Ground-based Interferometry: the past decade and the one to come

John E. Baldwin

Cavendish Astrophysics Group, Cambridge University, Cambridge CB3 0HE, United Kingdom

ABSTRACT

This paper aims to give a broad view of the progress achieved in ground-based interferometry over the past ten years and to assess quantitatively the factors determining the types of object that can be observed with high resolution over the next ten.

1. INTRODUCTION

Where have we come from in science and where are we going are always interesting questions to ask. But how far should we look in each direction? In selecting the title for this talk, the organisers of this meeting chose a decade, perhaps some indication that they see the field of ground-based interferometry as mature. Certainly many working in interferometry have been browsing in the field for a long time. This meeting seems the right time to take a critical look at the past and the future prospects, rather than simply to give an upbeat award of accolades for past work and visionary glimpses of future triumphs.

So for the first part of the discussion we need to answer several questions. What was interferometry ten years ago? Where are we now? What have we learned? What progress has there been overall? And for the future, which is only ten years for this talk, how can we set realistic expectations for interferometry rather than promise too much? What limiting sensitivity may be achievable? What is likely to be the contribution of adaptive optics on that timescale? What will be the observational characteristics of targets of interest? What filling factor should arrays have? The most important question, what new science will come out of the next ten years, can only be answered by the interferometric community going out and doing the work.

2. INTERFEROMETRIC SCIENCE 10 YEARS AGO AND NOW

The astronomical problems that could be attacked by interferometry ten years ago were strictly limited in scope. All of the long-baseline interferometers then in operation would permit observations only on a single baseline at any one time. Most would allow later data from other baselines to be combined in any analysis, but in all cases the data comprised fringe visibility amplitudes with no phase information. A direct consequence of this was a separation of astronomical programmes into two distinct categories. In the first, the data can be fitted by a very small number of parameters giving an adequate description of the target object, such as the angular diameter of a star treated as a uniformly bright disk or the angular separation, position angle and relative brightness of the two components of a binary, each of which was individually unresolved by the interferometer. Such studies give precise determinations of important quantities provided that the model assumptions are correct. In the second category are observations aimed at attempting to understand some aspect of the physical situation, for example the basic geometry in cases where there is no existing good model, such as the spatial extent and radial distribution of circumstellar matter near a star. The key advance sought here is essentially qualitative, not quantitative. The sharp distinction between the two categories was an essential feature of interferometry ten years ago but it is still important and will remain so, in my view.

Brief summaries of the astronomical programmes being pursued using long baseline interferometry in 1991 and at the present time are presented in Tables 1 and 2. Material for the former was taken from the 1991 ESO Conference on High-Resolution Imaging by Interferometry and for current topics from the invaluable OLBIN website lists of interferometric papers sorted by year of publication edited by Peter Lawson.

Further author information:
J.E.B.: E-mail: jeb@mrao.cam.ac.uk, Telephone: +44 1223 337299

Table 1. Astronomical programmes in 1991

Stellar diameters	$\sim$ 1% precision
Binary orbits	tens of mas major axes
	$\sim$ 0.2 mas scatter
Wide angle astrometry	$\sim$ 10 mas precision
Limb darkening	
Atmospheric extension	aperture mask imaging
Stellar shapes	"
Stellar surface features	"
Circumstellar envelopes	mid-IR dust, Hα

Table 2. Current astronomical programmes

Stellar diameters	dwarfs,giants,Miras,Cepheids
Stellar oblatenes	
Limb darkening	
Binary orbits	down to ~ 4 *mas*
Stellar pulsation	Miras,Cepehids
Stellar surface features	
Symbiotic stars	
Be star envelopes	
YSO's	
Dust in late evolutionary and WR systems	
(+ many technical issues)	

Obvious features of the Tables are the diversification of the programmes into new areas and the larger range of types of object studied within one area. These advances are mainly associated with the use of longer baselines, which have also improved the precision of angular diameters and orbital parameters but by relatively modest factors.

3. INTERFEROMETRIC ARRAYS 10 YEARS AGO AND NOW

Most interferometric arrays undergo a continous programme of commissioning and development, so it is difficult to be precise about the physical setup and performance of each array exactly 10 years ago and now. An attempt has been made using the description of each array as it appears in papers describing astronomical observations or instrumentation conferences close to those times. The lists in Tables 3 and 4 must therefore be read in that way, not as the planned performance of the completed instrument. But they do give the best way of judging actual progress made over 10 years.

