Analog And Digital Signal Processing, Calibration And Image Processing In Radio Telescopes

Rohan Chabukswar
04026007

Under the direction of

Prof. R. K. Shevgaonkar
Indian Institute of Technology Bombay

November 11, 2006

Indian Institute of Technology Bombay

Department of Physics

Contents

List Of Figures And Tables

List of Figures

List of Tables

Chapter 1

Introduction

Radio Interferometry in Astronomy was developed following World War II, during which attention was drawn to the possibilities it could offer by the effects of solar activity on radar receivers. A radio analog of the Michelson Interferometer was constructed by Ryle and Vonberg in 1964, using dipole antenna arrays at 175 MHz.

A somewhat different instrument, with which a number of important results were obtained, was the sea interferometer. It was used in Australia where horizon-looking radar antennas had already been installed. Radiation from sources rising over the eastern horizon was received reflected from the sea as well as directly, causing interference. The frequencies were 40-400 MHz, the middle of the range being most satisfactory because of ionospheric effects at low frequencies, and sea roughness at high frequencies, even though the low altitude maximized ionospheric fluctuation and refraction effects. The sudden appearance of a rising source was also useful in separating individual sources.

During the 1950s and early 1960s most large radio telescopes were of the non tracking array type. In non tracking arrays, the antenna beams are usually directed toward the meridian at some desired declination and the rotation of the earth scans the beams in right ascension. The Ooty Radio Telescope in Udhagamandalam, India is of this type, with an array of more than 1000 dipoles approximately 0.5 kilometer long in the North-South direction.

These were succeeded by one dimensional tracking, and then two dimensional arrays. In 1976 a new technique of interferometry was developed in which the receiving elements were separated by such a large distance that it was expedient to operate them independently with no real time communication link. This technique is called Very Long Baseline Interferometry (VLBI).

The largest two dimensional tracking array in the meterwave range, is the Giant Meterwave Radio Telescope, situated at Khodad near Pune, India. Fourteen of the thirty dishes in GMRT are located more or less randomly in a compact central array in a region of about 1 square kilometer. The remaining sixteen dishes are spread out along the 3 arms of an approximately 'Y'-shaped configuration over a much larger region, with the longest interferometric baseline of about 25 km. The design is based on what is being called the 'SMART' concept - for Stretch Mesh Attached to Rope Trusses, which is an important technological breakthrough achieved by Indian scientists and engineers.

The voltages induced by cosmic-source radiation are generally referred to as signals, although they do not contain information in the usual engineering sense. Such signals

are generated by natural processes and have the form of Gaussian random noise, that is, the voltage as a function of time at the terminals of a receiving antenna can be described as a series of very short pulses of random occurrence which combine as a waveform with Gaussian-distributed amplitude. For most sources these characteristics are stationary with time, and the signals are ergodic, that is, ensemble and time averages are equivalent. The noise power in a receiving system is often specified in terms of the temperature of a matched resistive load that would produce an equal power level in an equivalent noise free receiver.

The flux of signals coming from the source as a function of $(\zeta,\ \eta)$ which are coordinates on celestial sphere, is called the brightness distribution of the region. The flux as a function of spatial frequency in wavelength units or $(u,\ v)$, is called visibility. The signals from different antennas, when multiplied together with a specific delay (called cross-correlation) gives us the visibilities for antenna spacings corresponding to $(u,\ v)$. These can be converted to the brightness distribution by just a Fourier Transform, for small regions across the sky for which the curvature of the celestial sphere can be neglected.

Here certain aspects of system design, analog and digital are considered. The calibration and Fourier transform of visibility data, getting a map of the area being studied, and techniques of processing that further improve the image are also discussed.

Chapter 2

Design Of The Analog Receiving System

This chapter is concerned with the analog equipment between the antennas and the correlators and, in particular, those characteristics that are critical to the accuracy of visibility measurements. They include phase stability, frequency responses, spurious signals and automatic level control. Specifications of tolerances on system parameters that are consistent with goals of sensitivity and accuracy can be accrued from this analysis.

2.1 Principal Subsystems Of The Receiving Electronics

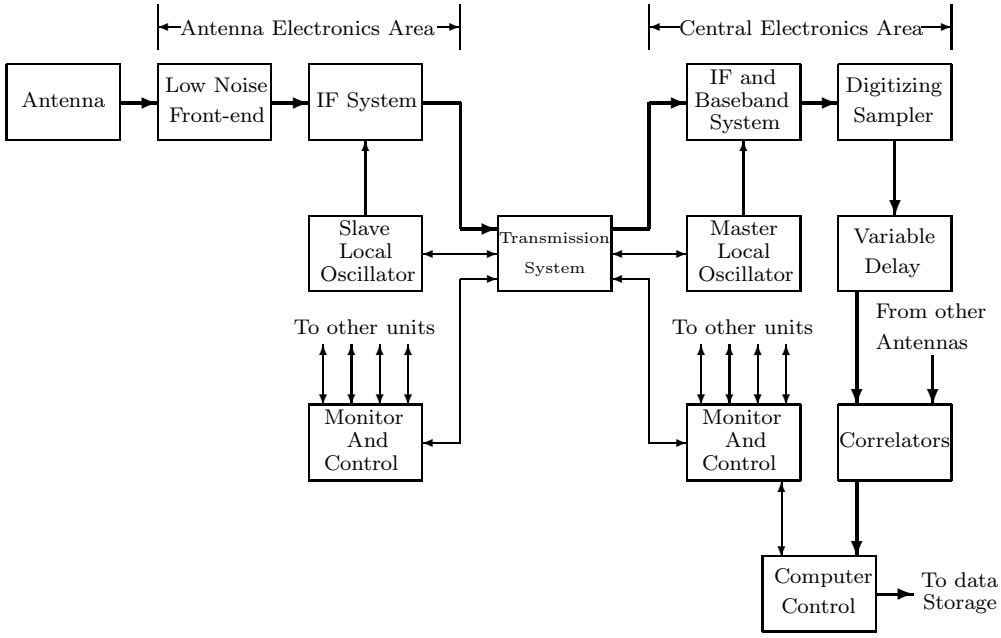

Figure 2.1: Schematic Diagram Of The Receiving System Of A Typical Synthesis Array

Figure 2.1 shows a simplified schematic diagram of the receiving system of a typical

synthesis array. The monitor and control blocks which comprise the digital communication system through which the computer monitors voltages, sets local oscillator frequencies, etc. Monitor and control interconnections to the other blocks have been omitted for simplicity. Except for the master local oscillator and the computer, one of each block is required per antenna. The heavy line shows the path of the received signal.

2.1.1 Low-Noise Input Stages

To achieve high sensitivity in a receiving system, the noise in the antenna and the input stages of the receiver be as low as possible. The signals and noise levels being specified in terms of temperature,

$$T_S = T'_A + (L_1 - 1)T + T_{R1}L_1 + T_{R2}L_1G_1^{-1} + T_{R2}L_1(G_1G_2)^{-1} + \ldots.$$

where, T_S is the overall system temperature, T_A, known as the antenna temperature, is used to express power received from radio source. If P_r is the power received from the source, T_A is defined as

$$P_r = kT_A\Delta\nu.$$

$L1$ is power loss factor of transmission line from antenna to first stage, T is the temperature of the line, T_{Ri} is the noise temperature of i^{th} receiver stage and G_i its power gain. If the first stage is a mixer, G_1 may be less than 1, in which case the second stage noise temperature becomes important. Modern methods to minimize noise temperature involve cryogenic cooling of amplifiers and mixer stages from input to the point where noise from succeeding stages is unimportant. This is referred to as the front-end subsystem.

