6

Interferometric observation of faint objects

The ability to observe faint objects is a key requirement for an astronomical instrument. Many types of object are intrinsically rare and therefore the closest examplars are far away and hence faint. For objects which are less rare, being able to 'go fainter' means that more exemplars of the class can be studied to give statistical validity to any findings.

The exposure times used in interferometry are typically much less than those used in other types of astronomical instruments – usually milliseconds instead of minutes or hours. As a result, a target which would be considered bright for many astronomical observations is considered faint in interferometric terms. This means that the majority of potential astronomical targets are likely to be in the faint-object regime for interferometry, because they will have been discovered using techniques which have intrinsically better faint-object sensitivity.

This chapter takes a quantitative look at the limitations of interferometry when observing faint objects. It looks at the trade-offs involved in adjusting the parameters of an interferometric observation such as the exposure time, with the aim of (a) determining the best parameter settings to use for observing faint objects and (b) determining the faintest possible object which can be observed under a given set of conditions.

Adaptive optics (AO) systems can be used to increase the signal-to-noise ratio (SNR) for an interferometric measurement, because they allow the use of a large-aperture telescope while ameliorating the negative effects of atmospheric wavefront errors on interferometric SNRs. In the same way, cophasing fringe trackers allow the use of long exposure times to improve the SNR of observations of faint objects.

Unfortunately, it is precisely on faint objects that the assumption that the atmospheric correction provided by active correction systems such as AO and fringe trackers breaks down. These systems need a sufficient number of

photons from a reference object to accurately sense the wavefront perturbations, and the reference object is typically the object under study itself. If the object being studied does not provide enough photons then the level of correction will be worse, degrading the SNR of interferometric measurements, which were already low due to the faintness of the source.

This vicious cycle means that the limits to observing faint objects with interferometers often come from the limits of the active correction systems rather than the ability to build larger telescopes. Later sections explore these limits in more detail to understand the faint-object limits for interferometry. In the discussion that follows, the examples given will concentrate on the measurement of the power spectrum. The scaling of the SNR of the bispectrum will behave in a similar way.

6.1 The optimum exposure time

It was shown in Section 3.3 that changes of the atmospheric 'piston' phase with time cause fringe motion during a finite-length exposure and this causes the fringe pattern to smear out, with a resultant loss in fringe visibility. Thus, when observing bright objects, the exposure time should be reduced as far as possible in order to minimise the amount of calibration required of the visibility loss due to fringe smearing. When observing faint objects, the main concern is maximising the SNR, and this favours longer exposure times.

A fringe tracker can be used to mitigate the piston fluctuations but, as mentioned above, the assumption that perfect correction of atmospheric seeing is possible becomes less valid for faint sources. In the rest of this section it will be assumed that for faint sources a fringe tracker is a coherencing one, keeping the fringe envelope centered to within a few wavelengths but not having any effect on the fringe motion during an exposure. The assumption of a coherencing-only fringe-tracker operation at faint light levels is justified further in Section 6.4.

The optimum exposure time to use with a faint object needs to balance a number of competing effects. Longer exposures allow more photons to be collected in an exposure, and this will increase the SNR of the fringe measurements. At the same time, longer exposures are also subject to greater visibility losses due to fringe smearing, and this will tend to reduce the SNR of the fringe measurements. Finally, the shorter the exposure time, the greater the number of exposures that can be averaged in a given overall observing time and this will tend to increase the SNR of the incoherently averaged result.

To take all these effects into account, we need to find the exposure time τ_{exp} which maximises the SNR of an incoherently averaged observable such as the power spectrum or the bispectrum over some fixed observation time τ_{obs}. If an exposure time of $\tau_{\text{exp}} \ll \tau_{\text{obs}}$ is used then $\tau_{\text{obs}}/\tau_{\text{exp}}$ exposures can be incoherently averaged and hence the averaged SNR is given by

$$\text{SNR}_X(\tau_{\text{obs}}) = \text{SNR}_X(\tau_{\text{exp}}) \sqrt{\tau_{\text{obs}}/\tau_{\text{exp}}} \tag{6.1}$$

where $\text{SNR}_X(\tau_{\text{exp}})$ is the SNR of the estimator X (for example the power spectrum P or the bispectrum T) for a single exposure of length τ_{exp}.

The scaling of $\text{SNR}_X(\tau_{\text{exp}})$ with τ_{exp} will depend on whether photon noise or read noise is the dominant noise source. We will make the simplifying assumption that the exposures are in a low-light-level regime where $\text{SNR}_X(\tau_{\text{exp}}) \ll 1$ for all the exposure times considered. This is clearly the regime in which optimising the exposure time is most important.

For photon-noise-limited measurements in the low-light-level regime the SNR of the power spectrum given in Equation (5.40) is

$$\text{SNR}_P(\tau_{\text{exp}}) \propto \langle |\gamma(\tau_{\text{exp}})|^2 \rangle \overline{N}_{\text{phot}}(\tau_{\text{exp}}), \tag{6.2}$$

where $|\gamma(\tau_{\text{exp}})|$ is the reduction of the visibility of the fringe due to fringe smearing over an exposure of length τ_{exp} and $\overline{N}_{\text{phot}}(\tau_{\text{exp}})$ is the mean number of photons received during the exposure. Thus, since $\overline{N}_{\text{phot}}(\tau_{\text{exp}}) \propto \tau_{\text{exp}}$, the SNR after incoherently integrating for a time τ_{obs} is

$$\text{SNR}_p(\tau_{\text{obs}}) \propto \langle |\gamma(\tau_{\text{exp}})|^2 \rangle \sqrt{\tau_{\text{exp}}}. \tag{6.3}$$

The results from Section 3.3 can be used together with Equation (6.3) to yield a graph of the variation in SNR_P with τ_{exp} as shown in Figure 6.1. It can be seen from this graph that, as expected, the SNR increases with exposure time for $\tau_{\text{exp}} \ll t_0$, but that fringe-visibility losses overwhelm the increase in the number of photons collected in the exposure for $\tau_{\text{exp}} > 1.6t_0$. Thus an exposure time of around $1.6t_0$ is optimal for low light levels and a fixed overall observation time τ_{obs}.

This exposure time will also be optimal in background-limited observations, such as in the mid-infrared where the dominant source of noise is noise from the background photons rather than photons from the source itself. In read-noise-limited situations, the variation of SNR with exposure time has a different functional form. The SNR of the power spectrum is given by

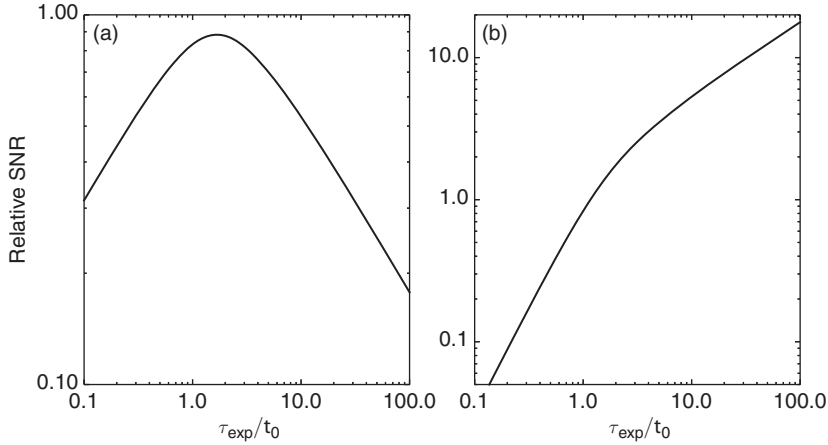

Figure 6.1 The SNR of the power spectrum as a function of integration time $\tau_{\rm exp}$ for photon-noise-limited measurements (a) and read-noise-limited measurements (b) in the low-light-level regime. The SNR is normalised such that a system with a perfect fringe tracker that freezes the fringe motion would give an SNR of unity for an exposure time of t_0.

$$\mathrm{SNR}_P(\tau_{\rm exp}) \propto \langle|\gamma(\tau_{\rm exp})|^2\rangle \overline{N}_{\rm phot}(\tau_{\rm exp})^2, \tag{6.4}$$

so the SNR after incoherent integration over a time $\tau_{\rm obs}$ is

$$\mathrm{SNR}_P(\tau_{\rm obs}) \propto \langle|\gamma(\tau_{\rm exp})|^2\rangle \tau_{\rm exp}^{3/2}. \tag{6.5}$$

Figure 6.1(b) shows a plot of this function. It can be seen that in the read-noise-limited case the averaged SNR always rises as the exposure time is increased, albeit more slowly for exposure times greater than a few t_0. In practice this rise does not continue indefinitely because at some point the number of photons collected during an exposure becomes large enough that the exposure is photon-noise-limited and/or the assumption that $\mathrm{SNR}_P(\tau_{\rm exp}) \ll 1$ becomes invalid. It must also be borne in mind that using exposure times such that $\tau_{\rm exp} \gg t_0$ means that the fringe visibility will be low due to fringe smearing, and this increases the danger that other effects such as the dynamic range of the detector, i. e. the ability to see small fluctuations in the intensity on a large background, may compromise the fringe measurement.

