

# 5

## Measurement noise

Fringe visibility measurements and derived observables such as the power spectrum and bispectrum are subject to both systematic and random errors. Chapter 3 showed how a systematic reduction in the fringe contrast can arise from the effects of atmospheric seeing. This chapter concentrates on the main sources of random errors or *noise*.

Noise on fringe parameters can arise both from the effects of atmospheric seeing and also from noise sources associated with measuring the light intensity levels in the fringe pattern. The main aim of this chapter is to derive robust ‘rule-of-thumb’ estimates for the noise levels in a given observation. These rules of thumb can then be used to determine which noise sources are dominant and whether random errors or systematic errors provide the fundamental limitation to accuracy in any given case.

### 5.1 Atmospheric noise

#### 5.1.1 Power spectrum

As discussed in Chapter 3, the spatial and temporal wavefront phase fluctuations which occur due to atmospheric seeing cause the mean fringe contrast to decrease as the apertures go from being point-like to being comparable in size to the Fried parameter of the seeing  $r_0$ , and as the exposure times go from being infinitesimal to being comparable to the coherence time of seeing  $t_0$ . Under these conditions the fringe contrast will also fluctuate on an exposure-to-exposure basis. Exposure times and aperture sizes need to be as large as possible in order to get more light, so it is helpful to be able to quantify what the trade-off between these experimental variables and the atmospheric noise level is.

An idea for how the noise varies as a function of integration time can be obtained using the ‘random-walk’ model for the visibility reduction used in Section 3.3.2. In this model, the coherent flux is given by a random walk consisting of  $n$  steps so that

$$F_{ij} = F_0 \sum_{k=1}^n e^{i\Phi_k}, \quad (5.1)$$

where  $F_0$  is the coherent flux in a single ‘step’ and  $\Phi_k$  is the fringe phase at step  $k$ . The mean power spectrum is therefore given by

$$\langle P_{ij} \rangle = \langle |F_{ij}|^2 \rangle = |F_0|^2 \left\langle \sum_{k=1}^n e^{i\Phi_k} \sum_{l=1}^n e^{-i\Phi_l} \right\rangle. \quad (5.2)$$

The order of summation and averaging can be exchanged to give

$$\langle P_{ij} \rangle = |F_0|^2 \sum_{k=1}^n \sum_{l=1}^n \langle e^{i[\Phi_k - \Phi_l]} \rangle. \quad (5.3)$$

In this summation, there are  $n$  terms for which  $k = l$  and  $n^2 - n$  terms for which  $k \neq l$ . The former terms are unity, while the latter terms are random phasors, which average to zero. Thus,

$$\langle P_{ij} \rangle = n|F_0|^2, \quad (5.4)$$

which is the familiar result that a random walk of  $n$  steps has a mean-squared length which is  $n$  times the step length.

The variance of the power spectrum can be calculated from

$$\text{var}(P_{ij}) = \langle P_{ij}^2 \rangle - \langle P_{ij} \rangle^2 \quad (5.5)$$

and expanding the summations as before. The second-order moment of the power spectrum is given by

$$\langle P_{ij}^2 \rangle = |F_0|^4 \sum_{k=1}^n \sum_{l=1}^n \sum_{m=1}^n \sum_{p=1}^n \langle e^{i[\Phi_k - \Phi_l + \Phi_m - \Phi_p]} \rangle. \quad (5.6)$$

The non-zero terms in this expansion occur when  $k = l = m = p$  ( $n$  terms), when  $k = l \neq m = p$  ( $n^2 - n$  terms) and when  $k = n \neq m = l$  ( $n^2 - n$  terms). Thus,

$$\langle P_{ij}^2 \rangle = |F_0|^4 (2n^2 - n), \quad (5.7)$$

and so

$$\text{var}(P_{ij}) = (n^2 - n)|F_0|^4. \quad (5.8)$$

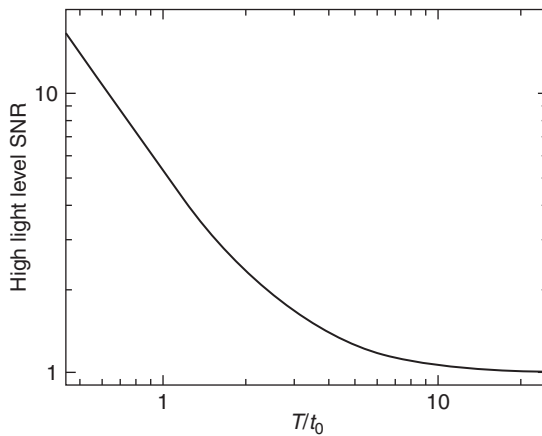


Figure 5.1 The SNR of atmospheric-noise-limited measurements of fringe amplitude as a function of exposure time. The SNR is defined as the mean-squared visibility modulus divided by the standard deviation of the squared modulus. From Buscher (1988a).

The ‘signal-to-noise ratio’ (SNR) is a common way of denoting the noise level in a dimensionless form and the SNR of the power spectrum will be given by

$$\text{SNR}(P_{ij}) = \frac{\langle P_{ij} \rangle}{\sqrt{\text{var}(P_{ij})}} = \sqrt{\frac{1}{1 - 1/n}}. \quad (5.9)$$

Thus, the atmospheric-noise-limited SNR is infinite for small exposure times (in practice it is limited by detection noise) and decreases towards unity for large  $n$ . It is notable that the SNR does not decrease to arbitrarily small values but saturates at a finite value – this is not true for detection-noise-limited observations or for atmospheric-noise-limited bispectrum measurements.

Figure 5.1 shows a plot of the SNR of atmospheric-noise-limited measurements computed using simulated temporal phase fluctuations corresponding to a Kolmogorov model. The results agree qualitatively with the results from the random-walk model in that the atmospheric noise SNR starts out at infinity for small integration times and then decreases to unity for integration times much greater than  $t_0$ .

The random-walk model can also be applied to evaluating the noise on apertures larger than  $r_0$ . This is borne out by simulations as shown in Figure 5.2, which show that the SNR falls to unity for apertures of a few  $r_0$  in diameter (somewhat unexpectedly the SNR falls to slightly below unity and then rises back towards unity, an indication that the random-walk model does not reveal all the complexities of the more realistic models).

