Bulletin de la Société Royale des Sciences de Liège, Vol. 74, 5-6, 20005

A high-level technical overview of prototype and first-generation optical arrays

D. F. Buscher

Cavendish Laboratory
J J Thomson Avenue
Cambridge CB3 0HE, UK

Abstract: More than a dozen optical/IR interferometers have been in operation in the period 1985-2000. The lessons that can be learned from these arrays to guide the building of future interferometers are discussed, including their design, the practical problems encountered in their construction and operation, and the strategic and fundamental technical issues relating to the overall scientific productivity of an array. The interferometers discussed in detail are mainly those which the author has had personal experience of designing, building, and/or operating.

1 Introduction

In a conference on next-generation optical/IR arrays, it is important to step back and reflect on the legacy from previous generations. A large part of this legacy is scientific, in the form of the astrophysical results which guide future directions of exploration, but another legacy is the technical one, which is often less well documented but is perhaps more valuable in the long run to the progress of the discipline. Such technical "lessons learned" include hard-won experience of what has worked and what has not, what was more difficult or more easy than first imagined, what was forgotten but turned out to be important and what was thought to be important but in fact didn't matter. Here I attempt to document briefly some of these lessons. These mostly come from a very limited point of view, namely my own, but also incorporate the input from others who have worked on the current generation of arrays, who were kind enough to discuss their "lessons learned" with me.

2 Context

It is hard to separate interferometers into "prototype" versus "first generation" arrays. For the majority of this paper I will restrict the discussion coverage to the arrays which were operational in the period from 1985 to 2000, and for simplicity I shall group all of these under the heading "first generation interferometers". I will also briefly mention the current "facility-class" arrays which came into operation in the period since 2000, namely Keck Interferometer and VLTI, which should be considered to be a generation on their own.

The history of first-generation arrays worldwide is treated in some detail by Monnier (2003). Table 3 of Monnier's paper serves to summarise the broad characteristics of this generation (the table also includes the "facility-class" interferometers, and does not include interferometers which were operational in 1985-2000 but no longer operational in 2003: the I2T, the Mark III, the Sydney University 11.4m, IRMA and FLUOR/McMath). A number of "population" aspects of the first-generation arrays can be noted:

- The sheer number of arrays which have been operated in the period 1985-2000 is very large: 14 interferometers over a 15-year period. This testifies to a level of intense and sustained interest in optical interferometry within the optics and astrophysics communities and also hints at the fact that interferometry cannot be as difficult as outsiders sometimes think, since all of these interferometers have produced both fringes and science.

- The interferometers operational during this period which have closed are outnumbered approximately 2:1 by those which were still in use in 2003. This indicates that the useful lifetime of even a "prototype" interferometer is relatively long. A cynic may perhaps say that this statistic demonstrates that it takes quite a long time for interferometers to enter their scientifically useful period!

- The majority of these interferometers have individual apertures of $\leq$ 50cm, an aperture size which is significantly smaller than those of the majority of astronomical telescopes in general use, indicating that the scientific utility of an interferometer is a much weaker function of aperture size as compared with conventional telescopes.

A detailed commentary of the lessons that could be learned from this entire generation of interferometers is beyond the scope of this communication. In the next section, I shall instead concentrate on the subset of interferometers which I have had direct experience with, whether in the design, construction, operation, or some combination of these tasks. In the following section I outline some lessons which can be inferred taking this set of interferometers as a whole, and in addition incorporating some of the comments from other interferometrists concerning interferometers I have not been privileged to work on.

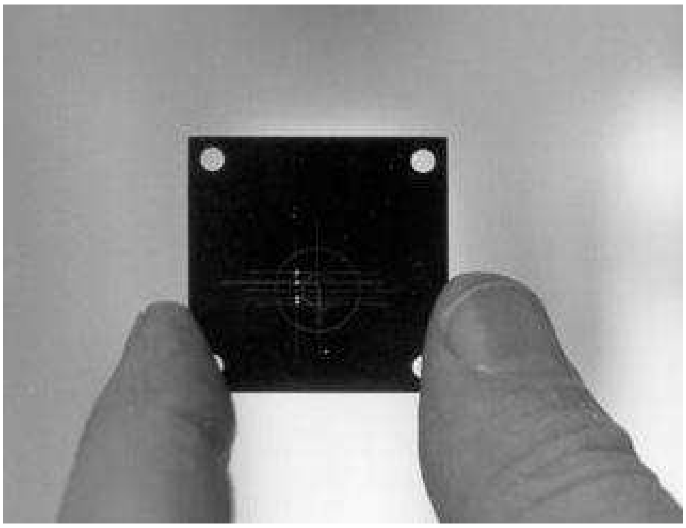
3 Lessons learned — individual interferometers

The following subsections outline the interferometric systems I have worked on, ordered chronologically by the periods on which I have experience of them rather than by their date of first operation.