A similar analysis can be performed for the bispectrum showing that the exposure time which minimises the phase error in the photon-noise-limited case is about $2t_0$ (Buscher, 1988a).

6.2 The optimum aperture size

Few observers are able to choose what size of telescope to use for their observations. Nevertheless, it is instructive to understand the factors connected with the aperture size of the individual telescopes in an interferometer that allow the observation of faint objects. With a non-interferometric observation, using the largest available telescope is likely to guarantee the best faint-object performance, but the situation is more complex for an interferometer.

Using a larger aperture allows more photons to be collected from a given object, but also increases the root-mean-square (RMS) wavefront perturbation across the aperture due to atmospheric seeing. The level of wavefront perturbation can be reduced by use of an AO system, but there are practical limitations to the effectiveness of such a system. The main one, discussed below in Section 6.3, is that the AO system needs a bright reference source for wavefront sensing, and this bright source is usually the interferometric target itself. On fainter sources there is only enough light to sense the lowest-order modes of the atmospheric wavefront perturbations, so that the effective order of the AO system can be less than the maximum possible number of corrected modes that the AO system was designed for. In addition, AO systems are expensive and many interferometers only have tip–tilt correction available.

Thus, the use of larger apertures is usually associated with larger residual wavefront aberrations, which have the opposite effect on the SNR to the increase in flux collecting area. This is similar to the optimisation of the exposure time in the presence of temporal perturbations. There are, however, some differences. The first is that there is no incoherent averaging advantage to using smaller apertures as there is to using shorter integration times – this would correspond to using many parallel interferometers of smaller telescopes instead of a single interferometer with larger telescopes, and this is not usually an option within the budget constraints of most interferometric facilities.

An additional difference to the temporal optimisation problem is that spatial wavefront aberrations can have two types of effects on the SNR of the fringe measurements, depending on whether spatial filtering is used. If atmospherically perturbed beams are spatially filtered (using, for example, single-mode fibres), the fringe contrast can be increased at the expense of a compensating loss in the number of photons. To the author's knowledge, there is no temporal filtering equivalent (note that taking shorter exposures is not the temporal equivalent to spatial filtering but rather to using smaller apertures).

From Equation (5.37), the SNR of a power spectrum measurement for an aperture size D is given by

$$\mathrm{SNR}_P(D) = \frac{\left\langle |F_{ij}(D)|^2 \right\rangle}{\sigma_P(D)} \tag{6.6}$$

where $\left\langle |F_{ij}(D)|^2 \right\rangle$ is the mean-squared coherent flux in a fringe pattern formed using aperture size D and $\sigma_P(D)$ is the RMS noise on the power spectrum for an observation with aperture size D. Equation (6.6) needs to be evaluated in at least six possible regimes depending on the type of hardware present and the wavelength range: two regimes corresponding to a spatially filtered and non-spatially-filtered beam combination combined with three regimes corresponding to photon-noise-limited detection, read-noise-limited detection and background-limited detection.

The variation of the numerator of the fraction in Equation (6.6) with D is similar when either a spatially-filtered combiner or a non-spatially-filtered combiner is used. In the case of a spatially-filtered combiner, the mean-squared coherent flux can be derived using Equation (3.60). The coherent flux will scale as

$$\left\langle \left|F_{ij}\right|^2 \right\rangle \propto \left\langle |\eta_i|^2(D) \right\rangle \left\langle |\eta_j|^2(D) \right\rangle D^2, \tag{6.7}$$

where $\left\langle |\eta_i|^2(D) \right\rangle$ and $\left\langle |\eta_j|^2(D) \right\rangle$ are the mean coupling efficiencies into a fibre from an aperture of diameter D and the factor of D^2 corresponds to the scaling of the total light collected with telescope area. The mean coupling efficiencies are given in Figure 3.17 and will typically be the same for both telescopes.

For a non-filtered beam combiner, the mean-squared coherent flux will be given by

$$\left\langle \left|F_{ij}\right|^2 \right\rangle \propto \left\langle |\gamma_{ij}|^2(D) \right\rangle D^2, \tag{6.8}$$

where $\left\langle |\gamma_{ij}|^2(D) \right\rangle$ is the mean-squared visibility loss for an aperture of diameter D and is plotted in Figure 3.14. By comparing Equations (6.7) and (6.8), and Figures 3.14 and 3.17, it can be seen that in both the filtered and unfiltered beam combiners the coherent flux will rise rapidly with diameter for small D but will rise less steeply or even fall after a 'cut-off' value of D/r_0, which depends on the order of correction.

The denominator in Equation (6.6) depends on which source of noise is dominant but can also depend on whether a filtered or unfiltered combiner is used. If the observation is photon-noise-limited, then in the low-light-level regime the noise is proportional to the mean number of photons per exposure

$$\sigma_P(D) \propto \overline{N}_{\mathrm{phot}}(D). \tag{6.9}$$

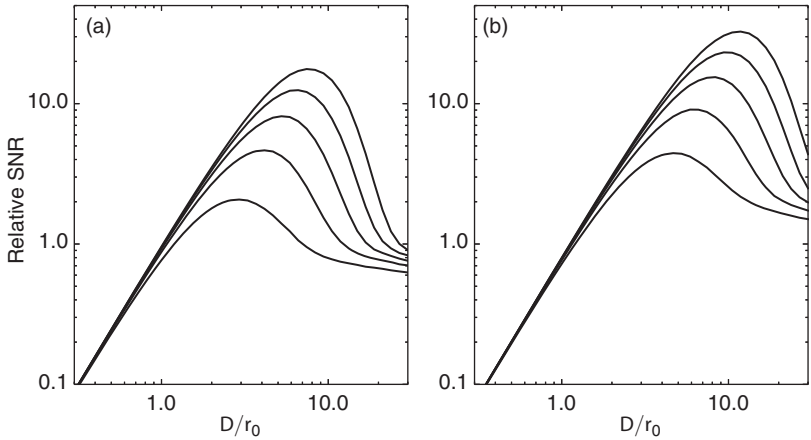

Figure 6.2 The SNR of photon-noise-limited fringe power spectrum measurements as a function of the interferometer aperture diameter D. The SNR is calculated for an interferometer using no spatial filter (a) or a single-mode spatial filter (b). Different lines show the effects of different radial orders of AO correction: the lowest line corresponds to the correction of Zernike wavefront modes up to radial order 1 (tip–tilt correction) and the uppermost line to a correction up to radial order 5. The SNR plotted is normalised to unity for a perfectly corrected aperture of diameter r_0, under the assumption that the source is sufficiently faint that the true SNR per exposure is much less than unity (the faint-source regime).

