise uncluttered field of view, one CLEAN window will do,

but multiple boxes may be needed when CLEANing more complicated sources, or for a field containing many sources. In the latter case, the presence of weak sources may be revealed only after the sidelobes of the stronger sources have been removed; more boxes may therefore be required as the CLEAN progresses. Note that such *a posteriori* definition of CLEAN boxes considerably complicates any possible noise analysis.

The practical implications of Schwarz's observation that the number of degrees of freedom should not exceed the number of independent constraints are difficult to gauge. In the presence of noise (u, v) points should be judged independent if the differences in visibility due to the size of structure expected are much greater than the noise level. Counting visibility points in such a way, the aggregate area of the CLEAN boxes in pixels should be less than twice the number of *independent* visibility points. If the FFT is used (see Lecture 7) then the number of independent visibility samples cannot be greater than $\mathcal{O}(N_l N_m)$, and so the use of CLEAN boxes is certainly advisable.

Given the uncertainty in determining the number of independent data points, and hence the number of constraints, caution dictates that boxes should always be placed tightly around the region to be CLEANed.

2.7. Number of iterations, loop gain and the beam patch size

The number of CLEAN subtractions N_{CL} and the loop gain γ determine how deep the CLEAN goes. In particular, for a point source the residual left on the dirty image is $(1-\gamma)^{N_{CL}}$. Hence, to minimize the number of CLEAN subtractions (and so to minimize the CPU time) γ should be unity; one then finds, however, that extended structure is not well represented in the corresponding CLEAN image. In typical VLA applications a reasonable compromise lies in the range $0.1 \leq \gamma \leq 0.25$. (Incidentally, this dependence of the CLEAN image upon the loop gain is a nice demonstration of the multiplicity of solutions to the convolution equation.) Lower loop gains may be required in cases where the (u, v) coverage is poor, but experience suggests that the improvements in deconvolution for $\gamma \ll 0.01$ are generally minimal. If one is in any doubt then it is wise to experiment (e.g., by decreasing γ and increasing N_{CL}). One exception to the use of low loop gain is in the removal of confusing sources; it is preferable to remove them with high loop gain, as their structure is usually not of interest.

The choice of the number of iterations depends upon the amount of real emission in the dirty image. One should aim at transferring all brightness greater than the noise level to CLEAN components (some implementations of CLEAN allow one to specify a lower intensity limit to the components instead of N_{CL}). CLEANing deep into the noise is usually a waste of time unless you specifically wish to analyze the extended, low surface-brightness emission. For high dynamic range imaging, the highest deconvolution fidelity will occur when CLEANing very deeply, but the very act of CLEANing noise will alter the statistics and risks making the image appear better than it really is.

Examination of the list of CLEAN components, and, in particular, of the behavior of the accumulated intensity in the model, is useful in detecting divergence; sometimes the accumulated intensity diverges. As discussed above, divergence of the Högbom CLEAN is always due to a computational problem. Possible culprits are the gridding process, aliasing, and finite precision arith-

metic. In the case of the Clark or the Cotton–Schwab algorithms, the truncated dirty beam patch that is used in the minor cycles of these algorithms must violate Schwarz's conditions. Therefore both may be subject to instability or divergence if the minor cycle is prolonged unduly. The default size of the beam patch in deconvolution programs was often set at a time when computer memory was at more of a premium than it is today. For large sources, it is often a wise idea to override this choice and use a larger beam patch than the default. In addition to obviously helping to delay divergence in the minor cycle it can also improve the overall level of the deconvolution by ensuring that there are smaller errors to be corrected after each major cycle than otherwise.

2.8. The problem of short spacings

Implicit in deconvolution is the interpolation of values for unsampled (u, v) spacings. In most cases CLEAN does this interpolation reasonably well. However, in the case of short spacings the poor interpolation is sometimes rather more noticeable since very extended objects have much more power at the short spacings. The error is nearly always an underestimation and is manifested as a "bowl" of negative surface-brightness in which the source rests. In such a case, introducing an estimate of the zero-spacing flux density into the visibility data before forming the dirty image will sometimes help considerably. The appropriate value of this flux density would be that measured by a single element of the array. In practice, however, single array elements rarely have sufficient sensitivity or stability to provide this estimate accurately. Values estimated from surveys made with larger, more sensitive, and more directive elements are therefore frequently substituted. Choosing the weight for the zero-spacing flux density is difficult; the best estimate seems to be simply the number of unfilled cells around the origin of the gridded (u, v) plane. However, the results obtained are fairly insensitive to the value used *provided that the CLEAN deconvolution goes deep enough.*

The CLEAN windows or boxes may also be viewed as providing crude estimates of the shape of the visibility function near the zero spacing $u = v = 0$. For this reason, careful choice of CLEAN windows may also minimize problems associated with the short spacings.