Obvious features of Table 3, describing the situation 10 years ago, are the small apertures used. With one or two notable exceptions, they operate on only a single baseline, the maximum baseline is a few tens of metres and poor sensitivity limits observations to the brightest stars. There is no evident preference for operation at shorter visible or longer IR wavelengths, the better quality of the atmosphere at long wavelengths being offset by the poorer performance of detectors in the IR. An entry in [ ] for the WHT corresponds not to a separate element array but aperture masking on a single large mirror. The short maximum baseline is offset by the large number of baselines, which allow true imaging to be carried out.

The corresponding Table 4 for today shows improvements in all areas but much the same pattern overall. New arrays are being commissioned and some of the earlier ones have been taken out of operation. The main changes are the first commissioning trials of large apertures, the first baselines greater than 100 m and three arrays with several simultaneous baselines. An entry, as before, for aperture masking on a large single-mirror telescope shows the very large gain in baseline coverage attainable by this technique. The improvements in limiting magnitude relative to 1991 are useful but modest.

Table 3. Array parameters in 1991

Array	Aperture/m	max baseline	no.	mag limit/band
I2T	0.26	67	1	~ 3 V
MkIII	0.05	31	1	~ 5 V
GI2T	1.52	65	1	6 V
ISI	1.65	32	1	0 N
[WHT	0.3	4	15	5 R]
Fluor	0.8	5.5	1	~ 2 K
IRMA	0.2	16	1	~ 0 K
SUSI	0.14	80	1	3.5 V
COAST	0.16	2.8	1	2 I

Table 4. Array parameters in 2002

Array	Aperture/m	max baseline	no.	mag limit/band
ISI	1.65	32	1	0 N
GI2T	1.52	65	1	6 V
SUSI	0.14	80	1	3.5 V
COAST	0.40	47	9	6.4 I
IOTA	0.45	38	3	6.1 K
NPOI	0.5	40	15	4.5 V
PTI	0.4	110	1	5.6 K
[KECK	0.3	10	210	0 K]
MIRA-1.2	0.3	30	1	
CHARA	1.0	330	1	
VLTI	8.2	102	1	4.9 K
KECK	10	85	1	2 K

4. DEVELOPMENTS IN BASIC TECHNOLOGY 1991-2002

Some fields of research are transformed by the developments in technology not directly related to that particular subject. That has not really been the case in optical interferometry over the last ten years. No fundamental changes have occurred in the design of telescopes, beam transport, delay lines, spectroscopy or metrology. Developments in piezo-electric devices have improved the implementation of tip-tilt correction of the input wave-fronts. Fibre couplers in the near-IR have been used with great effect for beam combination (but note that they were already in use by 1991) and integrated optics offers interesting prospects for the near future. The main developments of importance have been in computers for all aspects of control of arrays, data acquisition and processing and in infrared array detectors where the improvement in quantum efficiency and readout noise have shifted the balance of avantage strongly towards wavelengths of 1-2μm. The recent introduction of LLL-CCD's offers the prospect of replacing the avalanche photodiodes currently in use by a much larger number of pixels with comparable sensitivity at lower cost, an important development for spectroscopy.

5. 10 YEAR SUMMARY

As a broad generalisation it is probably true to say that progress in interferometry over the past ten years has been good but not astonishing. The output of astronomical results in papers has increased by roughly a factor of 10, including large increases in the number of measured diameters of differing stellar types and of binary orbital parameters. New fields have been opened up with the detection of stellar oblateness in a rapidly rotating star, precise comparison of limb darkening with stellar models, direct measures of pulsation in Mira and Cepheid variables, of the shape of Hα envelopes in Be stars and dust envelopes of supergiants, together with spectacular images from aperture masking of envelopes of WR stars and systems at a very late evolutionary stage. Of almost equal importance has been the improvement in quality of the data from interferometers, so that arguments now are not about the validity of the data itself but the astronomical interpretation. In that sense interferometry has become part of the mainstream of astronomy.

