

5. CALIBRATION AND EDITING

EDWARD B. FOMALONT AND RICHARD A. PERLEY

1. INTRODUCTION

Lecture 1 showed that, under certain well-justified assumptions, a two-element correlation interferometer measures the spatial coherence function of the radiation field at a location given by the antenna separation, or baseline, measured in wavelengths. Under other, generally well-justified assumptions, all these measurements of the coherence function can be considered to lie upon a plane, so that a two-dimensional Fourier transform suffices to obtain the sky brightness as a function of angle. (See Lecture 14 for a discussion of the consequences of having the measurements not lie sufficiently near a plane.) This spatial coherence function is usually referred to as the *true* visibility function and is denoted here by V_{ij} , with the subscripts indicating which pair of antennas is involved. An array of antennas samples this visibility function at many discrete locations. Lectures 2, 3 and 4 describe how the signals are collected, amplified, converted, transported, correlated and averaged. The data from each antenna pair are then recorded; the resulting ensemble of numbers is called the *observed* visibility data. In general, the recorded quantities are different from the desired visibilities. The intent of calibration is to recover the true visibility. We will use the symbol \tilde{V}_{ij} to denote observed visibilities.

The observed visibilities differ from the true visibilities for a multitude of reasons. The purpose this lecture is to discuss the origins and effects of the various mechanisms which affect the observed visibilities, to describe the methods used to measure the errors, and to describe how these derived parameters are used to estimate the true visibilities. We term this endeavor the *calibration* of the data.

The process of identifying and discarding discrepant and severely corrupted data is called *editing*. Discarding data is also commonly referred to as *flagging*, although the original meaning of the term was to note, or mark questionable data. Interference from terrestrial sources, antenna tracking inaccuracies, inclement weather, malfunctioning receivers, incorrect observing parameters, and data recording errors are but a few of the problems that may require flagging the affected data.

This lecture will emphasize calibration and editing problems and techniques in use at the VLA, with some digression on VLBI techniques. The discussion will emphasize general problems and not get too involved with specific VLA applications and software. The general organization of the lecture is:

1. Introduction
2. Basic Considerations
3. Initial Calibrations

4. Routine Corrections
5. Bandpass Calibration
6. Editing
7. Final Calibration Using Radio Sources
8. Application of the Calibration Values
9. Polarization Calibration
10. Additional Topics.

2. BASIC CONSIDERATIONS

2.1. Synthesis equation and phase delay.

The most useful array synthesis formulation for calibration purposes is Equation 1–13 from Lecture 1:

$$V(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A_{\nu}(l, m) I_{\nu}(l, m) e^{-2\pi i(ul + vm)} dl dm, \quad (5-1)$$

where

- ν is the frequency of the radiation,
- (l, m) specifies the direction cosines with respect to the phase-tracking center,¹
- (u, v) denotes the projected baseline coordinates, measured in wavelengths,
- $V(u, v)$ is the visibility function, evaluated at (u, v) ,
- $A_{\nu}(l, m)$ is the normalized reception pattern of the antenna, and
- $I_{\nu}(l, m)$ is the intensity distribution of the source.

Since the visibility is sampled at discrete times for each antenna pair, it is often useful to write the above equation as

$$V_{ij}(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A_{\nu}(l, m) I_{\nu}(l, m) e^{-2\pi i(u_{ij}(t)l + v_{ij}(t)m)} dl dm, \quad (5-2)$$

where we have explicitly expressed the visibility and the spatial frequency coordinates as functions of antenna pair (i, j) and time t . The term $(ul + vm)$ in the

¹ There are many ‘centers’ in radio interferometry, and we attempt to define them here. The *pointing center* is the direction of maximum antenna gain. The *phase-tracking center* is the direction for which the fringes have been stopped—that is, a point source in this direction will produce a constant measured phase (except for the influences of the atmosphere and electronics). This is also the direction for which the (u, v, w) coordinates are calculated, which thus defines the origin of the (l, m) coordinates. The *delay center* is that direction for which the coherence is maximized by inserting delay into one element of an interferometer to compensate for the geometrical and instrumental differential delay. It is important to note that these various centers are independent. The pointing and delay centers are fixed at the time of observation, but the phase-tracking center can be changed through recomputation of the baseline coordinates and phase-shifting the visibility data accordingly. This process is also known as moving the tangent point.

exponential is the geometric phase difference $\Delta\phi_g$ produced by the differential path length between the radiation from the source located at (l, m) to each antenna, compared with a fictitious source at the *phase-tracking center* with the assumed baseline between the two antennas.

The total geometric phase difference ϕ_g , or geometric group delay τ_g , for an interferometer with baseline components (L_x, L_y, L_z) is

$$\phi_g = 2\pi\nu\tau_g = \frac{2\pi}{\lambda}(L_x \cos H \cos \delta - L_y \sin H \cos \delta + L_z \sin \delta). \quad (5-3a)$$

This equation follows directly from Equation 2-30 with the realization that the delay is simply $2\pi w$. Then, to first order, the differential geometric delay between radiation from another direction and the reference direction, or when the baselines differ from those assumed, is:

$$\begin{aligned} \Delta\phi_g = 2\pi\nu\Delta\tau_g = \frac{2\pi}{\lambda} & (\Delta L_x \cos H \cos \delta - \Delta L_y \sin H \cos \delta + \Delta L_z \sin \delta \\ & + \Delta\alpha \cos \delta (L_x \sin H + L_y \cos H) \\ & + \Delta\delta (-L_x \cos H \sin \delta + L_y \sin H \sin \delta + L_z \cos \delta)), \end{aligned} \quad (5-3b)$$

where

- $\Delta\tau_g$ is the differential geometric delay between the phase tracking center and the point of interest,
- (L_x, L_y, L_z) is the assumed baseline separation for the antenna pair (i, j) ,
- $(\Delta L_x, \Delta L_y, \Delta L_z)$ is the (true-minus-assumed) baseline,
- (α, δ) is the true source position, and
- $(\Delta\alpha, \Delta\delta)$ is the (true-minus-assumed) source position.

This relation is derived from Equation 2-31, using the astrometric coordinate grid shown in Figure 2-11. Note that an error in time is equivalent to an error in α .

Equation 5-3b is the basis for all interferometric analysis of antenna position coordinates and the basis for determination of astronomical positions.

2.2. Calibration formalism.

An interferometric array generates the observed visibilities $\tilde{V}_{ij}(t)$. In principle, the relationship between V and \tilde{V} can be arbitrary. Fortunately, however, adherence to sound engineering practices ensures that most arrays, to a good approximation, are linear devices: the output is a linear function of the input, or very nearly so. Furthermore, the individual measurements are well-isolated: the response associated with one antenna pair does not depend on the response of any other antenna pair. The basic calibration formula can therefore be written as

$$\tilde{V}_{ij}(t) = \mathcal{G}_{ij}(t)V_{ij}(t) + \epsilon_{ij}(t) + \eta_{ij}(t), \quad (5-4)$$

where

t is the time of the observation,
 $\mathcal{G}_{ij}(t)$ is the baseline-based complex gain,
 $\epsilon_{ij}(t)$ is a baseline-based complex offset, and
 $\eta_{ij}(t)$ is a stochastic complex noise.

Recall that the use of complex numbers is a convenience—it describes the combination of two correlator outputs (often termed the ‘Real’ and ‘Imaginary’, or ‘Cosine’ and ‘Sine’ correlator outputs) into one complex quantity. Thus the ‘complex offset’ and ‘complex noise’ are merely the complex resultants of the offsets and noises of two independent correlators.

In practice, the visibilities are averaged over a time interval during which they, or the complex gains, are not expected to change enough to perceptibly lower the coherence. Since the field-of-view of the antenna primary element limits the region from which signals are received, the maximum rate-of-change of phase will come from points at the edge of this region. It can easily be shown that an estimate of the maximum integration time in seconds is given by $\Delta t_{\text{int}} \sim 2D_{(\text{m})}/B_{(\text{km})}$, where $B_{(\text{km})}$ is the length of the baseline in km and $D_{(\text{m})}$ is the diameter of the antenna in meters. Longer time intervals are permissible if the angular extent of the source is less than the primary beam size. The permissible integration-time scales with the ratio of the primary beam size to the source angular size.

2.3. Editing.

Data editing can be considered a part of calibration, in the sense that there may be periods of time when \mathcal{G} cannot be determined accurately for some or all of the antenna pairs. These data should then be removed.

There are several forms of editing: editing out the gross errors obvious in the uncalibrated data represents one extreme; and editing subtle, or suspected errors, where it is uncertain whether inclusion or exclusion of the data is the better choice, represents the other. Various handy software display tools aid in discovering the faulty data.

2.4. Calibration methods.

Particular calibration methods depend on the detailed design of the array, the severity of the problem, and the ingenuity and desperation of the engineers and astronomers. Nevertheless, calibration methods can be divided into three basic categories.

- (1) *Direct calibrations:* The tolerances in the design of modern arrays must meet exacting specifications to ensure stable and linear operation. For example, the VLA is designed to have amplitude stability of better than 1%, and phase stability of better than one degree of phase per gigahertz of observing frequency. Where instabilities and variations cannot be avoided, special feedback circuitry and the monitoring of critical parameters within the array system and in the environment around the array can be used to correct

for these changes as the observations progress. Some examples are given in Section 4.

- (2) *Calibrator sources in the sky:* An interferometer measures phase differences, so there is no absolute phase reference. For any given observation, we wish to reference the visibility phases to the phase-tracking center, which is generally the same position as the center of the primary antenna beam. To determine the antenna phase offsets, observations of a sky calibrator are required. Further, if the array is not completely phase- or gain-stable, periodic observations of calibrators are used to monitor these changes. Finally, the atmosphere will cause time-variable phase changes to occur in the data (mimicking the effect of unstable electronics), and observations of a calibrator source are often made in an attempt to remove this effect. Calibrator observations are not concurrent or co-located in the sky with the actual observations.
- (3) *Self-calibration:* In some cases the source being observed can be used as a test signal to calibrate the instrument. This type of calibration requires strong signals and particular array properties. This is the subject of Lecture 9 and will not be discussed in much detail in this lecture.

2.5. Calibration of amplitude and phase.

A detailed look at the calibration Equation 5-4 and the use of calibrator sources will be given in Section 7. Most data corruption occurs before the signal pairs are correlated, so that the baseline-based complex gain $\mathcal{G}_{ij}(t)$ can be approximated by the product of the two associated antenna-based complex gains $g_i(t)$ and $g_j(t)$,

$$\mathcal{G}_{ij}(t) = g_i(t)g_j^*(t) = a_i(t)a_j(t)e^{i(\phi_i(t)-\phi_j(t))}, \quad (5-5)$$

where $a_i(t)$ is an antenna-based amplitude correction and $\phi_i(t)$ is the antenna-based phase correction. Observations of calibrator sources determine $\mathcal{G}_{ij}(t)$ for each of the $N(N-1)/2$ baselines, where N is the number of antennas. There are simple algorithms which then solve for the N values of $g_i(t)$.

2.6. Multi-frequency and dual-polarization capability.

Many systems have multi-frequency (widely separated frequency bands) and dual-polarization capabilities. Most calibration operations that deal with the internal array system must be performed separately for each frequency and polarization channel, since the respective propagation paths differ. Calibrations associated with the geometry of the array or with the troposphere and ionosphere either are frequency independent or scale with known parameters. Such calibrations need to be done once, at a convenient frequency or polarization.

It is often useful to measure the visibility function at many frequencies within a narrow frequency range. These are called *spectral line* observations. These require an additional calibration, discussed in Section 5, in which the relative complex gains between the channels are determined. Except for this calibration, all of the other calibrations are identical.

If both orthogonal polarizations are obtained at one frequency, then there are four visibility functions that sample the spatial coherence function of the complete electromagnetic radiation field (see Sec. 5.2 of Lecture 1). Most of the correlation will be in only two of these—those including the I Stokes parameter. These correlations are used to calibrate the array parameters and the system gain. The remaining correlations (which depend on the linear and circular polarization of the field) require a special calibration, called *polarization calibration*, if polarization imaging is to be done. This calibration is discussed in Section 9.

3. INITIAL CALIBRATIONS

Before usable data can be taken, many instrumental parameters must be determined. These include the antenna pointing, delays and positions. Another related parameter is the accurate position of the calibrator to be used for monitoring system gain. This section describes these calibrations. In almost all cases, the parameters describing these effects must be determined and applied before useful observing can begin.

3.1. Antenna pointing and gain.

The fundamental synthesis formulation of Equation 5-1 assumes that the primary beam distribution $\mathcal{A}_\nu(l, m)$ is independent of time and identical for each antenna. Small deviations occur, and some of the resulting effects can be corrected at later stages of observations. The antennas, of course, must be able to track accurately the diurnal motion of the source. Arrays composed of more than two elements with significantly different primary beam shapes can properly image only those sources which are small compared with the angular size of the primary beam, unless sophisticated imaging software is used.

For accurate imaging, the tracking of the center of the primary beam for all antennas must follow the intended sky position with an accuracy better than one-tenth of the full-width to half-power (FWHP) of the primary beam. Significantly larger errors can reduce the sensitivity of the observations and introduce distortions in extended objects. The 0.1 FWHP tolerance should be even smaller if a large part of the primary beam contains emission, either from one large-diameter source or from a high density of background sources.

The *antenna pointing error* is the difference between the actual pointing position (the location of the center of the primary beam) and the desired position. This difference has a complicated directional dependence because of misalignment of the polar or elevation axis, gravitational deformation of the structure, non-perpendicularity of the two axes, atmospheric refraction, and other effects (Clark 1973b). The antenna pointing error is also a function of time, because of differential heating of the structure. This pointing change can be estimated with tiltmeters placed on appropriate parts of the antenna structure, and it can be controlled to some degree by proper insulation of the antenna structure against solar heating. Wind-loading on the antenna also causes pointing errors, with time-scales of seconds. Most arrays shut down operation when the wind becomes excessive.

Surveying methods and optical alignment of the antenna axis with Polaris and other bright stars can determine the pointing accuracy on the sky to within

about ten arcminutes; better accuracy generally requires observations in many parts of the sky of radio sources of known position, to determine the directional dependence of the pointing errors. These are then fitted either to a physically-meaningful mathematical model describing the antenna defects, or simply to *ad hoc* trigonometric dependencies in the sky. If all of the pointing parameters are to be determined (typically five for each coordinate), then at least twenty observations, well-distributed around the sky, are needed to obtain sound estimates of the parameters.

Two types of observations are used to determine the pointing position from a calibration source, total-power observations and interferometric observations. Total-power observations have the advantage that they can be done by a single telescope observing bright extended sources. Interferometric observations have better stability, however. They also allow antenna-based pointing solutions, which provide greater accuracy.

The gain of an antenna decreases when observations are made near the horizon, because of deformation of the antenna structure and surface. Although it is often not a critical step in calibration, the dependence of gain upon zenith angle can be determined via the pointing error analysis. If the intensity of a calibrator source is known, then one can measure, as a function of zenith angle, the ratio of visibility amplitude to intensity. Often the antenna surface is designed so that maximum gain occurs at a zenith angle of about 30° . A decrease of more than a factor of two in the gain occurs only at high frequencies, near the limit of usefulness of the antenna.

3.2. Delay calibration.

Equation 5-2 is the *monochromatic* synthesis equation. Since modern arrays are designed to operate over large instantaneous bandwidths, the frequency response over the desired bandwidth must be coherent. Large bandwidths, of course, are required to obtain sufficiently high signal-to-noise from the weak celestial sources. For a finite bandwidth, the integrated visibility function is

$$V_{ij}(t) = \int_0^\infty \left(\int_{-\infty}^\infty \int_{-\infty}^\infty A_\nu(l, m) I_\nu(l, m) e^{-2\pi i \nu \Delta \tau_g} dl dm \right) e^{2\pi i \nu \Delta \tau_r} \mathcal{G}_{ij}(\nu) d\nu, \quad (5-6)$$

where

(i, j) denotes an antenna pair,

ν is the frequency,

$V_{ij}(t)$ is the visibility function integrated over the finite bandwidth,

$\mathcal{G}_{ij}(\nu)$ is the complex gain as a function of frequency,

$\Delta \tau_g$ is the differential geometric delay for the i - j baseline (delay relative to the delay-tracking center), and

$\Delta \tau_r$ is the residual instrumental delay for the i - j baseline (the error in inserted delay for the delay-tracking center).

If $\Delta \nu$ is used to denote the spanned bandwidth of the observations, then the phase difference $\Delta \phi$ between the ends of the band, that results from a net

residual delay $\Delta\tau_g - \Delta\tau_r$, is given by

$$\Delta\phi = 2\pi\Delta\nu(\Delta\tau_g - \Delta\tau_r). \quad (5-7)$$

A significant loss of coherence across the band will occur if $\Delta\phi$ is greater than about one radian. Thus, the delay between signals must be less than about $(160/\Delta\nu)$ nanoseconds, assuming $\Delta\nu$ is expressed in MHz.

Note that there are two different origins for a delay error. The first is geometric—only the delays for the delay-tracking center can be offset through insertion of delay in the system. Emission from other directions will suffer some loss of coherence. This the origin of the ‘bandwidth-loss’ effect noted in Lectures 2 and 13. The other origin is errors in the inserted delays, either due to mis-calculation or mis-calibration of the required delay, or due to the inevitable quantization of inserted delay.

Although the geometric delay can be precisely calculated, assuming knowledge of the antenna positions, a small residual delay error will generally remain, due to differences in signal propagation times through the different electronic paths. Note that these will vary between frequencies and polarizations within a single antenna, as well as between antennas. Delay calibration normally refers to the measurement of these small, but important, residual delays. The usual procedure is to observe a strong, isolated source of emission, and vary the delays in small steps until a maximum in the coherence is found. Another method, which in principle is much more sensitive, is to utilize the phase slope caused by a delay error (*cf.* Eq. 5-7). In this method, the delays are varied until the phase slope across the passband is zero. A spectral line correlator is required.

An often forgotten calibration is equalization of the delay between the two orthogonal polarization channels. Failure to correct for this delay will result in decorrelation of the cross-hand response, which normally provides the information on source polarization. Calibration of this delay is straightforward: observe a strong calibrator that is highly polarized (3C 286 or 3C 138 is commonly used), and adjust the delays until maximum signal is obtained on a cross-channel visibility function. This additional delay is the propagation difference between the two orthogonal polarization channels, and this amount should be added into all antennas of one polarization channel. It is best to do this after the parallel-hand delays have been set.

3.3. Time and place.

The fundamental formula (Eq. 5-1) depends on the direction of the radiation (time, since the Earth rotates) and on the separation of the two antennas (place). If the baseline separation of the two antennas is in error or the source is not at the phase-tracking center, or if there is an error in the time, then the observed visibility phase $\varphi_{ij}(t)$ will vary with time, as given in Equation 5-3. The time reference, antenna locations, and calibrator location must be sufficiently well-determined that the visibility phase does not vary by more than about a radian over a characteristic time-scale. For the purpose of calibration, this time-scale will be the time between calibration observations, say, ten minutes. (This is a rather short time—one would prefer to lengthen the interval in order to increase

integration time on the target source. Lengthening the time interval will make the baseline and time conditions more stringent.) A more fundamental time-scale is the basic integration interval, typically 10 to 30 seconds. Errors in these basic parameters which cause the phase to significantly change within this time window will cause serious error in the measured coherence. Another fundamental problem is in the calculation of the projected baseline (u, v, w) . Incorrect antenna positions or time will cause errors in (u, v, w) , and thence errors in the image due to incorrect gridding. The importance of this error will depend on the field-of-view and the required fidelity. See Section 3 of Lecture 16 for an estimate of the required accuracy. Finally, another ramification of a clock error will be erroneous pointing of the telescope. An error of a few seconds can cause the telescope to completely miss the source.