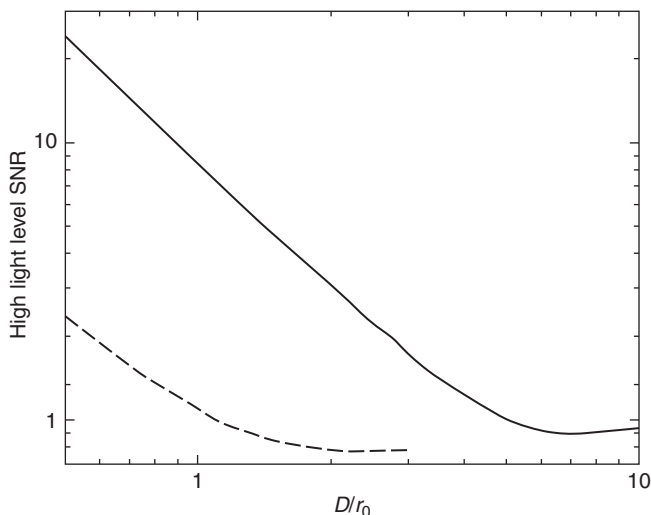


Figure 5.2 The SNR for atmospheric-noise-limited observations through apertures of different diameters (from Buscher, 1988a). The solid line corresponds to a tip-tilt-corrected aperture and the dashed line to an uncorrected aperture.

### 5.1.2 Bispectrum

For large aperture sizes or long integration times, the closure phase will also be subject to noise. The random-walk model again gives insight to the form the noise takes. The bispectrum measured in an integration time of  $nt_0$  or an aperture of area  $nr_0^2$  will be given by

$$\begin{aligned}
 T_{123} &= \frac{F_{12}}{n} \sum_{j=1}^N e^{i[\Phi_{1j}-\Phi_{2j}]} \frac{F_{23}}{n} \sum_{k=1}^N e^{i[\Phi_{2k}-\Phi_{3k}]} \frac{F_{31}}{n} \sum_{l=1}^N e^{i[\Phi_{3l}-\Phi_{1l}]} \\
 &= \frac{F_{12}F_{23}F_{31}}{n^3} \sum_{j=1}^N \sum_{k=1}^N \sum_{l=1}^N e^{i[\Phi_{1j}-\Phi_{2j}+\Phi_{2k}-\Phi_{3k}+\Phi_{3l}-\Phi_{1l}]} .
 \end{aligned} \quad (5.10)$$

The summation can be split into three sets of terms (Readhead *et al.*, 1988). There are  $n$  terms where  $j = k = l$ , which yield the unperturbed bispectrum. There are  $3n(n-1)$  terms ('terms of the second kind') of the form  $j = k \neq l$  or  $j = l \neq k$  or  $j \neq k = l$ . These terms can be paired as complex conjugates, for example the terms for  $j = q, k = q, l = s$  have the form

$$e^{i[\Phi_{1q}-\Phi_{1s}-\Phi_{3q}+\Phi_{3s}]} \quad (5.11)$$

while the terms for  $j = s, k = s, l = q$  have the form

$$e^{-i[\Phi_{1q}-\Phi_{1s}-\Phi_{3q}+\Phi_{3s}]} . \quad (5.12)$$

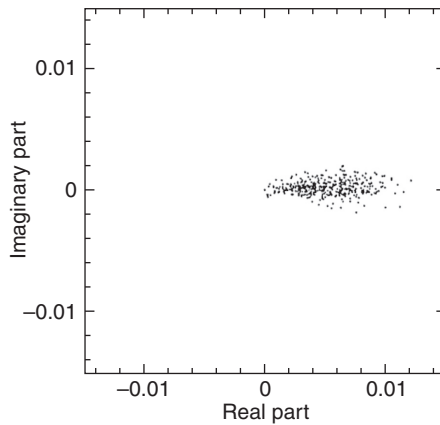


Figure 5.3 Sample bispectrum values plotted on an Argand diagram. The values were measured using aperture masking on a bright point source ( $\alpha$  Boo) with sub- $r_0$  subapertures. The scatter is larger in the direction parallel to the mean bispectrum value than in the perpendicular direction, suggesting that the exposure time used was much greater than  $t_0$ . From Haniff and Buscher (1992).

The sum of these two terms has no imaginary component and so adds no noise in the direction perpendicular to the mean bispectrum. As a result, the main effect of these terms is to cause fluctuations in the amplitude of the bispectrum and not of the phase (though they can cause the phase to be in error by  $180^\circ$ ). There are  $n(n-1)(n-2)$  terms of the third kind where  $j \neq k \neq l$ ; these have random phase and so add noise equally to the directions parallel and perpendicular to the mean bispectrum vector.

The above model predicts that for small  $n$  there will be more noise in the direction parallel to the mean bispectrum (primarily caused by the terms of the second kind) than in the direction perpendicular to the mean bispectrum (which would affect the phase). This is indeed what is seen in real data as shown in Figure 5.3.

If the bispectrum is averaged over  $N_{\text{exp}}$  exposures so that the noise on the averaged bispectrum is small, then the error on the closure phase, i. e. the phase of the bispectrum, will be approximately

$$\sigma = \frac{\sigma_{\perp}}{|T_{123}| \sqrt{N_{\text{exp}}}}, \quad (5.13)$$

where  $\sigma_{\perp}$  is the standard deviation of the noise in the direction perpendicular to the direction of the mean bispectrum. It is convenient to define the phase error  $\sigma_{\phi}$  for a single exposure

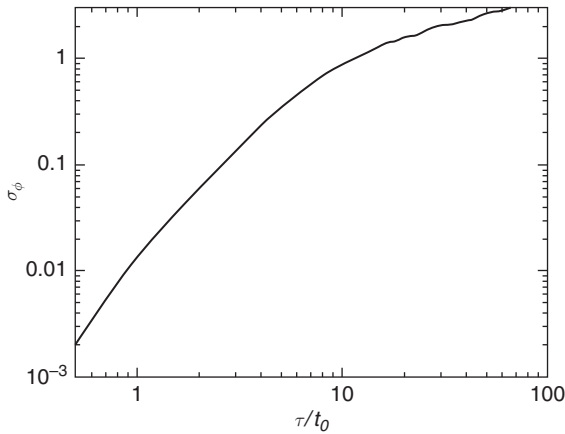


Figure 5.4 The phase error  $\sigma_\phi$  on the bispectrum as a function of exposure time  $\tau$  for atmospheric-noise-limited measurements. From Buscher (1988a).

$$\sigma_\phi = \frac{\sigma_\perp}{|\langle T_{123} \rangle|}. \quad (5.14)$$

Note that it is possible for  $\sigma_\phi$  to be much larger than  $2\pi$  but still have meaning: if the bispectrum is averaged over  $N_{\text{exp}}$  exposures, then the noise on the averaged bispectrum will be a factor  $\sqrt{N_{\text{exp}}}$  smaller than the noise on a single frame, and if sufficient exposures are averaged such that  $\sigma_\perp / \sqrt{N_{\text{exp}}} \ll |\langle T_{123} \rangle|$  then the noise on the phase of the averaged bispectrum will be  $\sigma_\phi / \sqrt{N_{\text{exp}}}$ . Thus, the phase error  $\sigma_\phi$  can be interpreted in terms of the number of exposures needing to be averaged in order to get a given phase standard deviation.

