

# An introduction to the theory of interferometry

Chris Haniff

*Astrophysics Group, Cavendish Laboratory, University of Cambridge,  
J.J. Thomson Avenue, Cambridge, CB3 0HE, UK*

---

## Abstract

An understanding of some of the physical bases of interferometric imaging can be helpful both in designing and understanding interferometric telescope arrays and in planning, executing and interpreting astronomical interferometric observations. This paper presents a brief introduction to some of these key principles at a level suitable for those who are new to interferometry, and with particular emphasis on four key areas: classical imaging theory, coherence functions, interferometric observables, and interferometric imaging. These topics underpin the practice of much of interferometric astronomy (at optical as well as other wavelengths), and provide a valuable basis from which to develop a better understanding of the operation of the VLTI in later chapters.

*Key words:* Interferometry, image formation, coherence functions, interferometric imaging

*PACS:* 07.60.Ly, 42.25.-p, 42.25.Hz, 42.30.-d, 42.30.Kq, 95.55.-n, 95.55.Br, 95.75.Kk

---

## 1 Introduction

It is a truth universally acknowledged — at least amongst most conventional astronomers — that the use of interferometric methods at optical and infrared wavelengths requires specialised and somewhat advanced knowledge. While the former may indeed be true, the basic physical principles that underpin interferometric astrophysics are rather straightforward. Furthermore, an understanding of some of these physical bases can be helpful both in designing and understanding interferometric telescope arrays and in planning,

---

*Email address:* [cah@mrao.cam.ac.uk](mailto:cah@mrao.cam.ac.uk) (Chris Haniff).

executing and interpreting astronomical interferometric observations. In this paper, I present a brief introduction to four of these key ideas, chosen so as to benefit new users of interferometric methods in astrophysics. Such a choice will necessarily be restrictive, but my aim has been to draw attention to themes that focus on aspects of interferometry that are likely to be the most novel to newcomers. For simplicity, this introductory chapter mentions only in passing two features of the VLTI that are clearly important for observers but can be ignored in an idealised description of interferometric practice: (i) the presence of a turbulent atmosphere and (ii) the presence of a significant thermal background. These additional elements will be covered elsewhere in this volume.

At the risk of offending those used to more mathematical rigour, the flavour of this presentation has deliberately been adjusted to be largely heuristic. In my experience, confusions and misunderstandings of interferometry usually derive from conceptual mistakes rather than from any lack of mathematical fluency. My hope here is that by focusing on the physical bases of a small number of topics, readers may gain enough familiarity and “feeling” for interferometric methods that they will subsequently be able to take advantage of more detailed and comprehensive treatments elsewhere.

By way of a final remark, readers will note that this paper contains little mention of the themes of astrometric interferometry nor the use of interferometry for extra-solar planet detection. The exclusion of these topics should not be interpreted as any measure of their scientific merit, but merely that time and space constraints did not permit any sensible treatment to be presented here.

## 2 Classical imaging theory

Fluency with the concept of the Fourier decomposition of an image is one of the most important aspects of spatial interferometry. This is less familiar than the usual description of an image as the convolution of the true source brightness distribution with a point-spread function (PSF), but the two are very closely related. We can see this as follows.

Under most conditions<sup>1</sup> the fundamental relationship between the brightness distribution of a source,  $O(\alpha, \beta)$ , and the image delivered,  $I(\alpha, \beta)$ , can be written as (e.g. Goodman, 1996):

$$I(\alpha, \beta) = \iint P(\alpha - \alpha', \beta - \beta') O(\alpha', \beta') d\alpha' d\beta', \quad (1)$$

---

<sup>1</sup> More specifically, when the source is spatially incoherent, i.e. there is no fixed phase relationship between the electric fields emitted from different parts of it, and when the performance of the imaging system does not vary with field angle.

where  $\alpha$  and  $\beta$  are co-ordinates on the sky, and  $P(\alpha, \beta)$  is the assumed space invariant PSF. For the case of an unaberrated telescope with a circular pupil, the PSF will be an Airy pattern. If we take the Fourier transform of each side of this equation we obtain:

$$\tilde{I}(u, v) = \tilde{T}(u, v) \times \tilde{O}(u, v), \quad (2)$$

where  $\tilde{I}(u, v)$  and  $\tilde{O}(u, v)$  refer to the Fourier transforms of the real space brightness distributions,  $u$  and  $v$  are the reciprocal co-ordinates to  $\alpha$  and  $\beta$ , and the properties of the imaging system are now encapsulated in a multiplicative transfer function,  $\tilde{T}(u, v)$ . The co-ordinates  $u$  and  $v$  have dimensions of inverse radians, and are usually referred to as spatial frequencies.

In general, the transfer function, which is simply the Fourier transform of the PSF, is obtained from the auto-correlation of the complex pupil function,  $A$ :

$$\tilde{T}(u, v) = \iint A^*(x, y) A(x + u, y + v) dx dy. \quad (3)$$

Here  $x$  and  $y$  represent co-ordinates in the pupil, and the function  $A(x, y)$  has an amplitude between unity and zero (depending on the strength of the electric field it transmits at position  $(x, y)$ ) and an argument that is related to the relative phases of the wavefront transmitted by different parts of the pupil. Importantly, for each spatial frequency,  $u$ , there is a corresponding physical baseline in the pupil of length  $\lambda u$ .

An example of a typical transfer function and its associated PSF are shown in Fig. 1. In this case, for a circularly symmetric aperture, the transfer function has been written as a function of a single co-ordinate,  $T(f)$ , with  $f^2 = u^2 + v^2$ .

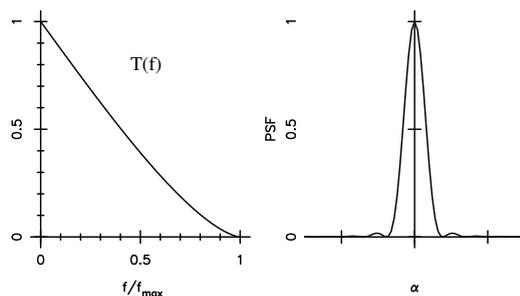


Fig. 1. Schematic plots of the transfer function (left) and point-spread function (right) of a diffraction-limited circular aperture. Note how the transfer function falls smoothly to zero at a spatial frequency,  $f_{max} = D/\lambda$ , determined by the aperture diameter,  $D$ , and the wavelength of observation. The form of the PSF, e.g. its characteristic width, sidelobe levels etc, are uniquely determined by the transfer function. For a circular pupil the PSF has a full width at half maximum of  $\sim 1.0\lambda/D$ .

