

### 3

## Atmospheric seeing and its amelioration

The interferometer described in Chapter 1 is an idealised one, where all the optical delays are known and stable. In such an interferometer the phase of the interference fringes measured for a point-like object at a chosen position on the sky (the phase centre) will be zero, and the fringe phase is a good ‘observable’. For a number of practical reasons, no existing optical interferometer even approaches this ideal.

First of all, the mechanical tolerances required to know the optical path difference (OPD) internal to the interferometer are formidable. If the position of a single mirror in the optical train is in error by a fraction of a millimetre then the phase of the fringes will be hundreds of radians from the value measured by an ideal interferometer.

While these mechanical errors could in principle be overcome through increasing the precision with which the interferometer is built, there are even more serious phase errors that are introduced external to the interferometer which no amount of increased construction accuracy will overcome. These errors are induced by the passage of starlight through the Earth’s atmosphere on its way to the interferometer. The atmosphere introduces phase errors, which are large (many radians) and rapidly varying (on timescales of milliseconds), and so these errors present a fundamental limitation to interferometry from the ground.

The optical effects which cause these rapidly changing phase perturbations are known in astronomy as *seeing*. Seeing can be observed in single telescopes as the ‘boiling’ of stellar images, which looks similar to the ‘shimmering’ of images that can be seen on a hot day.

The effects of seeing on interferometers are of a different character to those on a single telescope: whereas on a single telescope the seeing affects the angular resolution which can be obtained, in an interferometer the resolution can be relatively little affected by the presence of seeing but the sensitivity

can be dramatically altered. Understanding atmospheric seeing, how it affects interferometers, and how these effects can be ameliorated is the object of this chapter.

### 3.1 The wavefront perturbation model

The phase delay experienced by a light beam depends on the refractive index of the material it is passing through. The refractive index of air varies with temperature and humidity, so the delay experienced by the light propagating towards the interferometer depends on the detailed temperature and humidity structure of the atmosphere along the two paths.

To first order, the atmosphere is stratified, consisting of a series of horizontal layers. It was shown in Section 1.8 that if the atmosphere's refractive index is uniform within such a set of horizontal layers then the effects of the atmosphere's refractive index along the paths to the collectors of the interferometer cancel one another. However, any horizontal inhomogeneities in refractive index of the atmosphere can introduce disturbances to the OPD and hence to the fringe phase.

The most troublesome refractive index inhomogeneities come from turbulent mixing of atmospheric layers, which can cause random refractive index perturbations, and this will lead to a random perturbation to the delay experienced by the light passing through it. Wind will blow these refractive index perturbations past an observer causing the delay to evolve with time.

The standard model used to describe astronomical seeing therefore has three main ingredients: turbulence, its interaction with the existing gradients of refractive index, and its time evolution. These three ingredients are described to the level required to understand the main impacts of seeing on interferometry; more detailed insight can be gained from numerous papers and books on the subject (see, for example, Fried, 1978; Roddier, 1981; Wheelon, 2001).

#### 3.1.1 Turbulence model

Turbulence in the atmosphere is readily experienced by passengers in an aircraft unlucky enough to fly through it. It can be caused by wind shear at the interface between regions of air moving at different velocities or, near to the ground, by wind striking rocks, trees or buildings. Whatever its cause, the turbulence has a characteristic structure, consisting of random 'eddies' or 'whorls' of moving air, with larger eddies breaking down into smaller and

smaller eddies in a fractal type of structure. The statistical properties of this structure under quite general conditions were described by Kolmogorov (1941) and turbulence corresponding to this model is called *Kolmogorov turbulence*.

Turbulence by itself does not cause refractive index variations, as the mechanical energy present in the wind is not large enough to cause significant heating of the air. Instead ‘optical turbulence’ (that is turbulence which has a significant effect on a propagating beam) occurs only where turbulence and pre-existing gradients of refractive index coincide, for example where turbulence exists at the interface between two layers of air of different temperature.

In such environments, the turbulent motion mixes the air, creating regions of higher and lower refractive index whose structure mimics the structure of the turbulence itself. Figure 3.1 illustrates schematically the effect of these inhomogeneities in refractive index on a beam of starlight. The light arriving at the top of the atmosphere from a star is very close to a plane wave since the star is point-like and very distant. As the light travels through the turbulently mixed air, different parts of the wave are slowed down by different amounts depending on the line of sight through the turbulence, and so the wavefront (which describes a surface across which the light wave has a constant phase at

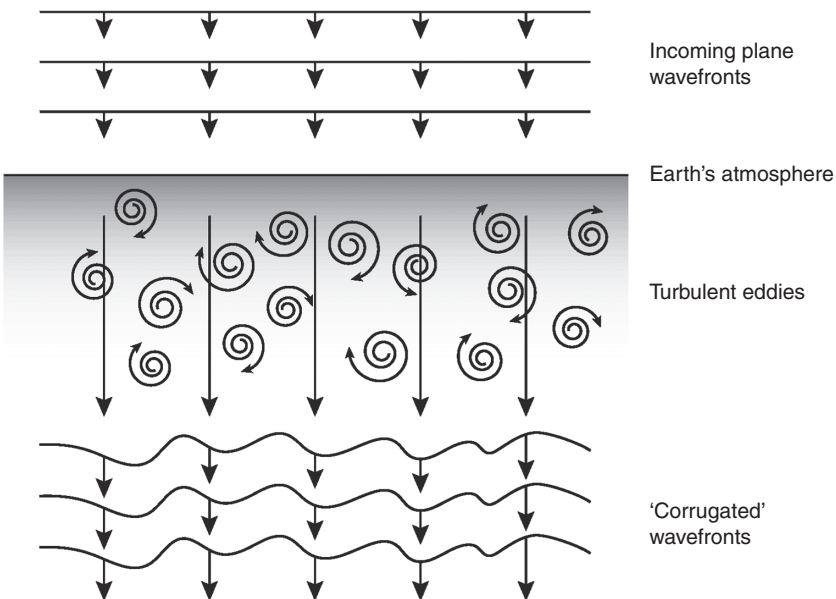


Figure 3.1 Schematic diagram of an initially plane wavefront propagating through the atmosphere.

any given instant) changes from being planar to being randomly ‘corrugated’ by the time the light arrives at a point of observation near the ground.

### 3.1.2 The spatial structure of the wavefront

The wavefront corrugations are random, and so the spatial structure of the corrugations can only be described statistically. There are a number of complementary ways of describing this structure. One description uses a quantity called the *structure function*  $D_\phi$  of the phase perturbations, which measures how the mean-squared difference in the phase perturbations at two locations on the wavefront varies as a function of their separation. It is defined as

$$D_\phi(\mathbf{r}, \mathbf{r}') \equiv \left\langle \left| \phi(\mathbf{r}' + \mathbf{r}) - \phi(\mathbf{r}') \right|^2 \right\rangle, \quad (3.1)$$

where  $\phi(\mathbf{r})$  is the value of the perturbation to the phase at a location denoted by the vector  $\mathbf{r}$ , which is in a plane perpendicular the average direction of propagation of the light beam under consideration, and  $\langle \rangle$  denotes taking the average value of a quantity.

If the process that is being described is *homogeneous*, in other words if its statistical properties do not depend on absolute position, only relative position, then it is possible to write  $D_\phi(\mathbf{r}, \mathbf{r}')$  as  $D_\phi(\mathbf{r})$ . The assumption of statistical homogeneity is a reasonable one for atmospheric turbulence, since in the free atmosphere there is typically no privileged position within the turbulence where the atmosphere can be expected to behave differently than elsewhere in the turbulence.

Tatarski (1961) showed that, for electromagnetic waves propagating through Kolmogorov turbulence, the structure function has the form

$$D_\phi(\mathbf{r}) = 6.88(r/r_0)^{5/3}, \quad (3.2)$$

where  $r = |\mathbf{r}|$  and  $r_0$  is the so-called Fried parameter. The factor of 6.88 arises from the original definition of  $r_0$  (Fried, 1966), which derives from considerations of the effect of the perturbations on optical imaging systems.

An alternative spatial description of the wavefront perturbations is to decompose the perturbations into sinusoidal ripples or corrugations on different scales. This can be done by taking a two-dimensional Fourier transform of the wavefront corrugations and considering how the amplitude and phase of each Fourier component varies as a function of the wavelength (or equivalently spatial frequency  $s$ ) of the component.

For any type of homogenous wavefront perturbations, the phases of all the Fourier components will be randomly distributed between 0 and  $2\pi$ , but the corrugations at different frequencies may have systematically different

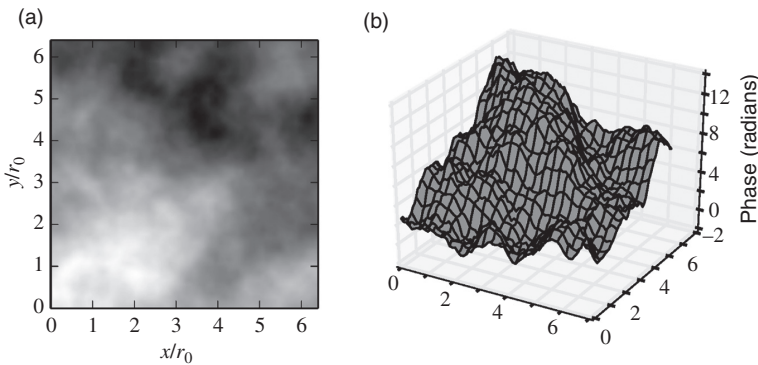


Figure 3.2 Atmospheric wavefront perturbations generated using a numerical simulation. Image (a) is a greyscale image of the surface plotted in (b).

amplitudes from one another. This can be expressed in terms of the *power spectrum* of the ripples, which expresses the mean-squared modulus of a Fourier component as a function of its spatial frequency  $s$ . Kolmogorov–Tatarski wavefront perturbations have a power spectrum of the form

$$\Phi(s) = 0.0229 r_0^{-5/3} |s|^{-11/3}. \quad (3.3)$$

It can be seen from this that the low-spatial-frequency (i. e. large-spatial-scale) corrugations have significantly higher amplitudes than their high-frequency counterparts.

The structure function and power spectrum are representations of the same information about the structure of the random wavefront, and one representation can be converted into the other by way of a Fourier transform. For Kolmogorov–Tatarski wavefronts, both representations are isotropic in that they depend only on the magnitude and not on the direction of  $\mathbf{r}$  or  $\mathbf{s}$  because it is assumed that the turbulence is itself isotropic.

Both representations show a power-law dependence, which means that the wavefront has a fractal structure, in other words the corrugations on larger scales look like scaled-up versions of the corrugations seen on smaller scales. Figure 3.2 shows a plot of corrugations with the Kolmogorov–Tatarski power spectrum: the similarity in appearance to clouds is not a coincidence as clouds are also caused by turbulence and thus have a similar fractal structure.