2.1.2 Local Oscillator

Local Oscillator signals required at antennas and at several points along signal paths to correlators. Frequencies of the oscillators at different antennas must be maintained in phase to preserve coherence. The phases of the oscillators need not be identical, but the differences must be stable enough so that calibrating would be necessary only once in several hours. This can be achieved by transmitting one or more reference frequencies from central master oscillator, which is used to phase lock other "slave" oscillators. These oscillators then synthesize mixer frequencies by heterodyne frequency conversion.

Mixing enables major part of signal processing to be performed at an intermediate frequency (IF) appropriate for amplification, transmission, filtering, delaying, recording etc. The device which performs this operation is called a mixer, which adds/subtracts the local oscillator frequency to the signal frequency. This is done by applying the signal to be converted plus the local oscillator waveform to a circuit element with nonlinear I-V characteristics, e.g., diode. The second order term in V gives rise to product terms which have frequencies of $\nu_{LO}-\nu_0$ and $\nu_{LO}+\nu_0$. Other terms lead to other combinations, but the signal is filtered ahead so that only signals from bands centered at $\nu_{LO} - \nu_0$ and $\nu_{LO} + \nu_0$ are admitted. One of them can further be removed or both can be used depending upon whether the system under design takes in a single-sideband or a double-sideband.

To control frequency of sinusoidal fringe variations in correlator output, a continuous phase change can be inserted into one of local oscillator signals, called fringe rotation.

Such a reduction of output frequency reduces quantity of data to be processed, since output has to be sampled at Nyquist Frequency to prevent data loss. Here, reduction to zero frequency, called fringe stopping, is preferred. Special techniques like using a complex correlator instead of a simple one are then required to extract amplitude and phase of output. These special phase shifts are also required at mixers, which can be best implemented by digital synthesis methods, e.g., by dividing a frequency of 50 MHz to 100 kHz by a counter. Phase can then be controlled in steps of 500^{ths} of a rotation by loading initial count. The input stage of counter can also be designed to accept one pulse as zero or two, which can be controlled by a variable rate generating circuit, for even more accurately controlled frequency offsets.

2.1.3 IF And Signal Transmission Subsystems

After being frequency-shifted, signals pass through various IF Amplifiers and a transmission system to reach correlators. This transmission is usually effected by coaxial or parallel-wire lines, a waveguide, optical fibers or a microwave radio link. However, cable attenuation for long distances require many line amplifiers in a cable layout, so waveguides or radio links are more commonly used. These elements can be buried at 1 - 2 meters to reduce temperature variations. Transmission frequencies range from 10 - 100MHz in coaxial cables, to 10 GHz in waveguide or radio and to optical frequencies in fibers.

At the receiving end signals are converted to a final IF where band filters and time delays are added. As phase errors due to temperature and delay setting errors can be minimized by using lowest possible IF, here amplifiers here are baseband. Transmission system may also be used to distribute reference frequencies for local oscillator subsystems, in which case time or frequency multiplexing has be used to separate outgoing and incoming signals.

2.1.4 Delay And Correlator Subsystems

The delay and correlator subsystems at the central electronics area can be either analog or digital, depending upon the system. The usual method of implementing an analog delay system contains series of binary valued switchable delay units, so that all delays from 1 to $2^n - 1$ times the smallest delay unit can be achieved from n such units.

Analog multiplying circuits basically take the logarithms of two signals, add them, and take the antilogarithm of the sum. This can be done by e.g., an opamp with diodes connected appropriately, as diodes give approximately exponential behavior. Another implementation of a multiplying circuit, using bipolar junction transistors, has been analyzed in [4].

A spectral line correlator, used for spectral analysis, contains a bank of filters, where each filter defines a channel.

Now-a-days, digital circuitry operating at frequencies upward of 100 MHz led to the practice of digitizing final IF signal, so that delay and correlation can be done digitally with greater precision.

2.1.5 VLBI Systems

VLBI systems use independent frequency standards at local oscillators and record signals at antennas themselves. The output from each antenna electronics area is recorded on a storage device, such as magnetic tapes. These recording devices can be considered as transmission-and-delay units of a variable delay. However, the discussions below pertaining to the phase stability of filters, frequency responses of signal channels and automatic level control, also apply to VLBI systems.

2.2 Local Oscillator And Phase Stability

2.2.1 Phase Measurement Schemes And Phase Stability Of Filters

The phases of the local oscillator signals are affected by temperature variation, diurnal and annual, rotating joints, flexible cables, etc. Path length variations can be monitored by phase of a signal that traverses the path and back, called round-trip phase measurement systems. Examples of these include Swarup And Yang system, frequency offset round-trip system, automatically correcting systems and phase locked loops with reference frequencies.

Tuned filters used for selecting local oscillator frequencies are also a source of temperature-related phase variations. The phase response of a filter changes across the 3 dB bandwidth, so filters of narrow bandwidth should be avoided. To pick out a particular frequency from a series of closely spaced harmonics it may be preferable to use a phase-locked oscillator rather than a filter. This also applies to filters which carry the receiving signal at observing frequency or IF.

2.3 Frequency Response Of The Signal Channels

2.3.1 Optimum Response

When signals in a synthesis array pass through a large number of amplifiers, filters, mixers and transmission lines, the characteristics of these instruments are impressed upon the signals, which necessitates characterization of frequency response of these elements to get the overall frequency response of the receiving channel, and its effect on sensitivity and accuracy of visibility measurements.

To examine the tolerable deviations of the responses, the effects on decrease in signal-to-noise ratio and determining gain factors, a factor D can be defined as the signal-to-noise ratio relative to that of identical rectangular responses of width $\Delta\nu$:

$$\mathscr{D} = \frac{\int_0^\infty H_m(\nu)\, H_n^*(\nu)\, d\nu}{\Delta\nu \int_0^\infty |H_m(\nu)|^2 |H_n(\nu)|^2\, d\nu}.$$

The distortions considered include amplitude slope across passband, sinusoidal amplitude ripple, displacement of center frequency, variation in phase response and delay setting error. Table 2.1 shows the values of these parameters for loss of sensitivity no more than 2.5% and and gain error less than 1%.

Type Of Variation	Criterion	
	2.5% Degradation In SNR	1% Maximum Gain Error
Amplitude Slope	3 dB edge-to-edge	2.7 dB edge-to-edge
Sinusoidal Ripple	2.9 dB peak-to-peak	2.0 dB peak-to-peak
Center-Frequency Displacement	$0.05\Delta\nu$	$0.007\Delta\nu$
Phase Variation	12.8° RMS	9.1° RMS
Delay-Setting Error	$0.12/\Delta\nu$	$0.05/\Delta\nu$

Table 2.1: Tolerances On Frequency Response Variations

2.3.2 Delay-Setting Tolerances

Inaccuracies in adjustment of compensating time delays result from errors in calibration of antenna position or delay devices. They can be minimized by using a calibration source close to source being mapped. They can also result from discretely adjustable delays. The error in this case has equal probability of being anywhere from $-\tau_0/2$ to $\tau_0/2$ for each, leading to a triangular distribution of phase difference, with a maximum error $\pm\tau_0$, and RMS error $\tau_0/\sqrt{6}$. One method of eliminating this error is to make delay step $\tau_0 = 1/\nu_d$, so that adjustments involve integral number of rotations of phase. This, of course, requires that the IF bandwidth be small compared to center frequency, so that latter does not vary much with bandpass distortions.