After CLEANing, the emission should be, but is not guaranteed to be, distributed sensibly over the CLEAN image. Failure of the interpolation is indicated by the presence of a "pedestal" of surface brightness within the CLEAN box upon which the source rests. Such a pedestal all over the CLEAN image can be caused by insufficient CLEANing of the dirty image; one can experiment by simply increasing N_{CL}. Ultimately, it may actually be necessary to measure the appropriate data!

2.9. The CLEAN beam

The CLEAN beam (more generally called the *restoring beam*) is used to suppress the higher spatial frequencies which are poorly estimated by the CLEAN algorithm. There are two competing opinions on this in the radio astronomy community: some object that it is purely *ad hoc* and is undesirable—in the sense that the equivalent predicted visibilities do not then agree with those observed. Others defend it as a way of recognizing the inherent limit to resolution.

In practice, it does appear to be necessary in order to produce astrophysically reasonable images.

The magnitude of just how poorly CLEAN extrapolates past the sampling envelope has only recently been appreciated—the errors in the restored image comes almost always from the extrapolated region in the (u, v) plane and rarely from interior holes in the sampling. Thus there is a straightforward tradeoff in resolution against image fidelity controlled by the size of the restoring beam. The most common method of choosing the restoring beam is to fit an elliptical Gaussian to the central region of the dirty beam, but this default is not mandated. A smaller restoring beam allows more of the erroneously extrapolated model into the final solution and yields poorer fidelity in the name of higher resolution. Conversely, a larger than default CLEAN beam can produce a highly accurate deconvolution. This tradeoff is explored pragmatically in Chapter 5 of Briggs (1995).

Various attempts have been made to improve the selection of the CLEAN beam. The dirty beam, truncated outside the first zero-crossing, is appropriate in some applications since it lacks the extended wings of a Gaussian, but we emphasize that, after convolution with such a beam, the CLEAN image does not agree satisfactorily with the original visibilities. An ideal CLEAN beam might be defined as a function obeying three constraints:

1. Its transform should be unity inside the sampled region of the (u, v) plane.
2. Its transform should tend to zero outside the sampled region as rapidly as possible.
3. Any negative sidelobes should produce effects comparable with the noise level in the CLEAN image.

Constraint (1) is usually the first to be relaxed, and then only positivity of the transform is necessary. It may be that in typical applications CLEAN performs so poorly that these constraints do not allow an astrophysically plausible CLEAN image, however such a topic is probably worth further consideration.

One very important consequence of a poor choice for the CLEAN beam is that the units of the convolved CLEAN components may not agree with the units of the residuals. The units of a dirty image are not very well defined but can be called "Jy per dirty beam area". The only real meaning of these units is that an isolated point source of flux density S Jy will show up in the dirty image as a dirty beam shape with amplitude S Jy per dirty beam area. An extended source of total flux density S Jy will be seen in the dirty image convolved with the dirty beam, but the integral will not, in general, be S Jy. However, convolved CLEAN components do have sensible units of Jy per CLEAN beam, which can be converted to Jy per unit area since the equivalent area of the CLEAN beam is known. Careful control of the dirty beam shape with weighting parameters as described in Lecture 7 can often produce a more Gaussian-like dirty beam than the typical defaults, resulting in a better match in the flux scale between convolved components and residuals. A few imaging programs will also rescale the residuals with the ratio between the CLEAN beam and the fitted beam before adding these to the convolved components, but neither approach is a perfect correction. This issue is most important when significant flux remains in

the residuals, so when using very non-Gaussian dirty beams and/or a non-fitted restoring beam, it is best to run CLEAN quite deeply and transfer as much flux density to the components as possible. In this limit, the integral of the CLEAN image will often provide an accurate estimate of the flux density of an extended object—surprisingly often better than that of MEM—usually failing when the u-v coverage is incomplete on the spacings required. If convergence is not attained then both flux density and noise estimates taken from the CLEAN image can be in error.

2.10. Use of *a priori* models

A priori models of sources can be used to good effect in CLEAN. Perhaps the best example is in the CLEANing of images of planets; in this case the visibility function of a circular disk can be subtracted from the observed visibilities before making the dirty image. CLEAN then needs only to find the small perturbations from the disk model, and so both the image quality and speed of convergence should be improved.

2.11. Non-uniqueness

Perhaps the biggest drawback to the use of CLEAN is the way in which the answers depend upon the various control parameters: the CLEAN boxes, the loop gain and the number of CLEAN subtractions. By changing these one can, even for a relatively well-sampled (u, v) plane, produce somewhat different final CLEAN images. In the absence of an error analysis of CLEAN itself one can do nothing at all about this problem. Awareness of the possible effects discussed in this section should however keep you from becoming over-confident in the final CLEAN image, as will experience of applying CLEAN to a wide range of different images.