The rate of progress in the field seems to have been steady, both in the technical development of interferometry and in its astronomical exploitation. There have been no major obstacles (though it is certain that individual groups will feel they have had to surmount some!) but equally there have been no major breakthroughs transforming the way interferometry is carried out or in the astronomical programmes which can be pursued. A comparison might be made here with the corresponding development of radio interferometry in which improvements by a factor of 100 per decade in both resolution and sensitivity continued over a period of more than four decades, with correspondingly dramatic advances in the astronomical output. Factors which may have influenced the very differing rates in the two fields are that interferometric angular resolution was essential for the development of radio astronomy whereas, at optical wavelengths, interferometry is in direct competition for resources, both funding and people, with the development of large single-mirror telescopes. Interferometry is at present certainly the junior partner. Most of the arrays listed in section 3 have suffered serious delays in completion of construction and/or lack of support for operations because of funding problems. The consequent slowdown in the rate of progress has been very unfortunate. It raises the question of what realistic expectations we can put forward for interferometry in the future.

6. WHAT WILL WE *NOT* BE DOING WITHIN THE NEXT 10 YEARS?

There are at least two ways to try to define the field of interferometry over the next ten years. The traditional way is to set a number of goals and then define the field as everything which lies within the fence set by these goals or boundary posts. The problem with this strategy is that there is a strong temptation to set the most extended and exciting goals, which may prove extremely difficult or impossible to reach. The result can be disillusion for those within interferometry and, importantly, for the rest of the community. An alternative approach may be to draw up a list of those topics or techniques which we believe will *not* be implemented within the next ten years. The field of interferometry is then defined as everything which has *not* been excluded. The advantages of this approach are that desirable, but perhaps unrealistic, goals are examined in detail first and rejected if there are persuasive reasons, rather than being left as implied goals for the future. I think there is a strong sense of release in not having to worry about issues that are definitely more than ten years in the future. Furthermore the field remaining has, by its definition, more flexibility. As an example of this procedure, Table 5 lists a number of items which I believe we should exclude at this stage, and the reasons for so doing.

Table 5. Techniques not implemented in the next 10 years?

technique	difficulty	cost	motivation
arrays+AO+laser guide star	X	X	
array baselines >1-2km			X
mas imaging of asteroids	X		
nulls on sky <10-4	X		

For instance, the present state of development of laser guide stars and AO on single telescopes suggest that implementation on an array of several widely spaced telescopes would be hard to justify on ground of both diffivulty an expense when the unexplored field of interferometry at bright magnitudes is so extensive. Baselines > 1-2km are certainly achievable without excessive difficulty, (remember that Michelson used a one-mile evacuated pipe as early as 1930), but there seems no very strong scientific case for implementing them now.

Table 5 is intended simply as an incomplete and non-controversial starting list. Readers are encouraged to construct their own. I strongly recommend it as an exercise that clarifies the major issues.

7. IS THERE A LIMITING MAGNITUDE FOR INTERFEROMETRY?

If we accept the argument of the previous section and, in particular, exclude the possibility of arrays equipped with both AO and laser guide stars within the next 10 years, then it is certain that most targets will continue to be stars and the material in their close neighbourhood. What fraction of objects of such objects will be accessible to observation? Taking a simple model of a star as a black body of angular diameter 1 mas at temperature T, the apparent magnitude as a function of wavelength is plotted in figure 1 for a number of values of T. The choice of 1 mas (the resolution of a 200 m aperture at 1 μm) is designed to be a good match to the angular resolution of several arrays over the next 10 years.

Wavelengths longer than K have been omitted from the figure but it is evidently straightforward to extrapolate all the curves to the mid-IR. The omission is to avoid compression of the shorter wave bands on the figure, and in part because magnitudes are less familiar and less used at mid-IR wavelengths. The effects of atmospheric extinction and thermal background also differ markedly from the short wavelength regime.

Two implications of the diagram are significant: first, that all stars which are larger than 1 mas, and hence individually resolvable, are brighter than $m = +6$ and essentially observable with the current sensitivity of interferometric arrays. Secondly, objects that are much fainter than, say $m = +8$, will be unresolved and it will be only the material surrounding them which can be imaged usefully. For those objects, the presence of the unresolved stellar flux at all baselines solves the otherwise difficult problem of boot-strapping many short baseline measurements to allow long integrations of long baseline data. It serves exactly the same role as unresolved nuclear sources in VLBI imaging at radio wavelengths, which were crucial to the rapid development of imaging in that subject.

