The application of interferometry to optical astronomical imaging

By John E. Baldwin and Christopher A. Haniff

Department of Physics, University of Cambridge, Madingley Road, Cambridge CB3 0HE, UK

In the first part of this review we survey the role optical/infrared interferometry now plays in ground based astronomy. We discuss in turn the origins of astronomical interferometry, the motivation for its development, the techniques of its implementation, examples of its astronomical significance and the limitations of the current generation of interferometric arrays. The second part focuses on the prospects for ground-based astronomical imaging interferometry over the near to mid-term (i.e. 10 years) at optical and near-infrared wavelengths. An assessment is made of the astronomical and technical factors which determine the optimal designs for imaging arrays. An analysis based on scientific capability, technical feasibility, and cost argues for an array of large numbers of moderate sized ($\sim$ 2m-class) telescopes rather than one comprising a small number of much larger collectors.

Keywords: Interferometry, aperture synthesis, optical astronomy

1. The origins of astronomical interferometry

The apparent angular sizes of stars seen through conventional telescopes at high magnification are vastly greater than their real dimensions. William Herschel knew this as early as 1779 (Herschel 1805) but did not know why. The key lay in Thomas Young's experiments that demonstrated interference and the wave nature of light unambiguously. As he stated in his Bakerian lecture of 1803: "The proposition on which I mean to insist at present, is simply this, that fringes of colours are produced by the interference of two portions of light" and later: "that homogeneous light, at certain equal distances in the direction of its motion, is possessed of opposite qualities, capable of neutralising or destroying each other, and of extinguishing the light, where they happen to be united" (Young 1804).

But it was not Young's researches that prompted Herschel to investigate the origin of the spurious diameters of stars. Instead it was the exactly contemporaneous discovery of the first minor planets: Ceres in 1801, Pallas in 1802 and Juno in 1803. Were their apparent diameters as real as those of planets or spurious as for stars? To address this question Herschel conducted an extensive series of experiments in his garden in Slough, examining through his telescope small globules of differing sizes and materials placed in a tree some 800 feet away (Herschel 1805). His observations showed that for the smallest globules the diameters were all spurious and all of the same size. Furthermore, he found that, if just the inner part of the aperture of the telescope were used, the spurious diameters, whether of globules or of stars, were larger. If the whole aperture was employed, the diameters were

smaller, and if only an outer annular aperture was used the diameters were smaller still. This experimental discovery, that unfilled apertures can be used to obtain high angular resolution, remains today the essential basis for interferometric imaging in astronomy.

The theoretical justification of this result came only after thirty years with Airy's analysis of the diffraction pattern of a circular aperture (Airy 1835), and it took a further thirty years before the idea of using multiple apertures was developed. In an early study the Rev. W. R. Dawes noted that he had "frequently found great advantage from the use of a *perforated whole aperture*" and that when observing Venus this produced "a central image of the planet perfectly colourless, and very sharply defined" (Dawes 1866). But it was left to Fizeau, in his submission to the Commission for the Prix Bordin the following year, to remark on "une relation remarquable et nécessaire entre la dimension des franges et celle de la source lumineuse" and suggest that by using an interferometric combination of light from two separated slits "il deviendra possible d'obtenir quelques données nouvelles sur les diamètres angulaires de ces astres" (Fizeau 1868).

Steps towards the practical implementation of these techniques for optical astronomy were taken by Michelson who defined the "visibility" of interference fringes obtained from a source of finite angular size (Michelson 1890) and followed this a year later with the measurement of the angular diameters of Jupiter's satellites (Michelson 1891). Finally, roughly thirty years later Fizeau's predictions became a reality when the direct interferometric measurement of a stellar diameter was realized by Michelson & Pease with their 20-foot interferometer mounted on the 100-inch Mount Wilson telescope (Michelson & Pease 1921).

2. Interferometric imaging in astronomy

(*a*) *Radio and optical development*

Interferometry provides access to very high angular resolution observations. It also importantly separates the issues of angular resolution and limiting sensitivity. A single mirror of diameter D has an angular resolution, $\theta \simeq \lambda/D$, and a collecting area for the flux of photons $\sim D^2$, so that there is a well-defined relation between resolution and sensitivity. Since astronomers necessarily study objects beyond their control, it is unlikely that this fixed relationship will be a good match to all but a subset of their observational requirements.

It was for this reason that the growth of radio astronomy after 1945 depended so dramatically on the use of interferometric methods. The extreme imbalance between the excellent sensitivity of small filled apertures at long radio wavelengths and their poor angular resolution, which could be many tens of degrees, led naturally to the development of sparse arrays of widely separated telescopes. In these the telescope size could be chosen to provide adequate sensitivity whilst large interferometer baselines would give the necessary angular resolution. For the incoherent sources that populate the astronomical sky, measurements of the spatial coherence function, i.e. the amplitude and phase of the interferometric fringes, would then provide a direct measurement of the Fourier transform of the sky brightness distribution. Initially the baselines used were tens to hundreds of metres, but these were extended successively to thousands of kilometres over the next forty years. Aperture

synthesis imaging became well established, with arrays involving many hundreds of simultaneous interferometer baselines, and exploiting the rotation of the Earth to move those baselines steadily to provide sensitivity to many other angular scales on the sky. The quality of the images in both resolution and sensitivity were improved progressively by a factor $> 10^9$ and the dynamic range (the ratio of the brightest to the faintest believable feature) in the best images now exceeds 10^5.

In contrast, at optical/infrared wavelengths the opposite situation obtained: even early telescopes had good angular resolution and the principal driver for technical development was to achieve better sensitivity. However, at these wavelengths larger telescopes do not directly give sharper images. Atmospheric turbulence ("seeing") limits their angular resolution to $\sim \lambda/r_0$ radians, where r_0 is Fried's parameter, the diameter over which the rms wavefront fluctuation is approximately 1 radian. At good observatory sites $r_0 \sim 15$ to 20 cm at 500 nm, giving a seeing-limited angular resolution of order 0.6 arcseconds. Though far poorer than the diffraction limit this was more than adequate for a wide range of astrophysics.

Technical difficulties also delayed the growth of optical interferometry. The human eye is a sensitive detector but is not capable of quantitative photometric assessment of interferometric fringe patterns. When coupled with the need to record data at kHz rates, progress in optical synthesis imaging had to wait for the development of sensitive photon-counting detectors. Furthermore, the limits of mechanical stability had been reached with the fifty foot beam interferometer constructed by Michelson and Pease in 1930 (Pease 1931). Further advances would require the interferometric combination of light from separate telescopes spaced by many tens of metres (Labeyrie 1975). This in turn later demanded micron-level metrology of variable optical delay lines, which was not feasible until access to stabilized lasers had become routine.

