A Brief Historical Introduction to Very Long Baseline Interferometry

B. Benjauthrit
TDA Engineering Office

This article provides a short historical account of Very Long Baseline Interferometry, including the rationale, development, and experiments.

I. Introduction

The technique of Very Long Baseline Interferometry (VLBI) has been employed to achieve extremely high angular resolution in the study of radio sources. Unprecedented accuracies may be obtained for values of various geophysical quantities. As an example, it is now possible to compare time scales on separate continents with an uncertainty of only one nanosecond (10^{-9} sec). The main purpose of this article is to provide a short history of VLBI starting from the Michelson-Pease experiment of 1920 up to the present time. This includes the rationale, development, and various experiments behind the subject of VLBI. However, due to a large volume of references[1] on the subject, only certain aspects are covered here.

II. The Rationale Behind Radio Interferometry

Until recently images of celestial objects obtained from radio telescopes lacked the detail of those obtained from optical telescopes. The reason is that the resolution of a telescope increases with the ratio of its aperture to the wavelength of the received signal, and radio wavelengths are about a million times longer than light wavelengths. Thus, radio telescopes are fundamentally confined to a poor angular resolution when compared to optical telescopes. However, in practice this is not the case for two reasons (Ref. 164):

(1) The resolution of large optical telescopes is confined not by their size but by irregularities in the earth's atmosphere. The limit is about 1 arc second. This is about 100 times better than the unaided human eye. Also, due to longer wavelength, the fluctuation of the incoming signal path length through the atmosphere at radio frequencies is much smaller than that at light frequencies.

(2) The signal must be coherent, or in phase, over the entire dimensions of the telescope. Coherent radio waves are much easier to obtain than coherent light signals, so that radio telescopes can operate much closer to the theoretical resolution limit than optical telescopes.

III. Early Need for Radio Interferometry

Although resolution increases as wavelength decreases, the performance begins to deteriorate significantly when the wavelength approaches the dimensions of the structural imperfections in the antenna. Further, the largest antennas

---

[1] All references in this article refer to "An Extensive Bibliography on Long Baseline Interferometry" following this article.

available were paraboloidal type; they had the least precise surfaces. Thus, the best resolution that had been obtained with paraboloidal antennas did not depend strongly on wavelength and was about 1 arc minute. Although it is possible to build a more precise large antenna, it would not provide better resolution than 0.1 arc minute (Ref. 177). This is the reason why radio astronomers turned to interferometry, where in effect two relatively small antennas function as opposite edges of a single huge radio telescope.

The development of interferometry began as early as 1890 when Albert A. Michelson published a paper (Ref. 237) describing a technique of modifying an optical telescope to receive star light along two paths, and determining the angular diameter of a star by means of interference of its own light waves. However, it was not until 1920 that his technique was successfully demonstrated on the 100-inch telescope of Mt. Wilson, in collaboration with F. G. Pease. Figure 1 shows a sketch of the Michelson-Pease stellar optical interferometer. Here, a light wave is intercepted by two separated mirrors which reflect two beams to a common point where the beams are combined. If the path of one beam is made slightly longer or shorter than that of the other beam, the light waves in one beam will be out of phase with the waves in the other beam. When the beams are combined, one sees a pattern of alternating light and dark "fringes."

Attempts to extend the optical interferometer to longer baselines in order to achieve higher resolution have not been very successful, mainly due to the difficulty in obtaining coherent light waves from the system and in maintaining the alignment of the mirrors within a fraction of the wavelength. It is much simpler to solve this problem by using radio waves. Two or more antennas can be used to synthesize large apertures and achieve high angular resolution. With such an instrument, one can determine not only the size and shape of discrete radio sources but also their precise positions in the sky.

IV. Early Development of Radio Interferometry

Early radio interferometers were essentially radio analogs of the above well-known optical device. However, in radio interferometers, the signals are combined and compared (or cross-correlated) electrically. This idea had evolved from the discovery of extraterrestrial radio-frequency radiation by Jansky[2] in the early 1930's. Figure 2 depicts the radio equivalent first employed by Ryle and Vonberg for solar observations in the late 1940's.[3] In the same period, a radio Lloyd's mirror interferometer was set up in Australia. Meanwhile, McCready and colleagues (Ref. 230) used a similar technique to study solar radiation and its relation to sunspots.