An important question is what limiting sensitivity can be expected as a function of wavelength for an optimum array on a good site? The dashed line on Figure 1 represents an estimate based on the following realistic assumptions:

1. For $\lambda_0 = 500\ nm$, $r_0 = 0.15m, t_0 = 5ms$.
2. Tip-tilt corrected apertures of the optimum diameter $2.75 r_0 (\lambda/\lambda_0)^{1.2}$ and integration time $2t_0(\lambda/\lambda_0)^{1.2}$ are used to give the best signal-to-noise ratio.
3. The target is unresolved.
4. A signal-to noise ratio in a single integration, S/N$\simeq$ 2.5. This allows successive integrations to be added coherently, giving a final S/N which improves as the square root of the number of integrations, rather than much more slowly if S/N $\sim$ 1.
5. The detector has zero read noise and dark count. This is a fairly accurate description of the behavior of avalanche photodiodes and the new generation of low-light-level CCD's. It is not yet the case for the near-IR but is expected to be true for the new generation of Rockwell detectors.
6. A fractional bandwidth of 15% is usable at all the standard wavebands.
7. The overall throughput from the top of the atmosphere to detected photons is 10%.

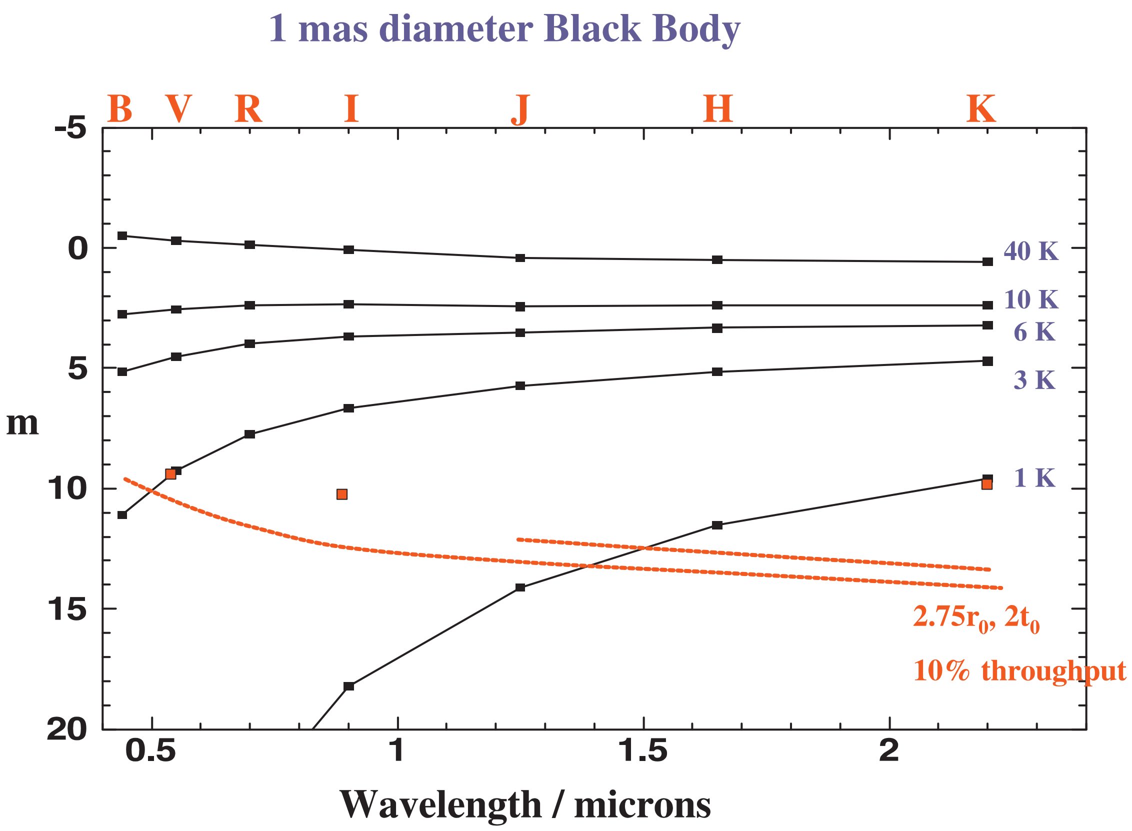

Figure 1. The apparent magnitudes of 1 *mas* diameter black bodies and the limiting sensitivity for tip-tilt corrected interferometers. For discussion of the assumed atmospheric and instrumental parameters, see text.

