VLA Scientific Report 105
June 4, 1973

THE CONTINUUM MAPPING PROBLEM FOR THE VLA

R. M. Hjellming

Purpose: To describe some preliminary thoughts on probable and possible mapping tasks for the VLA asynchronous computer.

I. THE MAPPING EQUATIONS

1. Theoretical

The starting point for any discussion of interferometer synthesis mapping is the fact that measurements of the complex visibility function, V, for baseline components u and v projected on the plane of the sky, are related to the sky brightness, B, in the (x,y) plane, by

$$V(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)B(x,y)e^{2\pi i(ux+vy)}dxdy$$

where f(x,y) is a normalized "sensitivity" function reflecting the effects of the antenna pattern, delay beam, etc. For our purposes we need discuss only the determination of an apparent intensity distribution

$$I(x,y) = f(x,y)B(x,y)$$

so that

$$V(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(x,y)e^{2\pi i(ux+vy)}dxdy. \tag{1}$$

Some mapping techniques would use equation (1) as an equation of constraint in the determination of I(x,y). However, most current methods, and all of those that we will discuss in detail, rely upon the reverse Fourier transform relationship between I(x,y) and V(u,v):

$$I(x,y) = \int_{\infty}^{\infty}\int_{\infty}^{\infty} V(u,v)e^{-2\pi i(ux+vy)}dudv. \tag{2}$$

The practical methods by which one utilizes approximations for equations (1) and (2) constitute the entire mapping problem.

2. Approximate Equations Assuming Unsampled V(u,v) Are Zero

All synthesis interferometers are forced to utilize assumptions rather than data in dealing with unsampled (u,v) cells. Most presently used mapping techniques are based upon approximating equation (2) as follows:

$$I(x,y) \cong \frac{\sum_{DATA} V(u,v)e^{-2\pi i(ux+vy)}w(u,v)}{\sum_{DATA} w(u,v)} \tag{3}$$

where w(u,v) is a weighting function which describes which aspects of the V(u,v) measurements one wishes to emphasize. In addition to the arbitrariness of the weighting function, two different additional assumptions are involved in equation (3). The first, and least bothersome, is approximating the Fourier integral by a Fourier series. The second, and most troublesome assumption, is that

$$V(u,v) = 0$$

for all points where V has not been measured. This assumption is the major source of the side-lobe features apparent in what are called synthesized beams. The dominant feature of the VLA is the minimization of the effects of this assumption by sampling the (u,v) plane as extensively as possible.

3. Possible Alternate Methods

(a) Interpolation in the (u,v) Plane

One technique that has been partially explored is a simple one whereby V(u,v) is assumed to be represented by some smooth function of u and v and measurements of V(u,v) are used to determine the free parameters in this interpolation function. This determines all values of V(u,v) within certain portions of the (u,v) plane and equation (2) can be used to map I(x,y) with very few further numerical approximations. My impression is that those who have investigated these methods have concluded that the choice of the functional form of V(u,v) has too strong an effect upon the results to believe that any resulting maps are unique. My reaction to this is that all present mapping techniques suffer from analogous uniqueness problems--therefore some effort should be made by us to develop and evaluate mapping programs based on this technique.

(b) The Maximum Entropy Method (MEM)

Another approach which may show promise, but has not yet been developed to the point where any of us can praise or condemn it, is that of using measurements of V(u,v), in conjunction with equation (1), as constraints upon some condition on I(x,y). In the MEM as discussed by Ables an entropy-like function involving I(x,y) is maximized in conjunction with equation (1) and the method of Lagrangian multipliers to determine I(x,y). Philosophically this approach seems attractive since it amounts to adopting an interpolation function for rational rather than arbitrary reasons. During the next couple of years, I think we must thoroughly research and test the use of this method.

In any case, lacking anything better to do, we proceed at the present time using equation (3) to determine I(x,y).

II. EFFECTS OF GRIDDING IN THE (u,v) PLANE

1. Reasons for Gridding

As we are all aware, data on V(u,v) gathered by a particular pair in a synthesis array gives us coverage described by portions of an ellipse in the (u,v) plane. The 351 pairs in the VLA will provide us with coverage along 351 different elliptical tracks in the (u,v) plane. There are then two possible approaches to the use of equation (2) to determine I(x,y). Either we evaluate the sum effects of 351 line integrals (sums) along these elliptical tracks, or we use the measured V(u,v) to determine V at points in the (u,v) plane chosen for the convenience and speed of numerical calculation.