## 2.1 Image formation - key ideas

As far as helping our understanding of interferometry, there are three key lessons to be learnt here:

- The formal equivalence of the decomposition of an image into a series of spatially separated PSFs or a series of non-localised sinusoids (i.e. Eq. 1 and Eq. 2 are a Fourier transform pair). Note that in the former of these cases “good” point spread functions will necessarily be compact, whereas in the latter the basis functions (sines and cosines) are completely non-localised.
- The description of an image in terms of its Fourier components, and the action of an incoherent imaging system as a filter of the spatial Fourier spectrum of a source. In this sense, all images — even diffraction limited ones — are merely filtered, and hence imperfect, representations of what is in the sky.
- The association of each Fourier component (or spatial frequency) measured by an imaging system with a real physical baseline (or baselines) in the aperture that samples the radiation from the source, and an understanding that the form of the PSF arises from the relative sampling — and hence weighting given to — the different spatial frequencies measured.

As we will see later, interferometry as applied to astrophysics can be understood as the process by which the Fourier components that describe the source brightness distribution are measured individually. From a user’s perspective one may ask: “Why follow this tedious procedure at all?” The answer is quite simple: as well as permitting investigations on angular scales much smaller than could be realised with any monolithic optical/infrared telescope, the use of an interferometer allows the astronomer to selectively probe any particular Fourier component of interest by choosing an appropriate interferometer baseline. How exactly the amplitudes and phases of these complex quantities might be measured will be the subject of the next two sections.

## 3 Coherence functions

The fundamental basis for the use of interferometry in astronomy arises from two physical laws, the Weiner-Khinchin theorem and the van Cittert-Zernike theorem (e.g. Born & Wolf, 1999). While the derivations of these two theorems are interesting in themselves, we need only familiarise ourselves of their content for the purpose of this review.

Consider a simple two-element interferometer interrogating the light from a distant source, as shown in Fig. 2. In the most general case, the radiation

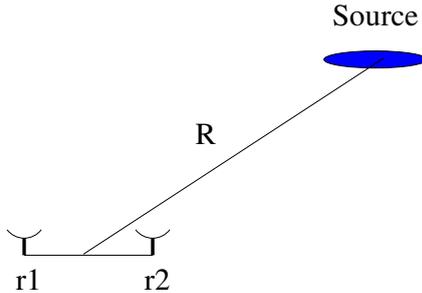


Fig. 2. Schematic cartoon of a two-element interferometer sampling the radiation from a distant source. The two sample points are located at  $r_1$  and  $r_2$ , and the distance to the source,  $R$  is assumed large enough that the source is in the far-field.

is assumed to be sampled at two different locations,  $\vec{r}_1$  and  $\vec{r}_2$ , and at two different times,  $t_1$  and  $t_2$ , and the quantity of interest is the correlation of the measurements of the two electric fields, i.e.  $\langle E^*(\vec{r}_1, t_1) \times E(\vec{r}_2, t_2) \rangle$ . Here, angle brackets refer to a time average over a period that is long compared to the oscillation time of the electric field. Note that the correlation function will in general be complex, and that each of the detected  $E$  fields is itself a summation of elemental contributions from each part of the source, which we assume to be temporally and spatially incoherent with respect to each other.

We will be interested in two simplified cases of this experiment; the first when  $\vec{r}_1 = \vec{r}_2$  but  $t_1 \neq t_2$ , and second, where the measurement points are spatially separated, but the measurements occur at the same time. In both of these cases it can be shown that the correlation (or coherence) function will be a function only of the difference in the time or space co-ordinates.

### 3.1 The temporal coherence function

In the case of the measurement of the electric field at the same location, but at different times, we can define the so-called temporal coherence function as:

$$\langle E^*(\vec{r}_1, t_1) \times E(\vec{r}_2, t_2) \rangle = V(\vec{r} - \vec{r}, t_1 - t_2) = V(\vec{0}, \tau). \quad (4)$$

This function measures the extent to which the electric fields along a given wave train are correlated in time and is basically the quantity that a laboratory based Michelson interferometer measures. Its importance for astronomers arises from the Weiner-Khinchin theorem, which states that the normalised value of the temporal coherence function is equal to the Fourier transform of the normalised spectral energy distribution of the source, i.e.:

$$\frac{V_t(\tau)}{V_t(0)} = \frac{\int B(\omega) \exp(-i\omega\tau) d\omega}{\int B(\omega) d\omega}, \quad (5)$$

where we have use the shorthand notation  $V_t$  to denote that the spatial argument of the coherence function is zero. Measurement, and subsequent correlation, of the electric fields at a point, thus allows the recovery of the source *spectrum* via an inverse Fourier transform if measurements of  $V_t(\tau)$  are made for a suitable sample of delays,  $\tau_i$ .

Since the temporal coherence function and source spectrum are a Fourier pair, a broad spectral energy distribution will lead to a coherence function that decays rapidly, while a monochromatic source will have a coherence function that is equal to unity for all values of  $\tau$ . It is useful to define a ‘‘coherence time’’ for the radiation such that  $t_{coh} \sim 2\pi/\Delta\omega$ , where  $\Delta\omega$  is its spectral bandwidth. This timescale basically measures the maximum time delay allowed between the measurements of the radiation field such that the coherence function will still have a non-zero value.

### 3.2 The spatial coherence function

In the case associated with measuring the electric field from a source at two locations but at the same time we can define the so-called spatial coherence function as:

$$\langle E^*(\vec{r}_1, t_1) \times E(\vec{r}_2, t_2) \rangle = V(\vec{r}_1 - \vec{r}_2, t - t) = V(\vec{\rho}, 0). \quad (6)$$

This function measures the extent to which the electric fields perpendicular to a wave train are correlated in space and is basically the quantity that a Young’s double slit experiment measures on axis. Its importance for astronomers arises from the van Cittert-Zernike theorem, which states that for sources in the far field the normalised value of the spatial coherence function is equal to the Fourier transform of the normalised sky brightness distribution, i.e.:

$$\frac{V_r(\vec{\rho})}{V_r(0)} = \frac{\int I(\vec{\alpha}) \exp\left(-i2\pi\frac{(\vec{\alpha}\cdot\vec{\rho})}{\lambda}\right) d\alpha}{\int I(\vec{\alpha}) d\alpha}, \quad (7)$$

or in the notation of Equations (1) and (2):

$$V_{r,norm}(u, v) = \frac{\iint I(\alpha, \beta) \exp(-i2\pi(u\alpha + v\beta)) d\alpha d\beta}{\iint I(\alpha, \beta) d\alpha d\beta}. \quad (8)$$

As before,  $\alpha$  and  $\beta$  are co-ordinates in the sky, whereas  $u$  and  $v$  are the components of the vector baseline between the two sampling points projected onto a plane perpendicular to the source direction and measured in wavelengths.