The limiting sensitivity improves slowly with increasing wavelength from $m = +10$ at B to $m = +14$ at K. The upper dashed line between J,H and K corresponds to the current detector performance with $6e^-$ readout noise. Most observations to date have been taken with apertures much smaller than the optimum for S/N or in significantly poorer atmospheric conditions. If currently observed limits are scaled to the standards listed above, the best values are indicated by square symbols in Figure 1. They fall short of predictions by 2-3 m, a factor which is currently not explained. It accords qualitatively with informal discussions I have had at this meeting with members of several interferometry groups whose common experience is that the throughput of their own interferometer falls short of that expected on the basis of the known reflectivities of the components by a significant factor, though not as large as 2-3 m. A possibility mentioned more than once is that of weak large-angle scattering. Such light from many component surfaces in the optical train may be included in standard measurements of reflectivity but not reach the fringe detector.

8. HIGHER-ORDER AO IN INTERFEROMETRY

All current interferometric arrays use tip-tilt correction of the wavefronts at the individual telescopes. An important question is whether the use of higher-order adaptive optics provides benefits for interferometry. Here I assume that laser guide stars are excluded, so that an AO system must operate on the target object itself, or on a nearby bright star if one is available. For the case of bright targets, AO can provide two benefits. First, the distorted optical wavefronts across existing apertures can be corrected, thus giving higher visibilities of the detected fringes. Second, larger apertures can be used, providing a larger detected photon flux. Each improves the signal-to-noise ratio and these improvements would allow an observation to be made in a shorter integration time or, alternatively, would permit higher spectral resolution to be used without penalty.

A more controversial question is whether, as is often assumed, higher-order AO will enable interferometry to reach fainter limiting magnitudes. I think there are strong arguments, both qualitative and quantitative, that it will not. Consider first the correction of a sub-aperture within some larger aperture which is to be adaptively corrected. The aim of AO is to correct both the tilt of the wavefront across the subaperture and its piston displacement from some chosen reference surface. That reference may be the mean surface of the complete aperture or indeed any other reference. In particular it may be an entirely separate wavefront of similar extent, for example that emerging from the second element of an interferometer. The procedures of establishing the wavefront tilt and the difference in piston value between them for AO are exactly the same as those required to carry out interferometry. Given identical atmospheric and instrumental parameters, each will require the same photon flux to achieve the same signal-to-noise ratio. So the limiting magnitudes for AO and interferometry will be the same. Put another way, if there are enough photons to carry out AO on a subaperture, then there are sufficient photons to do interferometry with that same subaperture. So correcting a larger aperture is not essential.

This qualitative argument indicates that the sensitivity limits for AO and interferometry are very similar but leaves some aspects a little vague. Consider now the optimum aperture size for interferometry to give the best signal-to-noise ratio, using tip-tilt correction only. This is well-established to be $2.75r_0$ and gives a mean fringe visibility, $V = 0.49$. In a single measurement the signal-to-noise ratio $\mathrm{S/N} = \frac{1}{2}V\sqrt{N}$, where N is the number of photons in the sample. If we set a requirement for the S/N ratio of a single measurement of $\mathrm{S/N} \geq 2.5$, which is necessary for coherent fringe summing, then the number of photons, $N \geq 100$ in the optimum integration time $2t_0$. How does this number compare with the number of photons required to carry out the tip-tilt correction with adequate precision? Assuming an identical detector, wavelength, band and optical throughput, and that the tip-tilt system behaves perfectly in the sense that both the measurement of the tilt and the correction are applied instantaneously at the end of the integration, then the only errors in the correction are due to photon noise in the measurement and the change in the tilt during the integration. The best precision is obtained if the errors from these two causes are made equal, which occurs for an integration time of $\simeq 2t_0$. The precision in the tilt correction necessary to avoid a reduction in the fringe visibility of $\leq 10\%$ is $\frac{1}{4}\lambda$ across the diameter of the aperture or $\simeq 0.2$ of the FWHM of the diffraction-limited peak in the image (which contains only half of the light). The number of photons needed to achieve this precision is $\simeq 50$, which implies that the limiting sensitivity for the tip-tilt correction in the ideal case is slightly better than for fringe detection, but certainly by less than 1 magnitude.