For an unfiltered beam combiner, the number of photons is given by

$$\overline{N}_{\text{phot}}(D) \propto D^2 \tag{6.10}$$

whereas for a filtered combiner the coupling efficiency reduces the number of photons present:

$$\overline{N}_{\text{phot}}(D) \propto \left\langle |\eta(D)|^2 \right\rangle D^2. \tag{6.11}$$

This means that the noise level is lower in the spatially filtered beam combiner compared to an unfiltered combiner, particularly for larger apertures where the coupling efficiency is poorer. This is borne out by the graphs of SNR versus diameter derived using Equations (6.6)–(6.11) and shown in Figure 6.2. It can be seen that in all cases there is an optimum diameter to use for a given level of AO correction. Above this diameter using a larger telescope actually decreases the SNR of the fringe measurements. In systems using spatial filtering, this optimum diameter is larger and the SNR at this optimum diameter is also larger, so if large telescopes are to be used, spatial filtering is of significant benefit.

At mid-infrared wavelengths and for the observation of very faint sources at shorter wavelengths, the dominant source of noise is the photon noise from thermal or other background radiation sources. The scaling of the noise from these sources with aperture diameter is different from the scaling for photon noise from the source. This is because, as the diameter of the telescope is increased, the increase in collecting area of the telescope is exactly compensated for by a corresponding decrease in the angular size of the diffraction-limited 'patch' on the sky from which radiation is received. In the case of a spatially filtered combiner the patch which couples into the spatial filter is almost exactly the same size as the diffraction limit of the telescope, whereas any well-designed but non-filtered combiner will use a focal plane 'cold stop', which is a few times the diffraction limit in size to reduce the amount of background radiation reaching the detector. This means that $\overline{N}_{\text{phot}}(D) \propto$ and hence $\sigma_P(D)$ is roughly independent of D.

A noise level that is independent of D is also seen when the fringe detection is read-noise-limited in the low-light-level regime. The resulting graphs of SNR versus D are shown in Figure 6.3. It can be seen that the behaviour in this case is quite different to when the main noise source is photons from the object itself. In the spatially filtered case, the SNR is maximised at a finite diameter as in the photon-noise-limited case. However, in the unfiltered case, the SNR can show a peak at some critical value of D, but going to even larger values of D causes the SNR to start to rise again, albeit more slowly, and thus the unfiltered system performs better at the very largest diameters.

The situation is slightly more complex in the background-limited case, as for very large values of D/r_0 the correction of the wavefront errors will be sufficiently poor that the majority of the flux will not be concentrated in a diffraction-limited core. Instead, the majority of the flux will be in a seeing-limited 'halo', and as a result the angular diameter of the cold stop needed to collect all the light from the source will need to be of order r_0/λ rather than a few times D/λ. This will increase the amount of background light allowed through the cold stop and means that for the largest apertures the SNR behaviour for the unfiltered combiner will be more like the source-photon-noise-limited case shown in Figure 6.2.

In all cases, the SNR benefit of using larger apertures is lower if the level of AO correction available is limited. As discussed in Section 6.3 this limitation is set not only by the expense and complexity of using higher-order AO systems but also by the light available for wavefront sensing, and this is particularly relevant for faint objects.

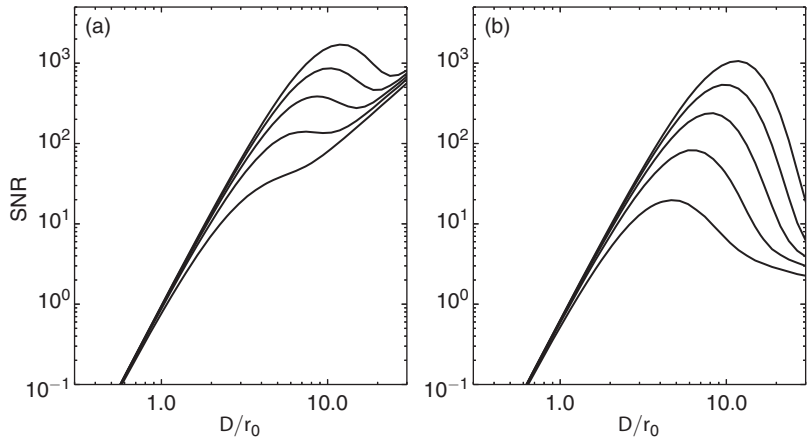

Figure 6.3 The SNR of read-noise-limited (a) or background-noise-limited (b) fringe power spectrum measurements as a function of the interferometer aperture diameter D. The SNR is calculated for an interferometer using no spatial filter (a) or a single-mode spatial filter (b). All other details are as in Figure 6.2.

6.3 AO on faint objects

The limitations of AO systems when observing faint targets arise from the limitations of wavefront sensing with faint reference sources. There are a number of different types of wavefront sensor used in AO, but for simplicity a single type of wavefront sensor will be used to illustrate these limitations. The Shack–Hartmann (or Hartmann–Shack sensor – there is some controversy about the naming) wavefront sensor consists of an array of lenslets placed at an image of the telescope pupil. These lenslets break the aperture of the telescope into a contiguous set of subapertures, and the light from each subaperture is focussed onto a different area of a detector, to form an array of spots as shown in Figure 6.4.

In the presence of an unabberated wavefront, all the spots will be on-axis for their respective lenslets, but in the presence of an aberrated wavefront the x and y location of the spots will be displaced by an amount which depends on the local tip and tilt across the subaperture sampled by the corresponding lenslet. These displacements therefore correspond to samples of the local wavefront gradient, and these gradient measurements can be integrated to form an estimate of the complete wavefront across the aperture. One way of performing this integration is to use the wavefront slope measurements to derive the coefficients of the Zernike modes and then to sum these modes together in the proportions given by the estimated coefficients.

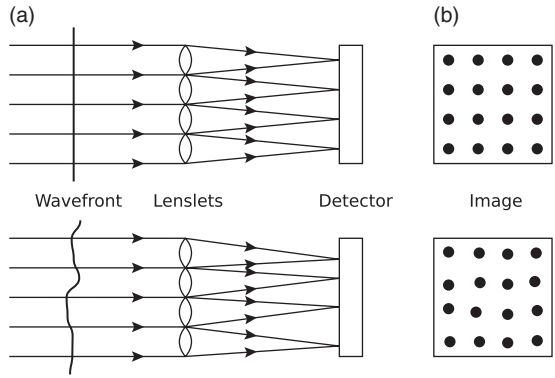

Figure 6.4 Schematic of a Shack–Hartmann wavefront sensor consisting of an array of lenslets (a) and the spot pattern seen on the detector (b). The upper row is for an unaberrated wavefront, which yields a regular spot pattern corresponding to the centres of the lenslets, while the lower row illustrates that the spots are displaced when the lenslets are illuminated with an aberrated beam. Diagram adapted from 'Adaptive optics tutorial at CTIO' by A. Tokovinin, `http://www.ctio.noao.edu/~atokovin/tutorial/`.

The larger the aperture to be corrected, the higher-order the AO correction needs to be in order to be effective (i. e. to reduce the residual wavefront phase aberrations to an acceptable level, e. g. of order a radian or less). A rule of thumb is that an aperture of size $D \gg r_0$ can be corrected using approximately $(D/r_0)^2$ degrees of freedom, so for example, a 4-m-diameter telescope would require the correction of about 1600 modes in order to give adequate correction when $r_0 = 10\,\text{cm}$. There must be at least as many wavefront slope measurements as modes to be corrected and this leads to the result that the subapertures of the wavefront sensor need to be less than or of the order of r_0 in diameter.

The wavefront needs to be sampled before it has had a chance to change by of order a radian, and so the integration time needs to be less than about t_0. Thus the light available for each Shack–Hartmann spot is just the light which can be collected from the reference source over a patch of area approximately r_0^2 and during an integration time of t_0. Typically, about 100 photons are needed per spot in order to estimate the wavefront slope with adequate accuracy. As a result, a reference source of a certain minimum brightness is needed in order to give a certain minimum level of AO correction, and, importantly, this brightness is independent of the size of the telescope.