Once these technical considerations had been addressed, all of the principles used at radio wavelengths could be taken over with almost no modification. This included the use of imaging software developed for VLBI at radio wavelengths which was used to reconstruct the first image from an array of optical telescopes, that of the 50 milliarcsecond binary star Capella (Baldwin *et al.* 1996). The only significant differences between the two wavelength regimes are the small temporal and spatial scales of the atmospheric fluctuations at optical wavelengths. For example, the characteristic timescale for these fluctuations is measured in milliseconds at optical wavelengths rather that minutes, and the spatial scale is typically smaller than the telescope mirror diameter, whereas at centimetric radio wavelengths this scale can be as large as 20 km. An important consequence of this small spatial scale is that the area of sky over which the atmospheric phase path is constant, the isoplanatic patch, is at most a few arcseconds at visual wavelengths.

(*b*) *Practice and problems of optical/infrared arrays*

The modern development of Young's original experiments with light is manifest in a number of optical interferometric arrays that have been or are being constructed for astronomy. While they differ in their precise implementations and their detailed scientific goals, their overall design is broadly similar. All share the following basic components (see figure 1):

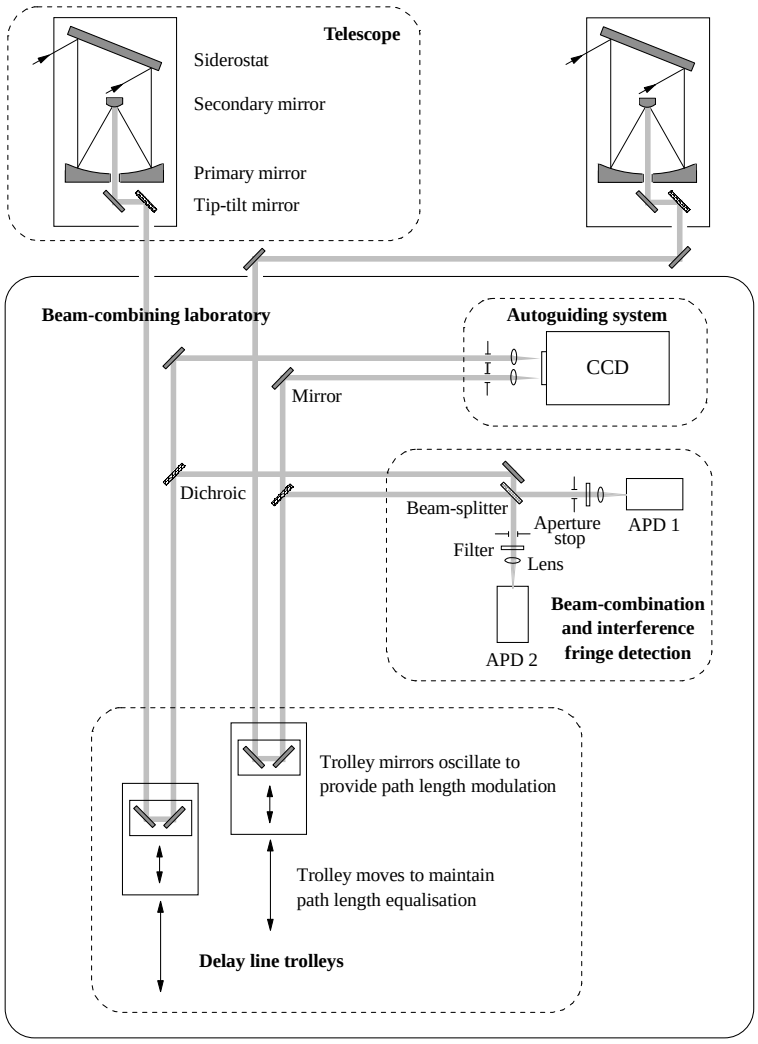

Figure 1. A schematic view of a modern optical/infrared interferometer. Only two of the array elements (in this case siderostat-fed two-mirror telescopes) are shown here.

- An array of movable telescopes, acting as the primary collection elements, each provided with rapid auto-guiding on a star,
- A beam transport system delivering afocal beams from the telescopes to a central beam-combining laboratory,
- A system of optical delay-lines to compensate for the geometric path delay introduced when observing a source away from the zenith,
- A beam combiner, or correlator, to mix the input signals,

Figure 2. An external view of a typical modern optical/infrared interferometer, the Cambridge Optical Aperture Synthesis Telescope (COAST). All five of the telescope enclosures are visible, together with the alloy beam transport pipes, and the grass-covered beam combining laboratory in the background. The maximum baseline in this configuration is approximately 25 m.

- A fast low-noise detector system with the ability to record the interferometric fringes on a timescale short with respect to the fluctuation timescale of the atmosphere.

In practice, the methods of implementing these functions are also common to many arrays. Each of these is described in brief below.

In cases where telescopes have been specifically designed for use as part of an interferometric array, their aperture sizes have generally been matched to the local seeing conditions. In practice, this means the use of telescopes no larger than $2.75r_0$ in diameter. This maximizes the signal-to-noise ratio for fringe detection provided that the telescopes have rudimentary correction for atmospheric wavefront errors (Buscher 1988). More precisely, this is the optimum telescope size if atmospherically-induced fluctuations in the tilt of the wavefront of the incoming beam, or instantaneous image position, can be fully corrected with a fast autoguider. At optical wavelengths current synthesis arrays operate with apertures of between 15 and 50 cm, while in the near-infrared, e.g. at $2.2\,\mu$m where r_0 is much larger ($r_0 \propto \lambda^{6/5}$), aperture sizes of up to 2.4 m can be usefully exploited. This is considerably smaller than the aperture sizes typical of large modern optical/infrared telescopes.

Existing synthesis arrays all use identical beam transport and delay compensation techniques. The preferred method for transport is free-space propagation of a collimated beam in an air-filled or evacuated pipe. To limit diffractive beam spreading for paths longer than about a hundred metres, large diameter (~ 10 cm) optical beams are usually employed. However, under realistic seeing conditions it can be advantageous to allow a moderate amount of diffraction to spatially filter the beam profile and "strip-out" components of the light that will not provide a useful interferometric signal (Horton, Haniff & Buscher 2001). Once the signals from the individual collectors have arrived at a central beam combining laboratory, equalization of the optical paths is accomplished with "optical-trombones", i.e. arrangements of retro-reflecting optics located on moveable carts. These run along precision tracks, either in air or in vacuum pipes, and allow for compensation of the optical path differences associated with the layout of the array elements and the

movement of the star across the sky. They also permit the introduction of deliberate path modulations that are used to encode the interference signals.

Finally, after optical path matching to within the coherence length of the radiation being examined ($\sim \lambda^2/\Delta\lambda$), the beams from the array elements are mixed and an interference pattern detected. Unlike Young's original two-slit experiment, in which the signals were combined in an image plane and produced an output with a spatially encoded fringe pattern, most optical interferometers use correlators comprising beam-splitter plates which combine afocal beams in a manner identical to a laboratory Michelson interferometer. In this case the combined beams can be focused on single element low-noise detectors, usually photon-counting silicon avalanche photodiodes, and the resulting fringe signals are realized as a time-varying intensity arising from appropriate modulation of the optical paths in the interfering beams.

