B. Clark

GENERAL CONSIDERATIONS FOR A VERY LONG BASELINE INTERFEROMETER

I. Introduction

It is proposed to build a very long baseline interferometer (VLBI) in order to investigate the structure of radio sources having features with the dimensions of the order of 10^{-3} to 10^{-2} seconds of arc. In order to achieve such high resolution at convenient radio frequencies, baselines of several thousand miles are required. In conventional interferometer systems coaxial cable or radio links are used both to carry the local oscillator signal to provide coherent oscillators at the two elements and to return the i.f. signal to a common correlator. Obviously, for baselines of thousand of miles the cost of coax cable is prohibitive while the use of existing telephone lines does not provide adequate bandwidth. Microwave radio links could be used, but they are very expensive to operate and difficult to procure. Also little is known about their overall phase stability.

Clearly, it would be desirable to be able to operate the individual elements independently which would avoid the cost and inconvenience of an elaborate communications link. This can be achieved by using atomic standards both as local oscillators and as a time reference, and by recording the intermediate frequency outputs on high speed magnetic tape separately at each end of the interferometer to be later correlated in a digital computer.

Such a scheme is not a new idea and it has been discussed in rather general terms by several radio astronomy groups. Recently a short article appeared in the Soviet journal Radiophysica 8, 651, 1965, which briefly discussed the requirements for this type of interferometer. Although the feasibility of such an arrangement has been realized for some time, it was thought until very recently that all radio sources would be resolved by rather modest baselines of a few hundred thousand wavelengths which could be obtained by using more or less conventional techniques. This conclusion was based on the results of several earlier interferometer studies.

The Nancay and CalTech groups, using baselines of a few thousand wavelengths found that about half of all extragalactic radio sources had dimensions greater than one minute of arc. At the same time workers at Jodrell Bank achieved baselines up to 68,000 wavelengths by using radio links between stations 72 miles apart. The long baseline studies indicated that only one to two percent of the extragalactic radio sources were unresolved at this spacing and it was thought that few sources had structure under one second of arc.

However, recently four independent lines of research have indicated that there are a number of sources which have components well under 0″01. Firstly, the English group have extended their baselines first to 180,000 wavelengths and then to 600,000 wavelengths and have found sources, including in particular a number of quasi-stellar sources, which have components which are unresolved with a resolution of .1 seconds of arc. A second piece of evidence comes from studies of scintillations of radio sources in the interplanetary medium made at Cambridge, Arecibo, Australia, and in the Soviet Union. From this work it is recognized that in general only sources having significant features under one second of arc show scintillations and that other conditions being equal the scintillation amplitude increases with decreasing angular size. While the sources known to be approximately 0.1 seconds of arc or less, are all good scintillators, some scintillate much more readily than others indicating dimensions one or possibly even two orders of magnitude smaller. Thirdly, studies of the intensity of extragalactic radio sources as a function of frequency have disclosed sharp low frequency cut-offs in some of their radio frequency spectra. These cannot be due to simple synchrotron radiation and it is thought that the cut-off is due to synchrotron self-absorption. If this is indeed the case the brightness temperature must be exceedingly high and the corresponding source dimensions in the range 0".001 to 0".01. Sources with spectra of this type are generally associated with distant quasi-stellar sources or with very active regions in the nuclei of galaxies. Angular dimensions of 0".001 imply linear sizes of only a few parsecs no matter how distant the source. Finally, the finding of rapid time variations in the radio and optical luminosity of some of the quasi stellar sources in the United States and in the Soviet Union again indicate linear dimensions of only a few parsecs. These very small dimensions are important in that they require exceedingly high densities of relativistic particles, resulting in the radiation of some 10^{45} ergs/sec from a very small volume. Direct confirmation of the existence of such small radio sources is important and investigations with resolutions $\sim$ 0".01 will no doubt play a major role in any attempt to understand the origin and evolution of radio galaxies and quasi stellar sources.