It is instructive to calculate the necessary accuracies in the baselines and in the time to keep the phase changes to less than a radian over 10 minutes. This condition requires the initial calibration of the baseline to be accurate to about four times the wavelength of the radiation, independent of the baseline length, or about 24 cm error at 5 GHz frequency. The accuracy of the time depends on the baseline length. For a frequency of 5 GHz, on a baseline of 10 km, a time error of 0.3 seconds will produce a phase error of about one radian over ten minutes. Smaller corrections are discussed in Section 7.7. Modern time-keeping can maintain the time to within a few milliseconds, equivalent to ~ 0.1 arcsecond. This is sufficient for coherent observations even with 30 km baselines at 24 GHz.

Calibrator sources with accurately known positions which are precessed accurately (see Clark 1973a, Clark 1982a) are observed over the sky in order to determine the antenna positions and timing error. For multi-element arrays, the data used are phase-referenced to a single antenna, whose position is assumed known. The errors $(\Delta L_x, \Delta L_y, \Delta L_z)$ in the baseline separations relative to this antenna, those in the source positions $(\Delta\alpha, \Delta\delta)$, and the time error Δt (which is equivalent to a zero-point shift in right ascension) can be determined by fitting the time and angle variations of these phases to Equation 5-3 (*cf.* Wade 1970; Brosche, Wade and Hjellming 1973). If the antenna locations are in error by many wavelengths, observations of closely spaced sources should be performed first. Alternatively, observations at a much lower frequency should suffice.

3.4. Amplitude check.

A trivial but important initial calibration is to compare the approximate visibility amplitude from a radio calibrator for all of the baselines. If the calibrator is unresolved, the amplitude should be about equal for all baselines. If one antenna consistently shows a low amplitude with all its baselines, check the r.m.s. noise while observing a weak source. If the ratio of signal to noise is normal, then the antenna is probably working properly, but with the amplification somewhat different. If the signal-to-noise ratio is more than a factor of five lower than that for other antennas, then that antenna must be malfunctioning, and the data may not be useful.

4. ROUTINE CORRECTIONS

The initial calibrations described above are generally applied on-line. Parameters describing the pointing characteristics of each antenna, the delay centering, the antenna locations, and the clock setting are inserted into the appropriate control computer. These allow the antennas to accurately track the source, and the phase and delay compensation for the phase-tracking center to be correctly calculated. However, these are not the only adjustments made while observing is in progress. Other corrections, which we call *routine corrections* are performed during observations. Although most observers do not deal with the routine corrections, these are important in determining the quality of the final image. These corrections include adjustments needed to track the phase and amplitude variation introduced by the electronics and signal transportation systems, and adjustments necessitated by the presence of the atmosphere. This section describes these adjustments.

4.1. Special array systems.

Most arrays use sophisticated internal sensing circuits to measure and correct temporal changes in the instrument as the observations proceed. Some of the most important examples follow:

- (1) Automatic Level Control: It is important that digital correlators receive a constant power level from each antenna. Most modern arrays use automatic gain control (AGC, or ALC) to maintain this condition. This is equivalent to making the receiver gain, when multiplied by the system noise temperature, invariant. However, the latter quantity may change for perfectly natural reasons, such as increased ground- or atmospheric pickup, or due to moving from one region of the sky to another. Under these conditions, the receiver gain will change, causing the resultant correlation coefficient to change. It is desirable to monitor the variations in receiver gain, and this is done by injecting a small amount of noise into the receiver and synchronously detecting the increase. These variations can then be used to correct the measured visibility. Gain changes outside of the receivers (e.g., antenna gain changes with elevation, and atmospheric attenuation) must be calibrated separately.
- (2) Round-trip phase measuring schemes: Section 7.2 of Thompson *et al.* (1986) describes several schemes designed to measure the phase path length along the cables and waveguide that carry the local oscillator signal from the master oscillator to each antenna. It is important to monitor the variations of this path, since variations of the phase of the local oscillator at each antenna will cause loss of coherence in the observed visibility if the changes are large enough. The path length changes are corrected in the observed visibility while the data are taken.
- (3) Tiltmeters on the axes of the telescopes can monitor changes in the pointing direction caused by wind and by thermal deformations. With sufficiently fast feedback loops, the pointing position of the

antenna can, in principle, be corrected to within a few seconds of arc (Dewdney 1987).

The VLA and other arrays continually monitor temperatures, voltages, and other parameters in all parts of the system. When some parameters are outside of normal operating ranges, automatic editing of the affected data can occur. As experience with an array grows, clever uses of the monitoring data can be devised to better calibrate and edit the data during the observations.

4.2. Externally determined parameters.

Some calibration parameters can be obtained from government monitoring programs or from other observatories whose instruments might be better suited to determining the relevant information. Weather-related parameters will be discussed in the next section. Earth-rotation and Earth axis orientation are monitored by concerted VLBI observations; up-to-date parameters are available from, e.g., Carter *et al.* (1985). Accurate positions of radio sources which are suitable for use as calibrators are also determined by VLBI observations. The longer-term motion of the Earth's axis, precession and nutation are encompassed in accurate precession algorithms that have been developed over the last several years. Smaller effects of Earth-tides, ocean loading, solar gravitational bending and motion of tectonic plates are now readily available.

The absolute gain of most arrays cannot be measured to better than 5% accuracy, but the relative gain stability is usually very much better than this. Thus, the relative flux densities of sources can be measured accurately, often to better than 0.5%. The absolute flux densities of the strongest sources are measured by single antennas with stable, well-determined gains. These values are adopted in most array work. An unfortunate problem is that most of the strongest sources whose absolute flux densities are known are also heavily resolved by modern interferometers. The commonly used flux density references are 3C 286, 3C 48, 3C 147, and 3C 295. Of these, only 3C 295 can safely be assumed to be non-variable.

4.3. Path length changes in the troposphere and ionosphere.

Propagation of radio signals through the Earth's atmosphere causes a modification of the phase of the signal due to refraction occurring within the medium. Accurate interferometry requires correction of the visibilities for this perturbation. The effects can be ascribed to two causes: the global, large-scale structure; and the turbulent, small-scale structure. The turbulent effects are difficult to determine from direct measurements—calibrator observations are only marginally effective—except through self-calibration techniques. Most of the large-scale effects can, fortunately, be predicted and corrected in real-time.

It is important to remember that the phase of an external signal measured by an interferometer on level ground is unaffected by propagation through a plane-parallel atmosphere. This is because such an atmosphere affects equally all signals propagating through it, so that although the antenna pointing must be corrected for the effects of refraction, the interferometer cannot sense the presence of the atmosphere. However, a real interferometer is not on level ground on a flat Earth. Phase corrections are required, because of two effects: First,

the antennas will be located at different heights. Second, the separated locations of the antennas mean that the objects observed will be seen at different elevations. Even though these differences are very slight, they are enough to require significant correction.

Refractive effects are usefully considered to have two origins, a dry component, about 6 km in thickness and a variable wet component, about 2 km in thickness. The decrease in the propagation speed of the radio signal is about one part in 3000, and independent of frequency up to about 300 GHz. This produces an additional time delay of the signal through the troposphere of about 8 ns (at the zenith), which is equivalent to an extra path length of 2.3 meters. From ground-based observations of the pressure P_{tot} , water-vapor partial pressure P_v , and temperature T , the excess propagation path can be roughly estimated. An approximate formula is

$$L = 0.228P_{\text{tot}} + 6.3w, \quad (5-8)$$

where

- P_{tot} is the total ground-level pressure in millibars,
- w is the vertical column water-vapor content above the array
(measured in centimeters), and
- L is the total zenith excess path in centimeters.

A typical value of P_{tot} is 1010 millibars (at sea level). The water-vapor content is highly variable, ranging from as low as 0.1 cm to perhaps 5 cm. These zenith excess-paths can be converted into a phase path length between two elements of an array at the zenith angle of the observations. For zenith distances less than about 70° , a $\sec z$ law can safely be assumed. For antennas separated by more than 100 km, in which case the tropospheric conditions are essentially independent, a simple difference of the two path lengths is taken. For antennas that are closer, as in connected-element interferometry, a good approximation of the differential excess path length ΔL is

$$\Delta L = \left(\Delta h + \frac{c\tau_g L}{r_0} \sec z \right) \sec z, \quad (5-9)$$

where r_0 is the radius of the Earth, Δh is the height difference between the two antennas, and τ_g is the geometric delay between the elements. The second term is typically 10^{-5} of the baseline length, or a few centimeters for baselines of a few kilometers. It obviously becomes much larger for low elevations, especially when the antennas have a large differential delay—e.g., when they observe at the same azimuth as the baseline. These corrections are usually made as the observations progress. At zenith distances greater than $\sim 70^\circ$, the curvature of the atmosphere cannot be neglected, and the dependence on zenith distance becomes more complicated.

Although the wet component of refraction is a small fraction of the dry component, the irregular distribution of water-vapor means that the highly variable phases which we see are dominated by this component. In the last few years,

radiometers have been built to measure the integrated water-vapor emission in small regions of the sky, so as to allow estimation of the wet-term from this integrated emission. See the discussion in Thompson, Moran and Swenson (1986) for further details. There is some hope that such emission measurements may produce estimates of the wet-term excess path accurate to a few millimeters.

The ionosphere is a magneto-active plasma in the region 60–2000 km above the surface of the Earth. It affects the propagation of radio waves in two ways: the signal is refracted by the ionosphere, and its plane of polarization is rotated (this effect is discussed in Sec. 9.3). The excess path L_i (in meters) in the zenith direction is

$$L_i = -40\nu^{-2}N_e, \quad (5-10)$$

where ν is the frequency in GHz and N_e is the electron column density in units of 10^{18} m^{-2} . The excess path length is negative because the phase is *advanced* with respect to a wave *in vacuo*. At 5 GHz using a typical daytime value for N_e of $3 \times 10^{17} \text{ m}^{-2}$, a path length difference of about 0.5 meter is obtained. A typical nighttime value of the electron density is about a factor of five less. The differential excess path ΔL_i between two antennas observing at zenith angle z is

$$\Delta L_i = \frac{L_i c \tau_g}{r_0 \cos^2 z + 2h}, \quad (5-11)$$

where h is the height of the ionosphere. An estimate of the electron column density in the ionosphere can be obtained from the average solar activity or can be measured using either an ionosonde or dual-frequency satellite transmissions.

The path length difference of the ionosphere is about the same as that of the troposphere at a frequency of about 1.4 GHz. Since the ionospheric refraction is dispersive (i.e., frequency dependent) it can be removed using dual-frequency observations.

4.4. Absorption by the troposphere and ionosphere.

Atmospheric opacity has two effects on visibility amplitude: First, the strength S of the source signal is decreased to $Se^{-\tau_a \sec z}$. Secondly, the atmosphere's emission adds to the system noise a contribution equal to $T_{\text{atm}}(1 - e^{-\tau_a \sec z})$, where τ_a is the zenith opacity of the atmosphere and T_{atm} is the average tropospheric temperature. Opacity can be determined by observing calibrators of known flux density at various elevations. However, the flux densities must be accurately known, the system gain must be very stable, and the dependence of antenna gain on zenith angle may cause confusion. Tracking a strong calibrator over a large range of zenith distance removes the first objection, but the second and third requirements remain. A better method is to use *tipping curves*, in which the total power from an antenna is monitored as the antenna slews from the zenith to the horizon. The dependence of total power on zenith angle provides a measure of opacity. Stable receivers are needed for accurate tipping curves (*cf.* Uson 1986).

An approximate estimate of the opacity can be obtained from ground meteorological measurements via the formula

$$\tau_a = \alpha_0 + \alpha_1 P_v, \quad (5-12)$$

where P_v is the water-vapor partial pressure in millibars. The following table lists α_0 and α_1 as functions of frequency.

Table 5–1. Coefficients to Estimate Atmospheric Opacity

Freq.	α_0	α_1
15 GHz	.013	.0002
22 GHz	.026	.015
35 GHz	.039	.0039
90 GHz	.039	.013

Ionospheric absorption of radio signals is minimal. At a frequency of 100 MHz during periods of high ionospheric activity, the absorption is about 2% and varies as the inverse-square of the frequency.

5. BANDPASS CALIBRATION

We have, so far, ignored finite bandwidths, except for the discussion of delay calibration in Section 3.2. There it was assumed that the observer was interested only in the integrated visibility over as wide a bandwidth as possible, in order to obtain the best signal-to-noise ratio. These are called *continuum* observations. It is acceptable to sum the emission over the bandwidth prior to correlation, provided that the delays are properly set (meaning the phase slope is very small), that there are no important changes in the visibility within the integrated band, and that the effects of chromatic aberration are unimportant.

However, these conditions may not be met, in which case it is necessary to generate the coherence as a function of frequency. In particular, many interesting radio sources contain atoms and molecules that emit most of their radiation in narrow frequency ranges, so it is important to image the source at a large number of adjacent frequency channels that span the width of the molecular emission, and with sufficient resolution in frequency to separate emission regions at different radial velocities. Or, in many cases, the effects of chromatic aberration cannot be ignored, due to the presence of strong background sources far from the phase-tracking center. In this case, observations must be taken with narrow bandwidths to allow a larger field-of-view. If the sensitivity is to be maintained, many observing bandwidths (or channels) are required. Such observations are called *spectral line* observations.

Delay calibration (Sec. 3.2) equalizes the propagation difference of the individual signal paths at the input of the correlator and removes the large phase gradient across the observing band that would otherwise occur. To handle changes of antenna gain with frequency, we can consider the baseline-based complex gain to be a function of frequency— $\mathcal{G}_{ij}(\nu)$. Compensating for the change of gain with frequency is called *bandpass calibration*. For a well-engineered system, variations of amplitude and phase across the bandpass will be small. Nevertheless, for many observations, the required channel-to-channel accuracy will be greater than can

be provided without bandpass calibration. This section briefly describes these calibrations.

The relative frequency response of the set of frequency channels can be determined by observing a strong calibrator source for sufficient time to reach the required accuracy. It is important that the spectrum of the calibrator be flat over the frequency band. It need not be a point source. Since the true visibility is identical in all channels for a calibrator, the bandpass complex correlator-based gain function for the i - j baseline and the k^{th} frequency channel is the observed visibility $\tilde{V}_{ij}(\nu_k)$, divided by the correct visibility $V_{ij}(\nu_k)$. The latter can be determined by observing a very strong source with a continuum correlator. The frequency-dependent, baseline-dependent gain can be factored into products of antenna-based gains, so that the set of baseline-dependent bandpasses can be converted into antenna-based bandpass amplitude and phase calibrations. The bandpass function is not a strong function of time and is determined relatively infrequently. In principle each significant change of frequency requires a new calibration, but the tolerance depends on the frequency characteristics of the array and the accuracy of the observations. Certainly the small frequency changes that are made to compensate for the rotation of the Earth are insignificant as far as the calibration is concerned.

Because of the narrow channel-width, accurate calibration of the antenna bandpasses should only be done using very strong calibrators. Fortunately, the time variation of these bandpass functions is very slow, so that for most applications, one calibration per observing run is sufficient. Calibration of other array parameters, such as temporal phase and gain fluctuations which affect all the frequency channels equally, is accomplished using the entire bandpass signal (i.e., summing the response over frequency). Again, because of the slow variation of the system parameters with frequency, this calibration can be done before the bandpass corrections are made.

In modern synthesis telescopes, the frequency separation is provided by a digital cross-correlation spectrometer, rather than by a set of narrow-band filters. The method and theory are discussed in Lectures 1 and 4. The output from the Fourier transform of the cross-correlated spectrum is a set of visibility functions at a grid of frequencies. The bandpass function can be obtained from the same calibrator observations described above. A minor problem introduced by the finite number of time lags in the correlator is that the ripples in the bandpass function depend on the input phase of the observed visibility! The real part of the visibility function is convolved somewhat differently than the imaginary part. This can be remedied by lowering the frequency resolution by smoothing (thus decreasing the amplitude of the ripples), or by increasing the number of frequency channels (which increases the number of measured time lags).

6. EDITING

When parts of an array malfunction, data of poor quality are recorded, and their inclusion in the imaging process will deteriorate the results. Even if it is possible to calibrate the severely affected data, it may have too much noise or be unstable. Data of poor quality are usually worse than no data at

all. It is impossible to display all of the data from modern synthesis telescopes. By carefully choosing a subset of representative data, it is usually possible to discover faulty data in nearly all of the data set, except in pathological cases.

Editing of array data is usually done in four steps. The first step was mentioned in Section 3.5—determining which antennas were working. This section describes the second step of editing: removing outliers and looking for inconsistent data. Part of this editing is done by the automated, on-line monitoring system and part by the observer looking over the data. The third step is done during the calibration process, when longer-term problems come to light and more subtle data problems may surface. The last editing step comes after the image is made. If this image is of poor quality (noisier than it should be or containing peculiar artifacts; see Lecture 16), then more data editing may be required or the calibration may be in error.

6.1. External monitoring.

Modern synthesis telescopes have internal checks which are intended to determine if the data being collected are of good quality. If it is clear that the data are in error (for example, if an antenna is not pointed at the source), then the system should not even record the data. If there is any question about data validity, then the data should be recorded and flagged bad, rather than discarded.

6.2. Scan consistency.

Most observations consist of scans where the visibility function is sampled every few seconds for a duration of several minutes. Since the visibility function for any source is a slowly varying function of time, relatively simple tests of the data consistency in a scan can be made. If the source visibility is limited by noise (each sampled data point $< 5\sigma$) then only bursts of interference can be detected in this manner. If each data point has good signal-to-noise ratio ($> 10\sigma$), then instabilities in the amplitude and phase can also be detected. The easiest test is to determine the r.m.s. spread of the data about the mean visibility (remember, it is a complex number) of the scan. Unless the source is strong, in which case slight gain instabilities will be the more important contribution to the r.m.s. scatter, it should be consistent with the expected noise. For a strong source, if the phase scatter from the troposphere is large, then it is better to look at the r.m.s. scatter of the visibility amplitude (rather than that of the complex visibility) to find discrepant points. But also keep in mind that any visibility variations of a strong source can easily be corrected through self-calibration. One wants to retain those data that can be corrected, and delete only those that are incorrigible.

The second method can be used if both polarizations at one frequency are simultaneously observed with a telescope; then the difference between the two parallel-hand visibility measurements should be consistent with noise. All atmospheric fluctuations should cancel. The difference in the phase of the two polarization channels is a good diagnostic of phase jumps in the system. Large weather-related phase changes will cancel, while phase jumps in the system will often affect the two channels differently.

Once a bad scan has been detected, it is relatively straightforward to find

the offending data point (or points) that produced the large r.m.s. scatter. Most errors are associated with an antenna, hence data on all baselines associated with that antenna at the relevant times must be edited out. Interference tends to occur only on short baselines.