In the case of the random-walk visibility model,  $\sigma_\perp \propto \sqrt{N^3}$  for large  $N$  while  $\langle T_{123} \rangle \propto N$ , so the phase error will increase as  $\sqrt{N}$ . Figures 5.4 and 5.5 show that this model is a good guide to the results of more accurate simulations, with the error rising as  $\sqrt{\tau}$  for long exposure times and as  $D$  for large aperture sizes. For shorter exposure times and smaller apertures the phase error decreases more steeply, and the phase error remains below a radian for integration times of nearly  $10t_0$ . This can be explained in terms of the random-walk model as being due to the fact that the only noise terms which give rise to a finite phase error are of the third kind, and there are no such terms for  $n < 3$ .

## 5.2 Detection noise

In addition to atmospheric noise, the fringe measurements are affected by *detection noise*, which results when the light is converted into an electronic signal. Figure 5.6 shows the typical detection process schematically.

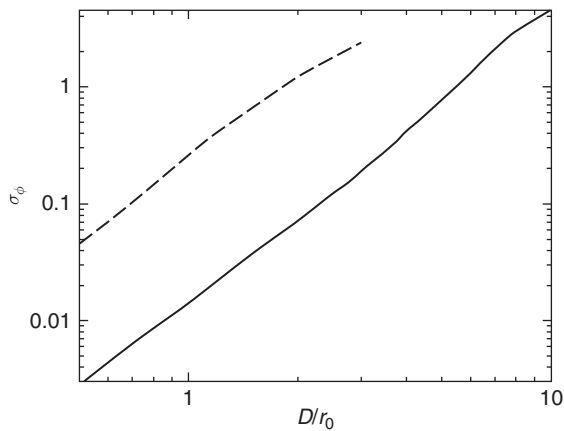


Figure 5.5 The phase error  $\sigma_\phi$  on the bispectrum as a function of aperture diameter for atmospheric-noise-limited measurements. The dashed line is for no adaptive optics correction and the solid line is for tip-tilt correction. From Buscher (1988a).

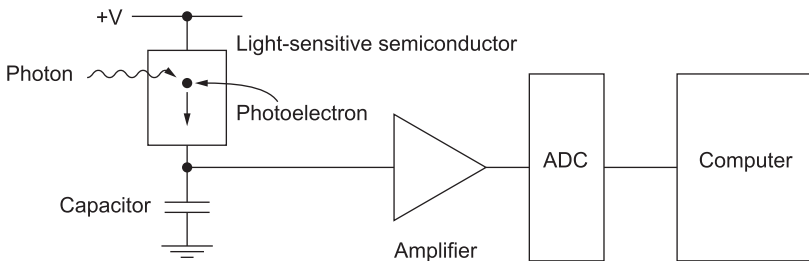


Figure 5.6 The signal-processing chain for a single pixel of a photoelectric detector. Light enters as photons on the left and exits as a digitised electrical signal on the right.

Detectors typically make use of the photoelectric effect in which incoming photons create free charge carriers in a material when it is exposed to light. A promising alternative detection mechanism makes use of superconducting materials in which the incoming photons break Cooper pairs rather than generating photoelectrons, but the noise issues present in the readout process are similar to those modelled here for photoelectric devices. The photoelectric detection material is usually a semiconductor, with semiconductors with different bandgaps being used at different wavelengths, for example silicon at visible wavelengths and mercury-cadmium-telluride (HgCdTe) at near-infrared wavelengths.

The photoelectric current is typically accumulated over the course of an exposure: this is represented in the diagram as the charging of a capacitor,

but the mechanism for charge accumulation varies with different technologies. At the end of the exposure, the voltage on the capacitor is read to determine the amount of charge accumulated. This typically involves several stages of charge and/or voltage amplification followed by digitisation using an analogue-to-digital converter (ADC). The digital signal is then sent to a computer for processing.

Most detectors consist of a matrix of pixels, each sensitive to the light falling on a small area, so that a two-dimensional intensity pattern can be read out. In order to reduce the amount of electronics in the system, these pixels are typically time-multiplexed: multiple pixel charge accumulators are connected one at a time to a single amplifier, though multiple amplifiers may operate in parallel to increase the readout speed.

The process of conversion of the light into an electronic signal introduces a number of sources of noise. The most important noise sources are discussed in the following subsections.

### 5.2.1 Photon noise

The quantum nature of electromagnetic waves means that, at the lowest light levels, light is detected as discrete ‘photo-events’. For semiconductor detectors, a photo-event consists of the generation of a single electron/hole pair in the material by the arrival of a photon. The number of photo-events fluctuates from exposure to exposure even if the light intensity is constant, and this fluctuation is called *shot noise* or *photon noise*.

In order to model phenomena such as ‘squeezed light’ correctly, the full quantum field theory of light, which treats the electromagnetic field as being quantised into photons from the start, is necessary. However, all the scenarios encountered in astronomical interferometry can be modelled to high accuracy using a simpler ‘semiclassical’ approximation (Mandel *et al.*, 1964), in which light waves propagate following classical laws (i.e. Maxwell’s equations), but are quantised on interaction with a detector.

In the semiclassical model, the number of photo-events in a given pixel in a given exposure follows a Poisson distribution with mean  $\Lambda_p$  given by

$$\Lambda_p = \int_{\text{exposure}} \int_{\text{bandpass}} \iint_{\text{detector surface}} \frac{\eta_p(x, y, \nu, t) I(x, y, \nu, t)}{h\nu} dx dy d\nu dt, \quad (5.15)$$

where  $I(x, y, \nu, t)$  is the classical specific light intensity (energy per unit area per unit time per unit frequency) falling on location  $(x, y)$  at frequency  $\nu$  and time  $t$ ,  $\eta_p(x, y, \nu, t)$  is the *quantum efficiency* (QE) of the detector pixel as a



function of location, frequency and time and  $h$  is Planck's constant (so  $h\nu$  is the energy of a photon at frequency  $\nu$ ). The QE is a value between 0 and 1 which relates the number of photo-events actually occurring in a detector to the maximum number of photo-events which could be recorded by an ideal device with the same incident light intensity. The mean of the distribution  $\Lambda_p$  will be referred to hereafter as the *integrated classical intensity* for a given pixel and exposure but it should be remembered that  $\Lambda_p$  is expressed in units of photo-events rather than joules.

The number of photo-events recorded in an exposure  $N_{\text{phot}}$  will have a mean of  $\Lambda_p$  and a standard deviation of

$$\sigma_p = \sqrt{\Lambda_p}. \quad (5.16)$$

For  $\Lambda_p \gtrsim 10$  the Poisson distribution can be approximated by a Gaussian distribution with the same standard deviation.