Extension of interferometer baselines using radio links cannot be continued much further. Already the Jodrell Bank-Malvern interferometer uses two repeater stations over an 80 mile path producing a baseline of 600,000 wavelengths, at 21-cm. It is likely that in the near future the resolution of this instrument will be doubled by going to 10 centimeter wavelength. Increases beyond this do not seem practical using direct radio links and it is likely that the Jodrell Bank group will also attempt correlating independent receivers, presumably one in the United Kingdom and one in the USSR.

The system proposed here, once perfected, could be used with already existing radio telescopes generally making use of existing instrumentation. Only the accurately controlled local oscillator and data recording systems need to be transported to the individual installations. This would provide a variety of baselines necessary to obtain a reasonably detailed structure of radio sources with resolutions up to 0.001 seconds of arc. Thus, in the fullest sense this would be a national and hopefully eventually an international operation and it seems that the NRAO is the logical institution to assume the responsibility for its development.

II. Proposed System

The system initially proposed would use the NRAO 140-foot telescope together with the 1,000 spherical dish at the Cornell University Arecibo Ionospheric Observatory in Puerto Rico. The distance between these two stations is 1600 miles giving a baseline of 5.2×10^6 wavelengths at an operating frequency of 611 megacycles. This would give a fringe spacing of 0!03 allowing the resolution of sources somewhat under 0!01.

Choice of Operating Frequency

The choice of frequency is a compromise between the longer effective baselines obtainable at higher frequencies and the greater signal strength and simplicity of radiometer systems at the low frequencies. 611 Megacycles has been chosen as an initial operating frequency for the following reasons:

a) It is close to the highest frequency that the Arecibo dish can presently achieve high efficiency, and the dish is presently instrumented to operate at this frequency.

b) A relatively simple radiometer and feed can be constructed for the 140-ft antenna at this frequency.

c) This baseline gives nearly one order of magnitude increase over the highest resolutions currently available.

d) 611 megacycles is in a protected radio astronomy frequency band.

e) Adequate signal-to-noise ratio is achieved with relatively simple receivers for a reasonable sample of sources.

f) At this frequency the combined effect of ionospheric ~~and~~ atmospheric and the atomic local oscillator phase fluctuations are near a minimum ~~although a somewhat higher frequency may be preferable in this respect.~~

It is hoped that in perhaps one year's time the 1000 ft dish may be usable at 20 cm at which time the system could really be adapted to this frequency giving a factor of 2-1/2 increase in resolution.

20 centimeters is a logical choice in that the NRAO already has a low noise system operating at this frequency. High frequency operation is desirable as there is growing evidence that at centimeter wavelengths many sources contain fine structure not detectable at longer wavelengths.

Local Oscillator System

The independent local oscillators will be derived from atomic standards. Rubidium controlled crystal oscillators are available for about 10,000 dollars which give a stability of one part in 10^{11} over a one week period. The output of these units are at radio frequencies which can readily be multiplied up to the desired operating frequency. At 611 megacycles the stability is approximately 6×10^{-3} cycles/second. It should be stressed that there will be no intent to recover any phase information. In order to measure the angular extent of the source the amplitude of the fringe visibility is sufficient. Only sufficient phase stability is needed to maintain coherence. Thus the r.f. phase shift must be small during the integration time required to detect the source. With a stability of 10^{-11}, coherent integrations up to ~ 50 seconds may be used.

Time Keeping

If the IF bandwidth is 300 kilocycles, it will be necessary to have the times at the two observatories synchronized to better than one microsecond. This can be achieved using the atomic standards used to derive the LO frequency, which will provide a time reference within an uncertainty of about 1 microsecond drift per day. Once the time difference between the two clocks is established by direct comparison, any subsequent small drift may be determined directly from the astronomical measurements. This is discussed further below.

