

## Chapter 2

# Propagation of Light

Propagating through space, light carries with it the information on the shape and the spectrum of the source. A straightforward way to extract this information is an optical system that forms an image, reproducing the source's spatial and spectral intensity distribution. However, there is a limit to the level of detail that can be resolved in the image of the source both due to fundamental physical boundaries – the diffraction at the aperture – and due to atmospheric turbulence.

While the increase in aperture size, improving the angular resolution, has its limitations, the combination of two or more apertures forming an interferometer has the maximum baseline as the limiting parameter. The imaging process in an interferometer requires to regard not only the amplitude of the light but also the coherence function, which is the correlation of the light amplitudes.

In this chapter, we will discuss the propagation of light, repeating briefly the scalar diffraction theory when the electromagnetic wave is represented by a scalar that is proportional to one component of the electric field vector. Using the Fresnel approximation we will introduce the thin lens as the simplest form of an optical system, and we will discuss the properties of the Airy disk.

The formalism for the propagation of the amplitude will then be expanded to the propagation of the coherence function, describing the coherence properties of the propagating wave. We will mainly restrict the discussion to incoherent sources and investigate partially coherent sources only as an exception. The main result is the van Cittert–Zernike theorem providing the link between the source intensity distribution and the coherence function in a plane far away from the source, e.g. in the aperture plane of our optical system.

Young's experiment with two pinholes will serve as an example for an interferometer, explaining the effect of the source shape on the coherence function by applying the theory developed in Sect. 2.3, and using a heuristic approach without coherence functions.

Finally, we will discuss the intensity interferometer involving higher order correlation functions of the amplitudes. We shall see that the intensity interferometer is less demanding in terms of opto-mechanical precision but in turn much less sensitive than an amplitude interferometer.

## 2.1 Preliminaries

### 2.1.1 Basic Properties of the Electromagnetic Wave

The electric field vector  $\mathbf{E}$  of the electromagnetic wave is a function of space and time. Assuming a monochromatic plane wave in vacuum propagating in the  $z$ -direction (see Fig. 2.1), the  $x$ -component of  $\mathbf{E}$  can be written as

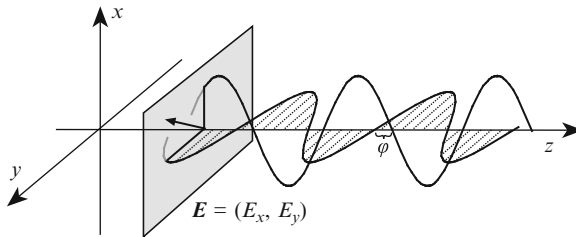
$$E_x = E_{x0} \cos(\omega t - kz), \quad (2.1)$$

with  $\omega = 2\pi\nu$ , time  $t$ , and  $k = 2\pi/\lambda$ , where  $\nu$  is the frequency and  $\lambda$  is the wavelength of the monochromatic wave. The phase velocity of the wave is  $\omega/k = \nu\lambda$ . The  $y$ -component can be described in the same way.

If the cosine functions of the  $x$ - and the  $y$ -component are not in phase, the resulting electric field vector  $\mathbf{E}$  rotates around the propagation axis, describing a circle if the components are of equal amplitude with a phase difference of  $\varphi = \pi$ . The light is then called *circularly polarized*. If the phase difference is zero, the two components are in phase and the resulting field vector  $\mathbf{E}$  oscillates in a plane. Then, the light is called *linearly polarized*. For all other phase differences, the light is *elliptically polarized*. When discussing polarization properties independent of a particular coordinate system one usually refers to the orthogonal components of the electric field vector as  $s$  and  $p$  polarization.

Since we will only describe the wave propagation in isotropic media the magnetic field  $\mathbf{H}$  could be used just as well to describe the electromagnetic wave.

The intensity is the quantity that we usually measure with optical detectors. It is related to the energy flow density given by the *Poynting vector*  $\mathbf{S} = \mathbf{E} \times \mathbf{H}$ . The Poynting vector is perpendicular both to  $\mathbf{E}$  and  $\mathbf{H}$ , and it points into the direction of propagation of the electromagnetic wave. If the electric field vector  $\mathbf{E}$  is linearly polarized in the  $x$ -direction then the magnetic field vector  $\mathbf{H}$  has a component only in the  $y$ -direction, and the wave propagates in the  $z$ -direction. Then, the Poynting



**Fig. 2.1** The electric field vector  $\mathbf{E}$ , split into two orthogonal components  $E_x$  and  $E_y$ , propagating in the  $z$ -direction. If one of the two components is shifted by a phase difference  $\varphi \neq 0$  the wave is polarized elliptically, i.e., the field vector – displayed as a *short arrow* – rotates around the  $z$ -axis describing an ellipse while propagating. If the phase difference  $\varphi$  is 0 the light is linearly polarized

vector has a component in the  $z$ -direction only, which can be written as

$$S_z = c\epsilon_0 E_{x0}^2 \cos^2(\omega t - kz). \quad (2.2)$$

$S_z$  has the dimension of  $\text{W/m}^2$ .  $c$  is the speed of light in vacuum with  $c = 2.998 \times 10^8 \text{ m/s}$ , and  $\epsilon_0$  is the permittivity of vacuum with  $\epsilon_0 = 8.854 \times 10^{-12} \frac{\text{A s}}{\text{V m}}$ . In a medium with refractive index  $n$ , the actual speed of light  $c/n$  has to be used, and  $\epsilon_0$  has to be replaced by the electric permittivity in the medium, usually expressed by  $\epsilon\epsilon_0$ , with  $\epsilon$  the dielectric constant of the material.

The Poynting vector oscillates with twice the frequency  $\nu$  of the electromagnetic wave, which is about  $10^{15} \text{ Hz}$  in the visible part of the spectrum. Since the temporal resolution of the available detectors is much lower than  $10^{-15} \text{ s}$ , one can only measure the time average of the Poynting vector defined as

$$\langle S_z \rangle = \lim_{T \rightarrow \infty} \frac{1}{2T} c\epsilon_0 \int_{-T}^T E_{x0}^2 \cos^2(\omega t - kz) dt = \frac{c\epsilon_0}{2} E_{x0}^2, \quad (2.3)$$

where  $\langle \cdot \rangle$  denotes the time average as defined above.

In practice, it is sufficient if the integration interval  $T$  is much longer than any of the processes involved, thus  $T \gg 1/\nu$ . Even time intervals down to  $10^{-12} \text{ s}$  fulfil this requirement.

**NB 1.** *In some cases, heterodyne detection can be applied to detect amplitude and phase of the light wave by mixing two waves of very similar frequency – one of them precisely defined in amplitude and phase – and measuring the beat frequency. This was done the first time in 1955 by mixing Zeeman components of a visible spectral line [70]. In astronomy, this is applied in the mid-infrared around a wavelength of  $10 \mu\text{m}$ , corresponding to a frequency of  $3 \times 10^{13} \text{ Hz}$ . Here, light from a local oscillator, typically a  $\text{CO}_2$  laser, is combined with the light to be detected producing a beat frequency signal in the  $\text{GHz}$  range [109]. This signal can be temporally resolved, and the amplitude and phase of the mid-infrared light can be determined. The limitation of heterodyne detection to the mid-infrared comes from the fact that at shorter wavelengths the sensitivity deteriorates very quickly [237], and that lasers with longer wavelengths are not available. However, if one aims at wavelengths beyond  $100 \mu\text{m}$ , in micro-wave and radio interferometers, suitable tunable oscillators are available and heterodyne detection is the standard measurement method.*

The time average of the Poynting vector is called the *flux* (in astronomy) or the *irradiance* (in radiometry) of the electromagnetic wave in units of  $\text{W m}^{-2}$ . The measurable quantity in an optical detector is the integral of the flux over the area of the detector, i.e., the power in units of Watt.

Since, throughout this book, we are above all interested in the spatial flux distribution and not in absolute values we will work with dimensionless quantities. At first, we introduce a dimensionless scalar, the *optical disturbance*  $v(\mathbf{r}, t)$  that is proportional to one component, e.g.,  $E_x$ , of the electric field vector with  $v = CE_x$  and

$C$  a suitable constant to make  $v$  dimensionless. The  $y$  component of the electric field vector can be treated independently.

As long as optical systems like, e.g., astronomical telescopes interact linearly with the amplitude and the phase of the incoming light wave we can use product relations between the light wave and a so-called *transfer function* of the optical system. For the mathematical treatment, it is very convenient to extend the optical disturbance by an imaginary part so that  $v$  becomes a complex quantity. A linear optical system is then described by a complex transfer function, and the outgoing wave can be calculated as the product of two complex functions, the complex optical disturbance and the complex transfer function.

For a plane wave propagating along the  $z$ -axis, the extension of the optical disturbance by an imaginary part reads as

$$v(z, t) = v_0 \cos(\omega t - kz) - i v_0 \sin(\omega t - kz) = v_0 e^{-i(\omega t - kz)}. \quad (2.4)$$

It is important to keep in mind that the introduction of the complex optical disturbance is just a convenience and that only the real part has physical significance, since only the real part represents the electromagnetic wave.

The time average of the product  $vv^*$  – the superscript  $*$  denoting the complex conjugate – can be used to define the *intensity* as

$$I(z) := \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T v(z, t) v^*(z, t) dt = v_0^2. \quad (2.5)$$

The intensity is a dimensionless quantity that is proportional to the flux  $\langle S_z \rangle$  and, thus, proportional to the signal that is measured with optical detectors. Throughout this book, we will always use the intensity  $I$  instead of the flux.

For the propagation of light in space it is convenient to introduce the time independent dimensionless *amplitude*  $V(\mathbf{r})$  at frequency  $\nu$  so that the monochromatic optical disturbance can be written as

$$v(\mathbf{r}, t) = V(\mathbf{r}) e^{-i2\pi\nu t}. \quad (2.6)$$

With (2.5) one sees that the intensity is  $I(\mathbf{r}) = |V(\mathbf{r})|^2$ .

In the case of **polychromatic light**, the linear superposition of individual monochromatic waves forms the polychromatic optical disturbance. We introduce the time independent *spectral amplitude*  $V(\mathbf{r}, \nu)$  such that it is

$$v(\mathbf{r}, t) = \int_0^\infty V(\mathbf{r}, \nu) e^{-i2\pi\nu t} d\nu. \quad (2.7)$$

$V(\mathbf{r}, \nu)$  has the dimension of  $\text{Hz}^{-1}$ . The integration is restricted to the positive frequency arm since the spectral amplitude has to be zero for negative frequencies. Only then the physically relevant real part of the optical disturbance contains all the information, and the imaginary part can be ignored. Since this was pointed out

by Gabor [76] the complex optical disturbance is also called *Gabor's analytic signal*. Formally, we can extend the integration to  $-\infty$  writing the optical disturbance  $v(\mathbf{r}, t)$  as the temporal Fourier transform of the spectral amplitude  $V(\mathbf{r}, \nu)$  (see Sect. A.1).

The *spectral intensity*  $I(\mathbf{r}, \nu)$ <sup>1</sup> is defined such that it is

$$I(\mathbf{r}) = \int_0^\infty I(\mathbf{r}, \nu) d\nu. \quad (2.8)$$

The dimension of  $I(\mathbf{r}, \nu)$  is  $\text{Hz}^{-1}$ . The spectral intensity is proportional to the *flux density* in units of  $\text{W m}^{-2} \text{Hz}^{-1}$ . In astronomy, a common unit for the flux density is 1 Jansky ( $\text{Jy}$ ) =  $10^{-26} \text{W m}^{-2} \text{Hz}^{-1}$ .

The spectral intensity is linked to the spectral amplitude through an averaging process similar to (2.5). The exact definition will be given in Sect. 2.3.1 in the context of coherence functions. When discussing the scalar diffraction theory in the following we will simplify this relationship by setting  $I(\mathbf{r}, \nu) = |V(\mathbf{r}, \nu)|^2$  without affecting the general conclusions of the discussions.

With the polychromatic intensity  $I(\mathbf{r})$  being the integral of the spectral intensities  $I(\mathbf{r}, \nu)$ , the propagation of polychromatic light through space and through optical systems can be treated by first considering the monochromatic case and then adding up the spectral intensities at the very end of the propagation process.

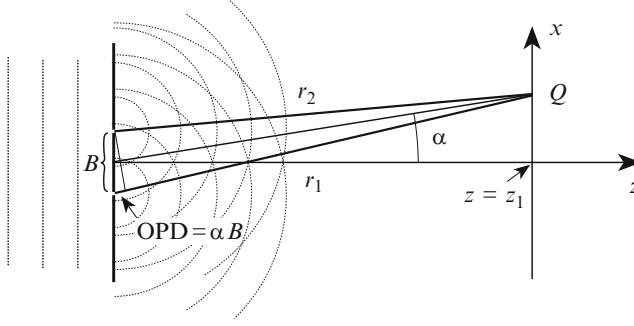
Throughout the book, we will use the monochromatic amplitude  $V(\mathbf{r})$  or the spectral amplitude  $V(\mathbf{r}, \nu)$  to discuss diffraction effects and the propagation of light in space, and we will use (2.8) to calculate the intensity. When discussing the coherence function in Sect. 2.3 we will follow a similar principle. This facilitates the mathematical treatment enormously. But we should keep in mind that the simple relationship between monochromatic and polychromatic intensity (2.8) is valid only because optical detectors measure the intensity as a time average. Otherwise the measured signal would show high frequency oscillations and the relationship would be more complicated.

### 2.1.2 Young's Experiment

This is the classical diffraction experiment that we will discuss very often in the following sections. It is named after Thomas Young who conducted it in 1802, providing the experimental cornerstone for the demonstration of the wave nature of light [256]. Describing Young's experiment, we will introduce diffraction effects and the concept of coherence.

---

<sup>1</sup> In the literature, this is sometimes called the *power spectral density*, since it gives the power per Hz.



**Fig. 2.2** The geometry of Young's experiment in a plane across the pinholes. A plane wave illuminates the two pinholes separated by a distance  $B$ . For small angles,  $\alpha$  is approximately equal to  $x/z_1$  with  $x$  the coordinate of  $Q$ , the point of observation. The difference  $r_1 - r_2$  of the distances from the pinholes to  $Q$  is called the optical path difference (OPD), which equals  $\alpha B$  for small angles  $\alpha$ . The OPD is related to the difference in arrival time  $\tau$  between the light from the two pinholes by  $\tau = \text{OPD}/c$

Young illuminated a screen with two pinholes with light from a single pinhole at a large distance. On passing through the pinholes the light was diffracted and the waves from the two pinholes interfered. On a second screen, the diffraction pattern could be observed showing the characteristic fringe pattern. In Fig. 2.2 the experiment is depicted schematically for an illuminating source at very large distance from the screen so that an approximately plane wave illuminates the two pinholes.

For the mathematical treatment we assume that the two pinholes are so small that they can be regarded as point sources of spherical waves. The spectral amplitude of a spherical wave at a distance  $r = |\mathbf{r}|$  from its origin can be written as

$$V(\mathbf{r}, \nu) = \frac{V_0}{r} e^{ikr}. \quad (2.9)$$

The spectral intensity is the squared modulus of the amplitude,  $I(\mathbf{r}, \nu) = V_0^2/r^2$ , decreasing with the square of the distance from the source.

The assumption that spherical waves originate from each pinhole comes across quite naturally in this context. However, the underlying concept is the Huygens–Fresnel principle of elementary waves, forming the basis of scalar diffraction theory that will be treated more formally in Sect. 2.2.

The amplitude  $V$  at point  $Q = (x, 0)$  in a plane at distance  $z_1$  is then the sum of the two spherical waves originating from the pinholes,

$$\begin{aligned} V(x, \nu) &= \frac{V_0}{r_1} e^{ikr_1} + \frac{V_0}{r_2} e^{ikr_2} \\ &= \frac{V_0}{z_1} e^{ik(r_1+r_2)/2} 2 \cos(k(r_1 - r_2)/2). \end{aligned} \quad (2.10)$$

$r_i$  is the distance between an individual pinhole and the point  $Q$ , with the approximation  $r_1 = r_2 = z_1$  for the amplitudes  $V_0/r_i$ .

The spectral intensity is the squared modulus of the amplitude,

$$I(x, \nu) = |V(x, \nu)|^2 = \left(\frac{V_0}{z_1}\right)^2 2(1 + \cos(k(r_1 - r_2))). \quad (2.11)$$

For small diffraction angles  $\alpha$ , it is  $\alpha = x/z_1$ , and  $r_1 - r_2 = \alpha B$  is the *optical path difference* (OPD) between the optical paths from each pinhole to the point of observation. A difference in optical path length is at the same time a difference in arrival time called the time delay  $\tau$  between the light from the two pinholes, with  $\tau = \alpha B/c$ . We will see later that the diffraction pattern for increasing diffraction angles  $\alpha$  – corresponding to increasing time delays  $\tau$  – is constrained by the temporal coherence of the incoming light.

We can now write the monochromatic intensity distribution of the diffraction pattern in its familiar form as a function of the diffraction angle  $\alpha$ ,

$$I(\alpha, \nu) \propto (1 + \cos(k\alpha B)). \quad (2.12)$$

This intensity distribution is also called the *fringe pattern* since the maxima of the cosine function appear as fringes on a two-dimensional screen. The first intensity minimum is at  $\alpha_{\min} = \lambda/(2B)$  as stated by Young as a result of his experiment. The OPD at the first minimum is  $r_1 - r_2 = \lambda/2$  and, consequently, the time delay is  $\tau = (\lambda/2)/c = 1/(2\nu)$ . Figure 2.3 shows the fringe pattern for a pinhole separation of 10 cm at wavelengths of 2.0, 2.2 and 2.4  $\mu\text{m}$ .

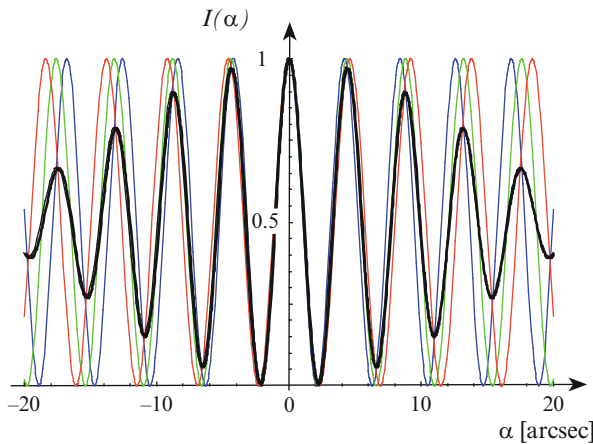
The monochromatic fringe patterns have excellent contrast since the intensity oscillates between 0 and 1. This can be expressed more formally by defining the contrast called the *fringe visibility* as introduced by Michelson [156] as

$$\mathcal{V} = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}. \quad (2.13)$$

With  $I_{\min} = 0$  and  $I_{\max} = 1$ , the contrast of the fringe pattern is  $\mathcal{V} = \infty$ .

Stellar interferometry is about measuring the contrast of fringes. Hence, we should have a closer look at the result of Young's experiment. A contrast of 1 in a fringe pattern, i.e., perfect constructive and destructive interference in its maxima and minima, implies that the light waves from the two pinholes are perfectly coherent. In fact, the monochromatic plane wave illuminating the screen with the two pinholes and, thus, the light waves emerging from the pinholes are perfectly coherent.

The term *coherence* is linked to the existence of interference phenomena in diffraction experiments like Young's experiment. A fringe pattern with a good contrast requires a good coherence between the light waves from the two pinholes. Light of perfect coherence causes a fringe pattern with a contrast of 1 as stated above. If there is no coherence between the light from the pinholes, there is no fringe pattern



**Fig. 2.3** Three individual monochromatic fringe patterns of Young's experiment, for  $\lambda = 2.0, 2.2$  and  $2.4 \mu\text{m}$  (blue, green and red lines), and the resulting polychromatic fringe pattern of the full  $K$ -band,  $I(\alpha) = \int I(\alpha, \nu) d\nu$ , displaying decreasing contrast for increasing diffraction angle  $\alpha$ . The pinhole separation is 10 cm. The first minima of the individual monochromatic fringe patterns are at  $\alpha_{\min} = \lambda/(2B) = 2.1, 2.3$  and  $2.5$  arcsec from the central maximum

but only a homogeneous illumination as a result of the diffraction of light at each individual aperture. Then, the light is called incoherent.

Implicitly, by discussing the behaviour of the amplitude only, we have used the fact that optical detectors perform a time average over a period much longer than  $1/\nu$ . If we had assumed that we have an ideal detector temporally resolving the oscillating electromagnetic wave we would then measure the instantaneous intensity of the light wave that is affected by the temporally varying parts  $\exp(i2\pi\nu t)$  of the optical disturbances in the two pinholes. The effect is that in any given moment, there is constructive or destructive interference at any point in the plane of observation regardless of the state of coherence. This diffraction pattern is oscillating typically with twice the light frequency. It settles to the familiar fringe pattern by the temporal averaging process when measuring the intensity with a real detector.

In Sect. 2.4, we will see in detail how the state of coherence of the light in the pinholes depends on the source properties. As an introduction, we now discuss the effect of a finite *spectral bandwidth*  $\Delta\nu$  of the light source. The resulting fringe pattern is formed by adding up interference patterns like (2.12) at different frequencies to obtain the observed intensity distribution. This is displayed in Fig. 2.3 for the  $K$ -band,<sup>2</sup> with  $\lambda = 2.2 \pm 0.2 \mu\text{m}$ , and  $\Delta\lambda = 0.4 \mu\text{m}$ , and a pinhole separation of 10 cm. For zero OPD, at  $\alpha = 0$ , all wavelengths have an intensity maximum. This is why this fringe is called the *white-light fringe*. The position of the first

<sup>2</sup> The atmosphere transmits only certain bands in the infrared. One frequently used band in the near-infrared is the  $K$ -band at  $2.2 \pm 0.2 \mu\text{m}$ , (see Sect. A.2). Most of the numerical examples will be given for this band.



minimum  $\alpha_{\min} = \lambda/(2B)$  or  $\text{OPD} = \lambda/2$  is then wavelength dependent, as well as the positions of the following maxima and minima.