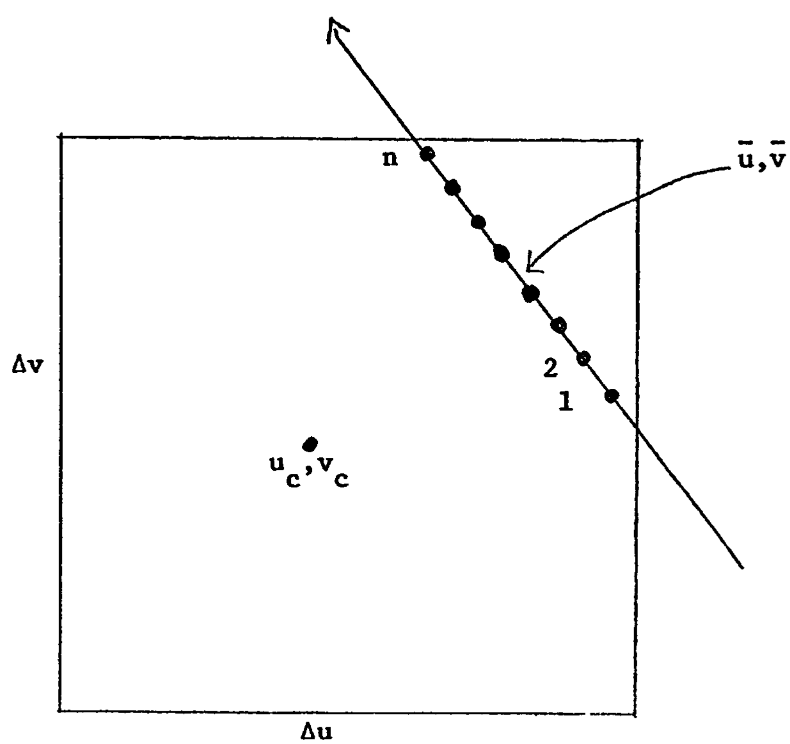
The latter approach is assumed by most people when they use the data to "determine" V(u,v) in a square grid of (u,v) points so the powerful and speedy Cooley-Tukey FFT algorithm can be used as a magic black box to make maps. However, there are a number of reasons why one may pay a high price for the use of this magic black box, and the main purpose of this present write-up is to explore its limitations and possible alternatives that can be used where its weaknesses can seriously limit the scientific results.

2. Assumptions Made in Gridding

In gridding visibility function data in the (u,v) plane to prepare to feed the Cooley-Tukey FFT algorithm, one chooses a square region in this plane and divides it into square boxes or cells. We assume that cells in the (u,v) plane have dimensions Δu and Δv. The gridding process then proceeds as follows.

If no measurments of V(u,v) are made within the cell, one sets V = 0 for that cell.

If data was taken within a cell that we can identify with an index c, let us assume that n successive measurements of V were made within the cell. Thus (u_k, v_k), k = 1, ..., n designates the locations of the n measurements in the (u,v) plane.



The first part of the gridding process is to obtain

$$\bar{V} = \sum_{k=1}^{n} V_k/n$$

$$\bar{w} = \sum_{k=1}^{m} w_k/n \ .$$

Normally one assigns a unit weight to each measurement within a cell, and then adopts a weight for the cell. In the so-called "natural" weighting, best suited for getting information about point sources dominating a field,

$$\bar{w} = n;$$

however, if one has considerable source structure and wishes to weight all size scales equally, one sets

$$\bar{w} = 1.$$

Between these two extremes there are a large number of possible compromises.

