

A Review of the History of VLBI

B. G. CLARK

National Radio Astronomy Observatory, P.O. Box O, Socorro, NM
87801, USA

Abstract. I briefly review the history of VLBI. I start with a review of the lines of evidence that lead to the belief that very long baseline interferometry was a desirable and necessary development of instrumentation. I then briefly mention the first efforts to construct VLBI instruments. I finally look back over the past thirty years and call out several themes that have developed over the years.

Much of the early history of VLBI has been reviewed by Kellermann & Moran (2001) and by Kellermann & Cohen (1988). These accounts have left me with little substantive to add, but only a viewpoint. And even in viewpoint, I restrict myself to things that particularly interest me. First, I discuss the reasons why, in the mid 1960's, three different radio astronomy groups, in the Soviet Union, Canada, and the US, felt that the development of a long baseline interferometer was a natural and pressing thing to do. I make a very few remarks about the development of the VLBI equipment, adding little to what is discussed in the references above. And I finally mention the areas in which it seems to me that VLBI interferometry has made substantive contributions to knowledge, and a few highlights in each area.

1. Why we needed VLBI

In the mid 1960s there were several indications that there were features of radio sources of angular size too small to resolve using conventional, connected element, radio interferometers of convenient size. First, large, connected element interferometers continued to find unresolved sources at ever increasing baselines. Second, the variability of radio quasars was discovered, and if one believed the cosmological distances to the quasars, then the variation time scale times the speed of light divided by the distance gave a very small indicated angular size. At this time also, the first of the Gigahertz peaked spectrum objects were being studied, and if the spectral turnover was interpreted as being due to synchrotron self absorption, they again indicated a very small angular size. And finally, the phenomenon of interplanetary scintillation of radio sources had been discovered, and while this was sensitive to angular scales rather larger than those probed by the other phenomena mentioned, it did indicate that sub-arcsecond structure was common in radio sources.

Once the sizes and structures of a few radio source's had been determined by various groups, the energetic group at the University of Manchester developed the technique of radio linked interferometers, and pushed them to ever increasing baselines. The natural limit of the radio link technology of the time

was just over 100 km. They pushed their 158 MHz interferometer to this distance in the early 1960s. Allen et al. (1962) report measurements of many sources on this 70,000 wavelength baseline. Having reached the natural baseline length supported by their technology, they then proceeded to raise the operating frequency, reporting measurements at 408 MHz, which gave a baseline of 180,000 wavelengths (Anderson et al. 1965). Operation at still higher frequencies required a large antenna with a sensitive receiver, and the radio link interferometer was next taken to the Royal Radar Establishment's antenna at Malvern, giving a baseline of 500,000 wavelengths at 20 cm (Adgie et al. 1965). This baseline was taken to yet higher frequencies, giving baselines of one million and two million wavelengths at, respectively, 11 cm and 6 cm wavelengths (Palmer et al. 1967). This last development was not yet known by the time that tape recorder based interferometry was undertaken in the US and Canada, but the prevalence of fringes in the earlier experiments already indicated that to properly explore and characterize the small scale structure already seen would involve the use of interferometers of such baseline that the tape recorder technique would be the more practical approach.

Variability has been claimed since the earliest days of radio astronomy. The earliest claims were later proven to be the result of ionospheric scintillation. The first modern claim of variation was that of Sholomitskii (1965) of the source CTA 102. The claim was greeted with some skepticism. However, a definitive paper demonstrating the variability of 3C 273 was soon published by Dent et al. (1965). This paper calculated the angular size required by the measured timescale of the variation, and rather apologetically noted that the varying region must be as small as two milliarcseconds.

Yet another clue to small angular size was given by source spectra. It was early realized that a power law distribution of relativistic electrons would result in a power law emitted spectrum of synchrotron radiation, but with a low frequency turnover due to the source becoming optically thick to its own radiation. If the spectral turnover known in a few sources such as PKS 1934–63 (Kellermann 1965) were identified as this synchrotron self absorption effect, very small, milliarcsecond, angular sizes were indicated.