It is important to be consistent about the editing. For example, if the first data sample for all calibrator observations is low because data were collected before the antennas reached the source, then you should assume that the same problem occurred for all sources, even though it is not possible to detect the decrease for the weak sources. If a phase jump occurred in an antenna during an observation interspersed between two calibrator sources, then, unless there is some way to ascertain the exact time of the phase jump, it is better to delete all data associated with that antenna for the intervening observations.

6.3. Data displays.

Displays using TV devices are useful for showing all of the data in a conveniently assimilated form, from which discrepant points can be recognized and discarded. The data are generally gridded into a two-dimensional matrix, with time as one axis and baseline the other. The amplitude is usually displayed as grey-scale intensity or color-coded in some convenient manner. The AIPS program 'TVFLG' provides this capability.

The most useful baseline ordering is that associated with one antenna, since almost all large perturbations of the observed visibilities are antenna-based. Such an ordering might be $(V_{12}, V_{13}, \dots, V_{1N}, V_{23}, V_{24}, \dots, V_{2N}, \dots, V_{N-2,N-1}, V_{N-2,N}, V_{N-1,N})$. Errors that are associated with any antenna should be obvious from the appropriate group. For sources that are extended and that have larger visibility amplitude at short spacings, ordering the baselines by increasing separation might be more useful. For noise-dominated sources, the grey-scale or color-coding should be chosen so that points greater than 5σ can be easily recognized. It should be convenient to use a cursor on the TV display to point to discrepant data points, which can then be automatically flagged in the data base.

The editing of spectral-line data is time-consuming. A simple scheme is to scrutinize and edit the associated wide-band channel and then edit all associated spectral channels. With a display in which time is represented on one axis and spectral channel on another, individual discrepant spectral data points can be easily found and flagged.

6.4. Shadowing.

When the projected spacing between two antennas is less than the diameter of the antenna, the radiation from the source to the far antenna is partially blocked by the near antenna. Although the geometrical blockage can be easily calculated, it is not generally safe to assume that the affected data can be corrected through an antenna-based multiplication, since the 'shadow' will be modified by diffraction effects. Furthermore, the blockage affects the primary beam pattern and the effective baseline length. Thus, a properly cautious procedure is to delete all data associated with the shadowed antenna. Be on the lookout for large, correlated signals, which can also occur under near-shadowing

conditions.

6.5. Long-term consistency.

Consistency of the data over long periods of time is difficult to determine before calibrating the data. Periodic observations of calibrator sources are important in determining long-term stability and discontinuous changes in the system. This will be discussed in Sections 7.8–7.10.

7. FINAL CALIBRATION USING RADIO SOURCES

The previous sections outlined initial calibration and editing of the visibility data, including the use of on-line monitoring. The final adjustment of the calibration is made by observations of radio sources in the sky. Calibrating the instrument using a radio source is an *ad hoc* method of calibration. It does not determine the cause of the calibration problem; it determines the complex gain of the entire system at a specific time and in a specific direction. It is therefore important to eliminate major directionally dependent calibration errors and short time-scale variations before using calibrator sources for temporal calibration. Calibrator sources are most useful for determining temporal variations over time-scales longer than about five minutes; most of these variations are associated with refraction variations over the array, caused by the troposphere and ionosphere.

For convenience, the complete calibration Equation 5–4 is reproduced here,

$$\tilde{V}_{ij}(t) = \mathcal{G}_{ij}(t)V_{ij}(t) + \epsilon_{ij}(t) + \eta_{ij}(t),$$

where

- t is the time of the observation,
- $\mathcal{G}_{ij}(t)$ is the baseline-based complex gain,
- $\epsilon_{ij}(t)$ is the baseline-based complex offset, and
- $\eta_{ij}(t)$ is stochastic complex noise.

In this section we will investigate how to fully exploit this relationship.

7.1. Calibrator source properties.

Radio astronomers are fortunate that there are hundreds of isolated, small-diameter radio sources with high flux density. These sources have ideal properties for a test signal: they are not significantly variable (over the time-scale of observation), they have accurately measured positions in the sky, their spectra are known and simple, they are strong enough to allow calibration in a short time, and they are isolated from nearby *confusing* sources. In short, the true visibility, $V_{ij}(t)$ is known for these sources; hence the various calibration gain terms in Equation 5–3 can be determined from the observed visibility, $\tilde{V}_{ij}(t)$. These *test signals* are especially useful because they traverse the entire array system, from the ionosphere and troposphere, through all of the electronics, to the sampling and recording devices. They have their limitations because calibrator observations cannot, in general, be made simultaneously and at the same location in

the sky as the source that is being calibrated. Thus, calibrator sources cannot calibrate gain variations that have time-scales shorter than about one minute, the typical time it takes to cycle the array between the calibrator and the source, or that have angular size-scales larger than a few degrees. A fortunate exception occurs if the target source itself is strong enough to calibrate the array (self-calibration), or there lies within the field-of-view a calibrator source (which is essentially the same thing as self-calibration).

7.2. The offset term.

The baseline-based offset term $\epsilon_{ij}(t)$ is generally negligible unless a correlator is malfunctioning, or unless there is significant cross-talk between the various antenna channels. The simplest method to determine this term is to observe a part of the radio sky that contains no emission and integrate many hours in order to decrease the contribution $\eta_{ij}(t)$ from the stochastic noise sources. An hour's integration time should suffice to find correlators that have significant offset problems. Twelve-hour integration is needed to determine offsets that might cause artifacts in the images made with many tens of hours of integration time. The measured visibility on each baseline is then an estimate of the offset term. Unfortunately, at low frequency and low resolution any random location in the sky will contain emission that can be detected in only a few minutes of integration and contaminate the measurement of any offset signal.

Two methods of reducing the presence of real signal are available. The first is to calibrate the data assuming the offset term is negligible and make an image of the emission in the field. Subtract the visibility function associated with the real structure from the observed visibility data and average the residual visibility data over time for each baseline. This will give a good measure of the offsets, which can then be subtracted from the original data.

If both polarizations are observed at one frequency, the difference in the offset term can be determined. For example, if the two channels are the right-circular polarization and the left-circular polarization, their difference, which measures the circular polarization of the emission, is insignificant for most radio sources. Residual visibility in the average signal from a correlator that is much larger than that expected from noise is a measure of the difference of the offset term between the two polarizations.

These methods assume that any offset signal is reasonably constant over many hours. If the variations are short-term, they are difficult to find; but they then act more like stochastic noise, which will average out to some extent. Once the baseline-based complex offset terms are determined, they should be subtracted from the observed visibility function or flagged from the database before any other calibrations are made.

7.3. Averaging of calibration observations.

In order to increase the accuracy of the computed complex gain, calibrator sources are observed for several minutes and a *suitable* average of the observed visibility is obtained. An unbiased estimate of the phase is determined by taking the vector average of the visibilities; an unbiased estimate of the amplitude is more difficult to obtain. Since the amplitude and phase fluctuations of most

instrumental effects are uncorrelated, separate averaging of the amplitude and phase produce the best estimate of each. Often the amplitude instability is less than one percent, while tropospheric turbulence produces visibility phase-winds of many degrees. Vector averaging of the visibility function will produce an amplitude that is low because of the decorrelation by the phase-wind. Under these conditions, an arithmetic average of the amplitudes alone is preferable. However, if the signal-to-noise ratio for the calibrator source for each time sample is less than about five, then this simple average of the amplitude will be increased by a noise bias (see Fig. 7-3 and related discussion). An arithmetic average is generally recommended, unless the signal-to-noise ratio is very low. In this case, the vector averaging time must be kept shorter than the decorrelation time-scale.

7.4. Baseline-based calibration.

The most straightforward use of calibrator observations is for determination of the complex gain \mathcal{G}_{ij} for each baseline. For the sake of simplicity, assume that the calibrator source is a point source of known flux density S and known position; thus the true complex visibility amplitude is S —i.e., the amplitude is S Jy, and the phase is zero degrees. At the time of observation of the calibrator, the complex baseline-based gain is

$$\mathcal{G}_{ij}(t) = \frac{\tilde{V}_{ij}(t)}{S} . \tag{5-13}$$

In words, the estimate of the gain is the observed complex visibility of the calibrator, divided by its flux density. The offset term $\epsilon_{ij}(t)$ is assumed to be negligible or already removed, and the noise $\eta_{ij}(t)$ is assumed to be negligible after proper averaging of the data in the scan. These assumptions are nearly always well-justified.

7.5. Antenna-based solutions for amplitude and phase.

In Section 2.5 we noted that most data corruption occurs before the signals are correlated, so it is convenient to write the baseline-based complex gain $\mathcal{G}_{ij}(t)$ as

$$\mathcal{G}_{ij}(t) = g_i(t)g_j^*(t)g_{ij}(t) , \tag{5-14}$$

where

$g_i(t)$ is the antenna-based complex gain for antenna i , and
 $g_{ij}(t)$ is the residual baseline-based complex gain.

Note that, if the baseline gain can be perfectly factored into a product of antenna gains, then $g_{ij} = 1$. For well-designed systems, this residual is within one percent of unity. This term is commonly (and unfortunately) called the ‘closure error’.

Equation 5-14 can then be separated into an amplitude and a phase to give

$$\mathcal{A}_{ij}(t) = a_i(t)a_j(t)a_{ij}(t) \tag{5-15}$$

and

$$\Phi_{ij}(t) = \phi_i(t) - \phi_j(t) + \phi_{ij}(t). \quad (5-16)$$

Let us represent the true and the observed visibilities as $V_{ij} = A_{ij}e^{i\varphi_{ij}}$ and $\tilde{V}_{ij} = \tilde{A}_{ij}e^{i\tilde{\varphi}_{ij}}$, respectively. Then, for a point-source calibrator of flux density S , we have $A_{ij} = S$ and $\varphi_{ij} = 0$. The calibration equations then become

$$\tilde{A}_{ij} = a_i a_j a_{ij} S \quad (5-17)$$

and

$$\tilde{\varphi}_{ij} = \phi_i - \phi_j + \phi_{ij}, \quad (5-18)$$

which can easily be solved for a_i and ϕ_i for all N antennas, provided that the ‘closure term’ g_{ij} is close to unity.

Several solution methods are available. For the phase, a linear least-squares solution can be used to estimate the antenna-based phases. Because the equations depend on differences between the antenna-based phases, one antenna phase is arbitrary, and is normally taken as zero. The antenna gains can be solved for by linearizing the gain equation through use of logarithms, then utilizing least-squares. This method, unfortunately, is susceptible to a biasing effect, due to noise—so that weak calibrators cannot be used to obtain reliable estimates of the gain amplitudes. A better approach is to solve the complex equation simultaneously for the amplitude and phase.

An estimate of the residual baseline-based error, often called the *closure phase* (unfortunately not the same as the structure-dependent closure phase), can be obtained from

$$\phi_{ij} = \tilde{\varphi}_{ij} - \phi_i + \phi_j. \quad (5-19)$$

This quantity should be less than one degree for a well-designed system. Similarly, the residual amplitude can be determined from

$$a_{ij} = \frac{\tilde{A}_{ij}}{a_i a_j S} \quad (5-20)$$

and should not exceed 1.01 for a well-designed system.

There are three reasons why antenna-based calibration values are preferred over baseline-based ones. First, most variations in the instrument are related to particular antennas, whether they arise from the medium above the antenna or the electronic components. The errors due to a well-designed correlator are always vastly smaller.

Secondly, since the number of baselines in an array of N elements ($N > 4$) is far larger than the number of elements, the computer capacity needed to store the calibration parameters and to apply them to the observed visibilities is reduced using antenna-based relations. Even if the baseline-based errors are significant, they are often constant with time—so a single determination of these errors can be used globally, to correct a large set of data.

Finally, antenna-based calibrations can be determined without the full set of baseline data. This is extremely important when using calibrators which are partially resolved, or confused by nearby sources. Such calibrators are often adequately approximated by a point source over a limited range of baseline lengths. Provided that all correlations between antennas are obtained, good calibration solutions can still be obtained in spite of stringent restrictions in the allowed range of spacings.

7.6. Imperfect calibrators.

Calibrators are, in general, not point sources. For high-resolution arrays, as in VLBI, all bright sources are resolved. At lower resolution and lower frequencies, many VLA calibrators can be considered point sources, but with extended emission present and with susceptibility to confusion from nearby sources. Partially resolved sources are still useful, provided the visibility is nearly enough constant over a sufficient range in baseline spacing. A good example of this is shown in Lecture 16.

If the calibrator source has a complex structure and cannot be approximated by a point source for any reasonable range of baselines, then self-calibration techniques are necessary to derive the antenna-based complex gain. The details of this process are given in Lecture 9. In that lecture the main impetus is to derive a good approximation of the source brightness distribution. However, the antenna-based complex gains must also be derived and it is these which are required for the calibration of the instrument. The calibration source should not be too complex, in order to avoid the ambiguities of determining both the structure and the complex gain.

7.7. Directional calibration—astrometry.

Although an important reason for frequent calibration observations is to determine the temporal changes of the array (particularly those caused by refraction in the troposphere and ionosphere), there is always a residual directional dependence of both the amplitude and phase of the gain. The effect of this dependence can be minimized if the calibrator source is close to the field that is being imaged, but it is best to determine the directional dependence as accurately as possible.

In most array operations, special calibrator observations are made to determine the directional dependencies, which are generally stable over periods of days. This often is done by observatory support staff, rather than observers themselves. Once these corrections have been determined, observers are free to concentrate their own calibration efforts on the problem of temporal variations (i.e., excluding directional effects). The typical observational scheme to determine the directional dependencies is to observe a large number of good-quality calibrator sources well-distributed over the sky, and packed into as short a period of time as possible—typically, about twenty observations in one hour. Observations at night and during clear, windless weather conditions, when tropospheric refraction variations are minimal, are the most productive. The sky coverage and the type of calibrator sources depend on the parameters that are to be calibrated. When carried to its extreme, this type of calibration is called *astrometry*.

and the following parameters can be determined:

1. Accurate locations of the antennas;
2. Antenna structure parameters (axis intersection displacement, for example);
3. More accurate positions of the calibrator sources;
4. Earth rotation (clock error);
5. Earth orientation (polar motion);
6. Zenith delay of the troposphere or the ionosphere;
7. Amplitude versus elevation dependence; and
8. Apparent displacement by the solar gravitational field.

When this type of observation is made periodically, changes in the above parameters can be monitored. Motion of the location of antenna which are on different tectonic plates are important geophysical data. Better values for the Earth precession, nutation, polar-motion traces, tidal motions, Earth-rotation changes, etc. can be determined.

Many of the above effects are of the order of milli-arcseconds and require very high-resolution arrays—particularly VLBI arrays, with antennas separated by many thousands of kilometers. Because these effects have similar directional dependencies, many observations of calibrators all over the sky are needed; complicated analysis methods are used, based on models similar to Equation 5–3, and embodying the geometric dependence of each relevant effect. When VLBI techniques are used, the delay (called *group delay* by VLBI'ers) is used rather than the phase delay, and a sophisticated form of the delay calibration of Section 3.2 is applied. The NASA software program known as 'CALC' makes initial estimates of these parameters, and the program 'SOLVE' then determines refined values from a set of antenna-based phases or group delays. See the VLBI lectures for more details.

7.8. Checking the calibration: closure errors.

An important step in calibrating and editing the data is to check the validity of the assumption that the complex gain of the observations can be approximated by antenna-based solutions—for each antenna, frequency and polarization. The soundness of this assumption is determined by the magnitude of the closure error. Some examples and suggested remedies are discussed below.

The closure errors are defined in Equations 5–19 and 5–20. They are, in essence, those residual amplitudes and phases for each baseline, averaged over the scan, which—if applied to the true visibility function after application of the best antenna-based gain—would reproduce the observed visibility function. For most arrays, amplitude closure errors should be less than 2%, and phase closure errors should be less than 2°.

If several baselines show significant, constant closure errors using different calibrator sources, then g_{ij} is different from unity and closure correction should be applied to all of the visibility measurements. If all of the large closure errors are associated with baselines to one antenna, the data from that antenna are suspect and should be closely scrutinized.

If closure errors are mainly associated with short baselines or with long base-

lines, then the calibrator source may not be sufficiently point-like. Structure in a radio calibrator can be confirmed if both polarizations at the same frequency show similar closure errors for the same baselines. Restricting the range of baselines can improve the solution quality. At last resort, self-calibration techniques that determine both the source structure and the antenna-based gain solution may have to be used.

Occasional large, anomalous closure errors on some baselines are indicative of interference or very discrepant data within a scan. Such data can usually be found by scrutiny of the individual points in the scan, for the relevant antennas or baselines. Interference spikes, extreme instability, contaminating signal from the Sun, or radar can be the cause.

Closure errors can be recognized only in calibrator observations, because their effects are subtle on a source with any degree of complexity. After the data problems have been determined, careful judgement must be used in applying closure error corrections derived from calibrator observations. For example, if just one antenna is producing large closure errors for many baselines for all calibrator data, it is best to flag all that antenna's data during the period of misbehavior. Under these conditions, it is likely the bad data affected the previously determined antenna gains, so it is wise to recalibrate.

Closure errors of a few percent are present in most observations at the VLA. They are caused by a variety of subtle problems: timing problems in the correlator, amplitude and phase variations over the frequency band, delay errors, and non-quadrature of the real and imaginary parts of the measured visibility function (e.g., Thompson 1980). All of these effects are stable over days. These closure errors produce artifacts in images at a level of about 5000:1. If you wish to image a very strong source then it is best to calibrate these closure errors as well as possible. There are several very bright point sources (3C 84 for example) that produce sufficient signal-to-noise ratio to allow determination of closure errors and their stability. Once determined, these closure corrections should be applied to the data before obtaining antenna-based calibrations or running the self-calibration algorithm.

7.9. Checking the calibration: amplitude stability.

During observing, many calibrator observations are made; afterward, the observer generates a table of the amplitude calibration of each antenna, for all frequencies and polarizations. If the array is operating well, the amplitude for any antenna should not vary over the observation period by more than a few percent; also, the amplitude ratio between antennas should be in the range 0.8 to 1.2. Problems frequently arise, however.

The flux density of some or all of the calibrators may be in error. Nearly all good point calibrators are variable, especially at high frequencies. More accurate relative flux densities for all of one's calibrators can be obtained from the ratio of the amplitude calibration from any or all of the antennas for the different sources. For example, if all of the antenna amplitude solutions for source A are 1.2 times greater than those for source B, then one can assume that the flux density of source A is a factor $\sqrt{1.2}$ less than that of source B. The absolute flux density

scale for the observations cannot be measured directly. Some radio sources have accurately-measured flux densities over a wide range of frequencies; one or more of these sources should be observed a few times during the run (3C 286, 3C 48, Planets, Planetary Nebulae).

Jumps or changes in the amplitude calibration for some antennas is an indication of real instabilities. Whether to flag these data depends on the type of experiment and the severity of the instability. If there is correlation of the instabilities in the two polarizations, then the problem may be due to shadowing or mis-pointing of the telescopes. The delay-zero may also be in error. If all of the antennas follow the same changes in amplitude, then residual gain vs. elevation problems (associated with sky absorption, antenna inefficiency, or ground pickup) may be occurring.

