4

Interferometers in practice

Previous chapters have assumed a rather generalised and abstract interferometer. This chapter looks at the how the functionality of this abstract interferometer is implemented in reality. This exposition will make use of examples from existing interferometers, with the aim of giving an idea of the diversity and ingenuity of the implementations of this functionality.

4.1 Interferometric facilities

The following is a brief summary of the interferometric facilities which were operational at the time of writing or expected to be operational within the next few years. The systems are listed in order of the date (or expected date) of 'first fringes' on each of these interferometers. More information can be found in the online supplementary material (see Appendix B).

Aperture-masking instruments Masking the aperture of a single telescope to convert it into an interferometer was used in the very earliest days of interferometry, and yet it is still a competitive technique for many astronomical measurements (Tuthill, 2012). Because the implementation challenges for aperture masking are in some ways different to those for separated-element interferometry, discussion of the practical features of this technique is deferred until Section 4.10.

SUSI The Sydney University Stellar Interferometer (Davis *et al.*, 1999) is sited near to the radio telescopes of the Australia Array in Narrabri, Australia. It operates at visible wavelengths and has baselines ranging from 5 m to 640 m (currently only baselines up to 80 m have been commissioned).

NPOI The Navy Precision Optical Interferometer (formerly the Navy Prototype Optical Interferometer) (Armstrong *et al.*, 1998) is sited on the

Lowell Observatory Anderson Mesa Station in Arizona, USA. It operates at visible wavelengths and is capable of performing wide-angle astrometric measurements as well as interferometric imaging. It has baselines from 2 m to 437 m (at the time of writing only baselines from 8.8 m to 79 m have been commissioned).

CHARA array The Center for High Angular Resolution Astronomy array (ten Brummelaar *et al.*, 2005) is sited on Mt Wilson, California, USA. It operates at visible and near-infrared wavelengths, and has baselines ranging from 34 to 330 m.

VLTI The Very Large Telescope Interferometer (Schöller, 2007) on Mt Paranal in Chile consists of four 'unit telescopes', which act part-time as independent 'single-dish' telescopes for conventional astronomical observations and four 'auxilliary telescopes', which are used full-time for interferometry. A suite of interferometric instruments allow operation at near-infrared and mid-infrared wavelengths with baselines from 8 m to 200 m. Narrow-angle astrometric observations will be possible with the GRAVITY instrument.

LBT The Large Binocular Telescope is a pair of telescopes on a common mount with a centre-to-centre separation of 14 m, sited on Mt Graham, Arizona, USA. When the beams from the two telescopes are combined, interferometric baselines up to 22 m are available (Angel *et al.*, 1998). The LBT interferometric configuration is in many ways more similar to a single telescope with an unusual aperture than to an interferometer.

MROI The Magdalena Ridge Observatory Interferometer is under construction in New Mexico, USA (Buscher *et al.*, 2013). It will operate at visible and near-infrared wavelengths and will have baselines ranging from 8 m to 340 m.

The discussion will generally follow the progression through the elements of the beam train of a typical optical interferometer. An example beam train, that for the MROI, is shown in Figure 4.1.

4.2 Siting

Any optical telescope needs to be situated on a site with adequate clear night skies. In addition, the performance of an interferometer is strongly dependent on having good atmospheric seeing conditions, as explained in Chapter 3.

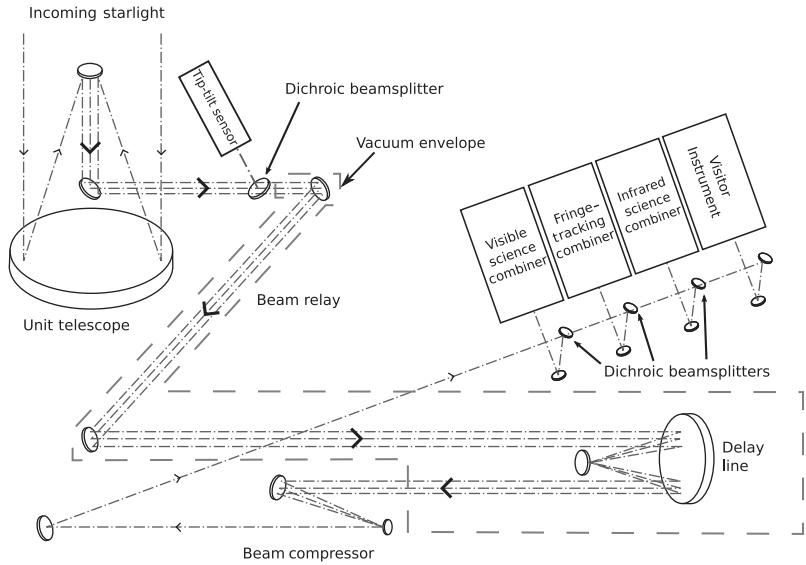

Figure 4.1 The optical beam train for one arm of the MROI.

The only way to completely get rid of the effects of seeing is to put the interferometer above the Earth's atmosphere, i. e. into space. Putting an interferometer into space provides its own formidable challenges, not least of which is cost: the likely cost of any interferometric space mission exceeds the cost of all the telescopes currently on the ground.

The alternative is to place the interferometer on a site with minimal seeing. The sites with the best seeing are typically on the tops of relatively sharply-peaked mountains, as this means that the wind over the site has not previously interacted strongly with the ground and therefore is less turbulent. This conflicts with another desirable characteristic for an interferometer, namely that it should be sited on a large flat area to allow for flexibility in positioning the telescopes. As a result the site chosen for most interferometers is a usually a compromise between these two desiderata.

Two examples of this compromise can be seen in the CHARA array and the NPOI. The CHARA array is situated on Mt Wilson, with excellent seeing, but the available baselines are restricted by the availability of suitable locations in the relatively hilly local topography. The NPOI is situated on the Anderson Mesa, which is quite flat, but although it is at a relatively high altitude the seeing is somewhat worse than at Mt Wilson.

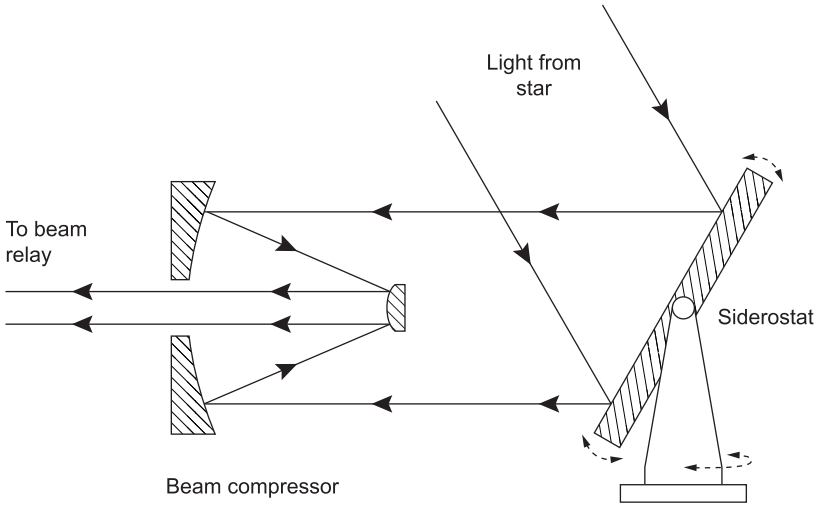

Figure 4.2 Schematic layout of a siderostat light collector.

4.3 Collectors

In radio interferometry, the interferometric system as a whole is often called a 'telescope' with the individual collectors being 'dishes' or 'antennae'. This causes some confusion at optical wavelengths where the collectors are themselves often telescopes, so the individual collectors are sometimes called 'unit telescopes' to distinguish them from the 'synthesis telescope' they form part of. In this book, the collectors will typically be called 'telescopes' and telescope arrays 'interferometers'.

The light-collecting apertures of the array serve to sample the wavefront of light from the target and inject the light into the rest of the system. The simplest collectors are 'siderostats', flat mirrors which can be rotated to reflect light from the selected target in a fixed direction. Siderostats are often followed by 'beam compressors' as shown in Figure 4.2. The beam compressors are fixed telescopes that serve to reduce the diameter of the starlight beam and hence reduce the size of the optics needed in the rest of the system.

An alternative is to use a conventional astronomical telescope to capture the light, and an arrangement of mirrors to send the light out in a fixed direction. Telescopes are preferred over siderostats for larger apertures, because the size of the siderostat mirror needs to be considerably larger than the required collecting area in order to take account of obliquity effects, and the

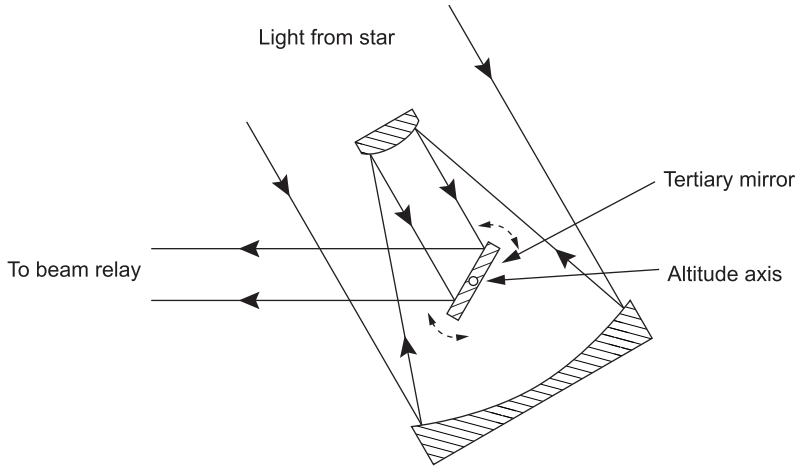

Figure 4.3 Schematic layout of an alt-alt telescope. The tertiary mirror rotates at half the angular rate of the primary-secondary pair in order to keep the output beam fixed.

cost and weight of such mirrors increases rapidly with size. However, using conventional telescopes has its own problems: the moving structure is less compact and so more prone to vibration.

One unusual geometry of telescope that is used in interferometry is the so-called *alt-alt* configuration as shown in Figure 4.3. This arrangement has a rotating tertiary mirror to allow the light to exit the telescope in a fixed direction after only three reflections, the same number of reflections as needed for a siderostat with a beam compressor (indeed, this arrangement can be considered as being equivalent to a beam compressor followed by a siderostat). A more conventional telescope design such as an *alt-az* design can be more compact, but requires seven or more reflections to achieve the same effect, leading to a loss in light and beam quality.

The diameter of the telescope needs to be large in order to collect as much light as possible, consistent with having adequate wavefront quality. Since the main determinant of wavefront quality is atmospheric seeing, the telescopes are typically a few r_0 in size if only tip–tilt correction is applied, but can be many r_0 for telescopes with higher-order AO.