### 5.2.2 Readout noise

The electronics used to amplify the accumulated photocurrent signal in the detector will introduce additional noise and this is characterised as a *readout noise* (also known as *read noise*). This noise usually has a Gaussian distribution, which is typically independent of the signal level and uncorrelated from pixel to pixel (noise which is correlated from pixel to pixel is often called *pattern noise* and in good detectors is kept at levels well below the readout noise). Values for the readout noise can range from 100 electrons to less than an electron in devices which have near-noiseless amplification.

### 5.2.3 Background noise

Background noise arises from signals arriving at the detector from sources other than the object being observed. These sources include stray light, sky emission or, at infrared wavelengths longward of about  $2\text{ }\mu\text{m}$ , by thermal emission from the optics. These arise from causes independent of the detector, but are included in 'detection noise' because the noise characteristics are similar to other noise processes arising in the detector. Thermally induced 'dark currents' in the detector are indistinguishable from photoelectrons and can also be included as a form of background.

The noise arises primarily from the shot noise of the additional radiation or dark current received during an exposure, and so can be modelled by including a constant offset in the expression for  $\Lambda_p$  given in Equation (5.15). Background

noise differs from ‘normal’ photon noise in that its variance is independent of the brightness of the target object, and it differs from readout noise in that its variance increases as exposure time is increased whereas the readout noise variance is usually independent of the exposure time.

### 5.2.4 Electron amplification noise

Some detectors use a process known as ‘impact ionisation’ in semiconductors to increase the number of charge carriers generated by a photo-event. In avalanche photodiodes, the photocurrent is amplified before charge accumulation and in electron-multiplying CCDs the charge is amplified in an ‘amplification register’ after charge accumulation. In both cases, the amplification introduces extra noise, called amplification noise or excess noise. These noise sources can be approximately modelled as increasing the level of photon noise by a fixed factor.

## 5.3 Alternative fringe detection methods

The detection process described above assumes that the interfering light beams are brought directly together on a detector to form fringes. In radio interferometry quite different technologies are used to form fringes and this has a large impact on the form of the detection noise. Almost all radio interferometers amplify the incoming signal at each telescope and also use heterodyne fringe detection. Although heterodyne techniques have been used at mid-infrared wavelengths, neither of these technologies is in use at shorter wavelengths. The reasons for this are discussed below.

### 5.3.1 Heterodyne interferometry

In a *heterodyne* system, the incoming electromagnetic wave is converted before beam combination. The conversion is achieved by interfering the incoming wave with a locally generated signal at a similar frequency, the so-called ‘local oscillator’. At optical wavelengths this local oscillator would likely be a laser. The interference takes place on a fast detector, which converts the rapid oscillations in intensity from the ‘beats’ between the astronomical signal and the local oscillator into an electrical voltage. This electrical signal is then amplified and transmitted to a central location where fringes are formed electronically in a mathematically equivalent way to a conventional optical beam combiner. The interference signal

can be derived by summing the instantaneous voltages from each pair of telescopes and squaring the result or alternatively by multiplying the voltages.

The interference signal derived using a heterodyne interferometer is identical (apart from the noise) to the signal, which would be seen in an equivalent direct detection (*homodyne*) interferometer, but with the advantage that the signal can be amplified before transmission and the signals can be transmitted via cables rather than light pipes. Heterodyne technology is used in the Berkeley ISI interferometer (Hale *et al.*, 2000), which operates at a wavelength of 11  $\mu\text{m}$ , but has not been used at shorter wavelengths due to a number of technical difficulties as explained below.

### 5.3.2 Optical amplification

An alternative to heterodyne fringe detection is to use direct detection of the fringes but to use an optical amplifier to increase the intensity of the light beams before beam combination. This would allow the light from each telescope to be split multiple ways to form pairwise interference fringes with multiple other telescopes without the accompanying loss of signal level in each interference pattern. Such an amplifier at optical wavelengths would likely be in the form of a laser, in other words a cavity containing ‘pumped’ atoms, which coherently amplifies the incoming signal. No such *pre-amplifier* technology has been used in an optical interferometer to date.

### 5.3.3 Bandwidth limitations

Both pre-amplifiers and heterodyne systems allow the signal to be amplified at the telescope before transmission to the beam combiner, and therefore overcome losses in the beam transmission and beam combination system. However, there are a number of both technological and fundamental problems, which mean that, at wavelengths shortwards of about 5  $\mu\text{m}$ , these technologies are unlikely to see the kind of near-ubiquitous use seen in radio-frequency interferometers.

The first disadvantage of these techniques is that with current technology they work over relatively restricted bandwidths. Heterodyne technology in particular has a spectral bandwidth which is less than or equal to the highest frequency of intensity fluctuation that the detector can respond to. For the best modern detectors, this is of order a few gigahertz. This may sound like a significant bandwidth but a few gigahertz corresponds to a fractional bandwidth  $\Delta\nu/\nu$  of order  $10^{-5}$  at optical frequencies. Compared with homodyne systems

whose fractional bandwidth can easily exceed  $10^{-2}$ , a heterodyne system can suffer a large loss in sensitivity.

A way around the bandwidth limitation is to disperse the light spectrally and run many systems in parallel at different frequencies. This is becoming more feasible with heterodyne systems because of the availability of optical-frequency combs, which can provide local oscillators at hundreds of different frequencies simultaneously (Ireland and Monnier, 2014). In this case the inherently narrow bandwidth of the heterodyne system can be turned to advantage when observing narrow spectral lines in the objects of interest.

### 5.3.4 Quantum noise limits

A more fundamental limitation arises from the fact that both heterodyne systems and pre-amplifier systems include a ‘coherent’ amplification process. All amplifiers of electromagnetic radiation are fundamentally limited by an irreducible level of quantum noise. Amplifiers which are coherent in the sense that they preserve the phase information in the electromagnetic field will have a noise level which can be much higher than the photon noise associated with non-phase-preserving amplification processes such as those which occur in photomultipliers.

Any coherent amplification process can be shown to have a noise level which is equivalent to one photon per second per hertz of amplifier bandwidth at the input to the amplifier (Caves, 1982). This assumes a single-spatial-mode amplifier as would be appropriate for an interferometer: the more general result is that there is one noise photon per amplified ‘mode’ of the electromagnetic field. The noise is unavoidable and can be interpreted as arising from the sensitivity of any amplifier of electromagnetic fields to zero-point quantum fluctuations or ‘vacuum’ fluctuations.

At optical wavelengths, this noise source will be significant. For example, an optical amplifier with a bandwidth of 1 GHz will have an equivalent noise level of  $10^9$  noise photons per second at its input. This can be compared to typical photon rates of perhaps  $10^6$  photons per second in an optical interferometer observing a relatively bright object using a bandpass of many hundreds of gigahertz.