In this second example, therefore, measurement, and subsequent correlation, of the electric fields at two different points, allows the recovery of the source *structure* via an inverse Fourier transform if measurements of  $V_r(u, v)$  are made for a suitable set of vector baselines,  $\{u_i, v_i\}$ .

### 3.3 Coherence functions - key ideas

Perhaps the most unusual aspects of interferometry are those summarised in the previous two subsections, i.e. that it is straightforward, in principle, to recover details of both the spectrum and structure of an astronomical source without any imaging or dispersing optics but simply by investigating correlations in the electric field the source delivers. This is a rather beautiful result, especially as the Fourier relationship between the source structure and spectrum and the two coherence functions is so simple. Were this relationship not linear and invertible, then it is possible that interferometric methods might not featured in astrophysics at all to date.

For the most part, we shall be interested in the spatial coherence or “visibility” function, as we will refer to it henceforth, since it is this that contains information about the structure of the source in the sky and which the VLTI has been designed to measure and exploit.

Summarising the previous two sections of this treatment, we can see then that, in a very simple way, the practice of astronomical interferometry involves nothing more than four key elements:

- The idea of representing a brightness distribution in the sky as a superposition of sine and cosine functions, i.e. a Fourier decomposition.
- The use of measurements of the spatial coherence function as a direct proxy for the strengths of these Fourier components, i.e. exploitation of the van Cittert-Zernike theorem.
- An understanding that different configurations of sample points with projected separations  $B_i$  will give rise to measurements of the value of the Fourier transform of the source brightness distribution at spatial frequencies  $u_i = B_i/\lambda$ .
- The idea of a final step — to be elaborated on in Section 5 — where the Fourier data are either interpreted or inverted to establish the details of the source morphology.

In the meantime, exactly how we can access the time averaged products of field quantities such as  $\langle E^*(\vec{r}_1) \times E(\vec{r}_2) \rangle$  will be the subject of the next section.

## 4 Interferometric observables

At first sight, measurement of the visibility function appears to require direct access to the amplitude and phase of the electric field. While this may be possible at radio wavelengths, optical and near-infrared detectors are generally square-law devices (i.e. they measure the intensity of the electric field and not its amplitude or phase) and so there is often confusion as to how it is actually possible to measure the visibility function at all. As we shall see below, both the amplitude and phase of the visibility function are encoded in an intensity pattern that can trivially be characterised by any square-law detector.

### 4.1 An aside on measuring coherence functions

We can best understand how optical coherence functions are measured by considering a simple Young's double slit experiment. In this case, at any given point on the image plane, for example at the on-axis point, the observed intensity will be given by the modulus squared of the summation of the electric field arriving from the two slits. If we call these  $E_1$  and  $E_2$ , we can write the detected intensity as:

$$\begin{aligned}
 I &= \langle (E_1 + E_2)^* \times (E_1 + E_2) \rangle \\
 &= \langle |E_1|^2 \rangle + \langle |E_2|^2 \rangle + \langle E_1 E_2^* \rangle + \langle E_1^* E_2 \rangle \\
 &= \langle |E_1|^2 \rangle + \langle |E_2|^2 \rangle + \langle 2 |E_1| |E_2| \cos(\phi) \rangle,
 \end{aligned} \tag{9}$$

where  $\phi$  is the phase difference between the electric field components  $E_1$  and  $E_2$  and the angle brackets refer to the usual time average. The first two terms of this equation obviously refer to the mean intensity seen in the double-slit fringe pattern, while the third term, which is associated with the modulation of the fringes from light to dark, clearly encodes the values of the complex products  $\langle E_1 E_2^* \rangle$  and  $\langle E_1^* E_2 \rangle$ , i.e. the visibility function and its complex conjugate. Thus the observed modulation of the intensity in the detector plane directly measures the complex visibility (coherence) function.

There are two key features of the detected fringe pattern that are of interest (see the left hand panel of Fig. 3). The first is the *fringe amplitude*. This is a measure of the fringe contrast, sometimes known as the ‘‘Michelson visibility’’, and is related to the measured maximum and minimum intensities in the fringe pattern by:

$$V_{\text{Michelson}} = \frac{I_{\text{max}} - I_{\text{min}}}{I_{\text{max}} + I_{\text{min}}}. \tag{10}$$

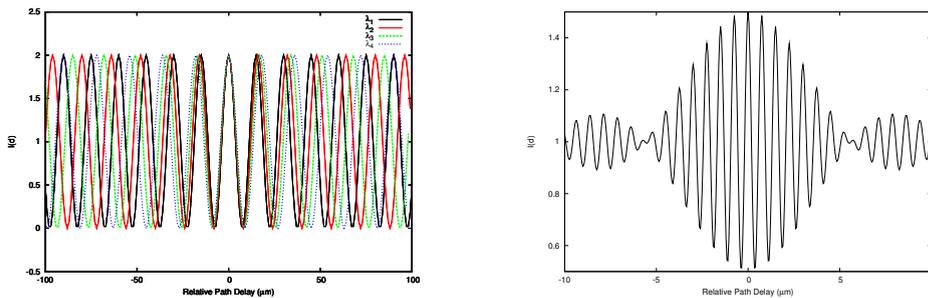


Fig. 3. Figures showing the individual monochromatic responses (left) and the resulting fringe pattern for a polychromatic source (right), for a simple two element interferometer. In both cases the intensity output is plotted as a function of optical path difference between the interfering beams. For the polychromatic fringe pattern a spectral bandpass of  $0.1 \mu\text{m}$  centred on  $0.75 \mu\text{m}$  has been assumed together with a visibility amplitude of 0.5. (Figure courtesy of D. Pearson.)

For example, for all the monochromatic fringes shown in the left hand panel of Fig. 3, the visibility amplitude is equal to unity.

The second key observable is the *fringe phase*. This is the location of the central fringe with respect to the location of the zero optical path difference (OPD) position. Typically the fringe phase is known modulo- $2\pi$ , with  $2\pi$  radians corresponding to an offset of a whole fringe period. In Fig. 3 both the left and right hand panels depict fringe patterns with phases of zero.

