8
Data reduction

The 'raw' data from an interferometer consist of the measurements of the fringe pattern plus auxilliary data required for calibration. These data need to be converted into calibrated power spectrum and bispectrum data or coherently-average visibility data for subsequenty model-fitting and image reconstruction. The exact details of the data-reduction process varies between interferometric instruments and typically software is provided for each instrument that can perform the major parts of the process. This chapter provides an outline of what is going on inside this software in order to provide an understanding of the processes and the rationale for choosing one process over another when analysing a given dataset.

8.1 Scientific inference

The data-reduction process is part of a larger process, which aims to gain some knowledge about the astronomical object under investigation based on measurement of fringe patterns, and it is helpful to consider the process as a whole to understand where data reduction fits in.

The process of gaining knowledge based on measurements is known as *scientific inference*. A conceptual model of scientific inference starts out with an existing state of knowledge. This can be cast in terms of a model of the object, which has a number of unknown parameters. An example model is a binary star system consisting of a pair of stars with unknown brightnesses and diameters for the constituent stars and an unknown separation between them.

A particular set of values for all the model parameters can be thought of as representing a single point in a multi-dimensional space known as the *model space*. For a particular point in model space, the set of fringe measurements that would be produced by a given interferometer represents a point in the *data*
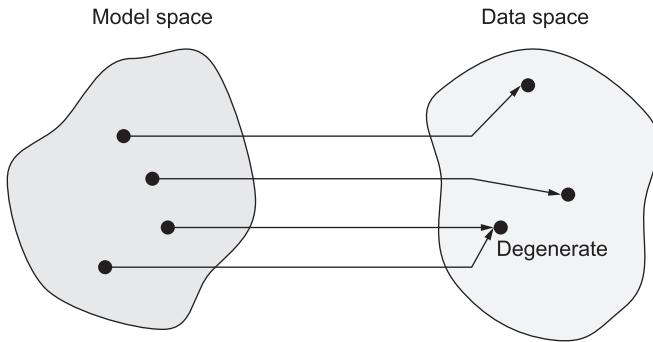
Figure 8.1 Schematic outline of the forward problem in scientific inference.

space of the instrument. This model of inference is shown diagramatically in Figure 8.1.

Computing the point in data space corresponding to a given point in model space is relatively straightforward and is known as the *forward problem*. The reverse process, that of taking a point in data space representing a set of measurements and inferring where in model space the data came from, is known as the *inverse problem* and is generally more difficult.

One difficulty is brought about by the presence of noise, which means that a single set of model parameters can lead to multiple possible measurement values. The problem may also be *degenerate* as shown in Figure 8.1: even in the absence of noise, multiple different sets of model parameters could result in the same values for the measured data. These difficulties mean that there may be no unique solution to the inverse problem, but instead a 'space' of possible solutions must be considered.

Solving the inverse problem lies at the heart of scientific inference, since the aim of the scientific method is to discern something about an unobservable model of the world based on its observable effects.

8.2 The forward problem

Understanding how to solve the inverse problem for an interferometer starts with studying the forward problem and this needs a model which includes both the object and the measurement process. The model used here attempts to balance generality against simplicity and so does not include all the important features of all interferometeric instruments. Nevertheless, it should provide a useful starting point for understanding the data-reduction process.

(a)

Source parameters

$\{\Theta_{\mathrm{m}}\}$

Source model

$I(\boldsymbol{\sigma})$

Ideal interferometer

$\{F(\boldsymbol{u}_{ij})\}$

Instrumental effects

$\{F_{ij}\}$

Beam combiner

$\{\Lambda_p\}$

Detector

$\{i_p\}$

Pixel gains

$\{r_p\}$

Raw data

(b)

Astronomy interpretation

$\{\hat{\Theta}_{\mathrm{m}}\}$ $\hat{I}(\boldsymbol{\sigma})$

Model fitting   Image reconstruction

Visibility calibration

$\{\langle\hat{F}_{ij}\rangle\}$

$\{\hat{P}_{ij}\}$ $\{\hat{T}_{ijk}\}$

Incoherent integration

$\{\langle\hat{F}_{ij}\rangle\}$

Coherent integration

$\{\hat{F}_{ij}\}$

Fringe decoding
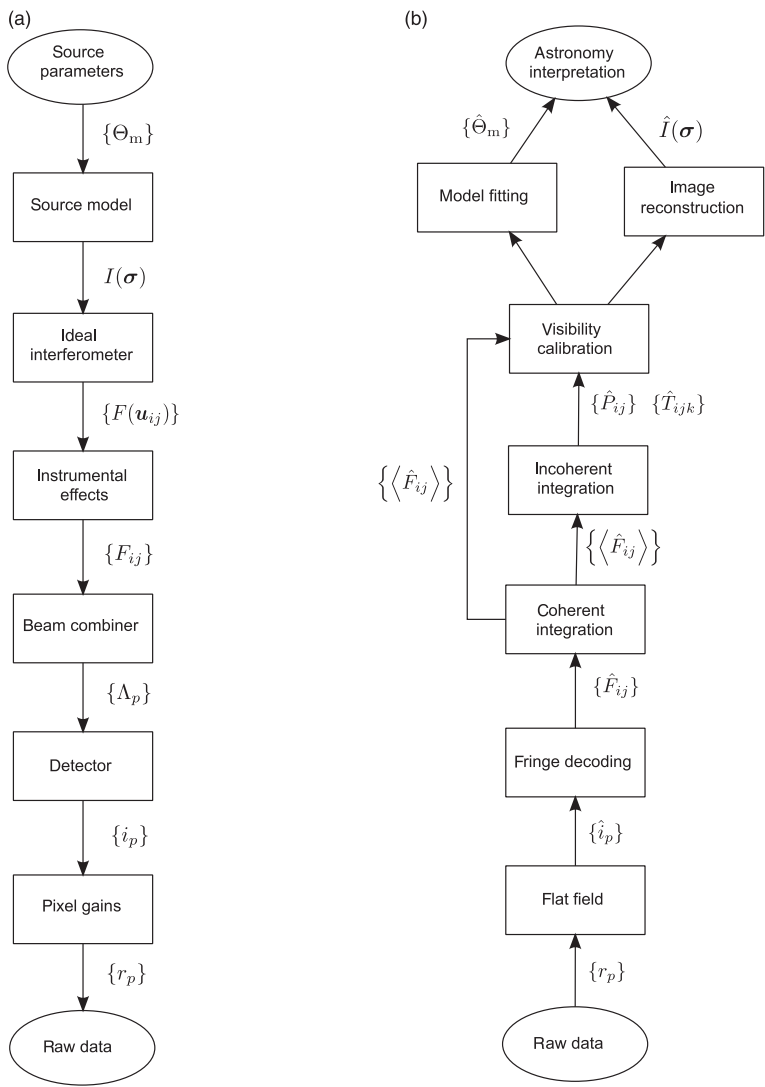
$\{\hat{i}_p\}$

Flat field

$\{r_p\}$

Raw data

Figure 8.2 The forward model of an interferometric measurement (a) and an outline of the data reduction process for interferometric data (b).