The mapping process then proceeds according to

$$I(x,y) = \frac{\sum_{c=1}^{n_c} I_c(x,y)w(u_c,v_c)}{\sum_{c=1}^{n_c} w(u_c,v_c)}$$

for all n_c cells in the square grid, where

$$I_c(x,y) = e^{-2\pi(u_c x+v_c y)} \sum_{k=1}^{n} V_{(uk,vk)} = \bar{V}e^{-2\pi i(u_c x+v_c y)}. \qquad (4)$$

3. <u>Evaluation of Errors Introduced by Gridding</u>

Obvious errors are introduced through the use of equation (4) which will affect the evaluation of $I(x,y)$. One can think of these errors in either of two ways: (1) the value $\bar{V}$ is a slightly erroneous approximation to the true value of $V(u_c,v_c)$; or (2) the proper transform factor for $\bar{V}$ should be $\exp[-2\pi i(\bar{u}x+\bar{v}y)]$ where

$$\bar{u} = \sum_{k=1}^{n} u_k$$

$$\bar{V} = \sum_{k=1}^{n} v_k \ ,$$

so an error factor of

$$\exp \{2\pi i[(\bar{u}-u_c)x + (\bar{v}-v_c)y]\}$$

is introduced during the transform process.

Let

$$\delta u = \bar{u}-u_c$$

$$\delta v = \bar{v}-v_c$$

and let us evaluate this error factor in terms of a phase error $\delta\Phi$, where

$$\delta\Phi = 2\pi(\delta ux+\delta vy). \tag{5}$$

The size of the most likely error in u and v will be

$$\frac{\delta u}{u} \approx \frac{\delta v}{v} \approx \frac{1}{4N}$$

where N is the transform size, usually taken to be 1024 for VLA planning. This evaluation is equivalent to noting that one will, on the average, be 1/4 of a (u,v) cell away from (u_c,v_c). The maximum errors introduced will then be of the order of twice the most likely error.

Taking N = 1024

$$\frac{\delta u}{u} \sim \frac{\delta v}{v} \sim 2.5\text{x}10^{-4}.$$

. Now for the VLA used at maximum 35 km resolution,

$$u_{max} \sim v_{max} \sim \frac{3.5\text{x}10^{6}}{\lambda_{cm}}$$

where λ_{cm} is the observing wavelength in cm. Hence, for the maximum resolution information,

$$\delta u_{max} \sim \delta v_{max} \sim \frac{900}{\lambda_{cm}}.$$

Now

$$\begin{aligned}\delta\Phi &= 2\pi(\delta ux+\delta vy) \\ &= 360° \ (2.9\text{x}10^{-4})(\delta ux_{arcmin}+\delta vy_{arcmin}) \\ &= 0.^{\circ}1 \ (\delta ux_{arcmin}+\delta vy_{arcmin}).\end{aligned}$$

taking $x_{min} \sim y_{min}$

$$\delta\Phi_{\text{most likely}} \approx \frac{(0.1)(2)(900)}{\lambda_{cm}} x_{arcmin}$$

$$\boxed{\begin{aligned}\delta\Phi_{\text{most likely}} &\approx \frac{180° \ x_{arcmin}}{\lambda_{cm}} \\ \delta\Phi_{\text{maximum}} &\approx \frac{360° \ x_{arcmin}}{\lambda_{cm}}\end{aligned}} \qquad (6)$$

Thus

Table 1

Errors Introduced Into Maximum Resolution Information*

x_{arcmin}/λ_{cm}	$\delta\Phi_{most\ likely}$	$\delta\Phi_{maximum}$
HPBW/2	150°	300°
HPBW/3	100°	200°
HPBW/4	75°	150°
HPBW/5	60°	120°
HPBW/10	30°	60°

Obviously the smaller the field one attempts to synthesize, the less the error introduced into the transform. Furthermore, the errors scale downwards with appropriate factors for smaller values of u and v. Therefore, the inner, lower resolution pairs of the VLA will be relatively unaffected, but the outermost ones--gained at fair expense in building railroad track--are strongly effected.

4. <u>Transform Size Needed to Reduce Transform Errors</u>

Many recent discussions have alluded to the possible need of a larger transform size for VLA mapping, particularly Ron Ekers at the recent Advisory Committee meeting. Rather than 1024x1024 transforms, sizes of up to 8192x8192 have been mentioned.

I believe that the results in Table 1 show the essential basis on which people have argued for a larger transform size for an instrument with the VLA resolution.

Everything depends on the standards one wishes to adopt. For example, to assert that at the maximum VLA resolution phase errors of no more than 20° should be introduced during the mapping process for fields synthesized out to HPBW/4, the only way to attain this in the context of square gridding for the Cooley-Tukey FFT is to increase N from 1024 to 8192.