The crucial point to note is that the fringe visibility amplitude and phase defined in this way directly measure the amplitude and phase of the complex visibility (or coherence) function. There is thus no fundamental difficulty in measuring the coherence function with a square-law detector, since its properties are encoded very straightforwardly in an observable intensity pattern.

#### 4.2 The output of a 2-element interferometer

In order to develop the mathematical formalism needed in the next subsection, it is helpful to focus again on a simple 2-element interferometer and explore how its output varies as a function of a number of key instrumental parameters. In Fig. 4 we can identify five primary functional components of the hardware associated with such an interferometer:

- A pair of collectors, located at positions  $x_1$  and  $x_2$ , whose role is to sample the radiation from the source. The direction to the source is characterised by the “pointing direction”,  $\vec{s}$ . It is usual to refer to the vector between the collectors as the “baseline vector”,  $\vec{B}$ .
- A “beam relay system” whose role is to transport the radiation sampled

to a central laboratory where the signals can be mixed and the resulting outputs detected. We will refer to the total optical paths from each collector to the location at which the signals are finally mixed as  $d_1$  and  $d_2$ .

- An apparatus to compensate for the “geometric delay”, i.e. the additional optical path experienced by the light travelling to the collector more distant from the astronomical source. The magnitude of this additional path is given by  $\vec{s} \cdot \vec{B}$ , where  $\vec{s}$  is a unit vector in the pointing direction. This additional optical path is usually introduced using a movable carriage carrying retro-reflecting optics, the whole mechanism being called a delay line. In Fig. 4 two delay lines are shown, allowing for variable extra optical paths to be introduced into the beams from each collector independently.
- A device to combine the electric fields sampled by the collectors. This is usually referred to as the “beam combiner”.
- A detector to sample the output of the beam combiner. This is usually some form of low-noise fast-readout detector.

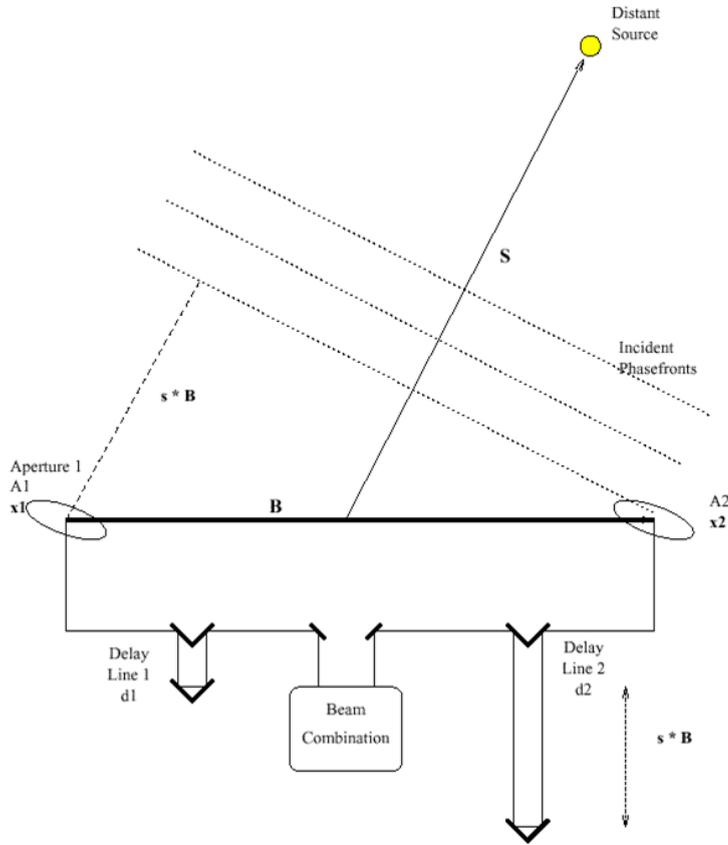


Fig. 4. A schematic cartoon showing the key functional elements of a simple 2-element interferometer. Light from the two collectors travels along the optical paths  $d_1$  and  $d_2$  and is interfered and detected at the “beam combiner” in the centre of the figure. Further details of the function of the hardware elements depicted here can be found in the main text.

Using the nomenclature introduced above, we can describe the monochromatic electric fields arriving at the beam combiner as:

$$\Psi_1 = A_1 \exp\left(ik\left[\vec{s} \cdot \vec{B} + d_1\right]\right) \exp(i\omega t) \quad (11)$$

and

$$\Psi_2 = A_2 \exp(ik[d_2]) \exp(i\omega t), \quad (12)$$

so that, assuming equally sensitive collectors (i.e.  $A_1 = A_2$ ), the resulting intensity can be written as:

$$\begin{aligned} I_{detected} &= \langle |\Psi_1 + \Psi_2|^2 \rangle \\ &\propto 2 + 2 \cos\left(k\left[\vec{s} \cdot \vec{B} + d_1 - d_2\right]\right) \\ &\propto 2 + 2 \cos(kD), \end{aligned} \quad (13)$$

where  $D = \left[\vec{s} \cdot \vec{B} + d_1 - d_2\right]$ .

This co-sinusoidal variation of the intensity is the quintessential feature of the interferometric output, and has maxima separated by  $\Delta d_{1 \text{ or } 2} = \lambda$ ,  $\Delta(\vec{s} \cdot \vec{B}) = \lambda$ , or  $\Delta k = 2\pi/D$ .

As a consequence, if we wish to visualise the interferometric fringes on a detector, at least three possible methods exist: (i) we can deliberately alter the optical path in one or both of the array arms — this is usually referred to as modulating the OPD (ii) we can wait for the source to move in the sky — in this case changes in the scalar product of the pointing vector and the baseline vector will produce an output that oscillates at a characteristic frequency known as the "fringe rate" or (iii) in the polychromatic case, for non-zero values of  $D$ , we can disperse the output and examine the fringes as a function of wavenumber — this is usually referred to as a "channelled spectrum" response.

We can extend our monochromatic treatment to the polychromatic case by integrating Eq. 13 over the spectral bandpass detected to give, for example, for a uniform bandpass centred on  $\lambda_0$  of  $\pm\Delta\lambda/2$ :

$$\begin{aligned} I_{detected} &\propto \int_{\lambda_0 - \Delta\lambda/2}^{\lambda_0 + \Delta\lambda/2} 2[1 + \cos(2\pi D/\lambda)] d\lambda \\ &\propto \Delta\lambda \left[ 1 + \frac{\sin(\pi D \Delta\lambda / \lambda_0^2)}{(\pi D \Delta\lambda / \lambda_0^2)} \cos(k_0 D) \right] \end{aligned}$$

$$\propto \Delta\lambda \left[ 1 + \frac{\sin(\pi D/\Lambda_{coh})}{(\pi D/\Lambda_{coh})} \cos(k_0 D) \right], \quad (14)$$

where we have introduced a length scale,  $\Lambda_{coh} = \lambda_0^2/\Delta\lambda$ , called the coherence length, and  $k_0 = 2\pi/\lambda_0$ . The essential difference between the monochromatic and polychromatic responses is thus a modulation of the fringes by the so-called “coherence envelope”<sup>2</sup>.