The precise form of this output illustrates how the conditions under which an interferometer is operating affect its performance. For a two-beam system, the detected intensity can be written as:

$$I(\Delta l, t) = I_1(t) + I_2(t) + 2\sqrt{I_1(t)I_2(t)}\ |\eta(t)|\ |\gamma| \cos\left[\frac{2\pi[\Delta l + \Delta l'(t)]}{\lambda} + \theta\right]\ M\left[\Delta l + \Delta l'(t)\right], \quad (2.1)$$

where $I_1(t)$ and $I_2(t)$ are the instantaneous intensities in the two input beams, Δl is the optical path difference (OPD) introduced by the delay lines, $M(\Delta l)$ is a bandwidth-dependent coherence envelope function that modulates the observed fringe pattern, and $|\gamma|$ and θ are the modulus and phase of the visibility function, i.e. the Fourier transform of the brightness distribution. The effects of the atmosphere are contained in the two additional factors $|\eta(t)|$ and $\Delta l'$. These represent the reduction in apparent fringe contrast due to spatial wavefront corrugations across the telescope apertures and the mean atmospherically-induced OPD error between the two telescopes.

Equation 2.1 implies that useful data will be secured only when the following conditions are met:

1. Measurements are made on a timescale short compared to the fluctuation timescale of the atmosphere, t_0. At good observatory sites this takes values of order 5 ms ($\sim r_0/v_{wind}$) at 500 nm and scales as $\lambda^{6/5}$. The upshot of this is a need to record individual measurements at kHz rates.

2. The coherence loss, $= 1 - |\eta(t)|$, due to residual spatial wavefront corrugations across the interfering beams is both small and known. In practice this forces the use of moderately sized telescopes and requires frequent observations of unresolved calibration targets. Larger telescopes can be used but only if they are combined with effective adaptive optics (AO) systems, i.e. systems that sense and correct for wavefront corrugations across the telescope apertures in real-time.

3. The total OPD between the interfering beams is sufficiently small that the value of the coherence envelope function, $M(\Delta l_{tot})$, is close to its peak value,

unity. This implies real-time monitoring of the atmospheric perturbations between the array elements and subsequent control of the path lengths internal to the interferometer to within a few wavelengths.

The principal difficulties associated with optical/infrared astronomical interferometry are associated with these three requirements. In particular, it is almost impossible to determine $\Delta l'$ well enough to extract values of the source visibility phase, θ, that are useful for direct Fourier inversion (i.e. with errors no larger than $\sim 10°$).

In the absence of such a capability, ground-based interferometers are limited to measuring the modulus of the source visibility function, together with visibility phases that are corrupted by unknown atmospheric errors. That these phases could, in fact, be put to good use was first established in the late 1950s (Jennison 1958) and subsequently demonstrated with a long baseline radio interferometer in 1974 (Rogers *et al.* 1974). Under conditions that obtain both for radio and for optical and near-infrared arrays, linear combinations of simultaneously measured visibility phases around closed loops of baselines are not subject to atmospheric corruption. The use of these so-called "closure" phases, rather than accurate visibility phases, for image recovery has now become a straightforward procedure and is referred to in the interferometric literature as "self-calibration" (see, e.g. Pearson & Readhead 1984 and references therein). While arrays that aim to measure closure-phases rather than uncorrupted visibility phases are much easier to build and operate, the extent to which they can match the imaging performance of phase-stable interferometers depends strongly on the number of array elements present. For an N-element array, this is characterized by an efficiency factor $f = 1 - 2/N$: arrays with fewer than five collectors are thus not really suitable for detailed imaging astrophysics.

3. Current arrays and their astronomical capabilities

A measure of the impact of interferometry at optical and infrared wavelengths for astronomy is the eight-fold increase seen between 1990 and 2000 in the annual publication rate of manuscripts dealing solely with astrophysical results from interferometry. This primarily reflects two main developments: first, a maturing of the methods and technologies associated with interferometric imaging, and, second, the commissioning of a number of modern separated-element arrays.

The former of these has been most successfully demonstrated by Tuthill and co-workers who have used a perforated mask over a single large aperture telescope to convert it to a multi-element interferometer (Tuthill *et al.* 2000) – in effect a generalized Young's multi-slit experiment. This approach, first demonstrated in 1986 (Haniff *et al.* 1987), closely mimics Michelson's early experiments in which the telescope superstructure maintains the necessary optical path equality and an interference pattern is formed in a common focal plane. Interferometric configurations with many tens of collecting elements can be realized with baselines of up to 10 m and using aperture sizes of several tens of centimetres. This latter restriction limits the sensitivity of this method, but it allows for detailed studies of complex astrophysical sources at near-infrared wavelengths (see, e.g. Tuthill *et al.* 2000; Tuthill, Monnier & Danchi 2001 and figure 3).

The availability of large 10 m-class telescopes, fast readout low-noise detectors, and modern imaging algorithms has meant that this technique is now fully mature.

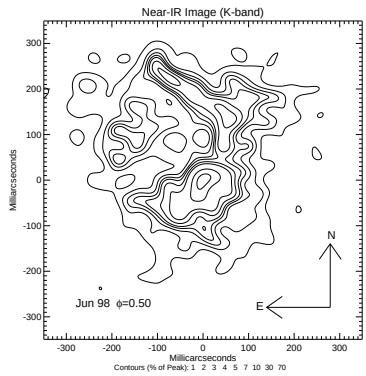

Figure 3. A typical near-infrared interferometric image of the dust-enshrouded evolved star IRC+10216. The star itself is hidden from view at this wavelength ($2.2\,\mu$m). The angular resolution is ~ 50 milliarcseconds, ten times better than a conventional image. Figure courtesy of J. Monnier and P. Tuthill.

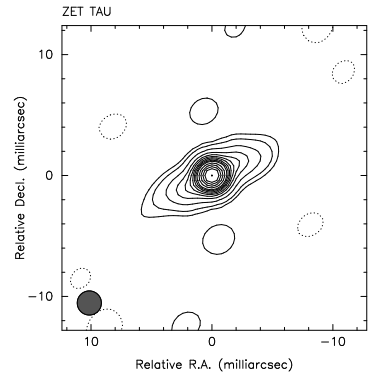

Figure 4. An interferometric image of the disk-shaped Hα envelope of the Be star, ζ Tauri. The emission is concentrated in an inclined disk and extends over ± 6 milliarcseconds. Interferometric baselines of 50 m were used to achieve an angular resolution of ~ 2 milliarcseconds, five hundred times better than a conventional image.