* In this table the HPBW is that of the antenna pattern.

One might argue that for the VLA many tracks will pass through many of the cells, so a better estimate of $V(u_c, v_c)$ will be obtained. This is not valid because the phase errors $\delta\Phi$ basically describe how much the phase of V can change across a cell. Thus the only way to reduce the errors in this context is to have many more smaller (u,v) cells.

In practice I doubt if we want to go in this direction. As we will discuss shortly, it is probably more intelligent to foreswear brute force transforms and make limited use of flexible direct transform methods where the FFT is insufficient.

5. "Reduction" of the Field of View

A common trick used to show more detail in a source so that nice contours can be drawn, or so that cleaning algorithms will "work more effectively", is to reduce the field of view from that where all the (u,v) coverage just fits within the square grid in the (u,v) plane to one in which the (u,v) coverage is a small fraction of the gridded (u,v) plane being transformed.

For example, for the NRAO interferometer using N = 256 and mapping $x_{arcmin}/\lambda_{cm} = 0.81 = \text{HPBW}/2$, the u_{max}, v_{max} just fit within the (u,v) grid obtained; however, in the (x,y) plane the cell size is then $0''.38\ \lambda_{cm}$, or $4''.2$ at 11 cm and $1''.4$ at 3.7 cm. The resulting maps are very "grainy" and do not contain position information anywhere near the level at which the instrument is capable. By reducing the field of view by factors of two, three or even more, one can obtain much "smoother" maps showing more "detail". One problem, however, is that the phase errors introduced into the transform by the gridding increase by the same factors of two, three or more.

Exactly the same effects will occur with VLA mapping as people attempt to reduce the "graininess" of their maps.

6. Acerbation of the Confusion Problem

Perhaps the worst effect introduced by square gridding in the (u,v) plane is the introduction of artificial periodicities which did not exist in the original data. This introduces serious aliasing errors when sources perceived by the antennas are outside the synthesized field, i.e., their images are "reflected" or "aliased" into the maps to varying degrees. With the great sensitivity of the VLA many mapping problems may be seriously hindered by this effect. As has been frequently mentioned by one member of the NRAO staff, a serious study of the probable effects of this problem should be made.

One should note that these aliasing effects will be considerably reduced by not introducing square gridding before performing the Fourier transform on the data.

7. Large Sources

Many sources that people will want to study with the VLA will be larger than the fields that can be synthesized. In this case there is no question that large phase errors introduced by the transform process, for radiation at the edges of the synthesized field, will not be acceptable. In addition, such sources will be "self-confusing".

8. Alternatives to Square Gridding

All of the above-mentioned problems make it imperative that many other mapping techniques be available to supplement use of the FFT in solving mapping problems. I think this is more intelligent than planning for 8192 x 8192 transforms.

The most obvious alternate technique is to directly transform the data without regridding. In this case no more errors are introduced into the maps beyond those inherent in the available data. We will therefore now turn to a more detailed discussion of this option. However, because of the limitations of the direct transform (slow!!), the best solutions may eventually be the maximum entropy or (u,v) plane interpolation methods.

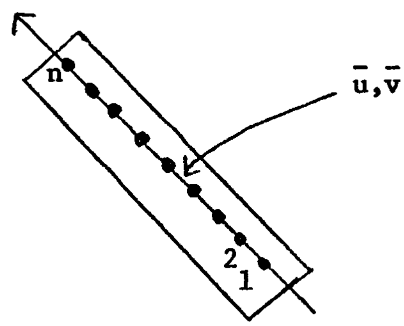
III. DIRECT TRANSFORM OF AVERAGED DATA

1. Direct Transform Procedures

The primary errors introduced into the mapping transformation by square gridding were basically due to treating the averaged V,

$$\bar{V} = \sum_{k=1}^{n} V_k/n$$

for each track through a cell as if it were the V for the point (u_c, v_c), rather than $(\bar{u}, \bar{v})$. There is, in practice, no way to avoid averaging V on some basis before applying the transform. The reason for this is the prohibitive cost in computing time if each data point were to be transformed separately. Hence any practical version of the direct transform will involve averaging as follows



with the transform equation

$$I(x,y) = \frac{\displaystyle\sum_{\substack{\text{averaged}\\ \text{segments}}} \bar{V}\, e^{-2\pi i(\bar{u}x+\bar{v}y)}\, w(\bar{u},\bar{v})}{\displaystyle\sum_{\substack{\text{averaged}\\ \text{segments}}} w(\bar{u},\bar{v})} \tag{7}$$

where

$$\bar{V} = \sum_{k=1}^{n} V_k/n$$

$$\bar{u} = \sum_{k=1}^{n} u_k/n$$

$$\bar{v} = \sum_{k=1}^{n} v_k/n$$

and one envisions the averages as successive segments of the elliptical tracks in the (u,v) plane.