This effect reduces the contrast of the resulting polychromatic fringe pattern for increasing diffraction angles  $\alpha$ . Since  $\alpha$  is related to the difference in arrival time, the time delay  $\tau$ , through  $\tau = \alpha B/c$ , this effect can be reformulated by stating that the contrast of the resulting fringe pattern is reduced with increasing  $\tau$ . The time delay that is related to the quasi loss of fringe contrast is called the *coherence time*  $\tau_c$ , which is proportional to the reciprocal of the spectral bandwidth  $\Delta\nu$ . The exact relationship depends on the form of the spectral band.

This introduces the connection between temporal coherence, quantified by the coherence time  $\tau_c$ , and the fringe contrast in a rather casual way. The spatial coherence of the incoming light, and its influence on the fringe contrast, is more difficult to come by. In Sect. 2.3 we will treat coherence as a statistical phenomena, and we will see that both temporal and spatial coherence are part of the more general coherence theory.

## 2.2 Scalar Diffraction Theory

The foundations of scalar diffraction theory have been discussed in most standard texts on optics [19]. Goodman [88] gave a very concise summary of the historical background and the mathematical deduction. We will briefly revisit the cornerstones of this theory in order to give an idea of how the propagation of a coupled vector field, i.e., of the electromagnetic wave can be reduced to the Fourier transform of a scalar.

### 2.2.1 The Rayleigh–Sommerfeld Diffraction Formula

Forming images in optical instruments, the propagation of light and the diffraction at physical boundaries needs to be dealt with. The fundamental equations treating the propagation of electromagnetic waves are *Maxwell's equations* [19]. They describe the propagation of both the electric and the magnetic field vectors  $\mathbf{E}$  and  $\mathbf{H}$ , respectively, through coupled partial differential equations.

We will make a number of assumptions and approximations that are adequate for the imaging process. First, light is treated as a scalar quantity as introduced in Sect. 2.1, by defining a dimensionless scalar  $v(\mathbf{r}, t)$ , the *optical disturbance*, that is proportional to one component of, e.g., the electric field vector  $\mathbf{E}$ . Second, one assumes that the propagation of the two orthogonal vector components of  $\mathbf{E}$  that are perpendicular to the direction of propagation can be treated independently.

This approximation is valid if (1) diffracting apertures are large compared to the wavelength, and if (2) the diffracted fields are not observed too close to the

diffracting apertures. Both conditions are easily fulfilled in optical systems in general, and in astronomical telescopes in particular.

One solution of the time dependence of the differential equation describing the propagation of  $v(\mathbf{r}, t)$  is given by the monochromatic complex harmonic wave  $V(\mathbf{r})\exp(-i2\pi\nu t)$  in (2.6). Then, Maxwell's equation can be replaced by the *Helmholtz equation*, which is the time-independent form of the wave equation, acting on the amplitude  $V(\mathbf{r})$ :

$$(\nabla^2 + k^2)V(\mathbf{r}) = 0, \quad (2.14)$$

with  $\nabla^2$  being the Laplace operator  $\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ , and  $k = 2\pi/\lambda$ . Solutions of this differential equation are a monochromatic plane wave, for example  $V(\mathbf{r}) = V_0 \exp(ikz)$  for a wave propagating along the  $z$ -axis, or a monochromatic spherical wave  $V(\mathbf{r}) = \frac{V_0}{r} \exp(ikr)$  with  $r = |\mathbf{r}|$ .

In optical systems, one is interested in the amplitude  $V$  in the plane of observation, the image plane, as a function of the amplitude in the aperture plane, or as a function of the amplitude of the object. Any solution for  $V$  needs to satisfy the Helmholtz equation at all points  $\mathbf{r}$ .

A starting point for tackling this problem is Green's integral theorem assuming two functions  $V$  and  $G$  that are solutions of the Helmholtz equation. The choice of function  $G$ , also called Green's function, is now crucial. Kirchhoff's choice of a spherical wave  $G = \exp(ikr)/r$  leads to the *integral theorem of Huygens and Kirchhoff* obtaining the amplitude  $V$  at point  $P_0$  as a function of the amplitude on a closed surface  $\mathbf{A}$  surrounding  $P_0$ :

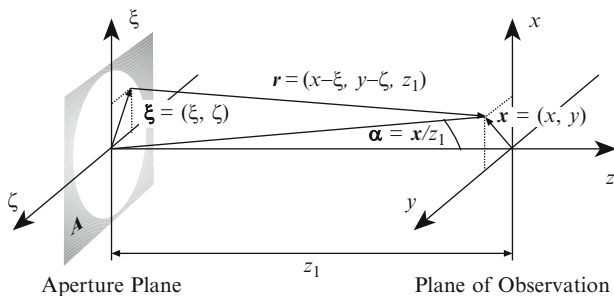
$$V(P_0) = \frac{1}{4\pi} \iint_{\mathbf{A}} (G \nabla V - V \nabla G) d\xi, \quad (2.15)$$

where  $\nabla = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z})$  is the gradient or nabla operator.

Kirchhoff further developed this formula by performing the step from a closed surface to a plane surface limited by an aperture. Thus, he described a situation that is more similar to the setup in optical systems. This led to the Fresnel-Kirchhoff diffraction formula. Although it has been found experimentally to yield remarkably accurate results, the choice of Green's function required a number of assumptions known as the Kirchhoff boundary conditions implying, when applied rigorously, that the amplitude  $V$  needs to be identical to zero everywhere in space.

This inconsistency was removed by Sommerfeld by assuming two point sources emitting spherical waves that are placed symmetrically to the aperture on either side of it. Then, two new Green's functions  $G_-$  or  $G_+$  can be defined as either the difference or the sum of the two spherical waves.

The final result known as the *Rayleigh-Sommerfeld diffraction formula* forms the basis of all our discussions on diffraction and propagation of light [88]. In the approximation of small diffraction angles, the propagation of the amplitude from the aperture plane  $\xi$  into the plane of observation  $\mathbf{x}$ , see Fig. 2.4, is computed by



**Fig. 2.4** The geometry for the diffraction at an aperture **A**. The aperture plane has coordinates  $(\xi, \zeta)$  and the coordinate vector  $\xi$ , the plane of observation is denoted by  $(x, y)$  with the vector  $\mathbf{x}$ . The distance between the planes is  $z_1$  and the vector between two points  $(\xi, \zeta)$  and  $(x, y)$  is  $\mathbf{r} = (x - \xi, y - \zeta, z_1)$ , with the notation  $r = |\mathbf{r}|$ . We call the  $z$ -axis of the coordinate system the *optical axis*. Usually, all elements of optical systems, i.e., apertures, lenses and mirrors, are centred on the optical axis

$$V(\mathbf{x}) = \frac{1}{i\lambda} \iint_{\mathbf{A}} V(\xi) \frac{e^{ikr}}{r} d\xi. \quad (2.16)$$

Let us assume for the moment that the incoming wave is a plane wave with  $V(\xi) = V_0$  in the aperture plane. Then the integration is performed over spherical waves  $\frac{V_0}{r} \exp(ikr)$  originating from each point in the aperture, and the superposition of these elementary waves forms the wave front<sup>3</sup> behind the aperture. This concept is called the *Huygens–Fresnel principle* of elementary waves. Huygens first expressed this principle out of intuition. In the context of seeking solutions for the integral theorem (2.15) the elementary waves come in as Green's functions.

In the extreme case of an infinitely large aperture **A** that is illuminated by a plane wave, a plane wave is expected to arrive in the plane of observation. In fact, the result of the integration (2.16) with infinite integration boundaries cancels with the factor  $1/(i\lambda)$  and a plane wave remains.

The other extreme is a single pinhole when a lone spherical wave originates from the pinhole. Since this is in conflict with the condition that apertures need to be large compared to the wavelength we will use the term *pinhole* if an aperture is small enough to describe the diffraction pattern for small diffraction angles (2.16) by a single spherical wave originating from the aperture. A pinhole diameter of a few wavelengths is sufficient to fulfil this condition.

One can now regard  $1/(i\lambda)$  as a weighting function of the spherical wave  $\frac{V_0}{r} \exp(ikr)$ . Then, the amplitude of the spherical wave is reduced by  $1/\lambda$  compared to that of the incident wave  $V(\xi)$ . This is intuitively comprehensible since long wavelengths are diffracted more strongly and the amplitude decreases over a given area in the plane of observation. The phase of  $V(\xi)$  is shifted by  $-\pi/2$  (since

<sup>3</sup> The term *wave front* describes the virtual surface of the same phase of a propagating wave.

$1/i = \exp(-i\pi/2)$ ) by the very process of diffraction at the pinhole. This is curious in a historical sense because Fresnel assumed that the secondary sources have this property in order to accurately calculate diffraction patterns. The deduction leading to the diffraction formula showed the mathematical necessity of this phase factor.

Returning to Young's experiment, we see that it represents the case when the aperture is reduced to two pinholes, and the resulting intensity distribution in plane  $\mathbf{x}$  is calculated by adding the two spherical waves originating from the pinholes as a consequence of the general diffraction formula (2.16).

### 2.2.2 Fresnel Approximation

For small angles of the distance vector  $\mathbf{r}$  with respect to the  $z$ -axis,  $|\mathbf{x} - \boldsymbol{\xi}|/z_1$  (see Fig. 2.4), its length  $r$  can be approximated by  $z_1$ . However, for the phase  $kr$  in (2.16) the requirement is more stringent. Here, the residual of the approximation for  $r$  needs not only be relatively small but the approximation of the product  $kr$  has to be accurate within fractions of a radian. Developing  $r$  into a power series and neglecting orders higher than quadratic for  $|\mathbf{x} - \boldsymbol{\xi}|^4/z_1^3 \ll \lambda$ , yields the *quadratic Fresnel approximation*:

$$r = z_1 + \frac{|\mathbf{x} - \boldsymbol{\xi}|^2}{2z_1} - \dots \quad (2.17)$$

This approximation is of fundamental importance for Fourier optics.

Inserted in (2.16) one obtains

$$\begin{aligned} V(\mathbf{x}) &= \frac{1}{i\lambda z_1} e^{ikz_1} \iint_{\mathbf{A}} V(\boldsymbol{\xi}) e^{ik|\mathbf{x}-\boldsymbol{\xi}|^2/(2z_1)} d\boldsymbol{\xi} \\ &= \frac{1}{i\lambda z_1} e^{ikz_1} e^{ik|\mathbf{x}|^2/(2z_1)} \iint_{\mathbf{A}} V(\boldsymbol{\xi}) e^{ik|\boldsymbol{\xi}|^2/(2z_1)} e^{-ik\mathbf{x}\cdot\boldsymbol{\xi}/z_1} d\boldsymbol{\xi}. \end{aligned} \quad (2.18)$$

The two exponential functions before the integral describe a constant phase  $kz_1$  due to the propagation between the planes, and a parabolic phase  $k|\mathbf{x}|^2/(2z_1)$  due to the quadratic approximation for spherical waves originating from the aperture. These two functions disappear when calculating the intensity distribution  $I(\mathbf{x}) = |V(\mathbf{x})|^2$  in a plane at  $z_1$ . We will omit them in the remainder of this section.

The shape of the aperture  $\mathbf{A}$  is called  $A(\boldsymbol{\xi})$ , and it is incorporated under the integral through  $V_{\text{ap}}(\boldsymbol{\xi}) = A(\boldsymbol{\xi})V(\boldsymbol{\xi})$ . Then, the amplitude for *Fresnel diffraction* [88] can be written as

$$V(\mathbf{x}) = \frac{1}{i\lambda z_1} \iint_{-\infty}^{\infty} V_{\text{ap}}(\boldsymbol{\xi}) e^{ik|\boldsymbol{\xi}|^2/(2z_1)} e^{-ik\mathbf{x}\cdot\boldsymbol{\xi}/z_1} d\boldsymbol{\xi}. \quad (2.19)$$

In the following, integration boundaries will only be given explicitly if they are finite.

At a very large distance from the aperture, i.e., for  $|\xi|^2/(2z_1) \ll \lambda$ , the argument of the first exponential under the integral in (2.19) goes to zero. This approximation is called the *Fraunhofer approximation*. Then the amplitude  $V(x)$  in the plane of observation and the amplitude  $V_{\text{ap}}(\xi)$  in the aperture are linked through a Fourier transform.

There is an apparent contradiction when considering the case of an infinitely large aperture. In the discussion following (2.16) we stated that an aperture of infinite extent lets a plane wave travel unaffected by any aperture boundaries into the plane of observation. This means that the diffraction pattern of the infinitely large aperture is a plane wave of infinite extent. This is in contrast to the diffraction pattern in Fraunhofer approximation that is computed by Fourier transforming the infinite aperture formally obtaining the Dirac  $\delta$ -function as an intensity distribution (see Sect. A.1). However, the Fraunhofer approximation  $|\xi|^2/(2z_1) \ll \lambda$  requires not only a large distance  $z_1$  but also a very small angular size  $|\xi|/z_1 \ll \sqrt{\lambda/z_1}$  of the aperture. If the aperture is of infinite extent the Fraunhofer approximation cannot be applied.

At the other extreme, we discussed in Sect. 2.2.1 that a very small aperture, a pin-hole, cannot be infinitely small but needs a diameter of a few wavelengths. Formally such an aperture can be expressed by a  $\delta$ -function since we investigate only small diffraction angles within the Fresnel approximation.

Returning to apertures of finite size, we regard an aperture with a diameter of 10 cm and an observing wavelength of 2.2  $\mu\text{m}$ . Then, the Fresnel approximation is valid for distances  $z_1 \gg 4$  m while in Fraunhofer approximation the distance has to be much larger than 2,500 m in order to properly compute the diffraction pattern with a Fourier transform (see Fig. 2.5).

Using a lens (or a parabolic mirror), Fraunhofer diffraction can be observed closer to the aperture. For our purpose, it is sufficient to regard the lens as a simple focusing element that converts an incoming plane wave into a spherical wave converging in the focus. The distance between the lens and the focus is called the focal length.

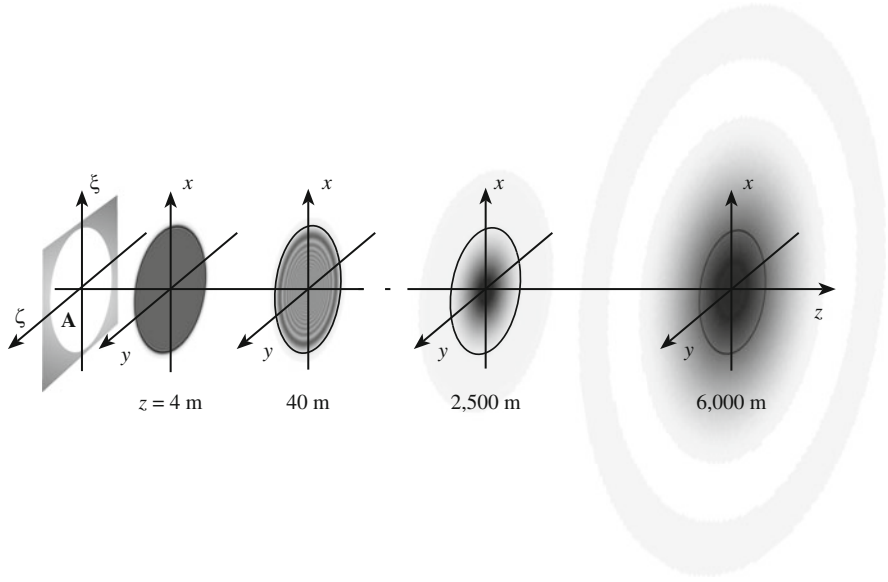
In quadratic approximation, the lens can be described by

$$L(\xi) = e^{-ik|\xi|^2/(2F)}, \quad (2.20)$$

with  $F$  the focal length [88]. The ideal lens has no absorption, hence  $|L(\xi)| = 1$ , it affects the phase of the incoming wave by adding a parabolic phase term  $k|\xi|^2/(2F)$ , and it is infinitely thin. A lens in this definition is called a *thin lens*, and a mirror a *thin mirror*.

Inserting  $L(\xi)$  in (2.19) one obtains

$$V(x) = \frac{1}{i\lambda z_1} \iint V_{\text{ap}}(\xi) e^{-ik|\xi|^2/(2F)} e^{ik|\xi|^2/(2z_1)} e^{-ikx \cdot \xi/z_1} d\xi. \quad (2.21)$$



**Fig. 2.5** Intensity distributions of the diffraction patterns of an aperture with 10 cm diameter and an observing wavelength of  $\lambda = 2.2 \mu\text{m}$  at different distances. 4 m after the aperture, the intensity distribution is very similar to the geometrical shadow pattern of the aperture that is indicated as a *circle* in each diffraction pattern. At about 40 m the Fresnel approximation describes the situation properly showing typical ringing effects inside the aperture and a fast decrease to zero outside. At about 2,500 m the intensity distribution resembles the Fraunhofer diffraction pattern, which is fully developed at 6,000 m. The diameter of the first diffraction ring is 12 cm at 2,500 m and 30 cm at 6,000 m. Note that this diameter is never smaller than the diameter of the aperture. For an observing wavelength of  $10 \mu\text{m}$  the Fresnel approximation is a proper approximation at 21 m (instead of 36 m), and the Fraunhofer diffraction pattern appears at 600 m (instead of 6,000 m). (Courtesy R. Wilhelm)

In the focal plane, with  $z_1 = F$ , the first two exponential functions under the integral cancel. Following common practice, we replace the spatial coordinate  $\mathbf{x}$  in the focal plane by the angle coordinate  $\boldsymbol{\alpha} = \mathbf{x}/F$ . Then, the amplitude  $V(\boldsymbol{\alpha})$  in the focal plane can be expressed as the Fourier transform of the amplitude  $V_{\text{ap}}(\boldsymbol{\xi})$  in the aperture [88]

$$V(\boldsymbol{\alpha}) = \frac{1}{i\lambda F} \iint V_{\text{ap}}(\boldsymbol{\xi}) e^{-ik\boldsymbol{\alpha} \cdot \boldsymbol{\xi}} d\boldsymbol{\xi}. \quad (2.22)$$

With this formula we can describe the situation in a telescope. The incoming plane wave stems from a point-like star approximately at infinity, i.e., the phase  $\varphi(\boldsymbol{\xi})$  of the complex amplitude is zero, and  $V(\boldsymbol{\xi}) = V_0 = \text{constant}$  so that  $V_{\text{ap}}(\boldsymbol{\xi}) = V_0 A(\boldsymbol{\xi})$ . The light is diffracted at the aperture, and the telescope optics form the Fraunhofer diffraction pattern of the aperture in the telescope focal plane with the

intensity distribution given by  $I(\boldsymbol{\alpha}) = |V(\boldsymbol{\alpha})|^2$ . Thus, this intensity distribution is the diffraction limited image of the point-like star, and the star is called *unresolved*.

In the theory of linear systems, the diffraction pattern represents the response of the optical system to an impulse, in this case the approximately point-like intensity distribution of an unresolved star. This response is called the *point-spread function* (PSF) of the optical system. The PSF is dimensionless and describes the spread of the intensity in the focal plane.

Aberrations of the telescope optics are incorporated in the phase  $\varphi(\boldsymbol{\xi})$  of the aperture function  $A(\boldsymbol{\xi})$ . Then, the subsequent PSF describes no longer the diffraction limited but the aberrated image of the point source.

**NB 2.** *Apart from placing the lens directly in the aperture plane, we can also place the aperture at a distance  $z_0$  in front of the lens. Then, the amplitude in the back focal plane of the lens can be written as [88]*

$$V(\boldsymbol{\alpha}) = \frac{1}{i\lambda F} e^{ik(1-\frac{z_0}{F})|\boldsymbol{\alpha}|^2 F/2} \iint V_{\text{ap}}(\boldsymbol{\xi}) e^{-ik\boldsymbol{\alpha}\cdot\boldsymbol{\xi}} d\boldsymbol{\xi}. \quad (2.23)$$

The amplitude in the back focal plane is again given as a Fourier transform of the amplitude  $V_{\text{ap}}(\boldsymbol{\xi})$  in the aperture, multiplied by a quadratic phase term. Following (2.18), we argued that this phase term needs not be considered since we are mainly interested in intensity distributions.

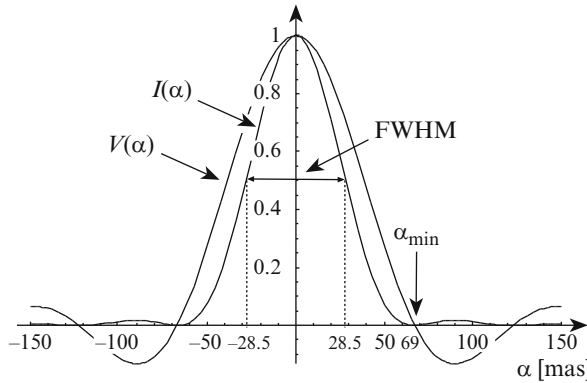
It is interesting, however, to discuss how this phase term is affected by the distance  $z_0$  from the lens. We see that for  $z_0 = F$ , i.e., with the aperture in the front focal plane of the lens, the phase term disappears, and we have a Fourier connection between the amplitude in the front focal plane and in the back focal plane without any further approximation.