Resolution obtained by radio instruments has been steadily upgraded by advanced technology in short wavelengths and increases in baseline lengths. Many radio sources are well resolved by an interferometer with a baseline of about one kilometer (km) and a resolution of the order of one arc minute. However, by the late 1950's, it was realized that further increases in resolution would be required to study the structure of those sources in more detail and to resolve the smaller sources.

V. Modifications to the Basic Interferometry

The basic interferometer system can be modified in many ways. One way is to employ several intermediate frequencies, each generated by an independent local oscillator. Another way is to use complicated local oscillator chains for greater frequency versatility and phase-compensating systems. Detailed discussions on four modifications, namely, lobe rotation, spectral-line interferometry, Very Long Baseline Interferometry, and intensity interferometry were given by E. B. Fomalont and M. C. H. Wright.[4] We deal here solely with the subject of VLBI.

VI. Type of Radio Source

It is now generally accepted that the radio emission of discrete sources such as radio galaxies and quasars is "synchrotron" radiation from electrons moving in a weak cosmic magnetic field at relativistic speeds, i.e., speeds close to the speed of light. Also, radio sources may essentially be classified into two different types. One has a large angular extent and is strongest at the longer wavelengths. The other is relatively compact and is strongest at the shorter wavelengths. There is no simple relation between the angular extent of the radio emission and the optical emission from galaxies and quasars, however compact radio sources are not confined to quasars; many are identified with the nuclei of galaxies. Furthermore, many quasars are large extended radio sources.

The large radio sources have a complex distribution of radio emission that typically extends over several hundred thousand light years of space, corresponding to angular dimensions of

---

[2] Jansky, K. G., "Electrical Disturbances Apparently of Extraterrestrial Origin," *Proc. Inst. Radio Engineers*, N.Y., Vol. 21, p. 1387, 1933.

[3] Ryle, M., and D. D. Vonberg, *Nature*, pp. 158–339, 1946.

[4] See Chapter 10 of *Galactic and Extragalactic Radio Astronomy*, edited by G. L. Verchuur and K. I. Kellermann, 1974.

between a few arc seconds to a few arc minutes. On the other hand, the compact radio sources are so small and their particle densities are so great that, since relativistic particles absorb radiation as well as emit it, the source becomes opaque to its own radiation at long wavelengths. The self-absorption cutoff frequency depends only on the flux density, the angular size, and weakly on the magnetic-field strength. The smaller the source, the shorter the wavelength at which it becomes opaque.

Since the compact sources are relatively weak at the long wavelengths, where most of the early observations of extragalactic radio sources were made, they remained essentially unnoticed for many years until sensitive receivers for short wavelengths became available. Further, William A. Dent (Ref. 95) discovered in 1965 that some compact sources such as the quasar 3C273 are variable as observed from their radio emissions.

VII. Development of VLBI

Longer baselines for conventional interferometers were not feasible partly because of the increased cost of the cable between the interferometer elements and partly because of human and natural obstructions such as roads, rivers, mountains, and ultimately oceans. To overcome such problems, radio relay links were employed (see Fig. 3). The idea is to join the radio telescopes by a microwave-relay link. Here the local-oscillator signal is transmitted to the mixers to reduce the high frequency signal down to an intermediate level for returning for correlation. In order for the intermediate frequency signals from each antenna to be coherent, they are generally derived from a common source. This technique was first used by Australian and British radio astronomers to obtain baselines of more than 100 km and resolutions better than one second of arc. By the early 1960's the efforts to extend the interferometer baseline were carried out successfully at the University of Manchester (Ref. 274). Various interferometric techniques have also been applied to polarization measurement[5,6] and the mapping of source brightness distributions (Refs. 142 and 321). However, in order to resolve the compact variable sources with expected dimensions of about 0.001 arc second, baselines of size comparable to the dimensions of the earth were needed. Moreover, large improvements in the resolution of radio link interferometers were not practical because radio links are limited to line-of-sight operation, and installation of large numbers of repeater stations was both costly and technically complex.