Objects of small angular size have an interesting observational property. If they are viewed through an irregular ionized screen, they will vary due to the relative motion of source and screen. This scintillation reaches a maximum at the radio frequency at which the phase variation through the scattering screen approaches a wavelength. It occurs for objects whose angular size is less than that of the first Fresnel zone. It was the low frequency ionospheric scintillation of Cygnus A that first indicated that it was an object of relatively small angular size, and thus began the study of discrete radio sources. The equivalent phenomenon of scintillation caused by the ionized interplanetary medium was first noted by Hewish et al. (1964). These authors noted that one can, in theory, derive actual brightness distributions from the scintillation properties of the radio source. The technique at convenient frequencies is sensitive to source sizes of about one second of arc. Hewish et al. suggested that it could be extended to look for smaller structure by observing at lower frequencies. (In retrospect, this would not have worked because synchrotron self absorption causes regions of small angular size to have very low flux densities at low frequencies.)

Cohen (1965), following up the work of Hewish et al., reported that scintillation in the interplanetary medium is a common phenomenon, and that therefore sub-arcsecond structure is common. There would be no shortage of sources for VLBI investigation.

Scintillation in the interstellar medium has become the most recent emergence of the phenomenon, and, as in the case of the interplanetary scintillation, is used for an indication of the angular size of objects smaller than the resolution of contemporary interferometers. Interstellar scintillation occurs for objects smaller than a few microarcseconds. Imaging with scintillations is more difficult, although Gwinn et al. (2000) have deduced the structure of the emitting region of a pulsar through analysis of scintillation.

2. Building the first tape recorder interferometers

The tape recorder interferometer appears to have first been discussed in the Lebedev Institute in Moscow. Despite attracting great interest in the radio astronomy group there, the idea was not taken up by the astronomical establishment in the Soviet Union, and the eventual publications were in journals which were not being translated, and the ideas did not readily pass to the West. The subject was broached with Sir Alfred Charles Bernard Lovell on a visit in 1964 (see discussion in Kellermann & Moran, 2001), and an agreement that was reached that the Lebedev group would work with the Manchester group to develop such an instrument. Work proceeded very slowly, however, because modern recording technology was not readily available to the Lebedev group, and because the Manchester group was working hard on the series of connected interferometers mentioned above. As resources became available, the Manchester group did complete a tape recorder interferometer back end, but not before the NRAO and Canadian efforts came to fruition, and the system was never deployed in practice.

Interest in long baseline interferometry arose in another quarter entirely at this time. The planet Jupiter emits short bursts at decameter wavelengths, and the question arose as to whether the bursts were a planet-wide event, or came from a small region on the planet. To settle the question required resolution of about one arcsecond at ten meter wavelength, a baseline of a few hundred kilometers. The Jupiter bursts are very intense, though of short duration and low duty cycle. The post-detection correlation technique (Brown & Twiss 1954) was well known at that time, and is well adapted for intense bursts. The bursts are so intense, and the required bandwidths so narrow, that a post-detection correlation interferometer could be realized using ordinary audio tape recorders (Carr et al. 1965). This group quickly realized that at this low frequency, a local oscillator could easily be made that would keep in phase between the two elements, and they went on to make observations with Michelson interferometry, though they did not publish the results until after the NRAO and Canadian interferometers were working (Brown et al. 1968).

The science drivers discussed above attracted the attention of groups at the Hertzberg Institute in Canada and at the NRAO in the US. The state of the art in stable oscillator construction was such that coherence could be attained over most of the radio astronomy bands. Readily affordable oscillators based on

the 6 GHz transition in rubidium vapor were seen to have sufficient stability to preserve coherence up to 1 GHz or more, and the expensive and finicky hydrogen maser could preserve coherence for many seconds at frequencies of 30 GHz. The time was ripe for the development of the tape recording interferometer.

The Canadian group began development of a system for use at 408 MHz, based on studio-type television tape recorders, implementing the delay tracking and fringe rotation functions in mostly analog form. After a couple of false starts, they achieved trans-Canada fringes at 408 MHz in 1967, reported by Broten et al. (1967).