7.10. Checking the calibration: phase stability.

For most observations the phase stability is determined by the tropospheric and ionospheric refraction turbulence over the array, the most important component being the water-vapor distribution. How well frequent calibrations correct for this phase error in a nearby source will be discussed in Section 8. If both polarizations at one frequency are simultaneously observed, their phase difference should cancel all refraction variations; this is a good diagnostic of the system phase behavior. The difference should not vary by more than 5° over twelve hours. Many system phase jumps are different at the two polarizations and can be seen in the difference. If the antennas are separated by more than 100 km and the ionospheric refraction is large, then differential Faraday rotation of the two circular polarizations may be significant. See Section 9.3.

There are several tests of the phase variations that will indicate whether they are likely caused by weather. First, the fluctuations should roughly scale with frequency if they are tropospheric or scale with wavelength if they are ionospheric. Secondly, the fluctuations will scale with baseline length, b . For baselines less than 2 km the scaling goes as $b^{0.8}$ which is indicative of Kolmogorov turbulence. For baseline between 2 km and 100 km the scaling goes as $b^{0.3}$. For longer baselines the variations do not increase much. These relations are only approximate (Sramek 1983). Of course, the general tendency of the phase fluctuations should be correlated with the weather. Dry, windless days should be stable; periods of afternoon thunderstorms should be dreadful; nights are better than days; solar minimum is better than solar maximum.

Look for phase drifts in the antenna phase solutions over a period of several hours. If phase drifts are noticeable in one or a few antennas it is likely that the assumed antenna positions are in error. Check with the array operator in order to determine the status of the antenna position calibration. If many of the antennas show phase drifts that scale with baseline length from the reference antenna, then the source position is probably in error. A better position can be obtained from imaging or by use of Equation 5-3. If the phase drift occurs for all of the sources, then the array time standard may be in error.

7.11. Calibrator choice.

The strategy of picking the correct calibrator for the observations is discussed in Section 7 of Lecture 24. In general, the closer the calibrator is to the source the better: direction-dependent gain terms then are less important. Some improvement in the calibration of temporal fluctuations caused by the troposphere occurs with a calibrator only a few degrees away from the source, but most of the benefit is lost because of the non-simultaneity of the calibrator and source observations. The most conservative calibration scheme is to observe in the sequence calibrator–source–calibrator. If you are particularly concerned about direction-dependent errors, then the use of several calibrators surrounding the source (or two calibrators collinear with the source) is recommended. These extreme configurations are needed for VLBI observations if phase coherence is desired. The faster the cycling time between sources, the better. However, most of the time will be used for calibration, and this may make the signal-to-noise ratio for the source too low.

8. APPLICATION OF THE CALIBRATION VALUES

Depending on their origin, the calibration values are applied to the measured visibility function with different time-scales and in different ways. After a major reconfiguration of the array or a substantial change in the electronics, approximate parameters are introduced into the system (antenna locations, delay zeros, antenna pointing parameters)—generally in the on-line computer that controls the telescope, collects the data, and records it on tape. Since these values may be in serious error and contaminate the data quality, initial calibrations are required, as described in Section 3. The relevant parameters must be changed in the on-line computer before serious observing is done. Other direct calibrations (Sec. 4) are applied as the data are collected—for example, the path length monitoring of the local oscillator and the automatic level correction.

8.1. Gain tables.

Some calibrations are uncertain, so they should not be applied directly to the data unless in a form which can be undone. For example, correction of the path length due to the tropospheric delay, as determined by ground-based weather monitoring, is uncertain and may be revised at a later time. If these calibrations are relatively slowly changing they can be stored in what is called a *gain table*. This is a tabular array of numbers that contains the antenna-based amplitude and phase calibration values (or baseline-based values) at convenient intervals of time. Whether each calibration step has its own gain table or all of them are lumped into one table depends on the software system. But, it is important to be able to reproduce the values associated with any calibration—in order to undo its effect, if necessary. At the VLA, the practice is to have gain table entries corresponding to the beginning and end of each scan, and entries at ten-minute intervals within long scans. For VLBI observations, gain tables are usually produced at shorter intervals. The changes in the amplitude between gain table entries should not exceed more than about 25%; the changes in phase should not exceed more than about 30°—otherwise interpolation of the gain entries to intermediate times will be too much in error.

8.2. Interpolation within the gain table.

The calibration of the array using radio sources, discussed in Section 7, produces a table of antenna-based amplitude and phase calibrations for each antenna, frequency, and polarization at the mean time of the calibration scan. These values are separated by the time between calibrator observations, typically ten to forty minutes. How should these values be interpolated for times between calibrator observations? If the calibration variations are small, it matters little what interpolation scheme is used to transfer the values into the gain table. Similarly, if the calibration variations are very large, it also makes little difference, since the detailed fluctuations in the calibrator may not be highly correlated with those in the source. The recommendation here is to use a relatively long averaging time of the calibration values (running means over periods that include about three observations) and apply these smooth values to all the data. This scheme will calibrate the longer-term variations, and the residual errors left in the calibrator observations themselves will be a reasonable indication of the residual errors in the source observations. Regardless of the interpolation scheme, the amplitude and the phase should be interpolated separately; not the real and imaginary parts of the antenna-based gain.

8.3. Large phase-fluctuations.

On long baselines and in periods of poor weather there are large differences in the antenna-based phase between successive calibrator observations. Unless the phase change is consistent with a linear drift (look for a drift during the calibrator scan), it may be difficult to (sensibly) interpolate between the calibrator observations. Lobe ambiguities will occur if the phase change between two successive calibrator observations is near 180° . With this ambiguity, interpolation may give entirely the wrong result. So, in the extreme case we say that the array is *incoherent* over the time-scale of the calibration gaps. It may be coherent for shorter time-scales and for shorter baselines.

One method of sorting out possible lobe ambiguities is to determine the magnitude and direction of the phase difference between the two calibrator observations at short baselines. Since phase fluctuations are coherent over many kilometers, lobe ambiguities of long baselines may be resolved by the phase change at the shorter baselines. Similarly, if observations are made simultaneously at another frequency, the frequency scaling of tropospheric fluctuations from the lower frequency may be an aid. Fluctuations from the ionosphere can be handled somewhat differently and are discussed in Section 10.

9. POLARIZATION CALIBRATION

As shown in Lecture 1, the complete state of the radiation field is most conveniently described by the superposition of two orthogonal polarizations. The complete spatial coherence of this vector field is described by four correlations. For example, if the dual-polarization feeds measure the right circular polarization (R) and the left circular polarization (L), then the four visibility functions are $R \star R$, $L \star L$, $R \star L$, and $L \star R$. The previous discussion concerned the calibration of the parallel hand correlations, $R \star R$ and $L \star L$. With these two

visibility functions, the complex gains of the two polarization channels can be calibrated. If the dual-polarization channels were precisely orthogonal (that is, if the voltage signals from the channels were precisely proportional to the orthogonal electric fields), then the cross-hands would also be calibrated, except for any phase difference between them and for possible correlator-based problems. In practice, however, this proportionality is not obtained, and further calibrations are necessary. This section describes the additional calibrations needed for the crossed-hand correlations, $L \star R$ and $R \star L$. For notational simplicity, we have assumed the parallel-hand calibration has been accomplished.

9.1. Stokes parameters.

Section 5.2 of Lecture 1 has described the polarization matrix and its relation to the Stokes' parameters, I , Q , U , and V . These are four real numbers describing the polarization state of radiation from a given region of sky, for a given frequency. Note that Q , U and V can be negative. It is quite correct to consider these as four images which completely characterize the polarization state of the radiation from the sky. Each of these has an associated visibility function, which we denote V_I , V_Q , V_U , and V_V . These cannot be measured individually, but appear as linear combinations of the four correlations produced by the interferometer. From these combinations, the four visibility functions can be obtained. With the use of R and L polarized feeds, the four visibility functions are

$$\begin{aligned} V[R \star R] &= V_I + V_V, \\ V[L \star L] &= V_I - V_V, \\ V[R \star L] &= (V_Q + iV_U) e^{-2i\chi}, \\ V[L \star R] &= (V_Q - iV_U) e^{2i\chi}, \end{aligned} \tag{5-21}$$

where χ is the parallactic angle, which determines the orientation of the feed with respect to the sky. Other combinations are given by Thompson *et al.* (1986, Sec. 4.9).

9.2. Leakage terms.

The feeds are not exactly orthogonal; that is, the antenna-feed combination causes a small amount of right-circular polarization to show up in the left channel, and *vice versa*. This 'leakage' is defined in the following way: If the provided voltages are v_R and v_L from the RCP and LCP channels, and the true electric fields are E_R and E_L , then $v_R = E_R e^{-i\chi} + D_R E_L e^{i\chi}$ and $v_L = E_L e^{i\chi} + D_L E_R e^{-i\chi}$. Note that the voltages are considered to be complex, corresponding to what is called the 'analytic signal' in the engineering literature. The indicated multiplications can be carried out, and assuming the leakage terms, and the circular and linear terms, are small so that products of them may be ignored, the four correlations can be expressed as (Bignell and Napier 1978):

$$\begin{aligned} V[R \star R] &= V_I + V_V, \\ V[L \star L] &= V_I - V_V, \\ V[R \star L] &= e^{-2i\chi}(V_Q + iV_U) + (D_{R1} + D_{L2}^*)V_I, \\ V[L \star R] &= e^{2i\chi}(V_Q - iV_U) + (D_{L1} + D_{R2}^*)V_I. \end{aligned} \tag{5-22}$$

For notational clarity, we have left off a phase rotation $e^{\pm i(\phi_R - \phi_L)}$ that should multiply the final two equations. This term is due to the arbitrary phase difference that occurs between the two orthogonally polarized channels. This phase difference is easily determined from observations of a strongly polarized source of known position angle, as described in the next section. χ is the relative angle of the feed orientation with the sky. For an equatorially mounted antenna, the parallactic angle is constant unless the feeds are purposely rotated. For an alt-azimuth mounted antenna, the parallactic angle is a calculable function of time and position: $\tan \chi = \cos \phi \sin h / (\sin \phi \cos \delta - \cos \phi \sin \delta \cos h)$.

The leakage terms D can be calculated by observing an unpolarized calibrator point source. In this case, the observed visibility in the cross-hand channels is a sum of two leakage terms. Using all of the baselines, the two leakage terms per antenna can be obtained. In general, the calibrator polarization is not zero, and must be included in the solution. For an alt-az antenna, both the source polarization and antenna polarization can be simultaneously calculated, since the change $\Delta\chi$ in parallactic angle causes the source and antenna polarizations to differentially rotate. The varying sum can be used to determine each term. The accuracy of this separation is dependent upon the strength of the source and the range of parallactic angle.

9.3. Phase difference between polarization channels.

The normal parallel-hand phase calibration calibrates the phase of one polarization channel of each antenna with respect to the phase of that same channel of the reference antenna. This is done independently for each polarization channel. In general, there will then remain a phase difference *between* the two polarization channels. The parallel hand calibration ensures that this difference will be the same for each antenna, and equal to that of the reference antenna. Since the crossed-hand phase carries information on the visibilities of Stokes' parameters Q and U , it is clear that this instrumental phase difference must be removed.

The phase difference of the reference antenna can be measured by observing a strongly linearly polarized radio source with known parameters $Q + iU$. Since we need only the phase difference for one antenna, only one baseline is needed; a choice of one of the shortest baselines will permit the use of highly polarized calibrators, which tend to be somewhat resolved.

In the ionosphere, the presence of a magnetic field causes the refraction to be different for right- and left-circularly polarized radio waves. This produces a rotation of the plane of the linearly-polarized signal and, equivalently, a phase rotation between the R polarization channel and the L polarization channel for the reference antenna. Since the phase calibration of the other antennas has been tied to the reference antenna, the difference is uniform over the array. If this differential Faraday rotation changes over the array, as it will for VLBI arrays, then the R and L phase calibrations for these antennas will differ. During periods of high solar activity the phase rotation can be as large as 20 rad m^{-2} , which produces an angle of 50° at 1.4 GHz. The angle scales with the square of the wavelength.

Ionospheric models are available to predict the density of electrons and the

strength of the magnetic field, from which estimates of the differential Faraday rotation can be made. Abrupt changes occur at sunrise and sunset, and significant changes occur during the day with time-scales of about one hour. For best results, at least once an hour observe a highly polarized calibrator source not more than about 20° from the source.

10. ADDITIONAL TOPICS

A *potpourri* of topics follows in this section. Some deal with fairly subtle and unusual calibration practices but are worth mentioning. Others deal with calibration of the entire field-of-view of the observations. Nearly all of the discussion here and elsewhere in the lecture deals only with the central region of the primary beam.

10.1. The antenna primary beam.

In Section 3.1 the calibration of the antenna pointing was discussed. Since the visibility function measures the spatial coherence function of a fictitious source $\mathcal{A}(l, m)I(l, m)$ (see Eq. 5-1), the effect of the primary beam sensitivity, $\mathcal{A}(l, m)$, must be removed from the image. Although a good theoretical estimate of \mathcal{A} can be made of the central region, above a relative sensitivity of 20%, direct observations are needed to determine \mathcal{A} in the regions $< 20\%$ and in the sidelobes. Observations are made of a very strong calibrator. Assuming all of the antennas are identical, you choose a convenient antenna pair, point one directly at the source and mis-point the other over the area of sky in which you want to determine the primary beam sensitivity. You take occasional observations with the second telescope pointed directly at the source, in order to determine variations of the antenna amplitude and phase with time. The complex ratio of the visibility measured at any desired location in the sky to visibility measured on-axis is the normalized reception *voltage* pattern. If two polarizations are observed, this procedure should be done for all four polarization correlations.

An interesting sidelight is that the Fourier transform of the complex relative visibility function, as a function of distance from the phase center, gives the distribution of the electric field on the aperture plane of the antenna. Departures of the phase from zero can be interpreted as surface deviations from a true paraboloid and thus measure the accuracy of the surface. This technique is commonly used for surface measurements of large telescopes.

The absolute value of the voltage pattern is $\mathcal{A}(l, m)$. The image made from the Fourier transform of the calibrated visibility function must be corrected for the primary beam sensitivity in order to determine the distribution of intensity in the sky. This correction becomes crucial when very large sources that cover several primary beam areas are to be combined into one image (see Lecture 15), when comparisons at widely different frequencies are made, and when the polarization properties of large sources are of interest.

Several other problems associated with the antenna pointing and finite size occur and will be briefly mentioned:

1. *Beam-squint between polarizations:* For off-axis feed systems like those of the VLA and VLBA, there is a slight displacement of about 0.05 FWHP

between the RCP and LCP electrical pointing centers. For accurate images of very extended sources, the primary beam correction of each must be done separately. The image difference between the two parallel hands will often show false circularly polarized sources where the two primary beam corrections differ.

2. Polarization characteristics across the beam: The calibration of the leakage terms D , described in Section 9, can be made at each location in the primary beam. It is found that the D 's substantially change outside of the 50% sensitivity region of the beam. Thus the relative linear combinations of the four Stokes parameters are mixed-up across the beam, invalidating the basic synthesis equation. And thus, outside of this area, it is difficult to reliably measure degrees of linear polarization lower than about 5%.

3. Alt-azimuth antenna mount: For the VLA and the VLBA the alt-azimuth mounting produces a rotation, measured by the parallactic angle χ , of the antenna relative to the sky. Unless the primary beam response is perfectly circular (i.e., identical along all radii), extended observations will cause an effective change in the primary beam sensitivity across the source. Such non-circularity is more pronounced for the cross-hand correlations than for the parallel-handed. A (tedious) solution is to image only that part of the data in which the parallactic angle has not changed by more than, say, 10° , correct for the instantaneous primary beam sensitivity, repeat, and then sum up all the corrected intensity distributions.

4. Different antenna in the array: In general, the synthesis properties of an array with more than two elements is invalidated if the antennas are dissimilar. This is because the effective primary beam sensitivity is different for different baselines and the sampled coherence function is not uniquely associated with one intensity distribution. For two-element arrays, the product of the voltage primary beam patterns applies. Many VLBI arrays are composed of widely different diameter antennas; since all of the sources under study are extremely small in angular size, they are all virtually at the beam center, so the details of the primary beam patterns are irrelevant.

10.2. Bandwidth smearing.

When observing over large instantaneous bandwidths, distortions occur in parts of the image far removed from the phase center. This effect, *chromatic aberration*, is discussed in detail in Section 2 of Lecture 13. As a rule-of-thumb, chromatic aberration becomes significant when the offset, measured in units of synthesized beamwidths, and multiplied by the fractional bandwidth, is of order unity. It produces a radial smearing whose shape is the effective bandpass convolved with the source structure. Deconvolution techniques, generalized to recognize this smearing, can be used to obtain an estimate of the intensity distribution. The smearing can be avoided only by separating the observations into a set of narrower-bandwidth channels by using filters, or by cross-correlation techniques in which many delay centers are used (see Sec. 5).

10.3. Dual-frequency ionospheric calibration.

Over most of the radio frequency spectrum, the ionospheric refraction varies as the wavelength squared, so the effect of the ionosphere on the visibility phase

is proportional to the wavelength. If a point source is observed, or the structure of the source is known or can be assumed to be invariant between two frequencies, the effects of the ionosphere on the phase can be removed (Fomalont and Sramek 1977). Given two observations of the source phase, ϕ_1 and ϕ_2 at frequencies ν_1 and ν_2 , the corrected phase ϕ_c , which would have been measured in the absence of the ionosphere, is:

$$\phi_c = \frac{\nu_2 \phi_2 - \nu_1 \phi_1}{\nu_2 - \nu_1}. \quad (5-23)$$

The corrected phase is the sum of the true visibility phase and an instrumental phase. The latter can be determined from an observation of any calibrator. This technique is used for geodetic VLBI, with the delay replacing the phase. Common dual-frequency pairs are (2.3 GHz, 8.1 GHz) and (1.3 GHz, 1.6 GHz).

10.4. Relative vs. absolute calibration.

If the radio source is strong and the signal-to-noise ratio of the visibility is greater than about five (for each integration time and for each baseline), then self-calibration techniques will likely succeed. In principle, there is no need to calibrate the data at all. It is advisable to observe a very strong source to determine the closure errors, and blank sky to determine the offset terms. If polarization information is required, proper calibration of the leakage terms is needed, using an unpolarized source and observations of the polarized source to calculate the differential Faraday rotation. An observation of a flux density standard is necessary to determine the flux density scale.

Alternating the observations between the source and a nearby calibrator aids in removing some of the tropospheric refraction errors and often produces an image with dynamic range $> 100:1$. The position of the source then is known with respect to that of the calibrator, but with some residual error caused by any long-term directional dependencies of the phase. Even if self-calibration techniques will be useful, this normal calibration will produce a good first image that may be then used for further self-calibration. When one uses the no-calibration scheme described in the last paragraph, a good first estimate of the source structure is not available: this is the usual case in VLBI observations. If the source is very complicated and the u - v coverage is poor, then ambiguities in the self-calibration process may become important, especially in the absence of a good starting model of the source. Relative motion of the source, or of parts of the source, over long periods of time can be determined if the same calibrator is used for all observations. Registration of images made at different frequencies, at different times, and with different configurations are made easier by using the same calibrator source. If the radio field contains many radio sources, then the images at different epochs can be registered by aligning the background sources in the final images, in order to determine relative motion of the source of interest.