The telescope aperture diameter tends to increase with increasing wavelength since r_0 scales as $\lambda^{6/5}$. This is illustrated by the collectors for SUSI and the VLTI shown in Figure 4.4: SUSI operates at visible wavelengths and has siderostats with an effective diameter of 14 cm, while the VLTI operates in the 1–10 μm range and so has telescopes with a diameter of 1.8 m.

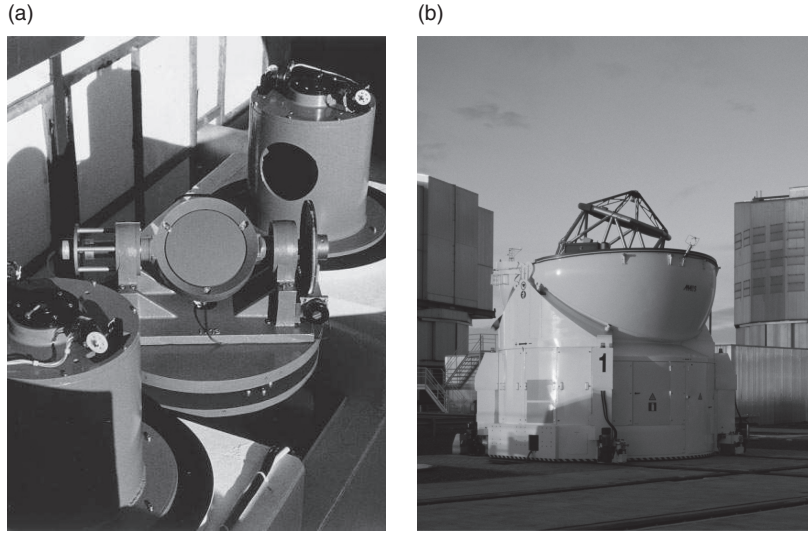

Figure 4.4 Collecting elements: the siderostats on SUSI (a, from Davis *et al.*, 1999) and the alt-az auxilliary telescopes of the VLTI (b, courtesy ESO).

4.3.1 Atmospheric dispersion correction

A number of other subsystems are usually sited at the telescopes. An atmospheric dispersion corrector can be used to make sure that beams at different wavelengths travel in the same direction: it corrects the chromatic refraction effects caused by the 'prism' of atmosphere in the stellar beam which occurs when the target is not directly overhead.

These effects on the tip and tilts of the beams are sometimes called 'angular dispersion' to distinguish them from the 'piston dispersion' or 'linear dispersion' effects caused by there being different amounts of air path between different telescopes and the beam combiner, as discussed in Section 1.8.

Because the atmospheric angular refraction at any given wavelength is the same at all telescopes, it is possible to do without an atmospheric dispersion corrector and still produce high-contrast fringes. The main negative effects of angular dispersion are to do with getting light at all wavelengths through the various apertures in the system. For example, the angular separation between the incoming beams at wavelengths of 500 nm and 800 nm from a target which is 45° from the zenith is about 0.6 arcseconds for a telescope located at an altitude of 2 km. This separation is then multiplied by the angular magnification of the unit telescope: the value of 10 for the VLTI is typical, so the separation of

the beams coming out of the telescope is perhaps 6 arcseconds. Over a 200 m path, this corresponds to a relative transverse displacement of the beams at the two wavelengths of about 6 mm, a substantial fraction of the 18-mm pupil size for the VLTI auxilliary telescopes.

Even more severe effects are apparent if spatial filtering is used, as the angular size of the pinholes or fibres used have typical angular sizes of a fraction of an arcsecond when projected on the sky, so if light at one wavelength is centred on the pinhole, light at another wavelength may not pass through at all.

4.3.2 Adaptive optics

All long-baseline interferometers have AO at least at the level of tip–tilt correction and the wavefront-correcting element is typically sited close to the telescope. This is because the largest aberrations must be corrected before they can have a significant effect on the propagation of the light towards the beam combiner. The most common example of this is that tip and tilt errors of many arcseconds (referred to the sky) can arise because of imperfect telescope tracking. If these were corrected at some considerable distance from the telescope then the starlight beams will likely miss any correcting element unless all of the intermediate optics were considerably oversized.

The wavefront-sensing element is often not sited at the telescope: this way aberrations further down the optical train can also be sensed and corrected. However, siting the wavefront sensor at the far end of the optical train from the telescope means that the field of view of the sensor is limited and less light reaches it because of losses in the rest of the beam train.

4.4 Beam relay

4.4.1 Free-space propagation

The signal collected by the unit telescopes must be transferred to a central point for processing and combining with the signal from other telescopes. In all the interferometers described above, this is done by propagating the light as collimated beams, using a set of flat mirrors to direct the beams along an appropriate path. The mirrors are polished to a high accuracy in order to minimise the wavefront distortion induced by the reflections, and have highly reflective coatings in order to minimise the loss of light. These coatings are usually based on metals such as aluminium, silver or gold, and can achieve reflectivities of 98% or more over a wide wavelength range.

4.4.2 Evacuated air paths

The path along which the beam propagates can be hundreds of metres in length, so the effects of air refraction along this path can be significant, causing chromatic dispersion and also seeing. To ameliorate some of the effects of refraction, the air path can be enclosed in ducts, as is done at the VLTI, but more often the beam propagates through evacuated tubes: this latter approach is used at the SUSI, NPOI, CHARA and MROI.

The level of vacuum used in the beam path need not be at the levels needed for cryostats or particle accelerators. These latter scenarios require pressures at the 10^{-8} bar level or lower, while pressures of order a millibar or so are sufficient to ameliorate the worst effects of air refraction in an interferometer. Reducing the air pressure inside the pipe to 1 mbar is equivalent to reducing 100 m of differential path to 10 cm of equivalent air path, and this is sufficient to reduce the differential dispersion effects discussed in Section 1.8 to tolerable levels.

As in normal air, atmospheric seeing can occur in vacuum pipes with residual air, if pockets of hot and cold air occur in the optical beam. We can consider an extreme example of a pipe 400 m long which is filled with random 'blobs' of air heated to about 20 °C above the ambient temperature. This could be due to some source of heat in the pipe, for example a motor operating a delay line.

If each blob is of order 20 cm in diameter and the pressure of the air in the pipe is 1 mbar, then the differential optical path introduced by one such blob of air is about 3.84 nm at optical wavelengths. If the pipe is 50% filled with such random blobs, then a beam of light will intercept about 1000 such blobs. Thus, the root-mean-square (RMS) optical pathlength perturbation will be of order $\sqrt{1000} \times 3.84 = 121$ nm, which is a substantial fraction of a wavelength.

However, the above scenario requires that the heated air blobs survive for long enough to fill most of the pipe. In fact, any deviations from temperature uniformity will decay away due to diffusion of heat from warmer regions to colder regions and the characteristic time constant for this decay will be small in a low-pressure environment. A sinusoidal temperature perturbation of wavelength λ decays with a time constant given by

$$\tau = \frac{1}{K}\left(\frac{\lambda}{2\pi}\right)^2,$$

where K is the thermal diffusivity given by

$$K = \frac{k}{C\rho},$$

with k being the conductivity, C being the specific heat capacity and ρ the density. The thermal diffusivity of air increases with decreasing air pressure because the conductivity remains roughly constant (providing the mean free path is smaller than the smallest dimensions being considered), and the specific heat remains roughly constant, but the density falls in proportion to the pressure. Substituting $\lambda = 40$ cm (i.e. a quasi-sinusoidal perturbation consisting of a 20-cm blob and a 20-cm region of unheated air) and a thermal difussivity appropriate for air at 1 mbar, we arrive at a time constant of around 200 ms. Thus, the heated air blobs are short-lived and unlikely to fill the pipe unless non-uniform sources of heating exist along the entire length of the pipe.

Polarisation

The reflections off mirrors in the beam train can induce changes in the state of polarisation of the beam on reflection. A quasi-monochromatic electromagnetic wave can be modeled as the superposition of two perpendicularly polarised waves denoted S and P; by convention, the electric field of the S vibration is parallel to the line of intersection of the propagating wavefront with the surface of an optical component. The amplitude and phase reflection coefficients for mirror coatings depend on whether the light is S or P polarised. Typically the phase effects are more pronounced. For example, the amplitude reflection coefficient for light at a wavelength of 630 nm incident at an angle of incidence of 45° on a bare silver mirror is about 0.5% greater for S-polarized light than for P-polarized light, and the P-polarized light is phase-retarded by about 160° when compared with the S-polarized light. It is clear from this that the polarization effects of the optical train can be significant.

A general interference pattern can be thought of as the superposition of the interference patterns seen in two orthogonal polarisations, and if these two patterns have different fringe phases, the fringe contrast will be reduced. If the interferometer is constructed so that the polarisation effects of all the beam trains are identical, then the interference patterns in both polarisations will have the same phase and the fringe contrast will be maximised. This condition can be achieved if every light beam experiences the same set of polarisation changes in the same sequence as every other light beam. In practice, this means that the direction cosines between the direction of the beam and the mirror normal must be same for the nth reflection that the beam from telescope m experiences as for the nth reflection that any other beam experiences (Traub, 1988).

This path symmetry condition has been adopted in all the interferometers described in Section 4.1, and leads to a preference for certain geometries in the beam transport system. The NPOI, CHARA and MROI interferometers

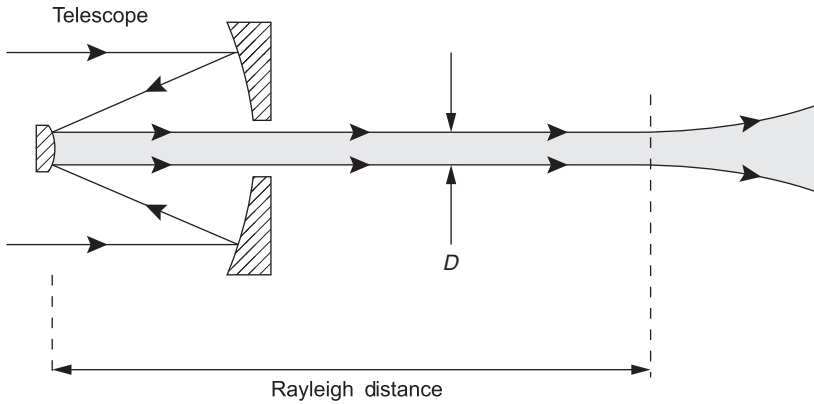

Figure 4.5 Diffraction of a beam propagating through the beam relay system.

have an equilateral 'Y'-shaped layout, so that the beams, after leaving the telescopes, experience reflections at 30° angles of incidence, while the beam transport within the VLTI has a rectangular geometry, with angles of incidence being generally 45°.