For many years radio astronomers had considered the possibility of completely eliminating the direct electrical connection between interferometer elements by separately recording the signals at each end on magnetic tape and later cross-correlating the two recorded signals. In order to utilize this technique successfully, the following requirements are necessary:

(1) The recordings of the two tapes must be synchronized so that the time when a given wavefront is received at each station is precisely known. Also, the time tags on the recordings should permit reasonably short searches for fringes in the correlation process.

(2) The actual radio signal frequencies are generally too high to be recorded directly on magnetic tape. Independent local oscillators must be employed to "heterodyne" the radio signal frequency (typically several gigahertz) to a much lower frequency so that it can be recorded on magnetic tape. If the two lower frequency signals are to be cross-correlated, then the oscillators must remain coherent over the observing time (or integration time). This implies that the relative phase change of the two oscillators must remain small during this period so that the change in frequency is less than the reciprocal of the recording time. For example, at a radio signal frequency of 10 GHz and a 100-second integration time, a frequency stability better than one part in one trillion ($\Delta f/f = 10^{-12}$) is required.

(3) Due to the time delay after heterodyning, phase rotation is also necessary.

A typical tape-recording interferometer system and a more sophisticated one may be seen in Ref. 164.

The idea of employing independent-local-oscillator tape recording interferometers was considered in the U.S.S.R. as early as 1961. However, stable frequency standards and wideband tape recorders needed for high sensitivity were not generally available then. The frequency standards generally perform two functions (Ref. 340):

(1) To generate sufficient phase stability that will ensure no significant loss of coherence in the two signal streams to be cross-correlated.

(2) To keep time with sufficient accuracy to insure that the clock offset error remains constant over the interval required for the determination of the instantaneous baseline vector.

---

[5] Ko, H. C., "Theory of Tensor Aperture Synthesis," *IEEE Trans. Antennas and Propagation*, Vol. AP-15, pp. 188–190, Jan. 1967.

[6] Morris, D. V., et al., "Preliminary Measurements on the Distribution of Linear Polarization Over Eight Radio Sources," *Astrophys. J.*, Vol. 139, pp. 560–569, Feb. 1964.

A tape-recording interferometer was first used in radio astronomy by a group at the University of Florida to study the dimensions of the radio storms on the planet Jupiter at a frequency of 18 MHz in 1965 (Ref. 40). Since the Jovian bursts are so intense, an integration time of considerably less than a second and a bandwidth of about one kilohertz (kHz) give an adequate sensitivity (requires $\Delta f/f = 10^{-3}$). Hence, a frequency stability of only one in a million and a time synchronization of about 1 millisecond from the National Bureau of Standards Station WWV were sufficient to maintain coherence at the two ends of the interferometer system. With this it was possible to determine on the surface of Jupiter the dimension of the radio emitting regions less than 1 arc second or about 320 km. This resolution is considerably better than the highest resolution in photographs made of Jupiter at optical wavelengths.

Tape-recording interferometers for studying the much weaker radio emission of radio galaxies and quasars had to wait until stable atomic frequency standards and wideband tape recorders were commercially available. At that time, two systems for tape-recording interferometry were developed independently in the United States (Refs. 12 and 248) and in Canada (Refs. 26 and 27).

A. Analog VLBI Development

The joint Canadian team from the National Research Council and the University of Toronto built an analog recording and processing system employing television tape recorders. At each site data were recorded at 4 MHz on video recorders (down converted from 408 MHz received frequency) on a 90-minute reel tape. Then rubidium frequency standards and a special analog correlator were used to process the data to simultaneously provide fringe patterns from a number of delays. These fringe patterns were then displayed on a multipen recorder.

B. Digital VLBI Development

A system built by two U.S. groups from the National Radio Astronomy Observatory and Cornell University (NRAO-Cornell) used a standard computer tape drive to record digital data in a 360 kHz band. Rubidium frequency standards were also employed here. The incoming signal (down converted from 610 MHz received frequency) was first bandlimited, infinitely clipped, sampled at 720 kilobits per second (kbps) and finally recorded on standard 7-track computer tape at a per-track density of 315 bits per centimeter; hence, a full reel of tape held three minutes of data. This system preserves only the zero crossing information (i.e., the phase) of the signal. The normalized cross-correlation function can still be estimated from this infinitely clipped data by applying a correction factor derived first by Van Vleck in 1943 (Ref. 382). A large general purpose computer was used to process the data.