True absolute measurements of radio source positions or antenna positions require careful calibration of the directional dependence of the phase, as described in Section 7.7. The origin of right ascension is a fundamental problem. Full use of Equation 5-3, with many additional terms, is needed. Often, corrections are made to the zenith path delay as a function of time.

REFERENCES

- Bignell, R. C. and Napier, P. J. (1978), "Preliminary polarization calibration of the VLA at 6 cm", VLA Test Memorandum No. 125, NRAO.
- Brosche, P., Wade, C. M., and Hjellming, R. M. (1973), "Precise positions of radio sources. IV. Improved solutions and error analysis for 59 sources", *Astrophys. J.*, **183**, 805–818.
- Carter, W. E., Robertson, D. S., and MacKay, J. R. (1985), "Geodetic radio interferometric surveying: application and results." *J. Geophys. Res.*, **90**, 4577–4587.
- Clark, B. G. (1973a), "VLA telescope pointing analysis", VLA Computer Memorandum No. 104, NRAO.
- Clark, B. G. (1973b), "Ephemeris routines for the VLA—specification considerations", VLA Computer Memorandum No. 105, NRAO.
- Clark, B. G. (1982a), "The new IAU precession procedures", VLA Scientific Memorandum No. 145, NRAO.
- Dewdney, P. (1987), "Investigating the use of tiltmeters to correct VLA antenna pointing—Part I. Initial measurements and analysis", VLA Test Memorandum No. 148, NRAO.
- Fomalont, E. B. and Sramek, R. A. (1977), "The deflection of radio waves by the Sun", *Comments on Astrophys.*, **7**, 19–33.
- Sramek, R. A. (1983), "VLA phase stability at 22 GHz on baselines of 100 m to 3 km", VLA Test Memorandum No. 143, NRAO.
- Thompson, A. R. (1980), "Closure accuracy and related instrumental tolerances", VLA Electronics Memorandum No. 192, NRAO.
- Thompson, A. R., Moran, J. M., and Swenson, G. W., Jr. (1986), *Interferometry and Synthesis in Radio Astronomy*, John Wiley & Sons, New York.
- Uson, J. (1986), "Atmospheric opacity at the VLA", VLA Scientific Memorandum No. 157, NRAO.
- Wade, C. M. (1970), "Precise positions of radio sources. I. Radio measurements", *Astrophys. J.*, **162**, 381–390.

6. IMAGING

RICHARD A. SRAMEK AND FREDERIC R. SCHWAB

1. FOURIER TRANSFORM IMAGING

A fundamental result of Lectures 1 and 2 was the existence of a Fourier transform (FT) relationship between the sky brightness I , the primary beam pattern \mathcal{A} , and the visibility V observed with an interferometer. From Lecture 2 (Eq. 2-26),

$$\mathcal{A}(l, m)I(l, m) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} V(u, v) e^{2\pi i(ul+vm)} du dv. \quad (6-1)$$

This simple relation holds if (a) $|\frac{\Delta\nu}{c} \mathbf{b} \cdot (\mathbf{s} - \mathbf{s}_0)| \ll 1$ and (b) $|w(l^2 + m^2)| \ll 1$. These conditions are met whenever the radiation to which the interferometer pairs respond originates in a suitably small (and confined) region of sky. Since the correction for the primary beam can be made trivially at the final stage of data processing¹ (as discussed in Lecture 1, Sec. 4.4), we shall use $I(l, m)$ to denote the *modified sky brightness*, $\mathcal{A}(l, m)I(l, m)$.

V is complex-valued and, after the usual calibration steps (see Lecture 5), is reckoned in units of flux density (say, $\text{W m}^{-2} \text{Hz}^{-1}$), while I has units of surface brightness (flux density per unit of solid angle). A standard unit for I is Jy/beam area; sometimes Jy per square arc second is used instead. The units are determined by the normalization of Equation 6-1.

Equation 6-1 is used to obtain an estimate of the modified sky brightness from the observed visibilities, recorded at u - v points (u_k, v_k) , $k = 1, \dots, M$. In practice, M may range from ten to a few hundred with a two element interferometer, to over a million with a multi-element array like the VLA. With M small, model fitting is feasible—and sometimes useful (see Lecture 11). But for large M the usual method of estimating I is via the discrete Fourier transform (the DFT), because extremely efficient algorithms are known for numerical evaluation of DFTs.

The topics of some of the lectures to follow also fall under the broad category of ‘imaging’. But the discussion here is restricted to ‘simple-minded’ methods of estimating the sky brightness: that is, *directly* approximating the right-hand side of Equation 6-1, via only *linear* operations. The so-called ‘dirty image’ that results is a discrete approximation to I^D , where (from Lecture 1, Eq. 1-10)

$$I^D(l, m) \equiv \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} S(u, v) V'(u, v) e^{2\pi i(ul+vm)} du dv. \quad (6-2)$$

¹ This is assuming that \mathcal{A} has been carefully measured over a large enough region in (l, m) . Wide-field imaging, in cases in which a source covers, say, a larger region than the central lobe of the primary beam, is an especial problem. Antennas with azimuth-elevation mounts (as at the VLA) present a problem because the primary beam patterns rotate on the sky, as functions of parallactic angle. See Lecture 5.

Here, S denotes the u - v sampling function and V' the observed visibility; the prime indicates that the visibility data are noise-corrupted measurements. (For conciseness, I^D has been left unprimed, but it too is noise-corrupted whenever V is.)

1.1.1. The ‘direct Fourier transform’ and the FFT.

Either of two methods is commonly used to numerically approximate the Fourier transform in Equation 6–2. The first, called the ‘direct Fourier transform’ method,² approximates $I^D(l, m)$ by brute-force evaluation of the sum

$$\frac{1}{M} \sum_{k=1}^M V'(u_k, v_k) e^{2\pi i(u_k l + v_k m)}. \quad (6-3)$$

If this ‘direct Fourier transform’ is evaluated at every point of an $N \times N$ grid, the number of real multiplications required is $4MN^2$ (the number is halved, though, assuming Hermitian data). In practice M is usually of the same order as N^2 , so the number of multiplications goes roughly as N^4 . The number of sine and cosine evaluations required is also $\mathcal{O}(N^4)$, as is the number of additions/subtractions.

The second method requires interpolating the data onto a rectangular grid, so that a fast Fourier transform (FFT) algorithm can be used. The process of interpolation is referred to as *gridding*. (Gridding may require sorting the data into order of decreasing $|u|$ or decreasing $|v|$.) The number of elementary arithmetic operations required by the technique most often used for gridding is $\mathcal{O}(M)$. The number of such operations required by an FFT algorithm (say, the Cooley–Tukey algorithm) is only a few times $N^2 \log_2 N$ —not $\mathcal{O}(N^4)$! This saves much computing time for large databases, and large N especially, if an economical method of interpolation is used. However, for making small images (i.e., for N small) from small databases (M small), the ‘direct Fourier transform’ may be faster than the combination of gridding and FFT.

In the following sections we first discuss weighting and selection of u - v data and how it affects the resulting images. This applies no matter how the Fourier transform is approximated. Then we touch upon the problems that are introduced by gridding the data to permit use of the FFT—the problems of aliasing and correction for gridding.

² This choice of terminology is unfortunate. The natural abbreviation for the term—‘DFT’—is used almost universally (by everyone except radio astronomers) to stand for something else: the ‘discrete Fourier transform’. For example, the 2-D discrete FT of an $M \times N$ matrix (x_{ij}) is the $M \times N$ matrix (y_{kl}) given by

$$y_{kl} = \sum_{p=1}^M \left(e^{2\pi i(p-1)(k-1)/M} \sum_{q=1}^N x_{pq} e^{2\pi i(q-1)(l-1)/N} \right).$$

The major distinction between the two usages is that in one case the data are regularly spaced, and in the other they are not. Also, the ‘direct FT’ is generally not invertible, whereas the ‘discrete FT’ is invertible; usually, of course, the term ‘transform’ is reserved for invertible transformations.

2. THE SAMPLING FUNCTION, AND WEIGHTING THE VISIBILITY DATA

The sampling function S and its Fourier transform, the synthesized beam B , were introduced in Lecture 1. In practice, the data are variously weighted, according to their reliability and to control the shape of the synthesized beam.

2.1. The sampling function.

S is a ‘generalized function’, or ‘distribution’, which may be expressed in terms of the two-dimensional Dirac delta function, or ‘ δ -distribution’,

$$S(u, v) = \sum_{k=1}^M \delta(u - u_k, v - v_k). \tag{6-4}$$

It is useful to introduce a second generalized function, called the *sampled visibility function* or, alternatively, the *u - v measurement distribution*,³

$$V^S(u, v) \equiv \sum_{k=1}^M \delta(u - u_k, v - v_k) V'(u_k, v_k). \tag{6-5}$$

That is, $V^S = SV'$. Let \mathbf{F} denote the Fourier transform operator. Equation 6-2 can be rewritten

$$I^D = \mathbf{F}V^S = \mathbf{F}(SV'). \tag{6-6}$$

By the *convolution theorem*, which says that the Fourier transform of a product of functions is the convolution of their FTs (see, e.g., Bracewell 1978),

$$I^D = \mathbf{F}S * \mathbf{F}V', \tag{6-7}$$

where $*$ denotes convolution. For a point source of unit strength, centered at position (l_0, m_0) , $|V'(u, v)| \equiv 1$, and $\mathbf{F}V'$ is the (shifted) Dirac δ -function: $\delta(l - l_0, m - m_0)$. So the point source response of the array, i.e., the *synthesized beam*, is given by $B = \mathbf{F}S * \delta = \mathbf{F}S$. Equation 6-7 is the familiar result (Lecture 1, Eq. 1-11) that the observed brightness is the true brightness convolved with this ‘beam’.

It should be apparent that the so-called ‘direct Fourier transform’, as defined by Expression 6-3, is *exactly* I^D . That is to say, that—assuming δ -function sampling— $I^D(l, m)$, as defined by Equation 6-2, is given exactly by a discrete summation, Expression 6-3, and that Equation 6-7 holds for the ‘direct Fourier transform’ method (an analogous relation is given below for the FFT method). Of course, a computed ‘direct Fourier transform’ image is indeed an approximation, but only in the sense that it is inevitably a discretely sampled version of I^D and that the sums are computed in finite precision arithmetic.

³ Note that the visibility measurements are not, in actuality, point samples of the inverse Fourier transform of the modified sky brightness $\mathcal{A}I$, but that instead they represent *local averages* of it. Time- and frequency-averaging, which are discussed in Lecture 2, are the dominant averaging effects. One should try to choose observing parameters (integration time and bandwidth) that make relatively safe our assumption here about δ -function sampling. This matter is further discussed in Lectures 12, 13, and 24.

2.2. Weighting functions for control of the beam shape.

In analogy to Equation 6-4, a *weighted sampling function*, or *weighted sampling distribution*, can be written as

$$W(u, v) = \sum_{k=1}^M R_k T_k D_k \delta(u - u_k, v - v_k). \quad (6-8)$$

And, in analogy to Equation 6-5, one can define a *weighted, sampled visibility function*, or *weighted and sampled measurement distribution*, V^W according to $V^W = WV'$, or, explicitly,

$$V^W(u, v) = \sum_{k=1}^M R_k T_k D_k \delta(u - u_k, v - v_k) V'(u_k, v_k). \quad (6-9)$$

The coefficients R_k , T_k , and D_k are weights assigned the visibility points. These data points may represent time-averages of visibility measurements spaced along the loci of the interferometer u - v tracks. R_k is a weight that indicates the reliability of the k^{th} visibility datum. It may depend on the amount of integration time, the system temperature, and the bandwidth used for that data point. There is no control of R_k in the image formation, so no further mention is made of it here.

The density weight D_k and the taper T_k can be specified in many Fourier transform imaging programs, to 'fine-tune' the beam shape. If S were a smooth, well-behaved function—say, a Gaussian—then B would have no sidelobes, just smooth 'wings'. In practice, S is a linear combination of many δ -functions, often with gaps in the u - v coverage corresponding to missing interferometer spacings. There is always a finite limit to the extent of the u - v coverage, corresponding to the largest (projected) spacing of interferometer elements. In addition, for many arrays more data points fall in the inner region of the u - v plane than fall further out. This tends to give higher weight to the low spatial frequencies. The natural sampling may impair effective deconvolution or mask interesting features of I .

The D_k and the T_k are used to control, to some extent, the beam shape. The T_k are used to weight down the data at the outer edge of the u - v coverage, and thus to suppress small-scale sidelobes and increase the beamwidth. The D_k are used to offset the high concentration of u - v tracks near the center, and to lessen the sidelobes caused by gaps in the coverage; i.e., to simulate more uniform u - v coverage. We shall discuss these forms of weighting separately.

2.2.1. The tapering function. The T_k are specified by a smooth function T : $T_k = T(u_k, v_k)$. T is usually separable, so that $T(u, v) = T_1(u)T_2(v)$; and often it is a radial function (i.e., a function with circular symmetry): $T_k = T(r_k)$ where $r_k \equiv \sqrt{u_k^2 + v_k^2}$. Although functions whose radial profiles follow a power-law or powers of a cosine are occasionally used, the most prevalent form is the Gaussian. The dispersion, or the half-width at half amplitude, or the half-width at 0.30 amplitude are used in different data reduction programs to specify the characteristic width (or widths) of T (see Fig. 6-1).

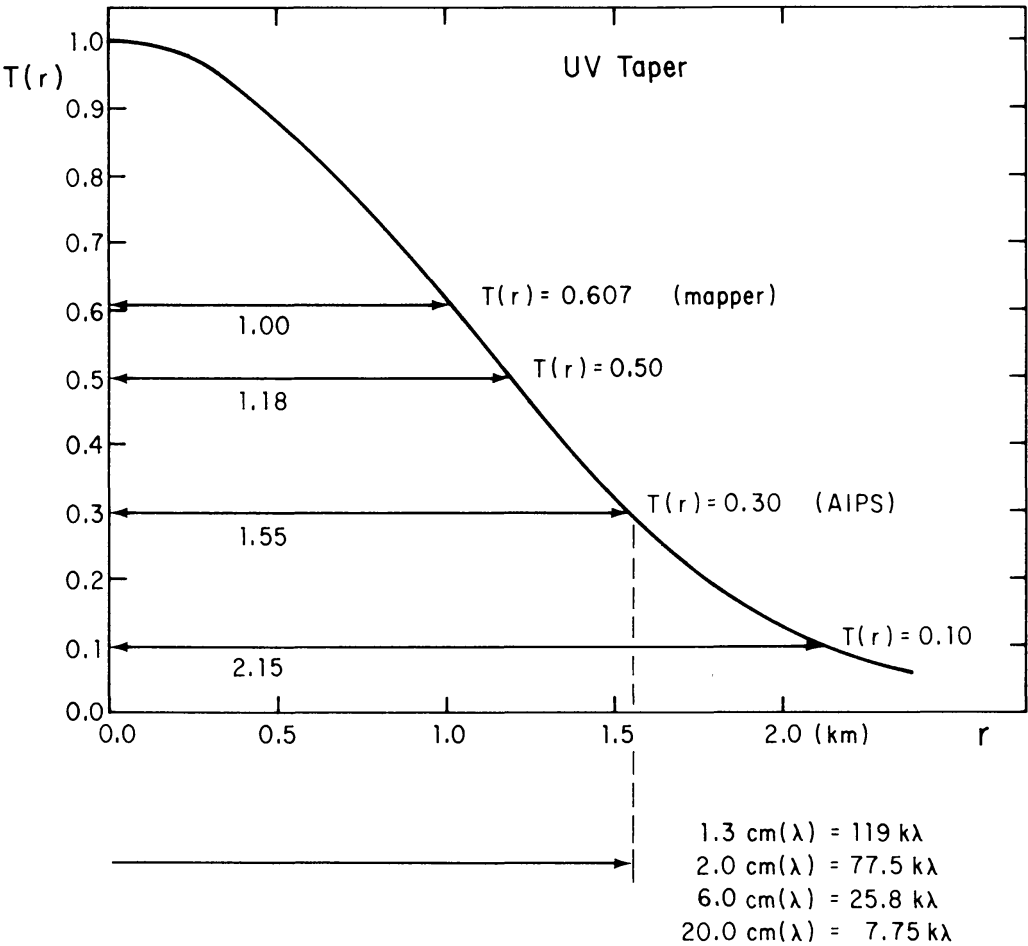


Figure 6-1. A Gaussian u - v taper with dispersion $\sigma = 1$ km.

For a Gaussian taper, $T(r) = \exp(-r^2/2\sigma^2)$, the half-power beamwidth (i.e., the width of the synthesized beam, measured between half-amplitude points) is $\theta_{\text{HPBW}} = 0.37/\sigma$ with θ in radians and σ in wavelengths. Translated into common units, $\theta_{\text{HPBW}} = 0.77\lambda_{(\text{cm})}/\sigma_{(\text{km})}$ arc-seconds. This holds only for a densely sampled Gaussian that is not truncated by the edge of the u - v coverage. When the taper is negligible at the edge of the u - v coverage (assuming dense coverage), one can use a filled circular aperture approximation, for which $\theta_{\text{HPBW}} = 2.0\lambda_{(\text{cm})}/a_{(\text{km})}$ arc-seconds, where a is the radius of the aperture. Real-life observational geometries and u - v coverages often produce larger θ_{HPBW} and, frequently, elongated beams. Examples of the VLA point source response with different u - v tapers are shown in Figure 6-2.

Instead of de-emphasizing data near the outer boundary of the u - v coverage, it is sometimes desirable to downweight the data near $u = v = 0$. An under-sampled large-scale emission region may introduce large undulations in image intensity that are hard to remove. These can present a problem for detecting a weak point source embedded within a region containing extended emission. Minimum u - v limits and other forms of downweighting are often used to diminish the effect of these low spatial frequency data points.