Radiometer Systems

By illuminating 600 feet of the Arecibo dish and using the 140-ft telescope, the interferometer will have a sensitivity equivalent to two 290 ft antennas. This gives a sensitivity of about 1.2 degrees antenna temperature per flux unit. The Arecibo Ionospheric Observatory is currently instrumented at 611 megacycles. NRAO has no facilities for this frequency but a simple feed and crystal mixer receiver could easily be constructed. A parametric amplifier would give about a factor of two increase in the signal to noise at the NRAO and about $\sqrt{2}$ increase in the interferometer sensitivity. It is not clear if this is worth the trouble. With a root mean square system temperature of 300 degrees and an IF bandwidth of 300 kilocycles, RMS noise fluctuations will be equivalent to about 1.2 flux units in a one second integration period. Thus with integration times of the order of one minute we should be able to detect quite a number of sources even if they are significantly resolved. At 1400 mc the situation is reversed. NRAO is fully instrumented at this frequency while AIO is not. Until it is determined whether or not the Arecibo dish will work at 20 cm, it seems best to plan on initial operating frequency near 600 megacycles. Any existing radiometers will have to be modified to provide a second conversion to bring the intermediate frequency down to the range 0 to 300 kilocycles in order to obtain reasonable sampling rates.

Calibration

No attempt will be made to directly calibrate the fringe visibility. Instead the relative intensities of the fringe amplitude of sources suspected to be unresolved will be compared with their known flux densities at zero spacing. Those showing the same relative intensities at the two spacings may be assumed to be unresolved. Some check is provided by checking sources to see if the fringe amplitude remains constant with hour angle. However, due to the limited checking range of the Arecibo dish, this is not a very sensitive test.

In the unlikely event that there are no unresolved sources at this spacing, a rough absolute calibration may be achieved by the measurement of the ratios of the source antenna temperatures to the receiver temperatures and multiplying this factor times the absolute correlation coefficient to obtain the flux corresponding to the observed fringes. This technique is probably much less accurate than the method of ratios, but may be resorted to in case of need.

Effect of Atmosphere and Ionosphere

Phase fluctuations due to variations of the differential path length to the atmosphere or ionosphere can cause loss of coherence. The former will be most serious at high frequencies while the latter becomes an important consideration at low frequencies. The total delay in the ionosphere will be of the order of 100 to 200 nanoseconds. The differential delay between the two stations may be expected to be some reasonable fraction of this and may fluctuate with a period ~ 10 seconds. Experimental and theoretical studies of these effects are in progress at the

NRAO with regard to the design of a very large array. Although no detailed experimental results are yet available, there is no evidence from interferometer systems operating in this country or the 600,000 wavelength interferometer in England that phase fluctuations will be a limiting factor at short decimeter wavelengths over the baselines discussed here.

Initial Operation

Initial testing of the VLBI should be performed using existing interferometer systems allowing separate testing of the local oscillator system and the intermediate frequency data recording as well as the digital correlation. The NRAO two element interferometer is ideally suited for this prupose and only after the system is thoroughly operational on this instrument will operation over a longer baseline be attempted.

III. Data Output

The *i*.f. output of the individual radiometers may be recorded in one of three different forms: as a direct analogue tape recording, as the output of an A/D converter operating at a frequency faster than the IF frequency, and as the output of a clipper recorded as digital bits (essentially the output of a one bit A/D converter).

In all of these methods the basic difficulty is synchronization. In order that not too large an area of tape must be searched to find the necessary delay setting, the absolute synchronization of the clocks must be known to an accuracy of a few tens of times the reciprocal of the IF bandwidth. In order that this necessarily time consuming search need not be repeated for every observation, the clocks, once synchronized for the day, should remain synchronized to an accuracy of one or two times the IF bandpass for the few hours that may pass between the observation of unresolved sources.

Let us consider the analogue method of recording the IF signals. Then, when the signals are brought to a central point for correlation, there are two ways to proceed: the signals may be processed in a machine constructed especially for this purpose, or it may be immediately converted by an A/D converter and then processed digitally. In the latter case, the considerations are much the same as in the case of direct digital recording, and will be considered together with that case. In either case the output of the master clock would be recorded on the tape along with the IF information to supply the synchronization.