A flow diagram of the forward model is shown in Figure 8.2(a) and starts with a model for the object under study. The details of this model are clearly dependent on the particular object and so it is represented as an abstract model with a set of parameters $\{\Theta_m\}$.

Assigning a particular set of values to the object parameters corresponds to selecting a single point in the multi-dimensional model space. From the object parameter values the model will predict a brightness distribution on the sky $I(\sigma)$. This brightness distribution is observed by an idealised interferometer, which samples the object coherent flux at known locations $\{u_{ij}\}$ in the (u, v) plane, yielding coherent fluxes $F(u_{ij})$.

Instrumental and atmospheric effects will disturb the visibility modulus and phase from their ideal values. These disturbances are represented by introducing complex fringe degradation factors γ_{ij}, which relate the fringes seen in an idealised interferometer to those observed in the presence of these effects. The coherent fluxes of the measured fringes are given by

$$F_{ij} = \gamma_{ij} F(u_{ij}). \tag{8.1}$$

The beam combiner converts the coherent fluxes F_{ij} into a fringe pattern, which is represented as a set of discrete intensity samples $\{\Lambda_p\}$ where Λ_p is the integrated classical intensity falling on the detector as defined in Section 5.2. These samples could be pixel values at different spatial locations on a detector or temporal samples when fringes are scanned across a single pixel, or a mixture of the two. The individual samples will be called 'pixels' whether or not the samples actually come from a two-dimensional detector.

The samples are divided into a series of independent *interferograms* each consisting of N_{pix} samples. For temporally coded fringes, the interferogram could consist of a single 'sweep' of the pathlength modulation or it could consist of a subset of the sweep which contains sufficient data to extract the fringe parameters.

The sampled intensities $\{\Lambda_p\}$ are assumed to be related to the coherent fluxes F_{ij} by a linear superposition of 'carrier waves' (or 'coding waveforms') w_{ijp} so that

$$\Lambda_p = \sum_{i \leq j} \text{Re} \left\{ F_{ij} w_{ijp} \right\}, \tag{8.2}$$

The quantity F_{jj} represents the flux from an individual telescope and so w_{jjp} corresponds to the shape of the 'background' intensity pattern while w_{ijp} for $i \neq j$ corresponds to the modulation of the fringes on baseline ij. Note that the carrier waveform implicitly includes any pixel-to-pixel variations in the quantum efficiency of the detector, so that the intensity Λ_p is in units of photo-events.

By choosing appropriate forms for w_{ijp} most beam combiners can be modelled using the *fringe composition equation* given in Equation (8.2).

For example, to model an ideal two-telescope interference pattern consisting of a uniformly sampled sinusoidal one-dimensional fringe pattern with a uniform background, the waveforms can be represented by $w_{11} = w_{22} = \frac{1}{2}$ and $w_{12p} = e^{2\pi i sp}$, where s is the spatial frequency of the fringes expressed in cycles per pixel so that

$$\Lambda_p = \tfrac{1}{2}(F_{11} + F_{22}) + \text{Re}\left\{F_{12}e^{2\pi i sp}\right\}. \tag{8.3}$$

The detector converts the intensity values into electronic levels, which are digitised. The detection process is modelled as two independent steps, whereas in reality these steps are coupled to one another. First, the noise-free interferogram Λ_p is corrupted by zero-mean additive noise n_p, yielding noisy values i_p given by

$$i_p = \Lambda_p + n_p. \tag{8.4}$$

In the second step these values are multiplied by a pixel-dependent gain g_p (known as a 'flat field') and a systematic background offset b_p is added to give the 'raw' data values:

$$r_p = g_p i_p + b_p. \tag{8.5}$$

Non-linearity of the detector response could also be included in the model but is ignored here for simplicity.

8.3 The inverse problem

The inverse problem for interferometry is to take a set of raw intensity measurements $\{r_p\}$ and use these to learn something about the model of the object that produced these measurements in terms of the values of the parameters $\{\Theta_m\}$. One method for solving inverse problems is known as Bayesian inference, which is explained in more detail in Section 9.1.

A fully Bayesian solution to the inverse problem involves fitting the complete forward model of the system, which includes the object unknowns and all the instrumental unknowns such as the atmospheric wavefront effects and noise to the all of the raw data and any additional calibration data, to compute a multi-dimensional probability distribution of the model parameters. The inference process described here is not fully Bayesian; it is composed from a number of separate steps most of which are non-Bayesian. A similar multi-step process is used for analysing data at most interferometers, because it is mathematically and computationally more tractable for dealing with large amounts of data than a fully Bayesian solution.

Many of the steps in the data reduction process rely on the use of statistical *estimators*. An estimator is a function which maps a single point in data space (i. e. the actual set of measurements) to a single point in model space. This point represents an estimate for one or more model parameters based on the data. This can be contrasted with the Bayesian method, which uses a single set of measurements to derive probability estimates for every possible set of model parameter choices. The choice of which estimator to use for a given problem is somewhat ad-hoc, but well-chosen estimators can give results similar to those given by Bayesian analyses.

A schematic diagram of the inference process described here is shown in Figure 8.2(b), which starts from the raw data and computes a number of intermediate quantities before ending up with either an image of the object or a set of model parameters for the object. These steps are discussed briefly below and then expanded upon in later sections.

The first step of the process is to remove detector artefacts such as non-uniform flat fields and background levels to produces a set of estimates of the noisy intensities $\{\hat{i}_p\}$ (a circumflex is used distinguish the value of the estimator from the 'true' value of i_p). The intensity estimates are then used to derive estimates of the coherent flux $\hat{F}_{ij}$ on an interferogram-by-interferogram basis.

The coherent fluxes can be coherently integrated over a number of exposures to increase the signal-to-noise ratio (SNR) of the estimate. The coherent-flux estimates then are converted into power spectrum and bispectrum estimates, $\{\hat{P}_{ij}\}$ and $\{\hat{T}_{ijk}\}$, which are independent of the effects of atmospheric piston errors.

The data can be 'edited' at this stage to remove outliers before being incoherently averaged. The incoherently averaged power spectrum and bispectrum data are then calibrated for the effects of residual visibility corruption factors using data from observations of calibrator stars.