### 3.1.3 Seeing measures

The Fried parameter  $r_0$  is a measure of the ‘strength’ of the seeing. It denotes a spatial scale across which the phase differences are large enough to have

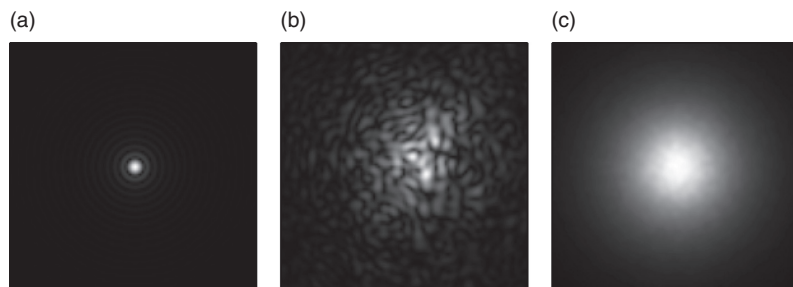


Figure 3.3 Simulated images of a point source of light: (a) seen through a diffraction-limited telescope of diameter  $d$ ; (b) seen in a short exposure through the same telescope affected by atmospheric seeing with a Fried parameter given by  $r_0 = 0.1d$ ; and (c) seen in a long exposure through the same telescope and seeing.

a significant optical effect. At visible wavelengths and at a good astronomical site  $r_0$  may be about 10 cm in size, so telescopes any larger than this in size will be significantly affected by atmospheric seeing. A point source of light seen through a telescope much larger than  $r_0$  in diameter will appear as a rapidly changing cloud of random ‘speckles’ similar to that shown in Figure 3.3.

In a long-exposure image, the speckles seen at different instants average out into a broader blur and what is seen is an approximately Gaussian-shaped peak with an angular full width at half maximum (FWHM) of about  $\lambda/r_0$  – this can be compared with the diffraction-limited spot size of  $\lambda/d$  where  $d$  is the diameter of the telescope. The FWHM angular size of a long-exposure image is often used as a measure of atmospheric seeing, and so conditions under which  $r_0 = 10$  cm at a wavelength of 500 nm can be described as ‘1 arcsecond seeing’.

The refractive index of the air is sufficiently constant with wavelength that the perturbations introduced to the optical delay are approximately independent of the wavelength of the radiation passing through it. Since  $r_0$  characterises the spatial scale on which a given magnitude of phase perturbations is seen at a given wavelength, the value of  $r_0$  scales with wavelength as  $\lambda^{6/5}$ . Thus, for example, the value of  $r_0$  at an infrared wavelength of  $2.2\ \mu\text{m}$  will be approximately six times greater than it is at a visible wavelength of 500 nm. The FWHM of a seeing-limited image scales as  $\lambda/r_0(\lambda) \propto \lambda^{-1/5}$  and so is a much less strong function of wavelength: 1 arcsecond seeing at 500 nm wavelength corresponds to 0.74 arcseconds seeing at  $2.2\ \mu\text{m}$  wavelength.

Seeing is like any other ‘weather’ condition in that the strength of the seeing can vary from place to place and as a function of time. At a good astronomical

site such as Mauna Kea in Hawaii or Paranal in Chile, the median seeing can perhaps be as good as 0.5 arcseconds, corresponding to an  $r_0$  value of 20 cm at a wavelength of 500 nm. Sites with median night-time seeing worse than about 2 arcseconds, corresponding to 5 cm  $r_0$  values, are typically considered unsuitable for siting interferometers.

### 3.1.4 Temporal seeing

The description of the wavefront perturbations has so far concentrated on the variation of the optical wavefront as a function of lateral position at a single instant in time. To complete the description requires a model of how the wavefront varies as a function of time.

The *frozen turbulence* hypothesis, often called the Taylor hypothesis, assumes that the majority of the temporal variation comes from the bulk motion of the turbulent fluctuations past the point of observation due to the wind. In this scenario, the turbulence is blown past the observer before the structure of the turbulence has time to evolve and, as a result, the temporal evolution of the perturbations to the phase at a given point on the wavefront can be derived straightforwardly from the spatial structure of the perturbations.

The temporal structure function is given by

$$D_\phi(t) \equiv \left\langle |\phi(\mathbf{r}, t' + t) - \phi(\mathbf{r}, t')|^2 \right\rangle = (t/t_0)^{5/3}, \quad (3.4)$$

where  $t_0$  is the *coherence time* of the seeing, corresponding to the timescale over which the phase perturbation typically changes by a radian. Note that there are multiple definitions for the coherence time of the seeing (Buscher, 1994; Kellerer and Tokovinin, 2007), but the above definition is perhaps the most common.

If there is only a single layer of turbulence moving at a constant speed  $v$ , then the coherence time will be related to the Fried parameter via

$$t_0 = 0.314 r_0 / v. \quad (3.5)$$

For the more realistic case of multiple layers of optical turbulence moving with different speeds and directions,  $v$  needs to be replaced by an effective windspeed, which is an appropriately weighted average of the speeds of the different layers. A typical windspeed for an atmospheric turbulence layer might be about  $10 \text{ m s}^{-1}$ , so that if  $r_0 = 10 \text{ cm}$  then  $t_0 \approx 3 \text{ ms}$ .

As with the spatial structure, the temporal structure of the perturbations can be represented as a power spectrum, which is given by

$$\Phi(f) = 0.00560 t_0^{-5/3} f^{-8/3}, \quad (3.6)$$

where  $f$  is the frequency of the perturbation, meaning that low-frequency oscillations in the phase have much greater amplitudes than those at higher frequencies.

### 3.1.5 Variation with zenith distance

If a telescope is looking at a star which is not directly overhead, the light received will have travelled through more turbulence between hitting the Earth's atmosphere and reaching the telescope. Thus, the seeing effects will get worse the further the object is from the zenith. Simple geometric arguments (Tatarski, 1961) lead to a prediction for the variation of  $r_0$  with zenith distance  $z$  as

$$r_0(z) = r_0(z = 0) \cos(z)^{3/5}. \quad (3.7)$$

The variation of  $t_0$  with zenith distance is more complex, as  $t_0$  depends on the relative orientation of the wind direction and the line of sight in each of the turbulent layers encountered. An experimentally measured dependence (Buscher, 1994) is

$$t_0(z) \propto \cos(z)^{1/5}, \quad (3.8)$$

which is a significantly less severe variation with zenith distance than that for  $r_0$ .

### 3.1.6 Scintillation

If the turbulence is sufficiently far from the point of observation, then diffraction effects can lead to the perturbed optical wavefront having variations in amplitude as well as phase. This gives rise to the phenomenon of stars twinkling when seen with the naked eye. These 'scintillation' effects are usually ignored when modelling interferometers, because under good astronomical seeing conditions the perturbations to the phase of the optical wavefront have a much larger effect on most interferometric measurements than the intensity fluctuations. Neglecting the effects of scintillation is conventionally called the *near-field* approximation (Roddier, 1981).

### 3.1.7 The outer scale of turbulence

Experimental measurements have confirmed that the Kolmogorov–Tatarski model and the Taylor hypothesis provide good models for the wavefront perturbations seen under astronomical observing conditions on spatial scales from



centimetres to metres (Breckinridge, 1976) and on timescales of milliseconds to seconds (Nightingale and Buscher, 1991).

However, the model predicts infinite amounts of wavefront perturbation if extrapolated to infinite scales. This is not surprising, as the Kolmogorov turbulence model only applies for scales that are smaller than the processes which are driving the turbulence.

There is limited evidence as to what this ‘outer scale’ is in the case of the turbulence causing seeing. The measurements that have been made show that the outer scale is of order 10–100 m in size (Buscher *et al.*, 1995; Davis *et al.*, 1995; Martin *et al.*, 1998; Dali Ali *et al.*, 2010). This means that the Kolmogorov–Tatarski model can be applied on the scale of individual telescopes, but that on the scale of the distances between telescopes in long-baseline interferometers the structure function and power spectrum need to be modified to take account of the effects of the outer scale.

The exact form of modification is unclear, but the evidence is that the structure function ‘flattens out’ at larger scales so that the rate of increase of phase-difference variance with separation falls below the  $r^{5/3}$  power law present on scales less than the outer scale.

### 3.2 First-order effects on interferometers

The most important effects of the turbulent wavefront perturbations on interferometers can be understood by first considering an interferometer where the individual light collectors are much smaller than  $r_0$  in diameter but the separation of the collectors is much greater than  $r_0$ , and where the exposure time used to measure the fringes is much less than  $t_0$ . In this case the wavefronts can be considered to have a constant phase across each aperture but the phase will be different between apertures.

The phase of the interference fringes seen in such an interferometer will be offset from the fringes that would be seen in the absence of the atmosphere, by an amount given by  $\epsilon_1 - \epsilon_2$  where  $\epsilon_1$  and  $\epsilon_2$  are the phase perturbations due to the atmosphere over telescopes 1 and 2, respectively. An idea of the typical size of this phase offset can be derived from the root-mean-square (RMS) value. For sources which are nearly overhead this will be given by

$$\sigma_\epsilon = \sqrt{\langle (\epsilon_1 - \epsilon_2)^2 \rangle} = \sqrt{D_\phi(B)}, \quad (3.9)$$

where  $B$  is the length of the baseline. For  $B = 10$  m and  $r_0 = 10$  cm then, assuming that the outer scale is much larger than 10 m, Equation (3.2) can be

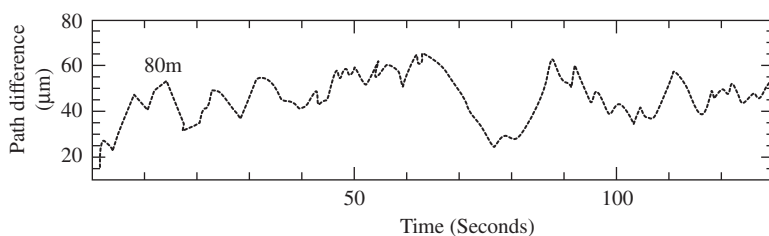


Figure 3.4 The motion of fringes seen on an 80-m baseline on the SUSI interferometer. From Davis *et al.* (1995).

used to derive a value for  $\sigma_\epsilon$  of approximately 120 radians, in other words nearly 20 wavelengths or  $10\text{ }\mu\text{m}$  of OPD at a wavelength of  $0.5\text{ }\mu\text{m}$ . Figure 3.4 shows the measured phase of the fringes seen on the SUSI interferometer showing that OPD fluctuations of many microns are not untypical.

### 3.2.1 Visibility modulus

The size of the random phase offset introduced by atmospheric perturbations is so large as to render the phase of the fringes meaningless as a measure of the object visibility phase. This is because, even after averaging many measurements, the error is likely to be much greater than a radian. However, the fringe visibility modulus remains unperturbed compared to the no-atmosphere case, providing that the interferometer apertures are much smaller than  $r_0$  and the effective exposure time of the fringe measurement is much less  $t_0$ .

Much can be deduced about an object from the visibility modulus alone, and for many years this was the primary observable in optical interferometry. For example, it is possible to measure the diameters and ellipticities of single stars, and to measure the separation and brightness differences of binary stars, purely from the changes in visibility modulus as a function of baseline.

Nevertheless, the lack of information related to the phase of the object visibility represents a serious loss of information about the object. For example, any object brightness distribution cannot be distinguished from the same distribution rotated by  $180^\circ$  using visibility-modulus information alone. This is because for any real function  $f(\mathbf{x})$ , the Fourier modulus is insensitive to reflections about the origin, i. e.

$$|\mathcal{F}\{f(\mathbf{x})\}| = |\mathcal{F}\{f(-\mathbf{x})\}|. \quad (3.10)$$

This clearly can cause problems in the astrophysical interpretation of a measurement, for example in determining the direction of evolution of a binary orbit from a sequence of snapshots of the star locations.