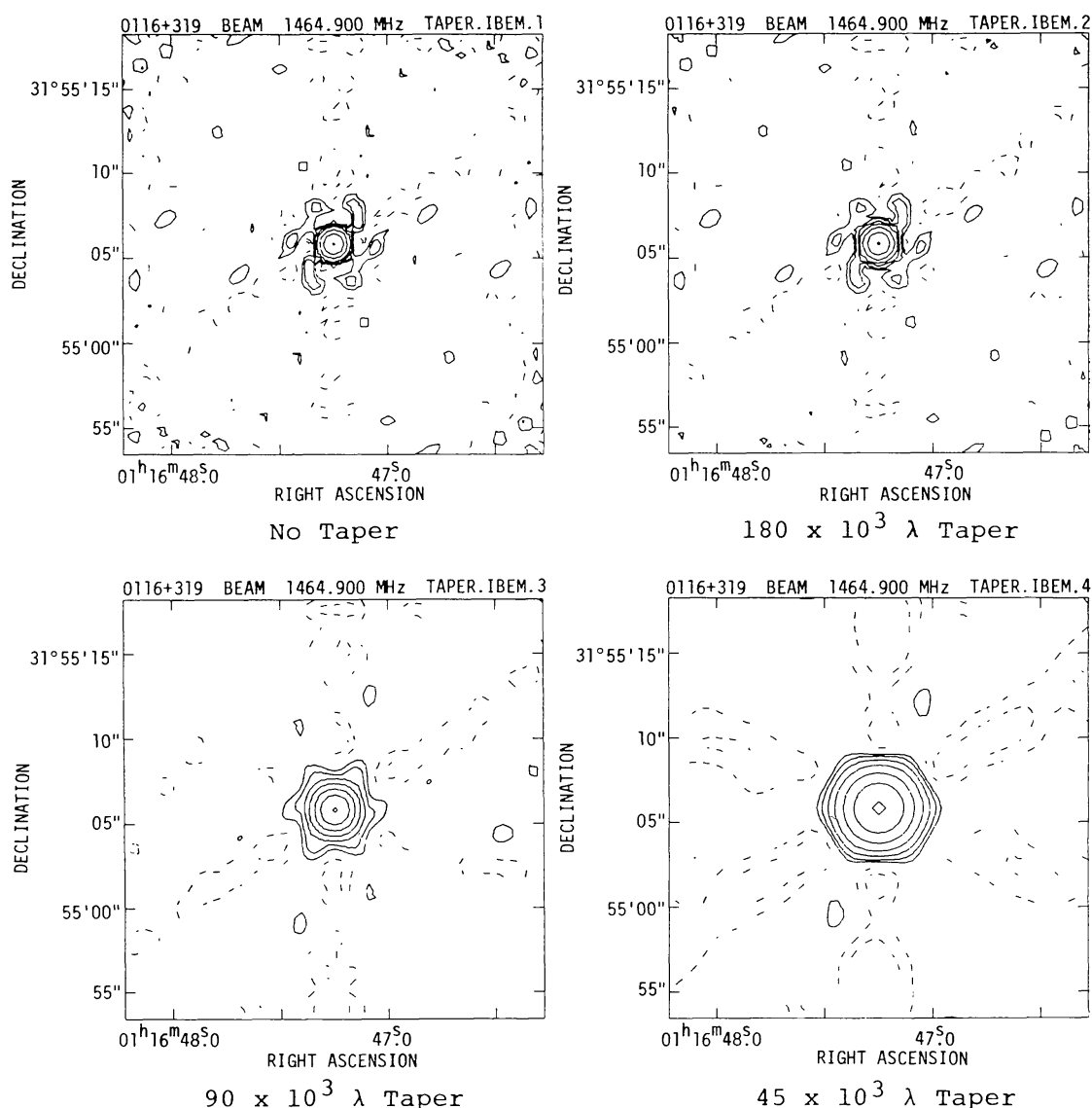


Figure 6-2. The effect of a Gaussian taper on the point source response of a VLA snapshot in the **A** configuration at 20-cm wavelength. As a narrower Gaussian taper (i.e., a heavier tapering) is applied, the half-power width of the point spread function increases and the inner sidelobes are reduced.

2.2.2. The density weighting function. The density weighting function can be used to compensate for the clumping of data in the u - v plane by weighting by the reciprocal of the local data density. Two choices for this weighting are commonly provided:

$$D_k = 1, \quad \text{called } \textit{natural weighting},$$

and

$$D_k = \frac{1}{N_s(k)}, \quad \text{called } \textit{uniform weighting},$$

where $N_s(k)$ is the number of data points within a symmetric region of the u - v plane, of characteristic width s , centered on the k^{th} data point. (s might be the

radius of a circle or the width of a square.) In many Fourier transform imaging programs s is a free parameter selected by the user.

Natural weighting, with all points treated alike, gives the best signal-to-noise ratio for detecting weak sources. However, since the u - v tracks tend to spend more time per unit area near the u - v origin, natural weighting emphasizes the data from the short spacings, and tends to produce a beam with a broad, low-level plateau. This latter feature is especially undesirable when imaging sources with both large-scale and small-scale structure.

With uniform weighting, a common choice for N_s is to count all the points that lie within a rectangular block of grid cells in the neighborhood of the k^{th} datum (gridding is discussed later).⁴ This produces a beam specified largely by the tapering function T .

Sometimes, especially in the VLA ‘snapshot’ mode of observing (see Appendix A7 of Lecture 24), uniform weighting may not be ‘uniform’ enough. Although all cells have equal weight, the filled cells are still concentrated toward the center and along the arms of the VLA ‘Y’. At the further expense of signal-to-noise ratio, the size parameter s can be increased. This ‘super uniform weighting’ gives lightly sampled, isolated cells weights comparable to those given cells in well-sampled parts of the plane. The result is again a beam shape controlled more by the tapering function and less by the arrangement of the sampled visibilities. Examples of the VLA point source response obtained with various weighting functions are shown in Figure 6–3.

3. GRIDDING THE VISIBILITY DATA

To take advantage of the extreme efficiency of the FFT algorithm, visibility values must be assigned to a regular, rectangular matrix or ‘grid’, usually with a power-of-two number of points along each side. Since the observed data seldom lie on such a grid, some procedure (an interpolation procedure comes most readily to mind) must be used to assign visibility values at the grid points, based on the observed values.⁵ There are many ways to achieve this interpolation (see, e.g., Thompson and Bracewell 1974), but with quasi-randomly placed observations a convolutional procedure in the u - v plane leads to an image with predictable distortions and to results that are easy to visualize. Convolution is not, in fact, a pure interpolation procedure, since it combines smoothing, or averaging, with interpolation. This should not be viewed as undesirable—given that there often are many noisy, possibly discrepant, data points in the neighborhood of a given grid point.

3.1. Gridding by convolution.

The idea is to convolve the weighted, sampled measurement distribution V^W with some suitably chosen function C , and to sample this convolution at

⁴ In the AIPS implementation, these blocks are called ‘uniform weight boxes’.

⁵ Some special array geometries (e.g., ‘T’s and Crosses, with elements aligned linearly N–S and E–W) can provide regularly spaced data. See, for example, the description of the Clark Lake array by Erickson *et al.* (1982). The assumption (mentioned below) of a sufficiently large number of data points in the neighborhood of each filled ‘cell’ is not required. However aliasing problems persist, because of the regular sampling.

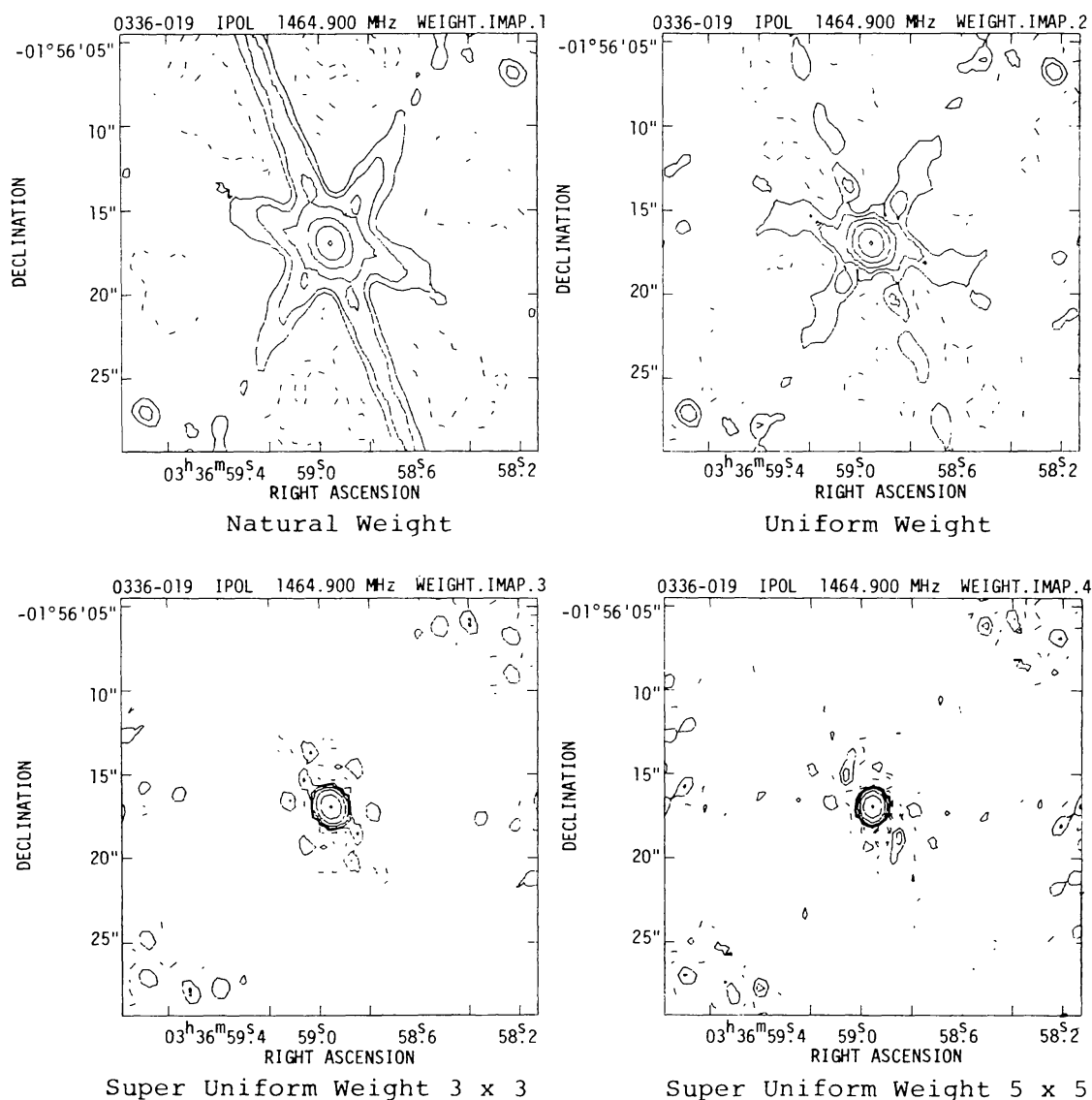


Figure 6-3. The effect of different weighting functions on a VLA 'snapshot' image of a point source.

the center of each 'cell' of the grid. For economy's sake—and because it seems reasonable for the value assigned at a given grid point to equal some local average of the measurements— C , in practice, is always taken to be identically zero outside some small, bounded region A_C . Since V^W is a linear combination of M δ -functions, this convolution $C * V^W$, evaluated at the grid point (u_c, v_c) , is given by

$$\sum_{k=1}^M C(u_c - u_k, v_c - v_k) V^W(u_k, v_k). \quad (6-10)$$

Note that, since the region A_C is quite small in area, there are generally *many* fewer than M nonzero terms in this sum.

Note also that Expression 6-10 does not, in fact, represent a *local average*

of the measurements in the neighborhood of (u_c, v_c) . For that, some sort of normalization would be required—say, multiplication by the area of A_C , followed by division by the number of data points whose shifted coordinates $(u_c - u_k, v_c - v_k)$ lie within the region A_C (and one would want C to integrate to unity). When this particular form of normalization is used, the normalized sum (ignoring weighting) approaches the *non-discrete, integral convolution* $C * V$ evaluated at (u_c, v_c) as the number of measurements increases without bound, provided that the measurements in the neighborhood of (u_c, v_c) are uniformly distributed, and provided that the noise in V' is well-behaved. In practice, this straightforward form of normalization is not always incorporated in imaging—so the matter of normalization becomes intertwined with that of ‘density weighting’, discussed above.

The operation of sampling $C * V^W$ at all points of the grid may be represented by the equation

$$V^R = R(C * V^W) = R(C * (WV')), \quad (6-11)$$

where (as usual) multiplication is indicated by juxtaposition and where R , a ‘bed of nails’ resampling function, is given in terms of Bracewell’s ‘sha’ function (denoted III) by

$$R(u, v) = \text{III}(u/\Delta u, v/\Delta v) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \delta(j - u/\Delta u, k - v/\Delta v). \quad (6-12)$$

Here, Δu and Δv define the cell size—i.e., the separation between grid points. This operation is called *resampling* (hence the R -notation) because, as you recall, the interferometer array earlier provided the samples embodied in V^S and V^W . Now, since V^R is a linear combination of regularly spaced δ -functions, a matrix of samples of its Fourier transform $\mathbf{F}V^R$ can be obtained by a discrete Fourier transform. Thus $\mathbf{F}V^R$ can be calculated by the FFT algorithm.

$\mathbf{F}V^R$ —after normalization, and after one simple correction—is what you have been seeking: a ‘dirty’ image—a cheap approximation to I^D . Denote $\mathbf{F}V^R$ by \tilde{I}^D .

Applying the convolution theorem to Equation 6-11, \tilde{I}^D is given by

$$\tilde{I}^D = \mathbf{F}R * [(\mathbf{F}C)(\mathbf{F}V^W)] = \mathbf{F}R * [(\mathbf{F}C)(\mathbf{F}W * \mathbf{F}V')]. \quad (6-13)$$

(Please refer now to Fig. 6-4 for a graphical interpretation of Eq. 6-13 and for an illustration of the operations that are described in the remainder of this section.) III is its own Fourier transform; R behaves similarly—by the dilation property of the FT (see Sec. 4.1),

$$(\mathbf{F}R)(l, m) = \Delta u \Delta v \text{III}(l\Delta u, m\Delta v) = \Delta u \Delta v \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \delta(j - l\Delta u, k - m\Delta v). \quad (6-14)$$

One effect of the resampling is to make \tilde{I}^D a periodic function of l and m , of period $1/\Delta u$ in l and period $1/\Delta v$ in m . Another effect, called *aliasing*, is also introduced. It, too, arises because of the convolution with the scaled sha function \mathbf{FR} (more on this later, in Sec. 3.2).

The FFT algorithm generates one period of (a discrete version of) \tilde{I}^D . To image a rectangular region of width $N_l\Delta\theta_l$ radians in l and $N_m\Delta\theta_m$ in m , one chooses grid spacings satisfying $N_l\Delta u = 1/\Delta\theta_l$ and $N_m\Delta v = 1/\Delta\theta_m$ wavelengths. An $N_m \times N_l$ FFT yields the discretely sampled version of \tilde{I}^D . Let P denote the region over which \tilde{I}^D is computed—i.e., P , which is called the *primary field of view*, is given by $|l| < N_l\Delta\theta_l/2$, $|m| < N_m\Delta\theta_m/2$.

The net effect of the gridding convolution is to multiply the sky brightness by a function $c(l, m)$, the FT of the convolving function C (i.e., $c \equiv \mathbf{FC}$). The tapering function T , introduced earlier for control of the beam shape, has the effect of a convolution in the image domain.

An image representing the point source response of the array, or the ‘dirty beam’ B^D , can be obtained by setting all the measurements $V'(u_k, v_k)$ to unity and following the steps outlined above. Denote the image so obtained by \tilde{B}^D .

Normally, \tilde{I}^D and \tilde{B}^D are corrected for the effect of the gridding convolution by pointwise division by c : The so-called ‘grid-corrected’ image is given by

$$\tilde{I}_c^D(l, m) = \frac{\tilde{I}^D(l, m)}{c(l, m)}, \quad (6-15)$$

and the ‘grid-corrected’ beam by

$$\tilde{B}_c^D(l, m) = \frac{\tilde{B}^D(l, m)}{c(l, m)}. \quad (6-16)$$

The commonly used term ‘grid corrected’ is, in a way, a misnomer, since one is actually correcting for the effect of the convolution function C . The grid correction is not an exact correction, except in the limit of a large number of well-distributed visibility measurements. It also is not exact due to the presence of R in Equation 6–11 and \mathbf{FR} in Equation 6–13. It could be so only if $c(l, m)$ were identically zero outside of the region being imaged; this is impossible because C is confined to a bounded region A_C .⁶

⁶ The FT of any nontrivial (i.e., nonzero) function which is confined to a bounded region has features extending to infinity. By a theorem of Paley and Wiener (see, e.g., Dym and McKean 1972) the FT of such a function is extremely well-behaved, in the sense that it can be analytically extended to an entire function in the complex domain (i.e., in the case of 2 dimensions, from \mathbf{R}^2 to \mathbf{C}^2). In particular, the FT cannot vanish over any open set (this is why the synthesized beam has sidelobes that ‘never go away’).

Finally, \tilde{I}_c^D and \tilde{B}_c^D both are normalized by a scaling factor selected so that the peak of \tilde{B}_c^D is of unit flux density. One may as well not alter the notation to reflect this, since it is a trivial operation.

Figure 6–4 (*pp. 128–129*). A graphical illustration of the steps in the imaging process is shown in this one-dimensional example. At the top, in panels (a) and (b), a model source and its visibility are displayed side-by-side; the results of successive imaging operations are displayed vertically. The image domain is shown on the left, and the visibility domain on the right. Horizontally opposed panels represent Fourier transform pairs. The units on the vertical axes were chosen arbitrarily—i.e., we have not bothered with normalization. The horizontal axes are in radians for the image domain plots, at left; the baselines are expressed in wavelengths for the visibility domain plots, at right.

The model source, shown in panel (a), is the sum of a Gaussian-shaped extended source and four symmetrically placed point sources. The total flux density of the Gaussian is 1.5 times the sum of the fluxes in the point sources. This symmetry was chosen to ensure that the visibility function, shown in panel (b), is real-valued and even, allowing a simpler display. Panel (d) shows the telescope transfer function, or sampling function S , which includes a central ‘hole’. We have chosen a smooth function for simplicity, but one should note that no array would in fact produce a smooth sampling function. In reality, S is a sea of closely- and irregularly-spaced δ -functions, as in Equation 6–4. The triangular sampling density was chosen to mimic the fall-off in the density of samples with increasing spacing. The telescope beam B corresponding to (d) is shown in panel (c). The data available for imaging are shown in panel (f); this product of the true visibility function and the sampling function corresponds to V^S , as defined by Equation 6–5. The image which a direct transformation of (f) would yield is shown in panel (e). This image is equal to the convolution of the beam (c) with the true sky brightness (a). This image shows a large amplitude oscillation, reaching a negative peak centered on the position of the extended source. This effect, which is of much larger amplitude than the oscillation seen in (c), is due to the missing central spacings in the u - v sampling and to the fact that the visibility of an extended source is relatively highly concentrated near $u = v = 0$. With sufficient computing resources (mammoth resources would often be required), one might use the ‘direct Fourier transform’ method of Section 1.1; (e) is the image that would result.

Extra steps are required to make use of the FFT: First, the data are convolved with some suitably chosen function, and then they are resampled over a regularly-spaced grid (in practice the convolution is evaluated only at the grid points). For illustration, a simple, and crude, convolution function C was employed, as shown in (h). The sharp drop-off in C creates large, oscillating wings in its Fourier transform, shown in (g) (the reciprocal of the ‘grid-correction function’). The data, after convolution, are shown in panel (j). If a (continuous) Fourier transform were applied at this stage, the result would appear as in panel (i). The important effect to note is that the outermost point sources have been inverted in amplitude. This occurs because the convolution function that we have chosen is too wide. The inner point sources have been slightly reduced in amplitude, though not inverted in sign. As the FFT requires regularly spaced data, the data in (j) must be sampled. The (re-)sampling function R is shown in panel (l), and its transform, the replication function, in panel (k). The resampled, convolved visibility is shown in panel (n). These are the data that the FFT actually sees. The FT of this is the image shown in panel (m); it has been replicated at the various points shown in panel (k). Notice that aliases of the outermost point sources appear just outside the positions of the innermost point sources. This aliasing occurs because the resampling function, shown in panel (l), undersamples (i.e., takes fewer than 2 samples per cycle) of the transform of the outermost point sources. The final operation is correcting for the effect of the convolution. This is done by dividing the image by the Fourier transform of the convolution function. The result is shown in panel (o). This is the end product, the ‘dirty image’ that is supplied to the deconvolution programs.