An example of this effect can be gleaned from the performance parameters of a typical AO system. A common measure of the performance of such a system is called the *Strehl ratio*. This measures the intensity of the light at

the centre of the image of a point source seen through telescope and AO system, and compares this with the intensity that would be seen with perfect AO correction, i. e. a diffraction-limited image. In the AO system for the Keck II telescope (van Dam *et al.*, 2007), the Strehl ratio of the images at a wavelength of 2.2 μm is about 60% in good seeing (characterised by $r_0 = 20$ cm at a wavelength of 0.5 μm) when guiding on a bright reference source (a source brighter than 8th magnitude at the wavefront sensor wavelength, which is roughly in the R band centred around 600 nm). When guiding on a source which has an R-band magnitude of 14 the Strehl ratio is about 40% in the same conditions and at R=15 this ratio has dropped to about 15%.

At shorter science wavelengths the effect is more pronounced, in that the Strehl ratio typically begins to fall off even for much brighter reference sources. This is because r_0 and t_0 are smaller at shorter wavelengths so less light is available per subaperture and per exposure of the AO system.

6.3.1 Anisoplanatism

The reference source is often the science target itself, but if the science target is not bright enough then another reference source can in principle be used to drive the wavefront sensor. This reference source could be a star which is nearby in the sky to the science target. However, Figure 6.5 illustrates the problem with using an off-axis reference: the wavefront perturbations encountered by the light propagating from the reference source are not the same as those encountered by the science target and so the wavefront correction will not be perfect.

The larger the angle between the reference source and the larger the height of the turbulence layer which is causing the seeing, the larger the difference between the reference and target perturbations will be. This variation of the wavefront perturbations across the field is called *anisoplanatism* and the separation between two stars which experience a wavefront perturbation difference of 1 radian RMS is called the *isoplanatic angle*. It can be seen from the figure that this will occur when the separation of equivalent rays from the target and reference is of order r_0 at the height of the relevant layer of turbulence. Thus, the isoplanatic angle is given approximately by

$$\theta_{\text{isoplanatic}} \sim \frac{r_0}{h}, \tag{6.12}$$

where h is the height of the dominant turbulent layer. Taking a typical value for h of 5 km and for r_0 of 50 cm (corresponding to good seeing and a near-infrared science wavelength) then $\theta_{\text{isoplanatic}} \sim 10^{-4}$ radians or about 20 arcseconds.

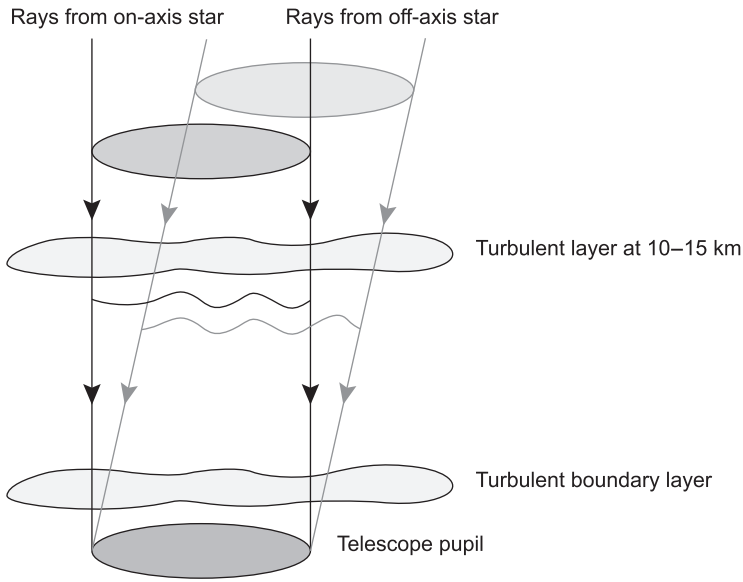

Figure 6.5 Illustration of the effects of angular anisoplanatism.

In order to provide acceptable correction, a bright star needs to be found within the *isoplanatic patch* around the target. The probability of such a star happening to be within the isoplanatic patch around a randomly selected target is about 50/50 if the isoplanatic patch is an arcminute in diameter and if reference stars with an R magnitude of 15 are considered 'bright'. However, both these conditions are rarely fulfilled and so for most interferometric targets, the best wavefront reference source is usually the science target itself.

6.3.2 Laser guide stars

An alternative to using off-axis reference stars is to use an artificially created reference star. This is the so-called *laser-guide-star* technology as shown in Figure 6.6, which relies on focussing a laser at some relatively high altitude in the atmosphere and using the light which is scattered back from the focus to sense the atmospheric wavefront perturbations.

Using a laser guide star overcomes the problem of finding a bright enough 'natural' guide star, but brings its own problems. The technology for laser guide stars is expensive and the wavefront correction provided by laser-guide-star techniques is imperfect in two ways. First, the lowest-order wavefront perturbations, tip–tilt and defocus, cannot be easily sensed using an LGS and

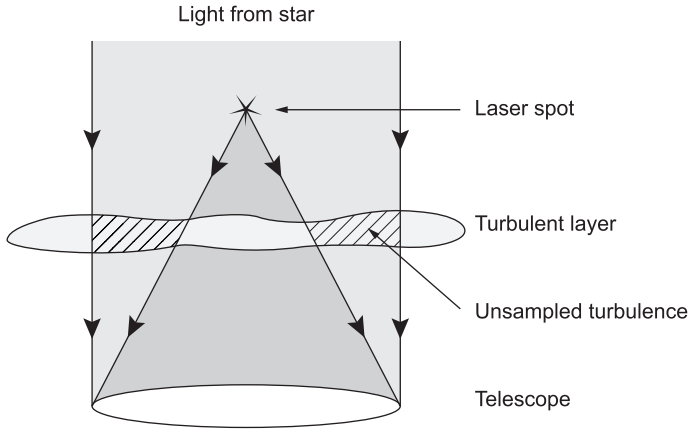

Figure 6.6 Cone effect for laser guide stars.

so an natural-guide-star system is needed in addition to the laser-guide-star system in order to correct for these terms. However, the low-order correction system does not need as many photons to provide the correction, and so fainter natural guide stars can be used.

A second problem of wavefront correction with laser guide stars arises because the guide 'star' is not at infinity. The wavefront perturbations seen on this reference will not be the same as those experienced by the target due to what is known as the *cone effect* or *focus anisoplanatism*. As shown in Figure 6.6, the backscattered light from the laser guide star samples a cone of the atmosphere, while the light from the science target samples a cylinder of the atmosphere. The differences between the turbulence sampled in these two volumes lead to differences in the wavefront perturbations, which grow worse as the telescope grows larger and causes the vertex angle of the cone to increase.

Currently, no interferometer is equipped with laser-guide-star AO on all the telescopes and so the majority of interferometric targets are observed using the target itself as the AO reference. When the reference is faint, the wavefront correction of large telescopes is less than perfect, and this provides a limit as to the faintness of the target which can be observed.

6.4 Fringe-tracking limits

Fringe trackers provide correction of temporal variations in the piston mode of the wavefront. Cophasing fringe trackers stabilise the fringes to allow the

science fringes to be integrated for long periods but, just as with an AO system, the fringe tracker itself must have enough light to follow the fringes.

In the case of the fringe tracker, the wavefront sensor consists of a beam combiner which forms fringes and this fringe sensor unit will be prey to the same sorts of SNR issues when observing faint sources as any other interferometric instrument. One of these issues is sensitivity to spatial wavefront errors: any residual wavefront errors which are not corrected by the AO system cause either a loss of coupling efficiency of the starlight into the fringe tracker (for spatially filtered fringe trackers) or a loss in fringe visibility.

The fringe tracker must employ a minimum exposure time in order to measure the fringe motion. The shorter the exposure time on the fringe tracker, the less fringe motion there will be between exposures, but the less light is available to measure the fringe position and hence the greater the effects of noise on the fringe measurement. As a result there is an optimum exposure time for a given fringe-tracker light level, which balances the sensing noise against undersampling of the high-frequency fringe motion.