### 3.2.2 Closure phase

It will be shown in Section 9.6.2 that phase information becomes even more important in complex scenes and is critical to making true images. It turns out that there exists an observable which does capture some object visibility phase information in the presence of large atmospheric disturbances to the phase: this observable is called the closure phase, and was first used to cope with phase instabilities in radio interferometry (Jennison, 1958).

The existence of such an observable can be postulated from considering the amount of phase information measured by an interferometer consisting of  $N$  telescopes. If all possible pairwise combinations of the telescopes are made, then there will be  $\frac{1}{2}N(N-1)$  independent fringe phase measurements  $\{\Phi_{pq}\}$ , which are related to the object visibility phases  $\{\phi_{pq}\}$  via a set of linear equations of the form

$$\Phi_{pq} = \phi_{pq} + \epsilon_p - \epsilon_q, \quad (3.11)$$

where  $\epsilon_p$  is the atmospheric perturbation associated with telescope  $p$ . There are  $N-1$  unknown atmospheric phase perturbations (since it is only the *differences* between the perturbations that matter, one of them can be arbitrarily set to zero) and so if the  $\frac{1}{2}N(N-1)$  measurement equations represented by Equation (3.11) are linearly independent (and they are) and  $N \geq 3$ , then it would seem possible that the equations can be solved for  $\frac{1}{2}N(N-1) - (N-1) = \frac{1}{2}(N-1)(N-2)$  linearly independent quantities which are dependent only on the object phases and not on the atmospheric phases.

The simplest such quantity is the sum of the set of measured phases around the closing triangle of baselines formed from any three telescopes in the array. If the three telescopes are labeled  $p$ ,  $q$  and  $r$ , it is defined as

$$\Phi_{pqr} = \Phi_{pq} + \Phi_{qr} + \Phi_{rp} \quad (3.12)$$

and is called the *closure phase*.

Substituting Equation (3.11) into Equation (3.12) gives

$$\begin{aligned} \Phi_{pqr} &= \phi_{pq} + \epsilon_p - \epsilon_q + \phi_{qr} + \epsilon_q - \epsilon_r + \phi_{rp} + \epsilon_r - \epsilon_p \\ &= \phi_{pq} + \phi_{qr} + \phi_{rp}. \end{aligned} \quad (3.13)$$

Thus all the atmospheric phase error terms cancel, so we are left with an observable that is dependent only on the object visibility phases. In fact the above analysis can be repeated to show that closure phase is immune to *any* phase error which can be associated with individual collectors (in the nomenclature of radio astronomy, any ‘antenna-dependent’ phase error), for example any errors in the positions of the mirrors in the beam relay or the delay lines.

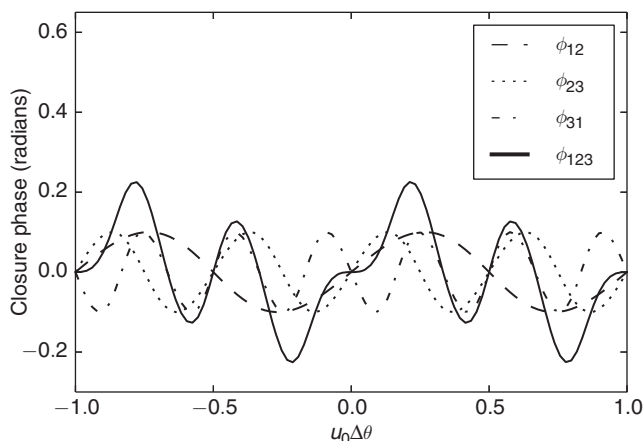


Figure 3.5 The closure phase measured on a binary star system as a function of the separation  $\Delta\theta$  of the pair of stars. The flux ratio of the pair is 10:1. In this one-dimensional example, the baselines form a 'linear triangle' with lengths  $u_0$ ,  $2u_0$  and  $3u_0$  and the angular separation of the pair of stars is assumed to be parallel to the direction of the baselines. The individual phases  $\phi_{12}$ ,  $\phi_{23}$  and  $\phi_{31}$  are shown, together with the closure phase  $\phi_{123} = \phi_{12} + \phi_{23} + \phi_{31}$ .

As a result using the closure phase eliminates many of the sources of phase error which are otherwise tricky to compensate for.

It is shown in Section 9.6.3 that the closure phase is insensitive to the position of the object with respect to the interferometer phase centre, but it is sensitive to the relative position of features *within* an object. Figure 3.5 shows an example of a binary star measured using a linear 'triangle' of telescopes. An important point illustrated in this example is that the closure phase changes sign if the stars are swapped (i. e.  $\Delta\theta$  changes sign). Thus, the closure phase can be used to determine 'which way round' an object is, whereas this would remain ambiguous if only visibility-modulus information were available.

It is shown in Chapter 9 that the closure phase, when combined with the visibility amplitudes, provides sufficient information to robustly reconstruct images in the presence of the atmosphere.

### 3.2.3 Spectral differential phase

The closure phase is a combination of phases from different baselines, which is independent of the atmospheric phase errors. Another object-dependent but atmospheric-independent phase which can be constructed is called the spectral differential phase (often shortened to 'differential phase'), which uses

a combination of phases measured on the same baseline but at different wavelengths.

Consider a two-element interferometer, which measures the fringe phase in  $N$  different spectral channels on a single baseline. Instrumental errors will introduce an unknown OPD error of  $c\tau_a$  and hence will perturb the fringe phase by  $2\pi\tau_a\nu$ , where  $\nu$  is the frequency. As shown in Section 1.8, the atmosphere will perturb the fringe phase by an amount which can be modelled over a limited wavelength range as  $\epsilon_0 + 2\pi\tau_b\nu$ , where  $\epsilon_0$  is a wavelength-independent atmospheric phase error and  $\tau_b$  is the atmospheric ‘group delay’. Thus, the measured fringe phase in spectral channel  $i$  will be given by

$$\Phi(\nu_i) = \phi(\nu_i) + \epsilon_0 + \epsilon_1 \nu_i, \quad (3.14)$$

where  $\phi(\nu_i)$  is the object visibility phase at frequency  $\nu_i$ , and  $\epsilon_1 = 2\pi(\tau_a + \tau_b)$ . If the fringes are measured *simultaneously* at three or more wavelengths, then there are more measurements than atmospheric/instrumental unknowns and so an atmosphere-independent measureable can be obtained.

The conceptually simplest such measurement can be derived if fringes are measured at three wavelengths which are equally spaced in frequency such that  $\{\nu_i\} = \{\nu_0 - \Delta\nu, \nu_0, \nu_0 + \Delta\nu\}$ , where  $\nu_0$  is the central frequency and  $\Delta\nu$  is the frequency spacing. Subtracting the average fringe phase measured at the outer two wavelengths from the phase measured at the central wavelength yields a quantity in which the atmospheric error terms cancel, leaving the object differential phase:

$$\Phi_{\text{differential}}(\nu_0) \equiv \Phi(\nu_0) - \frac{1}{2} [\Phi(\nu_0 - \Delta\nu) + \Phi(\nu_0 + \Delta\nu)] \quad (3.15)$$

$$= \phi(\nu_0) - \frac{1}{2} [\phi(\nu_0 - \Delta\nu) + \phi(\nu_0 + \Delta\nu)] \quad (3.16)$$

$$\equiv \phi_{\text{differential}}(\nu_0). \quad (3.17)$$

This is just one form of the spectral differential phase; with greater numbers of spectral channels, other differential phase measures can be defined. A more sophisticated way of computing a differential phase is to jointly estimate the values of  $\epsilon_0$  and  $\epsilon_1$  by fitting a phase change which is a linear function of frequency to the data, and attributing any residual higher-order changes of phase with frequency to higher-order variations in the object phase.

The fitting of instrument-introduced spectral phase variations to the measured data can be further extended to include estimation of higher-order terms. For example, atmospheric and instrumental chromatic dispersion (see Section 1.8) will introduce quadratic and higher-order phase variations with frequency. However, if higher-order instrumental phase terms are allowed to

be free parameters, then only the object phase variations which are even higher order than the unconstrained instrumental phase terms can be recovered. Thus, only a limited number of instrumental degrees of freedom are fitted to the data.

The interpretation of the object differential phase is simple in the special case where the object is expected to appear as an unresolved point source at the outer wavelengths while the object is expected to appear resolved on the interferometric baselines in question at the central wavelength. This is often because the central wavelength corresponds to a strong spectral emission or absorption line in the object. In this case the differential phase at the central wavelength is simply the object phase at that wavelength, since the object phase at the outer wavelengths is zero.

The differential phase can be used in other situations where there are a priori constraints on the variation of object shape with wavelength. For example, if the object can be assumed to be 'grey', i. e. its shape is independent of wavelength or is a slowly varying function of wavelength, then the differential phase can be used to provide empirical constraints on models of the object.

In many astronomical contexts, however, such assumptions are not appropriate because the variation of the object shape with wavelength cannot be constrained a priori. In such cases the differential phase has seen relatively limited application. This is because in such situations a model for the object shape at every wavelength must be inferred from the data, and the combined model will of necessity have many more parameters than the equivalent single-wavelength model. It is often constrained by many fewer measurements than model parameters, and so the scope for degeneracies, where many widely differing sets of model parameters fit the same data, is greater than in the case of closure phase imaging. Thus, except in a few special cases, imaging is usually obtained using closure phase information; no image reconstruction packages using the differential phase are generally available.

For simplicity, much of the rest of the discussion focusses on the closure phase as the primary source of object phase information for imaging, with the differential phase being reserved for special cases where its unique properties can be best exploited.

### 3.2.4 The power spectrum and bispectrum

The previous sections have shown that, while the visibility is corrupted by OPD perturbations, the visibility modulus and the closure phase are not (at least for small aperture sizes and for short integration times). These quantities can therefore serve as 'observable' quantities in optical interferometry in place of the complex visibility.

However, these quantities will in general be noisy because of the low light levels involved; therefore, it is advantageous to be able to reduce the noise by taking an average value over multiple independent exposures. Since the exposure time is typically measured in milliseconds, it is possible to take thousands of exposures in a few minutes, and so this averaging (called ‘incoherent averaging’) can serve to reduce the noise level by more than an order of magnitude. The quantities which are averaged are not the visibility modulus and the closure phase themselves, but rather two quantities from which the modulus and closure phase can be derived. These quantities are the ‘power spectrum’ and the ‘bispectrum’, respectively.

The power spectrum is the modulus squared of the fringe coherent flux  $F_{12}$ :

$$P_{12} = |F_{12}|^2. \quad (3.18)$$

This can be related to (and in the ideal case is equal to) the power spectrum of the object at spatial frequency  $\mathbf{u}_{12}$ , where the object power spectrum is given by

$$P(\mathbf{u}) = |F(\mathbf{u})|^2. \quad (3.19)$$

The modulus squared captures the magnitude of the coherent flux. It is easier to do analytic computations on the mean of the modulus squared rather than the mean of the modulus: as will be seen in Section 5.5.2, such computations are needed to remove bias terms caused by noise such as photon noise.