The final step is to use the calibrated data to constrain models or to reconstruct images of the object. This step is the subject of the next chapter; the data reduction process described in this chapter describes all the stages needed to go from the raw intensity data to the calibrated and incoherently averaged power spectrum and bispectrum data.

8.4 Flat fielding and background subtraction

The process for compensating for the detector gains g_p and offsets b_p relies on having accurate measurements for these values. These calibrations are usually in the form of 'darks' (i. e. exposures with the detector shuttered) and

'flats' (exposures where the detector is uniformly illuminated). The darks and flats can be analysed to produce estimates for $\hat{b}_p$ and $\hat{g}_p$, background and gain, respectively. The exact way these calibration measurements are taken and used are instrument-dependent and so further details for this process are not given here. Given these estimates, a set of corrected intensity values can be computed as

$$\hat{i}_p = \frac{\left(r_p - \hat{b}_p\right)}{\hat{g}_p}. \tag{8.6}$$

Note that, as defined in Equation (8.4), the units of i_p are the same as the units for Λ_p, i.e. detected photo-events, so the estimates for the gain need to be normalised on this basis. Scaling errors in this normalisation can affect the bias-correction stage (see Section 8.7).

8.5 Extracting the coherent flux

8.5.1 Sinusoidal model

In most beam combiners the carrier waveforms w_{ijp} for the fringes are sinusoids with slowly varying envelopes. For a one-dimensional pattern where the fringes are sampled in equal increments of phase, the waveforms can be represented as

$$w_{ijp} = a_{ijp}e^{2\pi \mathrm{i}s_{ij}p} \tag{8.7}$$

where s_{ij} is the frequency of the fringes caused by interference between telescopes i and j and a_{ijp} is the envelope function for those fringes. Typically, $s_{jj} = 0$ so that $w_{jjp} = a_{jjp}$, representing just a slowly varying background offset to the fringe pattern caused by the non-interfering light.

An example fringe pattern is shown in Figure 8.3. This fringe pattern consists of the superposition of sinusoids at three different spatial frequencies, each with a Gaussian envelope of a different width, on a flat background (i.e. w_{jjp} is constant for all p). This could represent, for example, the data from a temporally scanned beam combiner where the envelopes are due to longitudinal coherence effects (see Section 1.7) reducing the contrast of the fringes.

8.5.2 The discrete Fourier transform

With this data model it is natural to consider using a Fourier transform to extract the sinusoidal components from the data. The appropriate transform to use with discrete data is the discrete Fourier transform (DFT).
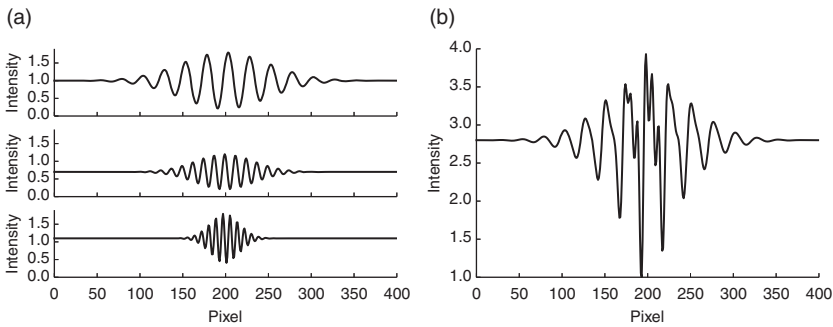
Figure 8.3 A discretely sampled fringe pattern with an envelope: individual fringe patterns (a) and the superposed pattern (b).
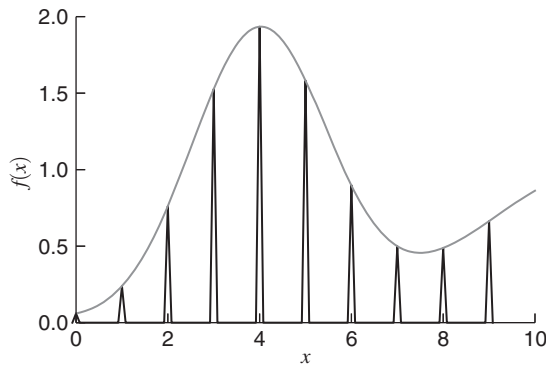

Figure 8.4 Discretising a continuous function represented as converting it to a series of Dirac delta functions.

For one-dimensional data the DFT is given by

$$I_k = \sum_{i=0}^{N_{\text{pix}}-1} i_p e^{-2\pi i k p/N_{\text{pix}}},\tag{8.8}$$

for $k = 0, 1, \ldots, N_{\text{pix}} - 1$. Equivalent two-dimensional discrete transforms are a natural extension of this but are not discussed in detail here.

The DFT can be understood as a conventional continuous Fourier transform in which the function to be transformed is represented as set of delta functions with unit spacing whose magnitudes (i. e. areas) are given by the sample values as shown in Figure 8.4. The DFT is then the continuous Fourier transform sampled in frequency space at intervals of $1/N_{\text{pix}}$ cycles per pixel.

The DFT is most commonly implemented on the computer using an algorithm known as the fast fourier transform (FFT), which, as its name suggests,
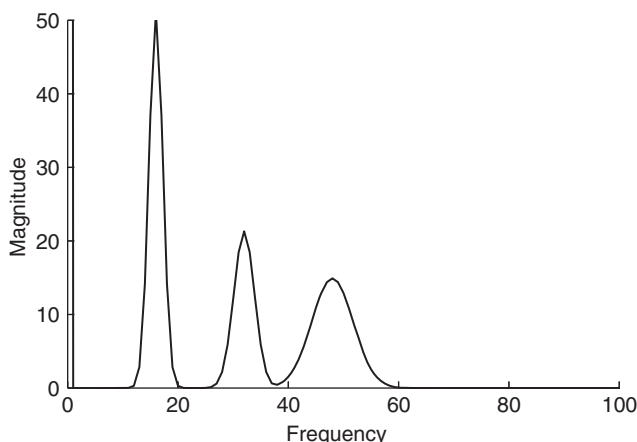
Figure 8.5 The magnitude of the DFT of the fringe pattern shown in Figure 8.3. The component at zero frequency has been truncated.

impements the DFT in a manner that is many times faster than a direct imple-mentation of the summation in Equation (8.8) if the Fourier coefficients at many frequencies are to be extracted. The FFT also has a numerical accuracy which is superior to more naïve methods.