Errors are introduced by the averaging, but they are minimized by the symmetry of the averaged points with respect to the point $(\bar{u},\bar{v})$ where the averaged $\bar{V}$ is transformed. Further, in principal, the results can be made as accurate as needed by shortening the averaging interval.

2. Reduction of Aliasing Errors

By direct transformation of data segments along (u,v) tracks no artificial periodicities are introduced other than those built in by sampling limitations. Only comparison of numerical examples, or practical experience with the VLA, will inform us about how much is gained by this.

3. The Price Paid in Doing Direct Transforms

The direct transform according to equation (7) has two built-in problems. First of all, the time required is directly proportional to the amount of data to be transformed, unlike the Cooley-Tukey FFT. Secondly, the time required is directly proportional to the square of the number of (x,y) points for which one wishes to evaluate I(x,y). However, unlike the FFT procedures, one need only evaluate the transform for the (x,y) that one is interested in; hence the principal use of the direct transform lies in achieving optimum analysis of the radiation distribution in small regions containing sources of interest.

During the last month the process of writing some direct transform mapping programs for sizes up to 127 x 127 has shown that the price of doing the direct transform is not as high as I originally expected. For example, at first look equation (7) seems to require that $M \cdot N^2$ sine functions and $M \cdot N^2$ cosine functions need to be evaluated to transform M averaged data points to determine N^2 values of I(x,y) in the (x,y) plane. However, since

$$\cos(\bar{u}x+\bar{v}y) = \cos \bar{u}x \cos \bar{v}y - \sin \bar{u}x \sin \bar{v}y$$

and

$$\sin(\bar{u}x+\bar{v}y) = \sin \bar{u}x \cos \bar{v}y + \cos \bar{u}x \sin \bar{v}y,$$

one need only evaluate $2 \cdot M \cdot N$ sine functions and $2 \cdot M \cdot N$ cosine functions. Furthermore, since

$$\cos \bar{u}x = \cos(-\bar{u}x)$$

$$\sin \bar{u}x = -\sin \bar{u}x$$

$$\cos \bar{v}y = \cos(-\bar{v}y)$$

$$\sin \bar{v}y = -\sin(-\bar{v}y)$$

one really need evaluate only $M \cdot N$ sine functions and $M \cdot N$ cosine functions to determine I(x,y) for N^2 points on the sky. This is enough to make the evaluation of sines and cosines no longer the major part of the computing time. All of the time of the transform then goes into performing $2 \cdot M \cdot N^2$ multiplications and $2 \cdot M \cdot N^2$ additions.

4. <u>Philosophy of Dilect Transform Mapping</u>

a. <u>Weighting</u>

In using the FFT algorithm the equal weighting of (u,v) cells is most natural and easily applied. In the direct transform it is most natural to weight each transformed point according to the amount of observing time involved. However, for both cases one can easily obtain any weighting by multiplying $\bar{V}$ by the appropriate weighting function just before the transform.

b. Study of Small Regions in the Field of View

The greatest utility of direct transform mapping will be the detailed study of regions involving particular sources of interest. Any level of detail can be obtained (given inherent resolution limits of the data) and aliasing errors are absolutely minimized. One can easily visualize a mode of operation whereby an FFT device provides a rough map subject to the errors previously discussed, thus informing the observer about what is probably present in the field; then the observer asks for and gets a detailed direct transform map for a small region of interest.