The bispectrum is sometimes known as the ‘triple product’ because it is the product of three coherent fluxes

$$T_{pqr} = F_{pq}F_{qr}F_{rp}, \quad (3.20)$$

where  $p$ ,  $q$  and  $r$  denote three different collectors.

The bispectrum of the object can be written

$$T(\mathbf{u}_{pq}, \mathbf{u}_{qr}, \mathbf{u}_{rp}) = F(\mathbf{u}_{pq})F(\mathbf{u}_{qr})F(\mathbf{u}_{rp}), \quad (3.21)$$

and  $\mathbf{u}_{pq}$  denotes the spatial frequency sampled by the baseline between collectors  $p$  and  $q$ . Since the relevant baselines form a triangle, then  $\mathbf{u}_{rp} + \mathbf{u}_{pq} + \mathbf{u}_{qr} = 0$  and so the bispectrum can be expressed as a function of two spatial frequencies (hence the term ‘bispectrum’):

$$T(\mathbf{u}_{pq}, \mathbf{u}_{qr}) = F(\mathbf{u}_{pq})F(\mathbf{u}_{qr})F(-\mathbf{u}_{pq} - \mathbf{u}_{qr}). \quad (3.22)$$

The bispectrum is a complex number whose argument (or phase) is

$$\arg(T_{pqr}) = \phi_{pq} + \phi_{qr} + \phi_{rp}; \quad (3.23)$$

in other words, the closure phase.

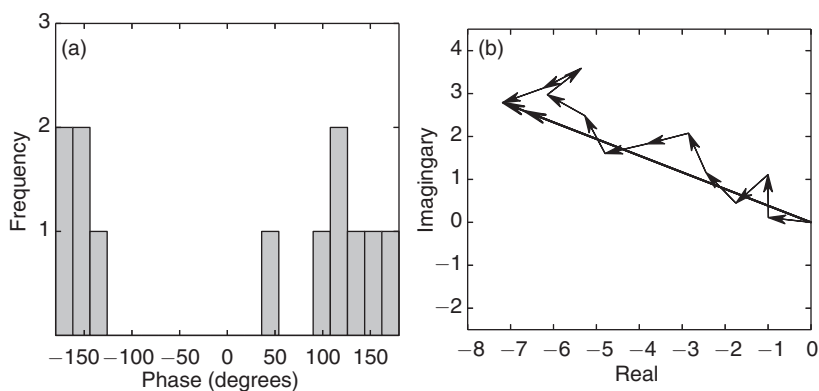


Figure 3.6 Histogram of a sample set of phases (a) and vector averaging of this set of phases (b). The arithmetic average of the phases is  $3^\circ$ , while the vector average phase is  $159^\circ$ .

There are a number reasons for averaging the bispectrum rather than the closure phase. The most important of these is that averaging a modulo- $360^\circ$  quantity like the closure phase is fraught with pitfalls, and averaging the bispectrum is a more robust way to obtain an average of such a quantity. This method of averaging is known as a ‘vector average’ of a complex quantity whose argument is the phase to be averaged.

The advantages of vector averaging can be understood by considering a sample of phase measurements such as those given in Figure 3.6. It can be seen that the samples cluster around a value close to  $150^\circ$ , but a mean of the sample phase values yields an average phase of about  $3^\circ$ . This is because around half of the sample phases wrap around from  $180^\circ$  to  $-180^\circ$ , and this biases the average towards zero.

If instead the sample phases are assigned to unit vectors (‘phasors’) in the complex plane as shown in Figure 3.6(b) then the sum of these phasors is a vector with a phase of  $159^\circ$ . This vector average is clearly a better central measure of the sample set of phases than the simple mean of the phases.

The argument of the average bispectrum is a form of the vector average. In this case, each phasor has a modulus whose amplitude is the product of the individual fringe amplitudes at the three spatial frequencies. With noisy data, this tends to weight higher-than-average signal-to-noise-ratio data more heavily and so converges faster to the ‘true’ average phase than a vector average where each phasor has the same length (Woan and Duffett-Smith, 1988).

The modulus of the bispectrum contains information about the magnitude of the object coherent flux at the three component spatial frequencies,



but this information is available in a more direct form by measuring the power spectrum at the relevant spatial frequencies, so the bispectrum amplitude information is not as critical to making interferometric images as the bispectrum phase.

### 3.3 The effects of finite exposure time

If sufficiently small apertures and short exposures are used to make measurements of the fringe pattern, the fringe visibility modulus will be close to the source visibility modulus. If the exposure time is comparable to or larger than  $t_0$  then the fringes will have changed phase during the exposure and so will ‘smear out’. As a result, the measured visibility modulus will be lower than the instantaneous fringe visibility modulus by an amount that depends on the fringe motion during the exposure. If the instantaneous fringe intensity pattern has the form

$$I(x, t) = I_0 \left( 1 + \operatorname{Re} \left\{ V_0 e^{i[2\pi s x + \phi_1(t) - \phi_2(t)]} \right\} \right) \quad (3.24)$$

where  $V_0$  is the object fringe visibility,  $s$  is the spatial frequency of the fringes and  $\phi_i(t)$  is the fringe phase perturbation over telescope  $i$  at time  $t$ , then the fringe pattern observed with a finite exposure beginning at time  $t = 0$  and ending at time  $t = \tau$  is given by

$$I(x, t, \tau) = \int_0^\tau I_0 \left( 1 + \operatorname{Re} \left\{ V_0 e^{i[\phi_1(t) - \phi_2(t)]} e^{2\pi i s x} \right\} \right) dt \quad (3.25)$$

$$= \tau I_0 \left( 1 + \operatorname{Re} \left\{ V_0 \gamma(\tau) e^{i[2\pi s x]} \right\} \right). \quad (3.26)$$

Thus the smeared fringe pattern is a sinusoid with the same fringe frequency  $s$  as the unsmeared fringe pattern but a visibility given by  $V_0 \gamma(\tau)$ , where  $\gamma(\tau)$  is a complex *blurring* factor given by

$$\gamma(\tau) = \frac{1}{\tau} \int_0^\tau e^{i[\Phi_{12}(t)]} dt, \quad (3.27)$$

and where  $\Phi_{12}(t) = \phi_1(t) - \phi_2(t)$  is the difference between the phase perturbations at the two telescopes at time  $t$ .

#### 3.3.1 Visibility loss for short exposure times

For exposure times such that  $\tau \ll t_0$  the phase difference  $\Phi_{12}$  can be expressed as

$$\Phi_{12}(t) = \Phi_0 + \Delta\Phi(t), \quad (3.28)$$

where  $\Phi_0$  is the average phase for the exposure and  $\Delta\Phi(t) \ll 1$  radian. In this case Equation (3.27) can be written

$$\gamma(\tau) = \frac{e^{i\Phi_0}}{\tau} \int_0^\tau e^{i\Delta\Phi(t)} dt \quad (3.29)$$

$$\approx \frac{e^{i\Phi_0}}{\tau} \int_0^\tau \left(1 + i\Delta\Phi(t) - \frac{1}{2}\Delta\Phi(t)^2\right) dt, \quad (3.30)$$

where the exponential has been expanded to second order in  $\Delta\Phi$ . Since  $\Phi_0$  is the average phase, then  $\int_0^\tau \Delta\Phi(t) dt = 0$  and so

$$\gamma(\tau) \approx e^{i\Phi_0} \left(1 - \frac{\sigma(\tau)^2}{2}\right), \quad (3.31)$$

where  $\sigma(\tau)$  is the RMS variation of the fringe phase during an exposure of length  $\tau$ , in other words

$$\sigma^2(\tau) = \frac{1}{\tau} \int_0^\tau [\Delta\Phi(t)]^2 dt. \quad (3.32)$$

Thus, as might be intuitively expected, the phase of the measured fringe is different from the phase measured in the absence of the atmosphere by an amount given by the average atmospheric phase difference over the exposure. At the same time the modulus of the fringe visibility decreases with increasing variation of the phase during the exposure as  $1 - \frac{1}{2}\sigma^2(\tau)$ .

It is usually more mathematically convenient to work with the loss in the *squared* visibility modulus, which will be given approximately by

$$|\gamma|^2 \approx 1 - \sigma^2(\tau). \quad (3.33)$$

This loss will vary from exposure to exposure; the mean loss will be given by

$$\langle |\gamma|^2 \rangle \approx 1 - \langle \sigma^2(\tau) \rangle. \quad (3.34)$$

An expression of this form appears in a well-known approximation for the intensity of an image formed by an optical system with phase aberrations, known as the Maréchal approximation. The similarity is not accidental, as the calculation to which the Maréchal approximation applies involves integrals similar to that in Equation (3.27).

For longer exposure times where  $\langle \sigma^2(\tau) \rangle > 1$ , Equation (3.34) is clearly invalid as it predicts a negative value for the squared modulus. A modified version of this expression, which does not have this problem, is

$$\langle |\gamma|^2 \rangle \approx e^{-\langle \sigma^2(\tau) \rangle}. \quad (3.35)$$

This expression agrees with Equation (3.34) in the region of applicability ( $\sigma \ll 1$ ) but gives more realistic answers for the visibility loss for  $\sigma > 1$ .

The equivalent expression used in aberrated image calculations is known as the ‘extended Maréchal approximation’ (Mahajan, 1983), and has been found both theoretically (Sandler *et al.*, 1994) and empirically to give an acceptable level of approximation when  $\sigma \lesssim 1$ .

Assuming a Kolmogorov–Tatarski–Taylor model for the phase perturbations, the mean variance of the phase during an exposure is given by Tango and Twiss (1980):

$$\langle \sigma^2(\tau) \rangle \approx \left( \frac{\tau}{2.6t_0} \right), \quad (3.36)$$

so the mean-squared visibility loss will vary with exposure time approximately as

$$\langle |\gamma(\tau)|^2 \rangle \approx e^{-\left( \frac{\tau}{2.6t_0} \right)^{5/3}}. \quad (3.37)$$

### 3.3.2 The random-walk model for longer exposures

The evolution of the visibility modulus at longer exposure times can instead be derived from a *random-walk* model. In this simplified model, the fringe phase remains roughly constant over some period  $\tau_0 \sim t_0$ , but changes randomly between successive periods. For an exposure consisting of  $n$  such periods, Equation (3.27) simplifies to

$$\gamma(n\tau_0) = \frac{1}{n} \sum_{k=1}^n e^{i\Phi_k}, \quad (3.38)$$

where  $\Phi_k$  is the value of the phase difference  $\Phi_{12}(t)$  in the time interval  $(k-1)\tau_0$  to  $k\tau_0$ . Since  $\Phi_k$  is randomly distributed in the interval 0 to  $-2\pi$ , the summation in Equation 3.38 can be seen as a random walk in the complex plane as shown in Figure 3.7. The standard result for a random walk is that the RMS length of the summed vector is  $\sqrt{n}$  times the step size, and therefore  $|\gamma|$  decreases as  $n^{-1/2}$ . As a result the mean-squared visibility loss decreases inversely with the exposure time

$$|\gamma(\tau)|^2 \approx \frac{\tau_0}{\tau}. \quad (3.39)$$

For phase fluctuations given by the Kolmogorov–Tatarski–Taylor model, the mean-squared visibility loss can be evaluated in terms of numerical integrals (Dainty and Greenaway, 1979; Buscher, 1988a) and Figure 3.8 shows the results of these calculations. It can be seen that approximations derived for small and large  $\tau$  give good predictions for the behaviour in their respective regimes of validity: the extended Maréchal approximation for  $\tau \lesssim t_0$  and the

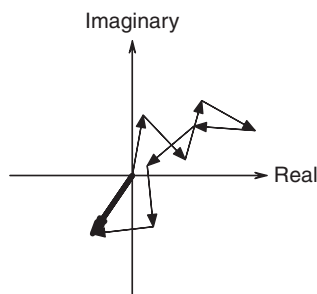


Figure 3.7 A simple random-walk model for the fringe visibility.