A more definitive idea of the effect of this noise at any wavelength can be obtained by comparing the amplification noise to the highest possible signal levels expected from typical astronomical sources (Prasad, 1994). These sources can be usually be approximated as thermal blackbody radiators at some temperature  $T$  and so the number of photons per mode of the radiation at frequency  $\nu$  from these sources will be given by

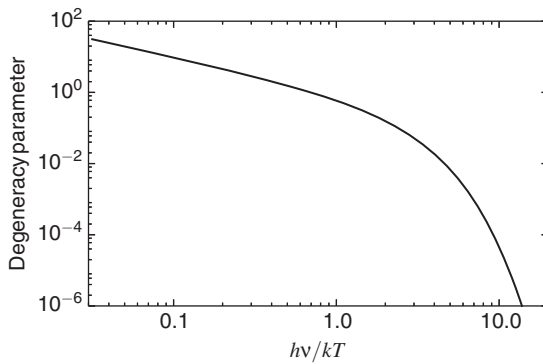


Figure 5.7 The photon degeneracy parameter (photons per unit frequency per unit time in a single spatial mode) for thermal radiation at frequency  $\nu$  from a black body at temperature  $T$ .

$$\delta = \frac{1}{e^{hv/(kT)} - 1}, \quad (5.17)$$

where  $h$  is Planck's constant and  $k$  is the Boltzmann factor (Mandel, 1961).

This 'photon degeneracy parameter' represents the maximum number of photons per unit bandwidth per unit time which can be injected into a single-mode optical fibre by a telescope observing a thermal radiator at temperature  $T$ . Note that this maximum is reached when the angular size of the radiator is larger than the diffraction limit of the telescope, which is unlikely to occur for most astronomical objects of interest to interferometry.

Since any amplifier will be subject to a noise input of one photon per unit bandwidth per unit time, then  $\delta$  is the maximum SNR of the output of an amplification system when observing a source at temperature  $T$ . It can be seen from Figure 5.7 that  $\delta$  falls below unity at frequencies  $\nu$  such that  $hv/kT > \log 2 \approx 0.69$ , and drops rapidly for  $hv/kT \gg 1$ . This means that at high frequencies the signal will be swamped by amplifier noise even when using the largest possible telescope and observing the largest possible object. For example, the radiation from an object at a temperature of 6000 K (typical for a star like the Sun) has a degeneracy parameter at a wavelength of 500 nm of  $\delta \approx 8.3 \times 10^{-3}$  and, under these circumstances, optical amplification systems are of less than no use.

The reason why using pre-amplification and heterodyne systems make sense in radio interferometry is that the much lower frequencies involved mean that the photon degeneracy parameter is significantly higher. For example, at a wavelength of 1 cm a 6000 K object has a photon degeneracy parameter of more than 4000. Even the coldest objects in the Universe, which are typically

at around the 3 K temperature of the cosmic microwave background, will have  $\delta > 1$  at this wavelength. Thus, the SNR penalty for amplification is negligible in almost all cases at radio frequencies and overwhelming in almost all cases at optical frequencies.

The crossover frequency at which amplification technologies become a practical option is somewhere in the mid- to far-infrared. For a ground-based interferometer, background radiation from the sky has an effective temperature of about 300 K and this background flux will comprise the majority of the radiation received by a telescope at wavelengths longer than about  $3\text{ }\mu\text{m}$ . This background flux will be greater than the noise introduced by a quantum-noise-limited amplifier for wavelengths longer than about  $70\text{ }\mu\text{m}$  and so at these wavelengths the noise penalty for using an amplifier may be outweighed by other practical advantages such as the ability to easily split the signal many ways.

These practical advantages might extend the usefulness of heterodyne interferometry to wavelengths as short as  $10\text{ }\mu\text{m}$  in interferometers with many telescopes (Ireland and Monnier, 2014), but it is unlikely that heterodyne systems will be useful at much shorter wavelengths because of the steep decrease in the photon degeneracy parameter with increasing frequency.

It should be noted that frequency upconversion, which is a process where the incoming photons are converted into photons at a higher frequency (Ceus *et al.*, 2013), is not subject to the amplification noise described above. Upconversion is useful because it can allow the use of optical fibres and detectors at frequencies where these technologies may be more mature than the same technologies at lower frequencies, but it does not result in a gain in the number of photons and so is not subject to the amplification noise of a true amplifier.

### 5.3.5 Intensity interferometry

An alternative but related form of fringe detection is intensity interferometry (Hanbury-Brown and Twiss, 1956). In this technique, the fluctuations in intensity of the light from a star are recorded using a fast detector at each telescope. The intensity signals are transmitted to a central signal combiner that measures the correlations in these fluctuations between telescopes, which are related to the modulus squared of the coherent flux on the projected baseline between the telescopes. The variation of the modulus of the coherent flux with baseline can be used to measure the diameters of stars (Hanbury-Brown *et al.*, 1967) and the separation of close binaries (Hanbury-Brown *et al.*, 1970).

An advantage of using an intensity interferometer over more conventional ‘amplitude’ interferometers is its insensitivity to mechanical or atmospheric

delays. This means that large but imprecise telescopes such as those present in the telescope arrays used for detecting Cherenkov radiation can be used as light collectors for intensity interferometry (Bohec and Holder, 2006).

However, even with large collectors intensity interferometry has so far been limited to the observation of bright stars. The reason for this is that, like heterodyne interferometry, the SNR of intensity interferometry is limited by the photon degeneracy parameter of the radiation, and for most astronomical objects this much less than unity at optical wavelengths as shown in Section 5.3.4.

An additional reason why intensity interferometry has not been developed as much as amplitude interferometry is that, although it may be possible to measure closure-phase-type quantities using intensity interferometry on three or more telescopes (Malvimat *et al.*, 2014), in most practical cases intensity interferometry is restricted to measurement of visibility-modulus information, which, as demonstrated in Section 9.6.2, severely reduces its applicability to the imaging of complex sources.

## 5.4 The interferogram

The following analysis assumes that the fringes are formed in by direct detection of un-amplified optical beams. The effects of detection noise on the measurement of fringe parameters can be understood from a simple model of the fringe pattern and how it is analysed. Different types of fringe pattern and different fringe analysis methods will be presented in Section 8.5, but the results from the simple model used here can be straightforwardly extended to apply to more complex practical examples.

The model of the detected fringe pattern consists of a one-dimensional sinusoidal light-intensity pattern sampled by a set of  $N_{\text{pix}}$  evenly spaced pixels. The integrated classical intensity on pixel  $p$  in a given exposure or ‘interferogram’ is given by

$$\Lambda_p = \frac{1}{N_{\text{pix}}} \left( \bar{N}_{\text{phot}} + \text{Re} \left\{ F_{ij} e^{2\pi i s_{ij} p} \right\} \right), \quad (5.18)$$

where  $\bar{N}_{\text{phot}}$  is the classical light intensity integrated over all pixels and expressed as a number of photons,  $F_{ij}$  is the coherent flux of the fringe expressed in photon units and  $s_{ij}$  is the spatial frequency of the fringes. It is assumed that there is an integral number of fringes across the interferogram, i. e. that  $s_{ij} N_{\text{pix}}$  is an integer.