Not only is the quality of the imaging comparable to that obtained by centimetric radio arrays – there is little difficulty in detecting features as small as the diffraction limit and with surface brightnesses a hundred times fainter than the brightest component in the image – but its fidelity has been verified by independent studies of the same sources. Perhaps its only shortcoming is the moderate sensitivity and angular resolution attainable.

The other key development has been the construction and commissioning of a large number (> 10) of separated-element arrays. In principle these offer much better sensitivity and resolution but this comes partnered with the full range of technical challenges outlined in §2. Most of these arrays have been technical prototypes, but a significant subset are currently being operated for regular astronomical observations. The most prominent of these are listed in table 1. There is a predominance of arrays comprising relatively small numbers of 1 m-class collectors arranged so as to deliver ultra-high angular resolution. A 230 m baseline (the mean for the arrays summarized in table 1) corresponds to an angular resolution of 1 milliarcsecond (mas) at $1\,\mu$m, an improvement of roughly one hundred times over that of the Hubble Space Telescope.

In this sense these are ultra-sparse interferometric arrays, sensitive to a limited range of angular scales on the sky for any given configuration of array elements, and hence lacking the versatility for studying complex sources. The fraction of phase information accessible with these arrays ($= 1 - 2/N$) is similarly restricted and so their imaging capability falls far short of those interferometers using masks on single telescopes that have given the clearest pointers to what interferometry can in principle deliver. A useful comparison of this is provided by figures 3 and 4. The latter shows a typical image obtained with COAST of the hydrogen nebula surrounding the hot Be star ζ Tauri. The Balmer line emission region is clearly resolved, but it has a total extent of only a few resolution elements as compared to

Table 1. *Major optical/infrared astronomical interferometer arrays*

(Major optical/infrared astronomical interferometer arrays. In this summary prototype instruments with a primarily technical role have been excluded. N, D, and B refer to the number of telescopes in the array, their aperture size, and the maximum baseline achievable. Both the Keck and VLTI arrays have the possibility of using a mixture of large and more modestly sized telescopes. All figures are approximate. Further details of these and other interferometric projects can be found in Quirrenbach (2001).)

Facility & location	N	D/m	B/m	Primary role
CHARA: Mt. Wilson, USA	6	1.0	350	Stellar astrophysics
COAST: Cambridge, UK	5	0.4	100	Imaging
GI2T: Calern, France	2	1.5	65	Spectroscopy
IOTA: Mt. Hopkins, USA	3	0.4	40	Stellar astrophysics
ISI: Mt. Wilson, USA	3	1.7	85	10 μm astrophysics
NPOI: Flagstaff, USA	6	0.5	440	Astrometry & Imaging
SUSI: Narrabri, Australia	2	0.1	640	Stellar diameters
KECK: M. Kea, Hawaii	$2+4$	$10.0+1.8$	140	Astrometry
VLTI: Paranal, Chile	$4+3$	$8.0+1.8$	200	Imaging & Astrometry

the large extent of the emission mapped in figure 3. This quantitative difference in the ability to map complex structures is perhaps the most distinctive trait of the existing generation of separated-element interferometric arrays.

For the discussion above one may infer the type of astronomical programmes that these arrays have focused on. All have studied the brightest possible targets, i.e. nearby stars, and have aimed to secure high angular resolution multi-wavelength data to compare with theoretical models. Amongst the range of projects that have been initiated are:

- Precise measurements of stellar diameters to investigate, for example, the effective temperature scale and mass-luminosity-radius relationship for normal stars (see, e.g. Perrin *et al.* 1998; Nordgren *et al.* 1999; Lane, Boden & Kulkarni 2001),
- Orbital studies of close binary systems to determine precise stellar masses (see, e.g. Hummel *et al.* 2001; Boden & Lane 2001),
- Measurements of the brightness profiles across individual stellar disks to test models for stellar atmospheres (see, e.g. Burns *et al.* 1997; Wittkowski *et al.* 2001),
- Measurements of the angular diameter changes in variable stars to assess pulsational models for these stars (Young *et al.* 2000b) or to calibrate the cosmic distance scale (see, e.g. Lane *et al.* 2000),
- Imaging studies to investigate the features on or near evolved stars, which are expected to have strong internal convection (Young *et al.* 2000a).

In most of these studies, the confrontation of theory with experimental data has taken place for the first time. It is this new approach to stellar astrophysics, allowing the direct and precise testing of hypotheses, that makes the future of interferometry so exciting.

4. Where should we go in the future?

Forecasting the future directions that astronomical interferometry will take may seem a dangerously unreliable activity for a subject in which advances in technology have determined progress in the field so completely. But the physical constraints set by the properties of the astronomical sources and of the Earth's atmosphere through which they must be observed, provide a clear indication of the possibilities for ground-based interferometers in the future. Space-based interferometers will not be subject to the same atmospheric restrictions, and, in principle, can realize unlimited resolution and sensitivity. However, their success will depend on solving a different set of technical problems, in particular those associated with controlling an array of free-floating spacecraft. The timescale associated with these technical challenges is likely to extend well beyond 2010, and for this reason we will hereafter consider only ground-based developments over the next ten or so years.

There are three main areas in which improvements to the capabilities of existing ground-based synthesis arrays would allow them to become as scientifically valuable as their radio counterparts. These are the angular resolution attainable, the level of imaging fidelity and image quality that can be realized, and the limiting sensitivity. We will assume here that, as for facilities at most other wavelengths, the broad priority for any future arrays will be to provide detailed images of complex astrophysical phenomena. Our discussion thus excludes interferometers designed, for example, for wide-angle astrometry and extra-solar planet detection (Marcy & Butler 1998; Woolf & Angel 1998), whose merit we acknowledge but whose application will be more specialized.

(*a*) *Angular resolution*

The most immediate scientific questions which evidently require high angular resolution measurements for their solution are those concerning stars. What are their sizes, the structure of their atmospheres, the existence and nature of features on their surfaces, and the processes involved in their formation? How do they lose mass from their surfaces in the late stages of their evolution, how do they interact with close companion stars, what are the mechanism and dynamics of jets, the structures of protoplanetary disks, and the distribution and properties of planetary systems? Finally, how do they end their lives, and how do their compact remnants – or their large scale equivalents, i.e. massive black holes in the centres of active galactic nuclei (AGN) – interact with their surroundings?

The angular scales associated with a sample of these phenomena are summarized in table 2. For some of these the scales on which detailed imaging would be of interest can be inferred on the basis of straightforward physical arguments: for example, knowledge of the sublimation temperature of dust gives an estimate of the minimum distance from a stellar surface at which it might exist. In other cases, we currently have only rudimentary estimates of the appropriate physical scales. This is of particular concern for active galactic nuclei where the evidence we have concerning the location of the clouds responsible for their Doppler-broadened Balmer lines or the sizes of their dust tori is indirect (see, for example, Capetti *et al.* 2000; Peterson 2001).