At a sufficiently low level, fringe tracking becomes ineffective because the fringe motion during the exposure time required to get sufficient SNR is too large. This light level is different for different types of fringe tracker. Two common types of fringe tracker are known as 'phase-tracking' and 'group-delay' fringe trackers, respectively, and these are discussed separately below.

6.4.1 Phase tracking

Measurements of the atmospheric optical-path-difference (OPD) perturbations in long-baseline interferometers such as those shown in Figure 3.4 show that the fringe phase excursions in any interferometer is likely to be many times 360°. Most cophasing fringe trackers rely on a technique known as 'phase unwrapping' or phase tracking to track such large excursions. In phase unwrapping, the fringe phase is sampled sufficiently often that the magnitude of the phase change between subsequent measurements due to atmospheric OPD fluctuations is never more than 180°. If this condition is achieved, then the atmospheric phase changes can be estimated by taking the difference between the fringe phases measured in consecutive exposures. The phase difference will be ambiguous to an integer multiple of 360° but the correct value of the phase difference can be inferred by choosing the value whose magnitude is less than 180°. By adding together these phase differences over multiple exposures, the phase change can be followed over arbitrarily many 360° cycles and so the science fringes can be stabilised.

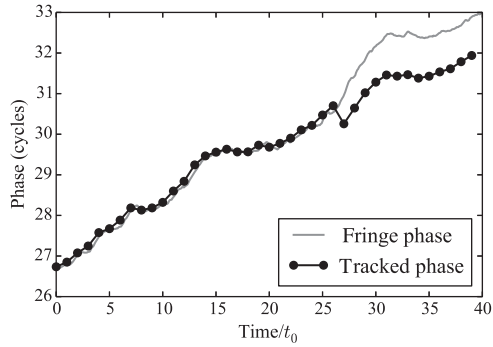

Figure 6.7 A sample fringe track showing a phase jump, which occurs when the fringe phase is changing rapidly compared to the sample time of the fringe sensor.

The maximum fringe-tracker exposure time allowable in order for phase unwrapping to work reliably is of order $t_0/2$ (Buscher, 1988b). If the exposure is longer than this then the phase shift between subsequent exposures occasionally exceeds 180° and an incorrect determination of the direction of fringe motion is made. If this happens the fringe phase estimated using the phase-unwrapping technique 'jumps' by 360° in a random direction as shown in Figure 6.7.

The fringe tracker will attempt to correct for the apparent phase jump by adjusting the internal delay by one fringe-tracker wavelength. If the science combiner is operating at a different wavelength, the science combiner will see a phase jump which is not a whole number of wavelengths. If the jump occurs during an exposure on the science combiner then the variation of the fringe phase will smear the fringe and cause a loss in fringe visibility.

A potentially more serious problem is that a series of phase-unwrapping errors will act as a random walk, tending to make the interferometer OPD wander away from the original tracked OPD value. Any polychromatic fringe pattern will have a fringe-visibility envelope (see Section 1.7), which decreases in modulus as the OPD moves away from the zero-group-delay location (see Section 1.8). Wandering away from the zero-group-delay location will therefore cause the modulus of the fringe visibility in the fringe tracker (as well as the science combiner) to decrease. Eventually, the SNR of the fringes in the fringe tracker will be so low that the fringe tracker will lose the fringes altogether and cease providing effective fringe tracking.

At low light levels, the SNR of the fringe-tracker phase measurements will be limited by detection noise. This noise will be significant because of the short exposure times used to avoid jumps in the phase unwrapping. At a sufficiently

low light level, the noise in the phase measurements will be large enough to cause additional fringe-unwrapping errors, leading to similar consequences to those caused by using too long an exposure time. This occurs when the SNR of the fringe sensor measurements is less than about 2 for a single exposure (Buscher, 1988b). Sources which are fainter than this are problematic to fringe track using phase-unwrapping methods because fringe jumps are frequent.

Conversely, if the source being observed is brighter than this limiting magnitude, then the fringe tracking will likely to be accurate at the sub-radian level, as low phase noise is essential to avoid unwrapping errors.

6.4.2 Group-delay tracking

Phase tracking fails on faint sources as a result of the 360° ambiguity of the fringe phase, which means that an unwrapping error in a fringe-sensor exposure cannot be easily be recovered from in subsequent exposures. An alternative measurement that does not have this ambiguity is the position of the envelope of fringe visibility for a polychromatic fringe packet, as described in Section 1.7. In most circumstances the fringe envelope will have a peak at a single location in 'delay space' corresponding to the location of zero net group delay at the wavelength corresponding to the centre of the bandpass. If this location can be tracked then atmospheric disturbances of arbitrary magnitude can be followed, without requiring the disturbances to be tracked on timescales short enough so that the disturbances are much less than a wavelength.

The conceptually simplest way to track this envelope is to scan the delay lines rapidly backwards and forwards and to measure the fringe-visibility modulus as a function of delay as shown in Figure 6.8. The peak of the visibility envelope can then be found and the centre of the delay scan adjusted to keep the fringe envelope in the centre of the scan.

In order to be effective, the delay scan must be larger than the fringe envelope. As a result, much of the observation time is spent in regions of delay space where the fringe visibility, and hence the SNR of the fringe measurement, is low. An alternative technique for finding the fringe-envelope peak, which has better SNR in general, is known as *group-delay tracking*.

The group-delay method can be thought of as the inverse of a Fourier-transform spectrograph: instead of measuring the spectrum of an object by scanning a delay and observing a fringe envelope, the group delay uses spectrally dispersed measurements of fringes to reconstruct the fringe-visibility envelope as a function of delay. The implementation of the group-delay technique takes a number of forms, but a convenient way to visualise how it

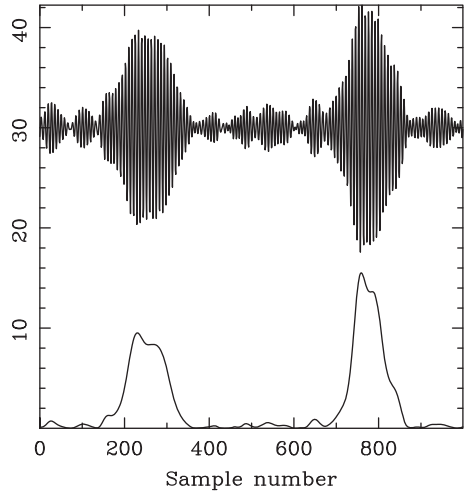

Figure 6.8 Fringe-envelope scans from the COAST interferometer. The change in OPD is approximately 60 μm during each 100-ms scan. The top trace shows the intensity as a function of time for two scans across the fringe envelope, and the bottom trace is the inferred envelope, which can be used to find the zero-OPD offset. From Thureau *et al.* (2003).

works is to consider a spectrally dispersed one-dimensional fringe pattern on a two-dimensional detector as shown in Figure 6.9(a).

The exact layout of the fringe pattern on the detector will depend on the design of the beam combiner, but in this example the x coordinate is a linear function of the delay difference τ between the beams interfering at that point on the detector and the y coordinate is proportional to the wavelength. The fringe pattern which would be seen on a target would be a sinusoidal pattern in the x direction at each wavelength whose spatial frequency is proportional to wavelength. For an unresolved source at the phase centre and with no net instrumental or atmospheric delay errors the peaks of all the sinusoids are aligned with each other at the centre of the fringe pattern, as shown in Figure 6.9.