I had similar ideas about how to implement the tape-recorder interferometer, and spent a day wandering around an electronics show in 1965, learning as much as I could about the state of the art in TV and instrumentation recorders. (The latter interested me more than the former, because of the TV recorder's built in dependence on the TV vertical and horizontal sync signals, which the Canadians actually very nicely exploited as an integral part of their system.) However, Sander Weinreb, newly the head of NRAO electronics, quickly informed us that digital was the way to do things, and that as much as possible we should exploit mass produced off-the-shelf digital technology for the recording system. It was quickly decided that the system would be based on a standard computer tape transport, with a correlator implemented in software. It was my job to provide the latter. Initial interferometer operation was to be demonstrated at 610 MHz, and then we were to see how high in frequency we could go. After a couple of false starts, the system produced fringes in 1967, first on the short baseline between the 140 foot and 85 foot radio telescopes in Green Bank. The next step in the quickly extending series of baselines was to observe with the interferometer formed by the Green Bank 140 foot telescope and Naval Research Laboratory 85 foot at Maryland Point. Then, at the urging of the Haystack group, who were interested in the interstellar maser OH line, we observed between the Green Bank 140 foot and the Haystack Observatory 120 foot telescope in Massachusetts (Clark et al. 1968a). We finally finished the sequence of rapidly increasing baselines with an observation between Green Bank and the 85 foot telescope at Hat Creek in California (Clark et al. 1968b), achieving transcontinental fringes at 18 cm wavelength.

The MIT/Haystack group, being interested in interferometric spectroscopy, independently developed a correlator for spectroscopy, using the NRAO hardware.

Soon after the first demonstration of the NRAO recording system it was clear that there was enough high resolution science to justify a more capable and easier to use system. The narrow bandwidth (360 kHz, limited by the 720 kbits⁻¹ that the computer tape transport could write) and the short running time of the tape (about 3 minutes) were clearly severe disadvantages. At the time, the first affordable helical scan TV tape recorders were available, and we decided to use them as a basis for a new generation of tape recorder interferometer. They could record a 2 MHz bandwidth (recorded digitally as 4 Mbit s⁻¹) on reels of tape that lasted as long as three hours. These early helical scan recorders were difficult to maintain and align. Making sure that the recordings at the telescope could be played back at the correlator was always a problem. These early recorders were replaced by later generation devices, eventually with

home video cassette recorders (VCRs), which were much more reliable and much less touchy. But the basic design, dubbed the Mark II interferometer, remained the workhorse of radio astronomy VLBI for many years.

The initial Mark II correlator at NRAO correlated a single baseline. We realized, of course, that multi-station observations could be conducted to advantage, but reasoned that the number of baselines would tend to be about the reciprocal of the percentage of time that large radio telescopes would devote to long baseline interferometry, so that a single baseline correlator running full time could keep up with the observations. The single baseline quickly proved inadequate in practice, especially a few years later, when advances in image reconstruction techniques demonstrated the importance of a large number of simultaneous baselines. The NRAO Mark II correlator was expanded to three baselines, and a large, 16 station Mark II correlator was built at Caltech.

The possibilities of using VLBI for astrometry and geophysics were realized quite early on (Gold 1967). After early experience with the NRAO Mark I system, the geodesists realized that phase connection was a very difficult thing to do indeed, and that a much easier route to precision measurement was through measuring delay. For that, one needs a wide spanned bandwidth, which is difficult to attain with the Mark I system. The group at Haystack observatory began development of the Mark III system specifically for geodesy, with fourteen frequency channels digitized and written to the fourteen independent heads of a high speed instrumentation-type tape transport. The fourteen channels could be positioned to cover a very wide spanned bandwidth. This system quickly became the standard for geodesy. For astronomical purposes, however, the figure of merit tends to be the number of bits recorded on a dollar's worth of medium. This number was much the same for Mark II and Mark III, so the easier-to-implement Mark II remained in use until the design and retrofit of Mark IIIa, which uses narrow heads to lay down parallel written tracks, so that up to fourteen times as much data could be written to a single tape as Mark III. The Mark IIIa and the very similar VLBA and Mark IV formats quickly became the standards.