Notwithstanding these uncertainties, the data of table 2 argue persuasively for

Table 2. *Potential targets for interferometric study and appropriate angular scales*

(Characteristic angular scales, θ, associated with potential targets for optical/infrared interferometry. For each type of source the distance shown has been chosen to allow observations of more than simply the brightest prototypes at an appropriate wavelength. Length scales are given either in solar radii ($1\ R_\odot = 7 \times 10^8$ m), astronomical units ($1\text{AU} = 1.5 \times 10^{11}$ m) or parsecs ($1\ \text{pc} = 3.1 \times 10^{16}$ m). The final column gives the interferometer baseline required to resolve the target. All figures are approximate.)

Source	Linear size	Distance	θ/mas	λ	B/m
Main sequence stars	3.0 $R_\odot$	100 pc	0.1	0.6	900
Evolved stars	200 $R_\odot$	250 pc	4	1.0	55
Stellar dust shells	400 $R_\odot$	250 pc	7	2.2	60
Stellar gas shells	20 $R_\odot$	250 pc	0.4	0.6	330
Spectroscopic binaries	0.2 AU	500 pc	0.4	0.6	310
Close binaries	10 $R_\odot$	250 pc	0.2	0.6	660
Protoplanetary disks	1 AU	500 pc	2	2.2	225
Recurrent novae ejecta	10 AU	2 kpc	5	2.2	90
Globular cluster proper motions	1AU yr^{-1}	10 kpc	0.1 yr^{-1}	0.6	1250
AGN broad line regions	0.05 pc	20 Mpc	0.5	0.6	240
AGN dust tori	0.5 pc	20 Mpc	5	2.2	90

interferometer baselines of at least 500 m. This is, roughly speaking, the minimum baseline for which the majority of these types of targets would be just resolvable, and in practice the preferred option would be to realize kilometric baselines. The technical difficulties associated with delivering a 1000 m baseline are not significantly greater than those already overcome in the current generation of arrays, and so as far as technical readiness is concerned this represents only a modest challenge.

(*b*) *Imaging fidelity*

The general principles governing the type and quality of imaging achievable with any type of synthesis array have been documented many times by radio astronomers (see, e.g. Thompson, Moran & Swenson 1991). The number of measured visibility data should be at least as great as the number of filled pixels desired in the reconstructed image, and the range of angular scales required in the map must be matched with an appropriate distribution of visibility samples. For a 10×10 pixel image, which seems to be a useful step beyond the capabilities of contemporary arrays (see figure 4), this implies a ratio of maximum to minimum baseline of ~ 10.

This could be realized by a sequence of measurements obtained with multiple re-configurations of an array with as few as three collectors. But the need to stabilize the interferometric output of optical and infrared arrays by controlling in real-time the OPD errors induced by the atmospheric perturbations introduces an additional complication. Sensing of the atmospheric perturbations relies upon the instantaneous signal-to-noise of the fringes being high since the fringes themselves provide the feedback signals. At low light levels, where Poisson statistics of the detected photons provides the main limitation to fringe detection, the signal-to-noise ratio for fringe measurements scales as $V\sqrt{N_{phot}}$, where V is the fringe visibility and N_{phot} the number of detected photons. Thus the measurement of the

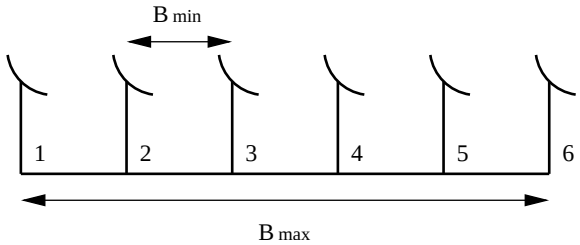

Figure 5. A schematic view of an array configuration that allows "phase bootstrapping". Simultaneous measurements of the fringes on all the short baselines (B_{min}) between neighbouring telescopes allows control of the atmospheric fluctuations and permits useful measurements to be made on the longest baseline (B_{max}). For this approach to work, the source of interest should not be significantly resolved on baselines of length B_{min}.

atmospheric OPD becomes increasingly difficult as the baselines of interest start to resolve the source. Unfortunately, this is precisely the situation that obtains when measurements for a map with many pixels across a source are being secured.

A favoured solution to this problem has been to use "phase bootstrapping", i.e. the simultaneous monitoring of the fringes on, for example, short baselines 1-2, 2-3, and 3-4, to assess what correction needs to be made on the baseline 1-4 (Armstrong *et al.* 1998, figure 5). In the context of the discussion here this has a supremely important consequence: the number of telescopes required to reach a given maximum baseline is significantly increased. For a "Y"-shaped configuration with equispaced telescopes along each arm and observing a source at the zenith, 18 telescopes are required to give a ratio of 10 : 1 between the longest and shortest baselines. Earth rotation allows for some leeway here, but the broad conclusion is clear: if future optical/infrared interferometers are to exploit bootstrapping and to deliver high quality imaging, then arrays of between 15 to 20 telescopes will be essential.

Another approach to interferometer stabilisation that has received attention is so-called "phase referencing". This uses measurements of an unresolved off-axis source at the same time as the science target so as to monitor the atmosphere simultaneously. It has been successfully demonstrated at the Palomar Testbed Interferometer (Colavita *et al.* 1999). However, as well as requiring additional optical and metrology hardware, it relies upon there being a suitable bright reference source within the isoplanatic patch. At 2.2 μm the angular radius of this is of order $20''$ and so for a 2 m-sized collector the sky coverage predicted is only ~ 0.1 percent. The suggestion that phase-referencing offers the better prospect for any general-purpose array rests squarely on this issue. There are a small number of specialized programmes that will be guaranteed a convenient reference source, in particular those that are focused on searching for companions and structures close to stars in the immediate solar neighbourhood. But these stars form a small fraction of those sources that are in principle amenable to interferometric scrutiny, and the trade off between the increased complexity and cost of phase referencing and the actual amount of new science it will deliver appears not to favour this approach.

If the above arguments are accepted and the requirements stated are met, how faithful an image can be delivered? For bright sources, the limit on the dynamic range in the image will be determined by the fractional error $\delta A/A$ and the phase

error $\delta\phi$, on the individual visibility amplitudes and phases, and their total number, n. As a rough guide, the dynamic range will be given by

$$\frac{\sqrt{n}}{\sqrt{\left[(\delta A/A)^2 + (\delta\phi)^2\right]}}\,. \tag{4.1}$$

Whilst errors on the visibility amplitudes of less than 1 per cent are certainly achievable (see, e.g. Perrin *et al.* 1998), phase errors may well be as large as 2 degrees, i.e. 3 percent of a radian, and hence be the dominant factor. For the 15 to 20-element arrays discussed above, the number of separate visibilities may readily be at least 1000; 10000 is perhaps realistic. These figures give a dynamic range of approximately 3000 : 1.