VIII. Further Developments and Experiments on VLBI

The Canadian system has a greater bandwidth advantage but requires more complex special equipment to synchronize the tape on playback. In the U.S. system digital data were simply fed into a large general purpose computer which stored and correlated the two data streams. In early experiments 90 minutes of computer time was typically required to process a pair of tapes over 15 delays. Lately, only 50 seconds of computer time has been required to search over 7 delays. By using special-purpose digital correlators further reductions in computer time are possible.

Both the Canadians and Americans had experienced difficulties in their early attempts. The Canadians obtained their fringes first, both on a baseline in Ontario between Shirley Bay and Algonquin Park (183 km) and on a much longer baseline between Algonquin Park and Penticton, B.C. (3074 km).[7] The NRAO-Cornell group obtained fringes at 610 MHz between Riverside, Md. and Green Bank, W. Va. (228 km), followed by a successful experiment at 1.665 MHz between Westford, Mass., and Green Bank, W. Va. (845 km), with the same digital equipment, by a group at the Massachusetts Institute of Technology. A hydrogen maser frequency standard was used at the Massachusetts site.

The MIT group (Ref. 408), in cooperation with the Northeast Radio Observatory Corporation, Haystack Observatory, Westford, Massachusetts, undertook a program to improve the instruments to allow accurate measurements of group delays (the frequency derivative of the phase delay) to make possible precise determinations of vector baseline (Ref. 138), radio source positions (Ref. 18), polar motion, universal time, and earth tides (Ref. 346).

To measure the group delays with an accuracy greatly exceeding the inverse of the recorded bandwidth, the method of bandwidth synthesis was devised (Refs. 310 and 408) where the receiver passband is switched to sample signals over a wideband. This bandwidth synthesis technique is also employed in the VLBI development at JPL (Refs. 273, 368-376).

First long-baseline measurements with tape-recording were made in 1967 on the baseline across Canada and the U.S. at

---

[7] Their experimental results were presented at the Ottawa meeting of the International Scientific Radio Union on May 23, 1967 (Refs. 26 and 27).

radio wavelengths of 75, 50, and 18 centimeters. The results confirmed many small size galaxies and quasars. But many were still unresolved and higher resolution was needed. This was rapidly extended in a series of cooperative intercontinental experiments conducted in 1968 and 1969 by American, Swedish, and Australian radio astronomers. With a 6-centimeter wavelength, a 0.001-arc second resolution was obtained on the longest baselines. In spite of this, some small sources still remain unresolved. The 10^4 km California-Australia baseline was already more than 80% of the earth's diameter, so further significant increases in the physical baseline were not feasible without expensive procedures of setting up stations in space or on the moon.

A cheaper and simpler alternative was to observe at shorter wavelengths. Outside North America, however, only two radio telescopes were suitable for operation at short wavelengths and large enough to yield sufficient sensitivity for long-baseline interferometry, both of which were in the U.S.S.R. Due to distance, government regulations, lack of exchange in technology, and other problems, arranging such joint experiments was both difficult and formidable.

The first U.S.-U.S.S.R. experiment was completed late in 1969, using the baseline between the 43-meter radio telescope at NRAO and the 22-meter radio telescope on the shore of the Black Sea in the Crimea (Ref. 388). Later in Spring of 1971, a second experiment was conducted, involving, in addition to the above, the 64-meter radio telescope at Goldstone,[8] California and the "Haystack" telescope in northern Massachusetts.[9] More than 20 investigators from eight institutions in both countries participated and it included observations of interstellar clouds of water vapor as well as of radio galaxies and quasars.

The data from the Goldstone-Crimea baseline obtained at a wavelength of 3.5 cm (X-band) provided the highest resolution (about 0.0003 arc sec) achieved so far in the study of radio galaxies and quasars. The measurements of the water vapor clouds which were made at shorter wavelengths gave even higher resolution.

From the various data taken, it was evident that some quasars appear to expand faster than the speed of light. Martin J. Rees (Ref. 297) of the University of Cambridge has suggested the "super light velocity" theory for interpreting such a phenomenon. If the radio source expands at a velocity close to the speed of light, then since it takes a finite time for radio signals to reach an observer, the signal arriving from the receding part of the source will have originated at an earlier time, when it was closer to the point of origin than the signals from the approaching parts. Under these conditions the apparent velocity of expansion may in fact exceed the speed of light.