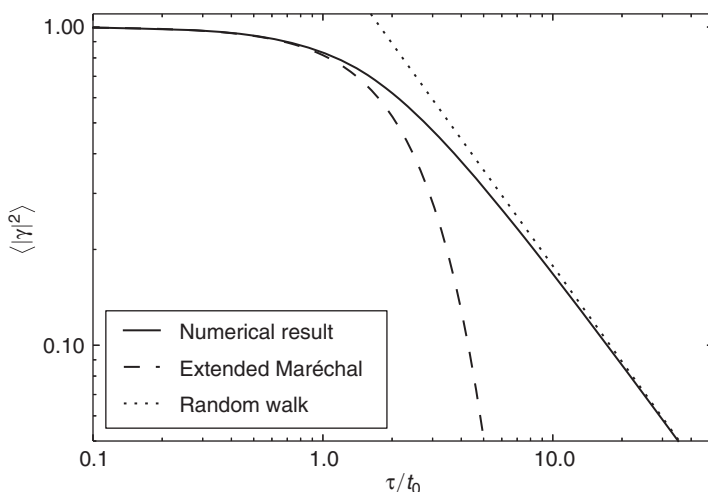


Figure 3.8 The mean-squared visibility loss  $\langle |V|^2 \rangle$  as a function of exposure time  $\tau$ . Also shown are approximate expressions for large and small  $\tau$ .

random-walk model for  $\tau \gg t_0$ . In the case of the random-walk model a value of  $\tau_0 \approx 1.78t_0$  gives the best fit.

### Visibility calibration

In order to collect enough photons in each exposure, the exposure times used in interferometry are often comparable to or greater than  $t_0$ . The loss in visibility due to the finite exposure time can be compensated for by taking many exposures and then normalising the sample RMS visibility using the relevant value given in Figure 3.8. However, there are complications:  $t_0$  is variable and may not be measurable directly; in addition there may be instrumental phase errors which further reduce the visibility. Therefore, the measured visibility is

usually calibrated by comparing it with the visibility measured on a nearby reference star with known visibility.

It is desirable that the change in RMS visibility for a given change in  $t_0$  is small so that any changes in  $t_0$  between the observation of the target and observation of the reference star have the least effect on the calibration. This is a function of the gradient of the curve in Figure 3.8 and hence it is desirable to work with short integration times from this point of view. Beyond an exposure time of about  $2t_0$ , however, a 1% change in  $t_0$  will always give about a 1% change in the mean-squared visibility, as predicted by the random-walk model.

### 3.4 The effects of finite aperture size

In the previous analysis, the interferometer was modelled as consisting of point-like light collectors. In an interferometer consisting of collectors with sizes comparable with  $r_0$ , the effects of the variation in the atmospheric phase perturbations within each collector aperture need to be considered.

For the ‘pupil-plane’ beam combination scheme used in the example interferometer introduced in Section 1.3, variations in the phase perturbations across the aperture will cause distortions in the fringe pattern. Including the atmospheric phase distortions into Equation (1.16) gives an instantaneous intensity pattern

$$i(x, y) = 2A^2 \left( 1 + \operatorname{Re} \left\{ V_0 e^{i[2\pi s x + \phi_1(x, y) - \phi_2(x, y)]} \right\} \right) \quad (3.40)$$

$$= 2A^2 \left( 1 + \operatorname{Re} \left\{ V_0 e^{i(2\pi s x + \Phi_{12}(x, y))} \right\} \right), \quad (3.41)$$

where  $(x, y)$  is a two-dimensional coordinate in the illuminated area of the detector,  $\phi_i(x, y)$  is the instantaneous phase perturbation due to the atmosphere at coordinate  $(x, y)$  within aperture  $i$  and  $\Phi$  denotes a phase difference  $\Phi_{12}(x, y) \equiv \phi_1(x, y) - \phi_2(x, y)$ . Thus, the positions of the maxima and minima of the fringe pattern shift to reflect the differential phase perturbations at a given location in the aperture as illustrated in Figure 3.9.

These fringe distortions have little effect on the fringe amplitude if the fringe measurement is made locally in small regions in the aperture. However, typically the light level is not high enough to allow accurate measurement of the fringe parameters in small regions, and instead the fringe signal needs to be averaged across the whole aperture. Different deviations to the fringe phase in different parts of the aperture will mean that the averaged fringe will be smeared in a similar way to that experienced in the finite exposure time case.



Figure 3.9 A simulated pupil-plane fringe pattern in the presence of atmospheric phase perturbations.

In Section 8.5 it is shown that there are multiple possible ways to determine the coherent flux from a fringe pattern. One way to do this is to extract the Fourier component in the intensity pattern at the spatial frequency  $s$  corresponding to the frequency of the fringes in the unperturbed pattern. This can be achieved by multiplying the intensity by a sinusoid at frequency  $s$  and integrating the result over the illuminated region  $S$

$$\hat{F}_{12} = \iint_S e^{-2\pi i s x} i(x, y) dx, dy, \quad (3.42)$$

where the circumflex on  $\hat{F}_{12}$  denotes an ‘estimator’ for  $F_{12}$ , the fringe coherent flux, and  $(x, y)$  is a coordinate in the plane of the detector. Substituting for  $i(x, y)$  from Equation (3.40) gives

$$\hat{F}_{12} \approx A^2 V_0 \iint_S e^{i\Phi_{12}(x, y)} dx, dy, \quad (3.43)$$

where it has been assumed that  $\iint_S e^{2\pi i s x} dx, dy \approx 0$  and that  $\iint_S e^{i[4\pi s x + \Phi_{12}(x, y)]} dx, dy \approx 0$ . Both conditions are achieved if the fringe spacing  $1/s$  is small compared with the size of the aperture and with  $r_0$ .

In the absence of atmospheric perturbations (i. e. when  $\Phi_{12}(x, y) = 0$ )  $\hat{F}_{12} \propto V_0$  and so  $\hat{F}_{12}$  provides an estimate of the object coherent flux. Equation (3.43) can be written as

$$\hat{F}_{12} \approx \gamma(D)F_{12}(0), \quad (3.44)$$

where  $F_{12}(0)$  is the coherent flux that would be observed in the absence of the atmosphere, and  $\gamma(D)$  is the complex visibility factor introduced by the atmosphere for an interferometer with apertures of diameter  $D$  and given by

$$\gamma(D) = \iint_S e^{i\Phi_{12}(x,y)} dx, dy. \quad (3.45)$$

The expression for the visibility factor  $\gamma(D)$  due to the atmosphere in Equation (3.45) is of the same form as for the factor  $\gamma(\tau)$  due to finite exposure times given in Equation (3.27) but with a two-dimensional integral over the aperture replacing the one-dimensional integral over the integration time. As a result, the techniques developed for calculating the change in visibility due to the temporal atmospheric phase perturbations can be used to derive the change in visibility due to the spatial perturbations.

### 3.4.1 Visibility loss for small- and large-aperture diameters

In the presence of Kolmogorov–Tatarski turbulent phase fluctuations, the mean phase variance across a single aperture of size  $D$  will be given by (Noll, 1976):

$$\langle \sigma^2(D) \rangle = 1.0299 (D/r_0)^{5/3}. \quad (3.46)$$

For two well-separated apertures the wavefront perturbations can be considered to be uncorrelated between apertures and therefore the phase *difference* between the apertures will have a variance twice as large. Using the extended Maréchal approximation given in Equation (3.35) leads to a mean-squared visibility reduction of

$$\langle |\gamma(D)|^2 \rangle \approx e^{-2.06 \left( \frac{D}{r_0} \right)^{5/3}} \quad (3.47)$$

for  $D \lesssim r_0$ .

In the case of larger apertures, a random-walk model for the phase perturbations from Section 3.3.1 can be used but now assuming that an aperture of diameter  $D = nd_0$  can be divided into  $n^2$  subapertures or ‘seeing cells’ of diameter  $d_0 \sim r_0$ . Across these cells the phase fluctuations are assumed to be small, but between cells the phases make random jumps of up to  $2\pi$  radians. For this seeing-cell model, the mean-squared visibility reduction is then given by

$$\langle |\gamma(D)|^2 \rangle \approx \left( \frac{D}{d_0} \right)^{-2}. \quad (3.48)$$

The mean-squared visibility loss in the general case can be evaluated numerically (Korff, 1973) to give the graph of visibility loss against aperture diameter

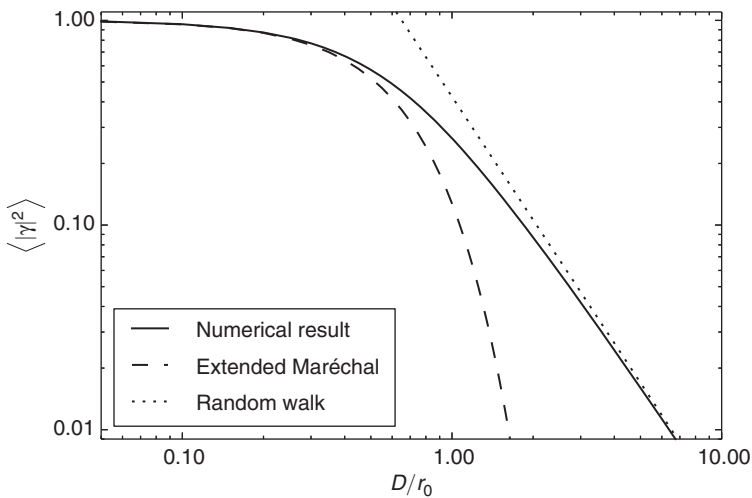


Figure 3.10 The mean-squared fringe visibility loss  $\langle |\gamma|^2 \rangle$  as a function of aperture diameter  $D$  for a pair of well-separated circular apertures in Kolmogorov turbulence. Also shown are approximate forms for  $\langle |\gamma|^2 \rangle$  when  $D \ll r_0$  and  $D \gg r_0$ .

shown in Figure 3.10, which confirms the applicability of the approximate models for apertures  $D \ll r_0$  and  $D \gg r_0$ . In the latter case a value of  $d_0 \approx 0.65r_0$  gives the best fit between the analytical approximation and the numerical integration.