In any one application, Monte Carlo tests of CLEAN can sometimes be illuminating, and, indeed, provide the only means of estimating the effects of various data errors and CLEANing strategies upon the final image.

2.12. Instabilities

One particular instability of CLEAN is well known: in CLEAN images of extended sources one sometimes finds modulations at spatial frequencies corresponding to unsampled parts of the (u, v) plane (see, e.g., Cornwell 1983 for an example). Convolution with a larger than usual CLEAN beam will sometimes mask this problem, especially when the unsampled region is in the outer parts of the (u, v) plane. Reducing the loop gain γ to very low values generally has little effect, but there is reason to believe that the instability is triggered by noise and hence that *temporarily* setting the loop gain equal to the noise-to-signal ratio when the instability begins may help (U. J. Schwarz, private communication).

Cornwell (1983) has developed a simple modification to the CLEAN algorithm that is sometimes successful in countering the instability. A small-amplitude delta function is added to the peak of the beam before CLEANing. The effect of the spike is to perform negative feedback of the CLEAN structure into the dirty image, and thus to act against any features not required by the data. Spike heights of a few percent, and lower loop gains than usual are usually

required. In view of the limited success of this modification, a better solution is to use another deconvolution algorithm, such as MEM.

The occurrence of the stripes is a natural consequence of the incorrect information about radio sources embodied in the CLEAN algorithm. Astronomers very rarely find convincing evidence for the existence of such stripes in radio sources and so they are skeptical about such stripes when found in CLEAN images. Unfortunately the only *a priori* information built into CLEAN, via the use of CLEAN boxes, is that astronomers prefer to see mainly blank images; there is no bias against stripes. Such considerations, and some others, have led to the development of deconvolution algorithms which either incorporate extra constraints on astrophysically plausible brightness distributions or are claimed to produce, in some way, optimal solutions to the deconvolution equation. In the next section we briefly consider one such algorithm.

3. The Maximum Entropy Method (MEM)

The deconvolution problem is one of selecting one answer from the many possible. The CLEAN approach is to use a *procedure* which selects a plausible image from the set of feasible images. Some of the problems with CLEAN arise because it is procedural so that there is no simple equation describing the CLEAN image. Thus, for example, a noise analysis of CLEAN is very difficult. By contrast, the Maximum Entropy Method (MEM) is not procedural: the image selected is that which fits the data, to within the noise level, and also has maximum entropy. The use of the term *entropy* has lead to great confusion over the justification for MEM. Even today there is no consensus on this subject evident in the literature (e.g., Frieden 1972; Wernecke & D'Addario 1976; Gull & Daniell 1978; Jaynes 1982; Narayan & Nityananda 1984, 1986; Cornwell & Evans 1985). We will use the "lowest common denominator" justification and define entropy as something, which when maximized, produces a positive image with a compressed range in pixel values. Image entropy is therefore not to be confused with a "physical entropy", although the logarithmic definition given in equation 8-9 parallels that of of the Boltzman H-function in statistical mechanics (see Cornwell 1984, Landau & Lifshitz 1980). The compression in pixel values forces the MEM image to be "smooth", and the positivity forces super-resolution on bright, isolated objects. There are many possible forms of this extended type of entropy, see e.g., Narayan & Nityananda 1984, but one of the best for general purpose use is:

$$\mathcal{H} = -\sum_k I_k \ln \frac{I_k}{M_k e}, \tag{8-9}$$

where M_k is a "default" image incorporated to allow *a priori* knowledge to be used. For example, a low resolution image of the object can be used to good effect as the default.

A requirement that each visibility point be fitted exactly is nearly always incompatible with the positivity of the MEM image. Consequently, data are usually incorporated in a constraint that the fit, χ^2, of the predicted visibility

to that observed, be close to the expected value:

$$\chi^2 = \sum_k \frac{|V(u_k, v_k) - \widehat{V(u_k, v_k)}|^2}{\sigma^2_{V(u_k,v_k)}}. \qquad (8\text{–}10)$$

Simply maximizing $\mathcal{H}$ subject to the constraint that χ^2 be equal to its expected value leads to an image which fits the long spacings much too well (better than 1σ) and the zero and short spacings very poorly. The cause of this effect is somewhat obscure but is related to the fact that the entropy $\mathcal{H}$ is insensitive to spatial information. It can be avoided by constraining the predicted zero-spacing flux density to equal that provided by the user (Cornwell & Evans 1985).

Algorithms for solving this maximization problem have been given by Wernecke & D'Addario (1976), by Cornwell & Evans (1985), and by Skilling & Bryan (1984). The Cornwell–Evans algorithm is coded in NRAO's Astronomical Image Processing System (AIPS) as `VM` or `VTESS`. It is generally faster than CLEAN for larger images; the break-even point being for images of about 1 million pixels.