If the fringe pattern intensities are remapped onto a new set of coordinates ($\phi \propto x/\lambda, \nu \propto 1/\lambda$) then all the horizontal fringe patterns will have the same spatial frequency and the fringe pattern with no delay errors will appear as a vertical sinusoidal pattern as shown. If an instrumental delay error τ_{ij} is introduced on baseline ij then the fringe patterns will shift in phase by an amount which depends on wavelength as

$$\phi(\nu) = \tau_{ij}\nu. \tag{6.13}$$

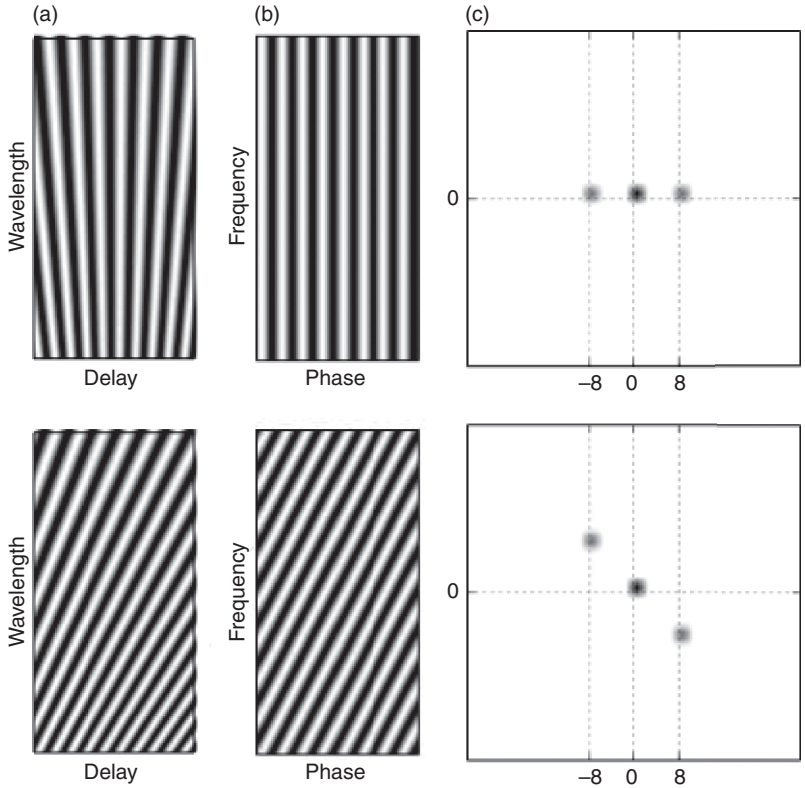

Figure 6.9 Spectrally dispersed fringe patterns (a), the fringe patterns remapped so that the fringe crests are parallel (b) and the power spectra of the remapped fringes (c). The upper row corresponds to zero OPD while the lower row corresponds to a finite OPD, which causes the fringe phase to change linearly as a function of wavelength.

This linear phase change with frequency ν appears as a change in the slope of the fringes as shown in the lower row of Figure 6.9. Thus, by determining the slope of the fringe phase with frequency the delay offset can be deduced.

The slope of the fringe crests is easy to identify 'by eye' at high light levels from the intensity patterns in Figure 6.9. At low light levels the slope can be determined by taking a two-dimensional Fourier transform of this pattern. The magnitude of the Fourier transform will have a peak at the origin corresponding to the 'DC' level of the pattern and a symmetric pair of peaks corresponding to the sinusoidal intensity modulation. As the slope of the fringes changes, the peak will move along a line with constant spatial frequency in the x direction. Taking a slice of the Fourier transform along this line will show a peak in

magnitude whose position depends on the delay error τ_{ij}. The group-delay estimate is then an estimate of τ_{ij} based on the estimated position of the peak.

This group-delay method uses all the flux received at all wavelengths simultaneously and so works well at low light levels. An alternative way of viewing the group-delay estimation process is one determining a best-fit value of the delay τ_{ij} as a two-step process (Basden and Buscher, 2005). In the first step (corresponding to Fourier transforming in the x dimension and taking the Fourier component corresponding to the fringe frequency) the fringes at each wavelength are analysed to yield a set of coherent flux estimates for all wavelength channels $\left\{\hat{F}_{ij}(\nu_k), k = 1, \ldots, N_{\text{chan}}\right\}$. In the second step a set of trial delays $\left\{\tau_{ij,m}, m = 1, \ldots, N_{\text{delay}}\right\}$ is generated, and for each trial delay the coherent flux estimates are 'phase rotated' to compensate for the assumed delay and then added together to give the coherent flux of the 'synthetic white-light fringe' at that delay, given by

$$F_{ij,m} = \sum_{k=1}^{N_{\text{chan}}} F_{ij}(\nu_k) e^{2\pi \mathrm{i} \nu_k \tau_{ij,m}}. \tag{6.14}$$

This latter step corresponds to a Fourier transform over the y dimension if the channels are sampled evenly in frequency. In the absence of noise and for an unresolved source the amplitude of the white-light fringe $|F_{ij,m}|$ is maximised when the trial delay equals the true delay. This is equivalent to finding the trial delay which best 'unwraps' the linear change of phase with frequency caused by the delay error τ_{ij}.

The precision with which the position of the peak in delay space can be determined is related to the SNR of the data and the width of the group-delay peak. The width of the peak is given approximately by

$$\Delta\tau \sim (\Delta\nu)^{-1}, \tag{6.15}$$

where $\Delta\nu = \nu_{\max} - \nu_{\min}$ is the total bandwidth of all the wavelength channels combined.

This result can be derived by considering Equation (6.14) as Fourier transform of the coherent fluxes over an infinite bandwidth multiplied by a 'top hat' of width $\Delta\nu$. Comparing this to the results in Section 1.7, it can be seen that the width of the peak is also the width of the fringe envelope of the synthetic white-light fringe, i. e. the fringe formed by summing pixels along the wavelength direction in the dispersed fringe pattern.

If $\Delta\nu \sim \nu_0$, where ν_0 is the central frequency of the spectral channels, the width of the peak is comparable to a wavelength and so precisions in the delay estimate of this order are possible. In order to get higher precision, the phase

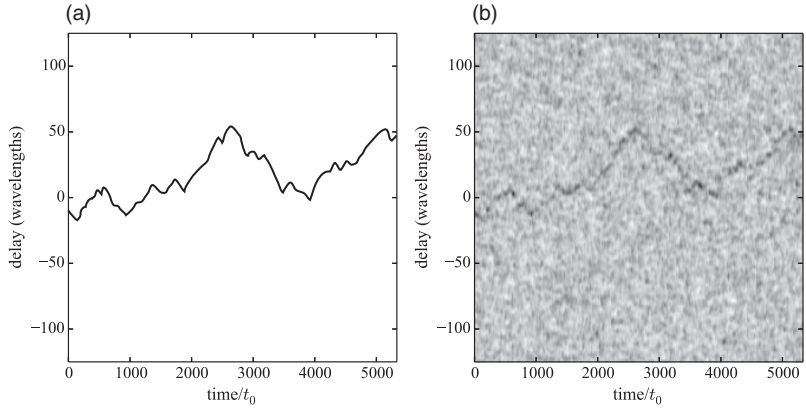

Figure 6.10 Simulated group delay signal at high light levels (a) and at a light level corresponding to the lowest SNR at which the fringe envelope can be reliably tracked (b). Each column of pixels corresponds to the time-averaged group-delay fringe power for different trial delays at a given instant in time.

of the synthetic white-light fringe phasor given in Equation (6.14) can be used to follow the sub-wavelength motion of the fringe, in a similar way to that used in phase tracking.

At lower light levels, group-delay tracking is generally used for coherencing and not cophasing.

The accuracy requirements for coherencing are lower and so the signal can be averaged for longer. In order to make peak identification more reliable the power spectrum as a function of trial delay can be integrated incoherently over many exposures in order to increase the SNR.

Simulations of group-delay tracking (Basden and Buscher, 2005) show that the group-delay power spectrum can be incoherently integrated for 20–50t_0. Figure 6.10 shows that the group-delay signal can be reliably extracted when $\mathrm{SNR}(\hat{F}_{ij}) \gtrsim 0.6$.

In Section 6.4.1 it was stated that phase tracking is possible when $\mathrm{SNR}(\hat{F}_{ij}) \gtrsim 2$. Taking into account the different exposure times ($0.5t_0$ for phase tracking and $1.6t_0$ for group-delay tracking), the corresponding different visibilities and the scaling of $\mathrm{SNR}(\hat{F}_{ij})$ with the number of photons in photon-noise-limited observations, this implies that group-delay tracking can be used on objects more than ten times fainter than the objects which are at the limit for phase tracking.