Diffraction

A circular parallel beam of light of diameter d and wavelength λ will propagate without significant diffraction effects over distances of up to

$$D_r = \pi d^2/4\lambda, \tag{4.1}$$

where D_r is known as the Rayleigh distance. At around this distance, Fresnel diffraction effects become important and the beam begins to spread, as shown in Figure 4.5. The eventual divergence angle of such a beam is given by the usual Fraunhofer diffraction angle $1.22\lambda/d$. For $\lambda = 1\mu$m, the Rayleigh distance for a 1-cm diameter beam is about 78 m, so diffraction effects are clearly important when propagating such beams over distances of hundreds of metres.

The effects of diffraction include both loss of light through spreading of the beam, and loss of fringe contrast because diffraction causes a mismatch in the phase and amplitude profiles of beams which have travelled different distances (Tango and Twiss, 1974). In order to keep these effects as small as possible, beam sizes of $d \gtrsim 10$ cm are typically used in optical interferometers. An alternative method of reducing the effects of diffraction is used in the VLTI, where a variable-curvature mirror is used to re-image the telescope pupil at the entrance to the beam-combining area (Ferrari *et al.*, 2000).

Diffraction occurs because of spatial non-uniformities in the wavefront. A finite beam diameter is a non-uniformity in the wavefront amplitude; wavefront distortions induced by atmospheric seeing can act as a non-uniformity in the wavefront phase. If the collectors are significantly larger than r_0 and have only partial adaptive correction, then the typical spatial scale of the seeing-induced phase non-uniformities will be smaller than that of the amplitude non-uniformity due to the beam size. Thus, the effective Rayleigh distance for these distortions will be less than for the aperture as a whole. Diffraction will cause the wavefront phase distortions to turn into amplitude distortions, and so a beam with an initially uniform intensity profile can take on a randomly changing 'speckled' appearance. If this beam then goes through an aperture which is smaller than the extent of the diffracted pattern, the total amount of light received will fluctuate randomly, which could cause problems in performing a photometric calibration of the system.

Although the effects of beam diffraction are mostly negative, they can be used to advantage: because any non-uniformities in the wavefront will tend to diffract out of the main beam, by choosing an appropriate size for the receiving aperture the wavefront distortions can be preferentially discarded while keeping the undistorted part of the beam; this is a form of spatial filtering (Horton *et al.*, 2001), which does not require any additional optics.

4.4.3 Fibre beam transport

In radio interferometry the signal from the collectors is sent to the central correlator via wires or waveguides. The equivalent for optical interferometry is the use of optical fibres.

The fibres used for beam transport are single-mode optical fibres, for different reasons than those leading to their use in the context of spatial filtering as discussed in Section 3.7. Multi-mode fibres cannot be used for beam transport as different modes have different propagation speeds and so modes which are initially in phase will end up out of phase after propagation in the fibre. This effectively corrupts the optical wavefront and so can destroy the quality of an interference pattern.

Single-mode optical fibres have benefited from many years of development for telecommunication systems, and fibres with very low losses allow the transmission of optical signals over many kilometres. Compared to free-space propagation, fibres offer the possibility to simplify the 'plumbing' of interferometers, as fibres are far narrower and much more flexible than vacuum tubes acting as 'light pipes'.

Using fibres to get the light out from a moving telescope to a fixed beam-transport system can potentially save many reflections of the beam, and using fibres instead of systems of vacuum pipes between the telescopes and the beam-combining laboratory offers the possibility to reconfigure the array rapidly – literally a 'plug-and-play' system.

Fibres have been used successfully for beam transport over hundreds of metres in interferometric experiments (Perrin *et al.*, 2006b). However, none of the interferometers mentioned in Section 4.1 use fibres for transport of beams over distances of more than a few metres (for example within beam combiners – see Section 4.7.3), and there are a number of technical issues to be overcome before the use of fibres for beam transport becomes widespread.

One issue is to do with the wide waveband range typically used in astronomical interferometers. The transmission of silica optical fibres is very good at the near-infrared telecommunications wavelengths, but away from these wavelengths the transmission falls off.

To have a throughput of at least 50% over a 300-m length of fibre, the fibre needs to have losses lower than 0.01 dB/m, or equivelently 10 dB/km. Low-OH silica fibres can achieve this figure over a wavelength region from about 600 nm to 1900 nm but losses rise rapidly at shorter and longer wavelengths, so that the throughput of a 300-m length of fibre at a wavelength of 2200 nm is less than 5%. Fibres made from flouride and chalcogenide glasses can have improved performance in the near- and mid-infrared region, but have poorer performance at shorter wavelengths.

Another issue with the step-index fibres typically used in telecommunications is that they only operate efficiently as single-mode waveguides over a restricted range of wavelengths. At wavelengths shorter than the 'cut-off' wavelength characteristic for the fibre, the fibre becomes multi-mode, and at wavelengths more than about 1.35 times the cut-off, losses due to any bends of the fibres become prohibitive.

This problem has to some extent been overcome with *photonic-crystal* fibres, whose structure consists of a regular grid of air holes running along the length of the fibre (Birks *et al.*, 1997). These fibres can operate in single mode over an octave (i. e. factor of two) in wavelength, but they are a relatively recent development and have not yet been deployed in an interferometer.

As a result, to equip a broadband interferometer with a fibre-based beam transport system will likely require multiple fibres to accomodate the different bandpasses. This introduces layers of complexity not seen in more broadband optical systems. For example, it requires an optical system to split light at the different wavelengths at the telescopes, and means that a fringe tracker using

the light in a bandpass going through one fibre will not take out any delay errors seen at wavelengths going through other fibres.

A second problem is that the intrinsic chromatic dispersion of the glass from which the fibres are made poses a problem when using the relatively wide bandwidths used to observe astronomical sources. This dispersion will cause a reduction in fringe contrast when the dispersion in the two arms of the interferometer being interfered is not equal. Nevertheless, a delay line introducing up to 2 m of differential delay has been built using fibres (Simohamed and Reynaud, 1997).

Finally, optical fibres are birefringent and so introduce large polarisation effects, which can reduce the fringe contrast by large factors. Static compensation of polarisation imbalances is possible for fibres in a thermally controlled environment but the birefringence depends on temperature and the level mechanical strain on the fibres caused by bends and twists, so dynamic control of the polarisation may be necessary if the fibres are outdoors.

4.5 Array layout

An array of M telescopes can simultaneously measure visibility information on $M(M-1)/2$ different baselines. The question of how best to arrange these telescopes to achieve the best imaging performance has been the subject of many studies, particularly at radio wavelengths. These studies have resulted in the development of a number of 'rules of thumb' as to the type of (u, v) coverage which is generally to be preferred, although this to some extent depends on the target being observed:

1. If a target has a certain overall size $\theta_{\max}$ and contains details of interest separated by $\theta_{\min}$ then the arguments in Section 2.3 suggest that the (u, v) plane needs to be sampled at a spacing of about $\Delta u \sim 1/\theta_{\max}$ up to a maximum (u, v) coordinate of magnitude $1/\theta_{\min}$.
2. Multiple (u, v)-plane samples spaced by much less than $1/\theta_{\max}$ are less useful than the equivalent number of samples positioned to 'fill in' less-well-sampled areas of the (u, v) plane. Samples spaced by $1/\theta_{\max}$ or more are termed 'independent (u, v) samples' whereas arrays that sample the same (u, v) coordinate multiple times are termed 'redundant arrays'.
3. The above rules can be relaxed for targets which are known to be 'sparse' in the sense that an image of the target at the desired resolution has relatively few resolution elements which contain flux. For example, a binary

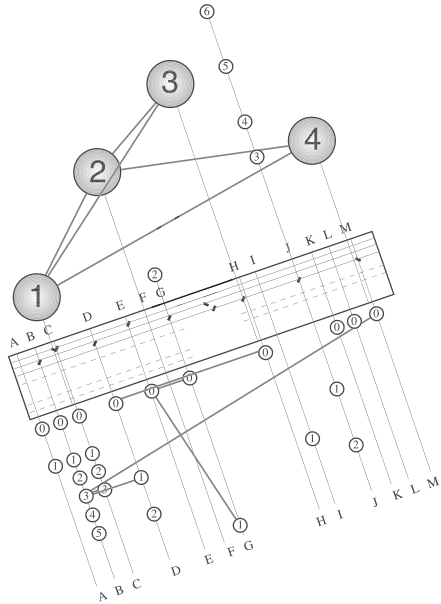

Figure 4.6 Array layout for the VLTI. The layout includes both the 8-m-diameter unit telescopes (shown as large circles) and the 30 pads for the 1.8-m-diameter auxiliary telescopes (shown as smaller circles). The baselines that were used for calibrator observations by Richichi and Percheron (2005) are overlaid on the diagram.

star system imaged at a resolution which allows the two stars to be distinguished but does not significantly resolve the stars themselves will have two 'filled pixels' in the image. In the case of a sparse target, image reconstruction can usually be accomplished when the number of independent (u, v) points sampled approximately equals or exceeds the number of filled pixels.

Despite the large number of studies, no single arrangement of telescopes has been found to be best in all cases and there are a large number of geometries to choose from which give adequate (u, v) coverage. The final choice of telescope layout tends to be made based on other considerations, for example the beam transport system symmetry concerns outlined in Section 4.4 or fitting into the local topography. Two such layouts are shown in Figure 4.6 and Figure 4.7, which are based on two different beam-relay geometries, one rectangular and one 'Y'-shaped.

If the telescopes in the array are relocatable, then the (u, v) coverage can be increased by rearranging the telescopes into a number of different arrays.

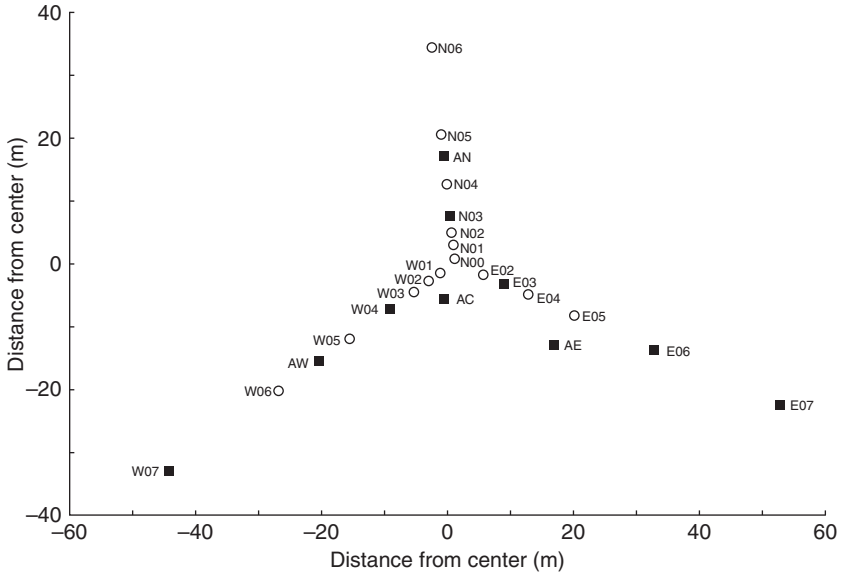

Figure 4.7 Array layout for the inner portion of the NPOI. The siderostat stations active in the late-2014 observing season are shown as filled squares. The fixed 'astrometric' stations have labels beginning with 'A', while the 'imaging' stations, which can take movable siderostats, have labels beginning with 'N', 'E' or 'W', depending on which arm of the array they are on. The outer imaging stations of the array are not shown: these extend to baselines of up to 450 m (image courtesy Don Hutter).