Evidence that this might be important in the quasar 3C279 was first obtained in a series of transpacific observations made between 1968 and 1970 by a joint team from Australia (at Tidbinbilla) and the California Institute of Technology (at Goldstone). This group found that a component of the source that had first appeared in 1966 had reached a diameter of about 0.001 arc second by the end of 1969. This corresponds to a linear diameter of about 12 light-years. Thus, the apparent expansion velocity was indeed about twice the speed of light, as theorized by Rees.

More detailed measurements were made of 3C279 in October, 1970, by a group from MIT, NASA, and the University of Maryland employing a transcontinental baseline. These observations were originally designed to measure the gravitational bending of the radio signals from 3C279 as it approached the sun on October 8, but they showed clearly that the compact source in 3C279 was complex and appeared to have at least two components separated by 0.00155 arc second, or about 20 light-years. This source was observed again in February 1971 by the same group and also by investigators from the NRAO, Caltech, and Cornell group, using the same baseline and techniques. The source appeared to double in only four months, but the separation was greater by two light-years. Hence the two components seem to be receding from a common point of origin with an apparent velocity about three times the speed of light.

Although the sensitive long-baseline interferometer systems were initially developed to study the compact extragalactic radio source, the technique has also been used to study the radio emission from interstellar clouds of hydroxyl radicals (OH) and water vapor. Typically the clouds are dispersed over a volume of several light-years across, although the individual components are as small as one astronomical unit (the average distance between the sun and the earth).

One limitation of tape-recording interferometry has been the lack of sufficient oscillator stability to determine the phase of the interference pattern; the improvement is to be expected from the use of hydrogen masers as frequency standards. Even with infinitely stable oscillators, there are still problems (Ref. 339) resulting from fluctuations in the path length through the atmosphere and ionosphere at the two widely separated observing sites.

---

[8] Managed by JPL for NASA.

[9] Managed by MIT.

Another limitation is the antenna flexure, but this type of limitation is often small and can be calibrated with a residual error of less than 1 mm.

IX. Limitations on Baseline Length

Extension of the baseline cannot always solve the resolution problem. The present evidence indicates that only for wavelengths less than about 10 cm can baselines much greater than the diameter of the earth be effectively used. Hence, hydroxyl emission regions, which radiate at 18 cm, and pulsars, which radiate most strongly at meter wavelengths, are not likely to be targets for a space interferometer.

There is an even more fundamental limit to the maximum baseline when dealing with the synchrotron sources. If the dimensions of the synchrotron system are below a critical size, the relativistic particles quickly lose all their energy by inverse compton scattering and so do not have time to radiate radio energy. This critical angular size is proportional to the wavelength at which the flux density is the highest. Since the resolution of a fixed length interferometer is inversely proportional to the wavelength of observation, the two effects cancel each other if the observations are made near the wavelength of maximum intensity (as they usually are to achieve the highest sensitivity). Thus, the maximum baseline needed to resolve synchrotron sources is nearly independent of the wavelength of observation; in fact, it is comparable to the diameter of the earth for the stronger radio galaxies and quasars. Baselines in space or on the moon will therefore probably not be necessary to study even the smallest radio nuclei and quasars.

Besides extending the baseline to achieve high resolution, intermediate baselines may be made very effective by linking together to form an ultra-high-resolution array (Refs. 74 and 171); but this can be difficult. Moreover, the task of collecting all the tape recordings at a common location and correlating all possible pairs of telescopes is formidable. An attractive alternative is to telemeter the data from each antenna to a common site by means of synchronously orbiting satellites, dispensing with the tape recording system completely. This technique was successfully demonstrated via the Harmes satellite as reported in Ref. 422.

X. Radio Interferometry for Earth Physics

The techniques of VLBI have also been applied to geophysical studies at JPL. The study of various faults, earthquakes, and other geophysical applications often requires a wide range of baseline lengths. A project that concentrates on this aspect of VLBI is referred to as the Astronomical Radio Interferometric Earth Survey (ARIES) Project (Ref. 218). The essence of the project is to use a pair of antennas, one fixed and the other portable, to provide the desired wide range of baseline lengths. The system characteristics, error sources, and other system details and requirements are given in Refs. 218, 272 and 273.