(c) *Sensitivity*

Perhaps the most demanding problem facing optical/infrared array designers today is how to achieve adequate sensitivity. Discussion of this issue is complicated for two main reasons. First, because the conventional coupling between resolution and sensitivity does not apply for a separated-element array, and second, because the most important considerations are those concerned with the instantaneous fringe signal, which is needed to control the array so that the data can be coherently integrated.

Since many of the scientific programmes proposed for future arrays concern stars and their immediate environments, is is useful to consider a "toy" model to clarify the key issues in the argument. We can model an isolated star as a black body of temperature T with an angular diameter of, say, 1 milliarcsecond. Figure 6 shows the apparent brightness of such a source as a function of observing wavelength for a range of temperatures. The brightness is plotted using the historical scale of magnitudes (m) where $m = -2.5\log(\text{intensity}) + \text{constant}$. The merit of this plot is that stars visible to the naked eye span a range in magnitude from $m \simeq -1$ for the brightest stars to $m \simeq +6$ for those at the limit of vision. So, the conclusion to be drawn from figure 6 is that essentially all stars larger than 1 milliarcsecond will be "bright". Conversely, all stars fainter than $m = +6$, and that can be adequately modelled as black bodies, will be smaller in diameter than 1 milliarcsecond.

A more quantitative criterion for judging what is "bright" is the photon flux necessary for interferometry to be successful. This limiting sensitivity will be determined by the number of detected photons, N_{phot}, necessary to give an adequate signal-to-noise ratio in a measurement, i.e. by the combination of aperture size and integration time which gives the best *instantaneous* signal-to-noise (S/N) ratio for detecting interference fringes. For telescopes with fast auto-guiders that correct the wavefront tilts of the incoming beams, the optimum aperture size and integration time have been computed by Buscher (1988). For atmospheric conditions described by a spatial scale r_0 and a coherence time t_0, the optimum aperture size and exposure time will be $2.75r_0$ and $2t_0$ respectively, and if observations are made using these values the measured fringe visibility, V, for an unresolved source will be 0.47. To enable measurements of the complex visibility to be integrated coherently, a value of S/N† $= \frac{1}{2}V\sqrt{N_{phot}}$ greater than about 2.5 is required. Hence the mini-

† The factor of $\frac{1}{2}$ pre-multiplying the visibility here assumes a simple two-beam combiner.

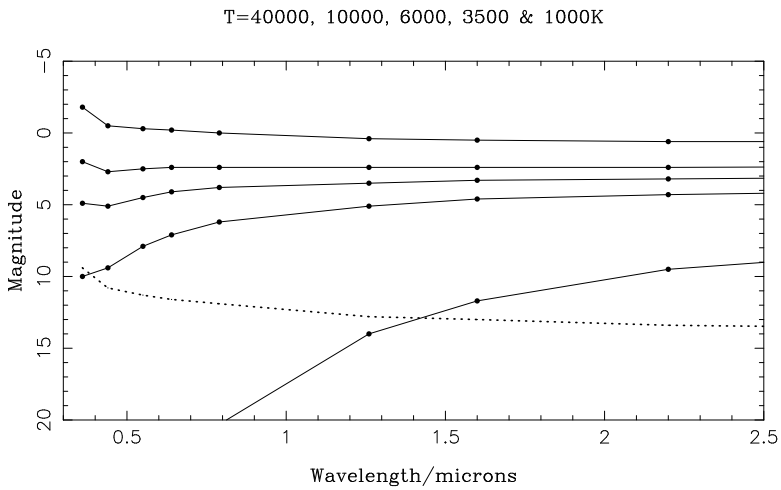

Figure 6. The apparent brightness of a 1 milliarcsecond black body at temperatures of 40000, 10000, 6000, 3500 and 1000 K. The units of brightness are magnitudes where $m = -2.5\log(\text{intensity}) + \text{constant}$. The black body temperature and apparent brightness decrease from top to bottom. The surface temperatures of most stars lie in the range 2000 to 50000 K. The dotted line at the bottom of the figure shows the limiting sensitivity for a shot-noise limited array using $2.75r_0$-sized telescopes and a measurement time of $2t_0$ (see text for details).

mum number of detected photons per short-exposure ($2t_0$) integration must be at least ~ 100.

To compare this to our stellar model some additional data are required. At good observatory sites r_0 and t_0 have median values of 15 cm and 5 ms at 500 nm (although published measurements of t_0 are sparse for many sites) and both scale as $\lambda^{6/5}$. The dotted line in figure 6 shows the resulting limiting sensitivity as a function of wavelength for a conservative value of the optical bandpass used for fringe detection of 50 nm and an overall optical throughput from source to detected output of 10 percent, and where we have assumed the availability of high efficiency array detectors that are limited solely by the shot-noise associated with the incoming signal. These have recently become available for wavelengths shorter than 1 μm (Mackay *et al.* 2001), but for the range 1 to 2.5 μm the readout noise of detectors currently under development implies limiting sensitivities about one magnitude poorer than those shown in the figure. The simple deduction to be made from figure 6 is that features much smaller than the star, perhaps as small as 0.1×0.1 mas, should be detectable.

Whilst some features of interest may exist on scales smaller than the stellar diameter, many others are expected to inhabit the circumstellar region, extending to many stellar radii. Their detectability will thus depend on whether they are heated or illuminated sufficiently by the star. Several arguments suggest that matter beyond ~ 100 stellar radii will be extremely difficult to detect interferometrically. Matter in radiative equilibrium at a distance R from a star of radius R_* and temperature T_* would be expected to have a temperature T such that $(T/T_*)^4 \simeq (R/R_*)^{-2}$. At $100R_*$, this will be a factor of 10 below that of the stellar surface, and so for the normal range of stellar temperatures (2000 to 40000 K) matter at larger radii will

be too cool to be detectable at visible or near infrared wavelengths. If, on the other hand, the circumstellar material simply reflects the stellar radiation, its surface brightness relative to the central star will be $\alpha(R/R_*)^{-2}$, where α is an albedo, which we can take to be of order 0.1. So, at $100R_*$ the surface brightness of any solid material will be only 10^{-5} that of the star. The problem of detecting and imaging these types of features is thus not so much one of sensitivity, but one associated with the maximum dynamic range that can be achieved. Dynamic ranges of 10^5 have been reached, albeit with some difficulty, for the very brightest sources at radio wavelengths, but it seems unlikely they will be realized at optical and infrared wavelengths in the near term.

Other mechanisms do exist for maintaining material at high temperatures at large distances from a star. Hot gas can be ejected sufficiently rapidly that there may be no appreciable cooling in the transit time. Alternatively, material at large radii may be heated in-situ through collisions with fast moving ejecta. However, for both of these situations the optical depth of the hot material is expected to be extremely small. Successful imaging is thus likely to hinge again on the maximum dynamic range that can be achieved.