This is done for example with the auxilliary telescopes (ATs) on the VLTI, where there are 30 different stations to place the four ATs on. A number of different 'quadruplets' of AT stations are offered in any given semester. Arrays such as the NPOI and MROI offer multiple array configurations, which are approximately scaled versions of each other. These allow a given level of imaging performance to be matched to similarly-shaped targets of different angular sizes.

4.6 Delay lines

Delay lines are also called 'path compensators' to indicate that their function is to compensate for any imbalance in the delays introduced elsewhere in the paths from the target to the beam combiner. The amount of differential delay that needs to be introduced can be separated into a static component and a dynamic component. The static component is typically dominated by the differential paths introduced by the beam transport optics. The dynamic

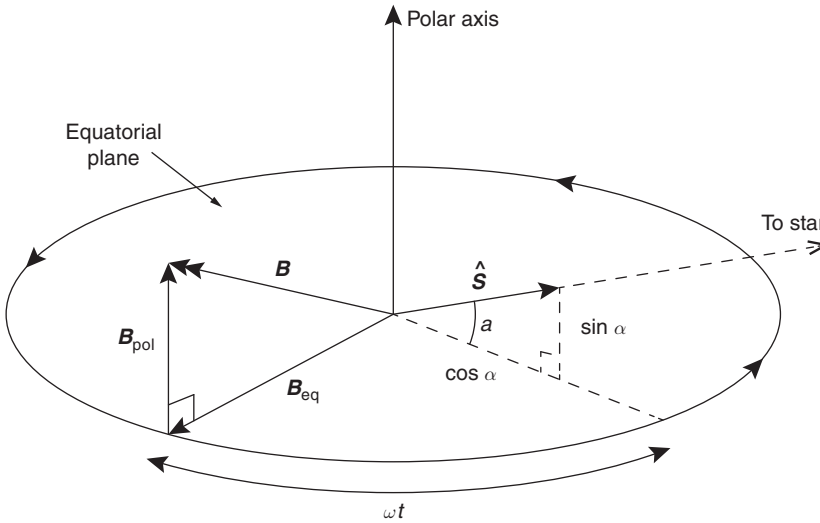

Figure 4.8 Geometry for calculating the change of delay with time due to Earth rotation.

component is dominated by the variation in the geometric delay when observing different parts of the sky, either when observing different targets or when observing a single target as the Earth rotates, but it also includes smaller components due to drifts and vibrations in the instrument and to atmospheric seeing.

In Section 1.4.5 it was shown that the geometric delay for light from a target in a direction $\hat{\boldsymbol{S}}$ and a pair of telescopes i and j with separation $\boldsymbol{B}_{ij}$ is $d_{\text{geometric}} = \boldsymbol{B}_{ij} \cdot \hat{\boldsymbol{S}}$. Thus, the maximum and minimum delays occur when the target lies in directions parallel or anti-parallel to the baseline, and so the delay ranges from $-|B_{ij}|$ to $|B_{ij}|$. In practice targets in these directions are too close to the horizon to observe; if targets are only observed if they are at least 30° above the horizon, then this range is reduced to $\pm\sqrt{3}|B_{ij}|/2$. Thus the delay range required increases linearly with the baseline and can easily reach hundreds of metres.

As explained in Section 2.3, any Earth-based baseline rotates with respect to a celestial target. This causes the required delay to change with time even when observing a single target. The maximum rate of change of delay can be deduced by decomposing the baseline vector $\boldsymbol{B}$ into a fixed component parallel to the polar axis of the Earth $\boldsymbol{B}_{\text{pol}}$ and a component in the equatorial plane $\boldsymbol{B}_{\text{eq}}$, which rotates as shown in Figure 4.8. The geometric delay is then given by

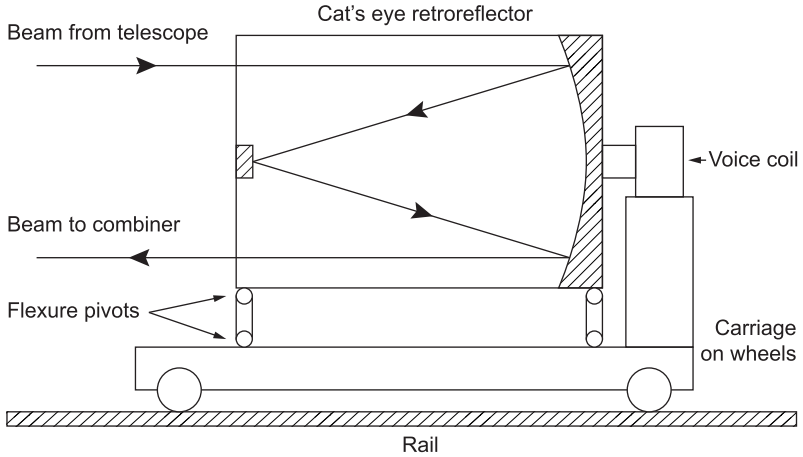

Figure 4.9 Schematic diagram of a delay line showing the path of the starlight through the cat's eye retroreflector. The path of the laser metrology beam has the same geometry as for the starlight beam, but in a plane perpendicular to the page.

$$d_{\text{geometric}} = |\boldsymbol{B}_{\text{pol}}| \sin\alpha + |\boldsymbol{B}_{\text{eq}}| \cos\alpha \cos\omega t, \tag{4.2}$$

where α is the declination of the object, ω is the rate of rotation of the Earth and t is zero at the moment that the angle between the projections of $\hat{\boldsymbol{S}}$ and $\boldsymbol{B}$ onto the equatorial plane is zero. The rate of change of delay is therefore

$$\dot{d}_{\text{geometric}} = \omega |\boldsymbol{B}_{\text{eq}}| \cos\alpha \sin\omega t, \tag{4.3}$$

which has a maximum/minimum value of $\pm\omega|\boldsymbol{B}_{\text{eq}}|$ when the object is at zero declination. Thus, for $|\boldsymbol{B}_{\text{eq}}| = 100$ m, the rate of change of delay reaches a maximum value of 7 mm/s. This change of delay is required to happen smoothly: the short-term 'jitter' in the delay due to vibrations etc. in each delay line has to be kept below $\lambda/20$ during an exposure in order to keep the visibility losses to less than 10%, which translates into a jitter of less than 15 nm RMS at a visible wavelength of 600 nm.

The requirement for a factor of order 10^{10} between the delay range and the allowable delay jitter is met using the multi-stage servo concept first developed by Pierre Connes (Connes and Michel, 1975). An example layout for such a system is shown in Figure 4.9. Starlight is reflected off a retroreflector, which is mounted via a flexible suspension system to a carriage that can be moved to adjust the distance travelled by the beams.

The retroreflector shown in the figure is a so-called 'cat's eye' arrangement consisting of a paraboloid primary mirror with a flat secondary mirror at its forcus. This has the advantage that the tilt of the reflected beam is insensitive to

changes in tilt of the retroreflector and so the tilt of the carriage does not need to be controlled at the sub-arcsecond level, as would be required if a plane mirror was used. However, any displacement of the cat's eye perpendicular to the beam propagation direction causes a displacement of the output beam; this needs to be controlled in order to make sure that the beams from different telescopes overlap correctly. As a result, the carriage is often made to run on rails, which are installed with sub-millimetre precision.

The motion of the retroreflector along the rails is monitored using a 'metrology system' – usually a laser beam is reflected off the same cat's eye as the starlight and the return beam is interfered with a reference laser beam to determine the change in position of the cat's eye. The position of the latter is controlled in real time based on feedback from the metrology system. Coarse position control is achieved using a motor driving the wheels of the carriage and fine control is achieved by moving the retroreflector relative to the carriage using a position actuator, typically an electromagnetic actuator (known as a 'voice coil') or a piezoelectric actuator. Using this technique, delays of hundreds of metres can be introduced with an error measured in nanometres.

Running the delay lines inside a vacuum system eliminates the chromatic dispersion described in Section 1.8 and becomes mandatory for delays of more than 100 m or so. Installing polished rails at high precision inside a vacuum system can become expensive as the length increases; therefore, the CHARA array and the NPOI use a pair of delay lines in series, a short length of continuously variable delay and a longer delay line, which can introduce discrete static delays. Because the longer delay line is static during an observation, it can be based on technologies which do not require precision rails such as 'pop-up' mirrors at discrete locations.

An alternative is to construct a continuously variable delay line, which does not require precision rails. At the MROI, the carriage runs directly on the inside of standard pipes acting as the vacuum enclosure (Fisher *et al.*, 2010). A control system monitors and corrects any errors in the transverse location ('shear') of the reflected beams caused by imperfections in the pipe manufacture. As a result, up to 400 m of delay can be introduced in a single stage. This arrangement allows delays to be changed more rapidly and has higher optical throughput than a two-stage system with switched mirrors.

4.7 Beam combiners

The beams from the different telescopes are mixed to form interference fringes in the beam combiner. There are two classes of information provided by beam

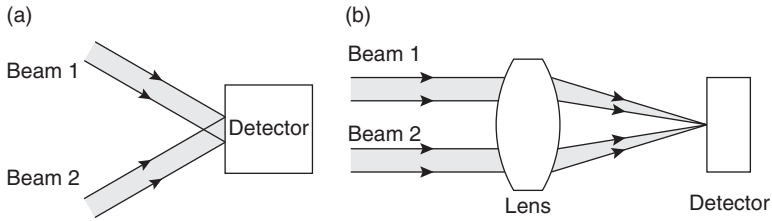

Figure 4.10 Pupil-plane (a) and image-plane (b) beam combination.

combiners. The first is information on the amplitude and phase of the object visibility function, and the second is real-time information on the piston disturbances due to atmospheric seeing and instrumental vibrations and drifts. Both types of information can be produced by a single beam combiner, but the discussion here will centre on the idea of using two specialised combiners, each producing one kind of information, called the 'science combiner' and the 'fringe-tracking combiner', respectively. The science beam combiner will be the main focus of the current section; fringe-tracking combiners will be discussed briefly at the end of this section.

There are a remarkably diverse set of ways of achieving the function of beam combination and so a text like this cannot hope to describe all the possibilities in depth. Instead, the 'zoo' of possible beam combiners is described here in terms of its composition out of a set of functional elements, followed by a few practical examples.