Thus, the position of the aperture with respect to the lenses in optical system does not affect the Fourier connection and, unless choosing extreme parameters for focal length and distance, the quadratic phase term can be disregarded.

### 2.2.3 The Airy Disk

As shown in the preceding section, the response of an optical system to a unresolved source is never point-like but enlarged by the diffraction at the telescope aperture. We now discuss the circular aperture as the most common case of a telescope aperture.

The circular aperture is described by the circ-function that is defined as  $\text{circ}(\frac{|\boldsymbol{\xi}|}{R}) = 1$  if  $|\boldsymbol{\xi}| \leq R$  and 0 elsewhere. A circular aperture with diameter  $D$  is then given by  $A(\boldsymbol{\xi}) = \text{circ}(\frac{|\boldsymbol{\xi}|}{D/2})$ , with the area of the circular aperture called  $A_0 = \pi(D/2)^2$ .  $A(\boldsymbol{\xi})$  is illuminated by a point source at infinity with  $V(\boldsymbol{\xi}) = V_0 = \text{const.}$  in the aperture plane.



**Fig. 2.6** Profiles of the amplitude  $V(\alpha)$ , normalised to unity, and of the subsequent intensity distribution  $I(\alpha)$  of the point-spread function (PSF) for a circular aperture with a diameter of 8 m, and an observing wavelength of  $2.2 \mu\text{m}$ .  $V(\alpha)$  has the form of the Besinc-function. The first minimum of  $I(\alpha)$  at  $\alpha_{\min} = 1.22 \lambda/D = 69 \text{ mas}$  (we use the notation mas for milliarcsec), and the full width at half maximum (FWHM) of  $\alpha_{\text{FWHM}} = \lambda/D = 57 \text{ mas}$  are indicated. For the computation of the angle in mas, the formula  $\alpha_{\text{FWHM}}/\text{mas} = 206 \frac{\lambda/\mu\text{m}}{D/\text{m}}$  is useful, and for the FWHM as a length in the focal plane, we can use  $x_{\text{FWHM}} = f_{\#} \lambda$ , with  $f_{\#}$  the *f-number* of the optical system defined as  $f_{\#} = F/D$ . Note that the diffraction rings in the intensity distribution at  $\alpha \approx \pm 90 \text{ mas}$  reach only about 2% of the peak intensity

Using (2.22), the diffraction limited amplitude in the focal plane can be written as

$$\begin{aligned} V(\alpha) &= \frac{V_0}{i\lambda F} \iint \text{circ}\left(\frac{|\xi|}{D/2}\right) e^{-ik\alpha \cdot \xi} d\xi \\ &= \frac{V_0}{i\lambda F} A_0 \text{Besinc}(k\alpha D/2), \end{aligned} \quad (2.24)$$

with  $\alpha = |\alpha|$ .

The result of the Fourier transform (2.24) of the circ-function is called the *Besinc-function* (see Fig. 2.6) defined as  $\text{Besinc}(x) := 2J_1(x)/x$ , with  $J_1(x)$  the first order Bessel function. The Besinc-function is a real function that is point-symmetric with respect to  $x = 0$ . With  $A_0$ , the area of the circular aperture, and  $\lambda F$  both having the dimension of  $\text{length}^2$ , the quotient is dimensionless and the amplitude  $V(\alpha)$  has the same dimension as  $V_0$ , the amplitude of the incoming plane wave.

Splitting the amplitude  $V(\alpha)$  in (2.24) into modulus and phase, one finds that the modulus is simply the absolute value of the Besinc-function, and the phase takes on distinct values of  $-\pi/2$ , with  $1/i = \exp(-i\pi/2)$ , when the Besinc-function is positive and  $+\pi/2$  when it is negative. A phase function in the focal plane jumping between values of  $+\pi/2$  and  $-\pi/2$  means that the wave front in the focal plane is not curved but essentially flat with jumps of  $\pi$ .

This represents the transition from a converging spherical wave immediately after the optical system with the centre of the sphere in the focal plane, and a diverging spherical wave after the focal plane. However, despite the flat wave front the



amplitude in (2.24) is not a plane wave because of the phase discontinuities at the phase jumps and the modulus following a Besinc-function. This causes the wave to expand into a diverging spherical wave in the course of the propagation. Further calculations of the phase just before and after the focal plane show that there is a phase jump from  $-\pi$  to 0 when moving along the optical axis through the focal plane [19].

The intensity distribution of the diffraction limited PSF of telescopes with circular aperture is called the *Airy disk* [19]. The squared modulus of the amplitude  $V(\alpha)$  yields the Airy disk as

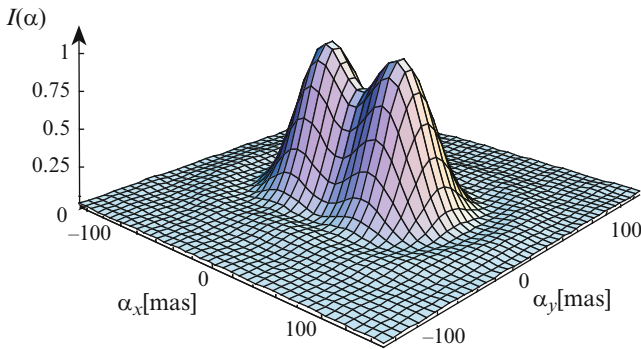
$$I(\alpha) = V(\alpha)V^*(\alpha) = \frac{V_0^2}{(\lambda F)^2} A_0^2 \text{Besinc}^2(k\alpha D/2). \quad (2.25)$$

The dimensionless PSF describing the shape of the intensity can now be defined as:

$$\text{PSF}(\alpha) := \frac{I(\alpha)}{V_0^2} = \frac{1}{(\lambda F)^2} A_0^2 \text{Besinc}^2(k\alpha D/2). \quad (2.26)$$

As long as we discuss monochromatic light, the wavelength appears as a constant parameter in the formula, and we write the PSF as a function of  $\alpha$  only. Later, in Sect. 3.3.3, when we will investigate the imaging process in polychromatic illumination we will write  $\text{PSF}(\alpha, \nu)$ .

The first minimum of the PSF, the first *dark ring* is at  $\alpha_{\min} = 1.22 \lambda / D$ . 84% of the total intensity are confined to the inside of the first dark ring. For a binary star with a separation  $\alpha_{\min}$ , the resulting image, which is the sum of two individual Airy disks, shows a local minimum between the peaks of the Airy disk (see Fig. 2.7). Therefore the two stars of the binary can be identified as individual objects in the



**Fig. 2.7** The joint intensity distribution in the diffraction limited image of a binary star separated by the Rayleigh limit of resolution of 69 mas for an aperture diameter of 8 m and an observing wavelength of  $2.2 \mu\text{m}$ . The local minimum between the two peaks is 19% lower than the intensity of the peaks

image. This criterion of angular resolution, when the smallest resolvable angle is  $\alpha_{\min}$ , is called the *Rayleigh criterion* of resolution of a telescope [19].

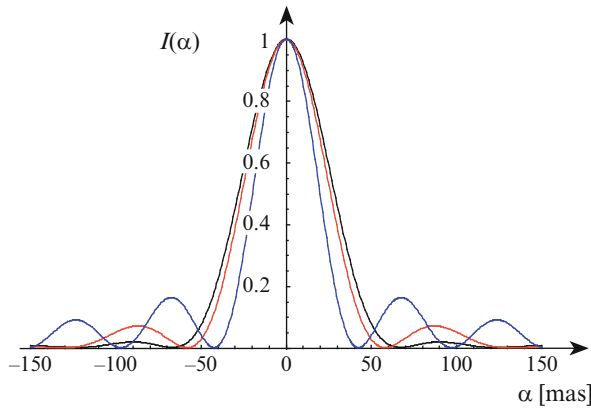
This definition is, however, somewhat arbitrary. Why should a local minimum with a larger value than 84% between the two peaks not be sufficient to resolve them? Therefore, another resolution criterion is often used. This criterion is related to the *full width at half maximum (FWHM)*, i.e., the diameter of the Airy disk at half its maximum intensity, which is in good approximation  $\alpha_{\text{FWHM}} = \lambda/D$  (see Fig. 2.6). About 50% of the total intensity is within the FWHM of the Airy disk.

If two stars are separated by  $\alpha_{\text{FWHM}}$  the intensity distribution between the two peaks of the Airy disks is almost flat so that the FWHM represents the limit to identify the peaks as separate. One could, however, argue that a much smaller separation could be spotted by precisely measuring the deviation of the resulting intensity distribution from the perfectly circular shape of an individual Airy disk. There is virtually no limit for the resolution unless the reality is taken into account with measurements that are never perfect [56, 87]. Implicitly, we have assumed a certain measurement quality in the discussion of Rayleigh criterion and FWHM that is sufficient for the identification of local intensity maxima, which can then be attributed to maxima of the object intensity.

Thus, the Rayleigh criterion has its virtue in providing a conservative estimate for the angular resolution that is linked to a prominent feature, the first minimum, of the Airy disk. We will see in Chap. 3 that the matter of angular resolution is more complex depending on many parameters of the imaging situation.

It should be noted that both quantities,  $\alpha_{\min}$  and  $\alpha_{\text{FWHM}}$ , depend on the shape of the telescope aperture that usually has a central obscuration due to the telescope design with the secondary mirror centrally above the primary mirror. For comparison, Fig. 2.8 displays three PSF: without central obscuration (black curve), with a central obscuration of 40% of the telescope aperture (red) and with an annular aperture (blue). The latter represents the theoretical limit with close to zero transmission. One can see that for the practical cases between 0 and 40% obscuration the PSF varies only slightly. While the position of the first minimum is reduced from  $1.22 \lambda/D$  to about  $\lambda/D$ , the FWHM is practically unchanged. Therefore we will always discuss the case of an unobscured aperture.

**NB 3.** *The possibility of increasing the angular resolution considerably beyond the Rayleigh limit was first recognised in the context of the development of microwave antennas by G. Toraldo di Francia in 1952 [58]. Combining nested ring apertures he showed that the intensity in the first diffraction rings could be suppressed so that the PSF was almost zero around its core up to a very bright diffraction ring limiting the field of view. The price to pay was a reduced peak intensity of the central core. Soon after, the connection between increase of angular resolution and decrease of peak intensity was analysed on general grounds [144], and more recently a quantitative analysis was presented [205].*



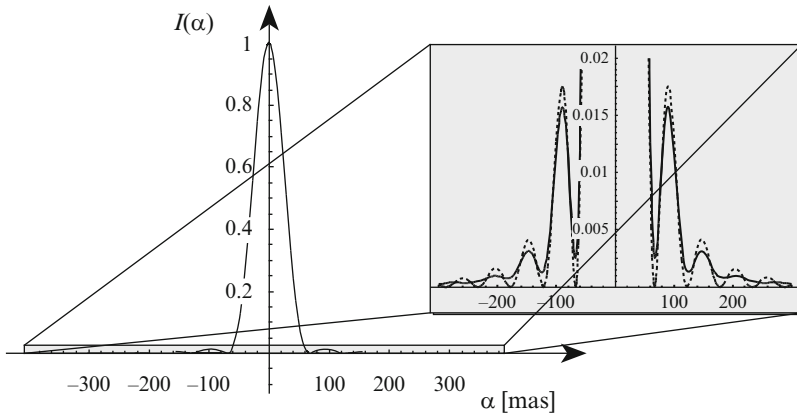
**Fig. 2.8** Examples of PSF with different central obscuration and a circular 8-m aperture for an observing wavelength of  $2.2\,\mu\text{m}$ . In *black* the PSF without central obscuration is displayed, as in Fig. 2.6. The PSF with a circular central obscuration of 40% of the telescope aperture is shown in *red*, and the PSF of an infinitely thin annular aperture is shown in *blue*. The first minimum of the PSF at  $1.22\,\lambda/D = 69\text{ mas}$  (*black curve*) is reduced to about  $\lambda/D = 57\text{ mas}$  (*red curve*) and further to  $0.76\,\lambda/D = 43\text{ mas}$  (*blue curve*). For comparison, all PSF are normalised to unity

### Polychromatic Airy Disk

So far, we have treated the case of monochromatic light. We now discuss the case of polychromatic illumination before moving on to the coherence function in the next section. The polychromatic intensity distribution  $I(\alpha)$  of an Airy disk can be calculated as the integral of the monochromatic Airy disks (2.25) over the spectral band. Using the atmospheric *K*-band ( $2.2 \pm 0.2\,\mu\text{m}$ ), the result of the *K*-band Airy disk as well as the monochromatic Airy disk are displayed in Fig. 2.9. Around the peak the difference is barely noticeable. The ringing, however, is reduced in the minima – not reaching zero any more – as well as in the maxima – being lower than in the monochromatic case – both due to the wavelength dependence of the minima and maxima positions. Beyond the third sidelobe this effect wipes out any diffraction rings in the polychromatic Airy disk, albeit at extremely low light levels.

The influence of the wavelength on the diffraction pattern was much more prominent in Young’s experiment (Fig. 2.3) when the contrast of the fringe pattern was reduced considerably. However, the physical process, namely the wavelength dependent diffraction, is the same in both cases.

This exercise of moving from monochromatic to polychromatic illumination by integrating over the spectral band will come back to us when treating the diffraction process in a stellar interferometer. There the situation is more complex since the properties of spatial coherence need also to be treated.



**Fig. 2.9** The Airy disk of an 8-m telescope, observing the full  $K$ -band,  $2.2 \pm 0.2 \mu\text{m}$ . In the *grey box*, the Airy disk is displayed with an enlarged intensity scale. Compared to the monochromatic Airy disk displayed by the *dashed line*, the sharp monochromatic minima disappear in the polychromatic Airy disk since their position is a function of wavelength

### Scalar Diffraction Theory: Summary

The Rayleigh–Sommerfeld diffraction formula (2.16) describes the propagation of the amplitude  $V(\xi)$  in space. Reducing the discussion to the propagation between two planes – the aperture plane and the plane of observation – in Fresnel approximation (2.17), the diffraction formula can be simplified considerably.

In addition, a lens with focal length  $F$  is introduced in the aperture plane. Then, the diffraction formula can be simplified further (2.21), and the Fraunhofer diffraction pattern of the aperture that usually appears at a very large distance can now be found in the back focal plane of the lens.

For Fraunhofer diffraction, the amplitude  $V_{\text{ap}}(\xi)$  in the aperture and the amplitude  $V(\alpha)$  in the focal plane are linked through a Fourier transform:

$$V(\alpha) = \frac{1}{i\lambda F} \iint V_{\text{ap}}(\xi) e^{-ik\alpha \cdot \xi} d\xi. \quad (2.22)$$

$V_{\text{ap}}(\xi)$  is the product of the amplitude of the incoming wave  $V(\xi)$  with the aperture function  $A(\xi)$ , the latter incorporating the shape of the aperture in its modulus and optical aberrations in its phase.

If the aperture is illuminated by a plane wave,  $V(\xi) = V_0 \exp(ikz)$ , from a point source at a very large distance (e.g., an unresolved star), then the intensity distribution  $I(\alpha) = |V(\alpha)|^2$  in the focal plane is the diffraction limited image of the point source. In telescopes, the diffraction limited intensity distribution is called the Airy disk. For a circular aperture of diameter  $D$  the

Airy disk is given by

$$I(\alpha) = |V(\alpha)|^2 = \frac{V_0^2}{(\lambda F)^2} A_0^2 \text{Besinc}^2(k\alpha D/2). \quad (2.25)$$

Treating the optical system as a linear system, it is customary to introduce the dimensionless *point-spread function* (PSF) describing the form of the Airy disk. The PSF is the squared modulus of the Fourier transform of the telescope aperture  $A(\xi)$  (2.24) linked to the intensity distribution  $I(\alpha)$  in the focal plane through

$$\text{PSF}(\alpha) = \frac{I(\alpha)}{V_0^2} = \frac{1}{(\lambda F)^2} A_0^2 \text{Besinc}^2(k\alpha D/2). \quad (2.26)$$

The Rayleigh criterion of angular resolution defines the resolution limit of a telescope as the angular distance between the peak of the Airy disk (or the PSF) and its first minimum,  $\alpha_{\min} = 1.22 \lambda / D$ , with  $D$  the diameter of the telescope aperture. Stars that are separated by this distance can still be identified as individual objects, i.e., they can be resolved. Another resolution criteria is given by the full width at half maximum (FWHM) that is approximately given by  $\alpha_{\text{FWHM}} = \lambda / D$ . The latter is easier to measure in real astronomical images, the former slightly more conservative measure has the historical merit of having introduced diffraction theory into observations with astronomical telescopes.

## 2.3 The Coherence Function

The nature of coherence was discussed in text books on optics [19], on statistical optics [87], and, very specialised, in books on coherence theory [147, 148]. The coherence function was introduced as a measure for the coherence of the light. If the light has good coherence, the value of the coherence function is high, and interference phenomena like the fringes in Young's experiment show a good contrast.

We will introduce the concept of the coherence function more formally in this section, we will present its different flavours, and we will provide the necessary detail for observing with stellar interferometers. By and large, we will follow the formalism laid out in [87] and [148].

Starting with this section, two-dimensional integrals will be denoted by a single integral sign. The vector as integration variable shall indicate the integration over two dimensions.

### 2.3.1 Varieties of the Coherence Function

Before defining the coherence function, the nature of the random process leading to coherence and to incoherence needs to be discussed. So far, the optical disturbance  $v$  has been regarded as a deterministic signal throughout the propagation and diffraction process, forming a plane or a spherical wave. Now, approaching large celestial bodies emitting (mostly) thermal radiation, their light cannot be regarded as monochromatic and only approximately as a plane wave. Very close to the surface of a star, it is almost impossible to define a wave front and a direction of propagation. At a very large distance from the star, a point is a good approximation for its shape, and a plane wave describes the situation rather well.

In any moment during the propagation process, the optical disturbance, fed by light from individual, independently radiating points on the star, takes on random values that fluctuate typically at timescales  $1/\nu$ , the reciprocal of the average frequency. If one could take a series of snapshots with femto second exposure time, the pictures would all look different. However, these fluctuations average out over time intervals longer than  $1/\nu$ . Thus, snapshots with longer exposure time would all look similar.

We will regard the individual wave fronts as possible realizations or members of the ensemble of the random process, and the optical disturbance  $v(\mathbf{x}, t)$  as the random variable. Figure 2.10 illustrates the situation when the wave fronts, composed by values of the optical disturbance, are individual realisations of the random process.

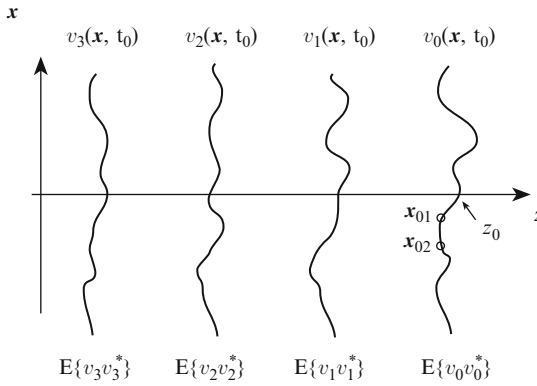
We will make two assumptions on the random process that will make our life much easier:

First, we assume that the random process is statistically stationary in time. This means that the statistical properties are the same over the ensemble, i.e., that the average is independent of the absolute moment in time  $t$  when it is taken and that the correlation only depends on the time difference  $t_1 - t_2$ .

Second, we assume that the statistics over one particular wave front at a given moment is the same as the statistics at a given point waiting a “long” time. In other words, the statistics, e.g., the average, over the complete wave front as an individual realization of the random process can be replaced by the average over many different realizations that appear in temporal succession. A “long” time providing the average over a sufficient number of realisations is defined as  $T \gg 1/\nu$ , i.e., much longer than the oscillation of the electromagnetic wave.

If these two conditions are met, the process is called *ergodic* and the ensemble average can be replaced by the time average. Figure 2.10 illustrates the situation.

Incidentally, the measurement with an optical detector provides a time average  $\langle v(\mathbf{x}, t)v^*(\mathbf{x}, t) \rangle$  (see (2.5)) that is sufficiently long to replace the ensemble average of the intensity. In Sect. 2.4 we will show how optical detectors – although only measuring the intensity – can be used to measure the coherence function, which is the correlation function of the optical disturbance as we shall see in the following discussion.



$$\text{Intensity: } I(\mathbf{x}_{01}) = \langle v_{z_0}(\mathbf{x}_{01}, t) v_{z_0}^*(\mathbf{x}_{01}, t) \rangle = E\{v_0(\mathbf{x}, t_0) v_0^*(\mathbf{x}, t_0)\}$$

Mutual Coherence Function (MCF):

$$\Gamma(\mathbf{x}_{01}, \mathbf{x}_{02}, \tau) = \langle v_{z_0}(\mathbf{x}_{01}, t + \tau) v_{z_0}^*(\mathbf{x}_{02}, t) \rangle = E\{v_0(\mathbf{x}_1, t_0 + \tau) v_0^*(\mathbf{x}_2, t_0)\}$$

**Fig. 2.10** Illustration of the propagation of light as an ergodic random process. The wave fronts of optical disturbances  $v_i$  are displayed at a given moment in time,  $t = t_0$ , and at different locations on the  $z$ -axis. The wave fronts were emitted at different times from an extended source at a large distance, having, thus, travelled different distances in the  $z$ -direction at  $t = t_0$ . Each wave front is an individual realisation of the random emission process. Due to the ergodicity of the random process the ensemble average  $E\{v_0(\mathbf{x}, t_0) v_0^*(\mathbf{x}, t_0)\}$  “across” the wave front for instance at  $z = z_0$  is identical to the time average  $\langle v_{z_0}(\mathbf{x}_{01}, t) v_{z_0}^*(\mathbf{x}_{01}, t) \rangle$  of different wave fronts passing through  $z = z_0$  if the integration time is sufficiently long, i.e.,  $T \gg 1/\nu$ . The correlation function of optical disturbances, the MCF  $\Gamma(\mathbf{x}_{01}, \mathbf{x}_{02}, \tau)$ , is computed as a time average  $\langle v_{z_0}(\mathbf{x}_{01}, t + \tau) v_{z_0}^*(\mathbf{x}_{02}, t) \rangle$ , (2.27), depending only on the time difference  $\tau$ . The ensemble average over pairs of points  $\mathbf{x}_1, \mathbf{x}_2$  across the wave front at  $z = z_0$  would yield the same result

Unlike the time averaged intensity, the time average of the optical disturbance  $\langle v(\mathbf{x}, t) \rangle$  and, thus, its ensemble average  $E\{v_i\}$  are zero since the optical disturbance fluctuates around zero.