### 3.5 Adaptive optics

Figure 3.10 shows that the loss in visibility for apertures of only moderate size is appreciable: the RMS visibility decreases by half when the aperture diameter is about  $r_0$ , and  $r_0$  is typically a few tens of centimetres while modern astronomical telescopes are typically many metres in size. Thus, the size of apertures which can be used and hence the faintness of the sources which can be observed is limited by the atmosphere and not the technology for building large-aperture telescopes. This limitation can be overcome if the spatial wavefront perturbations across the apertures can be reduced to an acceptable level, effectively increasing the size of  $r_0$ . This can be done using *adaptive optics* (AO) systems.

AO systems have become commonplace on large, single telescopes as a way to reduce the effects of atmospheric seeing. Figure 3.11 shows a schematic



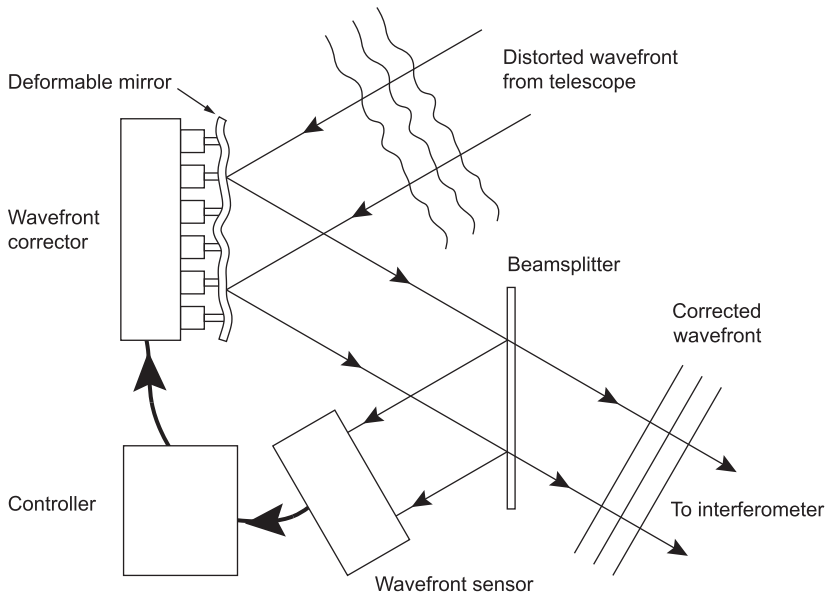


Figure 3.11 Schematic diagram of an AO system on a telescope.

layout for a typical AO system. It consists of a system for measuring the distortions to the incoming wavefronts, called a *wavefront sensor*, a *wavefront corrector*, which is typically a deformable mirror whose shape can be controlled electronically, and a *controller*, which takes the information from the wavefront sensor and attempts to cancel the errors in the wavefront by sending the appropriate signals to the wavefront corrector, causing it to adjust its shape to be the opposite to the atmospheric perturbations. The whole system runs in real time, at update intervals of order  $t_0$  or less, in order to compensate for the distortions before they have time to change.

### 3.5.1 Wavefront modes

In understanding the degree of correction possible with AO it is helpful to consider the atmospheric wavefront perturbations as being made out of a set of *spatial modes*, that is to say a set of distortion shapes that can be combined to form any possible wavefront perturbation across an aperture. A commonly used set of modes are the Zernike polynomials, which have a number of useful properties. One of these is that many of the lowest-order polynomials correspond closely to well-known aberrations of an optical system, such as coma, spherical aberration and so on.

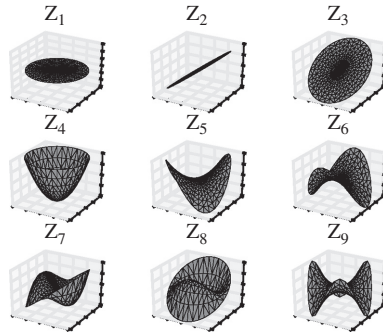


Figure 3.12 Surface plots of the lowest-order Zernike polynomials.

The Zernike polynomials  $Z_n^m$  are defined on a unit circle representing the aperture of a telescope and can be separated into radial and azimuthal components

$$Z_n^m(\mathbf{r}) = R_n^m(r)f_m(\theta), \quad (3.49)$$

where  $\mathbf{r}$  is a vector inside the circle,  $(r, \theta)$  are the corresponding polar coordinates, and  $n$  and  $m$  are integers with  $n \in \{0, 1, 2, \dots\}$  and  $m \in \{-n, -n+2, -n+4, \dots, n\}$ , known as the radial and azimuthal order of the Zernike polynomial, respectively.  $R_n^m(r)$  is a polynomial of order  $n$  and  $f_m(\theta)$  is either  $\cos(m\theta)$  or  $\sin(|m|\theta)$ , depending on the sign of  $m$ .

A conventional mapping of the two indices  $n$  and  $m$  to a single index  $j$  has been introduced by Noll (1976). This mapping of groups polynomials with the same radial order  $n$  into consecutive values of  $j$  such that higher values of  $j$  generally correspond to increasing numbers of oscillations within the unit circle.

Figure 3.12 shows a few of the lowest-order Zernike polynomials. The first mode  $Z_1 = Z_0^0$  is called the ‘piston’ mode and corresponds to a uniform phase offset over the telescope aperture. This mode is ‘invisible’ to a single telescope but piston-mode *differences* between pairs of telescopes are observable by an interferometer as a shift in the phase of the respective fringe pattern. The next two polynomials  $Z_2$  and  $Z_3$  correspond to linear phase ramps across the aperture in orthogonal directions and are conventionally called ‘tip’ and ‘tilt’ (which is tip and which is tilt is not well-defined). Changes in these modes correspond to changes in direction of the beam. Higher-order modes correspond to defocus ( $Z_4$ ), astigmatism ( $Z_5$  and  $Z_6$ ) and so on.

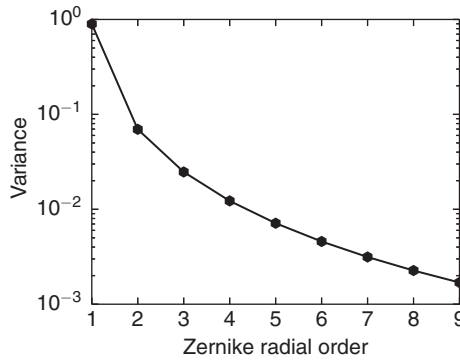


Figure 3.13 The variance of the coefficients of the Zernike modes in the wavefront perturbations produced by Kolmogorov turbulence. The variances for all modes with the same radial order are the same (Noll, 1976); what is plotted is the total wavefront variance contributed by all  $n + 1$  modes corresponding to a given radial order  $n$ . The variance is plotted for an aperture diameter of  $D/r_0 = 1$  and scales as  $(D/r_0)^{5/3}$ .

Any phase perturbation  $\phi(\mathbf{r})$  across a circular aperture can be expressed as a sum of Zernike polynomials with an appropriate set of weights  $\{a_k\}$  such that

$$\phi(\mathbf{r}) = \sum_{k=1}^{\infty} a_k Z_k(\mathbf{r}), \quad (3.50)$$

where  $Z_k(\mathbf{r})$  is the Zernike polynomial with index  $k$ , which is a function of the two-dimensional position  $\mathbf{r}$  in the aperture.

The statistical properties of the wavefront perturbations across an aperture can be analysed in terms of the mean-squared values  $\{\langle a_k^2 \rangle\}$  of the Zernike coefficients, and for Kolmogorov turbulence this ‘spectrum’ is heavily weighted towards the lowest-order modes, as shown in Figure 3.13. This reflects the fact that the spatial power spectrum of the Kolmogorov–Tatarski wavefront perturbations (see Equation (3.3)) has more power in the low-frequency corrugations.

In fact, for Kolmogorov turbulence with an infinite outer scale, the theoretical RMS amplitude of the piston mode is infinite. While an infinite outer scale does not exist in practice, infinite-outer-scale models are often used because they are mathematically convenient and represent a worst-case scenario. The infinite amplitude of the piston mode does not present a problem as only the difference of the piston modes between two apertures has any

impact on an interferometer, and this difference is finite even if the outer scale is infinite.

Only the non-piston spatial modes are sensed by conventional AO systems – correcting the piston mode requires a fringe tracker, and this is discussed further in Section 3.8. The practical implication of the rapidly falling amplitudes of the Zernike modes with mode order is that, for an AO system with a limited number of degrees of freedom, these degrees of freedom are best used in removing the lowest-order modes of the wavefront, leaving the higher-order modes uncorrected. All modern optical interferometers incorporate at least tip–tilt AO correction at each aperture, because removing just the tip and tilt modes reduces the variance of the wavefront by nearly 90% (Noll, 1976).

### 3.5.2 Interferometry with AO

Adding an AO system to each of the collecting elements of an interferometer will reduce the level of phase difference variation between the interfering wavefronts and so the integral in Equation (3.43) will be larger than the value obtained in the absence of AO. Thus, AO serves to increase the visibility of the fringes, by an amount which depends on the original level of wavefront distortion and on the level of correction provided by the AO.

The visibility loss due to the residual wavefront aberrations in an interferometer with an AO system can be calculated semi-analytically in some cases (Wang and Markey, 1978; Wilson and Jenkins, 1996) but a more flexible method to calculate these and other quantities that are affected by atmospheric wavefront perturbations is to use numerical simulations. This involves generating random wavefront perturbations across telescope apertures whose statistical properties are the same as those given by Kolmogorov–Tatarski theory, simulating the operation of the AO system in partially correcting these aberrations and simulating the combination of beams from two or more telescopes to form fringes. The simulation is run using a large number of different random realisations of the phase perturbations and the average values of quantities, such as the squared visibility, are computed. Typically a few thousand realisations are averaged to achieve results with accuracies of order a few percent.

The simulated perturbations are usually generated on a square grid representing a sampled wavefront; the sampling is usually at sub- $r_0$  intervals. One way to generate these *phase screens* is by filtering numerically generated ‘white noise’ so as to yield phase perturbations with approximately the correct power spectrum (McGlamery, 1976) but there are a host of other techniques which can be deployed. A wavefront perturbation simulation is included in the online material for this book (see Appendix B).

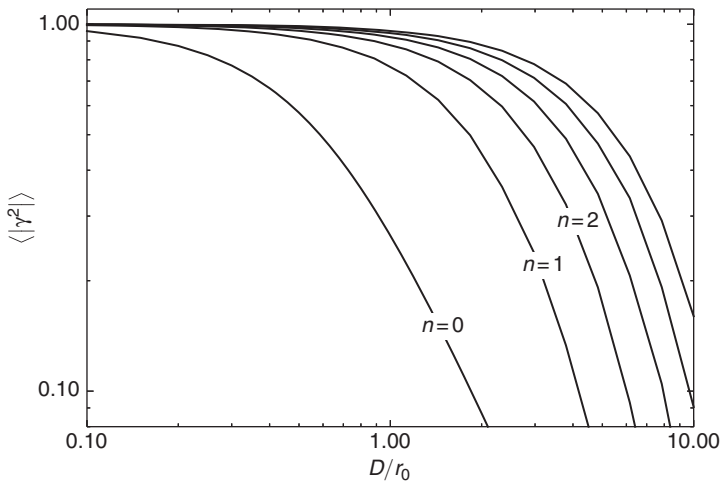


Figure 3.14 The mean-squared visibility loss  $\langle |\gamma|^2 \rangle$  as a function of aperture diameter  $D$  for an interferometer with an AO system at each telescope. AO systems which remove increasing numbers of low-order wavefront modes are shown and are labelled by the maximum Zernike radial order  $n$  corrected. An interferometer with no correction ( $n = 0$ ) is also shown for comparison. The graphs for  $n = 3 - 5$  are not labelled but follow the sequence of the lower-order graphs.