The left and right hand panels of Fig. 3 show plots of the monochromatic (left) and polychromatic (right) output of a 2-element interferometer as a function of  $D$ . In the monochromatic case, the fringe maxima for different colours all overlap when the OPD between the interferometer arms is zero, but each monochromatic response has a period that depends on the wavelength of observation. It is the “washing out” of the different fringe maxima and minima at non-zero OPD that leads to a polychromatic response with modulated fringe contrast, and it is this that gives rise to the real need for delay lines in any interferometric implementation. Importantly, unless the optical paths are matched to much better than  $\Lambda_{coh}$ , the measured fringe visibility amplitude will no longer be a faithful measurement of the source coherence function, but will instead measure the product of the true visibility amplitude with a scale factor associated with the value of the coherence envelope at the OPD obtaining when the data were collected.

#### 4.3 *Extended sources: a heuristic and a mathematical interlude*

Before presenting a somewhat more formal mathematical development of the response of an interferometer to an extended source — all we have mentioned previously has assumed an unresolved target — it is helpful to review what happens for a very simple extended source, e.g. a binary star consisting of two unequal unresolved components.

The basic argument can most easily followed with reference to Fig. 5. Here, each component of the binary produces its own fringe pattern, with unit visibility amplitude and a visibility phase that will be associated with the location of that element in the sky (see the two leftmost columns of the figure). For any spatially incoherent source, which is likely to be true for all optical/IR astronomical sources, the actual interferometer output for a extended source will be the intensity superposition of these individual “elemental” fringes: this is shown in the rightmost column of the panel. As is evident from Fig. 5, the resultant interferometric fringes will thus neither have unit contrast, nor a phase equal to the fringe phase of either of the elemental fringe patterns

---

<sup>2</sup> Radio interferometrists usually refer to this as the “delay beam”

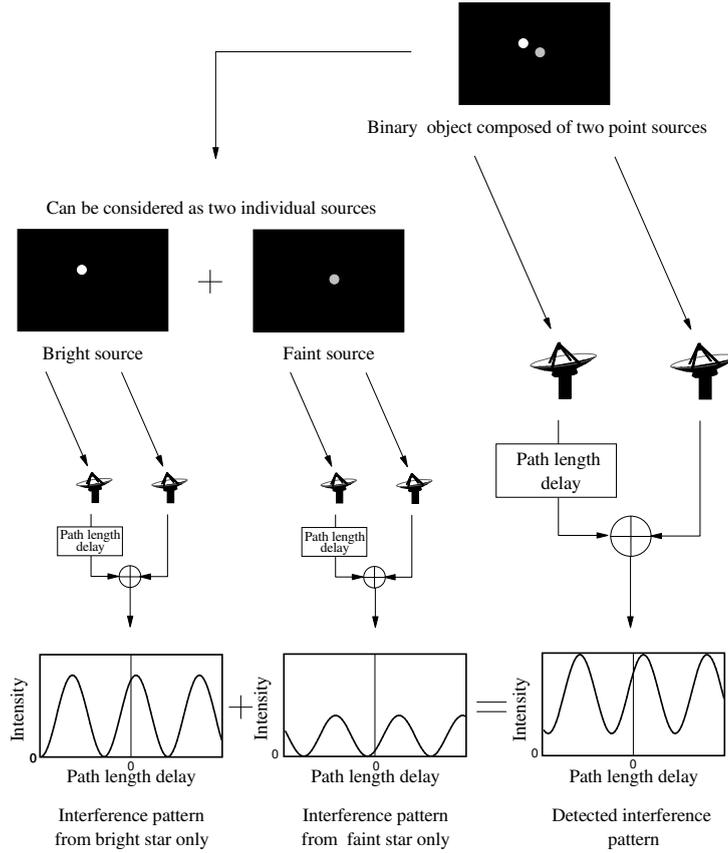


Fig. 5. A heuristic explanation of how the output of an interferometer encodes the source brightness distribution. The two left hand columns show the interferometer output produced by the two separate components of an unequal binary, while the rightmost column depicts the actual detected output (i.e. the superposition of the two fringe patterns). The resulting fringe contrast and location are clearly a function of the overall source brightness distribution (See text for details. Figure courtesy of T.R. Scott.).

from which they are formed<sup>3</sup>. Hence, in some yet to be quantified sense, the form of the overall brightness distribution must be encoded in the resulting non-unit-contrast interferometer output.

A more formal description of this process can be presented as follows. Consider the observation of an extended monochromatic (the extension to a polychromatic source is straightforward) source, whose brightness distribution on the sky is written as  $I(\vec{s}_0 + \vec{\Delta}s)$ , where  $\vec{s}_0$  is a vector in the pointing direction and  $\vec{\Delta}s$  is a vector perpendicular to this in the plane of the sky (see the left hand panel of Fig. 6). If we integrate the monochromatic response (Eq. 13) over the

<sup>3</sup> The careful reader will note that since the sum of two sine waves of the same frequency is another sine wave of the same frequency (but with a different amplitude and phase) the resulting fringe pattern has the same period as its component fringes.