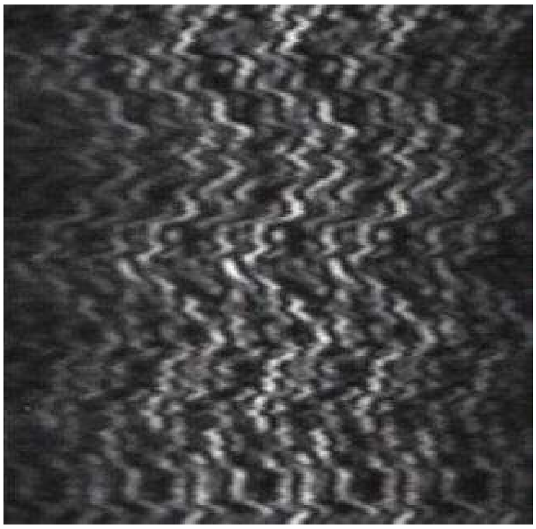
3.1 1987-1991: Aperture Masking

Aperture masking is not, strictly speaking, a separate-element interferometric technology. Nonetheless, it shares many features with separate-element arrays, and to date is has been the only way to gain practical experience of the interferometric use of more that 4 apertures at a time.

I joined the aperture masking programme which had earlier produced the first closure phases (Baldwin *et al.* 1986) and aperture-synthetic images (Haniff *et al.* 1987) at optical wavelengths. The essence of the technique is to convert a conventional monolithic telescope aperture — first the 2.5m Isaac Newton Telescope (INT) and later the 4.2m William Herschel Telescope (WHT) — into an interferometric array by placing an opaque mask containing an array of r_0-sized



(a)



(b)

Figure 1: Aperture masking experiments: (a) Example 4-hole aperture mask (b) Sequence of 1-d interferograms. The time axis runs in the vertical direction, with the spatial coordinate running in the horizontal direction. Multiple different fringe frequencies with different time evolution of different fringes can be discerned.

holes in a reimaged pupil plane, as shown in figure 1(a). The non-redundant arrangement of holes meant that the Fizeau fringe pattern formed in the image plane of the telescope contained the fringes formed from the interference of different hole pairs appeared at a different spatial frequencies, so that all baselines could be sampled simultaneously. Arrays with up to 6 holes were used on the WHT, giving up to 15 simultaneous baselines: later, this technique would be used on the Keck telescope with up to 21 apertures. On the INT, the detector used was a photon-counting imaging camera, the IPCS, while on the WHT a conventional CCD read out in a fast 1-dimensional mode was used, giving the two-dimensional time sequences shown in figure 1(b). Images of a range of sub-arcsecond binaries and of the surface of Betelgeuse (Buscher *et al.* 1990) were made from these data.

A number of lessons could be derived from these experiments. The first and most important was that there is no fundamental difficulty in making aperture-synthesis images at optical wavelengths. Scientifically useful images were routinely and reliably obtained using simple experimental apparatus (albeit with state-of-the-art detectors). A number of potential risks which had been identified turned out not to be problematic in practice. These include:

1. Problems with measuring closure phases at optical wavelengths. By using the phase of the mean bispectrum as an unbiased measure of the closure phase, problems with the noisy nature of optical closure phases could be overcome by the simple expedient of averaging over the plentiful exposures that were available as a result of the high frame rate. Also, by measuring all fringes simultaneously in the same fringe pattern, systematic errors in the closure phase were eliminated: in almost all cases when the systematic errors were measured, it was found that no calibration was necessary, even when the closure phase was measured to precisions of fractions of a degree. With this level of accuracy one could in principle measure the relative positions of two sources with a sensitivity measured in microarcseconds, something which would be exceedingly difficult to do using conventional phase measurements.

2. Problems in making interferometric measurements with only a few photons. The IPCS

detector was routinely used to make visibility and closure phase measurements, and subsequently images, when only a few photons per baseline per exposure were detected. Providing the correct photon bias corrections were made, the data reduction process was reliable and straightforward.