4.7.1 Free-space combination

A major subdivision of types of beam combiners is into free-space and guided-optics combiners. As might be expected from the name, in free-space beam combiners the beam combination takes place under conditions where the beams are propagating without confinement, while in guided-optics combiners, which are discussed in Section 4.7.5, the beams are combined while propagating in waveguides whose dimensions are of order a wavelength.

One element of a free-space beam-combiner design is whether it makes use of 'pupil-plane' or 'image-plane' interference as shown in Figure 4.10. A pupil-plane design was used in the example interferometer introduced in Chapter 1.3 and simply superposes images (usually demagnified) of the telescope pupils. In an image-plane design the light from multiple pupils is brought to a focus in a plane where images of the target as seen through the telescopes (for

point-like targets this will be the diffraction pattern of the telescope pupil) are superposed.

In both cases the interference results in a fringe pattern consisting of an envelope (for example the pupil image or the image-plane diffraction pattern) modulating a one-dimensional sinusoid. The spatial frequency of the sinusoid depends on the relative tilt between the beams at the plane of interference; for the image-plane design, this tilt is proportional to the spacing of the beams in the 'input pupil' of the focussing optic.

A potential advantage of the pupil-plane scheme is that tip–tilt errors introduced by atmospheric seeing change the fringe frequency but do not reduce the fringe contrast: in the case of image-plane interference, tip–tilt errors shift the images and the reduction in image overlap causes a reduction in fringe contrast.

In practice, this advantage of pupil-plane interferometry is limited because the SNR of the fringes is seldom high enough to determine the change in fringe frequency on a frame-by-frame basis. If the power is measured at a fixed frequency in the fringe pattern then atmospheric tilts cause a loss in fringe power comparable to that seen in the equivalent image-plane combiner.

A unique advantage of image-plane combiners occurs when they are operated in a so-called "Fizeau" mode (Faucherre *et al.*, 1990). In this mode the beams are arranged in the combiner as a scaled version of the input pupil of the interferometer as a whole, that is to say the wavefront sampled by the telescopes. This means that the vector separations of the beams at the focussing optic must be a scaled version of the interferometric vector baseline, with a demagnification the same as has occurred to the individual beams.

If this condition is satisfied, then fringes can be seen over a field of view which is considerably larger than the diffraction limit of the individual telescopes. This is because the optical path difference (OPD) introduced between the interfering beams an object not being at the phase centre is exactly cancelled by the change in OPD induced for the equivalent point in the image plane where the beams meet.

In practice, the field of view of an interferometer is usually limited to values considerably less than the diffraction limit of the collectors by the limited sampling of the (u, v) plane (see Section 2.4.3), and so no Fizeau-mode combiners have so far been implemented in long-baseline interferometers: all the Fizeau-mode combiners in use are in aperture-masking systems or in systems where the collectors are on a common mount such as the LBT.

One form of the pupil-plane combiner makes use of a beamsplitter to combine the pupil images as shown in Figure 4.11. In this configuration, the relative tilt between the combined beams can be made to be zero; in this 'co-axial' case

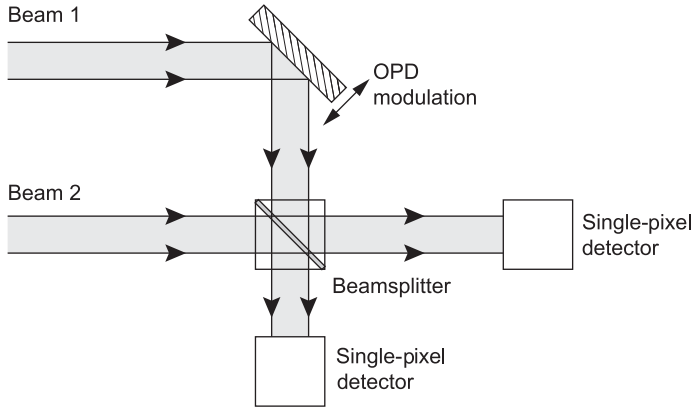

Figure 4.11 Co-axial temporally coded fringe pattern.

the spatial frequency of the fringes tends towards zero, in other words each beam is uniform in intensity and the fringes are said to be 'fluffed out'. Each output can be sampled using a single-pixel detector and the measured intensity at any moment depends on the phase difference between the beams: if the path length in one arm of the interferometer is changed linearly with time, the intensity will vary sinusoidally with time – a 'temporal fringe pattern', which can be captured by recording the detector output as a function of time. Thus, we can speak of fringes being 'spatially coded' or 'temporally coded', and it is also possible to have a mixture of the two.

Typically, the differential delay is scanned backwards and forwards rapidly enough to capture the fringe information before the fringe phase is altered by atmospheric seeing. The delay scanning (or delay *modulation*) is usually accomplished using mirrors attached to fast electro-mechanical actuators such as piezoelectric devices or 'voice-coil' actuators, although it is in principle possible to use electro-optic devices such as liquid crystals to perform the same function.

Note that this arrangement has two optical outputs: energy conservation means that when one output is bright, the other is dark (assuming a lossless beamsplitter). Thus, the outputs are said to be *complementary* and the temporal fringe patterns seen in the two outputs are 180° out of phase; subtracting the intensities of the two outputs gives an estimate for the sinusoidal fringe signal independent of any fluctuations in the total signal level, and so is useful in situations where there is a slowly-fluctuating background, for example in the thermal infrared.

4.7.2 Multi-baseline combiners

Interferometers with N telescopes can in principle measure the object visibility function on $N(N-1)/2$ different baselines. There are multiple ways to accomplish this for $N > 2$. The fringes on all baselines can be observed simultaneously or measurements can be made on a subset of the baselines at any given time and the beam combinations can be switched sequentially to build up the information on all baselines.

If fringes are to be measured on all baselines at the same time, two ways of doing this are shown in Figure 4.12. One option is to split each beam $N-1$ ways, and perform pairwise combination of the split beams to form a set of $N(N-1)/2$ two-beam interference patterns. A second option is to combine up to N beams into an 'all-in-one' fringe pattern, which contains the superposition of $N(N-1)/2$ fringe patterns corresponding to interference between all pairs of beams. Other options exist, for example combining beams in M-way fringe patterns, where $2 < M < N$.

For all-in-one combination, the superposed fringe patterns need to be 'multiplexed' (or 'coded') in some way, which allows the information corresponding to the visibilities on different baselines to be separated from each other, in other words 'demultiplexed' (or 'decoded'). The most common way of achieving this multiplexing is to arrange for the different fringe patterns to have different fringe frequencies.

Figure 4.13 shows a three-beam image-plane combiner where the input beams are arranged in a line with nearest-neighbour centre-to-centre spacings of 1:2. The fringe frequency for the interference between a given pair of beams will be proportional to the spacing between the beams, and so the fringe frequencies for the three possible baselines will be in the ratio 1:2:3. Fringe information for each baseline can be extracted by taking the Fourier components of the multiplexed fringe pattern at each of the three spatial frequencies as shown in Figure 4.13.

The arrangement of the input beams in this example is a 'non-redundant' one, in that none of the three spacings between the beams overlaps with any other spacing. Such non-redundant spacings are equally useful in choosing the best spacings for telescopes in an interferometric array, and so are a well-researched subject. Two-dimensional non-redundant arrays can also be used as well as the one-dimensional array shown in this example, but are less useful when spectral dispersion is incorporated, as described in Section 4.7.4.

A non-redundant pattern can be used similarly for a co-axial temporally multiplexed system as shown in Figure 4.14. In this case the fringe frequencies are proportional to the difference between the delay scanning *velocities* for

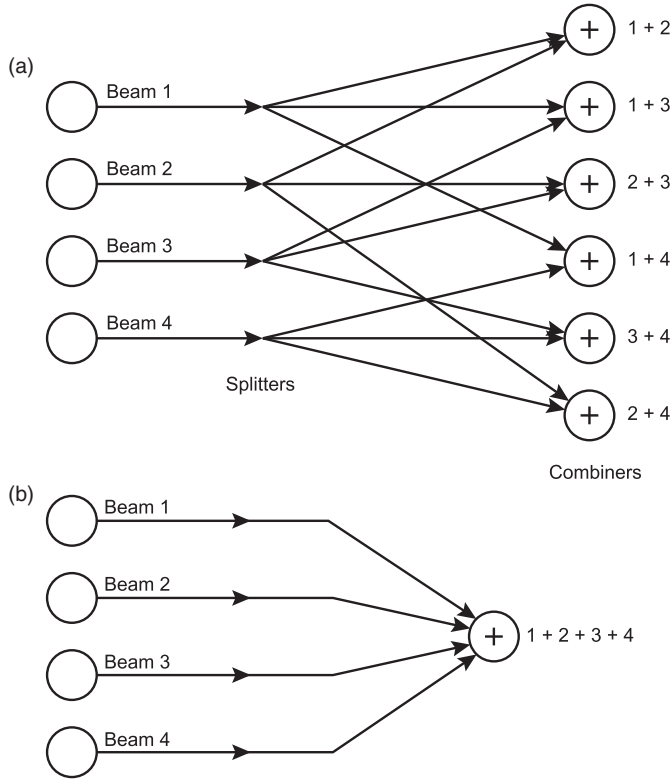

Figure 4.12 Schematic outline of two types of multi-telescope beam combiners, a pairwise combiner (a) and an all-in-one combiner (b).

the relevant pair of beams, and so the velocities must be arranged in a non-redundant fashion as shown in Figure 4.15.

An issue for multi-baseline combiners is *crosstalk* between the fringes on different baselines: if some of the fringe signal from one baseline leaks into the signal measurement for another baseline this can introduce errors in the visibility measurement, which are particularly difficult to calibrate as they will depend on the visibilities on different baselines. Crosstalk can arise if the fringe frequencies of different baselines are not sufficiently well separated or can be caused by atmospheric temporal piston variations during the scan in temporally multiplexed interferometers (Buscher, 1988b).