A mobile VLBI geodetic system[10] is currently under development at JPL for NASA application to earth crustal dynamics studies, as well as for technology transfer to the National Geodetic Survey (NGS), the U.S. Geological Survey (USGS), and other geodetic user communities. This system consists of a 4-meter high mobility station and a 13-meter transportable base station. The system characteristics are given in Table 1. The various error sources (both random and systematic) for the mobile VLBI geodetic system are tabulated in Table 2. Certain applications of such a geodetic system are:

(1) Environmental monitoring related to natural resource extraction such as subsidence resulted from geothermal field heat removal, oil drillings, and ground water pumping.

(2) Global tectonics and regional crustal deformations for uses in the development of a fundamental scientific understanding of geodynamic processes and earthquake precursory model development in the pursuit of a reliable productive capability.

Other possible VLBI applications to earth physics are global geodesy, tidal oscillations, crustal-block motions (including continental drift), polar motion, earth rotation, precession and nutation (including a test of general relativity), obliquity of ecliptic, shape of sea surface geopotential and global time and frequency synchronization. More detailed discussions on these applications are given in Ref. 339.

---

[10] Renzetti, N. A., "Radio Interferometric Geodetic System Applications," A JPL internal interoffice memo, dated June 1, 1978.

Table 1. Mobile VLBI geodetic system characteristics

Characteristic	Mobile	Transportable base
Antenna diameters	4 m	13 m
Antenna efficiency	50%	50%
Receiver temperature (X-band)	70 Kelvin	30 Kelvin
Bandwidth synthesized	400 MHz	400 MHz
Radio source strength	2 Jansky	2 Jansky
Digital data recording rate[a]	112 Mb/s	112 Mb/s
Observation time per source	720 sec	720 sec
Note: Signal-to-noise ratio	25:1	
Average delay precision	22 psec (7 mm)	
Baseline measurement precision limit	5 mm given 6 hr of data (single work shift)	

[a] 3-channel bandwidth synthesis

Table 2. Errors in mobile VLBI geodetic system

Errors	Distance equivalent, mm
Random Errors	
Delay precision	7
Time and frequency (H-masers, $\Delta f/f = 5 \times 10^{-15}$)	10
Troposphere	
Dry (surface meteorology)	10
Wet (water vapor radiometers)	10
Ionosphere (8.4 GHz, X-band)	5
Radio source catalog (0.003 arc sec)	7
Root sum square	21
Baseline vector precision given 12 hours of data	10
Systematic Error Sources (Baseline $\leqslant$ 500 km)	
UT1/polar motion (10-cm accuracy, supplied by external sources)	8
Radio source catalog (0.003 arc sec accuracy to be held essentially constant for monitoring missions)	4
Troposphere	
Dry (surface meteorology)	10
Wet (water vapor radiometry)	7
Ionosphere (X-band)	5
Antenna positioning relative to geodetic marks	5
Root sum square	17
Combined random and systematic error sources accuracy)	20