4. Practical Details of the Use of MEM

The following description relates to the AIPS MEM algorithm, `VM`.

4.1. The default image (prior distribution)

Examination of Equation 8–9 reveals that if no data constraints exist, the MEM image is the default image, so the MEM image is always biased towards the default. A reasonable "default default" image is flat, with total flux density equal to that specified. A low-resolution image, if available, can be used as the default to very good effect; this is a nice way of combining single-dish data with interferometric data. A spike in the default can sometimes be used to indicate the presence of an unresolved source, which could otherwise cause problems (see Sec. 4.5 below).

4.2. Total flux density

As described above, if the total flux density in the MEM image is not specified then the value found may be seriously biased if the signal-to-noise ratio is low. There is no real way around this at the moment, except by guessing a value and then adjusting it to get an image that looks "reasonable"—for example, possessing a flat baseline. For bright objects, only an order-of-magnitude estimate is required to set the flux density scale. Of course, then the estimated flux density is not fitted but is used only to set a reasonable default image.

4.3. Varying resolution

In the folklore, MEM is criticized for resolution that depends on the signal-to-noise ratio. In fact, there are sound theoretical reasons to believe that this effect is common to all nonlinear algorithms that know about noise (Andrews & Hunt 1977). If you want to "fix" the resolution in MEM, you basically have two choices:

1. Convolve the final MEM image with a Gaussian beam of appropriate width to smear out the fine scale structure and add the residuals back in.

2. Before deconvolution, convolve the dirty image with a Gaussian beam.

The advantages of (2) over (1) are that the algorithm usually converges faster, and that given the nonlinear nature of the deconvolution, the answer can be (and usually is) better. For example, sidelobes around a point source embedded in extended emission are not well removed by MEM, whereas scheme (2) often alleviates this effect. The advantages of (1) over (2) are that both image bias (see below) and errors in gradient representation are substantially alleviated by adding in the residuals.

There are occasions when the super-resolution exhibited by MEM images is reliable, although predicting this in advance is not feasible. With careful modeling of the source, however, it is possible to plausibly defend the physical reality of super-resolved features, as in Chapter 8 of Briggs (1995). MEM is in fact the algorithm of choice for super-resolution studies.

4.4. Bias

Another commonly heard complaint about MEM is that the answer is biased, i.e., that the ensemble average of the estimated noise is not zero. This is certainly true, and is the price paid by any method which does not try to fit exactly to the data as CLEAN does. Bias in an estimator is quite common and acceptable since it usually leads to smaller variance. Cornwell (1980) has estimated the magnitude of the bias, and has shown that it is much less than the noise for pixels having signal-to-noise ratio much greater than one. In fact, if the (u, v) coverage is very good then for bright pixels the effect of noise on an MEM image is very similar to that on a dirty image. The effect of bias can be substantially reduced by using a reasonable default such as a previous MEM image smoothed with a Gaussian; then only the highest spatial frequencies are biased. The effect of bias can also be eliminated by adding back the residuals after ensuring a similar flux scale via convolution of the MEM image with a Gaussian as outlined above.

4.5. Point sources in extended emission

Nearly all the power of MEM to remove sidelobes comes from the positivity constraint. Hence, if the source sits on a background level of emission, then the sidelobes will not be removed fully. The only consistently effective solutions are either (a) to remove the point sources using CLEAN or (b) to smooth the dirty image prior to deconvolution. MEM has difficult even with isolated point sources without a background, but only a small degree of resolution – say an intrinsic feature width of 1/5 of a beamwidth – is necessary for the algorithm to perform well.

5. Comparison of CLEAN and MEM

CLEAN has dominated deconvolution in radio astronomy since its invention nearly 25 years ago, but has not been widely applied in other disciplines. One of

the major reasons for this is the decomposition into point sources, which is often not permissible in other types of images. In contrast, MEM has spread to many different fields, probably because most of the justifications are independent of the type of data to which it is applied.

The philosophy behind MEM is intriguing and may convince some of you about the objectivity of MEM (see Jaynes 1982 for an exposition of MEM from its inventor). For those of you who do not become acolytes, the practical differences between CLEAN and MEM are probably more interesting.

CLEAN is nearly always faster than MEM for sufficiently small and simple images, because its approach of optimizing a relatively small number of pixels is simply more efficient. For typical VLA images, the break-even point is at around a million pixels of brightness. For very large and complex images, such as those of supernova remnants, which may contain up to 100 million pixels, CLEAN is impossibly slow and an MEM-type algorithm is absolutely necessary.