Mark II was extended by the Canadians to have comparable bandwidths to the Mark III/Mark IV/VLBA recorders, by operating several VCRs in parallel, each recording at 8 Mbit s^{-1} . This system is called the S2, with a further development, using studio quality VCRs, called S3.

The Haystack group meanwhile has noted that the cost to store a bit on computer disk storage is dropping much more rapidly than the cost to store a bit on tape. Within a couple of years, off-the-shelf computer disks will become the most cost effective media, even with the cost of motor, spindle, actuator included in the disk unit. There are also substantial operational advantages to disk units: easier maintenance of the mechanical parts, effectively instantaneous start and stop, essentially error-free operation. Haystack is now designing a system to use disks, to be called Mark V. A preliminary version of the system has been built and has been used for astronomical observations.

3. Science Themes of VLBI

As discussed in Section 1, VLBI as we know it was developed for studies of the quasars and bright radio galaxies — what are they, how do they work? But many uses not conceived at the birth of the technique have found the technique to be of importance. With the increase in sensitivity that came with the broader recorded bandwidths of the Mark III and successor systems, not only the bright quasars can be studied, but also the low luminosity active galactic nuclei. The early work on astronomical masers has been extended to many interesting results, in studies of OH, water, and SiO masers. And the early work on astrometry and geodesy continues. The accuracy of VLBI baseline determinations remains somewhat better than any other long-distance technique, and only VLBI gives the absolute orientation of the Earth in an inertial coordinate system to high accuracy.

Interestingly, although study of quasars and other AGN was the primary motivation for development of the modern VLBI systems, it has perhaps not provided the number of interesting results that some other themes have done. The early years did yield exciting results - that the quasars really were milliarc-second objects, that they are elongated, not round, and that the elongation is due to the core-jet nature of the emission. Of course, perhaps the most exciting early result was the discovery of superluminal motion by Whitney et al. (1971), quickly confirmed by other groups (for instance, Cohen et al. 1971). Since that time there has been a great deal of phenomenology, but little increase in understanding. Decades of effort have gone into the question of whether jet motion is ballistic, or follows an established path (answer: sometimes), but this has not yet resulted in a firm understanding of the processes at work. The most interesting modern work on AGN has been outside the main thrust of work on the brightest objects. This includes the discovery of the compact symmetric objects (Readhead et al. 1996), detection of the inner collimation region in M 87 (Biretta et al. 2002), and the quotidian use of VLBI as a means of separating active galaxies into starburst, AGN, and starburst+AGN systems. And our own AGN, Sgr A*, offers interesting possibilities, as we struggle to measure its intrinsic size, which will quite likely contribute some interesting physical insight if we manage to resolve it well (Falcke et al. 2000).

The astronomical masers have been interesting for both the intrinsic small sizes of the maser spots (Moran et al. 1968; Burke et al. 1971; and other authors), and for the applications of masers for inferring distances and motions. Maser proper motions have been used to give dynamic parallaxes for objects in our galaxy (a technique pioneered by Genzel et al. 1981). Shock excited masers have been used to outline the shock fronts in outflows from proto-stellar objects, and to measure their motions (for instance, Wootten et al. 1999). Perhaps one of the most interesting astronomical results of the whole decade of the 1990's was the detection and characterization of the Keplerian molecular disk surrounding the massive black hole at the nucleus of NGC 4258, outlined in a series of papers starting with that by Miyoshi et al. (1995). The SiO masers in the extended atmospheres of the AGB stars provide a picture of motion of the atmospheric material which already demonstrates that the outer atmospheres of these stars have more complex dynamics than was conceived prior to the measurement (Diamond et al. 1998).