Figure 8.5 shows the magnitude of the DFT of the fringe pattern shown in Figure 8.3. It can be seen that the three fringe patterns appear as peaks at dif-ferent spatial frequencies. These peaks do not just appear at a single frequency but are broadened, with the broadening being different for different peaks.

The reason for this broadening can be understood from the convolution the-orem. The component waveforms are the product of an envelope a_{ijp} and a sinusoidal component $e^{2\pi i s_{ij} p}$ and so the Fourier transforms of each waveform will be the convolutions of their Fourier transforms. The Fourier transform of the sinusoid will be a delta function at frequency s_{ij} and the transform A_{ijk} of the Gaussian envelope a_{ijp} will be a Gaussian whose width in the DFT is inversely proportional to the width of the Gaussian envelopes in the fringe patterns.

The value of the DFT at the discrete frequency k/N_{pix} which is closest to s_{ij} can be taken as an estimator for F_{ij}. More sophisticated approaches are possi-ble, for example interpolation of the DFT values, but these are not considered here. If s_{ij} is not known a priori, then it can be estimated from the DFT by looking for the frequency at which the Fourier magnitude is maximised for each peak in the spectrum.

The zero-spacing fluxes F_{ii} all contribute to the peak at the origin of the DFT, so if the envelopes a_{ii} are similar in shape then the individual fluxes

cannot be computed from the fringe pattern, only their sum. If it is desired to compute the individual fluxes, additional photometric channels are needed.

8.5.3 The *ABCD* estimator

The DFT estimator is widely used in interferometry. One particularly simple form of the DFT is the so-called *ABCD* estimator, This has been used in a number of interferometers which scan the fringes, such as the Mark III interferometer on Mt Wilson (Shao *et al.*, 1988) and the PTI (Colavita *et al.*, 1999).

This estimator is used when the fringe is sampled at intervals of $\pi/2$ in phase, and four successive samples are used to determine the fringe amplitude and phase. Labelling the four intensities as $i_0 = A$, $i_1 = B$, $i_2 = C$, $i_3 = D$, the fringe appears in the DFT at a frequency of $1/4$ so that

$$\hat{F}_{ij} = \sum_{p=0}^{3} i_p e^{-\mathrm{i}p\pi/2} \tag{8.9}$$

$$= A - C + \mathrm{i}(B - D). \tag{8.10}$$

Thus, the real and imaginary parts of the coherent flux are easily computed, and the amplitude and phase of the fringes can be computed as

$$\left| F_{ij} \right| = \sqrt{(A - C)^2 + (B - D)^2} \tag{8.11}$$

and

$$\arg F_{ij} = \arctan\left(\frac{B - D}{A - C}\right), \tag{8.12}$$

respectively.

8.5.4 Spectral leakage and windowing

Fringe-parameter estimators based on the DFT work well when the data can be accurately modelled by Equation (8.7). However, there are a number of practical situations where this model is not an accurate one.

One such situation is where the data are not evenly sampled in phase. This could be because the fringes are temporally scanned and the scan velocity is not constant, as might occur for a mechanical sawtooth-like scan when the scanning mirror is 'turning around' at the ends of the scan. Another cause for this uneven sampling could be if there are 'dead' pixels on a detector, which means that some of the samples are missing.

Both of these situations can be modelled by allowing the fringe envelope a_{ijp} to be a complex number, whose phase adjusts the phase of the sinusoid in an appropriate way to account for the uneven phase sampling. However, this will likely run into a second situation in which the DFT does not work well, which is when the fringe envelope is not sufficiently smooth and so 'spectral leakage' occurs.

As discussed in Section 8.5.2, the Fourier transform of the fringe pattern is the sum of a set of delta functions at the frequencies of the fringes s_{ij}, convolved with the Fourier transform A_{ijk} of the fringe envelopes. The Fourier transform of the fringe envelope can therefore be thought of as 'scattering' power from the fringe frequency to the neighbouring frequencies. If the envelopes are sufficiently smooth, A_{ijk} will be compact, and the fringe peaks will be well separated in the DFT. If the envelopes have high-frequency structure so that A_{ijk} is broad or if neighbouring fringe frequencies are too close together, then it is possible for the peaks to overlap and power to be scattered from one fringe peak into another, or from the peak at zero frequency to the lowest fringe frequencies.

A common example of such high-frequency structure is the sharp truncation of a sinusoidal fringe pattern due to sampling the pattern over a finite interval. If a sinusoidal fringe pattern is sampled over a integral number of fringe cycles, the DFT of this pattern will have a single peak at the fringe frequency. If there are a non-integral number of cycles, then the peak will be broadened so there is significant power at a range of frequencies as shown in Figure 8.6.

This effect can be seen as being due to the multiplication of an infinitely long sinusoid by a 'top-hat' function representing the finite length of the sampling. The Fourier transform of a top-hat of length T is a sinc function with zeros at frequency intervals of $1/T$. If the fringe frequency s_{ij} is an integer multiple of $1/T$, then the convolution of the sinc function with a delta function at s_{ij} has a zero at the origin, meaning there is no 'cross-talk' between the fringe and the zero-frequency level.

If the fringe frequency is not an integer multiple of $1/T$ then there is cross-talk arising from the sidelobes of the sinc function as shown in Figure 8.6. A similar cross-talk effect will also occur between neighbouring fringe frequencies if they are not separated in frequency by integral multiples of $1/T$.

This spectral leakage of power between nearby fringe frequencies due to truncation of the fringe pattern can be reduced by a technique known as 'windowing' or 'tapering'. This technique involves multiplying the intensity pattern $\hat{i}_p$ by a 'window function', which is tapered at the edges.