3. Problems in reconstructing images from optical data. In practice, standard radio-astronomical image reconstruction packages were used on optical data with little problem. Providing the standard rules-of-thumb concerning the number and distribution of u-v points measured were obeyed, images with a quality as predicted by the rules of thumb were obtained. Most of the fundamental problems of imaging with dilute apertures had been solved in the preceding years by the radio astronomical imaging community.

Another practical lesson was that detector artifacts were the limiting factor in the observations. A large part of both the data taking and the subsequent data analysis for the IPCS data was concerned with mitigating the problems that the image intensifier had with high photon rates and with non-linearities due to coincidence losses. The use of a CCD was far less problematic and far more scientifically productive, despite its greater read noise, because of its highly linear response.

A final observational lesson was that closure phases are more useful than visibility amplitudes in even only moderately complex situations. It is well known that the Fourier phase contains more information about complex images than the Fourier amplitudes. In the case of imaging Betelgeuse, it became clear that even in an image with only a low level of complexity, most of the properties of the distribution of flux in the image could be deduced "by eye" from the closure phase information, with only the gross features, e.g. diameter and asymmetry being deduced from the amplitudes.

3.2 1991–1992: The Mark III Interferometer

The Mark III interferometer on Mt Wilson (Shao *et al.* 1988) was originally built with the aim of doing global astrometry with a phase-tracking interferometer. It was subsequently converted to dual use of astrometry and astrophysical visibility measurements, increasing the complement of 3 fixed siderostats on ~11m baselines with a further 2 movable siderostats giving baselines up to 30m. The major innovations of the Mark III included (a) real-time white-light fringe tracking and (b) entirely automated night-time operation, requiring only one observer/operator who only intervened infrequently in the system's operation. Both of these features were essential to the fulfillment of the astrometric mission of the Mark III, but turned out to be almost as important for its success at visibility measurements. The combination of these two features meant that the interferometer could routinely make 150 "scans" a night, where a scan consisted of slewing to and acquiring a new target star, engaging the tip-tilt servo, finding the white-light fringe and initiating tracking, and accumulating 60-100 seconds of data. Remarkably, this feat was accomplished by a total of 2 Data General "Nova" computers, whose operating system booted with a copyright message dated 1968.

The impressive data rate contributed to the even more impressive scientific productivity of the Mark III (see Monnier 2003 and Quirrenbach 2001 for a list of references) in one obvious way and one less obvious one. The ability to scan many different targets per night meant that surveys of large numbers of stars of a particular class, whether binary stars or late-type stars, could be made and significant signal-to-noise and phase coverage built up through repeated observation. This meant that statistically significant and high-SNR astrophysical results could be obtained and rapidly-changing objects such as close binary stars could be measured in a timescale short comparable to their evolution timescale.

The more subtle advantage arising from the rapid acquisition of interferometric data was the hugely improved ability to characterise the interferometer itself, and thereby both debug the system and calibrate the visibility data to a high precision. The visibility degradation introduced by the seeing and the interferometer (the "system visibility") is subject to a host of varying and complex instrumental and atmospheric effects. The only reliable way to compensate for variations in the system visibility is to measure its value and how it changes. A system which is able to make hundreds of measurements a night allows for the variations in system visibility to be monitored and correlated with various system parameters such as time, zenith angle, seeing, stellar colour and so on, so that a multi-dimensional "calibration curve" can be built up. This technique was developed to a high degree of sophistication on the Mark III, and is also used with success on the intellectual successors to the Mark III, namely PTI and NPOI. On other interferometers where the number of stars observed per night could be counted on the fingers of one hand, use of this technique has been severely limited, and cruder calibration techniques, e.g. simply dividing the target visibility by the calibrator visibility measured closest in time to the target.

3.3 1992–1994: NPOI

The NRL/USNO Optical Interferometer project was responsible both for the running of the Mark III interferometer and also designing and building a successor interferometer, the Navy Prototype Optical Interferometer. Like the Mark III, this interferometer has an astrometric mission, but in addition an astrophysical mission, to include imaging with closure phase, was also included from the start. The project was considerably more ambitious in scope, involving an initial plan for two separate interferometers, which evolved to a single array. In addition to a larger number of larger siderostats and the first 6-way beam combiner for imaging, the array incorporates laser metrology which links together the pivot points of astrometric siderostats at the sub-micron level.