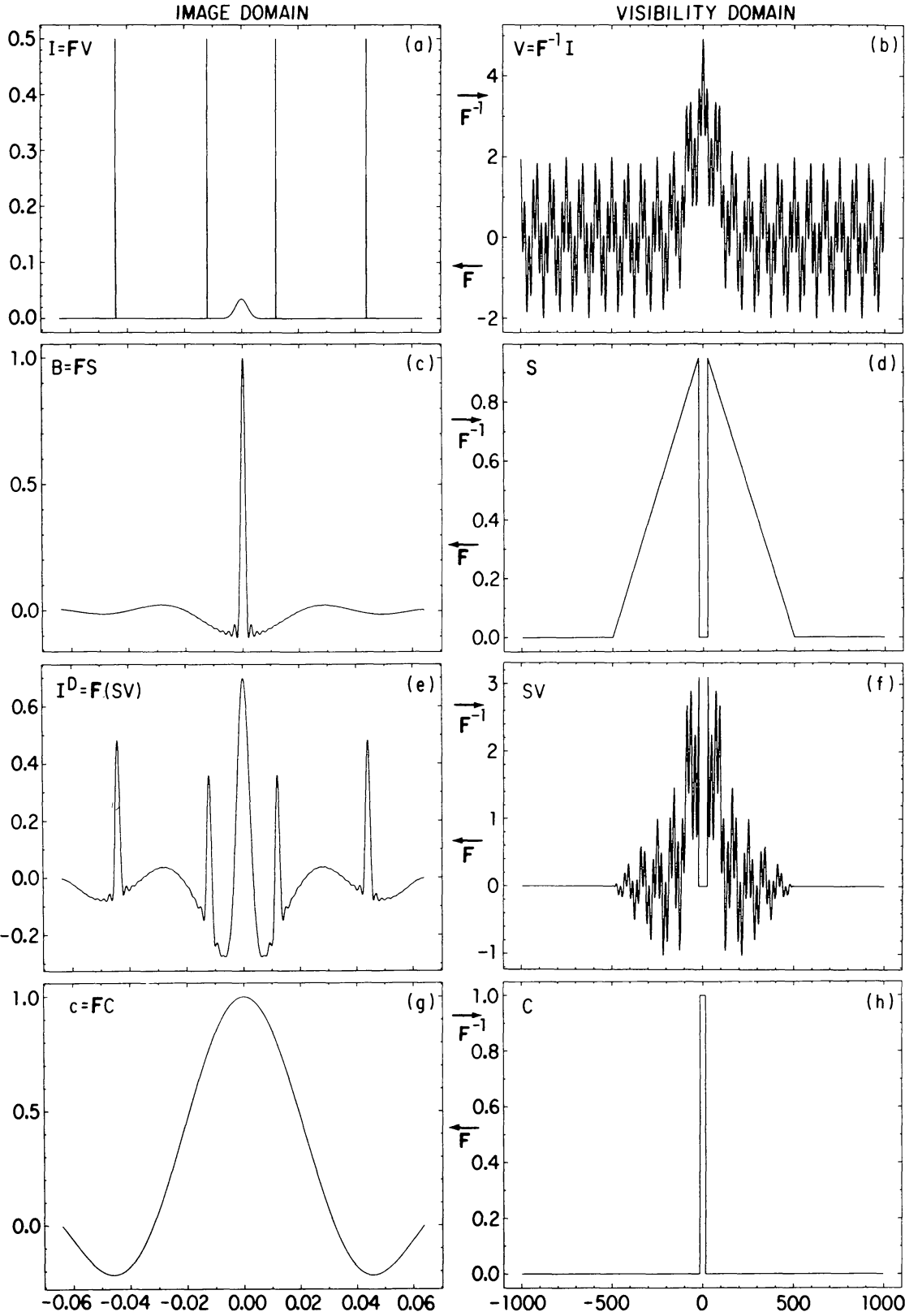


Figure 6-4 (Caption is on p. 127). (Continued on next page.)

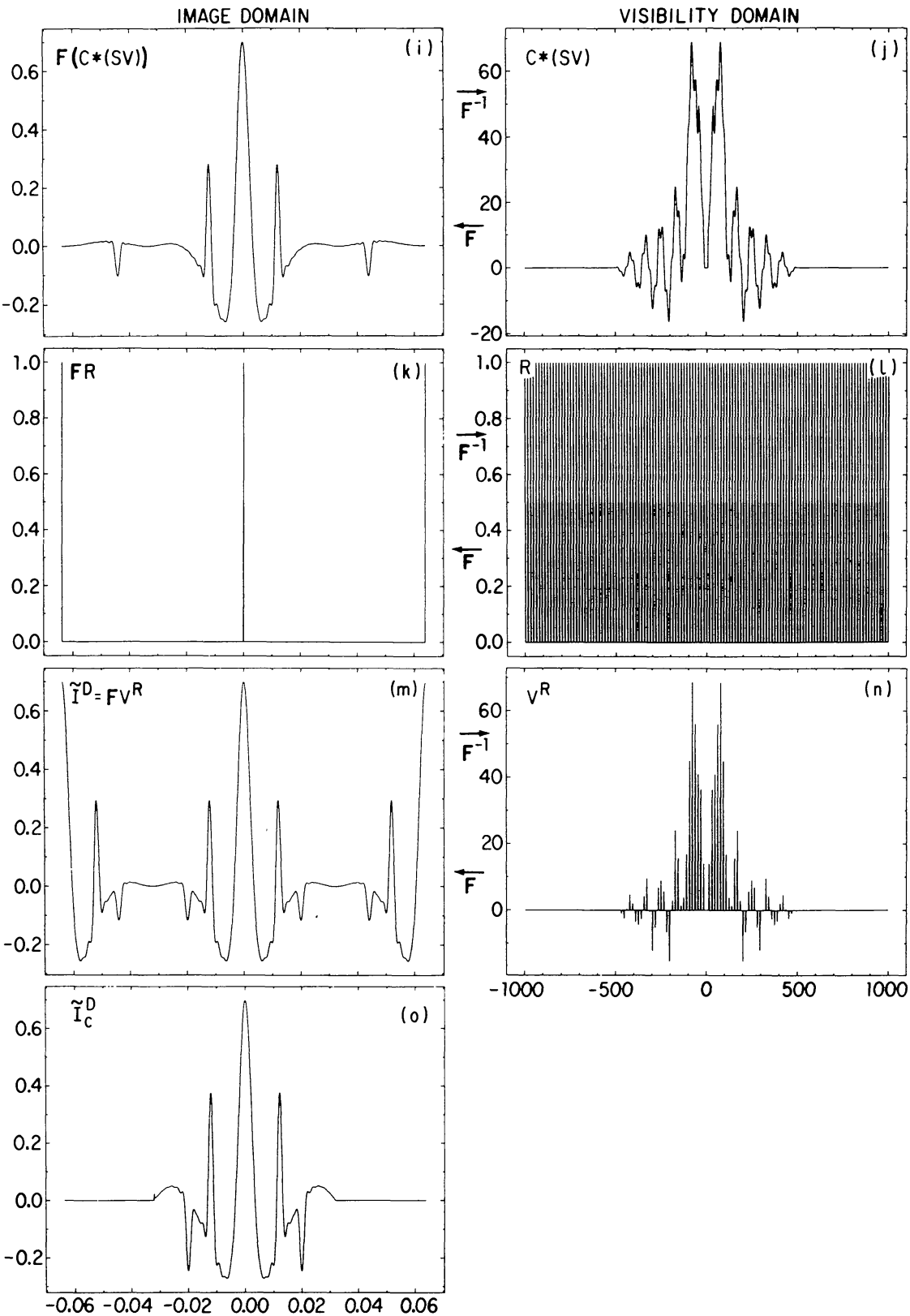


Figure 6-4 (Continued).

If $c(l, m)$ tends sufficiently rapidly to zero outside P , so that the resampling can be ignored, and if the u - v samples are well enough distributed for the gridding correction to be approximately valid, then \tilde{I}_c^D is a good approximation to I^D —that is, Equation 6–13 becomes

$$\tilde{I}_c^D = \mathbf{F}W * \mathbf{F}V', \tag{6-17}$$

—and then the usual convolution relation between I^D , B , and I is approximately valid with \tilde{I}_c^D and \tilde{B}_c^D substituted for I^D and B , respectively. Note, however, that \tilde{B}_c^D is usually computed only over a region of the same dimensions as the image \tilde{I}_c^D . For this reason, the deconvolution algorithms (described in Lecture 8) usually operate just on a region with one-quarter the area of the input image.

3.2. Aliasing.

Due to the presence of $\mathbf{F}R$ in Equation 6–13 and to the fact that c is not identically zero outside the primary field of view, parts of the sky brightness that lie outside P are aliased, or ‘folded back’, into P . Undersampling, and the truncation of the sampling at the boundaries of the u - v coverage, are the root causes of aliasing. (If the sky brightness I has features extending over a region of width Ω_l in l and width Ω_m in m , then its visibility function has been undersampled if the visibility samples are separated by more than $1/\Omega_l$ in u and $1/\Omega_m$ in v .) The amplitude of an aliased response from position (l, m) is determined by $|c(l, m)|$. The simplest way to tell whether a feature is aliased or authentic is to calculate images with different cell sizes $\Delta\theta$; an aliased feature then appears to move, while a real one stays the same angular distance from the image center. Additionally, an image covering the full main lobe of the primary beam may quickly reveal whether there is an aliasing problem in an image of a smaller region.

Aliasing of sources that lie outside the primary field of view is only part of the problem. Although it may be possible to obtain visibility samples that are closely enough spaced to avoid undersampling over the sampled region of the u - v plane, the finite physical size of the array sets a limit on how far the sampling can extend. For this reason, any authentic feature within P has sidelobes extending outside the image. These sidelobes are also aliased into P , effectively raising the background variance and resulting in a beam shape that depends on position. If, for example, the visibility function is well sampled over a square region of the u - v plane but no samples are obtained outside that region, then (assuming uniform weighting) the sidelobes in I^D are precisely those of Gibbs’ phenomenon, discussed in Lecture 4 (Sec. 4.6).

3.3. Choice of a gridding convolution function.

The best ways to avoid aliasing problems are (a) to make the image large enough that there are no sources of interest near the edges of the image, (b) to avoid undersampling, and (c) to use a gridding convolution function C whose Fourier transform c drops off very rapidly beyond the edge of the image. Desideratum (c) favors gridding convolution functions that are not highly confined in the u - v plane. But, in practice, computing time restricts one’s choice of C to

functions that vanish outside a small region, typically six or eight u - v grid cells across. A compromise must be struck between alias rejection and computing time.

C is always taken to be real and even. And, since C is usually separable—i.e., $C(u, v) = C_1(u)C_2(v)$ —we shall continue the discussion in just one dimension. Typical choices for C are:

- (1) a ‘pillbox’ function,
- (2) a truncated exponential,
- (3) a truncated sinc function ($\text{sinc } x \equiv \frac{\sin \pi x}{\pi x}$),
- (4) an exponential multiplied by a truncated sinc function, and
- (5) a truncated spheroidal function.

Each is truncated to an interval of width m grid cells, so that $C(u) \equiv 0$ for $|u| > m\Delta u/2$; thus $O(Mm^2)$ arithmetic operations are required for gridding. These functions are described below; for more discussion see Schwab (1978, 1980a):

- (1) *Pillbox*. $C(u) = \begin{cases} 1, & |u| < m\Delta u/2, \\ 0, & \text{otherwise.} \end{cases}$ For $m = 1$, convolution with this C is equivalent to simply summing the data in each cell. Calculation of these sums is fast, but the alias rejection is the worst of the five functions considered here. c is a scaled sinc function.
- (2) *Exponential*. $C(u) = \exp \left[- \left(\frac{|u|}{w\Delta u} \right)^\alpha \right]$. Typically $m = 6$, $w = 1$, and $\alpha = 2$. That is, a truncated Gaussian is often used, in which case c can be expressed in terms of the error function.
- (3) *Sinc*. $C(u) = \text{sinc } \frac{u}{w\Delta u}$. Typically $m = 6$, $w = 1$. c can be expressed in terms of the sine integral. If m is allowed to increase, c approaches a step function that is constant over P and zero outside. This is the intuitive justification for considering the use of this function, that the FT of a unit step function truncated at $\pm \frac{1}{2}$ is the sinc function.
- (4) *Exponential times sinc*. $C(u) = \exp \left[- \left(\frac{|u|}{w_1\Delta u} \right)^\alpha \right] \text{sinc } \frac{u}{w_2\Delta u}$. Typically⁷ $m = 6$, $w_1 = 2.52$, $w_2 = 1.55$, $\alpha = 2$; i.e., a truncated, Gaussian-tapered sinc function is often used. c can easily be computed by numerical quadrature, but it lacks a closed-form expression.
- (5) *Spheroidal functions*. $C(u) = |1 - \eta^2(u)|^\alpha \psi_{\alpha 0}(\pi m/2, \eta(u))$, with $\psi_{\alpha 0}$ a 0-order spheroidal function (Stratton 1935), $\eta(u) = 2u/m\Delta u$, and $\alpha > -1$. For $\alpha = 0$ this is the 0-order ‘prolate spheroidal wave function’, which is the optimal C (among all square-integrable functions of width m grid cells) in that the energy concentration ratio $\int_P |c(l)|^2 dl / \int_{-\infty}^{\infty} |c(l)|^2 dl$ is maximized. The other $\psi_{\alpha 0}$ are optimal in the sense of maximizing a *weighted* con-

⁷ For a gridding convolution function of this particular parametric form, these values of the characteristic widths w_1 and w_2 are an optimal choice, in the sense described below in the discussion of ψ_{00} .

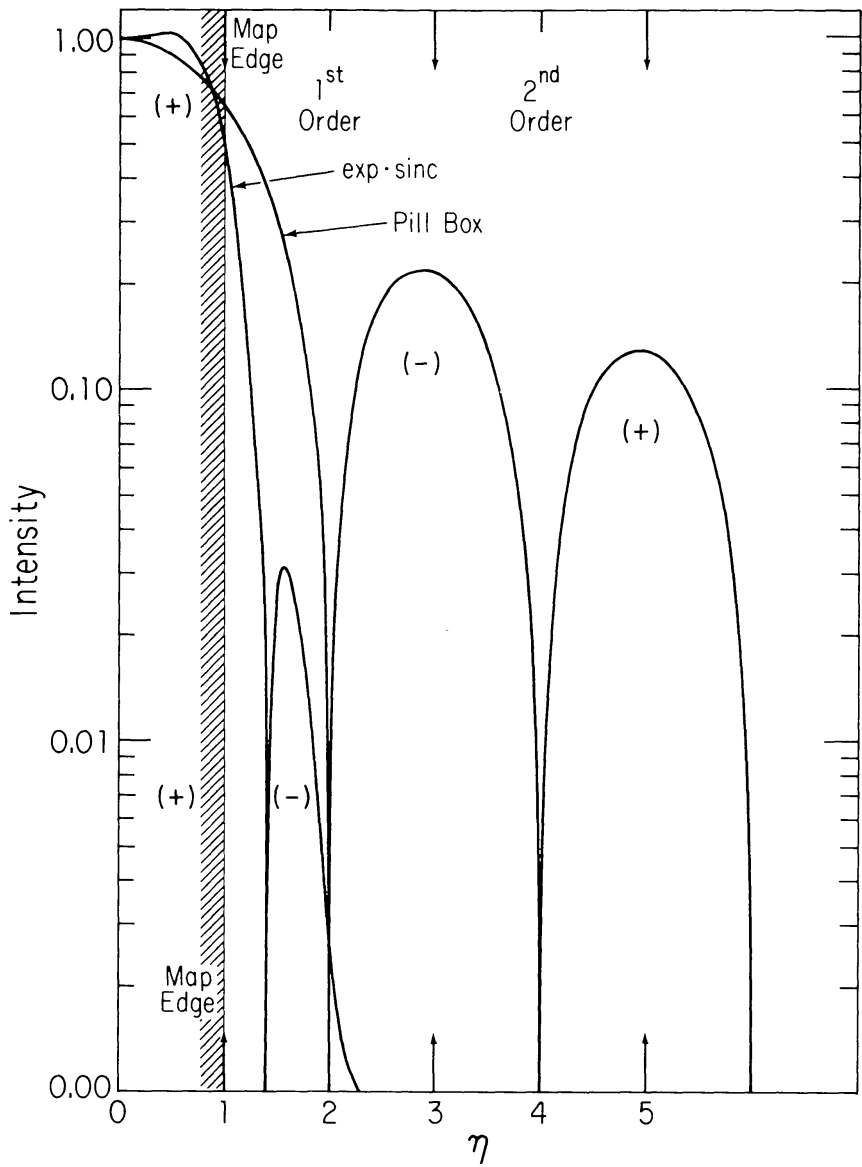


Figure 6-5. The response to a source, as a function of distance from the image center, for two typical u - v convolving functions.

centration ratio: for given α , $\int_P w(l)|c(l)|^2 dl / \int_{-\infty}^{\infty} w(l)|c(l)|^2 dl$ is maximized, where $w(l) = |1 - 2l\Delta u|^\alpha$. Choosing $\alpha > 0$ gives higher alias rejection near the center of the image, at the expense of alias rejection near the edges. ψ_{00} is its own FT, in the sense that if you truncate it as done here, and then take the FT, what you get back is ψ_{00} . Similarly, the other $\psi_{\alpha 0}$ are finite Fourier self-transforms, in the sense that if you so truncate one, weight it, and transform it, what you get back is $\psi_{\alpha 0}$. $\psi_{\alpha 0}$ is used at the VLA, with $m = 6$ and $\alpha = 1$ being typical. See Schwab (1984a) for further discussion and additional references.