2.3.3 Implementation Of Bandpass Tolerances

To implement all the tolerances derived above, the specifications of filters used should provide for required matching of responses and temperature effects. The frequency selectivity of the elements in earlier stages can then be held to the minimum required for rejecting interference, thus minimizing the effect on overall response.

2.4 Advantages And Implementation Of Phase Switching

An early method of obtaining the product of two voltages was to periodically reverse the phase of one, so that both $(V_1 + V_2)$ and $(V_1 - V_2)$ are obtained, and the difference gives the product. This method is still useful to eliminate small offsets in correlator outputs due to imperfection in circuit operations or spurious signals. This method reduces spurious signals by two or more orders of magnitude.

Periodical switching of phase can be done by using square waves, however, this is practical only for a small number of antennas. For a larger number of antennas, Walsh functions are used. Walsh functions are rectangular waveforms where the time interval between two transitions varies but remains an integral multiple of a fixed time interval. Walsh functions with even symmetry are represented by $\mathrm{Cal}\,(n,\ t)$ and those with odd symmetry by $\mathrm{Sal}\,(n,\ t)$, where n is the number of transitions that take place within a

time interval t. Figure 2.2 shows a few Walsh Functions. The waveform required at the

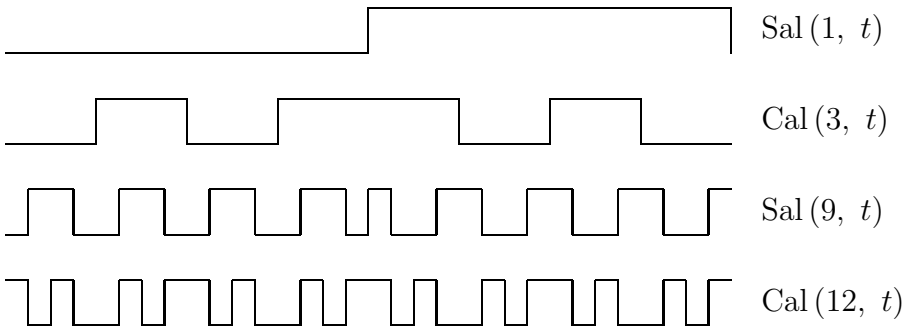

Figure 2.2: Walsh Functions

phase detector is the product of the functions at the two antennas involved. The product of two Walsh functions is a Walsh function, the sequency of which is greater than or equal to the difference between the sequencies of original functions.

2.4.1 Interaction Of Phase Switching With Fringe Rotation And Delay Adjustment

Effectiveness of phase switching is reduced if unwanted signals enter after phase switching. The following three cases illustrate the most important cases:

1. If the unwanted signal enters after phase switching, fringe rotation, and compensating delays, it is not suppressed by synchronous detection.

2. If the unwanted signal enters after phase switching but before fringe rotation and delay compensation, it appears at the correlator output and can cause a spurious response if it transiently matches with a frequency component of Walsh function.

3. If the unwanted signal enters after phase switching and fringe rotation, but before compensating delays, oscillations are lower and coincidence with a frequency component of Walsh function can be avoided.

2.5 Automatic Level Control And Gain Calibration

Automatic Level Control (ALC) is used to hold the level of total signal constant, by using a fraction of the total signal to control a variable gain element of the signal chain. Due to this, gain made to vary inversely as $(T_S + T_A)$, which can change substantially by ground radiation in side-lobes and atmospheric attenuation. A noise source can be switched on and off at a frequency of a few hundred hertz, and the resulting component and sampled and monitored. Of course, this does not cover gain changes due to mechanical deformation, which must be calibrated.

Chapter 3

Digital Signal Processing

The use of digital rather than analog instrumentation to introduce compensating delays and measure correlation has important advantages. Accuracy of delay depends upon the accuracy of timing pulses, which is achieved more easily in digital than analog delay lines. There is no distortion of the signal by digital units, except for calculable effects of quantization, while in analog systems, it is difficult to keep the shape of frequency responses within the tolerance required. Also, multichannel output is obtained more readily in digital systems, whereas it requires filter banks in analog systems, which can be a source of phase instability if subjected to temperature variations. Apart from that, digital circuits require less on-board adjustment than analog (except at high bit rates) and are better suited for replication, for the circuit elements which are required at all antennas.

3.1 Bivariate Gaussian Probability Distribution

As mentioned before, radio astronomy signals are generated by natural processes and have the form of random Gaussian noise. To analyze two such signals, the bivariate Gaussian probability distribution is applied:

$$p\left(x,\ y\right) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp\left[\frac{-\left(x^2+y^2-2\rho xy\right)}{2\sigma\left(1-\rho^2\right)}\right]$$

where, ρ is the correlation coefficient of the two signals,

$$\rho = \frac{\langle xy\rangle}{\sqrt{\langle x^2\rangle\,\langle y^2\rangle}}.$$

3.2 Periodic Sampling

3.2.1 Nyquist Rate

For baseband with upper cutoff frequency $\Delta\nu$, the function can be fully specified by samples with sampling frequency of $2\Delta\nu$, called Nyquist rate, or greater. This can also be applied to bandpass functions, that is, if the spectrum is nonzero within $n\Delta\nu$ to $\left(n+1\right)\Delta\nu$, the Nyquist rate is again $2\Delta\nu$. Sampling at frequencies greater or less than Nyquist frequency is referred to as oversampling or under-sampling.

3.2.2 Correlation Of Sampled But Unquantized Waveforms

Constructing a sampler which does not quantize the signals is not practical, but it has been studied for the sake of comparison to quantized versions. For such a signal,

$$\rho = \frac{\langle x(t)\, y(t)\rangle}{\sqrt{\left\langle [x(t)]^2\right\rangle \left\langle [y(t)]^2\right\rangle}}.$$

If x and y have the same variation σ^2,

$$\langle x(t)\, y(t)\rangle = \rho\sigma^2$$

Then the output of the correlator is

$$r_\infty = N_q^{-1} \sum_{i=1}^{N_q} x_i y_i.$$

Now, $\langle r_\infty\rangle = \rho\sigma^2$, as x_i, y_i follow same statistics as $x(t)$, $y(t)$. Hence, variance of the output is given by

$$\sigma_\infty^2 = \left\langle r_\infty^2\right\rangle - \langle r_\infty\rangle^2.$$

Thus,

$$\begin{aligned}\left\langle r_\infty^2\right\rangle &= N_q^{-2} \sum_{i=1}^{N_q}\sum_{k=1}^{N_q} \langle x_i y_i x_k y_k\rangle \\ &= N_q^{-2} \sum_{i=1}^{N_q} \left\langle x_i^2 y_i^2\right\rangle + N_q^{-2} \sum_{i=1}^{N_q}\sum_{k\neq i} \langle x_i y_i x_k y_k\rangle\end{aligned}$$