Involvement in a large scale project at an early stage allowed me to see much of the process by which such projects evolve and the technical compromises that must inevitably be made in the light of logistical and financial constraints. Perhaps the most profound lesson I learned there was that the main challenge of interferometry is not in any particular technical detail, but rather in managing the sheer complexity of the system. This complexity is less a function of the number of components in the system (although this can be substantial: Theo ten Brummelaar has pointed out before that the rails and support system for the delay lines for the CHARA array incorporate tens of thousands of components once one counts all the nuts and bolts) than of their variety. Thus building and installing six delay lines is more work but not significantly more complex than building and installing three delay lines. This can be contrasted with building and installing siderostats (of two kinds), enclosures, metrology systems, tip/tilt systems, vacuum systems, delay lines, beam combiners, detectors, control systems, temperature-controlled buildings, and so on, and then making them all work together, which is a problem of a different character all together. Building any one of these subsystems is a reasonable-sized project in itself; getting all of them designed, built, installed and debugged is a task which requires a particularly dedicated team.

A corollary of this is that simplicity in the design of an interferometer is a cardinal virtue: this is not just because reducing the number of surfaces in an interferometer reduces the degradation of wavefront quality and increases throughput, but also there is a significant danger that an interferometer can become just too complex to allow adequate quality control of the process of building and testing of the array.

3.4 1999–present: COAST

I was involved in the early stages of the design of COAST as a PhD student and postdoc in 1985-1990, but I had left for the U.S. by the time of its completion. Given its location (it is the lowest-altitude optical interferometer in the world, at 17m elevation), it was never designed to be a scientific workhorse but rather to demonstrating the basic principles of phase closure aperture synthesis. By the time I returned to Cambridge in 1999, COAST had succeeded in its basic aims by measuring the first optical closure phase from a separate-element array in 1993 and the first optical aperture synthesis image from an array in 1995.

What was slower in coming was a stream of images of astrophysically-interesting sources. There were a number of reasons for this, not least of which were the problems with the site alluded to before: the number of clear nights with reasonable (<1.5 arcsec) seeing was about 30 per year. A second problem was the lack of automation which meant that even the nights that were available were used at low efficiency. One problem that is obvious once one has experienced it is that slewing siderostats and acquiring stars on a 5-element array takes more than twice as long as doing so on a 2-element array if the system is human-operated, whereas if it had been automated, these procedures could have taken place in parallel on all siderostats simultaneously. Improvements were made to the automation over time, but it is difficult to build this into the array after the fact.

A much more fundamental problem however was something that is intrinsic to the science aims of imaging interferometry. By definition, an object can only really be said to have been "imaged" if it has been observed with baselines long enough to allow many resolution elements ("pixels") across the object's diameter. A consequence of this is that the intrinsic visibility of the source on the longest baselines is of necessity low, and since the visibility enters quadratically into the signal-to-noise ratio (SNR), the SNR is very low on the very baselines which have the high-resolution imaging information. This lack of SNR can be overcome by incoherent integration over a sufficiently long period, but in order to measure useful data to integrate, the position of the fringe coherence envelope must be found and tracked. If the instantaneous fringe SNR is too low, the fringes cannot be integrated for long enough to find the fringes before the fringe envelope has moved. Thus there is a hard lower threshold on instantaneous SNR below which it is impossible to observe fringes with any reliability, and it is easy to fall below this threshold on long baselines.

There are a number of solutions to this problem of finding and tracking fringes on baselines where the source is resolved. Probably the most robust of these is so-called "baseline bootstrapping", where a long baseline is made up of a "chain" of shorter baselines between nearest-neighbour telescopes. Fringes can be acquired and tracked on the short baselines and the fringes then integrated on the longest baselines to provide the necessary SNR. This procedure has been used many times on COAST, but what became quickly apparent was that baseline bootstrapping depends on having enough telescopes to form the "chain"; on COAST there are not enough telescopes to do a good job of bootstrapping, despite it having more telescopes than almost any existing array. Because of the 2-d nature of the COAST array layout, the longest chain that can be usefully formed is a 3-telescope chain, i.e. the longest baseline is twice the fringe-tracking baseline. This is sufficient to make rudimentary images with roughly 2×2 resolution elements. COAST has baselines which could in principle give resolutions many times greater on many of the science targets of interest, but what in fact limits the resolution is not the *baseline* but the *number of telescopes*. This is true on even the brightest sources, because of the quadratic nature of the dependence of SNR on visibility: a source would have to be 5 magnitudes brighter to compensate going from 100% to 10% in the source visibility on a given baseline, yet it is precisely those baselines which have 10% visibility which contain the useful imaging information.