The detected values for the integrated intensity in pixel  $p$  can be written

$$i_p = \Lambda_p + n_p, \quad (5.19)$$

where  $n_p$  is the noise value on pixel  $p$  for the given exposure. Note that Equation (5.19) implies that  $i_p$  and  $n_p$  are scaled to be in the same units as  $\Lambda_p$ , i.e. photons.

For simplicity,  $n_p$  will be assumed to have zero mean. The noise is assumed to be uncorrelated between pixels: this is an accurate approximation under most situations of practical interest. An additional assumption is that the noise has a Gaussian probability distribution, which is accurate in situations where read noise or background noise are the dominant sources of noise, or in situations where photon noise is dominant and  $\Lambda_p$  is larger than about 10.

The noise on pixel  $P$  is assumed to have a standard deviation  $\sigma_p$  that does not vary significantly from pixel to pixel, which is the case for many read-noise-limited detectors and for photon-noise-limited measurements if the fringe contrast is low so that  $|F_{ij}| \ll \bar{N}_{\text{phot}}$ . Results from this signal-independent Gaussian noise model give a good idea of the effects of noise in most situations of practical interest.

## 5.5 Noise on fringe parameters

### 5.5.1 Estimating the coherent flux

To derive a reliable estimate for a parameter such as the coherent flux  $F_{ij}$  from the noisy data  $i_p$  is a problem in statistical inference. Section 8.5 discusses the use of statistical ‘estimators’ in such problems and introduces a number of estimators for the coherent flux. For the purposes of this analysis the estimator that will be used is the discrete Fourier transform (DFT) of the intensity pattern at frequency  $s_{ij}$ .

The DFT estimator  $\hat{F}_{ij}$  for the quantity  $F_{ij}$  is given by

$$\hat{F}_{ij} = 2 \sum_{p=0}^{N_{\text{pix}}-1} i_p e^{-2\pi i s_{ij} p}. \quad (5.20)$$

Substituting Equations (5.19) and (5.18) into Equation (5.20) gives

$$\hat{F}_{ij} = F_{ij} + n_{ij}, \quad (5.21)$$

where  $n_{ij}$  is given by

$$n_{ij} = 2 \sum_{p=0}^{N_{\text{pix}}-1} n_p e^{-2\pi i s_{ij} p}. \quad (5.22)$$



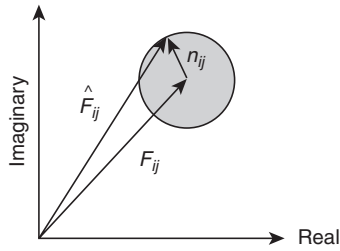


Figure 5.8 Argand diagram representation of the estimator  $\hat{F}_{ij}$  for the fringe coherent flux as a sum of the true coherent flux  $F_{ij}$  and the complex noise  $n_{ij}$ . It can be seen that the complex noise gives rise to both amplitude and phase errors in the estimate.

Thus, in the presence of noise  $\hat{F}_{ij}$  is offset from  $F_{ij}$  by a complex noise vector  $n_{ij}$  as shown in Figure 5.8. This complex noise causes both the amplitude and the phase of the measured coherent flux to deviate from the true value.

The sum in equation (5.22) is a two-dimensional random walk in the complex plane with a root-mean-square (RMS) ‘step size’ of  $\sigma_p$ . As the number of pixels becomes large, the probability distribution of  $n_{ij}$  will tend towards a two-dimensional circularly symmetric Gaussian. The variance of  $n_{ij}$  will be given by

$$\text{var}(n_{ij}) \equiv \sigma_{ij}^2 = 4 \sum_p \sigma_p^2. \quad (5.23)$$

For photon-noise-limited measurements,

$$\sigma_p = \sqrt{\Lambda_p}, \quad (5.24)$$

and so

$$\sigma_{ij} = 2 \sqrt{\bar{N}_{\text{phot}}}, \quad (5.25)$$

where

$$\bar{N}_{\text{phot}} = \sum_p \Lambda_p \quad (5.26)$$

is the integrated classical intensity summed over all pixels. The noise on the complex flux measurement is therefore independent of the number of pixels over which the light is spread.

For read-noise-limited measurements, assuming that the read noise is the same on all pixels so that  $\sigma_p = \sigma_{\text{read}}$ , then

$$\sigma_{ij} = 2\sigma_{\text{read}} \sqrt{N_{\text{pix}}}, \quad (5.27)$$

and so the noise gets worse as the light is spread out over more pixels. As a consequence, designs for beam combiners which feed read-noise-limited detectors try to concentrate the light into as few pixels as possible while those for photon-noise-limited detectors tend to be more relaxed in their pixel usage.

The SNR of the measurement is given by

$$\text{SNR}(\hat{F}_{ij}) = \frac{|\langle \hat{F}_{ij} \rangle|}{\sigma_{ij}}. \quad (5.28)$$

In the photon-noise-dominated regime the SNR is given by

$$\text{SNR}(\hat{F}_{ij}) \approx \frac{1}{2} |V_{ij}| \sqrt{N_{\text{phot}}}, \quad (5.29)$$

where  $V_{ij} = F_{ij}/\bar{N}_{\text{phot}}$  is the fringe visibility, while in the read-noise-dominated regime the SNR is given by

$$\text{SNR}(\hat{F}_{ij}) \approx \frac{V_{ij} \bar{N}_{\text{phot}}}{2\sigma_{\text{read}} \sqrt{N_{\text{pix}}}}. \quad (5.30)$$

Comparing Equations (5.29) and (5.30), it can be seen that, for photon-noise-limited measurements, the decrease in SNR caused by a given percentage loss in fringe contrast is larger than would be caused by the same percentage loss in photon rate, while the effects of visibility loss and photon loss are equal in the case of read-noise-limited measurements. These differences in the sensitivity to visibility loss versus photon loss have implications when optimising the SNR of a fringe measurement, and these implications are discussed further in Chapter 6.

### 5.5.2 Power spectrum bias and noise

If the object being observed is faint, the SNR in a single exposure is often less than unity. An obvious way to increase the SNR is to average multiple exposures, since with exposure times measured in milliseconds, thousands of exposures can be taken in a few seconds. However, the estimator  $\hat{F}_{ij}$  is unsuitable for such averaging because its phase will change randomly from exposure to exposure, and so the average will tend towards zero amplitude.

As the power spectrum and bispectrum are more stable in the presence of atmospheric phase disturbances, averaging these quantities over many exposures is possible. The effects of measurement noise on the power spectrum can be understood in terms of the simple model with Gaussian noise used above.