From above it can be shown that,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2 y^2 p(x,\ y)\, dxdy = \sigma^4\left(1+2\rho^2\right),$$

so first term is given by $\sigma^4(1+2\rho^2)N_q^1$. The second term can be evaluated using fourth order moment relation

$$\langle z_1 z_2 z_3 z_4\rangle = \langle z_1 z_2\rangle\langle z_3 z_4\rangle + \langle z_1 z_3\rangle\langle z_2 z_4\rangle + \langle z_1 z_4\rangle\langle z_2 z_3\rangle$$

and equals

$$\left(1 - N_q^{-1}\right)\rho^2\sigma^4.$$

So, finally,

$$\sigma_\infty^2 = \sigma^4 N_q^{-1}(1+\rho^2).$$

Thus the signal-to-noise ratio with unquantized sampling is given by

$$\mathscr{R}_{sn\infty} = \frac{\langle r_\infty\rangle}{\sigma_\infty} = \rho\sqrt{N_q},$$

if $\rho \ll 1$. If the data is sampled at βN_q, the equations are modified so that,

$$\mathscr{R}_{sn\infty} = \frac{\rho\sqrt{\beta N_q}}{\sqrt{1 + 2\sum_{q=1}^{\infty} R_{\infty}^2 (q\tau_s)}}, \tag{3.1}$$

where,

$$R_{\infty}(q\tau_s) = \frac{\beta \sin(\pi q/\beta)}{\pi q} \text{ (correlation coefficient for the samples).}$$

For $\beta = \frac{1}{2}, \frac{1}{3}, \frac{1}{4}$, etc., $R_{\infty} = 0$, and sensitivity drops with decreasing beta. For $\beta > 1$,

$$\sum_{q=1}^{\infty} R_{\infty}^2 (q\tau_s) = (\beta - 1)/2,$$

so there is no gain in sensitivity as expected, since no information is lost in Nyquist sampling.

3.3 Sampling With Quantization

In different sampling schemes, sampling is done after quantization, while in others, quantization is done after sampling. Here, ideal behavior of both the sampler and digitizer is assumed, so that both schemes lead to identical results. Quantization generates new frequency components in the signal, so the signal is not longer band limited. Hence sampling at Nyquist rate (of unquantized waveform) causes information losses.

3.3.1 Two-Level Quantization

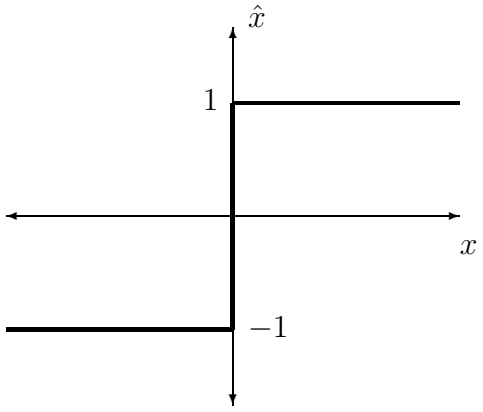

Figure 3.1: Two Level Quantization

A two level quantization as shown in Figure 3.1 can be achieved by simply amplifying and clipping the signal. The average correlation coefficient is given by

$$\rho_2 = \frac{(N_{11} + N_{\bar{1}\bar{1}}) - (N_{\bar{1}1} + N_{1\bar{1}})}{N},$$

where,

$N_{ij} =$ average number of products for which sample 1 is i, sample 2 is j $= NP_{ij}$.

Thus,

$$P_{11} = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \int_0^\infty \int_0^\infty \exp\left[\frac{-(x^2+y^2-2\rho xy)}{2\sigma(1-\rho^2)}\right] dxdy = \frac{1}{4} + \frac{1}{2\pi}\sin^{-1}\rho = P_{\bar{1}\bar{1}}.$$

and

$$P_{\bar{1}1} = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \int_{-\infty}^0 \int_0^\infty \exp\left[\frac{-(x^2+y^2-2\rho xy)}{2\sigma(1-\rho^2)}\right] dxdy = \frac{1}{4} + \frac{1}{2\pi}\sin^{-1}\rho = P_{1\bar{1}}.$$

Combining the results,

$$\rho_2 = 2(P_{11} - P_{\bar{1}\bar{1}}) = \frac{2}{2\pi}\sin^{-1}\rho$$

This is called the Van Vleck relationship. Calculating the signal-to-noise ratio as for the unquantized case,

$$\mathscr{R}_{sn2} = \frac{\rho\sqrt{N}}{\pi\sqrt{1+2\sum_{q=1}^\infty R_2^2(q\tau_s)}}, R = \frac{2}{\pi}\sin^{-1}\left[\frac{\beta\sin(\pi q/\beta)}{\pi q}\right]$$

which is the autocorrelation coefficient of signal after quantization. For $\beta = 1$, $\frac{1}{2}$, $\frac{1}{3}$, etc., $\sum_{q=1}^\infty R_2^2(q\tau_s) = 0$, and signal-to-noise ratio is 64% of that for unquantized sampling. For oversampling at $\beta = 2$ and $\beta = 3$, signal-to-noise ratio is enhanced from $\mathscr{R}_{sn2}$ by a factor of 1.17 and 1.21 respectively.

3.3.2 Four-Level Quantization

The use of two bits to represent the amplitude of each sample results in less degradation of the signal-to-noise ratio. A two bit, four-level quantization is shown in Figure 3.2, where the quantization states are $-n$, -1, 1, n, and the threshold values $-v_0$, 0, v_0. Thus, the correlation coefficient is given by

$$\rho_4 = \frac{2n^2N_{nn} - 2n^2N_{n\bar{n}} + 4nN_{1n} - 4nN_{1\bar{n}} + 2N_{11} - 2N_{1\bar{1}}}{(2n^2N_{nn} + 2N_{11})_{\rho=1}}.$$

Calculations by the method used for two level quantization or by using Price's Theorem gives

$$(\rho_4)_{\rho\ll1} = \rho\frac{2}{\pi}\frac{[(n-1)E+1]^2}{\Phi + n^2(1-\Phi)},$$

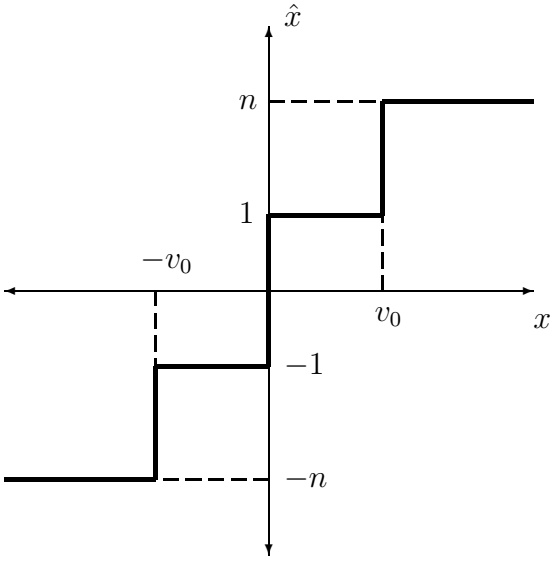

Figure 3.2: Four Level Quantization

where,

$$\Phi = \text{erf}\left(\frac{v_0}{\sigma\sqrt{2}}\right),$$

which leads to

$$\mathscr{R}_{sn4} = \frac{2\rho\left[(n-1)\,E+1\right]^2\sqrt{N}}{\pi\left[\Phi+n^2\,(1-\Phi)\right]\sqrt{1+\sum_{q=1}^{\infty}R_4^2\,(q\tau_s)}}.$$

Cooper (1970) [2] has examined values of n and v_0 that maximize $\mathscr{R}_{sn4}$ and has shown that values of $n = 3$, $v_0 = \sigma$ and $n = 4$, $v_0 = 0.95\sigma$ both result in the maximum signal-to-noise ratio of 0.88 of that for unquantized sampling. For over/under sampling,

$$R_4 = \frac{2\left[(n-1)\,E+1\right]^2 R_\infty}{\pi\left[\Phi+n^2\,(1-\Phi)\right]}.$$

For $\beta = 2$ signal-to-noise ratio comes out to be 0.94 times that for unquantized sampling.