As a result of being able to work on targets which are considerably fainter, group-delay tracking can be used on a far wider range of targets than phase tracking. The fact that group-delay tracking only does coherencing and not

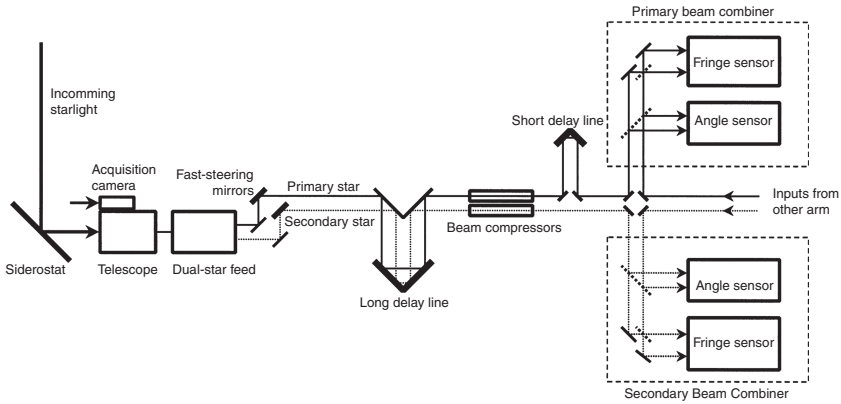

Figure 6.11 Schematic design of the dual-star system in the PTI. Two parallel beam combiners measure fringes simultaneously on a primary and secondary star. From Colavita *et al.* (1999).

cophasing at low light levels implies that the majority of faint targets will be observed using multiple short exposures on the science camera rather than relying on a cophasing fringe tracker to allow long coherent integrations.

6.4.3 Off-axis fringe tracking

The fact that fringe tracking only works on relatively bright targets restricts the range of targets which can be observed. A potential way around this restriction is to use off-axis stars as references for the fringe tracker, as is done in AO. Systems which accomplish this effect are known as 'dual-star' or 'dual-feed' systems.

A conceptual outline of a dual-star system is shown in Figure 6.11. The light from two different stars with angular separations of many arcseconds are split by a star separator at the focal planes of the unit telescopes and sent down parallel paths to two separate beam combiners. The brighter of the two stars is used to sense the atmospheric piston variations and these are used to provide corrections to a delay line which is common to the optical paths of both beam combiners, thereby stabilising the fringes in both combiners.

Dual-feed systems incorporate differential delay lines, which remove the additional geometrical delay difference between the two combiners due to the fact that the two stars are at substantially different locations in the sky

The dual-feed system implemented in the PTI was not primarily intended to allow the scientific observation of faint targets. The main reason for this is that, in order to provide good cophasing of a faint target, there needs to be a bright

enough reference star within the isoplanatic patch of the science target to drive the fringe tracker. Fringe trackers typically require brighter references in order to operate than AO systems, and so the probability of finding bright-enough fringe-tracking references is even lower than for finding wavefront references for an AO system, perhaps less than 1% in a typical interferometer.

Instead, dual-star systems are typically aimed at astrometric science targets on bright nearby stars. These science targets are used to drive the fringe tracking, and the second star is a faint star which is used as an astrometric reference. The cophasing provided by the target star serves to allow the SNR of the faint reference fringes to be increased by using long integrations. Faint reference stars are plentiful and therefore are not a problem to find within the isoplanatic patch.

6.4.4 Laser guide stars for fringe tracking

There is currently no equivalent of a laser guide star for single-telescope AO which can be used to provide an artificial signal for fringe tracking in interferometry. There are a number of reasons for this. One of these is the difficulty of providing a sufficiently small laser spot (milliarcseconds in size) so that the fringes from the spot have high contrast.

A second and more serious problem is that the 'cone effect' is difficult to overcome for long-baseline interferometers because the severity of the cone effect becomes larger, the larger the ratio between the baseline and the height of the laser spot becomes. Techniques for overcoming the cone effect for large telescopes, such as the use of multiple laser guide stars, rely on being able to 'map out' the turbulence as a function of altitude across the whole aperture, and this is difficult to do when the aperture consists of isolated patches as it does in an interferometer.

In the near term, the best hope for the use of laser guide stars in interferometry is to correct the non-piston atmospheric aberrations across each of the apertures of the array. If large-enough apertures can be corrected in this way, this will allow sufficient light to use fainter targets as natural references for fringe trackers aimed at stabilising the piston mode.

6.4.5 Bootstrapping

The previous sections show how important the SNR of fringe measurements are for successful fringe tracking and therefore for successful observations of a given target. This SNR depends on both the number of photons per exposure and on the fringe visibility. The latter is affected by both instrumental

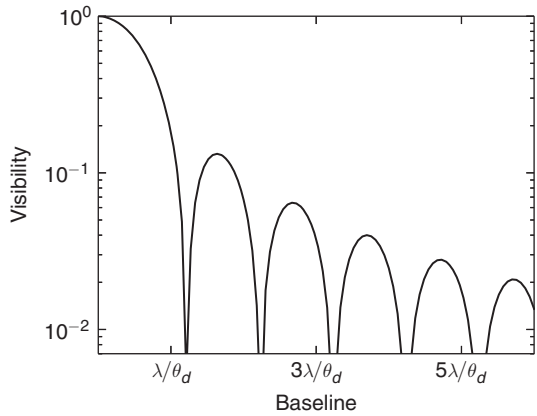

Figure 6.12 The object visibility modulus as a function of baseline when observing a uniform disc.

effects and by the visibility function of the object being observed. The object visibility function is a generally decreasing function of baseline length, so the SNR for fringe tracking will also decrease with baseline length. This can cause problems for fringe tracking on long baselines.

As an illustration of this problem we can consider the observation of features on the surface of a star, for example convection cells on the surface of a supergiant star such as Betelgeuse. To first order, the stellar surface can be modelled as a uniform disc of angular diameter θ_d, and so the fringe visibility as a function of baseline length will be an Airy function, depicted on a log scale in Figure 6.12.

In order to be sensitive to structures on the disc surface which are ten times smaller than the disc (in other words to be able to make a 10×10 resel image of the disc), the spatial frequency sampled by the longest baseline used must have at least five peaks and five troughs across the stellar disc. This corresponds to a longest baseline, which is at least $5\lambda/\theta_d$ in length. From Figure 6.12 it can be seen that the object visibility function in this region has a maximum of about 2%.

In the photon-noise-limited regime, the SNR is a monotonic function of $V\sqrt{\overline{N}_{\text{phot}}}$ and so the 50-fold reduction in fringe contrast compared to an unresolved source has the same effect on the SNR as a reduction in the photon rate by a factor of about 2500. In other words, over-resolving the stellar disc by this factor is equivalent in SNR terms to exchanging 8-m diameter telescopes for 16-cm telescopes. As a result, fringe tracking on such objects may fail on the long baselines when it works adequately on shorter baselines.

The net result is that it may not be possible to observe objects on the long baselines where the object shows interesting structure. What counts as a 'long' baseline for these purposes depends entirely on the size of the object under study. In the case of Betelgeuse with an angular diameter of $\theta_d = 40\,\text{mas}, 5\lambda/\theta_d = 12.9\,\text{m}$ and this is comparable to the *shortest* baselines available on many interferometers.

The situation of not being able to track fringes on baselines needed to resolve interesting structure is more common in optical interferometry than in radio interferometry. The main reason is that dependence of the SNR on the visibility is weaker in the background-noise-limited regime, which is prevalent at radio wavelengths, than in the photon-noise-limited regime, which can be encountered at optical wavelengths.