### The Mutual Coherence Function (MCF)

The correlation function,<sup>4</sup> also called the *second order moment*, of the optical disturbances  $v$  at positions  $\mathbf{x}_1$  and  $\mathbf{x}_2$  in a plane at times  $t + \tau$  and  $t$  is called the *mutual coherence function* (MCF) [87, 148]:

$$\begin{aligned} \Gamma(\mathbf{x}_1, \mathbf{x}_2, \tau) &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T v(\mathbf{x}_1, t + \tau) v^*(\mathbf{x}_2, t) dt \\ &= \langle v(\mathbf{x}_1, t + \tau) v^*(\mathbf{x}_2, t) \rangle. \end{aligned} \quad (2.27)$$

<sup>4</sup> Here, the correlation function  $\langle v_1 v_2 \rangle_{\text{cor}}$  is identical to the covariance function  $\langle v_1 v_2 \rangle_{\text{cov}} = \langle v_1 v_2 \rangle_{\text{cor}} - \langle v \rangle^2$ , as the optical disturbance has zero mean  $\langle v \rangle = 0$ .

Remembering that the optical disturbance is proportional to one component of the electrical field vector, one could say that we determine the correlation of two electrical field vectors at two points in space (e.g., at two telescopes) and two moments in time.

The MCF is dimensionless. Depending on the optical disturbance  $v$ , the MCF can be complex with the phase denoted by  $\phi_{\text{MCF}}(\mathbf{x}_1, \mathbf{x}_2, \tau)$ . Using (2.27), it can easily be shown that the complex conjugate of the MCF is  $\Gamma^*(\mathbf{x}_1, \mathbf{x}_2, \tau) = \Gamma(\mathbf{x}_2, \mathbf{x}_1, -\tau)$ .

The time average over the period  $T$  at two fixed points  $\mathbf{x}_1$  and  $\mathbf{x}_2$  replaces the ensemble average  $E\{v(\mathbf{x}_1, t + \tau)v^*(\mathbf{x}_2, t)\}$  “across” the wave front at a fixed time  $t$  since the random process is ergodic as discussed above (see Fig. 2.10).

The MCF at two identical points at the same moment provides the time average of the product of the optical disturbances, which is the intensity  $I(\mathbf{x})$  as defined in (2.5):

$$I(\mathbf{x}) = \Gamma(\mathbf{x}, \mathbf{x}, 0) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T v(\mathbf{x}, t)v^*(\mathbf{x}, t) dt. \quad (2.28)$$

The MCF inherently varies for different pairs of points if the intensity at these points varies. In order to obtain the pure correlation of optical disturbances, i.e., the probability to measure the same value, the MCF has to be calibrated by the intensities at the individual coordinates  $\mathbf{x}_1$  and  $\mathbf{x}_2$ . We call this function the *degree of coherence* or the correlation coefficient of the optical disturbances:

$$\gamma(\mathbf{x}_1, \mathbf{x}_2, \tau) = \frac{\Gamma(\mathbf{x}_1, \mathbf{x}_2, \tau)}{\sqrt{\Gamma(\mathbf{x}_1, \mathbf{x}_1, 0) \Gamma(\mathbf{x}_2, \mathbf{x}_2, 0)}}, \quad (2.29)$$

which is in general a complex function with  $0 \leq |\gamma(\mathbf{x}_1, \mathbf{x}_2, \tau)| \leq 1$ .

The dependence of the MCF and of the degree of coherence on the time difference  $\tau$  and on the spatial coordinates  $\mathbf{x}_1$  and  $\mathbf{x}_2$  characterises two different aspects, the temporal and the spatial coherence of the light. In Sect. 2.1.2, we gave an introduction to temporal coherence. In this section, we will treat both cases under the umbrella of a general coherence theory.

## The Mutual Spectral Density Function (MSDF)

We introduced the spectral amplitude  $V(\mathbf{x}, \nu)$  of the harmonic wave in Sect. 2.1.1 as a time independent function so that the Helmholtz equation (2.14) could be used to describe the propagation of light. The spectral amplitude  $V(\mathbf{x}, \nu)$  is related to the optical disturbance  $v(\mathbf{x}, t)$  through a temporal Fourier transform (2.7).

Here, in the domain of coherence functions, we follow the same path. We define the *mutual spectral density function* (MSDF) [148], also called the *cross spectral*



*density function* [87], as the correlation function of the spectral amplitudes:

$$\hat{\Gamma}(\mathbf{x}_1, \mathbf{x}_2, \nu) := \lim_{T \rightarrow \infty} \frac{1}{2T} E \{ V_T(\mathbf{x}_1, \nu) V_T^*(\mathbf{x}_2, \nu) \}, \quad (2.30)$$

where the subscript  $T$  indicates that the spectral amplitude for the interval  $-T$  to  $T$  is taken. The dimension of the MSDF is  $\text{Hz}^{-1}$ .

Thus, the MSDF describes the spatial coherence at the frequency  $\nu$  as a correlation of spectral amplitudes between two individual points  $\mathbf{x}_1$  and  $\mathbf{x}_2$ .

The MCF is related to the MSDF through a Fourier transform,

$$\Gamma(\mathbf{x}_1, \mathbf{x}_2, \tau) = \int \hat{\Gamma}(\mathbf{x}_1, \mathbf{x}_2, \nu) e^{-i2\pi\nu\tau} d\nu, \quad (2.31)$$

keeping in mind that the MSDF being deduced from the spectral amplitude is zero for negative frequencies.

In general, the MSDF is a complex quantity and its phase will be denoted by  $\hat{\phi}(\mathbf{x}_1, \mathbf{x}_2, \nu)$ . The complex conjugate of the MSDF is given by  $\hat{\Gamma}^*(\mathbf{x}_1, \mathbf{x}_2, \nu) = \hat{\Gamma}(\mathbf{x}_2, \mathbf{x}_1, \nu)$ .

We will use the MSDF to describe the propagation of the coherence properties in space and through optical systems. This is in analogy to the scalar diffraction theory when the propagation of the spectral amplitude  $V$  was discussed.

The MSDF at two identical points defines the spectral intensity:

$$I(\mathbf{x}, \nu) := \hat{\Gamma}(\mathbf{x}, \mathbf{x}, \nu) = \lim_{T \rightarrow \infty} \frac{1}{2T} E \{ V(\mathbf{x}, \nu) V^*(\mathbf{x}, \nu) \}. \quad (2.32)$$

Using the connection between MCF and MSDF in (2.31) we find that the intensity  $I(\mathbf{x})$  can be written as

$$I(\mathbf{x}) = \Gamma(\mathbf{x}, \mathbf{x}, 0) = \int \hat{\Gamma}(\mathbf{x}, \mathbf{x}, \nu) d\nu = \int I(\mathbf{x}, \nu) d\nu, \quad (2.33)$$

which is the integral of the spectral intensity  $I(\mathbf{x}, \nu)$  over the frequency band. This expresses in the language of coherence functions what we found under very general assumptions in (2.8).

In the following, we will call the intensity  $I(\mathbf{x})$  the *polychromatic* or *white-light intensity* in order to clearly distinguish it from the spectral intensity  $I(\mathbf{x}, \nu)$ .

### The Self-Coherence Function

The MCF  $\Gamma(\mathbf{x}, \mathbf{x}, \tau)$  at two identical points and arbitrary  $\tau$  is called the *self-coherence function* [87, 148]. With (2.31) it can be written as the Fourier transform of the spectral intensity,

$$\Gamma(\mathbf{x}, \mathbf{x}, \tau) = \int \hat{I}(\mathbf{x}, \mathbf{x}, \nu) e^{-i2\pi\nu\tau} d\nu = \int I(\mathbf{x}, \nu) e^{-i2\pi\nu\tau} d\nu. \quad (2.34)$$

The self-coherence function is a measure for the temporal coherence of the light source. If the spectral intensity distribution  $I(\mathbf{x}, \nu)$  is very narrow with respect to  $\nu$ , i.e., if the spectral bandwidth  $\Delta\nu$  is very small, the self-coherence function  $\Gamma(\mathbf{x}, \mathbf{x}, \tau)$  is very wide with respect to the time difference  $\tau$ . This can be quantified by defining the *coherence time* as that value  $\tau_c$  when the self-coherence function is reduced significantly.

Describing, e.g.,  $I(\mathbf{x}, \nu)$  by a rectangular function with a width of  $\Delta\nu$ , one finds that  $\Gamma(\mathbf{x}, \mathbf{x}, \tau)$  as the Fourier transform of  $I(\mathbf{x}, \nu)$  has the form of a sinc-function, the sine function divided by its argument, with a width of about  $1/\Delta\nu$ . A suitable choice of the coherence time would be  $\tau_c = 1/\Delta\nu$  when the self-coherence function has its first zero.

At the end of Sect. 2.1.2 we discussed the contrast of the fringe pattern in Young's experiment for different spectral bandwidths of the illuminating light. We saw that the contrast as a function of diffraction angle, which is proportional to the difference in arrival time  $\tau$ , depends on the spectral bandwidth. The contrast goes down faster, i.e., at smaller  $\tau$ , with increasing bandwidth  $\Delta\nu$ .

Here, in the context of coherence functions we deduced the mathematical relationship between spectral bandwidth and the self-coherence function. Both the contrast and the self-coherence function show the same behaviour with respect to the spectral bandwidth. This is intuitively comprehensible since the contrast in Young's experiment depends on the correlation of the interfering amplitudes. In Sect. 2.4 this will be discussed in full detail.

Very often the term *coherence length*  $l_c$  is used to express the permitted optical path difference before interference phenomena disappear. The coherence length can be defined as

$$\begin{aligned} l_c &= c\tau_c \\ &\approx c/\Delta\nu = \lambda\nu/\Delta\nu = \lambda^2/\Delta\lambda \end{aligned} \quad (2.35)$$

As a numerical example we regard again the *K*-band,  $2.2 \pm 0.2 \mu\text{m}$ , with  $\Delta\lambda = 0.4 \mu\text{m}$  and a spectral bandwidth  $\Delta\nu = 0.25 \times 10^{14} \text{ Hz}$ . The coherence time is then approximately  $\tau_c = 4 \times 10^{-14} \text{ s}$  and the coherence length  $l_c = 12 \mu\text{m}$ .

Thus, the physical quantity governing the temporal coherence is the spectral width of the light source. In Sect. 2.3.4 we will see that the spatial coherence is determined by the angular size of the source, which is an entirely different physical property.

## Special Cases: Coherence and Incoherence

Perfect coherence means perfect correlation of the optical disturbances throughout time and space. The requirement for coherence is that the degree of coherence (2.29) is unity,  $|\gamma(\mathbf{x}_1, \mathbf{x}_2, \tau)| \equiv 1$ , for all combinations of  $\mathbf{x}_1$  and  $\mathbf{x}_2$  in a domain  $D$ , and

for arbitrary large time differences  $\tau$  [87, 148]. In terms of fringe contrast in Young's experiment (see Sect. 2.1.2) it is easy to comprehend that perfect coherence means a contrast of unity. We will see soon that the degree of coherence is a measure for the contrast.

The requirement for coherence is fulfilled for instance by a monochromatic plane wave with

$$v(\mathbf{x}, t) = \int V(\mathbf{x}, \nu) \delta(\nu - \nu_0) \exp(-i2\pi \nu t) d\nu = V(\mathbf{x}) \exp(-i2\pi \nu_0 t), \quad (2.36)$$

since using (2.27) we obtain

$$\begin{aligned} \Gamma(\mathbf{x}_1, \mathbf{x}_2, \tau) &= \langle v(\mathbf{x}_1, t + \tau) v^*(\mathbf{x}_2, t) \rangle \\ &= V(\mathbf{x}_1) V^*(\mathbf{x}_2) e^{-i2\pi \nu_0 \tau}, \end{aligned} \quad (2.37)$$

and

$$\gamma(\mathbf{x}_1, \mathbf{x}_2, \tau) = \frac{\Gamma(\mathbf{x}_1, \mathbf{x}_2, \tau)}{\sqrt{\Gamma(\mathbf{x}_1, \mathbf{x}_1, 0) \Gamma(\mathbf{x}_2, \mathbf{x}_2, 0)}} = \exp(-i2\pi \nu_0 \tau), \quad (2.38)$$

with a modulus equal to one. In general, all solutions  $V(\mathbf{x})$  of the Helmholtz equation (2.14) provide a perfectly coherent wave.

Strictly speaking, only monochromatic fields, having an infinitely narrow spectral line by definition, are perfectly coherent. This is not a very practical definition. The measurement of the MCF would have to last infinitely long in order to determine perfect coherence. For measurements of shorter duration a spectral line of finite width would deliver a result that is undistinguishable from perfect coherence. Laser light with a coherence length of many hundred metres is a typical example for light that is practically coherent.

We call light incoherent if the optical disturbances are perfectly uncorrelated even for infinitely small distances in space and time. Then, each point of the source radiates independently of its neighbour.

In terms of the coherence functions, the requirement for incoherence is  $|\Gamma(\mathbf{x}_1, \mathbf{x}_2, \tau)| \equiv 0$  unless  $\mathbf{x}_1 = \mathbf{x}_2$  and unless  $\tau = 0$  [87, 148]. In incoherent illumination there would be no fringes in Young's experiment. Their contrast like the value of the MCF would be zero.

Although the MCF is infinitely narrow with respect to the coordinate difference  $\mathbf{x}_1 - \mathbf{x}_2$ , the spatial extension of the intensity distribution with respect to  $\mathbf{x}$  can be very large like, e.g., the incoherently radiating surface of a star.

The Dirac  $\delta$ -function seems to be ideally suited to express the MCF of an incoherent source, writing

$$\Gamma(\mathbf{x}_1, \mathbf{x}_2, \tau) = \chi I(\mathbf{x}_1) \delta(\mathbf{x}_1 - \mathbf{x}_2) \delta(\tau), \quad (2.39)$$

where  $\chi$  is a constant.

However, an incoherent light source that is defined as above cannot radiate because either the value of the MCF and, thus, the intensity is infinitely large,  $\delta(\mathbf{x}_1 - \mathbf{x}_2) \rightarrow \infty$  for  $\mathbf{x}_1 = \mathbf{x}_2$  and  $\delta(\tau) \rightarrow \infty$  for  $\tau = 0$ , which is physically impossible. Or, if the values of the  $\delta$ -functions are reduced to finite values, the integral over the infinitely narrow functions is zero, setting the emitted intensity to zero [13, 87, 148].

Therefore, we look for a more practical definition of incoherence. Instead of the MCF, we will discuss the MSDF  $\hat{I}(\mathbf{x}_1, \mathbf{x}_2, \nu)$  that is related to the MCF through a Fourier transform (2.31). The MSDF as a function of  $\nu$  contains the spectrum of the light source. If the spectrum were infinitely wide, the Fourier transform of the MSDF would be infinitely narrow with respect to  $\tau$ . In practice, a very wide spectrum will be sufficient to establish incoherence if the absence of interference effects in any practical experimental configuration, e.g., zero contrast fringes in Young's experiment, is regarded as a sufficient criteria.

For the spatial characteristics, we replace  $\delta(\mathbf{x}_1 - \mathbf{x}_2)$  in (2.39) by a narrow function that has non-zero values not only for  $\mathbf{x}_1 = \mathbf{x}_2$  but for an interval of the order of the wavelength  $\lambda$  with  $|\mathbf{x}_1 - \mathbf{x}_2| \leq \mathcal{O}(\lambda)$ . For example, a Gaussian function with a width of  $\lambda/\sqrt{2\pi}$  fulfils this requirement. The narrow Gaussian function has the same effect on the coherence function at some distance from the source as a  $\delta$ -function if regarding only small angles.

A similar argument was used in Sect. 2.2.1 when discussing a diffracting pinhole with a diameter of the order of the wavelength  $\lambda$ . We found that a wave front passing through this pinhole can formally be replaced by a spherical wave emerging from the centre of the pinhole due to the restriction to small diffraction angles.

We will therefore keep the notation with the  $\delta$ -function and choose  $\chi = \lambda^2$  in order to obtain the same integral value for the  $\delta$ -function as for the Gaussian function:  $\int \chi \delta(\mathbf{x}) d\mathbf{x} = \chi = \int \exp(-|\mathbf{x}|^2/(\lambda^2/\pi)) d\mathbf{x}$ .

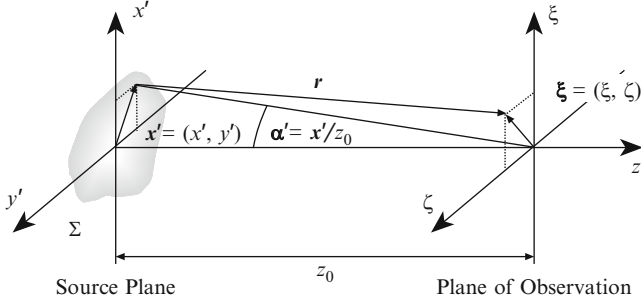
The MSDF of an incoherent source [27] now reads as

$$\hat{I}(\mathbf{x}_1, \mathbf{x}_2, \nu) = \lambda^2 I(\mathbf{x}_1, \nu) \delta(\mathbf{x}_1 - \mathbf{x}_2). \quad (2.40)$$

This definition of incoherence allows us to describe light that is spatially incoherent but that has an arbitrary spectrum. We can thus model for instance a spatially incoherent but monochromatic source. This makes it much easier to separate the effects of spatial and temporal coherence.

### 2.3.2 Generalised van Cittert–Zernike Theorem

The generalised van Cittert–Zernike Theorem [248, 259] describes how the statistical properties, the coherence, of light change when it propagates. In analogy to the scalar diffraction theory (see Sect. 2.2) using the spectral amplitude we discuss the propagation of the MSDF. In this section, we provide a formalism for the computation of the MSDF of the propagating light field in different planes along the path of propagation.



**Fig. 2.11** The geometry for the propagation of the coherence function of an incoherent source in plane  $\Sigma$ . The coordinate vector in the source plane is denoted by  $\mathbf{x}'$ , the plane of observation has the coordinate vector  $\boldsymbol{\xi}$ . The size of the source and the area of interest in the plane of observation are much smaller than the distance  $z_0$  allowing for small angle approximations,  $\alpha' = \mathbf{x}'/z_0$ . In a stellar interferometer the source would be a star or another celestial body, and the plane of observation would be the aperture plane of an interferometric array

The propagation of the MSDF can be derived in the same way as the Rayleigh–Sommerfeld diffraction formula (2.16) yielding [87, 148] the MSDF in the plane of observation with coordinate vector  $\boldsymbol{\xi}$  (see Fig. 2.11) as a function of the MSDF in the source plane  $\Sigma$  as

$$\hat{F}(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \nu) = \frac{1}{(\lambda z_0)^2} \iint_{\Sigma} \hat{F}(\mathbf{x}'_1, \mathbf{x}'_2, \nu) e^{ik(r_1 - r_2)} d\mathbf{x}'_1 d\mathbf{x}'_2. \quad (2.41)$$

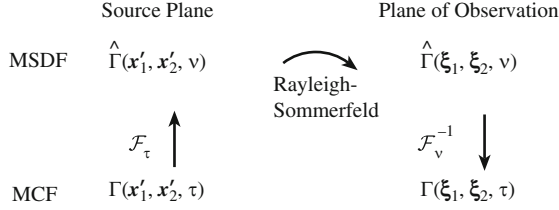
This is the basic equation for the propagation of the MSDF in space. It describes the propagation essentially as the product of two Rayleigh–Sommerfeld integrals (2.16) at individual points  $\boldsymbol{\xi}_1$  and  $\boldsymbol{\xi}_2$ . Equation (2.41) is valid if the angles involved are small, i.e., if the diameter of the light source and the observed area in the plane of observation are much smaller than the distance between them.