An initial guess at an estimator for the power spectrum would be

$$\hat{P}_{ij,\text{biased}} = |\hat{F}_{ij}|^2. \quad (5.31)$$

However, since this estimator is to be averaged over multiple exposures it is important that the average over a large number of exposures converges on the true power spectrum value. Substituting Equation (5.21) into Equation (5.31) and taking the average gives

$$\begin{aligned}\langle \hat{P}_{ij,\text{biased}} \rangle &= \langle |F_{ij}|^2 \rangle + 2\text{Re} \left\{ \langle F_{ij}n^* \rangle \right\} + \langle |n|^2 \rangle \\ &= \langle |F_{ij}|^2 \rangle + \sigma_{ij}^2,\end{aligned}\quad (5.32)$$

since the cross term  $F_{ij}n^*$  has random phase and therefore averages to zero. Thus, averaging the value of the estimator  $\hat{P}_{\text{biased}}$  over a large number of frames does not converge on the noise-free value and so this estimator is, as its name suggests, biased.

If the value of  $\sigma_{ij}$  is known, then an unbiased estimator can be constructed:

$$\hat{P}_{ij} = |\hat{F}_{ij}|^2 - \sigma_{ij}^2 \quad (5.33)$$

such that

$$\langle \hat{P} \rangle = \langle |F_{ij}|^2 \rangle. \quad (5.34)$$

The variance of the power spectrum estimator  $\hat{P}_{ij}$  can be estimated using the standard formula

$$\sigma^2(x) = \langle x^2 \rangle - \langle x \rangle^2, \quad (5.35)$$

where  $\sigma^2(x)$  is the variance of a quantity  $x$ . Substituting Equations (5.31), (5.33) and (5.21) into Equation (5.35), making use of the result that a circularly symmetric complex Gaussian distribution has a fourth-order moment  $\langle |n|^4 \rangle = 2\sigma_{ij}^4$ , and eliminating terms with zero mean gives

$$\sigma(\hat{P}_{ij})^2 = \text{var}(|F_{ij}|^2) + 2 \langle |F_{ij}|^2 \rangle \sigma_{ij}^2 + \sigma_{ij}^4, \quad (5.36)$$

where  $\text{var}(|F_{ij}|^2)$  is the variance of the squared modulus of the classical fringe amplitude due to atmospheric noise.

The SNR of the power spectrum is therefore given by

$$\text{SNR}(\hat{P}_{ij}) = \frac{\langle |F_{ij}|^2 \rangle}{\sqrt{\text{var}(|F_{ij}|^2) + 2 \langle |F_{ij}|^2 \rangle \sigma_{ij}^2 + \sigma_{ij}^4}}. \quad (5.37)$$

This expression gives the SNR of a single exposure: the SNR of the average of  $N_{\text{exp}}$  independent exposures is  $\sqrt{N_{\text{exp}}}$  higher than the single-exposure SNR. More general expressions for the noise on the power spectrum, which deal accurately with the combined effects of read noise and photon noise are given in Gordon and Buscher (2012), but the above expression gives a good idea of the SNR in most scenarios of practical interest.

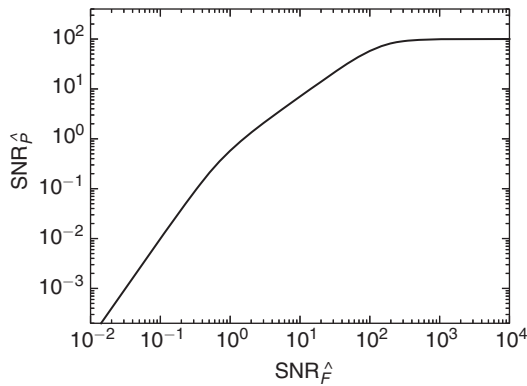


Figure 5.9 The SNR of the power spectrum as a function of the SNR of the coherent flux. The atmospheric-noise-dominated SNR has been assumed to be of order 100.

The expression for the SNR of the power spectrum given in Equation (5.37) can be split into three different regimes as evident from Figure 5.9. At the very highest light levels where the variation of the power spectrum due to atmospheric wavefront fluctuations dominates over the detection noise, i.e.  $\text{var}(|F_{ij}|^2)/\langle |F_{ij}|^2 \rangle^2 \gg \sigma_{ij}^2/\langle |F_{ij}|^2 \rangle$ , the SNR is independent of the light level:

$$\text{SNR}(\hat{P}_{ij}) \approx \frac{\langle |F_{ij}|^2 \rangle}{\sqrt{\text{var}(|F_{ij}|^2)}}. \quad (5.38)$$

This is the atmospheric-noise-dominated regime, and the SNR as a function of integration time and aperture diameter is given in Figures 5.1 and 5.2, respectively. At lower light levels where the detection noise is larger than the atmospheric noise but where  $\langle |F_{ij}|^2 \rangle \gg \sigma_{ij}^2$ , the SNR of the power spectrum is linearly proportional to the SNR of the coherent flux estimator given in Equation (5.28)

$$\text{SNR}(\hat{P}_{ij}) \approx \sqrt{\frac{\langle |F_{ij}|^2 \rangle}{2\sigma_{ij}^2}} \approx \frac{\text{SNR}(\hat{F}_{ij})}{\sqrt{2}}, \quad (5.39)$$

while at lower light levels where  $\langle |F_{ij}|^2 \rangle \ll \sigma_{ij}^2$  it varies quadratically with  $\text{SNR}(\hat{F}_{ij})$

$$\text{SNR}(\hat{P}_{ij}) \approx \frac{\langle |F_{ij}|^2 \rangle}{\sigma_{ij}^2} \approx (\text{SNR}(\hat{F}_{ij}))^2. \quad (5.40)$$

The quadratic fall-off in SNR at the lowest light levels means that there is a strong cut-off in the faintness of objects, which can be observed in practice. For example, in order to obtain an averaged power spectrum value with an SNR of 10 when  $\text{SNR}(F_{ij}) = 0.3$ , approximately  $10^4$  exposures need to be averaged. To obtain the same SNR at a lower light level where  $\text{SNR}(F_{ij}) = 0.1$ , approximately  $10^6$  exposures need to be averaged. If an exposure is taken every 10 ms then the former observation will take less than 2 min, while the latter will take nearly 3 h. In 3 h the  $(u, v)$  coordinate sampled by the pair of telescopes will have changed significantly due to Earth rotation, and so averaging the power spectrum over 3 h will be fraught with difficulty.

### 5.5.3 Bispectrum noise

In order to understand the noise on the bispectrum, the model of the fringe pattern from Section 5.4 needs to be extended to allow the measurement of the visibility at three spatial frequencies. The model used in this section will be of an interferometer consisting of three collectors and a set of three pairwise beam combiners. These combiners form fringe patterns on different detectors allowing independent measurements of the object visibility at spatial frequencies  $\mathbf{u}_{12}$ ,  $\mathbf{u}_{23}$  and  $\mathbf{u}_{31}$ . Some beam combiners instead multiplex all three fringe patterns in a single interference pattern as explained in Section 4.7.2, but the results derived using a pairwise model are a good guide to the results in more complex scenarios.