One example of a window function is a Gaussian whose values at the edges of the data are close to zero as shown in Figure 8.7. Multiplying by the window
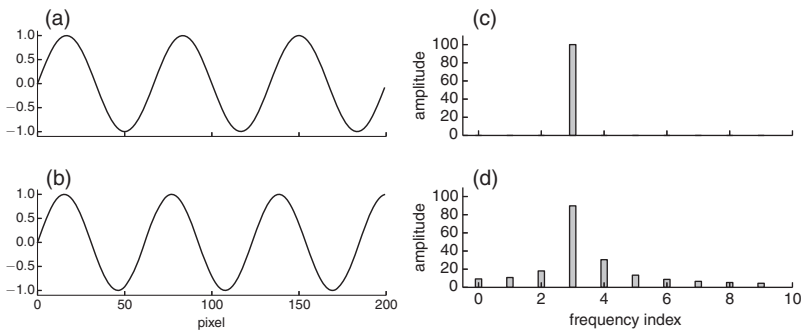
Figure 8.6 A sinusoid with an integral number of fringe cyles over the sampling interval (a) and with a non-integral number of fringe cycles (b) and their corresponding discrete Fourier transforms (c, d). The DC value of the fringe pattern has been set to zero to provide greater clarity.
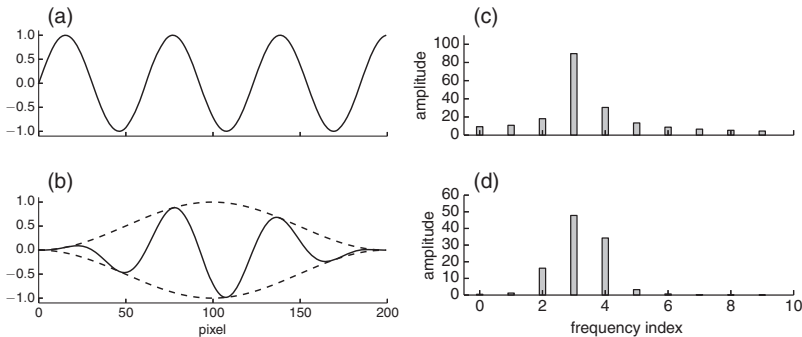

Figure 8.7 Windowing by a Gaussian (a, b) and its effect on spectral leakage (c, d).

function smooths the discontinuity at the edge of the sinusoid caused by truncation; in Fourier space, it corresponds to convolving the DFT with a function whose 'wings' are much more rapidly decaying than the sidelobes of the sinc function, and so the leakage to nearby frequencies is reduced.

8.5.5 The pixel-to-visibility matrix

Windowing only partially solves the problem of spectral leakage, and does so at the expense of SNR as it 'weights down' the data at the edges. In addition, there are other effects, such as uneven sampling, which cause spectral leakage and which cannot be solved by windowing. An alternative estimator that completely avoids spectral leakage can be found using matrix methods.

This method starts from noting that Equation (8.2) can be written in the form of a matrix equation,

$$\Lambda = Mf, \tag{8.13}$$

where $\Lambda = \Lambda_p$ is a vector of pixel intensities, $f = \{f_b\}$ is a vector of coherent fluxes and $M = \{M_{pb}\}$ is a matrix known as the *visibility-to-pixel matrix* or *V2PM*, since it converts visibilities (strictly speaking, coherent fluxes) into pixel intensities.

In order to make M a conventional two-dimensional matrix rather than a higher-dimensional tensor it is necessary to map the telescope index pairs to a unique one-dimensional baseline index $\{i, j : i, j = 1, \dots, N_{\text{tel}}\} \rightarrow \{b : b = 1, \dots, N_{\text{tel}}^2\}$. In order to make M and f real, it is necessary to also map the telescope pairs to a second index b' so that

$$f_b = \text{Re}\left\{F_{ij}\right\} \tag{8.14}$$

and

$$f_{b'} = \text{Im}\left\{F_{ij}\right\}. \tag{8.15}$$

The V2PM can then be written

$$M_{pb} = \text{Re}\left\{w_{ijp}\right\} \tag{8.16}$$

and

$$M_{pb'} = -\text{Im}\left\{w_{ijp}\right\}. \tag{8.17}$$

Given this representation of the forward equation as a matrix equation, the question arises if there exists an inverse matrix H such that

$$\hat{f} = H\Lambda \tag{8.18}$$

provides a 'good' estimate for the coherent fluxes f.

Since the V2PM is in general not square (but should rather be overdetermined, i. e. there should be more pixels than visibilities) there exists no unique inverse matrix. However, there does exist a *pseudoinverse* often called the Moore–Penrose pseudoinverse. This matrix has the property that it gives a least-squares solution to the problem; in other words, given a vector of noisy intensities $i = \{\hat{i}_p\}$, it provides a vector of coherent fluxes,

$$\hat{f} = Hi, \tag{8.19}$$

which minimises the squared difference χ^2 between the measured intensities and the intensities that would be predicted from the estimated coherent fluxes. In other words,

$$\chi^2 = |i - MHi|^2 \tag{8.20}$$

is minimised. This least-squares estimate is an optimal estimate of f from intensity measurements dominated by Gaussian read noise which is the same on all pixels.

The Moore–Penrose pseudoinverse can be computed from the V2PM by singular value decomposition (SVD), which comes as a standard package in many numerical libraries. The pseudoinverse in this case is is called the *pixel-to-visibility matrix* or *P2VM* since it converts pixel intensities to coherent fluxes or 'visibilities'.

The P2VM H can be computed providing that the V2PM M is known. This technique is used in the AMBER instrument on the VLTI (Tatulli *et al.*, 2007). As part of the AMBER calibration sequence, light is injected into the single-mode fibres to form fringes with known F_{ij} and this is used to estimate the matrix M, and hence to estimate H. AMBER also has photometric channels, which allow the fluxes from individual telescopes F_{ii} also to be estimated using a P2VM. In other instruments, it may not be possible to estimate the individual fluxes but only their sum.

The P2VM estimator is less directly applicable if the matrix M is unknown. This can occur, for example, if the carrier waveforms w_{ijp} change shape due to the effects of atmospheric turbulence. Nevertheless, it may still be possible to determine a suitable estimator for f, and if this estimator is linear it can be represented by a matrix.

An example of this is that the DFT method is robust to changes of shape in the envelope a_{ijp}, provided the envelope never has components at sufficiently high spatial frequencies that fringe power is scattered into nearby fringe components. The DFT estimator can be represented as a matrix H consisting of sinusoidal components

$$H_{bp} \propto \cos(2\pi s_{ij} p) \qquad (8.21)$$

and

$$H_{bp} \propto -\sin(2\pi s_{ij} p). \qquad (8.22)$$

In the following, it will be assumed that a linear estimator is used to estimate the coherent fluxes and that this estimator is represented by a P2VM H, independent of whether or not the matrix was computed using a direct pseudoinverse of M.

The P2VM formalism is a very general one. There is no requirement that the pixels are contiguous, so it can be applied to beam-combining systems where the fringe patterns for different baselines are superposed on a single camera, where the fringe patterns for different baselines are imaged by different cameras or where the information for a single baseline is extracted

from more than one camera, as may happen when complementary outputs of a beamsplitter-based combiner are used.

8.6 Coherent integration

8.6.1 Passive integration

The coherent fluxes estimated from a single interferogram may be noisy, and it was shown in Chapter 5 that at low light levels the power spectrum and bispectrum derived from a coherent flux with an SNR of less than one will be even more noisy. From this point of view it would be attractive to try to increase the SNR of the coherent flux estimate by averaging multiple exposures.