4.7.3 Spatial filtering

The use of spatial filtering to improve fringe measurement in the presence of atmospheric seeing has been described in Section 3.7. Spatial filtering is

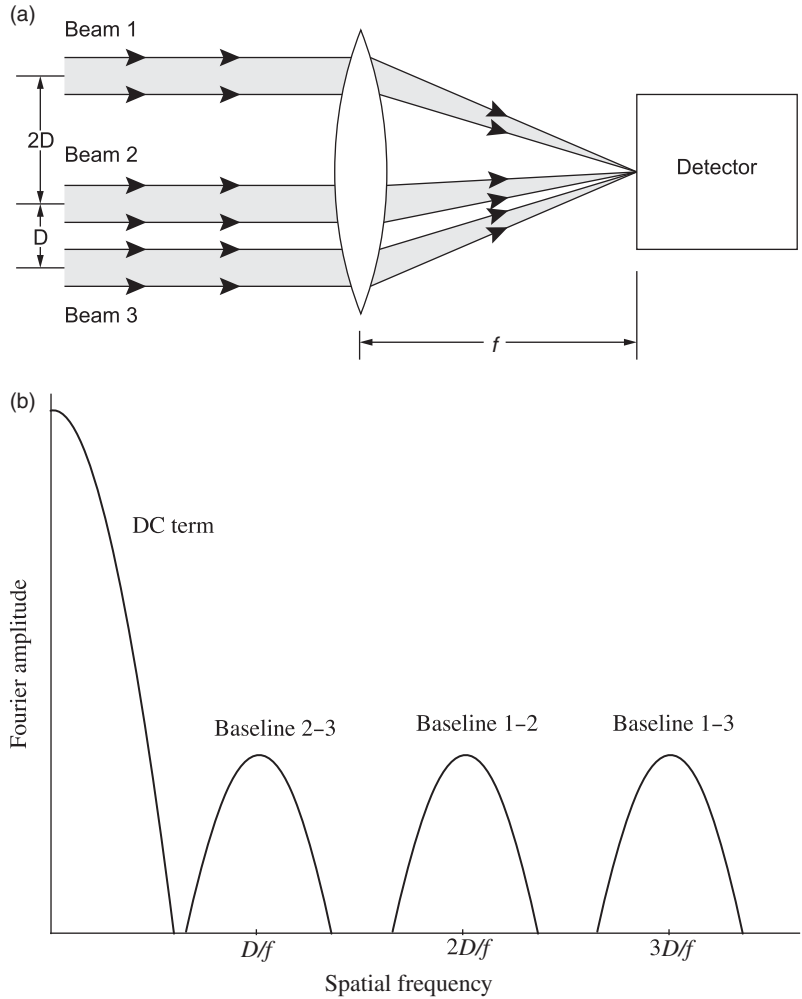

Figure 4.13 Multiplexing three beams into a single fringe pattern in the image plane (a) and the modulus of the Fourier transform of the fringe pattern showing peaks at three different frequencies (b).

usually accomplished within the beam combiner for a number of practical reasons.

In an image-plane combiner, a form of spatial filtering can be accomplished in software simply by setting to zero (or not reading) the intensity of any pixels outside a 'virtual pinhole' of a chosen size – usually the diameter is chosen to be comparable to the diffraction limit of the telescope. This is entirely

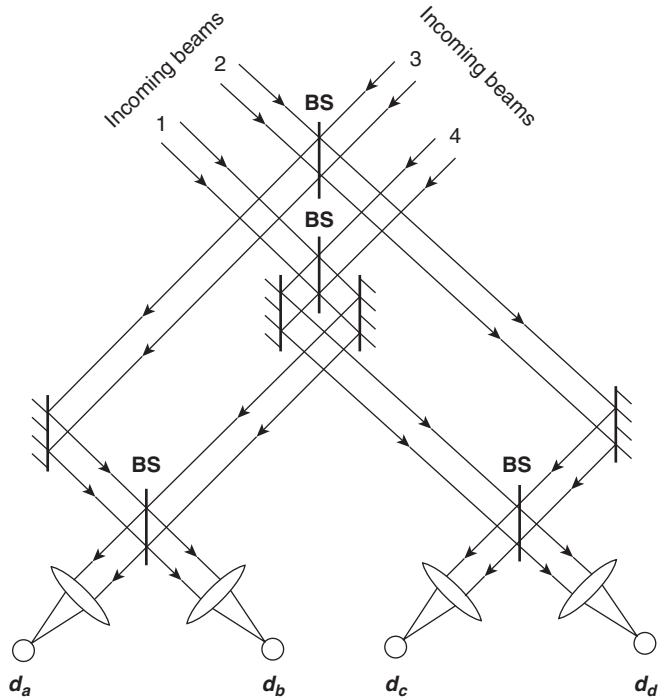

Figure 4.14 Schematic for the four-way temporally multiplexed beam combiner in the COAST interferometer (Baldwin *et al.*, 1996). BS indicates beamsplitters where the beams are combined. The output is on four detectors d_a–d_d, which all receive the superposition of all four input beams. OPD modulation takes place external to the beam combiner.

equivalent to passing each of the beams through a real pinhole of the same size before combination, but requires no additional hardware.

In a co-axial beam combiner, spatial filtering can be straightforwardly accomplished by focussing the combined light after beam combination onto a single-mode fibre. Since all the light is filtered by the same fibre, there is no need to match the lengths of the fibres to equalise chromatic dispersion and polarisation effects, as is needed when separate fibres are used for each input beam.

Single-mode fibres at the input of a beam combiner can be used to perform three functions simultaneously: spatial filtering, beam transport and beam reconfiguration. For example, in the AMBER beam combiner, the light from each of three telescopes illuminates a separate single-mode fibre. The fibres transport the light a few metres from the room-temperature environment into

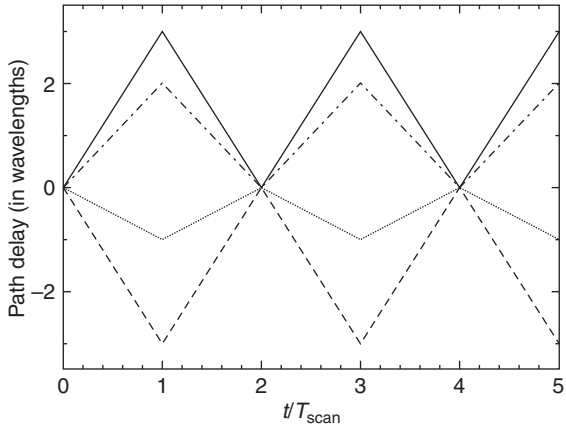

Figure 4.15 The scanning pattern for the temporally multiplexed combiner in Figure 4.14. The scanning velocities are at −3, −1, +2, and +3 times a base velocity. This non-redundant arrangement of velocity differences means that the interference fringes between the six pairs of input beams all appear at different fringe frequencies.

a cryogenic dewar and bring the beams into an arrangement where they are all arranged in a linear non-redundant pattern to illuminate an 'image-plane' combiner as shown in Figure 4.16 (the distinction between the image plane and the pupil plane is somewhat arbitrary after the beams have been through a single-mode fibre, as the beam intensity profile in all planes is roughly Gaussian).

4.7.4 Spectro-interferometry

Interference is a wavelength-dependent phenomenon, so there are a number of practical reasons, for example minimising the effects of chromatic dispersion, for observing interference in narrow spectral bands. In addition, much of the physical information about an astronomical object is derived from looking at the variation with wavelength of the emission from the object (or part of the object) under study. Thus, the combination of spectral information and interferometry can potentially provide better interferometric performance and more information about the object.

The simplest form of spectro-interferometry can be achieved by placing a narrow-band filter in front of the detector, observing the interference fringes, and then repeating the observation with a different filter. Spectrally dispersing the light by placing a prism or grating in the beam-combination optics allows

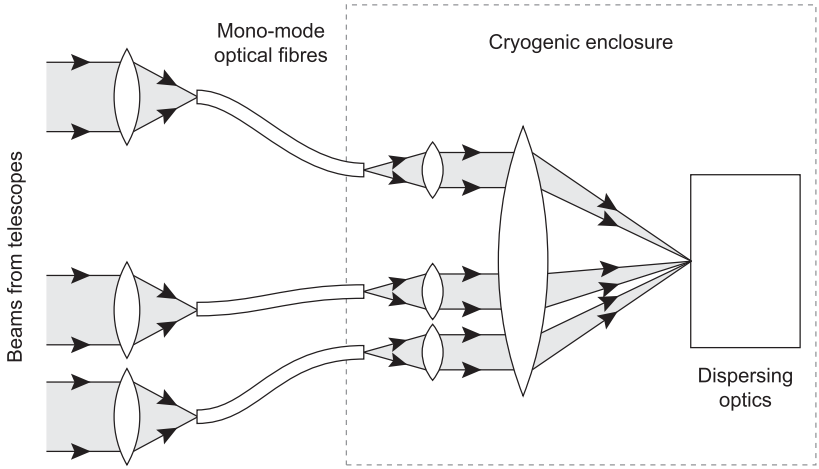

Figure 4.16 Schematic of the AMBER beam combiner.

interference to be observed simultaneously at many wavelengths and so spectro-interferometry of this form is now used in most optical interferometers.

Spectro-interferometry can be simply implemented in a co-axial temporally coded fringe pattern by placing a prism in the combined beam and focussing the resulting spectrum onto a one-dimensional detector as shown in Figure 4.17. Each pixel in the detector will show temporally coded fringes and can be analysed as a parallel collection of interferometers working in different bandpasses.

Spectral dispersion typically spreads the signal continuously along one dimension of a two-dimensional detector; therefore, in a spatially coded beam combiner the fringe modulation must be arranged to be in the perpendicular direction to this, in order that the fringe information is not 'smeared out'. Thus, the fringes for different baselines must run parallel to each other and so one dimension of the monochromatic fringe pattern contains no information. Most combiner designs will 'squash' the monochromatic fringe pattern in this dimension using anamorphic optics in order to reduce the number of pixels needed to sample the pattern, and then disperse the light along the direction in which the fringe pattern has been squashed to give a dispersed fringe pattern as shown in Figure 4.18.

An alternative form of spectro-interferometry combines two separate interferometric techniques: spatial interferometry and Fourier-transform spectroscopy. Such 'double-Fourier' interferometry (Mariotti and Ridgway, 1988) can be understood in the context of a co-axial, temporally coded beam

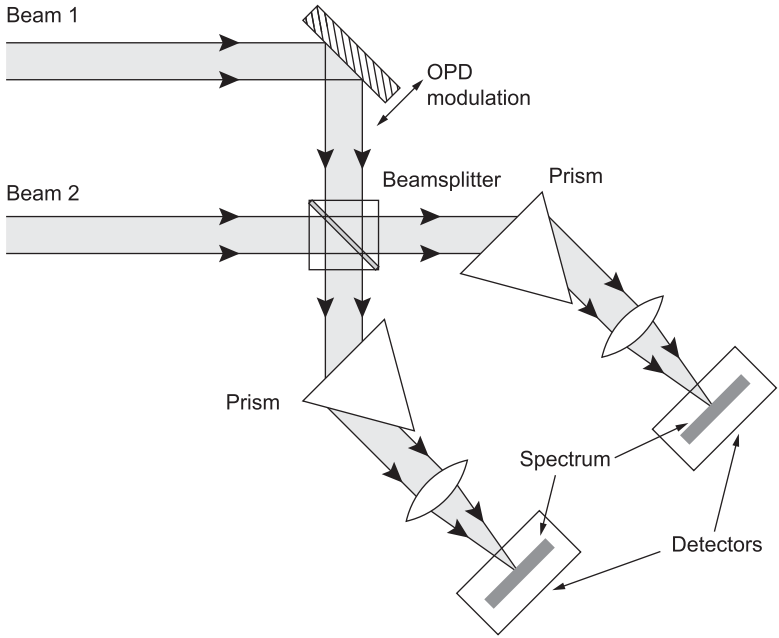

Figure 4.17 A spectrally dispersed temporally coded beam combiner.