3.3.3 Three-Level Quantization

If the lower products ± 1 are counted as 0 in four level quantization, a three level quantization is obtained. Proper adjustment of equations in four level quantization leads to

$$(\rho)_{\rho\ll 1} = \rho\frac{2E^2}{\pi\,(1-\Phi)}$$

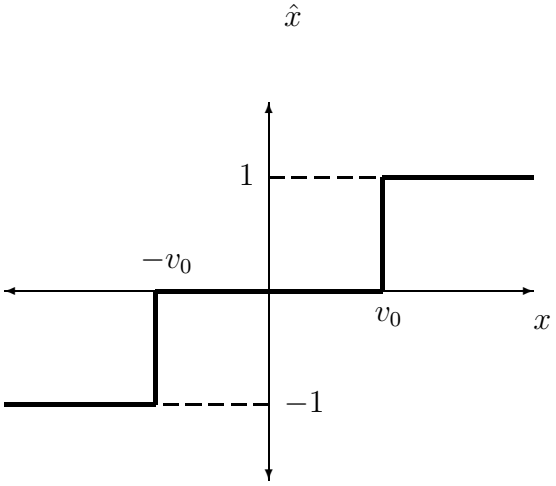

Figure 3.3: Three Level Quantization

and

$$\mathscr{R}_{sn3} = \frac{2\rho\sqrt{N\beta}E^2}{\pi\left(1-\Phi\right)\sqrt{1+\sum_{q=1}^{\infty}R_3^2\left(q\tau_s\right)}}.$$

In the three bit quantization, optimum v_0 is given by 0.612σ. For $\beta = 1$, signal-to-noise ratio comes out to be 0.81 times as much as from unquantized case. For $\beta = 2$, it increases to 0.89.

3.4 Comparison Of Quantization Schemes

Number Of Quantization Levels (Q)	Sensitivity Relative To Unquantized Case $\eta - Q$	
	$\beta = 1$	$\beta = 2$
2	0.64	0.74
3	0.81	0.89
4	0.88	0.94

Table 3.1: Efficiency Factors For Various Quantization Schemes

The efficiency factors for the three quantization schemes are given in Table 3.1. The two important factors that make lower level quantizations preferable are the complexity of the further digital circuits and the storage space required. The performance seen practically agrees to the theoretical prediction to a few percent. Higher quantization levels are now implemented, with a change in technology which facilitates massive storage. 16 level quantization results in sensitivity of 97% as much as unquantized, which is almost perfect reproduction, at a very reasonable storage of 2 bytes per sample.

3.5 Accuracy In Digital Sampling

3.5.1 Principal Causes Of Errors

The most serious error in digital sampling is a small offset voltage in definition of zero-level, causing small offsets in correlator output. These, however, can be largely eliminated by implementing phase shifting. Also, offsets can be measured by counters to compare the numbers of positive and negative samples, and assuming equal distribution of these, the correction to the offset can be controlled dynamically.

3.5.2 Tolerances In Three-Level Sampling

It can be calculated that if all threshold values are within $\pm 10\%$ of v_0, the output is always with 10^{-3} of expected correlator output. Gain errors can be corrected if the actual threshold values are known, by dividing the correlator outut by arithmetic mean of numbers of higher level signals. 10% errors in threshold then result in less than 1% error in ρ.

Another type of error that can be caused is by threshold level not being precisely defined, but affected by direction of change of signal, rate of change of voltage, hysteresis, etc. This error can be modeled by having an indecision region at the boundary,and if a signal falls in that region,it can take either of the values associated with the threshold randomly and with equal probability. Calculations reveal that for 1% maximum error, indecision region can be as large as 18% of threshold value. For maximum error of 0.1%, the indecision region has to be within 6%.

3.6 Digital Delay Circuits

To implement a digital delay circuit, a series of shift registers each with delays of 1, 2, 4, 8, etc. times the clock cycle can be used like binary cable lengths in analog delay system. Another way would be to use two shift registers to obtain a delay that is in variable increments of clock interval. Apart from these implementations, RAMs can also be used to store data for variable periods. Another useful technique is serial-to-parallel conversion, which divides the frequency of the digital signal. In all these schemes, the smallest delay increment is reciprocal of sampling frequency. Even finer delays can be obtained digitally by varying the timing of the sample pulse in number of steps, or by using small analog delay lines, which is quite practical for fractional delays.

3.7 Quadrature Phase Shift of Digital Signal

Complex correlators for digital signals can be implemented by introducing quadrature phase shift in the analog signal and then using separate samplers for the actual signal and the phase shifted version. The Hilbert transformation that represents the phase shift can also be performed directly on the digital signal, thus eliminating all quadrature networks. A truncated sequence of the form $(-\pi\tau)^{-1}$ provides convolving function for the digital data to introduce required phase shift. Such a truncation however, results in

ripples and degrades the signal-to-noise ratio by few percent. Also, summation process increases number of bits, and loss of information must be faced if the complexity of the correlators ahead is not to be increased.

3.8 Digital Correlators For Continuum Observations

Three distinct cases of quantization can be imagined, with none, one or both of the signals to be multiplied can be quantized at either 2 - 3 levels or higher levels.

1. Correlators with both signals 2 - 3 level quantized are very simple to construct, with digital multiplier circuits for 2 or 3 bits.

2. Correlators with one signal quantized at 2 - 3 bits and one quantized at a higher level can be implemented by a register which accumulates the higher quantized number, with the lower quantized number just specifying whether the other number is to be added, subtracted or ignored from the accumulator.

3. Correlators with both numbers quantized at higher levels can be best and fastest implemented by having a Read Only Memory (ROM) store the possible values of product, and the inputs are used to specify the address of the product.

Chapter 4

Calibration And Fourier Transformation Of Visibility Data

4.1 Calibration Of The Visibility

The purpose of calibration is to remove the effects of instrumental factors in the measurements. The stable instrumental factors like antenna position coordinates that specify the baselines, pointing corrections resulting from axis aligning tolerances and zero point settings of instrumental delays vary only as a result of major changes, like the complete relocation of an antenna, and can be calibrated by observing sources with known positions. It is assumed that they have been determined in advance and accurately known values are used in controlling the array.