Simulation of the AO system can be quite complex depending on the level of realism required. For simplicity, it is often assumed that the AO acts as a perfect ‘Zernike filter’, removing all Zernike polynomials below some given order in a wavefront, and leaving as residual perturbations all the constituent higher-order polynomials.

Using these simulations, the mean-squared visibility loss for an aperture of diameter  $D$ ,  $\langle |\gamma|^2(D) \rangle$ , can be calculated for different levels of AO correction. The results from such a simulation are shown in Figure 3.14, where it can be seen that, as expected, AO systems decrease the visibility loss for a given size of telescope, and that the larger the number of modes corrected the larger the telescope can be before the mean-squared visibility loss factor falls below a given value, for example 50%.

### 3.6 Spatio-temporal effects

The previous sections have examined the effects of finite exposure time on infinitesimal apertures or the effect of finite aperture size with infinitesimal exposure time. In reality, both the exposure time and the aperture size will be finite, and to derive the visibility loss in this case is complex because the

frozen-turbulence model means that the temporal variations in the phase are coupled with the spatial variations.

One extreme model that illustrates this cross-coupling is to consider an interferometer consisting of a pair of large telescopes with perfect AO correction of all Zernike orders above piston. In this case the temporal variation of the piston phase will still give rise to fringe smearing over a finite integration time, but the change of the fringe phase will be slower than that of a point-like aperture under the same seeing conditions. This is because the piston phase measures the average phase over the aperture; the averaging operation filters out high-spatial-frequency ripples in the wavefront, and this means that the high-temporal-frequency ripples are also removed.

In practice, however, the fringe smearing for short exposures is not significantly affected by telescope diameter (Tubbs, 2005). There are a number of reasons for this. The first of these is that a significant contributor to the fringe motion is low-frequency but large-amplitude ripples in the phase, which are much less affected by ‘aperture filtering’. A second reason is that most AO systems used in interferometers are not perfect correctors and as a result the corrected wavefront contains significant residual high-order modes. Although these modes have zero average phase, they can affect the fringe phase because of the non-linearity of the dependence of the fringe phase on the aperture phase (Buscher *et al.*, 2008). Since they can evolve more rapidly than the piston phase they can contribute significantly to the short-timescale evolution of the fringe phase.

As a rule, therefore, treating the effects of spatial and temporal wavefront fluctuations on fringe visibility separately and multiplying the visibility loss factors for the two gives a reasonable approximation to the visibility loss experienced in practice. To achieve higher-accuracy calculation of the loss requires running a realistic simulation of the atmospheric temporal and spatial variations together with a realistic simulation of the spatio-temporal performance of the AO system.

### 3.7 Spatial filtering

An alternative method to remove the spatial atmospheric perturbations is to use *spatial filtering*, which in contrast to AO is a passive method that does not require any moving parts. As its name suggests, spatial filtering selectively removes the high-frequency ‘noise’ in a wavefront caused by the atmosphere, leaving the unaberrated component of the beam behind.

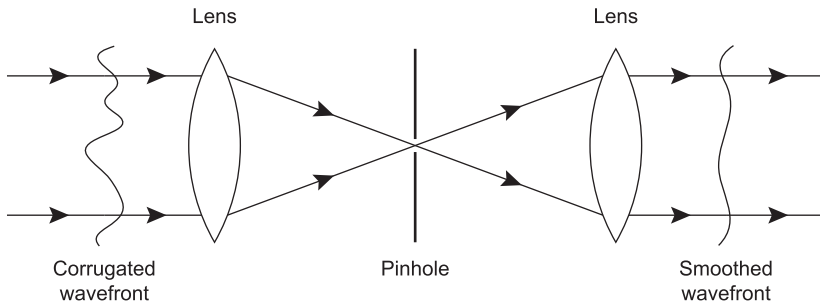


Figure 3.15 Layout of a spatial filtering arrangement using a pinhole.

The simplest form of spatial filter consists of a lens and pinhole arrangement as shown in Figure 3.15: this arrangement is commonly used for removing aberrations from laser beams. The diameter of the pinhole is chosen to be approximately the same as the first lobe of the Airy pattern from an unaberrated beam, i. e. approximately  $2.44f\lambda/D$ , where  $f$  is the focal length of the lens,  $\lambda$  is the wavelength of the radiation and  $D$  is the beam diameter. Aberrations to the beam cause some of the light to be scattered into a halo of speckles around the main lobe, but this light does not pass through the pinhole. At the output of the lens on the other side of the pinhole is therefore a filtered version of the input beam with only the low-frequency components of the input beam retained – the high-frequency components, including the wavefront corrugations, have been removed.

An even more effective form of spatial filtering can be obtained by replacing the pinhole with a length of single-mode optical fibre as shown in Figure 3.16. Optical fibres typically consist of a glass core surrounded by a lower-index glass cladding. Light is confined to the core by total internal reflection and so the fibre acts as an optical waveguide. As with all waveguides, the light propagates as a set of discrete ‘guided modes’, each of which is characterised by a distinct spatial pattern of the electromagnetic field within the waveguide. Single-mode (or ‘monomode’) fibres have cores only a few wavelengths across, and can support only one spatial mode (the “fundamental mode”) at the wavelength of operation.

Any input beam focussed on the fibre end can be thought of as being the sum of the fundamental mode and higher-order modes, and so a monomode fibre effectively filters out any component of the beam which does not match the fundamental mode; all other modes are either reflected from the fibre entrance or are dissipated in the cladding. As a result, the light coming out of the fibre has a

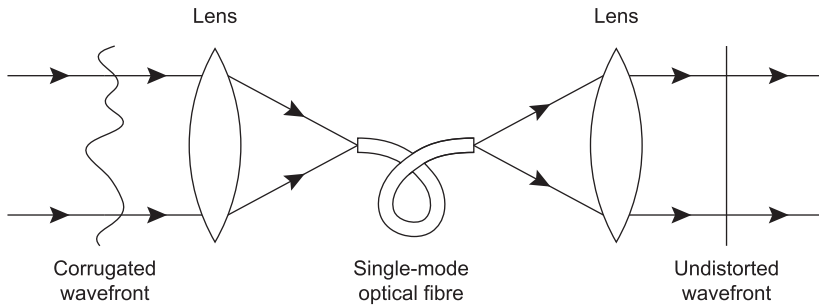


Figure 3.16 Layout of a spatial filtering arrangement using a single-mode optical fibre.

fixed amplitude and phase profile, which are independent of the amplitude and phase profile of the beam that is focussed on the fibre. Thus, a monomode fibre acts as a particularly ‘pure’ spatial filter, rejecting 100% of the ‘bad’ aberrated input beam and producing a beam with a fixed profile at the output.

Spatial filtering does not come for free, however, as components of the beam which do not match the filter profile go to waste. In the case of a pinhole filter, any light falling outside the pinhole is lost.

In the case of a monomode fibre, the loss calculation is slightly more complex. If a beam with a transverse amplitude and phase distribution given by the complex field amplitude  $\Psi(x, y)$  is focussed by a perfect lens onto a single-mode fibre then the amplitude and phase of the coupled fundamental mode is given by the complex ‘modal coefficient’,

$$a_0 \propto \iint_S \Psi(x, y) \Psi_0(x, y) dx dy, \quad (3.51)$$

where  $\Psi_0(x, y)$  is the so-called far-field pattern of the fibre fundamental mode (Wagner and Tomlinson, 1982). This pattern is equivalent to the wavefront which would be observed if a beam in the fundamental mode were to be propagated backwards through the fibre and the lens.

By the Cauchy–Schwarz inequality, the coupling efficiency of the beam into the fibre is therefore a maximum when  $\Psi(x, y) \propto \Psi_0(x, y)$  and decreases, the more dissimilar the incoming beam is to the fibre far-field pattern. The far-field pattern typically has a constant phase across the wavefront and so aberrated wavefronts will couple less light into the fibre.

Thus, spatial filtering is best used in combination with adaptive optics, with the AO providing the majority of the wavefront correction and the spatial filter ‘cleaning up’ the remaining aberrations. Figure 3.17 shows how the coupling efficiency of light from an AO-corrected telescope into a single-mode fibre



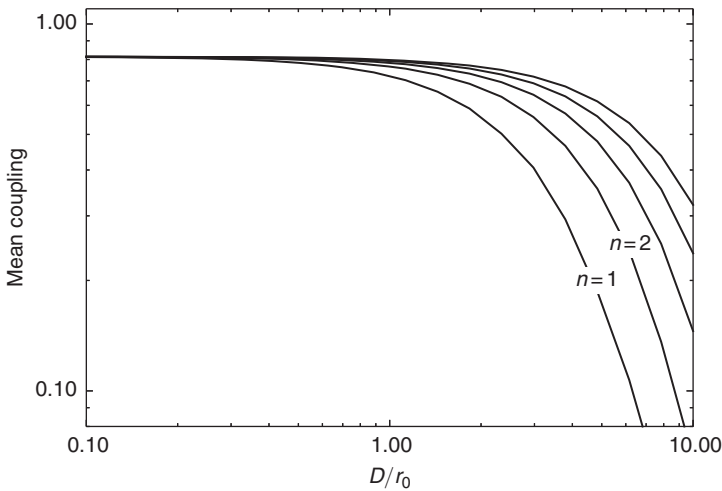


Figure 3.17 Fraction of starlight coupling into a single-mode fibre spatial filter as a function of the size of telescope for different levels of adaptive correction. The maximum Zernike radial order  $n$  removed by the adaptive optics system is indicated for each graph. The graphs for  $n = 3-5$  are not labelled but follow the sequence of the lower-order graphs. The fibre mode is assumed to have a far-field pattern that is Gaussian in shape and has a  $1/e$  radius which is 0.9 times the radius of the beam from the telescope: this choice gives approximately optimal coupling.

varies with aperture size and level of correction. These values were calculated using simulations similar to those described in Section 3.5.2. Even in the absence of atmospheric wavefront perturbations, the maximum coupling efficiency is approximately 80% because the uniformly illuminated beam from a circular telescope aperture does not match the Gaussian fibre far-field pattern (Shaklan and Roddier, 1988).

### 3.7.1 The effects of fibre coupling fluctuations

If atmospherically corrupted beams from two different telescopes are passed through single-mode fibres and then interfered, the fringe contrast will be high because all the spatial phase distortions in the incoming beams have been filtered out. However, even if the source is a point source (i. e. one with an object visibility modulus of unity), the instantaneous fringe contrast in any single short exposure of the fringes is unlikely to be unity. The reason for this is that the fraction of the incoming light coupling into each fibre is a fluctuating random variable, and so the two beams being interfered will, more likely than not,

have different intensities. The mismatch in intensity will cause a reduction in fringe contrast.