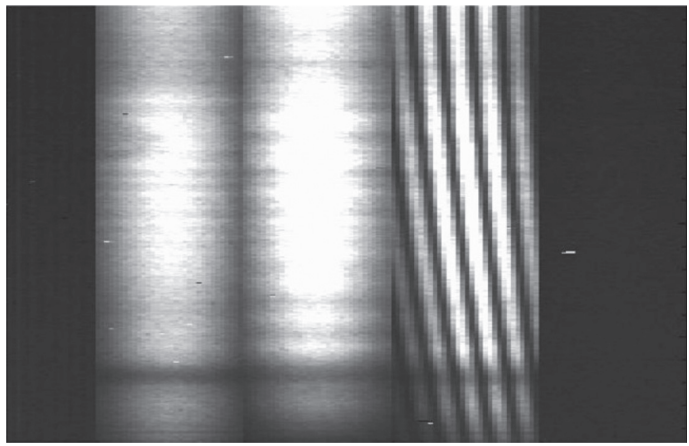

Figure 4.18 Spectrally dispersed fringes from the AMBER beam combiner. The spectral dispersion is in the vertical direction: each horizontal line on the right-hand side is a spatially coded fringe at a given wavelength. The left-hand pair of channels are photometric calibration channels and so have no fringes. From Tatulli and Duvert (2007).

combiner with no spectral dispersion, illuminated by polychromatic radiation. As explained in Section 1.7, the fringe pattern seen with wideband radiation depends on a combination of the object visibility function and a Fourier transform of the intensity spectrum of the source. This combination can be recovered by observing the fringe pattern over a wide range of OPD, and performing the Fourier transform of the intensity as a function of OPD. In other words, the spectral dispersion function is accomplished using the Fourier-transform spectrometer that is built in to any interferometer, rather than needing a separate spectral dispersing element.

4.7.5 Guided-optics combination

Beam combiners using lenses or beamsplitters often have strict alignment requirements in order to maintain high fringe contrast and as a result can require frequent and sometimes laborious realignment. One way around this is to make use of guided optics to keep the light confined right up to the point of beam combination.

One form of guided-optics combination uses single-mode optical-fibre couplers instead of beamsplitters. This form of beam combination is represented by the FLUOR (Coudé du Foresto *et al.*, 2003) and VINCI (Ségransan *et al.*, 2003) combiners on the CHARA array and the VLTI, respectively. Propagation in single-mode fibres provides spatial filtering and these combiners have secondary outputs for photometric calibration (see Section 8.8.2), and so both combiners are able to provide high-precision visibility measurements.

Both the FLUOR and VINCI combiners are two-telescope combiners. Multi-way combiners using fibre couplers are possible in principle but become complex. A more scalable technology is integrated optics, the optical equivalent to integrated circuits. Complete optical systems can be fabricated using microlithography techniques, and because all the optical components are built into a single 'chip' a few millimetres in size, then once the device is fabricated there is little opportunity for components to go out of alignment and the effects of thermal expansion are usually small.

Integrated optics are typically fabricated out of planar optical waveguides fabricated in a two-dimensional 'wafer'. Because the transverse dimensions of the waveguide are smaller than a wavelength, they act as monomode devices, similar to optical fibres, so provide spatial filtering to the beams.

An example of an integrated optics combiner is the PIONIER combiner used on the VLTI (Le Bouquin *et al.*, 2011). The layout of the combiner is shown in Figure 4.19. Light from one of the VLTI telescopes is injected into each of the four inputs to the combiners. The 'optical circuit' on the integrated optics chip

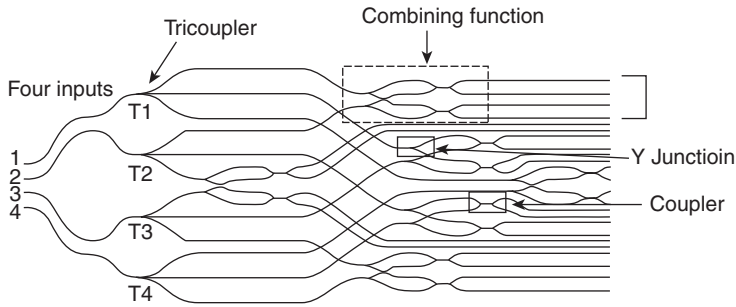

Figure 4.19 The layout of the integrated optics chip used in the PIONIER beam combiner. From Benisty *et al.* (2009).

implements a set of pairwise co-axial beam combiners to make interference for all six baselines simultaneously. In fact there are two beam combiners for each pair of telescopes: one combiner produces fringes in its complementary outputs at phase shifts (with respect to some arbitrary zero point of phase) of 0° and 180° and the second combiner produces fringes with phase shifts of approximately 90° and 270°. This so-called quadrature or 'ABCD' fringe information is sufficient to measure the amplitude and phase of the fringes (see Section 8.5.3) without requiring additional temporal fringe scanning.

4.7.6 Fringe-tracking combiners

Fringe-tracking beam combiners have similar requirements to science combiners, but the weighting between these requirements is different. Although in both cases the combiner must be designed to ensure accuracy in the measurements of fringe phase, in a fringe tracker the measurement of fringe amplitude is less important. In addition, the desirability of sampling the fringe visibility on all $N(N-1)$ baselines in an N-telescope interferometer is reduced in the case of a fringe tracker, since all the atmospheric OPD values can be sampled using $N-1$ appropriately chosen baselines. Finally, the scientific requirement for a large amount of spectral resolving power in order, for example, to isolate specific spectral features in the target is not present.

Instead, a primary requirement for the fringe tracker is to achieve as high an SNR on the fringe phase as possible in as short a time as possible in order to follow the atmospheric perturbations. Optimising for this requirement might lead to a design which has, for example, lower spectral resolution (or none at all) and covers fewer baselines than an equivalent science combiner.

An example fringe-tracking combiner is the CHAMP (Berger *et al.*, 2008; Monnier *et al.*, 2012) combiner at the CHARA array. Dichroic (i. e. wavelength-dependent) beamsplitters take light from all six of the CHARA

telescopes, but instead of measuring fringes on all 15 baselines, the combiner performs a set of six pairwise fringe combinations.

The beam combination is a co-axial, temporally modulated design. Each output of the combiner is focussed onto a single pixel of an infrared array – there is no spectral dispersion: the light from a whole astronomical band (J, H or K band) is used. The design is optimised for sensitivity at low light levels when using readout-noise-dominated detectors: the fewest pixel readouts are done in order to maximise the SNR in a single read.

4.8 Detectors

The beam combiner forms a fringe pattern on one or more light-sensitive devices. These 'detectors' are responsible for converting the light intensity pattern into a form suitable for analysis. The detector used in Michelson's day was the human eye, but modern interferometers use photoelectric devices, which convert the intensity into an electronic signal. These detectors are far more sensitive than the human eye, but nevertheless have limitations that can be critical to the performance of the interferometer.

Typical detectors are array devices, consisting of a two-dimensional grid of light-sensitive elements or 'pixels'. The light level in all the pixels are typically read out as a 'frame' at the end of each exposure. An important requirement for interferometric detectors is the ability to operate at frame rates of hundreds of times per second in order to avoid temporal smearing of the fringes. This is quite different from most detectors used in astronomical applications, which are typically read out less than one frame per second. The frame rate can often be increased by reading out fewer pixels; this requires efficient beam-combiner designs which put all the light into as few pixels as possible.

A second consequence of this high frame rate is that there will be relatively few photons captured per frame, so maximising the amount of light captured (known as the 'quantum efficiency' or QE) and minimising the amount of readout noise (see Section 5.2.2) is critical for interferometric detectors.

At optical wavelengths there are two types of high-frame-rate device with negligible readout noise and good QE. Silicon avalanche photodiodes (APDs) are photodiodes with internal electrical gain. When operated in so-called 'Geiger mode' the effective gain can be more than 10^6 so that they produce a macroscopically detectable pulse of current for the arrival of each photon. Photon-counting techniques mean that these devices can be entirely photon-noise-limited and the QE can be as high as 60% or more. A disadvantage of APDs is that they are usually only single-pixel devices (or effectively so – the trend is towards multiple diodes, which are wired together as a single pixel) and

so designs that use only a few pixels, such as co-axial combiners, are favoured when using silicon APDs.

A newer technology is electron-multiplying CCDs, in which the light detection is in a two-dimensional silicon array of thousands or millions of pixels, but which include amplification similar to that of APDs. These can achieve photon-counting performance under low-light-level conditions and can have quantum efficiencies of more than 90%; they are therefore becoming the detector of choice for interferometric applications.

Until recently the situation at near-infrared wavelengths was not as good as in the optical, because the read noise of high-QE detectors for this wavelength regime was at the level of 10 electrons per read or more. Furthermore, to keep the readout noise low, the readout had to be relatively slow so that designs which involved large numbers of pixel reads were less favoured.

This has begun to change because of the advent of detectors which include APDs made of HgCdTe, a semiconductor which has good QE (typically above 60%) at infrared wavelengths. Devices having read noises as low as 0.8 electrons RMS have been tested in the laboratory, so that it is looking likely that photon counting at these wavelengths may become possible in the near future (Finger *et al.*, 2012). Furthermore, these APDs can be straightforwardly incorporated into two-dimensional arrays and it is therefore likely that this technology will come to dominate near-infrared interferometry.

At wavelengths longwards of 3 μm the dominant source of noise is likely to be thermal background noise and so read noise is less of an issue. The main detector characteristics that are important under background-limited conditions are linearity and uniformity of the response between different pixels, so that the large background levels can be accurately subtracted.

4.9 Alignment

Alignment of the optical components in the beam train is a serious practical issue in all interferometers, because the mechanical tolerances to which the optical elements need to be placed are severe. The approach used in all interferometers is not to try to place components by ‘dead reckoning’ alone, but rather to adjust the positions of the components based on optical measurements of the error in position.

An interferometer needs to be aligned when it is initially assembled, but the componenents will inevitably drift out of alignment due to thermal and other long-term mechanical effects. The timescales before these drifts become noticeable can be as long as months or as short as minutes but inevitably

realignment is necessary, and so an ongoing programme of realignment is part of the operation of any interferometer.

4.9.1 Piston

The typical alignment needs of an interferometer can be broken down in terms of the Zernike components of the wavefront error affected. The lowest-order component, the piston component, is the one subject to the most severe drifts. Thermally induced expansion and contraction effects can mean that the pathlengths inside an interferometer and the baselines within a telescope can change by hundreds of microns over the course of a day, much greater than the RMS errors introduced by the atmosphere and typically larger than the coherence lengths over which fringes can be found.