The observational consequences of this discussion can be summarized as follows:

- For stellar systems it is likely that the central star will dominate the emission as long as it remains unobscured,
- For bright stars, i.e. those with angular sizes greater than 1 mas and hence brighter than $m = +6$, features should be detectable both on the stellar surface and out to, say, 100 mas,
- Stars ten times further away, i.e. with magnitudes as faint as $m = +11$, would remain detectable, as would features extending to 10 mas,
- Below the limiting sensitivity shown in figure 6, e.g. for $m \simeq +16$, stars would have dimensions of at most 10-microarcseconds and any surrounding structures that might otherwise be observable would not be expected to subtend an angle of greater than 1 mas,
- For an interferometer with a maximum baseline of $200\left[\frac{\lambda}{\mu\mathrm{m}}\right]$ m, corresponding to a resolution of 1 mas, all resolvable stellar systems should lie within the detection limits.

It is less straightforward to apply the type of argument presented above to active galactic nuclei (AGN). Here the physical conditions and geometries of the features of interest, in particular the Balmer line emitting broad-line region (BLR) and the nuclear dust tori, are less well characterized. However, there appears to be broad agreement that on sub-arcsecond scales the key question will be how much flux originates from a truly compact component, i.e. one located within some number of Schwarzschild radii from the central black hole ($R_s = 2GM/c^2 = 0.02$ to 20 AU for black hole masses in the range 10^6 to $10^9\,\mathrm{M}_\odot$). Any emission from this region will require micro-arcsecond resolution to resolve it, and so it is likely that high resolution images of AGN will be characterized by an unresolved core together with additional emission extended on scales of milliarcseconds. The very brightest AGN have core magnitudes, measured within a few arcseconds, of ~ 10 at $2.2\,\mu$m.

If ten percent of this emission comes from an unresolved component, then even these targets should be detectable with a $2.75r_0$-sized aperture array and the goal of imaging their broad-line regions and circumnuclear dust will become a realistic possibility.

The discussions above suggest that adequate sensitivity for most, if not all, the most exciting programmes for future ground based imaging interferometry can be achieved with relatively modest sized, i.e. $\sim 3r_0$ diameter, collectors. What then is the role for interferometers using larger aperture sizes ($D \gg 2\,\mathrm{m}$) than this? The potential advantage of using larger apertures is, in fact, not to enable imaging of much fainter sources, but rather to reduce the observing time needed for collecting data to achieve some given signal-to-noise ratio. This will be an issue, for example, when observing at very high spectral resolution in narrow atomic emission lines. This interesting conclusion arises because any benefits to be gained from a larger aperture can only be realized if the variance of the atmospheric phase fluctuations across it is no greater than that across an autoguided $3r_0$ diameter optic. In practice the limiting sensitivity for the adaptive optics (AO) correction required will be very similar to that discussed for fringe detection.

To illustrate this we can consider a simple example where the AO sensing is carried out at one wavelength and the interferometer fringes are to be sensed in an immediately adjacent band. The adaptive correction can thus be performed over a number of separate $3r_0$-sized patches of the larger mirror. There are two principal factors which distinguish the AO signal-to-noise ratio calculation from that for fringe detection. These are the usable optical bandwidth, which may easily be as large as 300 nm (c.f. 50 nm for fringe sensing), and the optimum exposure time, which is $1.5t_0$ (c.f. $2t_0$ for fringe sensing). Other differences, such as the throughput and collecting area will be identical. So, the sensitivity for the AO correction of each of the $3r_0$-sized sub-pupils of the larger aperture will be better by a factor of roughly $\left[\frac{300}{50}\right]\left[\frac{1.5}{2}\right] = 2.25$. However, this calculation does not provide for correcting the path delays between the separate sub-pupils and this must reduce the $\times 2.25$ advantage that the AO system has relative to direct fringe detection. The upshot of this is quite clear: adaptively corrected larger telescopes may well give a small improvement in the limiting magnitude for fringe detection, but unless laser guide star systems are introduced at each array element, this increase in sensitivity is likely to be no greater than a magnitude.

On the other hand, for relatively bright targets, i.e. where the instantaneous fringe signal-to-noise ratio $\gg 1$, successful adaptive correction should be relatively straightforward to achieve, and so the photon rate delivered to the beam combiner will increase as D^2. It is this advantage that large apertures ($D > 3r_0$) and natural guide star AO systems provide: the ability to collect photons faster and improve the overall observing efficiency of the array, rather than any major enhancement in limiting sensitivity.

5. Conclusions

The arguments presented above, and their implications for the future development of ground-based optical interferometry, can perhaps best be summarized by a sketching a picture of the type of facility array that we feel offers the most exciting prospects for the next decade. While the specific details of its implementation

are beyond the scope of this communication, three key aspects of its design seem secure. There should be:

- A large number (between 15 and 25) of 2 m-class telescopes, each with modest adaptive correction and distributed so as to give excellent snapshot Fourier plane coverage.
- Baselines extending to at least 1000 m, but with the opportunity for compact array configurations for sources subtending large angles on the sky, i.e. many tens of milliarcseconds.
- The presence of separate correlators optimized for real-time fringe stabilization and science imaging, but running in parallel and using different wavelength bands.

The fact that not one of the arrays currently operational or under construction (table 1) is likely to deliver on any of these requirements is both disappointing and invigorating. The type of array we have identified as being necessary is ambitious, but it quite clearly represents only a modest technical challenge beyond what has already been demonstrated. Simply put, the principal difficulties in delivering it are likely to be those of convincing non-interferometric astronomers of its moderate technical risk and subsequently in integrating its large number of interdependent parts.

What is undoubtedly clear, though, is that such an array would provide a unique new window on the universe, a massive stimulus to galactic and extragalactic astrophysics, and would be a fitting outcome to Thomas Youngs' prescient experiments.

The authors would like to thank all their colleagues within the COAST group for thought provoking contributions, and J. Monnier, A. George, and T. Scott for providing figures in advance of publication.

References

Airy, G.B. 1835 On the diffraction of an object-glass with circular aperture. *Trans. Camb. Phil. Soc.*, **5**, 283–291.

Armstrong, J.T., Mozurkewich, D., Pauls, T.A. & Hajian, A.R. 1998 Bootstrapping the NPOI: keeping long baselines in phase by tracking fringes on short baselines. *Proc. SPIE*, **3350**, 461–466.

Baldwin, J.E., *et al.* 1996 The first images from an optical aperture synthesis array: mapping of Capella with COAST at two epochs. *Astron. Astrophys.*, **306**, L13–L16.

Boden, A.F. & Lane, B.F. 2001 A preliminary visual orbit of BY Draconis. *Astrophys. J.*, **547**, 1071–1076.

Burns, D. *et al.* 1997 The surface structure and limb-darkening profile of Betelgeuse. *Mon. Not. R. ast. Soc.*, **290**, L11–L16.

Buscher, D.F. 1988 Optimizing a ground-based optical interferometer for sensitivity at low light levels. *Mon. Not. R. ast. Soc.*, **235**, 1203–1226.