IV. WEIGHTING OF DATA BEFORE MAPPING

1. General Philosophy

The weighting of one measurement of V(u,v) relative to another is a nontrivial matter and is not automatically choosable before a field has been mapped. The ideal weighting for any particular map depends upon both the nature of the sources within the map and on what particular features the observer wishes to study. As an indication of how slow we have been to fully realize this, it was only in the middle of 1972 that people at the NRAO began to realize how inappropriate the only weighting possible with the then available mapping program was for point source and detection problems, and that for these problems each cell should be weighted only according to the amount of observing time involved. In the other extreme, an "unbiased" map of a source with many size scales of structure demands equal weight for each cell in the (u,v) plane--even though much better signal to noise is attained for inner cells as compared to outer cells with large u and v.

Another case that can occur frequently in sources with extensive structure arises from that fact that considerably more flux is present in the large-scale structures compared to the smaller ones. Hence it can occur that the "noise" associated with large-scale structures will equal or exceed the flux contained in high signal-to-noise information about very small structures. As a result, real information about the small structures can "float" on a noisy sea of larger size scale noise. In this case it is useful to obtain maps suppressing all (u,v) data except that data on the small-scale structures.

Within certain limits the observer can and must produce maps that emphasize what he is interested in. No one can decide for the observer what weighting or mapping techniques is best for him.

2. Practical Effects of Weighting Ambiguities

In the operation of the VLA the observer must be supplied with the capability to choose, and change his mind about, maps with a diverse menu of weighting and mapping assumptions. This will necessitate a fairly large degree of interactive discourse between observer and computer, and hence a larger computer installation than would suffice if only one kind of weighting and one kind of mapping could satisfy all needs.

V. MODES OF OBSERVING AND MAPPING NEEDS

1. Possible Observing Modes

For the purposes of considering mapping needs and the associated computer requirements one can envision three types of observing modes: snapshot mode, source-weaving mode, and full-tracking mode.

a. Snapshot Mode

Quick observations, say 10-15 minutes of duration, of a large number of fields, or particular fields in advance of deciding which deserve further attention. As shown by the examples calculated by Dave Hogg, a surprisingly low side-lobe level is obtained in this mode because of the widely distributed (u,v) plane coverage of the VLA at any instant. As a result most observing programs will involve a fair amount of reconnaissance in snapshot mode before deciding what objects, regions, etc., are deserving of the full power of the VLA. Because of this the VLA in snapshot mode will be a survey instrument for the VLA in modes where more time is spent in each field.

b. Source-Weaving Mode

Many observing problems will be solvable by inter-weaving observations of a few different fields so that wide hour angle coverage is obtained with minimum over-kill (waste) in observing time. More sources per unit time can be studied in this way, and missing "holes" in coverage can always be filled at later times if preliminary results justify the investment.

c. Full Tracking Mode

Many sources or fields will require horizon-to-horizon tracking because of the complex structure and/or weak flux levels.

2. Effect of Observing Mode on Mapping Needs

The frequent use of snapshot mode as a prelude to observing in source-weaving or full tracking mode will cause the observer to want quick evaluation of maps with various weightings, level of detail, etc. It has always been felt that the VLA would produce maps that build up in time, but a more complex interactive system will be needed if the observers wish to exercise a number of mapping options as quickly as possible.

3. Related Observer Needs

a. Detection and Deletion of "Bad" Data

No matter what level of perfection is attained in the electronics and on-line software, there will always be at least some need to identify, flag, and leave out bad data. However, if this is too extensive the observer will be crushed under the load.

b. Display of "Hard" Information About Maps

Since simply looking at displayed maps provides only relative information, the observer will want the related capability to obtain "hard" numbers about intensities and positions for particular (x,y) positions. Related to this may be a demand for source model fitting capability or heaven forbid, cleaning algorithms.

c. Display of (u,v) Maps

The observer will want to display his (u,v) coverage. In addition, with proper development of display capabilities it may become very useful to display maps of the data in the (u,v) plane. An experienced observer will be able to look at a three-dimensional display of amplitude/or phase/or Re(V)/or Im(V) and gather information about the probable complexity of the field, or even its content. The evaluation of cases where strong sources are just outside mapped fields could possibly be made in this way. At the moment it is just guessing, but with the great (u,v) plane coverage of the VLA, maps of the data in this plane may be supremely valuable in analyzing fields, or particularly in detecting bad data. I can think of no other way an observer will be able to examine his data except in this form--it will be impossible to stare at all numbers for all correlators,but to find out which deserve close examination from display of (u,v) maps may be relatively easy.