Despite their low flux densities, active stars have proved a fertile area of research. The flaring regions in their outer atmospheres can be readily resolved, and comparisons to solar activity (which is thousands of times weaker) are leading to understanding of how these active stars radiate in the radio regime. And the loop back to quasars is closed by the discovery of the micro-quasars, stellar mass objects that demonstrate the same phenomena as quasars, on a vastly smaller spatial scale and vastly shorter time scale.

The precision measurements for astrometry and geodesy have been vigorously pursued by the Haystack, NASA-Goddard, and Naval Observatory groups. An early successful use of VLBI astrometry was the accurate determination of the relativistic displacement of radio source positions due to the gravitational influence of the Sun (Counselman et al. 1973). VLBI quickly grew to be one of the most accurate techniques in both astrometry and geophysics (for instance, Ma et al. 1986; 1998). The accurate determination of the Earth's orientation, and the length of day, have led to truly remarkable results in geophysics (for instance Marcus et al. 1998). The progress of radio measurements of position has been one of the great success stories of precision measurement. In 1960 there were few radio positions determined with sufficient accuracy to permit identification with even bright or unusual objects known optically. Forty years later, VLBI measurements of positions of extragalactic objects establish the celestial reference frame, in inertial coordinates, leaving optical techniques struggling to provide sufficient accuracy to register the images made at radio and optical wavelengths. In that forty years, the accuracy of common radio positions has improved almost a million fold.

References

- Adgie, R.L., et al., 1965, *Nature*, 208, 275
Allen, et al., 1962, *MNRAS*, 124, 477
Anderson, B. et al., 1965, *Nature*, 205, 375
Biretta, J. A., Junor, W., & Livio, M., 2002, *New Astron. Reviews*, 46, 239
Brotten, N. W., et al. 1967, *Nature*, 215, 38
Brown, G. W., Carr, T. D., & Block, W. F. 1968, *Astrophysical Letters*, 1, 89
Brown, R. H., & Twiss, R. Q. 1954, *Philosophical Magazine*, 45, 663
Burke, B. F., et al. 1970, *ApJ*, 160, L63
Carr, T. D., et al., 1965, *IEEE NEREM Record*, 7, 222
Clark, B. G., et al. 1968a, *ApJ*, 153, 705
Clark, B. G., et al. 1968b, *ApJ*, 153, L67
Cohen, M.H., 1965, *Nature*, 208, 277
Cohen, M. H., et al., 1971, *ApJ*, 170, 207
Counselman, C. C., et al. 1974, *Phys. Rev. Lett*, 33, 1621
Dent, W. A. 1965, *Science* 148, 1458
Diamond, P. J., & Kembal, A. J., 1998, *BAAS*, 30, 1349
Falcke, H., Melia, F., & Agol, E., 2000, *ApJ*, 528, L13
Genzel, R., et al., 1981, *ApJ*, 244, 884
Gold, T., 1967, *Science*, 157, 302
Gwinn, C. R., et al. 2000, *ApJ*, 531, 902
Hewish, A., Scott, P. F., & Wills, D., 1964, *Nature*, 203, 1214

- Kellermann, K. I., 1965, *AJ*, 70, 681
Kellermann, K. I., & Cohen, M. H. 1988, *JRASC*, 82, 248
Kellerman, K. I., & Moran, J. M., 2001, *ARA&A*, 39, 457
Ma, C., et al. 1986, *AJ*, 92, 1020
Ma, C., et al. 1998, *AJ*, 116, 516
Marcus, S. L., et al. 1998, *Science*, 281, 1656
Miyoshi, M., et al. 1995, *Nature*, 373, 127
Moran, J. M., et al. 1967, *ApJ*, 148, L69
Moran, J. M., et al. 1968, *ApJ*, 152, L97
Palmer, H. P., et al. 1967, *Nature*, 213, 789
Readhead, A. C. S., et al. 1996, *ApJ*, 460, 612
Sholomitskii, G. B., 1965, *AZh*, 42, 673 (translation in *Soviet Ast.*, 9, 516)
Whitney, A. R., et al. 1971, *Science*, 173, 225
Wootten, A., et al. 1999, *BAAS*, 31, 898