Calibration of visibility instruments for effects that vary during an observation principally include predictable effects like constant component of atmospheric attenuation as function of zenith angle, variation of antenna gain with altitude because of elastic deformation of antennas, shadowing of one antenna by another at low spacings and elevation angles, as well as directly estimable effects like variation of gain from ALC action because of variation in system noise, phase variations in LO system monitored by round trip phase measurement systems and variable component of atmospheric delay monitored using water vapor radiometers.

4.1.1 Use Of Calibration Sources

To calibrate the gains, an unresolved source can be observed, with the phase reference position (field center) chosen to coincide with the source being mapped. Then the response is

$$\mathscr{V}(u,\ v) = [\mathscr{V}(u,\ v)]_{\text{uncal}} \left(\frac{S_c}{\mathscr{V}_c}\right),$$

where $[\mathscr{V}(u,\ v)]_{\text{uncal}}$ is corrected only for predictable or directly measurable effects, and so is $\mathscr{V}$. Since gain varies with time, it is necessary to interrupt the mapping observations for periodic measurements of the calibrator, and interpolated for the observation time. The frequency of calibration depends upon stability of instrument and can range from around 15 minutes to several hours. The calibrator to be used should be strong, so that a good signal-to-noise ratio is obtained in short time, to minimize loss in $(u,\ v)$ coverage lost in

mapping while calibrating, which is more serious for linear arrays than two dimensional ones. The calibrator should also be unresolved so that precise measurements of visibility as a function of celestial coordinates are not necessary. Also, calibrating all spacings in an array are not necessary, just enough so that all antennas are included. The position of calibrator should be close to that of mapped source, so that effects of atmosphere or antennas varying with pointing angle can be minimized. This also reduces the time lost between changing antenna positions for mapping and calibrating is kept small.

4.1.2 Calibration Of Spectral Line Data

Calibration of spectral line data requires a bit more work. In principle, continuum methods could be applied to each channel, but integration time for each channel would need to be more for a good signal-to-noise ratio. The channel to channel differences are relatively stable with time, so they need not be calibrated as frequently as the gain. First, responses of all channels are summed up to simulate continuum response, and continuum methods applied equally to all channels. Then bandpass calibration observation is done to determine relative gains of all channels. The bandpass calibrator should be unresolved, provide a good signal-to-noise ratio in the channels and have flat spectrum, but it need not be close to mapped source and can be observed before or after the mapping observations.

4.2 Derivation Of Brightness From Visibility

4.2.1 Model Fitting

Fitting of brightness models for deriving the brightness was practiced extensively during early radio interferometry. However, sometimes the brightness map differs from that of a point source by only a slight broadening. Parameters then can best be determined by fitting model visibility function directly to observed values.

4.2.2 Mapping By Direct And Discrete Fourier Transformation

This is the most straightforward method of obtaining brightness distribution from visibility. For small celestial areas, the measured visibility is the Fourier transform of the brightness distribution of the source.

$$\mathscr{V}_m(u,\ v) = W(u,\ v)\, w(u,\ v)\, \mathscr{V}(u,\ v),$$

where $W(u,\ v)$ is transfer function and $w(u,\ v)$ is any weighting applied. The Fourier transform is given by

$$B_m(\zeta,\ \eta) = B(\zeta,\ \eta) * * b_0(\zeta,\ \eta),$$

where

$$b_0(\zeta,\ \eta) \rightleftharpoons W(x,\ v)\, w(x,\ v).$$

Discrete Fourier transformation can be implemented by the Fast Fourier Transform (FFT) algorithm. This algorithm drastically reduces the time complexity of Fourier transform, but it is necessary to evaluate visibility at points in rectangular grid, which brings in the possibility of aliasing of of parts of image from outside synthesized field.

It may be assume that the best technique for griding is interpolation, but the problem of aliasing still remains, and best way to deal with this is to convolve the data with a function which is stable in the map, but falls off rapidly toward the edges.

Functions which are separable into 1-D functions of the same form for u and v are the most convenient to consider as convolving functions. A rectangular function performs poorly in aliasing but the computation required is minimal. A Gaussian function was widely used in earlier days, but it is not much stable in the map. A combination of the two, a Gaussian sinc function is much better than either individually, and doesn't require much more computation.

Features aliased onto a map from outside the boundary include not only images of features of the sky but also random variations resulting from system noise. This noise increases toward edge of map, where the signal-to-noise ration degrades rapidly.

4.3 Practical Considerations

4.3.1 Possible Causes Of Errors In Maps

The distribution of visibility corresponding to a suspicious feature in a map may indicate a problem related to particular group of antennas for specific time. For example, a persistent error in an antenna pair will be distributed on an elliptic ring in visibility and elliptical feature with a Bessel function sort of radial distribution in brightness. On the other hand, a short duration error introduces two delta functions corresponding to measurement and its conjugate, and thus the effect is diluted in the brightness map. Thus a single short duration error may be negligible, but a persistent low-level error could compound into a spurious feature on the map. Other mapping errors result from interference, correlator offset errors, etc. as also response to radio emission of the sun from the side-lobes. Multiplicative errors include gain constants of antennas which can arise from calibration or phase errors introduced by the atmosphere. Distortion effects of non-coplanar baselines, bandwidth and visibility averaging increase with distance from the center of the map. Errors in pointing of antennas introduce errors in the map which increase toward the half-power points in the beam, where the rate of change of collecting area with angle is generally greatest.

4.3.2 Visibility At Low Spatial Frequencies

A problem common to all synthesis arrays is that the minimum antenna spacing cannot be less than diameter of antenna, and practically much greater. This limit does not apply to increments in longer spacings, but a central hole in $(u,\ v)$ coverage exists, resulting in broad negative side-lobes. The situation can be improved by inserting visibility at origin from independent measurement of flux and applying sufficient weight to them. Data from a smaller interferometer may also be added at the center.

Chapter 5

Image Processing And Enhancement

The limited distribution of spatial frequencies and errors in measurements are the deficiencies in visibility data that limit accuracy of synthesis maps. Such errors are usually in the phase in VLBI, so phase is almost always uncalibrated. Certain obvious errors like negative brightness are also introduced in the map, so it is possible to improve the result by injecting some a priori information about the source.

5.1 Limitation Of $(u,\ v)$ Coverage And Methods Of Compensation

5.1.1 CLEAN Algorithm

Observed brightness can be regarded as true brightness convolved with synthesized beam

$$B_0(\zeta,\ \eta) = B(\zeta,\ \eta) * *b_0(\zeta,\ \eta) \rightleftharpoons \mathscr{V}_0(u,\ v) = \mathscr{V}(u,\ v)\left[W(u,\ v)\, w_u(u,\ v)\, w_t(u,\ v)\right].$$

However, analytical deconvolution is not an option here, because the weighted transfer function contains a lot of zeros and cannot be divided out. Any improved procedure must involve placing non-zero visibility values in the unmeasured $(u,\ v)$ areas. As these can be filled with infinitely many values, which leads to an infinite number of solutions. Thus the procedure should also allow visibility at the unmeasured points to take in values most consistent with a reasonable brightness distribution. Judgments can be made as to what brightness distribution is reasonable and incorporated into the map, for example, extensive sinusoidal structure and negative brightness values are very obviously instrument artifacts. One of the more successful of such procedures is the CLEAN algorithm developed by Högbom in 1974. The procedure is to break down the point-source responses, and replace each one with the corresponding responses to a clean beam, i.e., a beam free of side-lobes.