4 Global lessons learned

A number of technical lessons can be elucidated by looking at patterns that emerge when looking globally at the experience gained from the interferometers in the above limited sample and first generation interferometers in general. In addition, as part of the preparation for this paper, I emailed a number of colleagues who have been involved in the design and construction of a number of first-generation arrays worldwide. The following brief points have resulted from the combination of their responses and my own observations.

Calibration The calibration of the system visibility of an interferometer is perhaps the most intractable problem facing observers trying to make a useful astrophysical inference from their data. Almost everything which can affect the system visibility will, as proven by the extensive investigations on the Mark III. Spatial filtering does help to reduce the effects of one degree of freedom, namely anything which affects the spatial wavefront quality, on the system visibility, but does not affect the temporal or bandwidth-dependent effects for example. The best solution is to have a high duty-cycle automated system which allows large amounts of visibility calibration data to be gathered.

On a related note, it is important to allow significant amounts of time between getting first fringes in an interferometer and routine science operations, in order to properly characterise the system visibility and the sources of its fluctuation.

Automation System automation is much more important to the success of an interferometer than in most other forms of astronomical observation. This point was raised by almost all my correspondents independently, citing experience from the GI2T, the Mark III, PTI, and Keck Interferometer. The reasons cited included the visibility calibration arguments cited above, science throughput, and real-time feedback of system performance.

Surprisingly, the automation performance of interferometers has in general been getting worse rather than better over the years, with the peak scan rate of 150 scans/night being attained by the Mark III in the early nineties using computer technology several thousand slower than modern-day machines.

Astrometry There is no doubt that astrometry in general is difficult, and that astrometry with interferometers perhaps more so. However, it is sometimes perceived that astrometry is a more direct route to interferometric science than imaging, because to do imaging well requires an array with a large number of telescopes. Experience tells a different story. To date, five different interferometers have been built with astrometric science as a prime science target, namely the Mark III, PTI, NPOI, KeckI and VLTI. Some of these interferometers are still in their early stages of development, but it is of interest to note that all of them have been much more successful in getting good "visibility" science results (a sibling of imaging science) than in doing astrometric science. The reasons for this can be debated: my explanation is twofold, namely (a) the atmosphere is a far worse enemy to astrometry than to visibility science and (b) somewhere within an astrometric

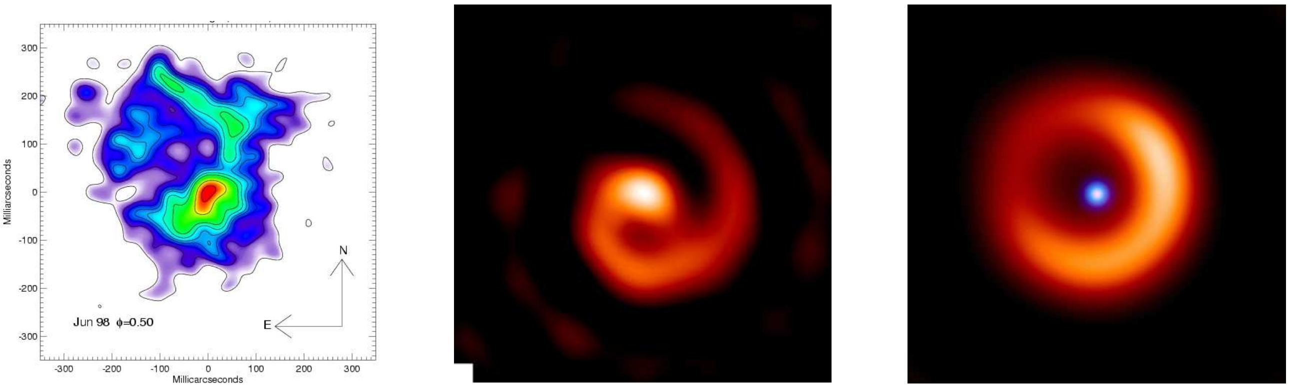


Figure 2: A sample of interferometric images produced by Keck aperture masking. Left to right: the extreme mass-loss AGB star IRC+10216 (Tuthill *et al.* 2000), the Wolf-Rayet interacting winds system WR104 (Tuthill *et al.* 1999) and the young star LkHα 101 (Tuthill *et al.* 2001)

system there is a mechanical subsystem which is not directly monitored by metrology and which must have nanometre-level mechanical stability over a time period of minutes to hours; imaging relies on a similar degree of stability but over timescales of a few tens of milliseconds at most. Nanometre-level mechanical stability is feasible on short timescales but the difficulty of achieving this get rapidly worse on longer timescales as thermal and creep effects become important. Whatever the reasons, the evidence that astrometry is a fundamentally more difficult route to science than imaging is hard to ignore.