Fig. 6. Figures showing the coordinate systems used in the text of subsection 4.3. The left hand panel refers to the co-ordinates describing the locations of each element of the source, with the origin at the centre of the interferometer baseline. The right hand panel shows the  $\{u, v, w\}$  coordinate frame in which the baseline vectors are measured and the  $\{\alpha, \beta, \gamma\}$  frame which is conventionally used to describe the brightness distribution on the celestial sphere. The  $w$  axis points in the direction of  $\vec{s}_0$  from the centre of the interferometer baseline.

extent of the source we find:

$$\begin{aligned}
I_{detected}(\vec{s}_0, \vec{B}) &\propto \int I(\vec{s}) [1 + \cos(kD)] d\Omega \\
&\propto \int I(\vec{s}) [1 + \cos(k \{ \vec{s} \cdot \vec{B} + d_1 - d_2 \})] d\Omega \\
&\propto \int I(\vec{s}) [1 + \cos(k \{ [\vec{s}_0 + \vec{\Delta}s] \cdot \vec{B} + d_1 - d_2 \})] d\Omega \\
&\propto \int I(\vec{s}) [1 + \cos(k \{ \vec{s}_0 \cdot \vec{B} + \vec{\Delta}s \cdot \vec{B} + d_1 - d_2 \})] d\Omega \\
&\propto \int I(\vec{\Delta}s) [1 + \cos(k \{ \vec{\Delta}s \cdot \vec{B} \})] d\Omega'. \tag{15}
\end{aligned}$$

Note that in the final step, we have assumed that  $d_1 - d_2$  is chosen to exactly cancel the geometric delay term,  $\vec{s}_0 \cdot \vec{B}$ , and we have changed the dummy variable of integration to signify summation over all possible values of  $\vec{\Delta}s$ .

Consider now, adding a small path delay,  $\delta$ , to one arm of the interferometer. In this case the response becomes:

$$\begin{aligned}
I_{detected}(\vec{s}_0, \vec{B}, \delta) &\propto \int I(\vec{\Delta}s) [1 + \cos(k \{ \vec{\Delta}s \cdot \vec{B} + \delta \})] d\Omega' \\
&\propto \int I(\vec{\Delta}s) d\Omega' + \cos(k\delta) \cdot \int I(\vec{\Delta}s) \cos(k \{ \vec{\Delta}s \cdot \vec{B} \}) d\Omega' \\
&\quad - \sin(k\delta) \cdot \int I(\vec{\Delta}s) \sin(k \{ \vec{\Delta}s \cdot \vec{B} \}) d\Omega'. \tag{16}
\end{aligned}$$

This expression can be rewritten in a shorter form as:

$$\begin{aligned}
I_{detected}(\vec{s}_0, \vec{B}, \delta) &\propto \int I(\vec{\Delta}s) d\Omega' + \cos(k\delta) \Re[Q] + \sin(k\delta) \Im[Q] \\
&\propto I_{total} + \Re[Q \exp(-ik\delta)], \tag{17}
\end{aligned}$$

where we have use  $Q$  to represent the quantity  $\int I(\vec{\Delta}s) \exp(-ik\vec{\Delta}s \cdot \vec{B}) d\Omega'$ , and  $\Re$  and  $\Im$  denote taking the real and imaginary part respectively.

If we look at Eq. 17, we can see that if we record the interferometer output for, say, two values of  $\delta$  (e.g. 0 and  $\lambda/4$ ), then this will allow recovery of the complex value of  $Q$  to within a multiplicative constant. The importance of this result can be understood by reminding ourselves what  $Q$  itself measures. If we look at the right hand panel of Fig 6 we can use the  $\{u, v, w\}$  and  $\{\alpha, \beta, \gamma\}$  coordinate systems to write  $\vec{s}_0$  as  $(0, 0, 1)$  and  $\vec{\Delta}s$  as  $(\alpha, \beta, 0)$  so that the quantity  $Q$  can be rewritten as:

$$\begin{aligned} Q &= \int I(\vec{\Delta}s) \exp(-ik\vec{\Delta}s \cdot \vec{B}) d\Omega' \\ &= \int I(\alpha, \beta) \exp(-ik[\alpha B_x + \beta B_y]) d\alpha d\beta \\ &= \int I(\alpha, \beta) \exp(-i2\pi[\alpha u + \beta v]) d\alpha d\beta, \end{aligned} \tag{18}$$

where  $u (= B_x/\lambda)$  and  $v (= B_y/\lambda)$  are the projections of the baseline on a plane perpendicular to the pointing direction. The astute reader will recognise the function  $Q$  as the same coherence (or visibility) function we introduced in Sec. 3.2, and the co-ordinates  $u$  and  $v$  as the same ‘‘spatial frequencies’’ we introduced much earlier when describing the Fourier decomposition of a source brightness distribution.

We have thus finally arrived at our key results:

- The interferometric output encodes the real and imaginary parts of the quantity  $Q$ .
- This quantity  $Q$  is nothing more than a sample from the 2-dimensional Fourier transform of the source brightness distribution.
- The particular sample of the Fourier transform selected is determined by the instantaneous projected baseline.

We can thus view interferometers simply as rather sophisticated ‘‘engines’’ that allow us to measure the Fourier content of the sky brightness.

#### 4.4 Interferometric measurements - key ideas

At this point it is probably worth reviewing what we have learnt: there are only two key ideas, but it will not harm us to reiterate them again here!

First, is the idea of the interferometric output encoding the complex Fourier transform of the source brightness distribution. This is as true at optical wave-

lengths as it is in the radio, and so identifies optical/IR interferometry as a true analogue of what has been commonplace in the radio for many years.

Second, we have the idea that extracting the complex visibility function *necessarily* involves examining the co-sinusoidal variation of the interferometric output in some way. In the example described above, we considered the introduction of two deliberate OPD settings, but in general many other implementations can be envisaged. The interested reader can find further details on these practical issues in other chapters in this volume.

## 5 Interferometric imaging

We have seen in the previous sections that the fundamental processes involved in interferometric astrophysics are relatively straightforward. Interferometers measure the Fourier transform of the source brightness distribution, and so, in principle, only three steps are involved in mapping the sky this way:

**Step 1** Measurement of the visibility function,  $V(u, v)$ , with as many different interferometer baselines as possible, i.e. at as many spatial frequencies  $u$  and  $v$  as possible.

**Step 2** Calibration of the measurements to remove any instrumental modification to the measured complex amplitudes and phases.

**Step 3** Fourier inversion of the calibrated data to recover the source brightness distribution,  $I(\alpha, \beta)$ .

In practice, of course, the details involved in these processes demand much attention, and a proper treatment of all of them would require a lengthy volume to do them justice. Instead, in the following subsections we will briefly examine three of these issues to get a feel as to what type of thinking might usefully be involved in preparing an interferometric observation.