The argument in the exponential can be simplified by using the Fresnel approximation (2.17):

$$r = z_0 + \frac{|\mathbf{x}' - \boldsymbol{\xi}|^2}{2z_0} - \dots, \text{ and} \\ r_1 - r_2 = \frac{|\boldsymbol{\xi}_1|^2 - |\boldsymbol{\xi}_2|^2}{2z_0} - \frac{\boldsymbol{\xi}_1 \cdot \mathbf{x}'_1 - \boldsymbol{\xi}_2 \cdot \mathbf{x}'_2}{z_0} + \frac{|\mathbf{x}'_1|^2 - |\mathbf{x}'_2|^2}{2z_0} - \dots$$

Then, the propagation of the MSDF in Fresnel approximation is calculated by

$$\hat{F}(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \nu) = \frac{1}{(\lambda z_0)^2} e^{ik(|\boldsymbol{\xi}_1|^2 - |\boldsymbol{\xi}_2|^2)/(2z_0)} \times \iint \hat{F}(\mathbf{x}'_1, \mathbf{x}'_2, \nu) e^{ik(|\mathbf{x}'_1|^2 - |\mathbf{x}'_2|^2)/(2z_0)} e^{-ik(\boldsymbol{\xi}_1 \cdot \mathbf{x}'_1 - \boldsymbol{\xi}_2 \cdot \mathbf{x}'_2)/z_0} d\mathbf{x}'_1 d\mathbf{x}'_2. \quad (2.42)$$



**Fig. 2.12** The principle of the generalized van Cittert–Zernike theorem. The MCF  $\Gamma(\mathbf{x}'_1, \mathbf{x}'_2, \tau)$  in the source plane is Fourier transformed yielding the MSDF  $\hat{\Gamma}(\mathbf{x}'_1, \mathbf{x}'_2, \nu)$ . Applying (2.41) yields the MSDF  $\hat{\Gamma}(\xi_1, \xi_2, \nu)$  in the plane of observation and a Fourier back-transform provides the MCF  $\Gamma(\xi_1, \xi_2, \tau)$  in the plane of observation

This formula describes the propagation between two planes that are separated by a sufficiently long distance  $z_0$  to justify the Fresnel approximation.

Treating the general case of polychromatic light requires discussing the mutual coherence function, starting with the source MCF  $\Gamma(\mathbf{x}'_1, \mathbf{x}'_2, \tau)$ . Using the relationship between the MCF and the MSDF (2.31), the first step is to calculate the MSDF  $\hat{\Gamma}(\mathbf{x}'_1, \mathbf{x}'_2, \nu)$  in the source plane, and then to apply (2.41) describing the propagation of the MSDF into the plane of observation. There, the MCF  $\Gamma(\xi_1, \xi_2, \tau)$  can be computed by Fourier transforming the MSDF  $\hat{\Gamma}(\xi_1, \xi_2, \nu)$  in (2.41). The MCF completely describes the coherence properties of the polychromatic light in the plane of observation. This process is visualised in Fig. 2.12. It is called the *generalized van Cittert–Zernike theorem* [87, 148].

In the following, we will see how the situation can be simplified by reducing the discussion to an incoherent source.

### 2.3.3 Incoherent Sources of Light: Stars

Observing celestial bodies, we will use  $\alpha' = \mathbf{x}'/z_0$ , the angle of observation, as a source coordinate (see Fig. 2.11). We also reduce our discussion to spatially incoherent sources since stars as thermal sources are spatially incoherent. This means that every point on the surface of the star radiates independently of its neighbour point as discussed in Sect. 2.3.1. This applies to all celestial bodies. In its simplest form a star is shaped like a disk with a diameter independent of wavelength.

If the shape of the source is independent of wavelength over the observed spectrum, the spectral intensity  $I(\alpha', \nu)$  can be split into the product of *source brightness distribution*,  $I_b(\alpha')$ , and *source spectrum*,  $G(\nu)$ :

$$I(\alpha', \nu) = I_b(\alpha')G(\nu). \quad (2.43)$$

$I_b(\alpha')$  is dimensionless and describes the shape of the source intensity assumed to be independent of wavelength.  $G(\nu)$  has the dimension  $\text{Hz}^{-1}$  and, unless stated otherwise, it is calibrated to unity  $\int G(\nu)d\nu = 1$ .

For ground-based observations the width of the spectrum is limited either by the width of the atmospheric bands like the  $K$ -band or by spectral filters in the astronomical instruments.

Similar to (2.40), the MSDF of the star in the source plane in angular coordinates now can be written as

$$\hat{\Gamma}(\alpha'_1, \alpha'_2, \nu) = \frac{\lambda^2}{z_0^2} I_b(\alpha'_1) \delta(\alpha'_1 - \alpha'_2) G(\nu). \quad (2.44)$$

Due to the change of variables, the factor  $1/z_0^2$  comes in, keeping in mind the discussion at the end of Sect. 2.3.1.

Whenever we speak about incoherent sources it is in this definition of a spatially incoherent source with arbitrary spectrum  $G(\nu)$ .

The expression for the MSDF of an incoherent source (2.44) is inserted into (2.42) replacing  $\mathbf{x}'_i/z_0$  by  $\alpha'_i$  and  $d\mathbf{x}'_i$  by  $z_0^2 d\alpha'_i$ . The quadratic exponential before the integral in (2.42) describes the planes of equal phase, similar to the wave front when describing the propagation of the spectral amplitude (2.18). In stellar interferometers, the source is usually a star at a fairly large distance  $z_0$ , and the plane of observation is on the surface of the Earth or, for space interferometers, in a near Earth orbit. Then the quadratic exponential can be approximated by unity, yielding the MSDF in the plane of observation

$$\begin{aligned} \hat{\Gamma}(\xi_1, \xi_2, \nu) &= G(\nu) \\ &\times \iint I_b(\alpha'_1) \delta(\alpha'_1 - \alpha'_2) e^{ik(|\alpha'_1|^2 - |\alpha'_2|^2)z_0/2} e^{-ik(\xi_1 \cdot \alpha'_1 - \xi_2 \cdot \alpha'_2)} d\alpha'_1 d\alpha'_2 \\ &= G(\nu) \int I_b(\alpha'_1) e^{-ik(\xi_1 - \xi_2) \cdot \alpha'_1} d\alpha'_1, \end{aligned} \quad (2.45)$$

which is the product of the spectrum  $G(\nu)$  with the Fourier transform of the source brightness distribution  $I_b(\alpha')$ . Thus, the MSDF at frequency  $\nu$ , i.e., the spatial coherence in the plane of observation, is determined by the Fourier transform of the source shape.

With the exponential kernel  $\frac{2\pi}{\lambda}(\xi_1 - \xi_2) \cdot \alpha'_1$ , the Fourier transform is performed from  $\alpha'$  to  $(\xi_1 - \xi_2)/\lambda$ . This means firstly that the Fourier transform does not depend on the individual coordinates  $\xi_1$  and  $\xi_2$  but on their difference only, and secondly that the Fourier transform is a function of wavelength,  $1/\lambda$ .

The consequence of working with an incoherent source is, thus, that the MSDF is a function of coordinate difference only. Therefore, we will write the MSDF and the MCF in the plane of observation<sup>5</sup> as functions of coordinate difference in the following:

<sup>5</sup> More generally speaking, the coherence function depends on the coordinate difference in those planes that are illuminated by an incoherent source at large distance when the van Cittert–Zernike theorem applies.

$$\hat{\Gamma}(\xi_1 - \xi_2, \nu) = \hat{\Gamma}(\xi_1, \xi_2, \nu) \text{ and } \Gamma(\xi_1 - \xi_2, \tau) = \Gamma(\xi_1, \xi_2, \tau). \quad (2.46)$$

It is also important to note that the quadratic phase term  $\exp(ik(|\alpha'_1|^2 - |\alpha'_2|^2)z_0/2)$  in (2.45) that stems from the Fresnel approximation disappears only because the light of the celestial source is spatially incoherent. There is no further approximation required.

The spectral intensity at frequency  $\nu$  is the MSDF for  $\xi_1 = \xi_2$ , i.e.,  $I(\xi, \nu) = \hat{\Gamma}(\xi - \xi, \nu) = \hat{\Gamma}(0, \nu)$ . Then, the Fourier transform in (2.45) is reduced to a simple integration  $I(\xi, \nu) = \hat{\Gamma}(0, \nu) = G(\nu) \int I_b(\alpha') d\alpha' = G(\nu) I_0$  yielding the constant spectral intensity  $G(\nu) I_0$  in the plane of observation.

In general, we are interested in **polychromatic light**. To determine the white-light intensity  $I(\xi)$  in the plane of observation, we integrate the spectral intensity  $I(\xi, \nu) = G(\nu) I_0$  over the spectral band (see (2.33)) yielding the constant value  $I_0$  since the integral over  $G(\nu)$  is unity. Thus, not very surprising, the star sheds its light homogeneously over the surface of the Earth. Only the coherence functions reflect the properties – in particular the shape – of the source.

When it comes to coherence properties of polychromatic light we have to deal with the MCF in the plane of observation. The relationship between the MSDF and the MCF is defined by (2.31). However, this is not completely straightforward since a complicated integral with two entangled Fourier transforms needs to be resolved:

$$\begin{aligned} \Gamma(\xi_1 - \xi_2, \tau) &= \int \hat{\Gamma}(\xi_1 - \xi_2, \nu) e^{-i2\pi\nu\tau} d\nu \\ &= \int G(\nu) \int I_b(\alpha') e^{-ik(\xi_1 - \xi_2) \cdot \alpha'} d\alpha' e^{-i2\pi\nu\tau} d\nu. \end{aligned} \quad (2.47)$$

A common simplification of this double integral is based on the assumption that the source brightness distribution  $I_b(\alpha')$  is independent of wavelength. This is, however, not sufficient for splitting the formula into an integral over  $\nu$  and an integral over  $\alpha'$  since – bearing in mind that the Fourier transform is done from  $\alpha'$  to  $(\xi_1 - \xi_2)/\lambda$  – the Fourier transform of  $I_b(\alpha')$  is still a function of wavelength.

One has to assume additionally, that also the Fourier transform of  $I_b(\alpha')$  is invariant over the frequency range set by the spectrum. Then the MCF can be split into the product of the Fourier transform of the source brightness distribution  $\int I_b(\alpha') \exp(-ik(\xi_1 - \xi_2) \cdot \alpha') d\alpha'$  and the Fourier transform of the spectrum  $\int G(\nu) \exp(-i2\pi\nu\tau) d\nu$ . While the first Fourier transform can be regarded as the purely spatial part of the MCF, the second Fourier transform is the temporal part of the MCF like the self-coherence function.

If the source were approximately point-like, described by a  $\delta$ -function, then its Fourier transform would be a constant,  $I_0$ , independent of the coordinate difference  $\xi_1 - \xi_2$  and of  $\lambda$ . In this case, the MSDF in the plane of observation reads as

$$\hat{\Gamma}(\xi_1 - \xi_2, \nu) = G(\nu) I_0. \quad (2.48)$$



For a source of finite size,  $\alpha_0$ , the Fourier transform would be approximately constant only either for very small coordinate differences  $\xi_1 - \xi_2$  or over a very narrow spectrum. Writing the coordinate in Fourier space as  $(\xi_1 - \xi_2)/\lambda$  we can also state that the Fourier transform is invariant over any coordinate range smaller than  $1/\alpha_0$ . This puts a constraint on the size of the source. For a larger source or for a wider spectrum the Fourier transform would vary over the observed spectrum, and the separation of the two integrals and the approximation no longer holds.

This approximation restricting the width of the spectrum and, implicitly, the size of the source is also known as the *quasi-monochromatic approximation*.

### 2.3.4 Quasi-Monochromatic Approximation

We simplify the double integral for the propagation of the MCF (2.47) to a manageable formula by assuming a narrow spectrum  $G(\nu)$  with  $\Delta\nu \ll \nu_0$  and by observing only time differences shorter than the coherence time,  $\tau \ll 1/\Delta\nu$  [148]. The latter means that the corresponding optical path difference (OPD) must be much smaller than the coherence length  $l_c$ .

A narrow spectrum in connection with a small source means that, as argued above, the Fourier transform of  $I_b(\alpha')$  is invariant over the spectrum and the Fourier transform is represented by the value at the average frequency  $\nu_0$  (and at the average wave number  $k_0$ ). The restriction to small time differences,  $\tau \ll 1/\Delta\nu$ , means that the second integral over  $\nu$  is reduced to the function value  $G(\nu_0) \exp(-i2\pi\nu_0\tau)\Delta\nu = \exp(-i2\pi\nu_0\tau)$  due to the calibration of  $G(\nu)$ .

Now the MCF in the plane of observation can be written in quasi-monochromatic approximation as

$$\begin{aligned} \Gamma_{\text{qm}}(\xi_1 - \xi_2, \tau) &= \int I_b(\alpha') e^{-ik_0(\xi_1 - \xi_2) \cdot \alpha'} d\alpha' e^{-i2\pi\nu_0\tau} \\ &= \Gamma_{\text{qm}}(\xi_1 - \xi_2, 0) e^{-i2\pi\nu_0\tau}. \end{aligned} \quad (2.49)$$

Not very surprisingly, the MCF in quasi-monochromatic approximation (2.49) has some similarity to the MCF of coherent light (2.37). While in the coherent case the product of two deterministic amplitudes  $V_{\nu_0}(\xi_1) V_{\nu_0}^*(\xi_2) \exp(-2\pi\nu_0\tau)$  formed the MCF for any time delay  $\tau$ , the quasi-monochromatic MCF is valid for small time delays  $\tau$  only.

However, unlike in the case of coherent light, there is not a product of deterministic amplitudes in quasi-monochromatic approximation but a correlation function  $\Gamma_{\text{qm}}(\xi_1 - \xi_2, 0)$  that is determined by the Fourier transform of the source shape.

With the MCF at  $\tau = 0$  we can now write down the fundamental *van Cittert-Zernike theorem* determining the spatial coherence in the plane of observation as

$$\Gamma_{\text{qm}}(\xi_1 - \xi_2, 0) = \int I_b(\alpha') e^{-ik_0(\xi_1 - \xi_2) \cdot \alpha'} d\alpha'. \quad (2.50)$$

The brightness distribution  $I_b(\boldsymbol{\alpha}')$  of an incoherent source is linked to the MCF  $\Gamma_{\text{qm}}(\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2, 0)$  in the plane of observation at  $\tau = 0$  through a Fourier transform. The spatial coherence depends on the shape of the intensity distribution of the source, and it is a function of the coordinate difference. The size of the source has to be sufficiently small so that its Fourier transform is invariant over the spectral band  $\Delta\nu$  and can be replaced by its value at the average frequency  $\nu_0$  and  $k_0$  respectively.

Dividing the MCF  $\Gamma_{\text{qm}}(\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2, 0)$  by the geometric mean of the intensities  $I(\boldsymbol{\xi}_i) = \Gamma_{\text{qm}}(\boldsymbol{\xi}_i - \boldsymbol{\xi}_i, 0)$  at position  $\boldsymbol{\xi}_1$  and  $\boldsymbol{\xi}_2$ , we obtain the normalised MCF, which is the degree of coherence,  $\gamma$ , (2.29) at  $\tau = 0$ . In quasi-monochromatic approximation, this is called the complex *visibility function*  $\mu_{\nu_0}(\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2)$ .

Then, the van Cittert–Zernike theorem can be written in its most common form as:

$$\mu_{\nu_0}(\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2) = \frac{\Gamma_{\text{qm}}(\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2, 0)}{\sqrt{I(\boldsymbol{\xi}_1)I(\boldsymbol{\xi}_2)}} = \frac{\int I_b(\boldsymbol{\alpha}') e^{-ik_0(\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2) \cdot \boldsymbol{\alpha}'} d\boldsymbol{\alpha}'}{\int I_b(\boldsymbol{\alpha}') d\boldsymbol{\alpha}'} \quad (2.51)$$

so that  $0 \leq |\mu_{\nu_0}(\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2)| \leq 1$  and  $\mu_{\nu_0}(0) = 1$ . The normalisation by the geometric mean of the intensities  $I(\boldsymbol{\xi}_i)$  is replaced by the integral over  $I_b(\boldsymbol{\alpha}')$  that has the constant value  $I_0 = \int I_b(\boldsymbol{\alpha}') d\boldsymbol{\alpha}'$ . However, observing with a real interferometer through atmospheric turbulence the intensity fluctuates randomly and the values at  $\boldsymbol{\xi}_1$  and  $\boldsymbol{\xi}_2$  are no longer constant and have to be accounted for individually. This calibration is a very important part of the interferometric data processing.

The phase of the visibility function will be denoted by  $\phi_{\nu_0}(\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2)$ , which is identical to  $\phi_{\text{MCF}}(\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2, 0)$ , the phase of the MCF at  $\tau = 0$ .

The definition of the visibility function in (2.51) can be used to formally rewrite the MSDF in (2.45), as

$$\hat{\Gamma}(\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2, \nu) = G(\nu) I_0 \mu_\nu(\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2), \quad (2.52)$$

at any frequency  $\nu$ . The phase,  $\hat{\phi}(\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2, \nu)$ , of the MSDF is then identical to the phase,  $\phi_\nu(\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2)$ , of the visibility function since  $I_0$  and the spectrum  $G(\nu)$  are real functions.

Thus, the MSDF is composed of the spectrum  $G(\nu)$  determining the temporal coherence, and of the visibility function  $\mu_\nu(\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2)$  describing the spatial coherence. This facilitates the interpretation of the MSDF when discussing polychromatic illumination in the following sections. Because of the exponential kernel,  $\frac{2\pi}{\lambda}(\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2) \cdot \boldsymbol{\alpha}'$ , of the Fourier transform,  $\mu_\nu$  varies with wavelength (indicated by subscript  $\nu$ ) although the source brightness distribution  $I_b(\boldsymbol{\alpha}')$  is independent of wavelength. In quasi-monochromatic approximation, regarding only the average frequency  $\nu_0$ , we use the subscript  $\nu_0$ .

Similar to the coherence time that was defined in the context of the self-coherence function in Sect. 2.3.1 as the reciprocal of the spectral bandwidth (2.34), we can now define the *coherence width* as that coordinate difference  $\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2$  when the value of the visibility function is reduced significantly. Due to the Fourier connection

between the visibility function and the source intensity distribution (2.51), the coherence width is inversely proportional to the angular size of the source.

The symmetry characteristics of the visibility function are driven by the fact that the source brightness distribution is a real and positive function by definition. Then, its Fourier transform is a Hermitian function (see Sect. A.1), meaning that the modulus of the visibility function is symmetric with respect to  $|\xi_1 - \xi_2| = 0$  and that the phase  $\phi_v$  of the visibility function is anti-symmetric:

$$|\mu_v(\xi_1 - \xi_2)| = |\mu_v(\xi_2 - \xi_1)| \text{ and} \\ \phi_v(\xi_1 - \xi_2) = -\phi_v(\xi_2 - \xi_1).$$

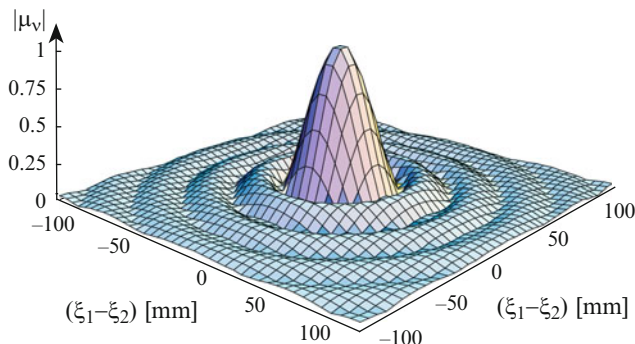
How does the propagation of the coherence function described by the van Cittert–Zernike theorem (2.50) compare to the propagation of the spectral amplitude as discussed in Sect. 2.2.2? Both the Rayleigh–Sommerfeld diffraction integral in Fresnel approximation (2.22) and the van Cittert–Zernike theorem link physical properties in two separate planes through a Fourier transform. While the Rayleigh–Sommerfeld diffraction integral connects the spectral amplitude in the plane of observation with the spectral amplitude in an aperture – which are the same physical quantities – the van Cittert–Zernike theorem links the intensity distribution of an incoherent source with the visibility function in the plane of observation, which are seemingly different physical quantities. However, this should be well understood by now since the intensity is a special case of the coherence function, which, as a product of two optical disturbances, has the dimension of an intensity.

Both formulae are valid in Fresnel approximation and quadratic phase terms cancel in both cases albeit for different reasons: The diffraction integral is reduced to a Fourier transform by introducing a lens (2.21), and the van Cittert–Zernike theorem uses a Fourier transform because the light source is spatially incoherent.

### The Visibility Function of Venus

An example illuminates the situation: we model Venus as a uniform disk with angular diameter  $\alpha'_0$  varying between about 15 and 45 arcsec depending on the mutual positions of Venus and Earth. The circular source brightness distribution is then represented by  $I_b(\alpha') = (\pi(\alpha'_0/2)^2)^{-1} \text{circ}\left(\frac{|\alpha'|}{\alpha'_0/2}\right)$ . The factor  $(\pi(\alpha'_0/2)^2)^{-1}$  is a normalization factor so that the integral of the intensity over the source and, thus, the denominator in (2.51), equals one, and we compute the visibility function as the Fourier transform of  $I_b(\alpha')$  yielding

$$\begin{aligned} \mu_v(\xi_1 - \xi_2) &= \frac{1}{\pi(\alpha'_0/2)^2} \int \text{circ}\left(\frac{|\alpha'|}{\alpha'_0/2}\right) e^{-ik(\xi_1 - \xi_2) \cdot \alpha'} d\alpha' \\ &= \text{Besinc}(k|\xi_1 - \xi_2|\alpha'_0/2). \end{aligned} \quad (2.53)$$



**Fig. 2.13** The modulus of the visibility function  $|\mu_v(\xi_1 - \xi_2)|$  of Venus in the plane of observation, e.g., on the surface of the Earth. The visibility function has the form of a Bessel-function since Venus is a light source at a very large distance, modelled as a uniform disk (2.53). If Venus has an angular diameter of  $\alpha'_0 = 15$  arcsec the visibility function has its first zero for a coordinate difference of  $|\xi_1 - \xi_2|_0 = 37$  mm with  $|\xi_1 - \xi_2|_0 = 1.22\lambda/\alpha'_0$  and for an observing wavelength of  $2.2 \mu\text{m}$  (in the visible, it is  $|\xi_1 - \xi_2|_0 = 8$  mm). The second zero of the visibility function is at  $|\xi_1 - \xi_2|_0 = 2.233\lambda/\alpha'_0 = 67.5$  mm. If Venus were closer to the Earth displaying an angular diameter of e.g. 45 arcsec, the coherence width would be reduced to 12 mm

Figure 2.13 displays the modulus of the visibility function for Venus' smallest angular diameter of  $\alpha'_0 = 15$  arcsec as a function of the coordinate difference in mm. A suitable measure for the coherence width is the first zero of the visibility function, which occurs at a coordinate difference of  $|\xi_1 - \xi_2|_0 = 1.22\lambda/\alpha'_0 = 37$  mm for an observing wavelength of  $\lambda = 2.2 \mu\text{m}$ . For larger coordinate differences the visibility function slowly oscillates between negative and positive values with decreasing amplitude.