Imaging The problems with getting "imaging" science have been generally less than people first imagined, except where people have underestimated the number of apertures required to make good images. A reason for underestimating the number of apertures is the bootstrapping problem explained earlier. Conversely, the imaging science with the largest number of apertures has been amongst the most spectacular as can be seen in figure 2. That some of these results simply look spectacular should not be dismissed as irrelevant: these are far more likely to capture the imagination of funding authorities than a visibility curve or a spectrum, and so producing images from interferometers are perhaps a key factor in the long-term health of interferometry research.

Throughput It is hard to overemphasise the importance of throughput on the performance of an interferometer, but it is often a neglected detail, precisely because it is a topic dependent on attention to detail. Very few interferometers publish their throughput figures, and most people who build interferometers will admit that the throughput of their system is less than expected. Simple systems with few surfaces will always beat more complex systems.

Detectors As with most instrument technologies in astronomy, the greatest performance boosts have typically come from improvements in detector technology, so one should always be on the lookout for new detector technologies, and if necessary design the system around them.

Heterodyne interferometry Heterodyne and other amplification methods have a fundamental disadvantage at optical wavelengths where the "photons per mode" from a typical astronomical source are significantly less than unity. At mid-IR wavelengths this technology is viable, as shown by the ISI group, and there is a potential future for mid-IR work when systems with larger number of telescopes are built, because there is little loss due to signal splitting between telescopes with the heterodyne technique.

Vacuum pipes I owe this one to Dave Mozurkewich: all interferometer designers start out by trying to avoid the use of vacuum systems for beam transport. All end up regretting not using them. Vacuum systems are very broadband, do not suffer from dispersion or polarisation effects and are more reliable than most people think.

Functionality In the design of an interferometer, it very is easy to add too much functionality to the system. Interferometry is a powerful technique and it is tempting to design interferometers to make use of every "trick" that interferometers are capable of, for example imaging, resolved spectroscopy, astrometry, differential phase etc. Conversely, to get interferometers to work at the level that is required to do robust science requires careful attention to detail. It is very difficult to keep this level of scrutiny on a system as complex as an interferometer, and this gets harder and harder the more functionality is added. In addition the cost of the interferometer goes up strongly with increased functionality.

The result is that many interferometers get stuck behind a "funding barrier": the functionality designed into the interferometer requires more funds and time than is available, so what gets built is an interferometer with lots of potential but only partly complete and only partly debugged. This partly complete interferometer cannot produce very much science, and more importantly, it can produce much less science than was promised in the initial plans for the interferometer. The gap between actual and promised science output is likely to reduce the chances of further funding, entering a vicious cycle which is hard to get out of once it is entered.

A better approach may be to promise less and deliver what is promised. This means that we should design interferometers which can do just enough science to get further funding, rather than one which can do everything given sufficient funds.

Acknowledgments

I would like to thank Mark Colavita, John Davis, Denis Mourard and Charles Townes for useful discussions on the topic of "lessons learned". Nevertheless I take full responsibility for the opinions expressed here and any factual errors.

References

Baldwin, J.E., Haniff, C.A., Warner, P.J., Mackay, C.D., 1986, Nature, 370, 595
Buscher, D.F., Baldwin, J.E., Warner, P.J., Haniff, C.A., 1990, MNRAS, 245, 7P
Haniff, C.A., Mackay, C.D., Titterington, D.J., Sivia, D., Baldwin, J.E., 1987, Nature, 328, 694
Monnier, J.D., 2003, Reports on Progress in Physics, 66, 789
Quirrenbach, A., 2001, ARAA 39, 353
Shao, M., 1988, A&A, 193, 357
Tuthill, P.G., Monnier, J.D., Danchi, W.C., 1999, Nature 398, 487
Tuthill, P.G., Monnier, J.D., Danchi, W.C., Lopez, B., 2000, ApJ 543, 284
Tuthill, P.G., Monnier, J.D., Danchi, W.C., 2001, Nature 409, 1012