The steps involved are as follows:

1. Compute the map and the response to a point source by Fourier transformation of the visibility and weighted transfer function, to get a dirty map and the dirty beam.

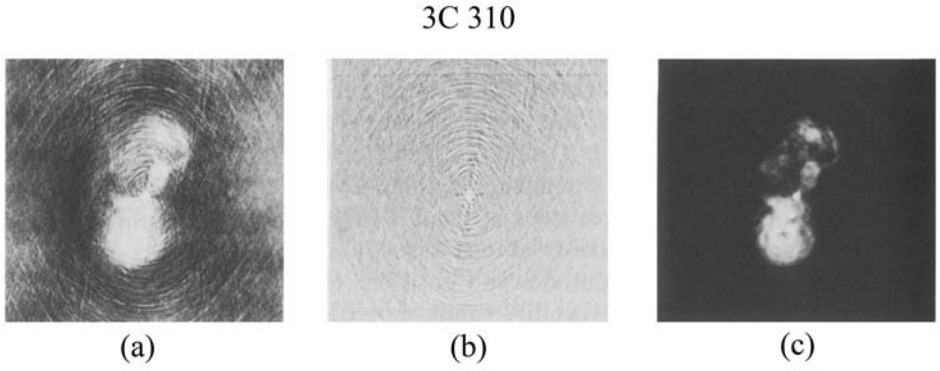

Figure 5.1: Performance of the CLEAN algorithm illustrated by images of the radio galaxy 3C10 from observations with the VLA. (a) Fourier transform of the visibility data. Side-lobe structure is at the 2% level. (b) The corresponding synthesis beam. (c) The image produced by the Cornwell (1983) version of the CLEAN algorithm. From [5].

2. From highest brightness point on the map subtract the dirty beam corresponding to a peak amplitude of γ times the peak amplitude in map. γ here is called loop gain.

3. Repeat Step 2 till all significant source structure is removed from map, indicated by comparison of highest peak with RMS level pf rest of map, first time the RMS level fails to decrease when a subtraction is made, or when significant numbers of negative component start to be removed.

4. Add the removed components, in the form of clean beam responses, to the residual brightness distribution to obtain map. Clean beam can be a Gaussian with FWHM equal to that of original beam.

Experience shows that CLEAN, developed initially for random point sources, also works well for extended sources. Step 4 de-emphasizes high spatial frequencies which may be spuriously extrapolated from delta function responses. At the point where subtraction is stopped it is assumed that residual brightness distribution consists mainly of noise. Retaining this distribution in the map is a non-ideal procedure, but necessary to prevent misinterpretation. The background noise indicates the level of uncertainty in brightness values.

5.1.2 Implementation And Performance Of The CLEAN Algorithm

Being highly nonlinear, CLEAN does not yield readily to a complete mathematical analysis. However, if number of delta function responses in map does not exceed the number of independent visibility data, CLEAN converges to a least squares fit of the Fourier transforms of delta function components to measured visibility (Schwarz, 1978,1979). To

try to maintain this condition of convergence, it is usual to apply CLEAN to only a small window in the map. To clean a map of given dimensions, it is necessary to have a beam pattern of twice the dimensions of the map so that a point source can be subtracted from anywhere in the map. In practical implementations, however, it is convenient to have the map and beam dimensions equal. In such a case it is recommended that the map obtained from initial Fourier transformation should have dimensions twice that of final map required.

Several arbitrary choices influence CLEAN — γ, window area and the criterion for termination. γ lies between 0.1 and 0.5, and gives better results when it is at the lower range, but at a cost of higher computation time. The number of subtraction cycles required is $-\log \frac{\mathscr{R}_{sn}}{\log(1-\gamma)}$.

A well known problem with CLEAN is the generation of spurious spots or ridges as modulation on broad features. This can be explained by the fact that negative side-lobes of the dirty beam add new maxima which are detected and subtracted in subsequent cycles, resulting in a tendency of component-subtraction points to be located at intervals equal to spacing of first side-lobe.

A modification of CLEAN by Cornwell (1983) minimizes $\sum w_i \left|\mathscr{V}_i - \mathscr{V}_i'\right|^2 - \kappa s$ instead of $\sum w_i \left|\mathscr{V}_i - \mathscr{V}_i'\right|^2$, where κ is adjustable and s is measure of smoothness, like negative mean squared brightness of the map. It was also shown that inclusion of the extra term only requires a modification of original synthesized beam before implementing CLEAN.

5.1.3 Constrained Optimization Techniques

Constrained Optimization Techniques include a class of algorithms which produce a map agreeing with observed visibility to within noise level, constrained by maximizing some measurement of image quality.

Maximum Entropy Method is one of the, where a function F' (called entropy) is defined which is determined entirely by B' as a function of solid angle. F' is maximized within the constraint that Fourier transform of B should match visibility values. The two most used forms of F' are $F_1 = -\sum_i \frac{B_i'}{B_s'} \log\left(\frac{B_i'}{B_s'}\right)$, $B_s' = \sum_i B_i'$ and $F_2 = \sum_i \log B_i'$. Maximum Entropy Method is clearly an effective mapping algorithm, and both forms of entropy work equally well. However, these are not the only functions which can be used. The first derivatives of an entropy function should tend to infinity as B' approaches zero, so that maximizing F produces positivity in brightness. Its second derivatives everywhere should be negative, to favor uniformity in brightness. A general class of functions which have the property $\frac{d^2F}{dB'^2} < 0$ and $\frac{d^3F}{dB'^3} > 0$ were studied in [6], many of them are effective as entropy. CLEAN is the most widely used algorithm, but Maximum Entropy Method and similar algorithms become more and more competitive for maps with broad smooth features for which CLEAN requires lot of subtractions, and don't have CLEAN's problem of introducing spurious structure to broad features.

5.2 Mapping With Incomplete Phase Data

5.2.1 Mapping With The Visibility Modulus Only

Circumstances in which phase data is completely missing include measurements with uncalibrated phase data, or with two antennas. These conditions do not usually occur in large arrays, but the problem is worth discussing for the general insight it provides. Also it is usually of importance in optical interferometry. The Fourier transform of the square modulus of the visibility is equal to the autocorrelation of the brightness distribution:

$$|\mathscr{V}(u,\ v)|^2 = \mathscr{V}(u,\ v)\,\mathscr{V}^*(u,\ v) \rightleftharpoons B(\zeta,\ \eta) * *B(-\zeta,\ \ -\eta) = B(\zeta,\ \eta) \star\star B(\zeta,\ \eta).$$

The question of whether it is possible to obtain a unique solution for B from its autocorrelation function has been a long standing one in image processing. Without phase data, the center of the field cannot be determined, and there is a 180° rotational ambiguity in the position of the map. Apart from this, it is indeed possible in two dimensions to obtain maps from visibility data alone, using constraints on positivity and confinement of image.