Let us consider the requirements of a device to perform the first correlation of the tapes from an interferometer operating at 600 Mc/s between Green Bank and Arecibo. Firstly, the synchronization must be accurate to a fraction of the reciprocal bandwidth. If no variable delay buffer (a difficult feat, especially for analogue signals) is provided, then one tape drive must be slaved to the other. It is unlikely that this slaving can be achieved to better than a degree or so of capstan revolution, about 20 microseconds at 150 ips. This imposes an upper limit of about 30 kc/s bandwidth, so that it may not be possible to utilize the full bandwidth inherent in the recorder.

Even given an electronic "signal stretcher," this correlating device must have several other features that are difficult to obtain.

Firstly, the delays must be variable between the two tapes at intervals (depending on the bandwidth) of a few seconds in order to keep the correlator in the "white light" fringe of the interferometer. Secondly, it must have an output every millisecond in order to avoid smearing the fringes, or else have built in a device which phase detects the fringes. These functions are sufficiently complicated that it would seem most profitable to have this correlator controlled by and linked to a small stored program digital computer. This special correlator would appear to be a sufficiently complicated device that it would be cheaper and easier, at least in this first attempt, to do the correlating in a large general purpose digital computer.

Doing the correlation in a digital computer consumes a great deal of computing time, perhaps much more computer time than telescope time. With this thought in mind, the experiment should be designed to require a minimum of computation, which is to say that the maximum amount of information be placed in each number that the computer must process. It is known that about half of the information in a noise-like signal is contained in the first bit, so that a one-bit correlator recovers all the correlation with a degradation in signal to noise ratio of about 1.4 at the advantage of handling a much smaller number of bits than any other procedure.

If the one-bit conversion is done before recording, it also solves some of the difficulties in the recording. The master clock may drive a shift register so that each bit comes from a precisely defined interval, and if the tape unit is controlled from a slower output of the master clock, it may be started at a given clock time. Further, if a standard digital recorder is used as the tape recorder, the tapes may be fed directly into the general purpose digital computer for processing with no further ado. This appears to require a minimum of equipment in terms of A/D conversion and in the tape controller.

Bandwidth Considerations

First, the number of bits of data required to detect a given source is independent of the bandwidth used and is $\gamma\left(\frac{T_a}{T_r}\right)^2$ where T_a and T_r are the antenna temperature of the source and the receiver temperature, respectively, and γ is a numerical coefficient of order 10 depending on the certainty required in the detection and the degree of correlation between adjacent bits. Since the computer processing and the physical length of the tapes are the controlling factors in the processing rather than the telescope time required, the bandwidths are determined by considerations other than the required signal to noise ratio. Firstly, there is a maximum bandwidth set by the maximum speed of commercially available recorders. The fastest computer type tape drives now on the market operate at a rate of 120 Kc/s with six bits plus parity, or 720 Kb/s. On the other hand, we would like the maximum integration time to be shorter than the time in which any system parameter may change appreciably. As mentioned earlier, if the local oscillator stability is one part in 10^{11}, they will become uncorrelated in a time $\sim 10^{11}/(6 \times 10^{8}$ c/s) ~ 3 minutes. The time during which the phase path through the ionosphere may change significantly is also of the same order. Since a reel of tape, holding perhaps 2×10^{7} characters, would last about 200 seconds at the rate of 120 Kc/s, this seems to be the optimum bit recording rate.

Given the bit recording rate, or more conveniently, Δt, the interval between bits, then the loss of correlation because of the finite bandpass of the receiver is

$$\int_0^\infty P(\omega) \frac{\sin\left(\frac{\omega\Delta t}{2}\right)}{\frac{\omega\Delta t}{2}} \, d\omega$$

where $P(\omega)$ describes the passband shape, and is normalized so that

$$\int_0^\infty P(\omega) \, d\omega = 1$$

There is a further requirement on bandpass, that the varying delay can be handled by simply shifting the two records relative to each other by an integer number of bits. At the worst possible case, when the delays are off by $\frac{\Delta t}{2}$, the loss in correlation is measured by

$$\int_0^\infty P(\omega) \cos \frac{\omega\Delta t}{2} \frac{\sin \frac{\omega\Delta t}{2}}{\frac{\omega\Delta}{2}} \, d\omega = \int_0^\infty P(\omega) \frac{\sin \omega\Delta t}{\omega\Delta t} \, d\omega$$