CLEAN images are nearly always rougher than MEM images. This may be traced to the basic iterative scheme: since what happens to one pixel is not coupled to what happens to its neighbors, there is no mechanism to introduce smoothness. MEM couples pixels together by minimizing the spread in pixels' values, so the resulting images look smooth although the entropy term does not explicitly contain spatial information.

Both MEM and CLEAN fail to work well on certain types of structure. CLEAN usually makes extended emission blotchy, and may introduce coherent errors such as stripes, while MEM copes very poorly with point sources in extended emission. Both work quite well on isolated sources with simple structure, and can produce meaningful enhancement of resolution, although MEM seems to do better in most cases.

Since MEM tries to separate signal and noise, it is necessary to know the noise level reasonably well. Also, as mentioned above, knowledge of the total flux density in the image helps considerably. Apart from this MEM has no other important control parameters, although it can be helped enormously by specifying a default image. CLEAN makes no attempt to separate out the noise, and so specification of the noise level is not required. The main control parameters are the loop gain γ, and the number of iterations $N_{\rm CL}$, both of which are important in determining the final deconvolution.

The default image of MEM is a very powerful mechanism for introducing *a priori* information. We have previously described the use of a simple image as a default; however, the default image need not be only a simple fixed set of numbers, but instead can be used to introduce functional relationships between pixels. For example, to further encourage smoothness, one might make the default for a pixel equal to the geometric mean of the brightness of its neighbors (S. F. Gull, private communication). Only the simple fixed default image can be easily mimicked by CLEAN: the default image is simply used as the starting point for the collection of CLEAN components. Thus the use of a disk model for a planet is an example of the use of a default in CLEAN.

6. Example

Figures 8–1 and 8–2 give an example deconvolution of a core jet source, adapted from Cornwell (1995). The data are synthetic, with the model resembling M87 and scaled to the VLBI size regime. The source was 'observed' with the VLBA at a frequency of 1.6 GHz, the declination was 50°, and the coverage was horizon to horizon down to 15° elevation. That is, the (u, v) coverage is superb by VLBI standards, and medium to poor by those of the VLA. The bright point source core and extended jet was designed to demonstrate the strengths and weaknesses of the two algorithms, with CLEAN performing better on the core, and MEM performing better on the extended emission. A small amount of thermal noise was added – below the lowest contour level – but the calibration was assumed to be perfect.

Figure 8–1 shows the parameters and truth image of the simulation. Panel **(a)** shows the model, smoothed to the same resolution as the restored images. The (u, v) coverage is in panel **(b)** and the visibility amplitude in panel **(c)**. Notice that the total flux density is dominated by the extended emission, yet the point source core will totally dominate the deconvolution in some respects. Panels **(d)** and **(e)** show the uniformly weighted beam used in this simulation. The former has lowest contour at ±2.2% and the latter is a typical slice though the central portion of the beam.

Figure 8–2 presents the results of different deconvolution strategies. Panel **(a)** is a simple Clark CLEAN with a loop gain of 0.1, run to 20,000 components without any constraint on their position. (The contours are roughly powers of two from a low of 0.05%.) The image has greatly improved from the dirty image (not shown), but there is still some evidence of sidelobes paralleling the jet. In panel **(b)**, components have only been allowed in a tight region surrounding the model source. The same 20,000 components now produce a very good image, showing the incorporation of information from the support constraints. Panel **(c)** is the same image, but contoured starting a factor of 10 lower.

The lower three panels are all MEM images. The first, panel **(d)** was generated with a flat default and the same support constraints as in panel **(b)**. 80 iterations were used, as compared to the more typical 30 and still MEM is having great difficulty with the point source core. (The image without the support constraint was even poorer.) In panel **(e)**, a point source model was fitted to the core of the CLEAN image, then subtracted from the visibility data. The residuals were imaged with MEM (and the support constraint), and then final image reassembled—the difference is dramatic. Panel **(f)** has this same image contoured down at the level of panel **(c)**. The best MEM image is smoother at the lowest contour levels than the best CLEAN image, and has a different characteristic error pattern. The images are of comparable fidelity. Remember that MEM has the most difficulty with *point* sources. If the core had been resolved so much as half a beam width, the initial MEM image would have been comparable to the CLEAN image without needing the subtraction.

7. Other Methods, Including Hybrids

Deconvolution in radio astronomy is still dominated by two *nonlinear* algorithms, CLEAN and MEM. Other nonlinear algorithms exist and may turn out to be

useful, at least in the sense that, as with CLEAN and MEM, their defects are orthogonal to those of other algorithms. This property of defect orthogonality also suggests the use of a combination of algorithms in the deconvolution of a single image, so that the virtues of each approach can be exploited. More novel approaches to deconvolution are under development, but have yet to be transfered into the mainstream of mundane reduction.