The simplest way of performing this *coherent integration* is to integrate the pixel data over N exposures.

$$\hat{i}'_{p,k_0} = \sum_{k=k_0}^{k_0+N} \hat{i}_{p,k},$$ (8.23)

where $\hat{i}_{p,k}$ represents the intensity estimate in pixel p on exposure k. This is essentially equivalent to taking longer exposures (except that there is more readout noise because there are more readouts), and so it suffers from the same problem as longer exposures – if the effective integration time is comparable to t_0 the fringes will smear out and therefore have lower visibility. This will lower the SNR of the fringe measurement and will tend to cancel out the intended effect of coherent integration.

8.6.2 Phase-referenced averaging

If there is a source of information as to how the fringes are moving between exposures, this can be used to compensate for the effects of fringe motion during a coherent integration using a technique known *phase-referenced averaging* or *software cophasing*.

The information about the fringe motion can come from several sources. One possibility is that the information comes from a dedicated fringe tracker. As described in Section 3.8, fringe trackers are typically used for *hardware cophasing*, in which the fringe sensor measures the atmospheric piston changes and compensates for these changes in real time using a movable mirror or similar hardware compensation element. This compensation will not be perfect, because the piston compensator will take a finite time to move to the required location, during which time the fringe will have moved. An alternative

to hardware cophasing (or something which can be used in conjuction with hardware cophasing) is to use the fringe tracker measurements for post facto compensation of the science combiner fringe phases in software. This can offer an improvement over hardware cophasing because software cophasing suffers from no 'lag' between phase measurement and correction. Indeed, because the correction is done in post-processing, the correction can make use of knowledge of the phase motion in the future as well as the past of the exposure being corrected.

Additional fringe-tracking information can come from the science combiner itself. Coherent integration is of value when the SNR of the fringes in a given science channel is low, and under these circumstances the data from that science channel has too low an SNR to provide the information for fringe tracking. There may, however, be other channels in the science combiner which do have higher SNRs. One example is if a wide-band channel is used to provide fringe tracking for a narrow-band channel. In a spectrally dispersed fringe pattern the wideband channel can be synthesised by adding together the data from all the narrow-band channels.

A common occurrence is that the fringe visibilities on different baselines can be widely different, and this can lead to a situation where two high SNR baselines form a closing triangle with a third baseline with much lower SNR. In this case, the phase changes seen on the two high-SNR baselines can be used to infer the phase changes on the low-SNR baselines, since the closure phase is constant – this is a form of 'baseline bootstrapping'.

Given an estimate of the fringe motion over a number of exposures, the fringe data can be compensated for this motion using a technique known as *phase rotation* or *fringe rotation*. To do this, the coherent flux $\hat{F}_{ij,k}$ is estimated for each exposure and then the phase of this coherent flux is adjusted based on an estimate ϕ_k of how the phase of the fringes in that exposure has changed compared to some reference exposure. This involves multiplying the visibility by a complex phasor

$$\hat{F}'_{ij,k} = e^{-i\phi_k}\hat{F}_{ij,k},\qquad(8.24)$$

which rotates the complex vector $F_{ij,k}$ in the Argand diagram.

Assuming that the phase change estimate ϕ_k is correct then the phase-rotated coherent fluxes for a sequence will all have the same phase. They can then be added together with no loss of visibility and a corresponding increase in SNR. If the read noise from the science combiner detector can be neglected, then the resultant sum has the same properties as the visibility from a long-exposure fringe measurement made with hardware cophasing.

Time-varying errors in the fringe-motion estimates can arise for a number of reasons, including low SNRs caused by periods of bad seeing ('drop-outs') and time-variable atmospheric chromatic dispersion effects between the fringe-tracking wavelength and the science wavelength. The longer the sequence of phase estimates, the greater the chance that there will be an error in the estimate of the fringe motion, which will cause a loss in fringe contrast, so coherent integration is often limited to the averaging of relatively short sequences of data. On the other hand, if it is possible to coherently integrate an entire sequence of data (Jorgensen *et al.*, 2010) then it may not be necessary to perform incoherent integration.

8.7 Incoherent integration

The coherently integrated data can be summarised by averaging quantities which are independent of any phase perturbations, such as the bispectrum and power spectrum.

8.7.1 Power spectrum

In Section 5.5.2 it was shown that a naive estimator for the power spectrum is biased in the presence of detection noise and that an unbiased estimator must be used. In deriving this estimator it was assumed that a DFT was being used to derive $\hat{F}_{ij}$ and that the noise was additive Gaussian noise, which was independent of the signal level in a given pixel. These assumptions can be relaxed to include the effects of Poisson noise (i.e. photon noise) and to allow a more general P2VM to be used at the expense of more complex mathematics (Gordon and Buscher, 2012).

The results from such an analysis are broadly in line with the results from the Gaussian model. The bias-free estimator based on a coherent flux estimator $\hat{F}_{ij}$ in the presence of photon noise and Gaussian read noise of σ_p on pixel p is

$$\left\langle \hat{P}_{ij} \right\rangle = \left| \hat{F}_{ij} \right|^2 - \sum_p \left(\hat{i}_p + \sigma_p^2 \right) \left| H_{bp} \right|^2 , \qquad (8.25)$$

where H_{bp} is the P2VM used in estimating $\hat{F}_{ij}$.

If there are uncertainties in the calibration of $\hat{i}_p$ in units of photo-events or in the value of the read noise σ_p then there may be difficulties in estimating the bias correction terms in Equation (8.25). In many cases a DFT is used for the P2VM and there are frequencies at which there is expected to be no fringe power. In these cases the power spectrum bias can be estimated from the 'background level' of the power spectrum at these frequuncies, since for a DFT $|H_{bp}| = 1$.