If $P(\omega)$ is square, with lower and upper limits ω_1 and ω_2 respectively, then the maximum and minimum loss of correlation are respectively

$$\frac{S_i(\omega_2\Delta) - S_i(\omega_1\Delta)}{\Delta(\omega_2 - \omega_1)} \quad \text{and} \quad \frac{S_i(\omega_2\Delta) - S_i(\omega_1\Delta)}{\Delta(\omega_2 - \omega_1)}$$

If we rather arbitrarily set $\omega_1 = 0$ and insist that the maximum loss of correlation be 30%, then the minimum loss of correlation is 10%, and the cutoff frequency is $2.5\Delta = \omega_2$. The encoding clipper and shift register could be used to set the effective bandwidth, but the variation of the coefficient of correlation with the delay offset might prove troublesome.

This completes the picture: feeding the digital tape recorder would be an IF with a frequency cutoff at about 280 kc/s, to supply a shift register at 720 kb/s, which in turn feeds a tape recorder at 120 kC/s.

The logic feeding the tape recorder would have to perform one more function, the generation of inter-record gaps. It would be desirable to break the records into groups of about 100,000 characters to facilitate machine processing, though this function could be performed in the digital computer at the cost of an extra pass to edit in the gaps, and a rather complicated program to handle the buffering of long records. However, it seems like a good idea to have a tape gap of 5 milliseconds or so each second in order to start the timer over again for the next record, so that the timing is not fouled up for the rest of the tape if a character is lost somewhere. This would require only a small modification to the commercial tape unit, to allow suppression of the parity bit to write blank tape.

IV. Data Processing

AIO has a CDC 3200 computer. By the time this system is put together NRAO will probably have an IBM 360/50 computer. These two appear to be about equally fast for this particular problem. There are three basic operations to be performed by the computer: the shifting to set the delays to the white light fringe, the bit-by-bit correlation, and the rectification and integration of the fringe pattern. The fringe rectification may be done before or after the correlation, by complementing the output of one IF for half a cycle of the fringe before correlation, or by multiplication by a sine wave after the IF's have been correlated and accumulated for about a millisecond. It is probably slightly more efficient to do the rectification after the correlation, but this can be looked into in more detail when the actual programming is to be done.

In any event, it appears likely that the computation of the amounts of the shifts and the shifting operations will take of the order of 1 microsecond per bit. The actual multiplication is simply an "exclusive or" and takes very little time to perform. The accumulation is probably most efficiently done by a table lookup to find the sums of 12 or 16 (depending on the machine) bits, and then adding to accumulate these sums. This can probably be done in about 1 microsecond per bit also. The rectification of the fringes and miscellaneous housekeeping will probably add another 1 microsecond per bit, so that the total processing time is about 3 microseconds per bit per pass. This must be repeated for at least two different delay settings to check the synchronization of the clocks, so at the end the total computer processing time may amount to an average of 10 microseconds per bit, or about 7 times the telescope time.

The initial synchronization of the clocks at the beginning of an observing session must be done by observing a strong unresolved or partially resolved source. It seems likely that unresolved signals with an antenna temperature of $\frac{1}{30}$ T_r are to be expected. These sources should be detected with only about 10^5 bits. With only 10^5 bits, a complete correlation may be done in only a few tenths of a second, so that one is willing to try of the order of 10^4 different correlations, which would suffice to search an area of time of about 10 milliseconds duration, so that 10 milliseconds absolute error in the clocks would be tolerated. Synchronization to this tolerance would appear to be very easy from observations of WWVL or similar means. Once the initial synchronization is determined, this value of the difference between the clocks is assumed for the next source, and if the source is detected, a new value of the synchronization is deduced, so that the clock synchronization is continually updated.

B. G. Clark
K. I. Kellermann

11/3/65