Applying the quasi-monochromatic approximation we would assume that the visibility function at the average frequency  $\nu_0$  is a suitable approximation for the visibility function at all frequencies within the spectral band. If the spectral band is too wide or if the size of the source is too large this approximation no longer holds and we have to compute the integral over the spectral band in (2.47) by properly considering the visibility function as a function of wavelength.

The angular diameter of stars is typically in the milliarcsecond range, i.e., a factor of 1,000 smaller than Venus. If a star is modelled as a uniform disk the shape of the visibility function is again described by a Bessel-function but the coherence width has values of several 10 m and not millimetres. Separating the two points  $\xi_1$  and  $\xi_2$  by only a few metres, the light coming from the star still has a very high coherence indicated by values of the visibility function that are close to unity.

Another way of putting this is to say that the incoherent light emitted by the star formally has acquired coherence by the very process of propagation. While the coherence function on the surface of the incoherent star (2.44) is zero unless the coordinate difference is zero, the increased coherence width of the light at a large distance from the star allows for a coordinate difference of a few metres before the coherence function in the plane of observation drops substantially.

In contrast, we will see later that light from a coherent source does not change its state of coherence when propagating in free space. Coherent light remains coherent throughout space and time. Thus, the visibility function of a coherent source is always unity independent of the coordinate difference.

For these examples of uniform disks the visibility function as a Besinc-function has the same mathematical form as the spectral amplitude in the diffraction pattern of a circular aperture (2.24). While the diffraction pattern can be observed with the naked eye, the visibility function describing the correlation of optical disturbances at different points cannot be observed directly. It can only be observed through an interferometer experiment.

Revisiting Young's experiment in the following section we will see how the contrast of a fringe pattern as a measurable quantity can be related to the visibility function.

**NB 4.** *Expanding the scope of our discussion to sources of arbitrary spatial coherence we replace the  $\delta$ -function in (2.44) by a function of finite extent describing the source's spatial coherence. Describing the spatial coherence by the visibility function  $\mu(\alpha'_1 - \alpha'_2)$  – without the subscript  $v$  since we are free to define the source's spatial coherence independent of wavelength – we write the MSDF of a partially coherent source as*

$$\hat{\Gamma}(\alpha'_1, \alpha'_2, v) = \frac{\lambda}{z_0^2} \sqrt{I_b(\alpha'_1)} \sqrt{I_b(\alpha'_2)} \mu(\alpha'_1 - \alpha'_2) G(v). \quad (2.54)$$

*Then the coherence width of the source is no longer zero but it is determined by the shape of the visibility function  $\mu$ .*

*In this notation, the visibility function is linked to the degree of coherence  $\gamma(\alpha'_1 - \alpha'_2, \tau)$ , using (2.29) and (2.31), through*

$$\gamma(\alpha'_1 - \alpha'_2, \tau) = \int \mu(\alpha'_1 - \alpha'_2) G(v) e^{-i2\pi v \tau} dv, \quad (2.55)$$

*like the MCF to the MSDF. This is why the visibility function is sometimes called the spatial degree of coherence.*

*Choosing a Gaussian function for both the visibility function  $\mu(\alpha'_1 - \alpha'_2)$  with a width of  $\sigma_\mu$  and for the brightness distribution  $I_b(\alpha')$  with a width of  $\sigma_{\alpha'}$ , we introduce Gaussian Schell-model (GSM) sources [147, Sect. 5.3 and 5.4] with an MSDF defined as*

$$\hat{\Gamma}(\alpha'_1, \alpha'_2, v) = \frac{\lambda}{z_0^2} \exp\left(-\frac{|\alpha'_1 + \alpha'_2|^2}{8\sigma_{\alpha'}^2}\right) \exp\left(-\frac{|\alpha'_1 - \alpha'_2|^2}{2\sigma_\mu^2}\right) G(v), \quad (2.56)$$

*with the condition  $\sigma_\mu \ll \sigma_{\alpha'}$ , i.e., the coherence width is much smaller than the size of the source.*

*The propagation of light from the GSM source is described by the van Cittert–Zernike theorem replacing the  $\delta$ -function in (2.45) by  $\mu(\alpha'_1 - \alpha'_2)$ . The spatial*

coherence in the plane of observation in Fraunhofer approximation is again described by the Fourier transform of the intensity distribution in the source plane that – being a Gaussian function – is itself a Gaussian function of width  $1/(2\pi\sigma_{\alpha'})$ . The coherence width is thus proportional to the reciprocal of the angular source diameter,  $\sigma_{\alpha'}$ , just like for an incoherent source.

It was shown by Carter and Wolf [27], that the intensity distribution  $I(\xi)$  in the plane of observation, which was constant in the case of an incoherent source, is the Fourier transform of the source's visibility function  $\mu(\alpha'_1 - \alpha'_2)$ . Then,  $I(\xi)$  is a Gaussian function of width  $1/(2\pi\sigma_{\mu})$ . In the incoherent limit when the coherence width in the source plane approaches zero,  $\sigma_{\mu} \rightarrow 0$ , the width of the intensity distribution steadily increases until it fills the plane of observation homogeneously.

Thus, the source visibility function determines the intensity distribution in the plane of observation and, vice versa, the source intensity distribution determines the visibility function in the plane of observation. This reciprocity theorem was first pointed out by Walther [251].

In practice, GSM sources can be used to describe the propagation of multi-mode laser light [89, 217, 240]. Observing celestial bodies we rarely come across multi-mode laser light but we do encounter partially coherent light during the imaging process. In the preceding section, discussing the visibility function of Venus we found that the coherence width in the plane of observation, e.g., on the surface of the Earth, is some 10 mm. The light is then partially coherent. The aperture of a ground-based telescope is filled by this partially coherent light and, computing the propagation of light from the aperture plane into the image plane, we treat a case similar to the one above. In Sect. 3.2.3, discussing the coherence properties in the image plane, we will return to this subject.

### The Coherence Function: Summary

The mutual coherence function is the critical quantity for understanding stellar interferometry. It is the second order correlation function of the optical disturbance as a function of time difference and spatial coordinates. The mutual coherence function (MCF) is defined as

$$\Gamma(\mathbf{x}_1, \mathbf{x}_2, \tau) = \langle v(\mathbf{x}_1, t + \tau) v^*(\mathbf{x}_2, t) \rangle, \quad (2.27)$$

where  $\mathbf{x}_i$  are the coordinate vectors and  $\tau$  is the time difference.

For the propagation of the MCF in space it is convenient to introduce the mutual spectral density function (MSDF) that is the correlation function of spectral amplitudes, defined as

$$\hat{\Gamma}(\mathbf{x}_1, \mathbf{x}_2, \nu) := \lim_{T \rightarrow \infty} \frac{1}{2T} E \{ V_T(\mathbf{x}_1, \nu) V_T^*(\mathbf{x}_2, \nu) \}. \quad (2.30)$$

The MSDF is related to the MCF through a Fourier transform,

$$\Gamma(\mathbf{x}_1, \mathbf{x}_2, \tau) = \int \hat{\Gamma}(\mathbf{x}_1, \mathbf{x}_2, \nu) e^{-i2\pi\nu\tau} d\nu. \quad (2.31)$$

The polychromatic or white-light intensity is then

$$I(\mathbf{x}) = \Gamma(\mathbf{x}, \mathbf{x}, 0) = \int \hat{\Gamma}(\mathbf{x}, \mathbf{x}, \nu) d\nu = \int I(\mathbf{x}, \nu) d\nu, \quad (2.33)$$

with  $I(\mathbf{x}, \nu) = \hat{\Gamma}(\mathbf{x}, \mathbf{x}, \nu)$ , the spectral intensity.

The propagation of the coherence functions in space is described by applying the generalised van Cittert–Zernike theorem (2.41) that was derived from the Rayleigh–Sommerfeld diffraction formula (2.16) yielding the MSDF in the plane of observation with coordinate  $\xi$  as the Fourier transform of the source brightness distribution  $I_b(\alpha')$ :

$$\hat{\Gamma}(\xi_1, \xi_2, \nu) = G(\nu) \int I_b(\alpha') e^{-ik(\xi_1 - \xi_2) \cdot \alpha'} d\alpha', \quad (2.45)$$

with  $\alpha' = \mathbf{x}'/z_0$ , and  $G(\nu)$  the source spectrum. The source brightness distribution is assumed independent of wavelength.

For simplification, a number of assumptions and approximations are made, which are suitable for the situation in a stellar interferometer:

- Fresnel approximation (all involved angles are small)
- Spatially incoherent light sources (e.g., stars)
- The quasi-monochromatic approximation when
  - (1) The spectral bandwidth  $\Delta\nu$  is assumed to be much smaller than the average frequency  $\nu_0$
  - (2) The time difference  $\tau$  is much smaller than  $1/\Delta\nu$  corresponding to an OPD much smaller than the coherence length  $l_c$

With the assumption of an incoherent light source the coherence functions in the plane of observation can be written as functions of coordinate difference only

$$\hat{\Gamma}(\xi_1 - \xi_2, \nu) = \hat{\Gamma}(\xi_1, \xi_2, \nu) \text{ and } \Gamma(\xi_1 - \xi_2, \tau) = \Gamma(\xi_1, \xi_2, \tau).$$

The van Cittert–Zernike theorem (2.50) in quasi-monochromatic approximation describes the propagation of the mutual coherence function under these conditions stating that the MCF  $\Gamma_{\text{qm}}(\xi_1 - \xi_2, 0)$  in the plane of observation, i.e., the spatial coherence, is given by the Fourier transform of the source brightness distribution  $I_b(\alpha')$ ,

$$\Gamma_{\text{qm}}(\xi_1 - \xi_2, 0) = \int I_b(\alpha') e^{-ik_0(\xi_1 - \xi_2) \cdot \alpha'} d\alpha'. \quad (2.50)$$

The Fourier transform of  $I_b(\alpha')$  is taken at the average frequency  $\nu_0$  and  $k_0$  respectively, under the assumption that the size of the source is sufficiently small so that its Fourier transform is invariant over the spectral band  $\Delta\nu$ .

The normalized MCF is called the *complex visibility function*  $\mu_{\nu_0}$ , defined as

$$\mu_{\nu_0}(\xi_1 - \xi_2) = \frac{\Gamma_{\text{qm}}(\xi_1 - \xi_2, 0)}{\sqrt{I(\xi_1)I(\xi_2)}} = \frac{\int I_b(\alpha') e^{-ik_0(\xi_1 - \xi_2) \cdot \alpha'} d\alpha'}{\int I_b(\alpha') d\alpha'}, \quad (2.51)$$

so that  $0 \leq |\mu_{\nu_0}(\xi_1 - \xi_2)| \leq 1$  and  $\mu_{\nu_0}(0) = 1$ . We often denote the integral over  $I_b(\alpha')$  in the denominator by  $I_0 = \int I_b(\alpha') d\alpha'$ , describing the homogeneous intensity in the plane of observation.

It is very important to note that the visibility function is a function of coordinate difference only. The absolute positions of the two points  $\xi_1$  and  $\xi_2$  are not relevant.

The visibility function  $\mu_{\nu_0}(\xi_1 - \xi_2)$  is a complex function with phase  $\phi_{\nu_0}(\xi_1 - \xi_2)$ . Due to the source brightness distribution being real and positive by definition, the modulus  $|\mu_{\nu_0}|$  is symmetric and the phase  $\phi_{\nu_0}$  is anti-symmetric:

$$|\mu_{\nu_0}(\xi_1 - \xi_2)| = |\mu_{\nu_0}(\xi_2 - \xi_1)| \text{ and} \\ \phi_{\nu_0}(\xi_1 - \xi_2) = -\phi_{\nu_0}(\xi_2 - \xi_1).$$

Using the definition of the visibility function, the MSDF in polychromatic illumination, (2.45), can be written as

$$\hat{I}(\xi_1 - \xi_2, \nu) = G(\nu) I_0 \mu_{\nu}(\xi_1 - \xi_2), \quad (2.52)$$

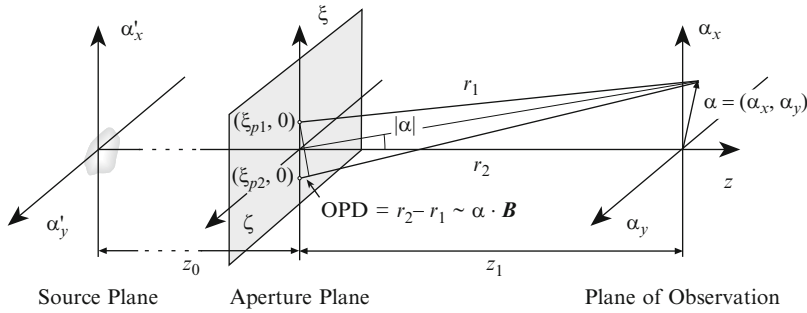
splitting the MSDF in a term  $G(\nu)$  determining the temporal coherence, and a term  $\mu_{\nu}(\xi_1 - \xi_2)$  describing the spatial coherence.

This facilitates the interpretation of the MSDF. The phase,  $\hat{\phi}(\xi_1 - \xi_2, \nu_0)$ , of the MSDF is then identical to the phase,  $\phi_{\nu}(\xi_1 - \xi_2)$ , of the visibility function since  $I_0$  and the spectrum  $G(\nu)$  are real functions.

## 2.4 Young's Experiment Revisited

So far, we have discussed the properties of the coherence functions under various conditions and in different planes along the direction of propagation. But how do we relate these properties to a measurable quantity? In Sect. 2.1.2 it was noted that coherence effects have an influence on the interference pattern in Young's two-pinhole experiment. This will be discussed now in the context of the coherence functions.





**Fig. 2.14** Young's experiment with an incoherent source at large distance  $z_0$ . The two pinholes are at positions  $\xi_{p1} = (\xi_{p1}, 0)$  and  $\xi_{p2} = (\xi_{p2}, 0)$  on the  $\xi$ -axis. This plane is now called the aperture plane. The vector between the pinholes is called *baseline vector* with  $\mathbf{B} = \xi_{p1} - \xi_{p2}$ . The diffraction pattern is calculated as a function of observing angle  $\alpha$  in the plane of observation. The source could be a star and the aperture plane could be on the surface of the Earth

### 2.4.1 The Coherence Function in Young's Experiment

We modify the experiment in Sect. 2.1.2 by replacing the plane wave illumination of the two pinholes<sup>6</sup> by an illumination from an incoherent source at a large distance from the pinholes (see Fig. 2.14). Therefore, we replace the constant amplitudes  $V_0$  in (2.10) by variable amplitudes  $\lambda^2 V(\xi_{p1}, \nu)$  and  $\lambda^2 V(\xi_{p2}, \nu)$ , with  $\lambda^2$ , the square of the wavelength, accounting for the size of the pinholes.

Then, the spectral amplitude  $V(\alpha, \nu)$  in the plane of observation (see Fig. 2.14) is the sum of two elementary (spherical) waves originating from the pinholes, weighted by the spectral amplitudes  $V(\xi_{pi}, \nu)$  in the pinholes,

$$V(\alpha, \nu) = \frac{\lambda^2 V(\xi_{p1}, \nu)}{i\lambda z_1} e^{ikr_1} + \frac{\lambda^2 V(\xi_{p2}, \nu)}{i\lambda z_1} e^{ikr_2}. \quad (2.57)$$

This result is identical to applying the Rayleigh–Sommerfeld diffraction formula (2.16) to an aperture with two pinholes of diameter  $\lambda$ .

We call the vector between the two pinholes the *baseline vector*  $\mathbf{B} = \xi_{p1} - \xi_{p2}$ , and the distance between the pinholes  $B = |\xi_{p1} - \xi_{p2}|$ . The optical path difference (OPD) between the light from the two pinholes is the scalar product of the baseline vector  $\mathbf{B}$  with the coordinate vector  $\alpha$ ,  $\text{OPD} = \alpha \cdot \mathbf{B}$ .

Then, the spectral intensity, i.e., the intensity distribution of the diffraction pattern at frequency  $\nu$ , is

<sup>6</sup> As discussed at the end of Sect. 2.2.1 we define a pinhole as a small aperture with a diameter of a few wavelengths with the approximation that diffraction effects can be calculated using a single spherical wave emerging from the centre of the aperture.

$$\begin{aligned}
I(\boldsymbol{\alpha}, \nu) &= \hat{I}(\boldsymbol{\alpha}, \boldsymbol{\alpha}, \nu) = \lim_{T \rightarrow \infty} \frac{1}{2T} E\{V(\boldsymbol{\alpha}, \nu) V^*(\boldsymbol{\alpha}, \nu)\} \\
&= \frac{\lambda^2}{z_1^2} \lim_{T \rightarrow \infty} \frac{1}{2T} E\{|V(\boldsymbol{\xi}_{p1}, \nu) e^{ikr_1} + V(\boldsymbol{\xi}_{p2}, \nu) e^{ikr_2}|^2\} \\
&= \frac{\lambda^2}{z_1^2} \left( I(\boldsymbol{\xi}_{p1}, \nu) + I(\boldsymbol{\xi}_{p2}, \nu) + 2\text{Re} \left[ \hat{I}(\mathbf{B}, \nu) e^{-ik\boldsymbol{\alpha} \cdot \mathbf{B}} \right] \right), \quad (2.58)
\end{aligned}$$

using  $\boldsymbol{\alpha} \cdot \mathbf{B} = r_2 - r_1$  (see Fig. 2.14), and

$$\begin{aligned}
I(\boldsymbol{\xi}_{pi}, \nu) &= \lim_{T \rightarrow \infty} \frac{1}{2T} E\{V(\boldsymbol{\xi}_{pi}, \nu) V^*(\boldsymbol{\xi}_{pi}, \nu)\}, i = 1, 2, \text{ and} \\
\hat{I}(\mathbf{B}, \nu) &= \lim_{T \rightarrow \infty} \frac{1}{2T} E\{V(\boldsymbol{\xi}_{p1}, \nu) V^*(\boldsymbol{\xi}_{p2}, \nu)\}.
\end{aligned}$$

In the discussion following (2.45), we stated that the intensity in a plane illuminated by an incoherent source is a constant independent of the position  $\boldsymbol{\xi}$  in the aperture plane. Assuming that the two pinholes are illuminated by an incoherent source at a large distance we find indeed that the intensity in the aperture plane is constant,  $I(\boldsymbol{\xi}_{p1}, \nu) = I(\boldsymbol{\xi}_{p2}, \nu) = G(\nu)I_0$ .

Now, we replace the MSDF,  $\hat{I}(\mathbf{B}, \nu)$ , by the product of spectrum and visibility function,  $G(\nu)I_0\mu_\nu(\mathbf{B})$ , see (2.52), and we write the real part in (2.58) as the product of its modulus and the cosine of its phase. We obtain the spectral intensity distribution of the diffraction pattern  $I(\boldsymbol{\alpha}, \nu)$  in the plane of observation as

$$\begin{aligned}
I(\boldsymbol{\alpha}, \nu) &= 2 \frac{\lambda_0^2}{z_1^2} (G(\nu)I_0 + |G(\nu)I_0\mu_\nu(\mathbf{B})| \cos(\phi_\nu(\mathbf{B}) - k\boldsymbol{\alpha} \cdot \mathbf{B})) \\
&= 2G(\nu)I'_0 (1 + |\mu_\nu(\mathbf{B})| \cos(\phi_\nu(\mathbf{B}) - k\boldsymbol{\alpha} \cdot \mathbf{B})), \quad (2.59)
\end{aligned}$$

with  $\lambda_0$  the mean wavelength, and  $I'_0 = \frac{\lambda_0^2}{z_1^2} I_0$ . The result is a fringe pattern essentially proportional to  $1 + \cos(\cdot)$  like in (2.12). The fringe spacing in the direction parallel to  $\mathbf{B}$  is  $\lambda/B$ . The modulus of the visibility function,  $|\mu_\nu(\mathbf{B})|$ , taking values between 0 and 1, acts as a damping factor on the cosine function, determining the contrast of the fringe pattern, and the phase of the visibility function,  $\phi_\nu(\mathbf{B})$ , determines the position of the white-light fringe.

While the simplifying assumption of equal intensities in the two pinholes permits to simplify the formulae without restricting their general validity, we should be aware that in practice, the intensities in two apertures are rarely constant. Writing the intensities in the individual apertures as  $I(\boldsymbol{\xi}_{pi}, \nu) = G(\nu)I_{pi}$ , with  $i = 1, 2$ , we modify (2.59), yielding

$$I(\boldsymbol{\alpha}, \nu) = G(\nu) \left( I'_{p1} + I'_{p2} + 2\sqrt{I'_{p1}I'_{p2}} |\mu_\nu(\mathbf{B})| \cos(\phi_\nu(\mathbf{B}) - k\boldsymbol{\alpha} \cdot \mathbf{B}) \right). \quad (2.60)$$

The contrast of the fringe pattern is now given by  $2(I'_{p1}I'_{p2})^{1/2}/(I'_{p1} + I'_{p2}) \times |\mu_v(\mathbf{B})|$ . We will use this formula when discussing observations through atmospheric turbulence in Sect. 6.1. In the absence of turbulence, we will apply the simplified approach in (2.59).