With modern computers, it is now possible to solve for directly the parameters of the point source model for some interesting objects, via brute force constrained least squares optimization. Briggs (1995) has applied the NNLS algorithm of Lawson & Hanson to solve the convolution equation for compact objects. At present this seems computationally feasible for images having non-zero flux in about 5000–6000 pixels. The quality of the deconvolved images is excellent for such sources, though it is actually the interaction of the algorithm with self-calibration which might prove the most important. A weakness of both CLEAN and MEM is that they produce highly correlated error patterns in the (u, v) plane. This correlated error pattern can prove a significant problem when the deconvolved model is used as input for a self-calibration correction, (see Lecture 10). CLEAN is better than MEM in this regard, though both algorithms can cause the deconvolution/self-calibration hybrid mapping cycle to stall. By contrast, NNLS appears to yield a much flatter and less correlated error pattern in the (u, v) plane and interacts extremely well with self-calibration. This approach might prove fairly useful for VLBI and for very high dynamic range VLA applications.

Multi-resolution algorithms are becoming more attractive. These all rely implicitly on the notion that deconvolution of simple objects is easier than complicated ones. In some cases, one actually iteratively solves related deconvolution problems at different scales, such as in the multi-resolution CLEAN of Wakker & Schwarz (1991). In others, the multi-resolution aspect is reflected in the decomposition of the problem into a wavelet domain. (See, e.g., Starck et al. 1994 or Pantin & Starck 1996). The wavelet based methods in particular seem very promising, but as yet available software is still an impediment to widescale exploration of these algorithms.

Hybrid techniques attempt to exploit the virtues, while avoiding the pitfalls, of a number of algorithms simultaneously. For example, the awkward but common circumstance of deconvolving compact structure on an extended background can be successfully approached with a shallow CLEANing of compact structure down to the level of the extended emission, followed by a MEM deconvolution of what remains. The component models of each method are then combined, restored, and added to the residuals. A further variant of this approach which is also effective for multi-pointing deconvolution problems consists of CLEANing the individual pointings at the full available resolution and forming the linear combination with appropriate weighting, while using MEM to simultaneously deconvolve the data at very low resolution. These results are then merged by extracting the inner Fourier transform plane of the MEM result and combining it (with appropriate normalization) with the outer Fourier transform plane of the CLEAN result and back-transforming. Surprisingly, while these techniques have now been used successfully for many years, there is still no streamlined datapath for the hybrid approachs. The scientist must still do