A secondary reason is due to the typical spatial structure of the targets at the two wavelengths. Many targets at radio wavelengths, for example radio galaxies, contain a bright unresolved 'core' surrounded by a much larger region containing interesting structure. The presence of the core means that the visibilities remain high even on baselines which are long enough to resolve the rest of the structure in the object. At optical wavelengths, many of the sources being observed have a structure where majority of the emission comes from an object, such as a stellar surface, which is on the same scale as the interesting features. Therefore, when the baseline is long enough to see detail, the 'core' is also resolved – such objects can be termed 'resolved-core' objects. Nevertheless, there are many objects which have a brightness structure at optical wavelengths that are like the 'compact-core' radio targets, for example extended discs around stars, and for such objects fringe tracking on long baselines is less of an issue.

There are a number of ways around the fringe-tracking limitations which occur when observing resolved-core objects. One technique, known as *wavelength bootstrapping*, is to track fringes at a different wavelength from the science wavelength, where the fringe visibility is higher on the same baseline.

If the object has roughly the same apparent size at all wavelengths, then the fringe visibility observed on a given baseline will tend to be greater at longer wavelengths than at shorter wavelengths. For example, using a fringe-tracking wavelength of 2 μm and a science wavelength of 500 nm would be advantageous, as the (u, v) coordinate sampled by the fringe tracker would be four times smaller than the (u, v) coordinate sampled by the science instrument. Thus, the fringe tracker could be inside the first lobe of the visibility function of a uniform-disc object such as Betelgeuse while the science instrument is sampling the fourth lobe at much lower visibilities.

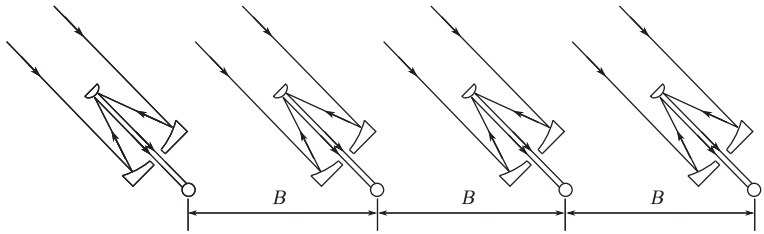

Figure 6.13 A set of telescopes arranged in a baseline-bootstrapping configuration.

Alternatively, the fringe tracker can be used at wavelengths where the apparent object size is much smaller and so unresolved. For example, when observing a star with a disc of warm dust around it, the disc may only be visible at mid-infrared wavelengths. A fringe tracker operating at a 2-μm wavelength might only see the star itself and would have high-visibility fringes if the star is unresolved. At the same time, a mid-infrared science instrument operating at 10 μm would see the disc emission and hence potentially resolve structure in the disc.

In both cases the science beam combiner would see quite low visibilities and hence the SNR in a short exposure would be low. However, because the fringe tracker is seeing higher visibilities it is able to provide effective fringe tracking, and the fringe measurements on the science camera can be integrated (coherently if the fringe tracker is a cophasing tracker and incoherently if not) over long periods to provide useful SNR data.

Another technique for allowing fringe tracking on resolved-core sources is called *baseline bootstrapping*. This requires an array of telescopes arranged in a 'chain' of short baselines making up a longer baseline such as that shown in Figure 6.13. Fringes can be measured on all the short baselines because the source is unresolved on these baselines and so the fringe visibilities are high. The fringe motions on the short baselines can be used to drive piston actuators to compensate for the piston phase differences between adjacent telescopes. As a result of this, the phase differences between all pairs of telescopes in the chain will also be compensated, and so fringe tracking on the shortest baselines provides fringe tracking on the longest baselines as a natural byproduct.

Baseline bootstrapping is most effective if the nearest-neighbour telescopes are as close together as possible to allow the highest possible visibilities for fringe tracking while keeping the longest baselines as long as possible, in order to resolve interesting structure in the target. These competing factors

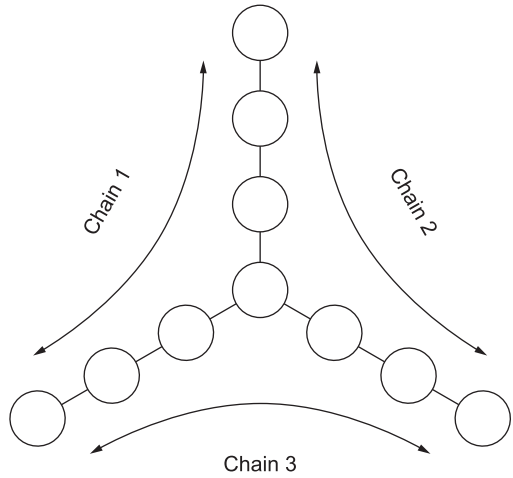

Figure 6.14 The baseline bootstrapping layout of the MROI.

tend to favour an array of equi-spaced telescopes, a 'redundant' arrangement which is less than optimal for covering a wide diversity of (u, v) plane spacings.

For a linear chain of telescopes, the ratio of the longest to the shortest baseline increases linearly with the number of the telescopes, whereas in a non-redundant array this ratio can increase approximately as the square of the number of telescopes. This is less of a problem in two-dimensional arrays of telescopes, as even with an optimal non-redundant design, which covers the (u, v) plane evenly, the ratio of the longest to the shortest spacing can at best grow linearly. Figure 6.14 shows a bootstrapping layout used in the MROI. The ten telescopes are arranged in an equispaced Y-shaped configuration, so the array can be viewed as a set of three 'bootstrapping chains' each consisting of seven telescopes, oriented at 120° angles to one another. The (u, v) coverage of the array is shown in Figure 2.6; only 20% of the baselines are redundant in this case.

6.5 Faint-object limits for interferometry

The above analysis has confirmed that atmospheric phase perturbations serve to reduce the signal-to-noise ratio of interferometric measurements on faint objects. The effects of the atmospheric perturbations can be ameliorated using active systems such as AO and fringe tracking, but these themselves do not

work well when the object under study is faint. As a result, for each of these major interferometer subsystems, whether the AO system, the fringe tracker, or the science instrument, the performance of the subsystem becomes unacceptable when the object being observed is below some limiting faintness, known as the magnitude limit.

If the performance of any one system falls, it can bring down the performance of the other systems. For example, poor performance of the AO system will lead to large residual wavefront perturbations and this will in turn bring down the performance of the fringe tracker and the science camera, potentially making the observation impossible.

Which subsystem limits the ability to perform a given observation depends both on the design of the subsystems and also the colour of the object being observed. Each subsystem typically takes light from different wavelength ranges; for example, AO systems often take visible-wavelength light while fringe trackers and science instruments might both be at infrared wavelengths. Many interferometric targets are quite 'red' in that they are significantly brighter at infrared wavelengths than at visible wavelengths, and so it is possible for the AO system to perform poorly on objects which are bright at the science wavelength.

The performance of all these subsystems depends on the number of photons received from the object during the coherence time of the atmospheric fluctuations t_0 and over a coherence patch of order r_0^2 in size. Thus, the magnitude limits strongly depend on the prevalent seeing, and so siting interferometers in places with intrinsically good seeing and scheduling observations of faint sources during the periods of best seeing can dramatically improve the magnitude limits.

Since r_0 and t_0 are wavelength-dependent, the magnitude limits are typically better at longer wavelengths. At wavelengths longer than about 2 μm, the background noise due to photons emitted by the atmosphere and the instrument becomes the dominant noise source and the noise increases rapidly with increasing wavelength. As a result, the sensitivity of interferometers tends to peak at near-infrared wavelengths, which have the best combination of large r_0 and t_0 and low backgrounds.

At the time of writing of this book, the best reported magnitude limits for interferometric observations were visible magnitudes of around $m = 7$ (Mourard *et al.*, 2009) and infrared magnitudes of around $m = 11$ (Petrov *et al.*, 2012). Of note is the fact that both of the faintest limiting-magnitude measurements reported above were made using the group-delay fringe-tracking method with the science instrument itself acting as the fringe tracker.

Improvements in limiting magnitude are likely to come from improvements in detectors and in the performance of the interferometer 'infrastructure', such as reducing vibrations and increasing throughput and these improvements are likely to result in targets some 2–3 magnitudes fainter being observable. Further improvements will likely need the use of laser guide stars or going into space.