Piston errors are typically compensated for by adjusting the delays introduced by the delay lines. In the crudest case, the delay lines are scanned backwards and forwards until fringes are found on each baseline on each star.

A more sophisticated approach makes use of a model which tries to predict where the fringes can be found, so that the time spent searching for fringes is minimised. This 'baseline model' typically includes parameters describing the three-dimensional locations of the telescopes and the internal delays from the telescopes to the beam combiner. This latter parameter is sometimes called the 'constant term' as the locations of the telescopes predict the external OPD due to the location of the star in the sky, which is changing, while the internal OPD is stable at least over the short term.

This model is updated using measurements of the delay offsets at which fringes are found on stars; after the measurement of the fringe delay offsets on at least four stars widely dispersed across the sky the whole model can be updated.

4.9.2 Tilt

The tip–tilt Zernike aberrations typically experience drifts, which are not as severe as those for piston, but are still at the level where regular readjustment is mandatory.

Any errors in the tilts of any mirror in the system will cause twice that error in the final tilt of the beams at the beam combiner. For a differential tilt error between two circular beams of θ radians the fringe contrast will be reduced by a factor

$$\gamma \approx 1 - (\pi\theta D/\lambda)^2/8, \tag{4.4}$$

where D is the diametre of the beam (Porro *et al.*, 1999).

With a beam diameter of 10 cm, a drift in the tilt of one mirror of about 0.05 arcsecond will cause a 1% degradation of fringe contrast for light at a wavelength of 500 nm. This order of magnitude of tilt stability is difficult to attain so some kind of tilt realignment is usually necessary on at least a nightly basis.

Realignment of the net tilt of the optical train can be accomplished by using a tip–tilt sensor close to the beam combiner, i. e. after most of the optical elements in the beam train. A complication is that full realignment requires more information than is available from beam tilt measurements made at one location in the beam train: because of the long pathlengths inside an interferometer, tilt drifts can cause the transverse position of the beam (the so called 'beam shear') to drift as well.

For example, a 1-arcsecond tilt error in a mirror 500 m from the beam combiner can cause the beam centre to move by 5 mm, enough to cause a significant loss of beam overlap at the beam combiner and hence a loss in fringe contrast. Larger beam-shear drifts can cause the beam to 'wander off' the mirrors in the beam train and so light will be lost due to vignetting. Mirrors which are tilted by the same amount but are different distances from the combiner will cause different amounts of beam shear, so without knowing which mirrors have drifted, it is difficult to know what combination of tilt and shear correction is required.

The solution usually adopted is to make measurements at multiple places along the beam train to constain the tilt errors of most or all of the mirrors in the train. There are perhaps as many ways of effecting this alignment as there are interferometers, but a common feature is that alignment of every element in the beam train is typically time-consuming and so is not performed very often. As a result, significant errors in beam tilt and shear can build up as a result of instrumental drifts between system realignments.

An approach to remedy this is to combine relatively infrequent 'coarse' alignment to correct any gross errors in individual optics, which could cause vignetting, and more frequent 'fine' realignment of a subset of the optics to correct for the drifts in the overall tilt and shear of the beams, which could cause reductions in fringe contrast. If the realignment procedure is automated, it can be repeated frequently without too much loss in observing time.

4.9.3 Higher-order errors

Higher-order wavefront errors such as defocus are less of a problem, as they are less sensitive to mechanical alignment issues. For example, in most systems wavefront defocus can be held at an adequate level by tolerancing of spacings

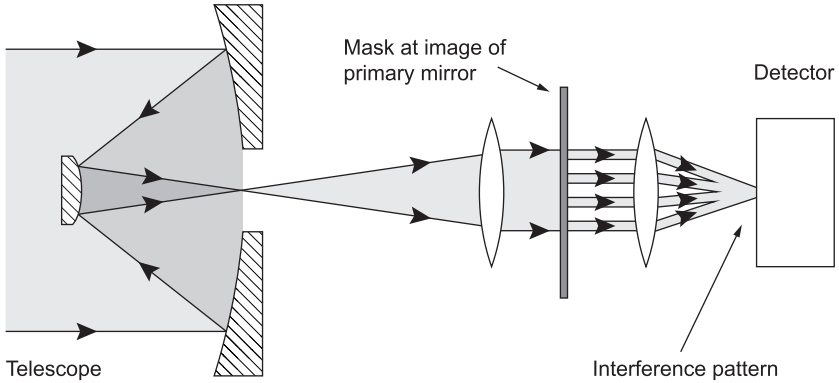

Figure 4.20 The optical setup for an aperture-masking experiment on a large single telescope.

at the micron level, while tilt errors can be significant when the mechanical errors reach a few tens of nanometres. As a result many of the higher-order errors can be removed at the initial alignment stage and no further realignment is necessary – 'set and forget'. An exception to this is the telescopes, which are in a thermally hostile environment and typically need to be focussed on at least a nightly basis.

4.10 Aperture masking

Interferometry with a single telescope sounds like a contradiction in terms, but at optical and infrared wavelengths this technique is currently one of the most scientifically productive ways of imaging using interferometry. The first interferometric measurement of an astronomical object was done this way (Stephan, 1874), and the technique was revived when closure phase was shown to be a practicable way of imaging at optical wavelengths in the presence of turbulence (Haniff *et al.*, 1987).

Aperture masking works by converting a single large aperture into multiple smaller apertures using a mask containing an array of small holes of order r_0 in diameter. This mask is nominally located at the entrance aperture of the telescope, but for practical convenience is usually mounted at a reimaged and demagnified pupil plane in an optical arrangement like that shown in Figure 4.20, or in some cases at the secondary mirror of the telescope (Tuthill *et al.*, 2000b).

The light passing through the holes comes together in the focal plane of the telescope and forms an image-plane interference pattern consisting of the diffraction pattern of the holes crossed by interference fringes. The holes are normally arranged in a non-redundant pattern (see Section 4.7.1) so that the interference between any pair of holes appears at a spatial frequency which is different from any other pair of holes. For this reason another common term for aperture-masking interferometry is 'non-redundant masking' or NRM.

A camera mounted in the focal plane is used to take images with exposure times of order t_0 in order to 'freeze' the fringe motions and the images are processed in an identical manner to the fringe patterns from a long-baseline interferometer. If data at sufficient (u, v) points are collected, images can be reconstructed from the power spectrum and bispectrum and will have a maximum resolution of order λ/D, where D, is the diameter of the telescope.

This may not seem any better than can be achieved with adaptive optics on the same telescope, but there are a number advantages over AO which make NRM scientifically competitive. The first of these is that NRM can work well at shorter wavelengths where AO struggles to perform.

Secondly, even at the near-infrared wavelengths that modern AO systems are optimised for, AO systems typically leave behind some uncorrected wavefront perturbations. Propagating these perturbations to the image plane results in an image of a point object which consists of a diffraction-limited 'core' surrounded by a more diffuse 'halo'.

If there is faint structure near to a bright object this can be lost in the halo, particularly as this halo is time-variable and therefore is difficult to calibrate. An aperture-masking system is able to image fainter structure near to a bright object because the 'transfer function' between the object and the observables, especially the closure phases, is more stable and so easier to calibrate than with AO.

A disadvantage of NRM over AO is that the mask blocks out most of the light (in many cases 99% or more) and only short exposure times can be used. As a result, most aperture-masking results have been on relatively bright targets. If aperture masking is used in conjunction with AO, the AO will reduce the level of piston phase difference and so the exposure time can in principle be extended to many t_0 and thereby allow aperture masking to be used on fainter targets (Bernat *et al.*, 2010).

Most of the design considerations which apply to long-baseline interferometers apply also to NRM setups, with the practical constraint that NRM instruments typically have to fit in with other instruments on the same telescope and so cannot be too large or complex. Array layout is more flexible

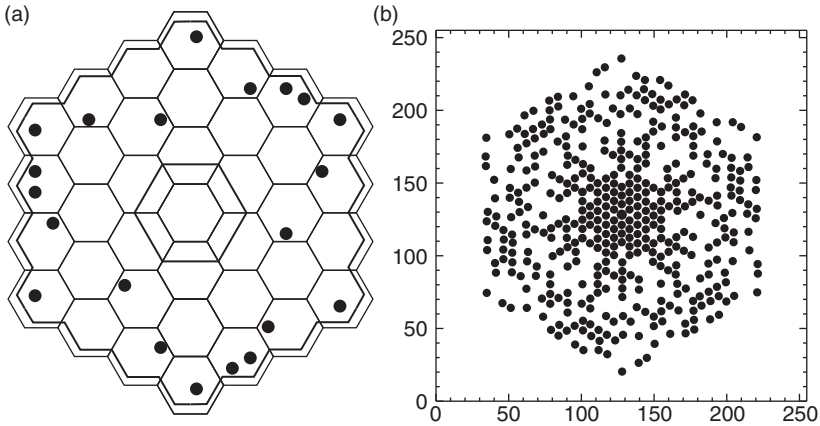

Figure 4.21 The mask layout for a Keck aperture masking experiment (a) and the resulting (u, v) coverage (b). From Tuthill (2012).

in NRM: in particular the number of 'telescopes' can be as large as is consistent with having a non-redundant layout. As a result it is possible for a single mask configuration to provide sufficient (u, v) coverage to make a high-quality image (see Figure 4.21), but the (u, v) coverage can also be increased by using different masks in succession or rotating a single mask.

Separate delay lines are not necessary in an NRM system because the optical delays are equalised at a gross level by pointing the telescope as a whole towards the object, and an AO system can function effectively as a fringe tracker to remove the fine errors in OPD.

The beam combination in an NRM system is by default an image-plane beam combination in a Fizeau mode (see Section 4.7.1). A disadvantage of this form of beam combination is that it couples the array layout to the beam-combination layout, and if spectral dispersion is required then only linear arrays can be used. A number of initiatives to do 'pupil remapping' are under way to overcome this problem (Perrin *et al.*, 2006a; Jovanovic *et al.*, 2012).

The Large Binocular Telescope (LBT) can be considered to be a special form of an aperture-masking system, with the two telescopes acting as two large 'holes' in a virtual mask. Because the telescopes are large compared with r_0 AO is required for each telescope, together with a fringe tracker to cophase the two telescopes.

The wide field of view offered by the Fizeau mode means that cophasing the two telescopes on a bright star within the field should stabilise the fringes on any star within the isoplanatic patch of the atmosphere. Rotation of the

baseline between the two telescopes comes from observing the target at different times of night, so that the paralactic angle rotation of telescope on its mount gives a form of 'Earth rotation synthesis'. Because there are only two apertures, closure phase techniques are not currently envisioned for use in the conventional two-telescope interferometry, but inserting an aperture mask into the beam has been proposed (Stürmer and Quirrenbach, 2012) and this would allow conventional 'dilute-aperture' interferometric imaging techniques to be employed.