8.7.2 Bispectrum

If the coherent flux F_{ij} is estimated using a DFT and the detection noise is purely Gaussian, there is no bias correction term on the bispectrum. However, Poisson noise has a non-zero third-order moment and so it can induce a bias on the bispectrum. In addition, using a non-DFT P2VM can also lead to biases. A bias-free estimator in the case of a general P2VM H and a mixture of photon noise and Gaussian read noise is (Gordon and Buscher, 2012).

$$
\begin{aligned}
\hat{T}_{123} = \hat{F}_{12}\hat{F}_{23}\hat{F}_{31} & \\
& - \hat{F}_{12}\sum_p \left(\hat{i}_p + \sigma_p^2\right)H_{23,p}H_{31,p} \\
& - \hat{F}_{23}\sum_p \left(\hat{i}_p + \sigma_p^2\right)H_{12,p}H_{31,p} \\
& - \hat{F}_{31}\sum_p \left(\hat{i}_p + \sigma_p^2\right)H_{12,p}H_{23,p} \\
& + \sum_p \left(2\hat{i}_p + 3\sigma_p^2\right)H_{12,p}H_{23,p}H_{31,p}.
\end{aligned}
\tag{8.26}
$$

When the DFT is used so that $H_{ij,p} = e^{-2\pi i s p}$ then a simpler form is recovered

$$
\begin{aligned}
\hat{T}_{123} = \hat{F}_{12}\hat{F}_{23}\hat{F}_{31} & \\
& - \left|\hat{F}_{12}\right|^2 - \left|\hat{F}_{23}\right|^2 - \left|\hat{F}_{31}\right|^2 + 2N_{\mathrm{phot}} + 3N_{\mathrm{pix}}\sigma_{\mathrm{pix}}^2,
\end{aligned}
\tag{8.27}
$$

where it has been assumed that the noise on all N_{pix} pixels has the value σ_{pix} and N_{phot} is the total number of photons given by

$$
N_{\mathrm{phot}} = \sum_{p=0}^{N_{pix}-1}\hat{i}_p.
\tag{8.28}
$$

It is important for accurate bias subtraction in the bispectrum that the intensities $\{\hat{i}_p\}$ are calibrated in units of photo-events and that the read noise is well characterised.

8.8 Visibility calibration

8.8.1 Antenna-based gains

The forward model includes a visibility corruption factor γ_{ij} due to atmospheric and instrumental effects, which can cause large discrepancies between the object visibility function $F(\boldsymbol{u}_{ij})$ and the measured fringe parameters F_{ij}.

Some of the degration effects on the fringes can be represented by associating a complex 'gain factor' η_i with each arm of the interferometer (in radio interferometry these factors are known as *antenna-based gains*), so that

$$\gamma_{ij} \approx \eta_i \eta_j. \tag{8.29}$$

For example, piston wavefront errors ϵ_i associated with the atmosphere above a telescope i and the beam train from the telescope to the beam combiner can be incorporated as an antenna-based gain $\eta_i = e^{i\epsilon_i}$. Light losses in single-mode fibre beam combiners can be represented as values of $|\eta_i|^2$ that are equal to the coupling efficiency of light power into the fibre.

The antenna-based phase errors cancel out in the closure phase, and so no further calibration is needed to remove the effects of these errors. In radio interferometry, quantities known as 'closure amplitudes', which involve multiplying and dividing the magnitudes and coherent fluxes measured on four or more baselines are sometimes used to cancel the errors caused when $|\eta_i| \neq 1$, but these have not gained a foothold in optical interferometry.

8.8.2 Photometric calibration

An alternative way of compensating for the errors in the measured visibility modulus due to antenna-based gain errors is to use so-called *photometric calibration* methods. These methods rely on measuring the fluxes from individual telescopes F_i and F_j by 'peeling off' a small fraction of the light after it has undergone coupling into a single-mode system but before beam combination, as shown in Figure 8.8. These 'photometric channels' allow the measurement of the instantaneous coupling efficiencies since $F_i = |\eta_i|^2 F(0)$ where $F(0)$ is the flux from the object, and the product of these efficiencies is related to the coherent flux. The simplest way to use these photometric channels is to estimate the root-mean-square (RMS) visibility using the ratio

$$
\begin{aligned}
\hat{V}_{\mathrm{RMS},ij} &= \sqrt{\frac{\left\langle \hat{P}_{ij} \right\rangle}{\left\langle \hat{F}_i \hat{F}_j \right\rangle}} \\
&= \sqrt{\frac{\left\langle |\eta_i \eta_j| \left| F(\boldsymbol{u}_{ij}) \right|^2 \right\rangle}{\left\langle |\eta_i|^2 F(0) |\eta_j|^2 F(0) \right\rangle}} \\
&= \left| V(\boldsymbol{u}_{ij}) \right|.
\end{aligned} \tag{8.30}
$$

Thus, the dependence of the visibility on the antenna-dependent terms has been eliminated. Single-mode beam combiners with photometric channels such as
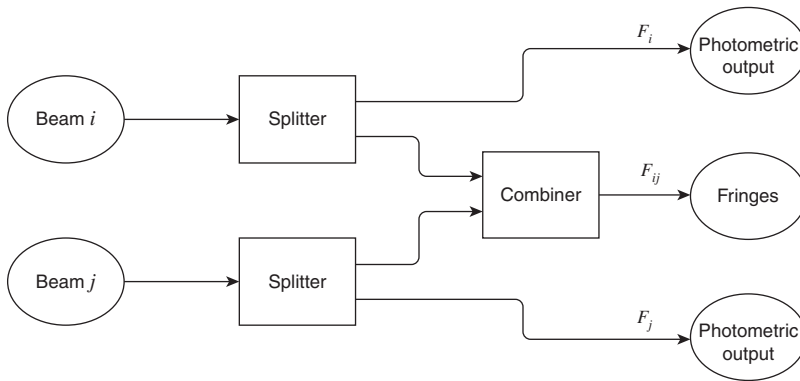
Figure 8.8 Schematic of a photometric calibration system.

the VINCI instrument on the VLTI have produced visibilities that have been calibrated at the 1% level or better (Ségransan *et al.*, 2003).

The fraction of the flux which is extracted in the photometric channels is usually quite small (a few percent in many cases) and so the instantaneous SNR of $\hat{F}_i$ may be less than unity for faint sources. In such a case the product term in the denominator $\hat{F}_i \hat{F}_j$ may have a low SNR, even after averaging over many frames. In these circumstances a preferable estimator to use might be

$$\hat{V}_{\text{RMS},ij} = \sqrt{\frac{\langle P_{ij} \rangle}{\langle \hat{F}_i \rangle \langle \hat{F}_j \rangle}}. \tag{8.31}$$

Using the methods employed in Section 3.7, it can be shown that

$$\langle \hat{F}_i \rangle \langle \hat{F}_j \rangle \approx \langle \hat{F}_i \hat{F}_j \rangle, \tag{8.32}$$

providing that the fluctuations in the coupling coefficients $|\eta_i|^2$ and $|\eta_j|^2$ are uncorrelated. This assumption will be true over short periods where the coupling fluctuations are due to changes in the instantaneous wavefronts across each telescope, but the fluctuations can become correlated if there is a systematic change in r_0 across both telescopes during the averaging period. Thus, it is best to apply the estimator given in Equation (8.31) to averages over timescales that are short compared to the expected timescale of any fluctuations in the value of r_0.