### 5.1 Visibility functions

The behaviour of the source visibility function with interferometer baseline and in particular how its amplitude changes, is of major interest when planning interferometric observations. This is because the amplitude of  $V(u, v)$  tells us the apparent contrast of the interference fringes, and our ability to measure the parameters of the fringes is a very strong function of the fringe contrast. For example, in the faint source photon-limited case, a reduction in the fringe contrast by a factor of 10 needs an increase in source brightness of a factor of 100 to maintain the same signal-to-noise on the fringe parameter

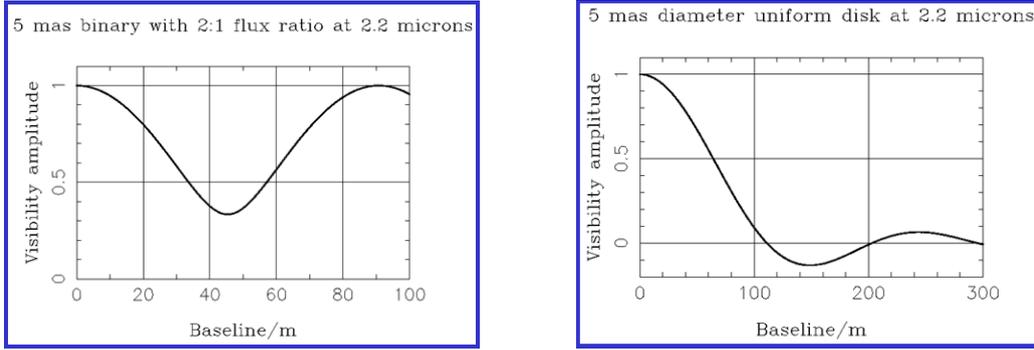


Fig. 7. Plots showing the amplitude of the visibility function for two 0.005 arcsecond sized sources as a function of interferometer baseline. The left hand panel is for a binary star comprising two unresolved components with a 2 : 1 intensity ratio while the right hand panel is for a uniform disk. In both cases an observing wavelength of  $2\ \mu\text{m}$  has been assumed, and the interferometer baseline has been oriented parallel to the direction in which the source is extended. Negative values for the visibility amplitude correspond to a phase of the visibility function of  $-\pi$  radians.

measurements. It is no wonder that most sensitivity calculations for interferometers assume a source producing unit contrast fringes!

Examples of the behaviour of two “typical” visibility functions are shown in Fig. 7. The left hand panel shows the visibility amplitude for a 5-milliarcsecond binary as a function of the interferometer baseline. The modulation of the visibility amplitude is the characteristic signature of a binary target. The period of the modulation encodes the binary separation — larger separations give a more rapid modulation — and the modulation depth the intensity ratio of the components. Although not shown here, the visibility phase also displays oscillations and these are similarly informative.

The right hand panel shows the equivalent data for a 5-milliarcsecond diameter uniform disk source. This again shows a modulated visibility function, but demonstrates a new feature that is typical of targets that are fully resolved, i.e. an overall reduction in visibility amplitude with increasing baseline. Most importantly, information on structures smaller than the disk size ( $\theta_{disk}$ ) will correspond to measurements of the visibility amplitude on baselines longer than  $\lambda/\theta_{disk}$ , i.e. where the visibility amplitude will be  $\ll 1$ .

In summary, then, the two panels of Fig. 7 draw attention to a number of important “home truths”.

- First, that distinguishing between different source morphologies requires measurements of the visibility function on many different baselines, with changes in visibility amplitude and phase encoding the source structure.
- Second, that measurements on baselines at least as long as  $\lambda/\theta$  are required for unambiguous study of a target that has an angular size of  $\sim \theta$ .

- Finally, that for resolved targets, the expected fringe contrast is likely to be much much lower than the unit visibility amplitude that point-like sources will produce, and hence difficult to measure with good signal-to-noise.

This last point is probably the most difficult one to address from an experimental point of view, and represents a significant challenge for the designers of modern facility arrays which must be able to operate efficiently when observing resolved, and hence, scientifically exciting targets.

## 5.2 The practice of imaging and some useful rules of thumb

The examples in the previous subsection show that simple features (e.g. separations and flux ratios) of a source can be inferred directly from measurements of  $V(u, v)$ . However, unambiguous recovery of a faithful high resolution image of the target is clearly an attractive goal to consider.

We can begin by reminding ourselves of the fundamental relationship between the visibility function and the normalised sky brightness distribution:

$$I_{\text{norm}}(\alpha, \beta) = \iint V(u, v) \exp(+i2\pi \{ \alpha u + \beta v \}) du dv . \quad (19)$$

In practice, however, we will only have a sampled version of  $V(u, v)$ , and so image recovered by a Fourier inversion will be the so-called “dirty map”:

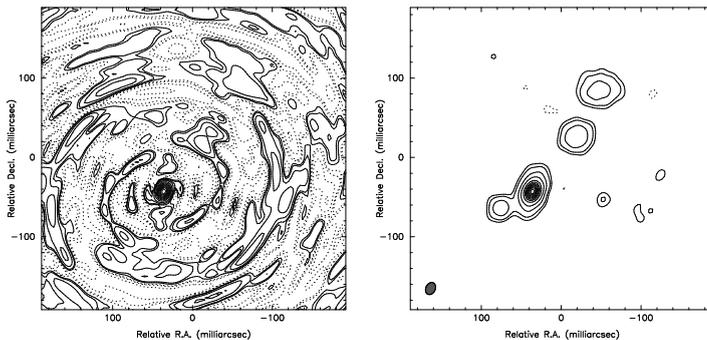


Fig. 8. Simulated dirty (left) and deconvolved (right) images for a target comprising a number of resolved components. Both panels have contours plotted at -10, -5, -2, -1, 1, 2, 5, 10, 20, 30, 40, 50, 60, 70, 80, and 90% of the peak flux. Negative contours are shown dashed. While the source is barely visible in the dirty image, it is straightforward to deconvolve the map and correct for the interferometer PSF. The noise in the right hand panel reflects the number and quality of the data used in the simulation. The small grey ellipse in the bottom left hand corner of the panel shows the size of the core of the PSF, i.e. the resolution in the restored map.

$$\begin{aligned}
I_{\text{dirty}}(\alpha, \beta) &= \iint S(u, v) \cdot V(u, v) \exp(+i2\pi \{\alpha u + \beta v\}) du dv & (20) \\
&= B_{\text{dirty}}(\alpha, \beta) * I_{\text{norm}}(\alpha, \beta) , & (21)
\end{aligned}$$

where  $S(u, v)$  is the sampling function describing which measurements of  $V(u, v)$  have been secured, and  $B_{\text{dirty}}(\alpha, \beta)$  is the Fourier transform of the sampling distribution. This is known as the “dirty beam” and is simply the PSF of the interferometer. It is usually far less attractive than an Airy pattern, e.g. exhibiting strong and numerous sidelobes, but it is completely determined by the known sampling of the Fourier plane.