This reduction in contrast can be calculated by considering the interference of two beams with complex wave amplitudes  $\eta_1\Psi_0$  and  $\eta_2\Psi_0$  corresponding to the output of two fibres where the inputs both have wave amplitude  $\Psi_0$  (corresponding to a point source at the phase centre) and the intensity coupling efficiencies are given by  $|\eta_1|^2$  and  $|\eta_2|^2$ . From Equation (1.44), the fringe intensity pattern will be given by

$$\begin{aligned} i(x) &= |\eta_1|^2 \langle |\Psi_0|^2 \rangle + |\eta_2|^2 \langle |\Psi_0|^2 \rangle + 2 \langle |\Psi_0|^2 \rangle \operatorname{Re} [\eta_1 \eta_2^* e^{-2\pi i s x}] \\ &= F_0 (|\eta_1|^2 + |\eta_2|^2 + 2 \operatorname{Re} [\eta_1 \eta_2^* e^{-2\pi i s x}]), \end{aligned} \quad (3.52)$$

where  $F_0 = \langle |\Psi_0|^2 \rangle$  is the incident intensity. The fringe visibility modulus will therefore be given by

$$|V| = \frac{2|\eta_1||\eta_2|}{|\eta_1|^2 + |\eta_2|^2}. \quad (3.53)$$

Writing the ratio of the coupling amplitudes as  $f = |\eta_1|/|\eta_2|$  then

$$|V| = \frac{2}{f + 1/f}, \quad (3.54)$$

which is unity when  $f = 1$  but lower than unity for all other values.

Since the visibility depends only on the coupling ratio  $|\eta_1|/|\eta_2|$ , then if it is possible to measure this coupling ratio on an exposure-by-exposure basis then the visibility can also be calibrated on an exposure-by-exposure basis. This is known as *photometric calibration* and is discussed further in Section 8.8.

Since the visibility depends on how well matched the instantaneous intensity coupling factors  $|\eta_1|^2$  and  $|\eta_2|^2$  are, the RMS fringe visibility modulus depends on the level of coupling fluctuations: the larger the fluctuations the lower the RMS visibility will be. However, it is straightforward to show that under relatively general conditions the RMS coherent flux is dependent only on the mean coupling into the fibres and not the exposure-to-exposure fluctuations in the coupling.

Equation (3.52) shows that the coherent flux in a given exposure is given by

$$F_{12} = 2F_0\eta_1\eta_2 \quad (3.55)$$

and so its mean-squared modulus will be given by

$$\langle |F_{12}|^2 \rangle = 4F_0^2 \langle |\eta_1||\eta_2|^2 \rangle, \quad (3.56)$$

where angle brackets are used to denote averaging over multiple exposures. The intensity coupling terms can be written as

$$|\eta_1|^2 = \langle |\eta_1|^2 \rangle + n_1 \quad (3.57)$$

and

$$|\eta_2|^2 = \langle |\eta_2|^2 \rangle + n_2, \quad (3.58)$$

where  $n_1$  and  $n_2$  are zero-mean fluctuation terms. In terms of this decomposition, the mean-squared coherent flux modulus will be given by

$$\langle |F_{12}|^2 \rangle = 4F_0^2 (\langle |\eta_1|^2 \rangle \langle |\eta_2|^2 \rangle + \langle n_1 \rangle \langle |\eta_2|^2 \rangle + \langle n_2 \rangle \langle |\eta_1|^2 \rangle + \langle n_1 n_2 \rangle). \quad (3.59)$$

Since  $n_1$  and  $n_2$  are by definition zero mean, the second and third terms in the above expression average to zero and, if  $n_1$  and  $n_2$  are uncorrelated, the last term also averages to zero, giving

$$\langle |F_{12}|^2 \rangle = 4F_0^2 \langle |\eta_1|^2 \rangle \langle |\eta_2|^2 \rangle. \quad (3.60)$$

The assumption of uncorrelated fluctuations is a good approximation for coupling the light from well-separated telescopes into fibres, and under this assumption the mean-squared coherent flux is independent of the coupling fluctuations  $n_1$  and  $n_2$ . This means that it may be possible to calibrate the coherent flux by measuring the mean intensity coupling, in an alternative form of photometric calibration. This is explored further in Section 8.8.2.

### 3.8 Fringe tracking

Conventional adaptive optics works in the domain of individual collectors in an interferometer and cannot measure the piston-mode wavefront perturbations. The form of AO used to correct the piston perturbations in an interferometer is called fringe tracking.

The ‘wavefront sensor’ in this case is called the ‘fringe sensor’ as it is some form of beam combiner that uses light from a reference source (which is usually the science target itself) to form fringes between two or more telescopes. Measurements of the fringe patterns are used to derive estimates of the differential piston errors between pairs of telescopes and these estimates are used to adjust internal delays in the interferometer, using, for example, the delay lines, to compensate for the measured errors.

The main beam combiner of the interferometer can be used as the differential piston sensor, but many interferometers use separate beam combiners for making the ‘science’ measurements of the object visibility and for fringe

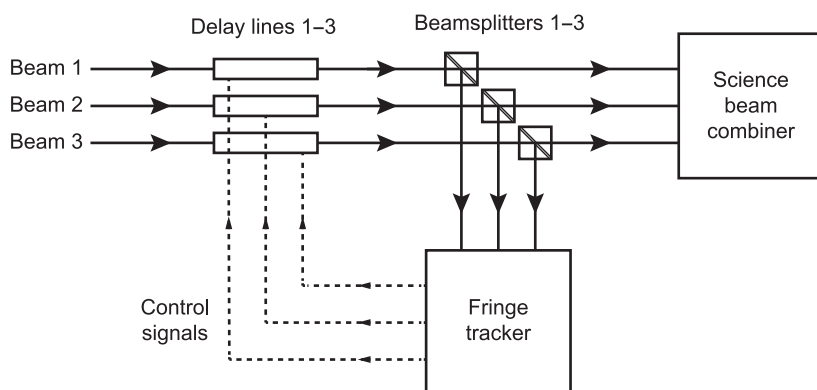


Figure 3.18 Example layout for an interferometer with a separate fringe-tracking beam combiner and science beam combiner.

tracking as shown in Figure 3.18. This is so that the two combiners can be optimised separately: the fringe-tracking combiner is typically optimised for speed and sensitivity at low light levels, while the science combiner can be optimised for fringe measurement precision, baseline coverage, spectral resolution and/or coverage and a host of other factors.

### 3.8.1 Cophasing and coherencing

Fringe tracking cannot entirely remove the effects of the atmospheric piston, as it is difficult to distinguish phase shifts due to the atmosphere from those due to the intrinsic structure of the reference source used for fringe tracking. Fringe tracking is therefore mainly concerned with reducing the effects of *changes* in fringe phase with time and with wavelength, and in practice the closure phase or spectral differential phase are still used as observables in order to remove ‘quasi-static’ errors in the visibility phase.

As described in Section 3.3, the change of the differential piston with time causes a visibility loss due to fringe smearing if the fringe exposure time is comparable to or greater than  $t_0$ . A fringe tracker can be used to compensate for this change in real time (i.e. on timescales  $\ll t_0$ ) so that the fringes on the science beam combiner are ‘frozen’. This freezing of the fringes is called ‘cophasing’ and means that the detector in the science beam combiner can be read out on timescales much longer than  $t_0$ . This allows cameras with long readout times to be used and also allows the effects of readout noise (discussed in Section 5.2.2) to be reduced.

The OPD errors due to the combination of atmospheric and instrumental effects in the interferometer can be of order  $100\mu\text{m}$  or more. Such a delay causes phase shifts between the fringes at neighbouring wavelengths and, as explained in Section 1.7, phase shifts within a spectral bandpass will cause a reduction in fringe visibility due to the finite size of the fringe envelope.

To take an example, the fringes observed using a bandpass corresponding to the astronomical infrared *J* band with a centre wavelength of  $1.25\mu\text{m}$  and a fractional bandwidth of 16% will have a fringe envelope which is approximately  $8\mu\text{m}$  wide. If the OPD error is uncertain at the level of many tens of microns, then most of the time no fringes will be visible. This implies that some form of active method of finding and tracking the coherence envelope is required unless considerably narrower spectral bandpasses are used.

The requirements of such a ‘coherencing’ system are not as severe as for a cophasing system, since the allowable error in the OPD is of the order of the fringe envelope width rather than being a fraction of a fringe. Since it takes longer for the OPD to change by an envelope width than to change by a fraction of a wavelength, the coherencing system can operate more slowly than a cophasing system. Coherencing methods will be examined in more detail in Section 6.4.

### 3.9 Wavelength dependence of atmospheric perturbations

In the above analysis it has been implicitly assumed that the OPD changes caused by turbulence are wavelength-independent. In reality, the OPD changes are caused by refractive-index changes in the air and the refractive index of the air changes with wavelength, as discussed in Section 1.8. Thus, atmospheric chromatic dispersion needs to be considered when considering the wavefront perturbations due to seeing.

The OPD changes due to atmospheric seeing are caused by the light traversing regions of differing temperature and humidity. The refractive-index change due to a small temperature offset  $\delta T$  is given approximately by

$$\delta n(\lambda) \approx \mu(\lambda) \frac{\delta T}{T_0}, \quad (3.61)$$

where  $\mu = n - 1$  and  $T_0$  is the ambient temperature in kelvins. At optical and near-infrared wavelengths the effects of temperature fluctuations are typically much larger than the effects of humidity fluctuations (Colavita *et al.*, 2004) and so if a delay perturbation of  $\tau(\lambda_1)$  is introduced at a wavelength  $\lambda_1$ , then at a wavelength  $\lambda_2$  the equivalent delay is

$$\tau(\lambda_2) = \frac{\mu(\lambda_2)}{\mu(\lambda_1)} \tau(\lambda_1), \quad (3.62)$$

and so the difference in delay at the two wavelengths is given by

$$\tau(\lambda_2) - \tau(\lambda_1) = \tau(\lambda_1) \left( \frac{\mu(\lambda_2)}{\mu(\lambda_1)} - 1 \right). \quad (3.63)$$

Figure 1.17 shows that  $\mu$  changes by about 2% over the visible region of the spectrum, and so this gives rise to a 2% error in the OPD calculated by assuming that the atmosphere is non-dispersive over this wavelength range. This error can be neglected when considering the turbulent fluctuations over an aperture of order  $r_0$  in size, where the phase perturbations at any given wavelength are of order 1 radian, as phase errors of 0.02 radians will typically have a negligible effect on the fringes.

In contrast, the turbulent perturbations to the fringes typically encountered in long-baseline interferometry are of the order of a few tens of microns, and so the error in assuming that the atmosphere is non-dispersive can be as large as a few hundred nanometres of OPD (i. e. several radians of phase) over the same wavelength range. When a fringe tracker is operated at a different wavelength from the science wavelength, it may therefore be necessary to compensate for the difference in phase change seen at one wavelength with that seen at another wavelength using knowledge about the chromatic dispersion properties of the atmosphere.

Problems can still arise because the magnitude of the turbulent water-vapour ('wet') fluctuations can become significant compared to a wavelength even though they are usually less than the temperature-induced 'dry' fluctuations (Colavita *et al.*, 2004). The wavelength dependencies of the wet and dry fluctuations are different, and so assuming that all the perturbations seen by the fringe tracker are 'dry' will lead to an error in the phase correction at the science wavelength.