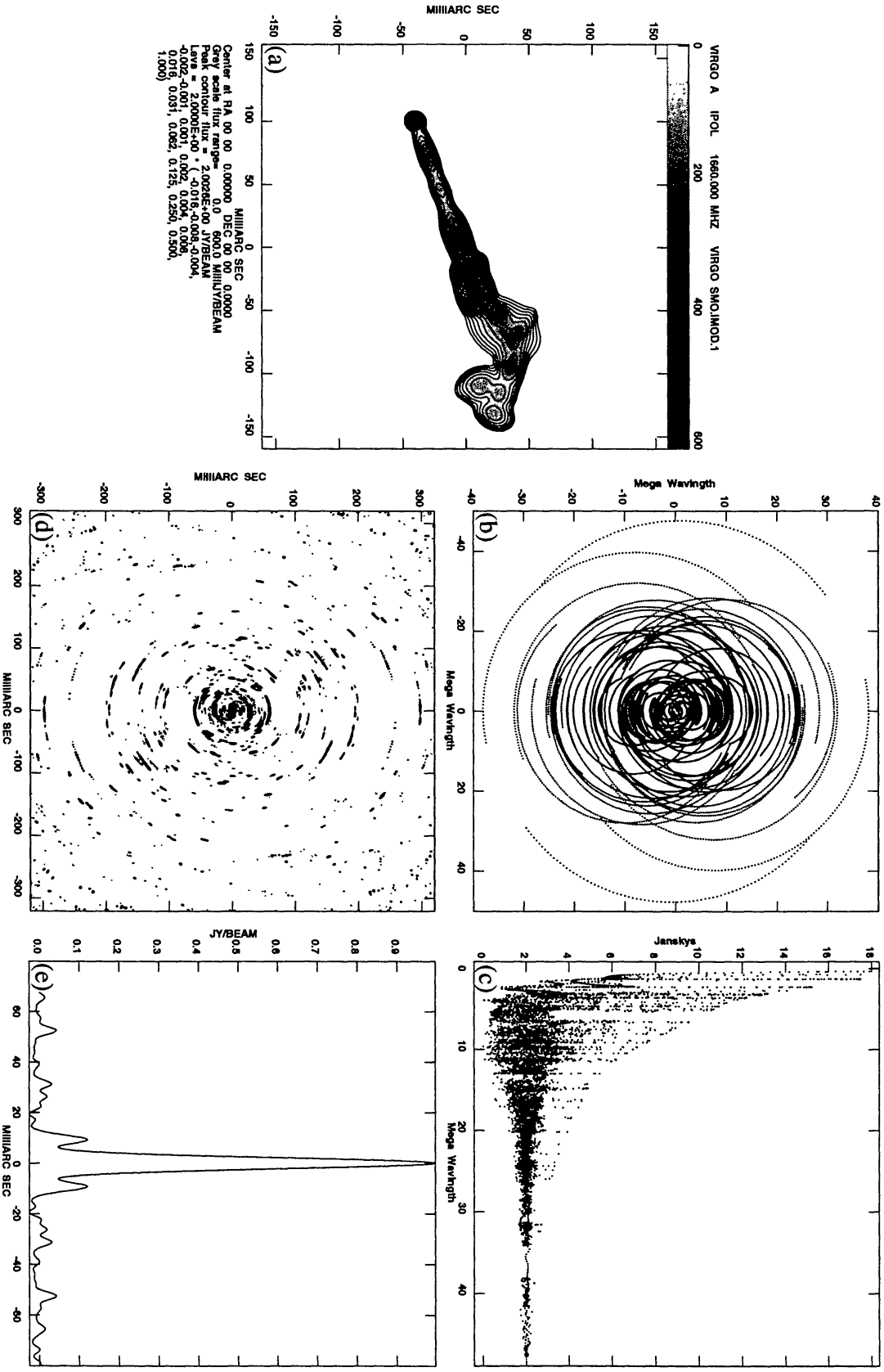

Figure 8–1. Example deconvolution: See text for details.