To extend the argument to correction of a larger aperture for operation at the same wavelength, we simply note that the large aperture will be divided into sub-apertures each $2.75r_0$ in size and hence that the limiting magnitude for AO correction of the whole will be identical to that for tip-tilt correction of each sub-aperture. For the circumstance where the AO correction might be based on wavefront sensing in the visible for correction of a larger aperture for use in interferometry in the IR, then the sensitivity for AO correction proves to be definitely poorer than that for interferometry. These arguments, which do not depend on the detailed properties of the AO system but assume it to be ideal, indicate that there seems no prospect of reaching fainter limiting magnitudes by the use of AO.

9. DYNAMIC RANGE OF TARGETS AND IMAGES

The issue of dynamic range in interferometric imaging has been widely discussed in regard to the problem of detecting planets close to a star and has led to the development of the whole field of nulling interferometry, which is not reviewed here since it has yet to go beyond the stage of first trials on a telescope. There has been little discussion of the less extreme, but still difficult to achieve, dynamic range required in the general situation where imaging of material in the neighbourhood of a star is planned. Simple arguments suggest that the imaging of material further from a star than 100 stellar radii will be hard to achieve.

There are two regimes to consider. First, at the shorter visible wavelengths, when material surrounding the star is detectable by reflected light and secondly, at longer wavelengths, where the detectable radiation is thermal radiation from the material in thermodynamic equilibrium with the stellar radiation. For the case of reflected light, the surface brightness of material at a distance R from a star of radius R_* will be $\alpha(R/R_*)^{-2}$ where α is the albedo of the material. For a typical value of α of 0.1, the surface brightness falls below 10^{-5} of that of the star beyond $100R_*$. The temperature, T of the material in thermal equilibrium with the stellar radiation at a distance R from a star of temperature T_* will be determined by $(T/T_*)^4 \simeq (R/R_*)^{-2}$. So that at $100R_*$ the material temperature will be a factor 10 below the stellar surface temperature giving a surface brightness in thermal emission only $\simeq 10^{-4}$ of that of the star, except at wavelengths $\geq 2\mu m$.

How do these estimates compare with the quality expected for interferometric images? The dynamic range in such an image will be determined by the errors in the visibility amplitudes and phases and the photon noise in the data. For bright enough targets, where photon noise can be neglected, and in which the errors in both amplitudes and phases are randomly distributed, it is expected that the dynamic range (the ratio peak in the image: rms sidelobe level) will be given by

$$n^{1/2}[(\delta A/A)^2 + (\delta\phi)^2]^{-1/2}$$

where n is the number of individual visibility measurements and $\delta A/A$ and $\delta\phi$ are the fractional error in visibility amplitude and the error in (closure) phase respectively. Assuming fairly optimistic values of $n \simeq 10^4$ and both $\delta A/A$ and $\delta\phi \simeq 10^{-2}$, then the dynamic range is $\simeq 10^3$. This example shows how difficult it may be to reach a dynamic range of 10^4; 10^5 may well prove to be unrealisable, almost certainly so in the coming decade. So detection of matter beyond $100R_*$ seems unlikely except at wavelengths in the mid-IR.

This conclusion may seem suprising, given that images showing extensive circumstellar material have already been obtained, in particular from aperture masking observations. The apparent conflict arises from the comparison made above of the surface brightness of the circumstellar material with that of the stellar surface. It is the appropriate ratio to adopt if the star itself is resolved in the image. But in a more compact configuration of the array, the image will be of lower resolution and the angular width of the synthesised beam may be say 10 times that of the star. The peak brightness of the star in the image is reduced by a factor of 100, whereas the surface brightness of extended surrounding material may be unchanged. So the dynamic range needed in the image is also reduced by a factor of 100 and may be well within current capabilities in imaging. This good surface brightness sensitivity of compact arrays with a high filling factor is well known from radio interferometry and will be of equal or greater importance in optical interferometry. It is notable that all current separated-element arrays may be described as sparse, having the chief aim of resolution rather than surface brightness sensitivity. Higher filling factors are currently achieved only in aperture masking experiments and in the near future by the LBT.

10. CONCLUSIONS

The purpose of this paper is simply to set the scene for the discussions of interferometry at this meeting by summarising very briefly how we have reached the current state of the subject and by raising issues which will be of importance in constraining the science which will be possible over the next ten years. As such, it is not appropriate to draw specific conclusions. The current position and visions of the future are best described by the papers which follow this one.