5.2.2 Mapping With Uncalibrated Phase Data

Mapping with uncalibrated phase data is of great practical interest, as instrumental limits sometimes inhibit calibration of phase and may even complicate amplitude calibration. Relative values of uncalibrated visibility measurements can be extracted from closure relationships between phases of three different antennas. The number of equations obtained by the phase closure relationships is $\frac{1}{2}(n-1)(n-2)$. Amplitude closure equations require 4 antennas each. The Readhead iterative mapping technique which incorporates closure relationships consists of the following steps:

1. Obtain an initial trial map based on inspection of visibility amplitudes and any a priori data like a map from a different wavelength or epoch.

2. For each visibility amplitude, determine a complete set of independent amplitude and/or phase closure equations. For each such set compute a sufficient number of visibility values.

3. Solve for complex visibility corresponding to each antenna spacing. Apply CLEAN and select positive components within window area as new trial map. Test for convergence and return to step 2.

Another group of programs involve what is called the self-calibration technique. With an iterative method it is possible to use almost any source as a self calibrator by using long spacings only to get relative phase data. The self calibration technique is as follows:

1. Make an initial map.

2. Calculate ratios of uncalibrated observed visibility to model visibility.

3. Determine antenna gains for each period of integration.

4. Use the gain to calibrate observed visibility values and make a map.

5. Use CLEAN to provide positivity and confinement of image.

6. Test for convergence and return to step 2.

The two techniques described are basically equivalent but differ only in details of approach and implementation. In both cases, any initial map, like a point source, converges to the solution, but inaccuracies in initial map need higher number of iterations.

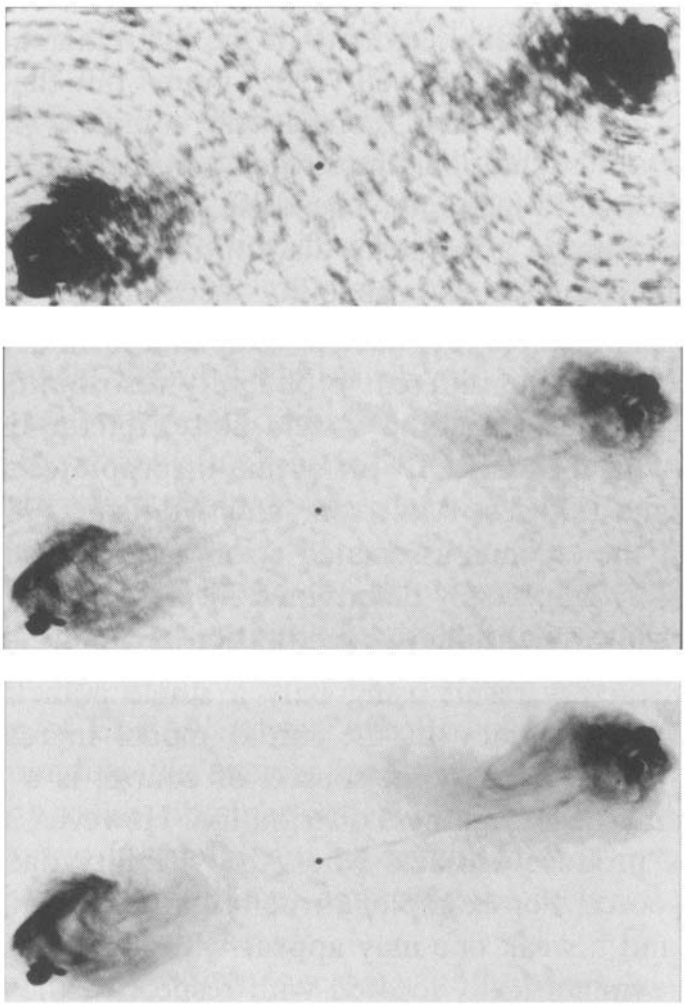

Figure 5.2: Three stages in the reduction of the observation of Cygnus A. The top image is the result of transformation of the visibility data using the FFT algorithm. The calibration source was approximately 3° from Cygnus A. The center image shows the effect of reduction using the maximum entropy algorithm. This compensates principally for the under-sampling in the spatial frequencies and thereby removes side-lobes from the synthesized beam. The result is similar to that obtainable using the CLEAN algorithm. The bottom image shows the effect of the self-calibration technique, the maximum entropy image being used as the initial model. This final step improves the dynamic range by 3. In observations where the initial calibration is not as good as in this case, self-calibration usually provides a greater improvement. The long dimension of the field is 2.1 arcmin and contains approximately 1000 pixels. From [7].

Chapter 6

Summary

The different ways of designing the receiving and analyzing systems, analog and digital have been seen. The way calibration is done of the received data, and the methods of obtaining a "dirty map" of the region being studied have also been discussed. The different algorithms and enhancements used to obtain a clean image of the area have also been studied. The systems described illustrate the principles involved, and methods of implementation have changed, and are expected to change more as digital technologies develop. The techniques of image enhancement described can be largely characterized as quasi-empirical since procedural details have not been fully analyzed, and include user selected parameters, the most effective choice of which rests on experience. This development is limited by available computing facility, and as this increases, new techniques continue to arise.

An important aspect of all electromagnetic radiation coming from outer space is its polarization. Studying the polarization of the radio signals gives valuable information about the kind of natural processes that lead to the radiation. For example, most synchrotron radiation shows a small degree of polarization, which increases with angular resolution and helps to reveal the distribution of magnetic fields within the source. Radio telescopes today have polarization receivers which record the four Stokes' parameter of the radiation. Different maps might then be made of different polarizations like linear circular etc. In considering the response of interferometers and arrays the question of polarization has been ignored. This simplification can be justified on the basis of the assumption that we have been dealing with completely unpolarized radiation, but the decoding and analysis, as also the system response and digitization of this information forms an entire chapter of system design.

Acknowledgments

I would like to thank Prof. R. K. Shevgaonkar for guiding me through this literature survey for my B. Tech. Seminar, and his constant encouragement and invaluable suggestions.

I would also like to thank Prof. S. H. Patil for being my co-guide for this project.

Rohan Chabukswar
04026007
November 11, 2006

Bibliography

[1] *Interferometry And Synthesis in Radio Astronomy*, A. Richard Thompson, James M. Moran, George W. Swenson, Jr., John Wiley & Sons, 1986.

[2] Cooper, B. F. C,. Correlators With Two-Bit Quantization, *Aust. J. Phys.*, **21**, 521-527, 1970.

[3] Cornwell, T. J., A Method Of Stabilizing The CLEAN Algorithm, *Astron. Astrophys.*, **121**, 281-285, 1983.

[4] Frater, R. H., Accurate Wideband Multiplier Square-Law Detector, *Rev. Sci. Instrum.*, **35**, 810-813, 1964.

[5] Napier P. J., A. R. Thompson and R. D. Ekers, The Very large Array: Design And Performance Of A Modern Synthesis Radio Telescope, *Proc. IEEE*, **71**, 1295-1320, 1983.

[6] Narayan, R. and R. Nityananda, Maximum Entropy Image Reconstruction - A Practical Non-Information-Theoretic Approach, *J. Astrophys. Astron.*, **3**, 419-450, 1982.

[7] Perley, R. A., J. W. Dreher And J.J. Cowan, The Jet And Filaments In Cygnus A, *Astrophys. J.*, **285**, L35-L38, 1984.

[8] Thompson, A. R., B. G. Clark, C. M. Wade and P. J. Napier, The Very Large Array, *Astrophys. J. Suppl.*, **44**, 151-167, 1980.