8.8.3 Transfer-function calibration

Many sources of visibility corruption, for example those due to fringe smearing during an exposure or polarisation mismatches, are baseline-dependent factors, that is to say they cannot be factorised into antenna-based gains as assumed in Equation (8.29). These effects therefore cannot be removed using closure-phase or photometric calibration. The most common way of calibrating these baseline-dependent factors is to use observations of 'calibrator stars' and this technique is described in outline in this section.

The following discussion will concentrate on the calibration of the power spectrum but most of the principles can be carried over straight-forwardly from calibration of the power spectrum to calibration of the bispectrum.

Calibration of the power spectrum at visible and near-infrared wavelengths is usually carried out in terms of the visibility rather than the coherent flux. The visibility is normalised by the zero-spacing flux, and this allows unknown values of the flux throughput of the system to be factored out. It also allows the use of calibrator stars whose diameters are well known but whose intrinsic brightness at the wavelength of observation are not well known.

In contrast, at mid-infrared wavelengths normalisation of the visibility is made difficult by the large thermal background levels, which make the zero-spacing flux less easy to determine than the coherent flux. As a result, a calibration approach is often adopted, which is similar to that used in radio interferometry: a set of calibrator objects is defined whose coherent fluxes as a function of baseline are well determined, and the calibration is done in terms of coherent fluxes.

For the purposes of this discussion, the convention used at shorter wavelengths, that of using the visibilities, will be used. The principles can be carried over straightforwardly to using coherent fluxes.

The averaged squared visibility can be written in terms of the object visibility and a multiplicative error $\left\langle \left| \gamma_{ij} \right|^2 \right\rangle$ known as the power-spectrum *transfer function* so that

$$\left\langle \left| \hat{V}_{ij} \right|^2 \right\rangle = \left\langle \left| \gamma_{ij} \right|^2 \right\rangle \left| V(\boldsymbol{u}_{ij}) \right|^2. \tag{8.33}$$

The square root of the power-spectrum transfer function is known as the *system visibility*. The simplest assumption is that the system visibility is constant. The transfer function can then be estimated by observing another object (the
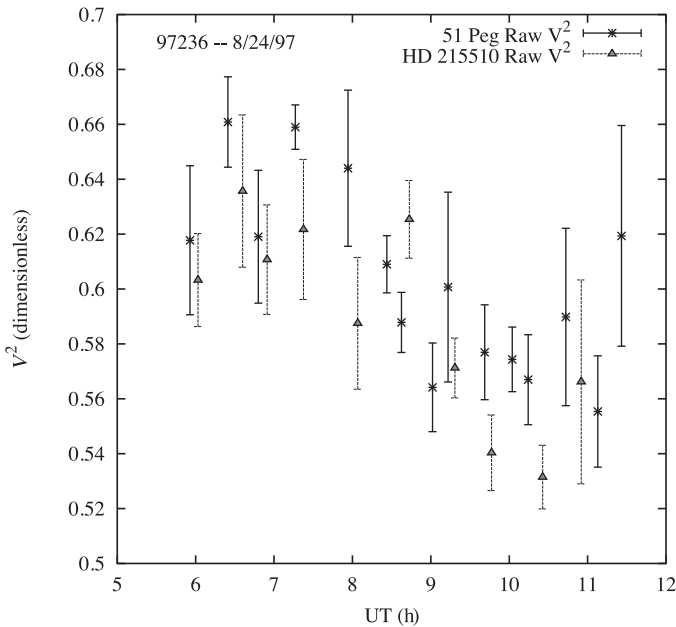
Figure 8.9 Squared visibility measurements from the PTI as a function of time of night. Measurements from both the target star (51 Peg) and the calibrator star (HD 215510) are shown. The authors of the paper conclude that there is little evidence for the target star being resolved, and so the majority of the visibility variation seen in both the target and the calibrator is due to variations in the system visibility. From Boden *et al.* (1998).

calibrator star) whose object visibility function $V(\boldsymbol{u}_{ij})$ is known, typically a star of known diameter. However, the assumption that the system visibility is constant is usually not a good one. An example set of visibility modulus data is shown in Figure 8.9; it can be seen that there may be large variations of the system visibility between different observations of the same star. This is because the system visibility is the product of a number of factors, each of which can vary with time.

The most obvious factor which can change with time is the seeing. Not only are the wavefront perturbations variable from exposure to exposure, but the mean parameters of the seeing r_0 and t_0 also vary with time and with pointing direction. Baldwin *et al.* (2008) report variations of r_0 by a factor of two in the space of less than a second, but typically significant variations in seeing take place on timescales of minutes. Slow variations in seeing can be compensated for by 'bracketing' the observation of the science target with two calibrator

observations and then interpolating the system visibility between the calibrator observations.

If there are many observations of calibrator stars during a given observing session, then this technique can be extended to model the system visibility as a function of time and to interpolate this model to the times of observation of the science targets. A further extension is to use the calibrator visibilities together with data from other sources in the interferometer to construct a model of the system visibility as a function of variables which may affect it. Any variable that can be monitored and which might correlate with system visibility can be used. For example, the strength of the temporal seeing can be estimated from the fringe-tracker data while the strength of the spatial seeing can be estimated from the adaptive-optics system data. The group-delay difference between the fringe tracker and the science combiner will likely correlate with spectral dispersion effects and the telescope elevation and azimuth will likely correlate with orientation-dependent telescope vibrations and polarisation effects. Environmental variables such as the temperature and pressure in the delay line can also be used in the model. In many interferometers, all variables that may have an effect on the system visibility are routinely recorded together with the interferometric data.

A multi-dimensional model of the system visibility as a function of these variables can be fitted to the calibrator data, and this model can then be used to estimate the system visibility during the observations of the science targets. The more calibrator observations that are available, the more variables can be used in the model, and so this technique has been used with most success on interferometers that are able to make large numbers of such observations a night, for example the Mark III interferometer on Mt Wilson and the PTI (Colavita *et al.*, 1999).

The error in calibration of the system visibility is often the largest source of measurement error for the visibility modulus, and so estimating this error is critical to later stages of data reduction. Large calibration datasets allow the calibration error to be determined from the misfit between the system visibility model and the data it is based on. With careful modelling these errors can be reduced to around 1% (Mozurkewich *et al.*, 2003).

8.9 OIFITS files

The data-reduction process for an interferometer will have many features which are instrument-specific, but the product at the end of this process has been standardised between interferometers, in the form of a file format.

This format is called OIFITS and is documented by Pauls *et al.* (2005). The format allows for the storage of calibrated power spectra and closure phases and also for the storage of calibrated averaged visibilities – these can be derived from coherent averaging or used to represent spectral differential phase data.