Capetti, A., Schreier, E.J., Axon, D., Young, S., Hough, J.H., Clark, S., Marconi, A., Macchetto, D. & Packham, C. 2000 Hubble Space Telescope infrared imaging polarimetry of Centaurus A: implications for the unified scheme and the existence of a misdirected BL Lacertae nucleus. *Astrophys. J.*, **544**, 269–276.

Colavita, M.M. *et al.* 1999 The Palomar Testbed Interferometer. *Astrophys. J.*, **510**, 505–521.

Fizeau, H. 1868 Prix Bordin: rapport sur le concours de l'annee 1867. *Comptes Rendus Hebdomadaires des Séances de l'Academie des Sciences*, **66**, 932–934

Dawes, Rev. W.R. 1866 A catalogue of micrometrical measurements of double stars. *Mem. R. ast. Soc.*, **35**, 137–502.

Haniff, C.A., Mackay, C.D., Titterington, D.J., Sivia, D. & Baldwin, J.E. 1987 The first images from optical aperture synthesis. *Nature*, **328**, 694–696.

Herschel, W. 1805 Experiments for ascertaining how far telescopes will enable us to determine the real from the spurious diameters of celestial and terrestrial objects: with an application of the result of these experiments to a series of observations on the nature and magnitude of Mr.Harding's lately discovered star. *Phil. Trans. Roy. Soc.*, **95**, 31–64.

Horton, A.J., Haniff, C.A. & Buscher, D.F. 2001 Diffraction losses in ground-based optical interferometers. *Mon. Not. R. ast. Soc.*, **327**, 217–226.

Hummel, C.A., Carquillat, J.-M., Ginestet, N., Griffin, R.F., Boden, A.F., Hajian, A.R., Mozurkewich, D. & Nordgren, T.E. 2001 Orbital and stellar parameters of omicron Leonis from spectroscopy and interferometry. *Astron. J.*, **121**, 1623–1635.

Jennison, R.C. 1958 A phase sensitive interferometer technique for the measurement of the Fourier transforms of spatial brightness distributions of small angular extent. *Mon. Not. R. ast. Soc.*, **118**, 276–284.

Labeyrie, A. 1975 Interference fringes obtained on Vega with two optical telescopes. *Astrophys. J.*, **196**, L71–L75.

Lane, B.F., Kuchner, M.J., Boden, A.F., Creech-Eakman, M. & Kulkarni, S.R. 2000 Direct detection of pulsations of the Cepheid star zeta Gem and an independent calibration of the period-luminosity relation. *Nature*, **407**, 485–487.

Lane, B.F., Boden, A.F. & Kulkarni, S.R. 2001 Interferometric Measurement of the Angular Sizes of Dwarf Stars in the Spectral Range K3–M4. *Astrophys. J.*, **551**, L81–L83.

Mackay, C.D., Tubbs, R.N., Bell, R., Burt, D, & Moody, I. 2001 Sub-electron read noise at MHz pixel rates. *Proc. SPIE*, **4306**, 289–298.

Marcy, G.W. & Butler, R.P. 1998 Detection of extrasolar giant planets. *Ann. Rev. Astron. Astrophys.*, **36**, 57–98.

Michelson, A.A. 1890 On the application of interference methods to astronomical measurements. *Phil. Mag.*, **30**, 1–21.

Michelson, A.A. 1891 Measurement of Jupiter's satellites by interference. *Pub. Astr. Soc. Pac.*, **3**, 274–278.

Michelson, A.A. & Pease, F.G. 1921 Measurement of the diameter of alpha Orionis with the interferometer. *Astrophys. J.*, **53**, 249–259.

Nordgren, T.E., *et al.* 1999 Stellar angular diameters of late-type giants and supergiants measured with the Navy Prototype Optical Interferometer. *Astron. J.*, **118**, 3032–3038.

Pearson, T.J. & Readhead, A.C.S. 1984 Image formation by self-calibration in radio astronomy. *Ann. Rev. Astron. Astrophys.*, **22**, 97–130.

Pease, F.G. 1931 Interferometer methods in astronomy. *Ergebnisse de Exakten Naturwissenschaften*, **10**, 84–96.

Perrin, G., Coude du Foresto, V., Ridgway, S.T., Mariotti, J.-M., Traub, W.A., Carleton, N.P. & Lacasse, M.G. 1998 Extension of the effective temperature scale of giants to types later than M6. *Astron. Astrophys.*, **331**, 619–626.

Peterson, B.M. 2001 Variability of active galactic nuclei. In *Advanced Lectures on the Starburst-AGN Connection,* (eds I. Aretxaga, D. Knuth & R. Mujica) pp. 3–67, Singapore: World Scientific.

Quirrenbach, A. 2001 Optical Interferometry. *Ann. Rev. Astron. Astrophys.*, **39**, 353–401.

Accepted for publication in Phil. Trans. A

Rogers, A.E.E., *et al.* 1974 The structure of radio sources 3C 273B and 3C 84 deduced from the 'closure' phases and visibility amplitudes observed with three-element interferometers. *Ap. J.*, **193**, 293–301.

Thompson, A.R., Moran, J.M. & Swenson, G.W. Jr. 1991 *Interferometry and synthesis in radio astronomy*, 2nd edn. Malabar; Krieger Publishing Press.

Tuthill, P.G., Monnier, J.D., Danchi, W.C., Wishnow, E.H. & Haniff, C.A. 2000 Michelson interferometry with the Keck I Telescope. *Pub. Ast. Soc. Pac.*, **112**, 555–565.

Tuthill, P.G., Monnier, J.D. & Danchi, W.C. 2001 A dusty torus around the luminous young star LkH alpha 101. *Nature*, **409**, 1012–1014.

Wittkowski, M., Hummel, C.A., Johnston, K.J., Mozurkewich, D., Hajian, A.R. & White, N.M. 2001 Direct multi-wavelength limb-darkening measurements of three late-type giants with the Navy Prototype Optical Interferometer. *Astron. Astrophys.*, **377**, 981–993.

Woolf, N. & Angel, J.R.P. 1998 Astronomical searches for earth-like planets and signs of life. *Ann. Rev. Astron. Astrophys.*, **36**, 507–537.

Young, J.S., *et al.* 2000a New views of Betelgeuse: multi-wavelength surface imaging and implications for models of hotspot generation. *Mon. Not. R. ast. Soc.*, **315**, 635–645.

Young, J.S., Baldwin, J.E., Boysen, R.C., Haniff, C.A., Pearson, D., Rogers, J., St.-Jacques, D., Warner, P.J. & Wilson, D.M.A. 2000b Cyclic variations in the angular diameter of chi Cygni. *Mon. Not. R. ast. Soc.*, **318**, 381–386.

Young, T. 1804 Experiments and calculations relative to physical optics. *Phil. Trans. Roy. Soc.*, **94**, 1–16.

Accepted for publication in Phil. Trans. A