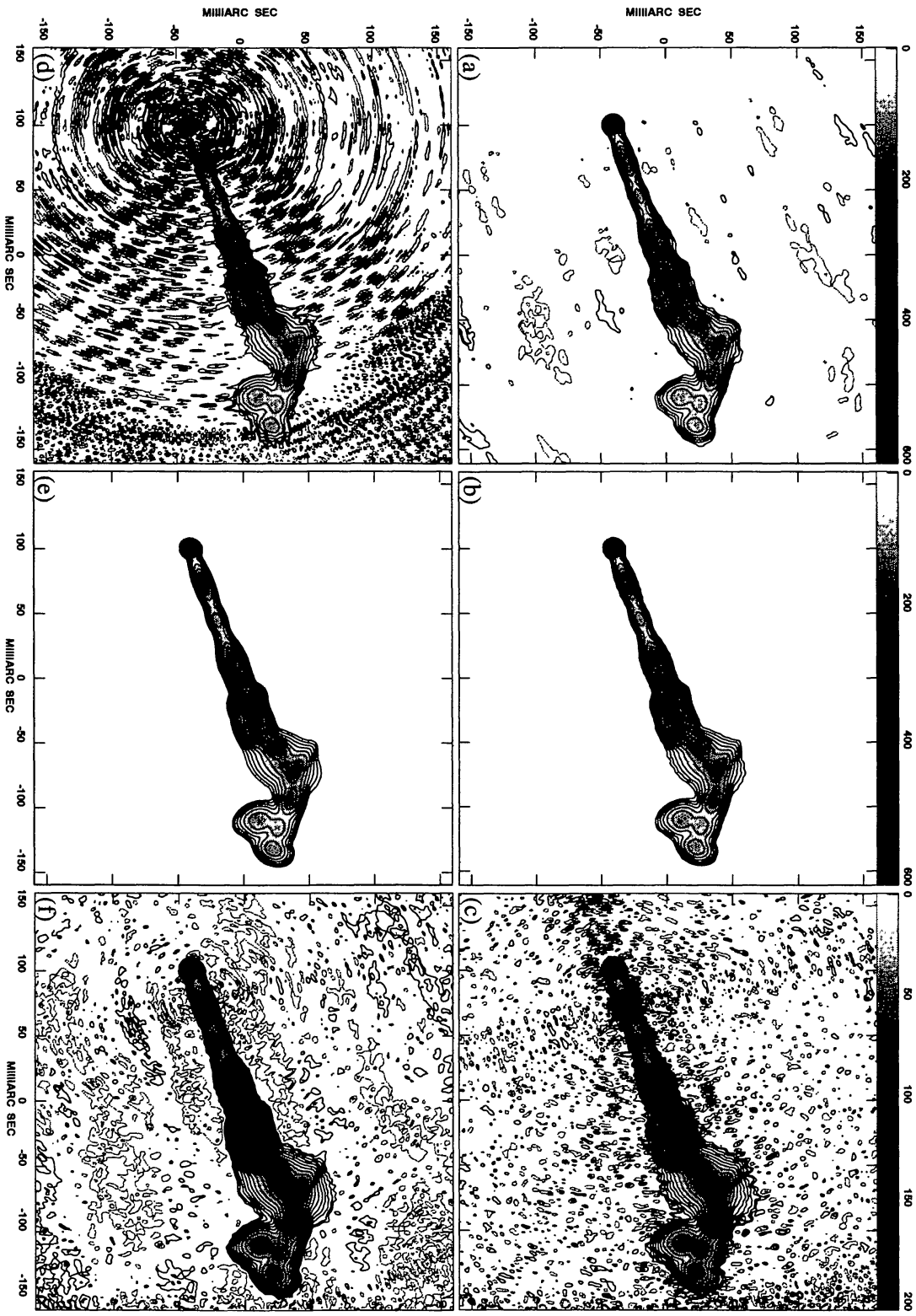

Figure 8–2. Example continued.

much of the bookkeeping involved in combining the results of the different sub-algorithms.

It is ironic that, formally, more is known about the type of images generated by MEM than by CLEAN (see e.g., Narayan & Nityananda 1986), since CLEAN is rather more widely used. Indeed many of the criticisms of MEM arise because certain of its properties, such as the bias, can be analyzed. Schwarz's analysis of CLEAN is incomplete in that it does not address the interesting underdetermined case in which there are fewer data than pixels. We hope that someday this problem might be investigated satisfactorily.

Although deconvolution algorithms are now as important in determining the quality of images produced by a radio telescope as the receivers, correlators and other equipment, they are far less well understood. A good description is that they are poorly engineered. Only further research and development of new and existing algorithms can redress this imbalance.

References

Andrews, H. C. & Hunt, B. R. 1977, *Digital Image Restoration*, Prentice–Hall (Englewood Cliffs, NJ).

Bracewell, R. N. & Roberts, J. A. 1954, *Aust. J. Phys.*, 7, 615–640.

Briggs, D. S. 1995, Ph. D. thesis, New Mexico Institute of Mining and Technology. Available via `http://www.aoc.nrao.edu/ftp/dissertations/dbriggs/diss.html`

Clark, B. G. 1980, *A&A*, 89, 377–378.

Cornwell, T. J. 1980, Ph. D. Thesis, University of Manchester.

Cornwell, T. J. 1983, *A&A*, 121, 281–285.

Cornwell, T. J. 1984, in *Indirect Imaging*, J. A. Roberts, Ed., Cambridge University Press (Cambridge, England), pp. 291–296.

Cornwell, T. J. & Evans, K. F. 1985, *A&A*, 143, 77–83.

Cornwell, T. J. 1995, in *Very Long Baseline Interferometry and the VLBA*, Zensus et al., Eds., San Francisco: Astronomical Society of the Pacific

Frieden, B. R. 1972, *J. Opt. Soc. Am.*, 62, 511–518.

Gull, S. F. & Daniell, G. 1978, *Nature*, 272, 686–690.

Högbom, J. 1974, *ApJS*, 15, 417–426.

Jaynes, E. T. 1982, *Proc. IEEE*, 70, 939–952.

Keel, W. C. 1988, *ApJ*, 329, 532–550.

Landau, L.D., & Lifshitz, E.M. 1980, *Statistical Physics*, New York: Pergamon Press, p. 119.

Marsh, K. A. & Richardson, J. M. 1987, *A&A*, 182, 174–178.

Narayan, R. & Nityananda, R. 1984, in *Indirect Imaging*, J. A. Roberts, Ed., Cambridge University Press (Cambridge, England), pp. 281–290.

Narayan, R. & Nityananda, R. 1986, *Ann. Rev. Astron. Astrophys.*, 24, 127–170.

Pantin, E. & Starck, J.-L. 1996, *A&AS*, 118, 575–585

Schwab, F. R. 1984, *AJ*, 89, 1076–1081.

Schwarz, U. J. 1978, *A&A*, 65, 345–356.

Schwarz, U. J. 1979, in *Image Formation from Coherence Functions in Astronomy*, C. van Schooneveld, Ed., D. Reidel (Dordrecht, Holland), pp. 261–275.

Segalovitz, A. & Frieden, B. R. 1978, *A&A*, 70, 335–343.

Skilling, J. & Bryan, R. K. 1984, *MNRAS*, 211, 111–124.

Starck, J.-L. et al., 1994 *A&A*, 283, 349–360

Steer, D. G., Dewdney, P. E., & Ito, M. R. 1984, *A&A*, 137, 159–165.

Tan, S. 1986, *MNRAS*, 220, 971–1001

Wernecke, S. J. & D'Addario, L. R. 1976, *IEEE Trans. Computers*, C-26, 351–364.