The fringe pattern of a point source on axis is centred at the position of zero OPD,  $\alpha = 0$ . In case of a non-zero phase  $\phi_v(\mathbf{B})$ , the fringe pattern would be shifted by  $\alpha_0$ , with  $k\alpha_0 \cdot \mathbf{B} = \phi_v(\mathbf{B})$ . A phase of  $\pi$  is equivalent to replacing the  $\cos(\cdot)$  by  $-\cos(\cdot)$ , replacing a fringe maximum by a minimum and vice versa.

Before we discuss this in detail we will step from the spectral intensity  $I(\alpha, \nu)$  to the white-light intensity by integrating over the observed frequency band,  $I(\alpha) = \int I(\alpha, \nu) d\nu$ . For a sufficiently narrow spectral band we replace the factor  $\lambda^2$  before the bracket by the average wavelength  $\lambda_0^2$  since we are primarily interested in the influence of the wavelength on the shape and not in its contribution to the absolute intensity.

For the discussion of the polychromatic case we start by assuming a **point source** on axis, with  $\mu_v(\mathbf{B}) = \int \delta(\alpha') \exp(-ik\mathbf{B} \cdot \alpha') d\alpha' = 1$  in the aperture plane.

Integrating monochromatic fringe patterns over the spectral band, the integration is applied individually to each term in (2.59). The first term yields the white-light intensity in the aperture plane  $\int G(\nu) I_0 d\nu = I_0$ . The second term is again written as the real part of a complex function as in (2.58). Integrating over the frequency  $\nu$ , we replace  $k\alpha \cdot \mathbf{B}$  by  $2\pi\nu\tau$ , writing the polychromatic intensity distribution as a function of time delay  $\tau$  as

$$I(\tau) = 2I'_0 \left( 1 + \operatorname{Re} \left[ \int G(\nu) e^{-i2\pi\nu\tau} d\nu \right] \right). \quad (2.61)$$

Replacing  $\tau$  by  $\alpha \cdot \mathbf{B}/c$  we would obtain the diffraction pattern as a function of  $\alpha$  again.

The spectrum  $G(\nu)$  has its centroid at the average frequency  $\nu_0$ . Introducing the centred spectrum  $G_c(\nu) = G(\nu + \nu_0)$ , we write

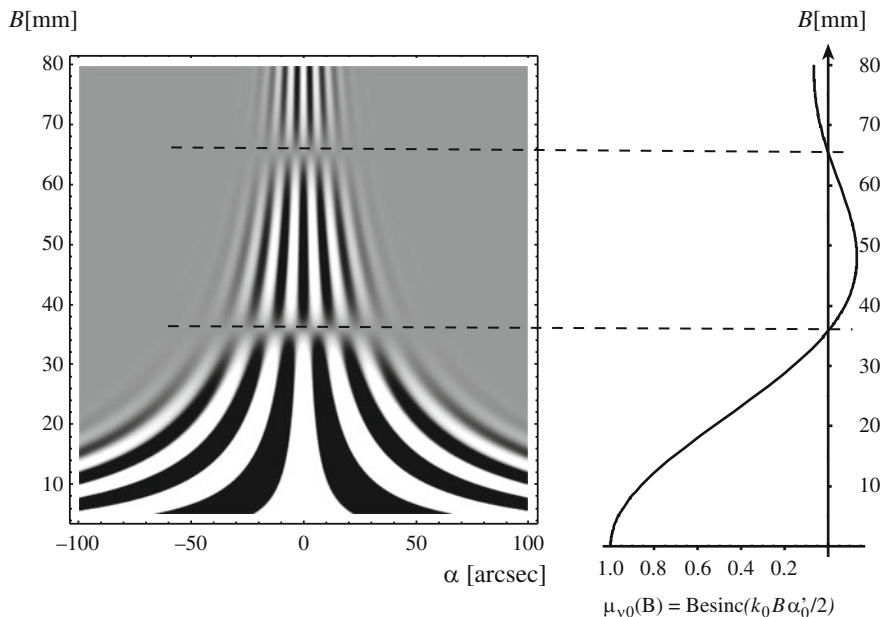
$$\begin{aligned} I(\tau) &= 2I'_0 \left( 1 + \operatorname{Re} \left[ \int G_c(\nu - \nu_0) e^{-i2\pi\nu\tau} d\nu \right] \right) \\ &= 2I'_0 \left( 1 + \operatorname{Re} \left[ \int G_c(\nu') e^{-i2\pi\nu'\tau} d\nu' e^{-i2\pi\nu_0\tau} \right] \right), \end{aligned} \quad (2.62)$$

using the coordinate transform  $\nu' = \nu - \nu_0$ , see Sect. A.1. The Fourier transform of the centred spectrum is called  $g(\tau)$  that is a real function if  $G_c(\nu)$  is symmetric, for example rectangular or gaussian.

The polychromatic fringe pattern of a point source can now be written as

$$I(\tau) = 2I'_0 (1 + g(\tau) \cos(2\pi\nu_0\tau)), \quad (2.63)$$

with  $g(\tau) = \mathcal{F}_\nu(G_c(\nu))$  and  $g(0) = 1$ .  $\mathcal{F}_\nu$  denotes the Fourier transform with respect to coordinate  $\nu$ .



**Fig. 2.15** Fringe patterns of Venus in Young's experiment. On the *left*, the intensity distributions along horizontal lines display the fringe pattern for pinhole separations  $B$  between 10 and 80 mm. The fringe spacing is  $\lambda_0/B$ . The light source illuminating the pinholes is Venus with a uniform disk diameter of 15 arcsec. The spectral band is the  $K$ -band ( $2.2 \pm 0.2 \mu\text{m}$ ), i.e.,  $\lambda_0/\Delta\lambda = 5.5$ . The finite spectral bandwidth makes the fringe visibility disappear for diffraction angles  $|\alpha|$  larger than about  $5.5\lambda_0/B$ , i.e., there are about 11 fringes in the fringe pattern (compare to (2.63)). The fringe spacing decreases with increasing pinhole separation, and the fringe visibility is reduced to zero at  $B = 37$  mm. For  $37 \text{ mm} \leq B \leq 67.5 \text{ mm}$ , the fringe pattern inverts its sign displaying a *black* fringe at  $\alpha = 0$ . On the *right*, the visibility function  $|\mu_{v_0}(\mathbf{B})|$  as a function of pinhole separation  $B$  is displayed following a Besinc-function (see also Fig. 2.13)

The variable part – the cosine function – of a fringe pattern with frequency  $v_0$  is multiplied by  $g(\tau)$ , the (real) Fourier transform of the (centred and symmetric) spectrum  $G_c(v)$ , that acts like an envelope on the fringe pattern. If the fringe pattern is observed in the  $K$ -band ( $2.2 \pm 0.2 \mu\text{m}$ ) with  $\Delta\lambda = 0.4 \mu\text{m}$  and  $\Delta\nu = 2.5 \times 10^{13} \text{ Hz}$ , then  $g(\tau)$  has its first zero for a time delay  $\tau = \pm 1/\Delta\nu = \pm 4 \times 10^{-14} \text{ s}$  (see Sect. A.1). This time delay corresponds to an optical path difference (OPD) equal to the coherence length  $l_c = \tau c = \pm 12 \mu\text{m}$  corresponding to  $\pm 5.5 \lambda_0$ , i.e., the fringe pattern has about 11 perceivable fringes (see Fig. 2.15).

Returning to the general case of an **extended source** we write the integral of the spectral intensity distributions  $I(\alpha, \nu)$  (2.59) over the observed frequency band in the complex notation as

$$\begin{aligned}
I(\boldsymbol{\alpha}) &= \int I(\boldsymbol{\alpha}, \nu) d\nu \\
&= 2I'_0 \left( 1 + \int G(\nu) |\mu_\nu(\mathbf{B})| \cos(\phi_\nu(\mathbf{B}) - k\boldsymbol{\alpha} \cdot \mathbf{B}) d\nu \right), \quad (2.64)
\end{aligned}$$

or, with  $k\boldsymbol{\alpha} \cdot \mathbf{B} = 2\pi\nu\tau$ ,

$$I(\tau) = 2I'_0 \left( 1 + \operatorname{Re} \left[ \int G(\nu) \mu_\nu(\mathbf{B}) e^{-i2\pi\nu\tau} d\nu \right] \right). \quad (2.65)$$

Disentangling the integral over the product of spectrum  $G(\nu)$  and the visibility function  $\mu_\nu(\mathbf{B})$  – both functions of  $\nu$  – we can apply the same reasoning as after (2.47), bringing back the quasi-monochromatic approximation with the assumption of a narrow spectral bandwidth  $\Delta\nu$ , implicitly of a source of size  $\boldsymbol{\alpha}_0$  that is larger than a point but still so small that its visibility function  $\mu_\nu(\mathbf{B})$  does not vary over the spectral band.

We obtain the intensity distribution of the diffraction pattern of Young's experiment in quasi-monochromatic approximation as

$$I_{\text{qm}}(\tau) = 2I'_0 \left( 1 + \operatorname{Re} \left[ \mu_{\nu_0}(\mathbf{B}) \int G(\nu) e^{-i2\pi\nu\tau} d\nu \right] \right), \quad (2.66)$$

with  $\mu_{\nu_0}(\mathbf{B})$  the visibility function at  $\nu_0$ , (2.51).

If the real part of the complex function is written as the modulus of  $\mu_{\nu_0}(\mathbf{B})$  times the cosine of the phase we obtain the intensity distribution of the fringe pattern

$$I_{\text{qm}}(\tau) = 2I'_0 (1 + g(\tau) |\mu_{\nu_0}(\mathbf{B})| \cos(\phi_{\nu_0}(\mathbf{B}) - 2\pi\nu_0\tau)), \quad (2.67)$$

or in the more familiar notation with diffraction angle  $\boldsymbol{\alpha}$ :

$$I_{\text{qm}}(\boldsymbol{\alpha}) = 2I'_0 (1 + g_B(\boldsymbol{\alpha}) |\mu_{\nu_0}(\mathbf{B})| \cos(\phi_{\nu_0}(\mathbf{B}) - k_0\boldsymbol{\alpha} \cdot \mathbf{B})), \quad (2.68)$$

with  $k_0\boldsymbol{\alpha} \cdot \mathbf{B} = 2\pi\nu_0\tau$ , and  $g_B(\boldsymbol{\alpha}) = g(\boldsymbol{\alpha} \cdot \mathbf{B}/c) = g(\tau)$  using  $\boldsymbol{\alpha} \cdot \mathbf{B} = \tau c$ .

Like for the spectral intensity distribution in (2.59) the modulus of the visibility function determines the contrast of the fringe pattern around the white-light fringe, when  $g_B(\boldsymbol{\alpha})$  is approximately constant, and the phase,  $\phi_{\nu_0}(\mathbf{B})$ , determines the position of the fringe pattern with respect to the position of zero OPD at  $\boldsymbol{\alpha} = 0$ .

This result is identical to (2.12) when the two pinholes were illuminated by a monochromatic plane wave, so that  $g_B(\boldsymbol{\alpha})$  is infinitely wide, since the modulus of the visibility function of a plane wave is  $|\mu_{\nu_0}(\mathbf{B})| \equiv 1$  and the phase is  $\phi_{\nu_0}(\mathbf{B}) \equiv 0$ .

In Sect. 2.1.2, the contrast was called the visibility  $\mathcal{V}$  of the fringe pattern and it was defined (2.13) as

$$\mathcal{V} = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}. \quad (2.69)$$

Comparing this definition with (2.68) it is easy to see that, in the ideal case,  $|\mu_{v_0}(\mathbf{B})| = \mathcal{V}$  around the white-light fringe. Thus, we can measure the modulus of the visibility function by determining the fringe contrast in Young's experiment. However, if the fringe pattern is affected by atmospheric turbulence or other disturbances, we distinguish  $\mathcal{V}$ , the contrast of the measured fringe pattern, from the source visibility function  $\mu_{v_0}$ .

Very often, the visibility is interpreted as the quotient between the *correlated flux* and the *total flux* of the light since the visibility function is defined (2.51) as the quotient of the MCF, which is the correlation function of the optical disturbances, and the integrated intensity.

We can now understand the effect of a visibility function on the fringe pattern. Regarding for instance the visibility function of Venus, displayed in Fig. 2.13, that is shaped like a Besinc-function (2.53) we see in Fig. 2.15 that the visibility function shows up in the varying contrast of the fringe patterns. Each horizontal line shows a fringe pattern for an individual pinhole separation  $B$ , the contrast of which is determined by the modulus  $|\mu_{v_0}(\mathbf{B})|$  of the visibility function.

The Besinc-function is a real function with zero phase. However, working with the modulus of the visibility function, negative values of the Besinc-function have to be accounted for by a phase of  $\phi = \pi$ , since  $|\mu_{v_0}(\mathbf{B})|e^{i\pi} = -|\mu_{v_0}(\mathbf{B})|$ . The fringe pattern is shifted by  $\pi$ , producing a black fringe at  $\alpha = 0$  for pinhole separations between 37 mm and 67.5 mm when the Besinc-function has negative values (see Fig. 2.15).

The two pinholes can be regarded as an instrument to measure the coherence properties by probing the wave front with two pinholes, and determining the spatial coherence as the visibility of the fringe pattern. This result is interesting in two respects. First, we found how to relate a measurable quantity to the complex visibility function. Second, by doing so, we derived the visibility function in the aperture plane from characteristics of the intensity distribution (fringe contrast and centre fringe position) in the plane of observation.

### 2.4.2 ABCD Method

A method to derive both modulus and phase of the complex visibility from the fringe pattern was described by Shao [214], originally proposed by Wyant [254]. It is called the *ABCD method* since it relies on measuring the intensity at four different points around the white-light fringe of the fringe pattern  $I_{\text{qm}}(\boldsymbol{\alpha})$ , (2.68), that are separated by  $1/4$  of the fringe spacing of  $\lambda_0/B$ . If we denote the intensities by  $I_A, I_B, I_C, I_D$  and their sum by  $I_{\text{tot}}$  we write

$$|\mu_{v_0}(\mathbf{B})| = \mathcal{V} = C \frac{\sqrt{(I_A - I_C)^2 + (I_B - I_D)^2}}{I_{\text{tot}}}$$

$$\phi_{v_0}(\mathbf{B}) = \tan^{-1} \left( \frac{I_A - I_C}{I_B - I_D} \right). \quad (2.70)$$

The value of the constant  $C$  depends on the measurement method. If the four intensities are determined at individual points on the fringe pattern, it is  $C = 2$ . If the intensities are measured by pixels that are  $\lambda/(4B)$  wide, integrating the intensity over  $1/4$  of the fringe spacing and, thus, damping the fringe pattern, it is  $C = \pi/\sqrt{2}$ . However, this value is only of academic interest since the visibility always has to be calibrated for instance by observing a point source with a nominal visibility  $\mu_{v_0} = 1$ . This will be discussed in Sect. 6.1.

Thus, we can determine the visibility by applying the ABCD method to the white-light fringe. However, if it is difficult to identify the white-light fringe and by mistake another fringe in the field is used, its contrast is affected additionally by the temporal coherence,  $g_B(\alpha)$ , so that the estimate for  $|\mu_{v_0}|$  is too small, while the phase estimate remains unaffected. Therefore, the ABCD method is particularly well suited to measure the fringe position.

As an alternative, the Fourier spectrum of the fringe pattern can be processed providing an estimator for the complex visibility function.

### 2.4.3 Power Spectrum of the Fringe Pattern

The fringe pattern can be regarded as a function of either the diffraction angle  $\alpha$  or of the time delay  $\tau$ . The two quantities are linked through the optical path difference  $\alpha \cdot \mathbf{B} = \text{OPD} = \tau c$ , with the baseline vector  $\mathbf{B} = \xi_{p1} - \xi_{p2}$ . In the following, we will choose the variable  $\tau$  and we will perform the Fourier transform between time and frequency space. The properties of the Fourier spectrum will be discussed in detail before we introduce the power spectrum, which is the squared modulus of the Fourier spectrum.

We will first treat a monochromatic fringe pattern to develop a feeling for the process, and then we will discuss the general expression for the polychromatic diffraction pattern of Young's experiment (2.65).

The monochromatic fringe pattern is calculated in (2.65) by integrating the spectral intensity  $I(\alpha, \nu)$  (2.64) with a monochromatic spectrum  $G(\nu) = \delta(\nu - \nu_0)$  as

$$I(\tau) = 2I'_0 \left( 1 + \text{Re} \left[ \int G(\nu) \mu_\nu(\mathbf{B}) e^{-i2\pi\nu\tau} d\nu \right] \right), \quad (2.71)$$

using  $I'_0 = \frac{\lambda_0^2}{z_1^2} I_0$ . Then the Fourier transform  $\hat{I}(\nu)$  of this fringe pattern reads as

$$\begin{aligned} \hat{I}(\nu) &= \int 2I'_0 \left( 1 + \text{Re} \left[ \int \delta(\nu - \nu_0) \mu_\nu(\mathbf{B}) e^{-i2\pi\nu\tau} d\nu \right] \right) e^{i2\pi\nu\tau} d\tau \\ &= I'_0 (2\delta(\nu) + \delta(\nu - \nu_0) \mu_{\nu_0}(\mathbf{B}) + \delta(\nu + \nu_0) \mu_{\nu_0}^*(\mathbf{B})), \end{aligned} \quad (2.72)$$

Note that, depending on the properties of  $I(\tau)$ ,  $\hat{I}(\nu)$  can be a complex function.

Despite the syntax with time delay  $\tau$  and frequency  $\nu$  this is not a temporal Fourier transform since the replacement of the diffraction angle  $\alpha$  by the time difference  $\tau$  does not make  $I(\tau)$  a temporal signal but a signal of time difference. However, in analogy to (2.34) when introducing the self-coherence function, the coordinate  $\nu$  is the frequency of the light, and we can draw conclusions on the shape of the spectrum  $G(\nu)$  as will be discussed in the following.

The peak of the Fourier spectrum at  $\nu = 0$ , sometimes called the *photometric peak*, describes the constant component of the fringe pattern proportional to the constant intensity  $I_0$  in the aperture plane. The  $\delta$ -peaks at  $\nu_0$  and  $-\nu_0$ , sometimes called the *interferometric peak*, are weighted by the visibility function  $\mu_{\nu_0}(\mathbf{B})$  at baseline  $\mathbf{B}$  determining the amplitude of the cosine pattern.

The fringe pattern of a **polychromatic point source** was discussed in Sect. 2.4.1, and we found the intensity distribution of the fringe pattern (2.61) to be

$$I(\tau) = 2I'_0 \left( 1 + \operatorname{Re} \left[ \int G(\nu) e^{-i2\pi\nu\tau} d\nu \right] \right), \quad (2.73)$$

or, in its more familiar form (2.63), using  $g(\tau) = \mathcal{F}_\nu (G(\nu + \nu_0))$ ,

$$I(\tau) = 2I'_0 (1 + g(\tau) \cos(2\pi\nu_0\tau)), \quad (2.74)$$

when the Fourier transform of the centred spectrum is the envelope of the cosine pattern reducing its visibility with increasing time delay  $\tau$ , or diffraction angle  $\alpha$ .

Using (2.61), it is straightforward to calculate the Fourier spectrum of the fringe pattern of a polychromatic point source as

$$\begin{aligned} \hat{I}(\nu) &= \int 2I'_0 \left( 1 + \operatorname{Re} \left[ \int G(\nu') e^{-i2\pi\nu'\tau} d\nu' \right] \right) e^{i2\pi\nu\tau} d\tau \\ &= I'_0 (2\delta(\nu) + G(\nu) + G(-\nu)), \end{aligned} \quad (2.75)$$

which is displayed in Fig. 2.16. If the width of the spectrum  $G(\nu)$  increases, the width of its Fourier transform decreases and the fringe pattern loses contrast after fewer and fewer periods. Another way of putting this is to say that each individual frequency  $\nu_i$  within the spectral band provides a cosine of frequency  $\nu_i$  in the fringe pattern. Adding up these cosine functions results in a cosine pattern of the average frequency  $\nu_0$  that loses contrast depending on the spread of contributing frequencies, i.e., the width of the spectrum.