The coherent fluxes  $F_{12}$ ,  $F_{23}$  and  $F_{31}$  can be estimated using Equation (5.20) and the three estimators are denoted  $\hat{F}_{12}$ ,  $\hat{F}_{23}$  and  $\hat{F}_{31}$  respectively. A bispectrum estimator can be constructed from these as simply

$$\hat{T}_{123} = \hat{F}_{12}\hat{F}_{23}\hat{F}_{31}. \quad (5.41)$$

Substituting Equation (5.21) into Equation (5.41) gives

$$\begin{aligned} \hat{T}_{123} &= (F_{12} + n_{12})(F_{23} + n_{23})(F_{31} + n_{31}) \\ &= F_{12}F_{23}F_{31} \\ &\quad + F_{12}F_{23}n_{31} + F_{23}F_{31}n_{12} + F_{31}F_{12}n_{23} \\ &\quad + F_{12}n_{23}n_{31} + F_{23}n_{31}n_{12} + F_{31}n_{12}n_{23} \\ &\quad + n_{12}n_{23}n_{31}. \end{aligned} \quad (5.42)$$

The mean value of  $\hat{T}_{123}$  is the noise-free object triple product  $\langle F_{12}F_{23}F_{31} \rangle$ , since the rest of the terms in the expression have zero mean. Thus, the estimator  $\hat{T}_{123}$  is an unbiased estimate of  $T_{123}$ .

The variance of  $\hat{T}_{123}$  can be calculated by combining Equations (5.35) and (5.42). Eliminating the zero-mean terms and assuming that the standard

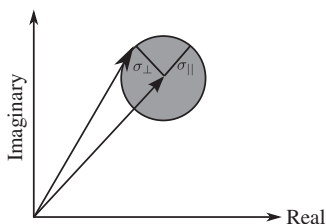


Figure 5.10 Schematic diagram of the noise on the bispectrum.

deviation of the noise is the same for all the fringe patterns,  $\sigma_{12} = \sigma_{23} = \sigma_{31} = \sigma$  gives

$$\begin{aligned} \sigma(\hat{T}_{123})^2 &= \sigma(F_{12}F_{23}F_{31})^2 \\ &+ \left( \langle |F_{12}F_{23}|^2 \rangle + \langle |F_{23}F_{31}|^2 \rangle + \langle |F_{31}F_{12}|^2 \rangle \right) \sigma^2 \\ &+ \left( \langle |F_{12}|^2 \rangle + \langle |F_{23}|^2 \rangle + \langle |F_{31}|^2 \rangle \right) \sigma^4 + \sigma^6. \end{aligned} \quad (5.43)$$

If the noise on each of the fringe coherent flux measurements has a circularly symmetric distribution in the complex plane, then it is straightforward to show that the noise on the bispectrum will also have a circularly symmetric distribution as depicted in Figure 5.10. Since the variance of a circularly symmetric noise distribution is given by  $\sigma^2 = \sigma_{\perp}^2 + \sigma_{\parallel}^2 = 2\sigma_{\perp}^2$ , then  $\sigma_{\perp}^2$  is half the value given in Equation (5.43).

Assuming that the seeing-induced fluctuations in the fringe amplitudes are small so that  $\text{var}(F_{12}F_{23}F_{31})$  is small,  $\left| \langle F_{ij} \rangle \right|^2 \approx \langle |F_{ij}|^2 \rangle$  and so on, then Equation (5.43) can be used to show that there will be two different SNR regimes, similar to those for the power spectrum.

When the SNR on all the baselines is sufficiently high that  $\langle |F_{ij}|^2 \rangle \gg \sigma_{12}^2$  and so on, then

$$\sigma_{\phi}^2(T_{123}) \approx \sigma_{\phi}^2(F_{12}) + \sigma_{\phi}^2(F_{23}) + \sigma_{\phi}^2(F_{31}), \quad (5.44)$$

where  $\sigma_{\phi}^2(F_{12})$  is a phase error on the coherent flux  $F_{ij}$  defined analogously to the phase error given in Equation (5.13).

Thus, in high-SNR conditions the phase errors of the component coherent fluxes add in quadrature to give the phase error on the closure phase (as might be expected if the component phases were simply added to give the closure phase).

When the SNR is low on all baselines,

$$\sigma_{\phi}^2(T_{123}) \approx \sigma_{\phi}^2(F_{12})\sigma_{\phi}^2(F_{23})\sigma_{\phi}^2(F_{31}); \quad (5.45)$$

therefore, in low-SNR conditions, the bispectrum phase error scales roughly as the cube of the phase error of a single coherent flux. In other words, there is an even more severe cut-off for using the bispectrum at low light levels than there is to using the power spectrum at these light levels.

## 5.6 Comparison of noise levels

The power spectrum and bispectrum are subject to a combination of atmospheric noise, detection noise and systematic calibration errors. The question as to which of these is most likely to be important is an obvious one when planning interferometric observations.

For sufficiently bright objects the atmospheric noise will dominate over detection noise. However, after incoherent averaging over relatively short periods the atmospheric noise can become less important than systematic errors such as calibration errors. To take an example, using a relatively long exposure time of  $3t_0$  will give an SNR for the power spectrum of approximately 1.5 for a single exposure (see Figure 5.1) and for the bispectrum the phase error will be approximately 0.1 radians (see Figure 5.4).

Assuming that the power spectrum and bispectrum are averaged over an observation of 100 s, if  $t_0 = 10$  ms then incoherent averaging will reduce the noise levels by a factor of  $\sqrt{100/0.03} \approx 58$  if the atmospheric errors are uncorrelated between exposures (which is a reasonable approximation providing that  $t_0$  and  $r_0$  are stable during the incoherent integration time). Thus, the atmospheric noise errors on the power spectrum will be at around the 1% level and on the closure phase at around the  $0.1^\circ$  level.

The systematic calibration errors in many cases are likely to be at least this large. Thus, atmospheric noise is not frequently an issue except in situations where extreme care has been taken to reduce calibration errors.

For faint objects, the atmospheric noise will be less than the detection noise. The faintest object which can be usefully observed will have an SNR after incoherent averaging of around 2 or so, meaning that if 1000 exposures are averaged the SNR per exposure is approximately 0.07. Thus, the detection noise in this case is a factor 20 larger than the atmospheric noise.

The conclusion that atmospheric noise is almost always negligible means that for most observations the dominant concerns when choosing parameters such as the exposure time are calibration accuracy and detection noise. The minimisation of the detection noise is the subject of the next chapter, and calibration issues are discussed in Section 8.8.