Figure 6-5 shows the Fourier transforms of two typical gridding convolution

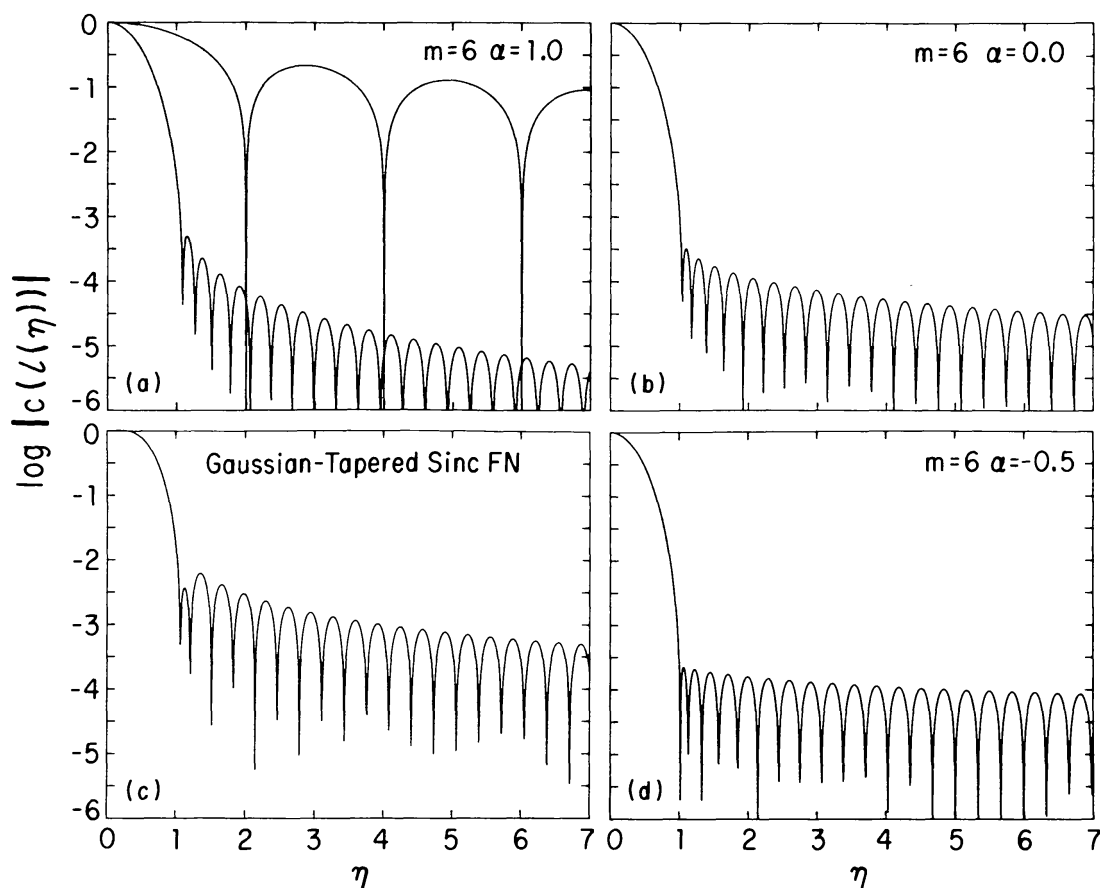


Figure 6-6. For some typical gridding convolution functions C , plots of the absolute value of the Fourier transform of C . (a) The spheroidal function ψ_{10} , for $m = 6$, compared with the pillbox function ($m = 1$); (b) the ‘prolate spheroidal wave function’ ψ_{00} , $m = 6$; (c) an optimized Gaussian-tapered sinc function, $m = 6$; (d) the spheroidal function $\psi_{-\frac{1}{2},0}$, $m = 6$. Adapted from Schwab (1984a).

functions, normalized to unity at $l = 0$. The abscissa on this plot is in units of image half-widths, $\eta = 2l\Delta u$, so that $\eta = \pm 1$ at the image edges. The image response is suppressed at the edge for both functions, however the $\exp \times \text{sinc}$ function is flatter inside P , and drops much faster past the image edge. The aliased response can, of course, be negative, producing an apparent ‘hole’ in the image. The plots in Figure 6-6 compare the pillbox function and the Gaussian-tapered sinc function with several spheroidal functions. The quantity of most direct importance is the ratio of the intensity of an aliased response to the intensity the feature would have if it actually lay within the primary field of view P , at the position of its alias: if η' denotes the position within P at which the aliased response of a source at position η appears, then this ratio is given by $q(\eta) = |c(l(\eta))/c(l(\eta'))|$. (And η' is given by $\eta' = ((\eta + 1) \bmod 2) - 1$; it is useful to sketch a plot to convince oneself of this.) Schwab (1978) and Greisen (1979) show plots of q for these convolving functions and for many others.

The pillbox, exponential, and sinc functions do not give as effective alias rejection as the $\exp \times \text{sinc}$ or the spheroidal. The $\exp \times \text{sinc}$ has somewhat smaller

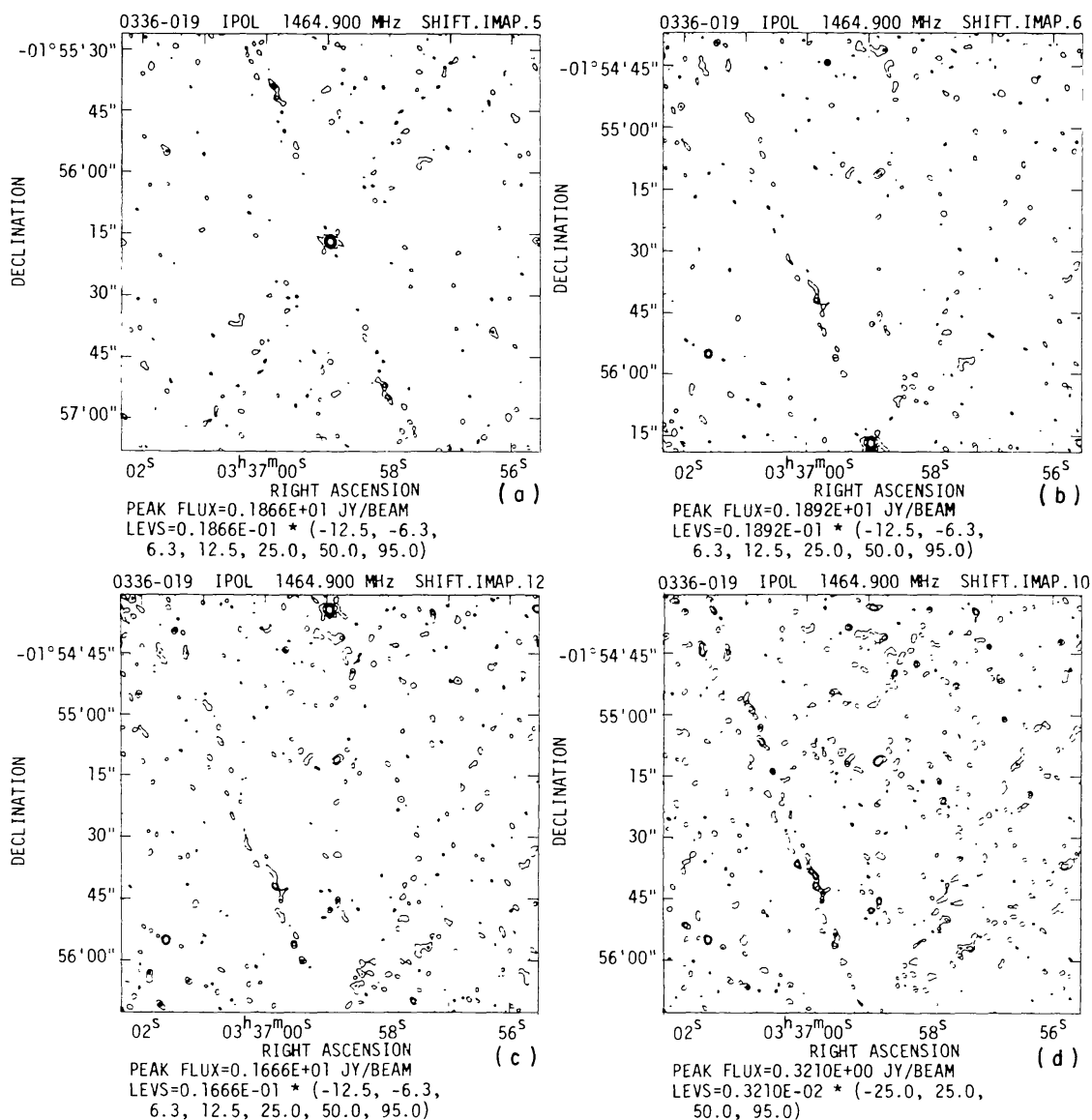


Figure 6-7. The effects of aliasing: (a) a point source at the field center; (b) the same source near the image edge; (c) the source below the lower image edge appears as an aliased image at the upper image edge, with pillbox convolution; (d) with $\exp \times \text{sinc}$ convolution, the aliased source is greatly attenuated, but the sidelobe response remains the same.

corrections and, thus smaller errors (due to round-off noise and to violation of the assumptions that make the grid correction valid), near the image edges, while the spheroidal has better rejection beyond the image edge (Schwab 1984a).

Remember that the convolution functions suppress only aliased responses. Sidelobes which legitimately fall within the primary field of view, whether from sources inside or outside P , are not suppressed (see Fig. 6-7). With alias suppression of 10^2 to 10^3 , at two or three image half-widths, it is these sidelobes which may cause the dominant spurious image features and impair effective deconvolution.

4. ADDITIONAL TOPICS

4.1. Translating, rotating, and stretching images.

The Fourier transform possesses three basic symmetry properties that are useful in radio interferometric imaging. The first important property is the behavior of the Fourier transform with respect to translation—that is, with respect to a shift of origin: namely, if you shift a function, i.e., replace $f(\mathbf{u})$ by $f(\mathbf{u} - \Delta\mathbf{u})$, and take the FT you get the same result as if you had first taken the FT and then multiplied by $e^{2\pi i \mathbf{x} \cdot \Delta\mathbf{u}}$ (here \mathbf{x} denotes the variable in the transform domain). Similarly, if you want a shift of origin $\Delta\mathbf{x}$ in the transform domain, all you need do is multiply, before transforming, by a factor $e^{-2\pi i \mathbf{u} \cdot \Delta\mathbf{x}}$. Thus, in imaging, all that is required to achieve a shift of origin in the image is to multiply the visibilities by the appropriate complex exponentials before transforming.

The second important property is that the Fourier transform commutes with rotations; that is, if you take the FT and then rotate the coordinate system in the transform domain, you get the same result as if you had first rotated the coordinate system and then taken the FT. Thus, to ‘turn an image around’, all that you need do is rotate the u - v coordinates of the visibility data. (It is easy to see why the FT has this property: the inner product $\mathbf{u} \cdot \mathbf{x}$ in the exponential kernel of the FT is invariant under rotation.) At the VLA, the visibility u - v coordinates are routinely rotated to correct the data for differential precession—i.e., to put the data into the coordinate reference frame of a standard epoch, say, J1950 or J2000. Data taken at two different epochs, say a year apart, need this correction for differential precession before they can be sensibly combined or compared; routine correction to a standard epoch automatically rectifies this problem. Additionally, it is sometimes convenient to rotate the coordinate system so that features in a source have a particular alignment in an image. For an elongated source, this can reduce the data storage requirements (by reducing the number of pixels needed to represent the source by a computed, discrete image) and therefore aid during deconvolution (see Lecture 8) by reducing the required number of arithmetic operations.

The third basic symmetry property of the FT is that it *anti*-commutes with dilations. That is, if you ‘stretch’ a function linearly and isotropically, then its FT ‘shrinks’ proportionately. (That is, the FT of $g(\mathbf{u}) = f(\alpha\mathbf{u})$ is given by $(\mathbf{F}g)(\mathbf{x}) = \alpha^{-n}(\mathbf{F}f)(\mathbf{x}/\alpha)$. The multiplicative constant α^{-n} depends on the dimensionality n .) Or, if you linearly stretch a function in just one coordinate, then its FT ‘shrinks’ proportionately, but in only one of the coordinate directions. This property is the reason that, for a fixed array geometry, the spatial resolution increases (i.e., the characteristic width of the synthesized beam decreases) with observing frequency—the reason that as the u - v coverage expands, the beam shrinks proportionately.

Following Bracewell (1978), the shift property is sometimes called the *shift theorem*, and the dilation property the *similarity theorem*.

4.2. Practical details of implementation.

Most Fourier transform imaging programs do not work quite as described above. Often the tapering, introduced in Equation 6–8, and specified by $T(u, v)$,

is applied after gridding. This would appear to make only a minute difference. But, in the same sense in which it is incorrect to ignore resampling to justify the grid correction, it is also incorrect to ignore the convolution with \mathbf{FT} , which, if inserted into Equation 6-13, would now appear outside the square brackets.

For economy, Fourier transform imaging programs often do not attempt to evaluate the gridding convolution function very accurately, but instead use a step function (tabular) approximation, with steps spaced at increments of, typically, $\Delta u/100$. This introduces another (not very serious) ‘replication’ effect like that due to \mathbf{FR} , but one with a very long period, $100/\Delta u$. The grid correction given by Equation 6-15 should be based now on the FT of the step function approximation to C rather than on the FT of C itself. For analysis, see Greisen (1979). (Schwab (1984a) gives cheap and accurate rational approximations to the spheroidal functions; the step function approximation is unnecessary.)

4.3. Non-coplanar baselines.

In Equation 6-1 the visibility samples are expressed as a function of two variables, u and v , rather than as a function of (u, v, w) . As shown in Section 6 of Lecture 2, Equation 6-1 is strictly valid whenever the visibility measurements are confined to a plane, as they would be if obtained with an interferometer array whose elements are aligned along an East–West line; and, again as shown in Lecture 2, this relation is approximately valid when $I(l, m)$ is confined to a small region of sky—that is, when our condition (b) holds, $|w(l^2 + m^2)| \ll 1$. In wide-field imaging with non-coplanar baselines, condition (b) is often violated.

Recall from Lecture 2 (Eq. 2-21) the relation

$$V(u, v, w) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\mathcal{A}(l, m) I(l, m)}{\sqrt{1 - l^2 - m^2}} e^{-2\pi i (ul + vm + w(\sqrt{1 - l^2 - m^2} - 1))} dl dm. \quad (6-18)$$

This can be rewritten as

$$V(u, v, w) e^{-2\pi i w} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\mathcal{A}(l, m) I(l, m)}{\sqrt{1 - l^2 - m^2}} \delta(n - \sqrt{1 - l^2 - m^2}) e^{-2\pi i (ul + vm + wn)} dl dm dn. \quad (6-19)$$

Now, by sampling V , weighting by $e^{-2\pi i w}$ and by the Fourier kernel, and integrating over (u, v, w) , one obtains an analog of Equation 6-2,

$$I^{3D}(l, m, n) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} S(u, v, w) V(u, v, w) e^{-2\pi i w} e^{2\pi i (ul + vm + wn)} du dv dw, \quad (6-20)$$

which (cf. Eq. 6-19) is equal to a three-dimensional convolution—the convolution of

$$I^3(l, m, n) \equiv \frac{\mathcal{A}(l, m) I(l, m)}{\sqrt{1 - l^2 - m^2}} \delta(n - \sqrt{1 - l^2 - m^2}), \quad (6-21)$$

with

$$B^{3D}(l, m, n) \equiv \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} S(u, v, w) e^{2\pi i(ul+vm+wn)} du dv dw. \quad (6-22)$$

Note that I^3 is a distribution confined to the celestial sphere and that B^{3D} is mostly concentrated near the origin, i.e., near $l = m = n = 0$.

Either of the methods described earlier for approximating I^D can be extended straightforwardly to Equation 6-20. In applying the ‘direct Fourier transform’ method, one simply uses a discrete summation, in analog to Equation 6-3. In the FFT method, w -terms need to be inserted into Expression 6-10, defining the gridding operation; a 3-D FFT yields a three-dimensional discretely sampled image⁸; and one interpolates this result to obtain data over a spherical cap, a portion of the surface $(l, m, \sqrt{1-l^2-m^2})$. Because usually the importance of the curvature effect is minor and the data cover a small range of w , N_n , the number of slices required in the w - and n -dimensions, is small—typically eight to sixteen. At the VLA, such a 3-D imaging capability was designed into the ‘pipeline’ imaging system, but it has seldom been used. Rick Perley, in Lecture 14, will report on more recent efforts to deal in software with the non-coplanar baseline problem.

5. THE PROBLEM WITH I^D —SIDELOBES

An astronomer is seldom satisfied with the approximation to I defined by I^D , or with the computed version thereof, \tilde{I}_c^D . This is because of the sidelobes which contaminate I^D . As you have seen, these are due to the finite extent of the u - v coverage and to gaps in the coverage. Sidelobes from bright features within an image are likely to obscure any fainter features. The process described here is usually just the first step in obtaining a better approximation to I . Because the convolution relation $\tilde{I}_c^D = \tilde{B}_c^D * I$, is approximately valid, this first step provides a starting point for the deconvolution (i.e., sidelobe removal) process described in Lecture 8. However, in cases of very low signal-to-noise ratio (as might occur in an observation to determine the detectability of a putative source) one would often choose not to proceed any further. This is often the case, too, in spectral line observing, primarily because spectral line data reduction is computationally very expensive, and because narrow bandwidths lead to low signal-to-noise ratios.

In wide-field imaging, deconvolution is the real problem in trying to cope with non-coplanar baselines. Because simple 2-D deconvolution itself is an extremely expensive operation, there has been little progress to date in obtaining high quality (deconvolved) images taking proper account of sky curvature and non-coplanar baselines, via any sort of three-dimensional deconvolution technique (to complement the 3-D imaging techniques described in Sec. 4.3). Data storage is another problem. Typically, non-coplanar baseline

⁸ In the FFT method, one normally would want a shift of origin, in order to get the plane tangent to the celestial sphere at $(0, 0, 1)$ shifted to the origin of the third coordinate axis of the grid. This involves multiplying the data by $e^{2\pi i w}$, which cancels the multiplication by $e^{-2\pi i w}$ in Equation 6-20.

effects are an important concern in the largest images; but computer storage is often barely adequate for the number of points, or ‘pixels’, required in just the l - and m -coordinates. A crude approach which has yielded some useful results involves *mosaicing*—constructing ‘patchwork’ images, each piece computed with the ‘ $w(n-1)$ ’-correction appropriate to the center of the patch. This approach, which is used in the AIPS program ‘MX’, combining linear imaging with deconvolution, is described in Lecture 8. Because sidelobes from a source in any one patch fall into each of the other patches of the mosaic, the deconvolution operation must work in parallel on the patches. This necessitates repeated re-gridding of data.

More recent developments in 3-D imaging and 3-D deconvolution will be described in Lecture 14 by Rick Perley.

ACKNOWLEDGMENTS

We would like to thank Alan Bridle and Rick Perley for numerous helpful discussions during the preparation of this lecture. Rick Perley kindly provided Figure 6-4.

REFERENCES

- Bracewell, R. N. (1978), *The Fourier Transform and Its Applications*, Second Edition, McGraw-Hill, New York.
- Dym, H. and McKean, H. P. (1972), *Fourier Series and Integrals*, Academic Press, New York.
- Erickson, W. C., Mahoney, M. J., and Erb, K. (1982), “The Clark Lake Teepee-Tee telescope”, *Astrophys. J. Suppl. Ser.*, **50**, 403–420.
- Greisen, E. W. (1979), “The effects of various convolving functions on aliasing and relative signal-to-noise ratios”, VLA Scientific Memorandum No. 131, NRAO.
- Schwab, F. R. (1978), “Suppression of aliasing by convolutional gridding schemes”, VLA Scientific Memorandum No. 129, NRAO.
- Schwab, F. R. (1980a), “Optimal gridding”, VLA Scientific Memorandum No. 132, NRAO.
- Schwab, F. R. (1984a), “Optimal gridding of visibility data in radio interferometry”, in *Indirect Imaging*, J. A. Roberts, Ed., Cambridge University Press (Cambridge, England), pp. 333–346.
- Stratton, J. A. (1935), “Spheroidal functions”, *Proc. Nat. Acad. Sci. U.S.A.*, **21**, 51–56.
- Thompson, A. R. and Bracewell, R. N. (1974), “Interpolation and Fourier transformation of fringe visibilities”, *Astron. J.*, **79**, 11–24.