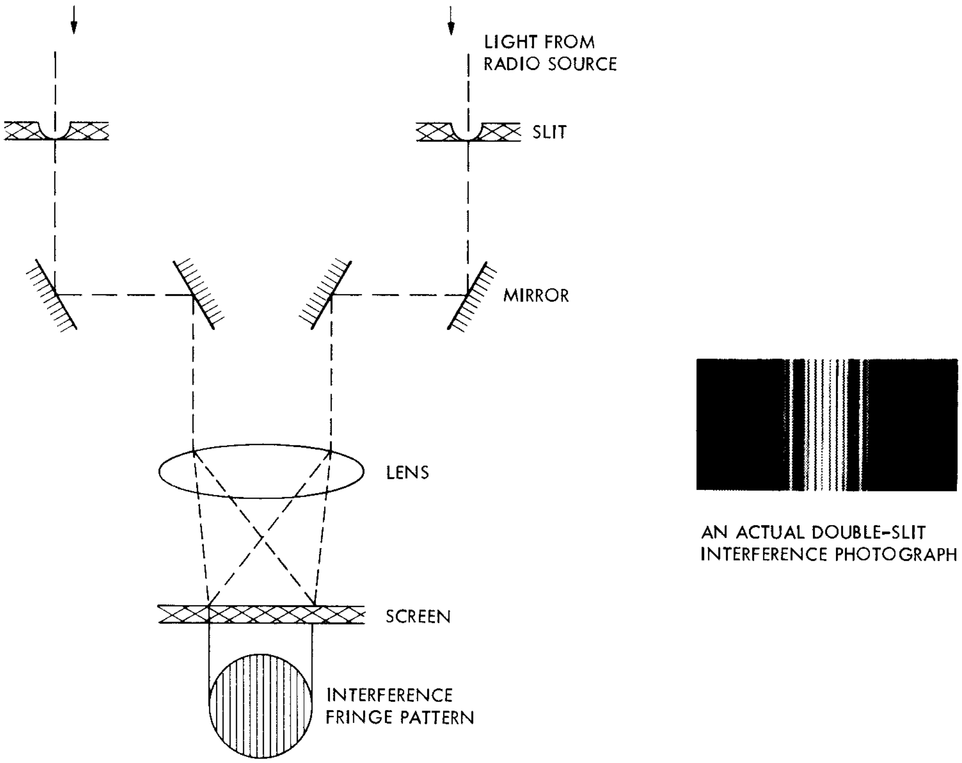

Fig. 1. Michelson–Pease Optical Interferometer. (Light from a distant radio source is reflected from the outer mirrors to the inner mirrors and then optically combined at the projection screen. Moving one of the mirrors makes one light path longer than the other. The two out-of-phase beams, when combined at the screen, will interfere with each other to provide "fringes." A photograph of actual double-slit interference fringes is shown on the right.)

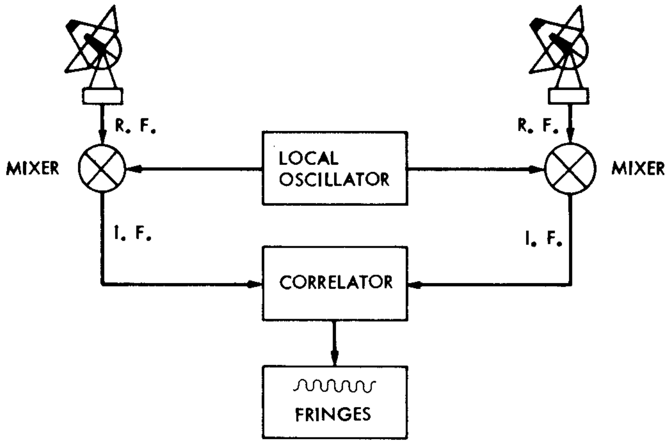

Fig. 2. Basic Radio Interferometer. (Two radio telescopes, separated by a few kilometers, are linked by cables. The high radio frequency (R.F.) signal from the radio source is mixed with a local oscillator signal to provide an intermediate frequency (I.F.) signal which is suitable for further data processing.)

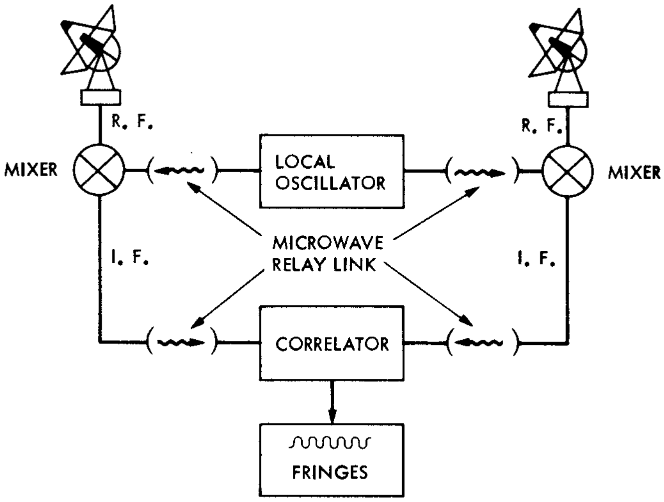

Fig. 3. General Radio-link Interferometer Concept. (The local-oscillator signal is transmitted to the mixers and the I.F. signals are returned via microwave-relay links. The local-oscillator signal at each end is derived from a common source so as to provide coherent (in phase) I.F. signals from each antenna.)