Thus, despite the unusual form of the interferometric PSF, its behaviour will generally be well understood and it can be accounted for very straightforwardly. This process of correcting an interferometric image for the Fourier plane sampling is known as deconvolution and can be performed using many schemes such as CLEAN, MEM, and WIPE (see, e.g., Cornwell et al. (1999); Lannes et al. (1994)). Fig. 8 shows an example of this type of deconvolution in practice with “before” and “after” images for a simulated observation of a resolved source. It is interesting to note how difficult the source is to discern in the dirty map, yet how successful the deconvolution is.

Fortunately, most of the rules-of-thumb developed for radio interferometric imaging can be translated almost unchanged to the optical/IR domain. We mention here two sets of guidelines that may be helpful to newcomers.

From the point of view of imaging in general, the following four rules-of-thumb are often useful:

- If one wishes to recover an image with some number of filled pixels, i.e. pixels containing source flux, then the total number of visibility amplitudes and phases secured should be at least as great as this number. This places a very high premium on collecting the Fourier data efficiently.
- The distribution of samples of  $V(u, v)$  should be as uniform as possible so as to aid deconvolution and deliver a representation of the sky that is not skewed towards any particular part of the spatial frequency spectrum.
- The range of angular scales expected in an interferometric image will be limited by the ratio of maximum to minimum baseline length. This can often be rather small for interferometers with small numbers of collectors, even using Earth rotation to alter the projected baseline lengths.
- For a source of maximum extent  $\theta_{max}$ , there will be no need to sample  $V(u, v)$  any more finely than  $\Delta u \sim 1/\theta_{max}$ .

Another area in which some general comments may be helpful concerns the field of view that one can hope to map with an interferometric array. These can be summarised as follows:

- At the largest angular scales, all interferometers are limited by the field of view of the individual collectors. This is usually referred to as the “primary beam” by radio interferometrists, where it has the more specific meaning of the beam pattern of an individual collector on the sky.
- In practice, vignetting along the optical train from the collectors to the beam combiner is likely to restrict the field of view much more tightly, usually to no more than a few seconds of arc for most optical/IR arrays.
- In the context of Fourier imaging, a further limitation will come from the shortest interferometer baseline,  $B_{min}$ . An array will not generally be sensitive to structures much larger than  $\sim \lambda/B_{min}$  and so these will simply not appear in any image reconstructed from the visibility data.
- Finally, for arrays that utilise non-homothetic beam combiners<sup>4</sup>, the field of view will be limited by the ability of the delay lines to adequately correct for the geometric delays associated with the directions of different parts of the source. In quantitative terms, this implies that the field of view will be no greater than approximately  $[\lambda/B_{max}][\lambda/\Delta\lambda]$ , i.e. the product of the spatial and spectral resolutions.

Finally, we can say something about the dynamic range, i.e. the ratio of the maximum intensity in an image to the minimum believable intensity, and the fidelity of interferometric images. Experience with radio interferometers has shown that dynamic ranges in excess of  $10^5:1$  are possible, but more typical values for snapshot images are of order a few hundred to one. A reasonable estimate for what can be expected can be obtained by multiplying square root of the total number of visibility data by the signal-to-noise per datum. For example, measurements on one hundred different baselines, each with a signal-to-noise of ten, could potentially be assembled to produce a map with a dynamic range of 100:1. Assessing the fidelity of the such a map is quite another matter, but would likely depend on issues such as the completeness of the Fourier plane sampling etc.

It is worth reiterating that these comments should not be taken as “truths” but rather as general guidance for use when preparing an interferometric observation. Furthermore, in many cases the recovery of an image may neither be necessary nor desirable: one can often make significant advances in understanding using much fewer data. No one should forget that the art of interferometric astrophysics, and indeed of many branches of experimental physics, is often tied to designing observational strategies that provide the most scientific leverage for the minimum expenditure of effort!

---

<sup>4</sup> This basically means that the output pupil of the interferometer optics is *not* a scaled version of the input pupil

### 5.3 Sensitivity

The question of the sensitivity of optical/IR interferometry is a very frequent one for newcomers. Give that arrays like the VLTI have to operate in the presence of atmospheric fluctuations, the answer draws heavily on the experience of adaptive optics experiments. If we interpret the question to mean: “What is the faintest source that can be detected in an interferometric map?”, then there will be three interconnected components to the answer:

- First, the “source” — this can either be the science target itself (at any suitable wavelength) or any suitable “reference” target — must be bright enough to allow stabilisation of the interferometric optical paths so that interference fringes can be usefully measured at all.
- Second, the science target must be bright enough such that a useful signal-to-noise can be built up on the visibility amplitudes and phases when averaged over time. The reader should note that the rotation of the Earth (and hence changes in the projected baselines) will mean that the maximum time allowed per Fourier plane measurement will be limited to only a few minutes.
- Finally, the faintest features detectable will be determined by the dynamic range in the recovered image, which will in turn depend on both the total number and quality of the visibility data.

In general, then, the issue of sensitivity is a complex one, and is best determined on a case-by-case basis. Suffice to say, at present, the typical dynamic ranges that have been demonstrated in the near-infrared are of order 100:1, and science targets with magnitudes of order 7 in the  $K$  band have been routinely observed. It is certain that both of these numbers will be revised significantly over the next 5 years, as the latest 2nd-generation facility interferometers reach maturity.

## 6 Summary

Hopefully this lecture has provided some useful pointers to “what matters” to those who are new to how interferometric arrays like the VLTI work. This has necessarily been an incomplete survey, but my hope is that this brief foray will have helped set the scene, and that subsequent contributions will help develop a better feel for the real scientific excitement that a new window on the Universe can bring.

## Acknowledgements

A didactic review such as this necessarily draws upon many contributions colleagues have made to my understanding of interferometry over the years, and it is a pleasure to record my appreciation to them here. In addition, certain components of my presentation have borrowed heavily from others', in particular those of Andy Boden, Tania Scott and Debbie Pearson. I am very grateful to them for allowing me to "borrow" their material. Finally, I should like to record my thanks to D.P. McManus, whose insightful comments have always helped clarify my thinking, and to the school organisers for inviting me to attend.

## References

- Goodman, J.W., *Introduction to Fourier Optics*, McGraw-Hill (New York, 1996).
- Born, M. & Wolf, E., *Principles of Optics*, Cambridge University Press (Cambridge UK, 1999).
- Cornwell, T.J., Braun, R. & Briggs D.S. 1999, ASP Conference series, 180, 151.
- Lannes, A., Anterrieu, E. & Bouyoucef, K. 1994, J. Mod. Opt., 41, 1537.