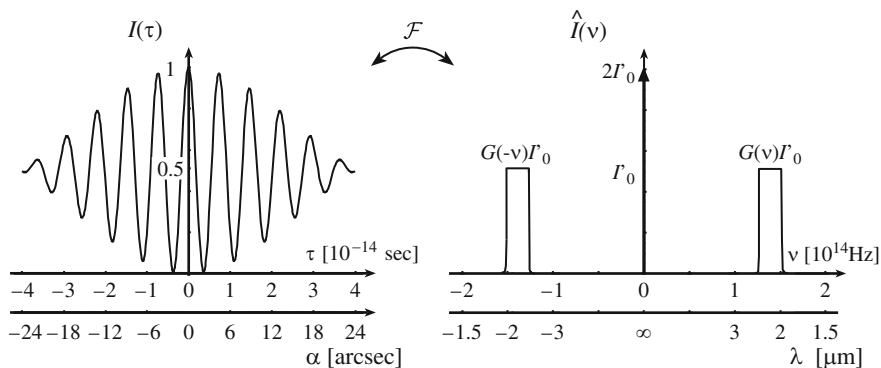
A similar exercise leads us to the Fourier spectrum of an **extended source**. We write the polychromatic fringe pattern of an extended source  $I(\alpha)$  (2.65) as

$$I(\tau) = 2I'_0 \left( 1 + \operatorname{Re} \left[ \int G(\nu) \mu_\nu(\mathbf{B}) e^{-i2\pi\nu\tau} d\nu \right] \right), \quad (2.76)$$

and perform the Fourier transform, yielding the Fourier spectrum as

$$\hat{I}(\nu) = I'_0 (2\delta(\nu) + G(\nu) \mu_\nu(\mathbf{B}) + G(-\nu) \mu_\nu^*(\mathbf{B})). \quad (2.77)$$





**Fig. 2.16** On the *left*, the  $K$ -band fringe pattern of a point source for a pinhole separation of 10 cm is displayed both as a function of time difference  $\tau$  and of diffraction angle  $\alpha$  (for  $\alpha \parallel \mathbf{B}$ ), linked through  $\tau = \alpha \cdot \mathbf{B}/c$ . On the *right*, its Fourier transform is shown both as a function of frequency  $\nu$  and of wavelength  $\lambda$ . The  $\delta$ -peak is at  $\nu = 0$ , and the shape of the spectrum  $G(\nu)$  is given by a rectangular function over the  $K$ -band,  $\nu = 1.25 \times 10^{14} - 1.5 \times 10^{14}$  Hz (respectively  $\lambda = 2.0 - 2.4 \mu\text{m}$ )

The peak of the Fourier spectrum at  $\nu = 0$  describes again the constant component of the fringe pattern, proportional to the white-light intensity  $I_0$  in the aperture plane. The width of the spectrum  $G(\nu)$  and its mirror function at  $-\nu$  is given by the width of the spectral band, and the values of the visibility function  $\mu_\nu(\mathbf{B})$  over the spectral band are determined by the shape of the spectrum and by the spatial coherence at the given baseline  $\mathbf{B}$ .

Now it seems to be straightforward to determine the values of the spectrally resolved visibility function – in modulus and in phase – in the Fourier spectrum. However, in reality the situation is not quite that simple. Celestial bodies at a large distance are observed through the Earth's atmosphere so that the fringe pattern is temporally unstable and weak, and the white-light intensity  $I_0$  in the aperture plane is neither temporally nor spatially constant. The implications will be discussed in detail in Chaps. 5 and 6.

Instead of exploiting the spectral information in the Fourier spectrum, one can do the opposite, averaging over the spectrum to improve the signal quality and directly measure the visibility function.

It was the idea of Roddier and Léna [196] to determine the average visibility function by calculating in the Fourier spectrum the ratio between the integral of the MSDF, the interferometric peak, over the spectral band and the integral of the  $\delta$ -peak, the photometric peak, at  $\nu = 0$ .

Collapsing the spectrally resolved visibility function into a single value means to lose the spectral information. If the spectral band is sufficiently narrow and if the source is sufficiently small we can replace the visibility function by its value at the average frequency  $\nu_0$  yielding  $\mu_{\nu_0}(\mathbf{B})$  in quasi-monochromatic approximation as given in the van Cittert–Zernike theorem (2.50).

Integrating the visibility function over the spectral band and dividing it by the integral of the spectrum at  $\nu = 0$  yields the estimator for the visibility function  $\mu_{\nu_0}(\mathbf{B})$  as

$$\frac{\int I'_0 G(\nu) \mu_\nu(\mathbf{B}) d\nu}{\int 2I'_0 \delta(\nu) d\nu} \approx \frac{\mu_{\nu_0}(\mathbf{B})}{2}. \quad (2.78)$$

Although for this theoretical deduction it is not required to explicitly account for  $I'_0$ , we must not forget that working with real data the spectrum at  $\nu = 0$  is given by  $I'_0$  and around  $\nu = \nu_0$  it is given by the product  $\hat{I}(\nu) = I'_0 G(\nu) \mu_\nu(\mathbf{B})$  so that both integrals in (2.78) have to be performed.

This conceptually simple computation provides a direct measure for the visibility of the fringe pattern. Due to the  $\delta$ -peak at  $\nu = 0$  being a real number we also measure the phase of the visibility function.

Discussing the quasi-monochromatic approximation in Sect. 2.4.1, we emphasised that  $\mu_{\nu_0}(\mathbf{B})$  describes the fringe visibility only for small diffraction angles, i.e., around the white-light fringe, (2.68). It is the integral over the spectral band in (2.78) that reduces the information about the visibility over the full fringe pattern to the visibility around the white-light fringe as discussed before. The advantage of this method is that in case of noisy signals more values of the fringe pattern contribute to the measurement than just a few points around the white-light fringe.

The treatment of noise in the interferogram requires to calculate the power spectrum of the fringe pattern in order to have an unbiased measurement of the visibility function [224]. The power spectrum reads as

$$|\hat{I}(\nu)|^2 = I_0'^2 (4\delta^2(\nu) + |G(\nu) \mu_\nu(\mathbf{B})|^2 + |G(-\nu) \mu_\nu^*(\mathbf{B})|^2), \quad (2.79)$$

and, using again the visibility value at the average frequency  $\nu_0$ , we obtain the squared visibility function as

$$\frac{\int |I'_0 G(\nu) \mu_\nu(\mathbf{B})|^2 d\nu}{\int 4I_0'^2 \delta^2(\nu) d\nu} \approx \frac{|\mu_{\nu_0}(\mathbf{B})|^2}{4}. \quad (2.80)$$

While this quadratic estimator has advantages when measuring noisy signals, the phase  $\phi_\nu(\mathbf{R})$  of the visibility function is lost (see Sect. 6.1.1). However, in practice it is very difficult to determine the position of the white-light fringe since atmospheric turbulence constantly moves the fringe pattern. Even when processing the fringe pattern directly instead of the power spectrum the phase cannot be determined unambiguously without further calibration methods (see Sect. 6.1).

The result of the Fourier transform of the fringe pattern is familiar from the *Michelson Fourier Spectrometer* measuring the source spectrum  $G(\nu)$ . There, the fringe pattern is created by splitting a light beam with a beam splitter and introducing an optical path length modulation in one of the two arms. The fringe pattern is then registered as a function of the OPD modulation after recombining the

beams with a second beam combiner. Formally, the two spatial coordinates are then identical,  $\xi_{p1} = \xi_{p2} = \xi$  and  $B = 0$ , and we obtain the spectral intensity  $I(\xi, \nu) = G(\nu) \int I_b(\alpha') d\alpha' = G(\nu) I_0$  with a homogeneous intensity distribution  $I_0$ .

In Young's experiment, we do not use beam splitters but two pinholes, and the spatial correlation between the two different points  $\xi_{p1}$  and  $\xi_{p2}$  across the wave front enters the result.

Thus, by Fourier transforming the polychromatic fringe pattern  $I(\alpha)$  we receive information not only on the spectrum  $G(\nu)$  but also on the spatial coherence properties of the light determined by the Fourier transform of the source brightness distribution. The term *double Fourier spatio-spectral interferometry* was phrased to emphasise this characteristic [150].

### Example: A Uniform Disk

Stars can be modelled as a uniform disk, writing the source brightness distribution  $I_b(\alpha') = \frac{1}{\pi(\alpha'_0/2)^2} \text{circ}\left(\frac{|\alpha'|}{\alpha'_0/2}\right)$  for a disk with diameter  $\alpha'_0$ . This formula implies that the disk diameter is independent of the wavelength over the observed spectral band, which is a common assumption when attempting for instance to determine stellar diameters.

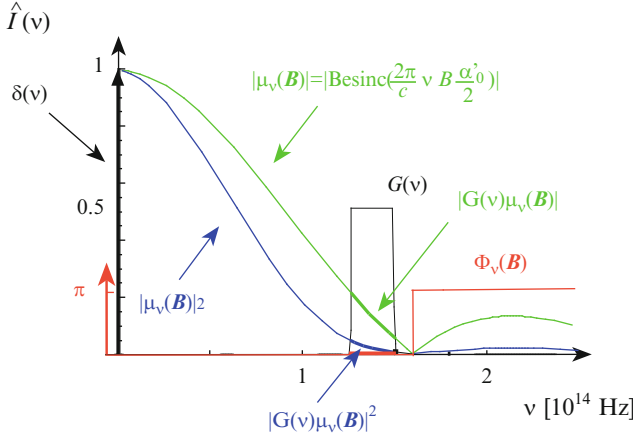
The visibility function is given by the spatial Fourier transform of the circular source brightness distribution (2.53) as

$$\mu_\nu(\mathbf{B}) = \text{Besinc}\left(\frac{2\pi}{c} \nu B \frac{\alpha'_0}{2}\right), \quad (2.81)$$

with  $\frac{2\pi}{c} \nu = \frac{2\pi}{\lambda}$ .

So far, we have always discussed how the visibility function varies with pinhole separation  $B = |\xi_{p1} - \xi_{p2}|$ , and we have taken the wavelength as a fixed parameter. Here, we regard  $B$  as constant and discuss the Besinc as a function of  $\nu$ , i.e., we regard the visibility function for a fixed pinhole separation at different frequencies. Although the disk diameter is wavelength independent, its Fourier transform, the Besinc-function, varies with wavelength since with the exponential kernel  $\frac{2\pi}{\lambda} \mathbf{B} \cdot \alpha'$  the Fourier transform is performed from  $\alpha'$  to  $\mathbf{B}/\lambda = \mathbf{B} \nu/c$ . If the stellar diameter varied with wavelength, the Besinc-function would be distorted reflecting the actual stellar shape at every frequency.

In Fig. 2.17 both the spectrum  $G(\nu)$  and the Besinc-function are displayed as a function of  $\nu$ . The consequences of the averaging process (2.78) over the spectral band are apparent in this figure. As long as the spectral band is sufficiently narrow for the visibility function to be linear within the band, the function value at the average frequency is a good estimate for the integral. It is very obvious that this approximation cannot be fulfilled with the same quality for the visibility function and for the squared visibility function.



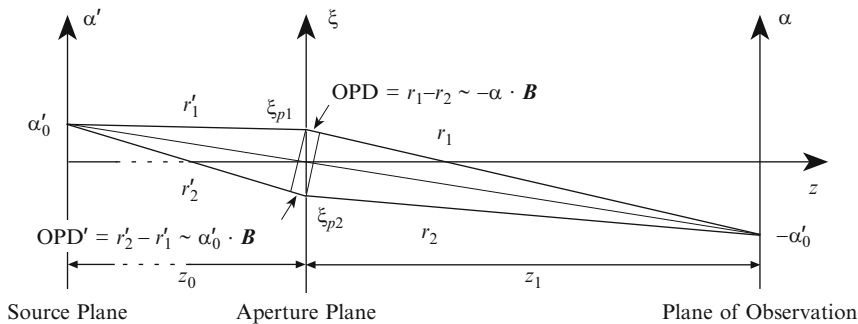
**Fig. 2.17** The spectrum  $\hat{I}(\nu)$  of a  $K$ -band fringe pattern of a uniform disk. The  $\delta$ -peak at  $\nu = 0$  and the visibility function  $\mu_\nu(\mathbf{B})$  at a fixed baseline are displayed. The effective visibility function is the product of the spectrum  $G(\nu)$  with the Besinc-function, which is the visibility function of a uniform disk. Since we discuss the visibility function as a complex function, the Besinc-function is split into its modulus (green lines) and its phase  $\phi_\nu(\mathbf{B})$  (red lines). The phase is zero up to the first minimum of the Besinc-function and jumps to  $\pi$  beyond when the Besinc-function has negative values. For this example the pinhole separation is chosen as  $B = 100$  m and the stellar diameter as  $\alpha'_0 = 4.7$  mas yielding the first zero of the Besinc-function at  $\nu_0 = 1.6 \times 10^{14}$  Hz, using  $\nu_0/\text{Hz} = 755/((B/\text{m})(\alpha'_0/\text{mas}))10^{14}$ . The squared modulus of the Besinc-function (blue lines) is also displayed forming the power spectrum of the fringe pattern

#### 2.4.4 Heuristic Approach

Since the connection between the coherence function and the intensity as a measurable quantity (2.68) is one of the fundamental pillars of stellar interferometry, we also look at it from a different perspective. Why is the fringe contrast affected by the shape of the source? We regard an individual point at angle position  $\alpha'_0$  on the surface of the source with intensity  $I(\alpha'_0)$ . Then, the incoming plane wave is slightly tilted and there is a difference in arrival time at the pinholes that can be expressed as an OPD of  $\alpha'_0 \cdot \mathbf{B}$ . The zero OPD position in the plane of observation is no longer on the optical axis but at position  $\alpha = -\alpha'_0$  so that the incoming and outgoing OPDs,  $\alpha'_0 \cdot \mathbf{B}$  and  $-\alpha \cdot \mathbf{B}$  in Fig. 2.18, cancel.

We regard the fringe pattern in the plane of observation as given in (2.68), with  $|\mu_\nu(\mathbf{B})| \equiv 1$  and  $\phi_\nu(\mathbf{B}) \equiv 0$  for a point source on-axis in quasi-monochromatic approximation with a spectral width  $\Delta\nu$  sufficiently narrow to allow for an OPD of several wavelengths without losing the fringe contrast. For an off-axis point source at angle position  $\alpha'_0$ , the visibility function is  $\mu_\nu(\mathbf{B}) = \exp(-ik\alpha'_0 \cdot \mathbf{B})$  (using (2.51) with  $\xi_1 - \xi_2 = \mathbf{B}$ ), and the intensity distribution reads as

$$I(\alpha, \alpha'_0) = 2 \frac{\lambda^2}{z_1^2} I(\alpha'_0)_\xi \left( 1 + \cos(-k(\alpha + \alpha'_0) \cdot \mathbf{B}) \right).$$



**Fig. 2.18** Young's experiment with a point source at angle position  $\alpha'_0$ . For the sake of simplicity, only one coordinate axis per plane is displayed. The optical path lengths between the source and the pinholes are slightly different with  $OPD' = \alpha'_0 \cdot \mathbf{B}$ . The position of zero OPD in the plane of observation, i.e., the position where the total optical path lengths are equal,  $r'_1 + r_1 = r'_2 + r_2$ , is then at angle position  $\alpha = -\alpha'_0$ . This means that the centre of the fringe pattern, the white-light fringe, is shifted to  $-\alpha'_0$

The homogeneous intensity distribution in the aperture plane due to a source point at  $\alpha'_0$  is denoted by  $I(\alpha'_0)_\xi$ .

The step from a point source to an extended source is done by summing up the fringe patterns of the individual points of the source, as in Fig. 2.19. This is correct since we assumed the light source to be spatially incoherent with every point radiating independently.

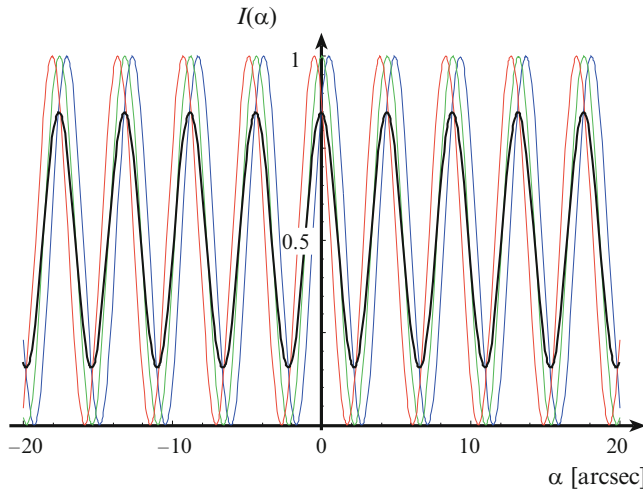
The subsequent integral over the fringe patterns of each point of an incoherent source with diameter  $|\alpha'_0|$  reads as

$$\begin{aligned}
 I(\alpha) &= \int_{-|\alpha'_0|/2}^{|\alpha'_0|/2} I(\alpha, \alpha') d\alpha' \\
 &= 2 \frac{\lambda^2}{z_1^2} \left( I_0 + \int_{-|\alpha'_0|/2}^{|\alpha'_0|/2} I(\alpha')_\xi \cos(-k(\alpha + \alpha') \cdot \mathbf{B}) d\alpha' \right),
 \end{aligned}$$

which is a convolution between the source intensity  $I(\alpha')$  and the fringe pattern of an on-axis point source.

The first term  $I_0$  is a constant intensity as a result of the integration over the source  $I(\alpha')$ . The second term is the real part of the complex Fourier transform  $\int I(\alpha') \exp(-ik(\alpha + \alpha') \cdot \mathbf{B}) d\alpha'$ . We extract  $\exp(-ik\alpha \cdot \mathbf{B})$  since this term does not depend on  $\alpha'$ . The remaining expression  $\int I(\alpha') \exp(-ik\alpha' \cdot \mathbf{B}) d\alpha' = \mu_v(\mathbf{B}) I_0$  is the van Cittert–Zernike theorem as given by (2.51) with  $\xi_1 - \xi_2 = \mathbf{B}$ . Reducing the discussion to the real part again we obtain the final result as in (2.68),

$$I(\alpha) = 2I'_0(1 + |\mu_v(\mathbf{B})| \cos(\phi(\mathbf{B}) - k\alpha \cdot \mathbf{B})),$$



**Fig. 2.19** Three individual monochromatic intensity distributions of the fringe patterns in Young's experiment for point source positions at  $|\alpha'| = -1$  arcsec, 0 and  $+1$  arcsec (*grey lines*), and the resulting fringe pattern of an extended source with diameter 2 arcsec.  $\alpha'$  and  $\alpha$  are parallel to the baseline vector  $\mathbf{B}$ . The pinhole separation  $B$  is 10 cm, the observing wavelength  $\lambda = 2.2 \mu\text{m}$  and the fringe spacing  $\lambda/B = 4.5$  arcsec. The resulting fringe pattern is reduced in contrast since it is the sum of the individual fringes. Note that the contrast would be even smaller if either the pinhole separation were larger – and the fringe spacing smaller – or if the source were larger in diameter

describing a fringe pattern following in principle a  $1 + \cos(\cdot)$  function with a contrast of  $|\mu_v(\mathbf{B})|$ . It is  $I'_0 = \frac{\lambda_0^2}{z_1^2} I_0$ .

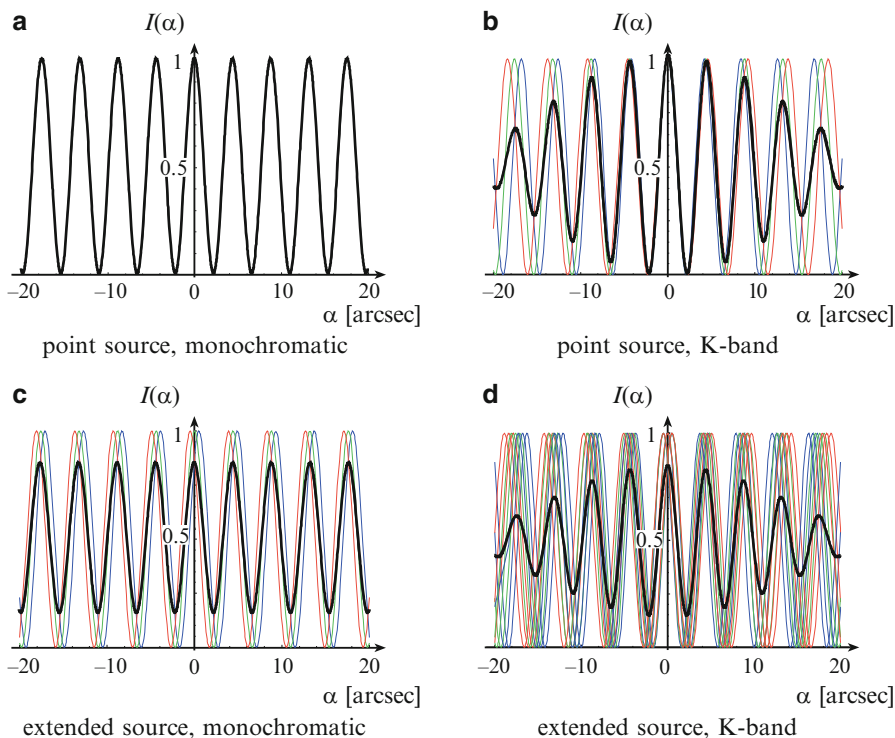
The interpretation of the phase  $\phi$  of the visibility function  $\mu_v$  is now rather simple. Returning to the process of adding up fringe patterns originating from individual points one can think of cases when the resulting fringe pattern does not have its white-light fringe at  $\alpha = 0$ . For example a source that is not extended between  $+$  and  $-|\alpha'_0|/2$  but between 0 and  $|\alpha'_0|$  would display the white-light fringe at  $\alpha = -\alpha'_0/2$ .

In the language of the coherence functions this is another way of saying that the phase of the visibility function is  $-k(\alpha'_0/2) \cdot \mathbf{B}$ . If the source were not constant in intensity but, e.g., be brighter on one side, then fringe patterns with varying intensities have to be added causing the resulting fringe pattern to be shifted sideways. Thus, it is essential to measure not only the modulus but also the phase of the visibility function, i.e., the position of the white-light fringe, if one wants to determine the position of the star or to obtain its shape, which means its image.

The answer to the question at the beginning of this section, why the fringe contrast is affected by the source shape, is thus, that fringe patterns with slightly different white-light fringe positions are overlaid losing contrast by this very process. The formal treatment of this integration reintroduces the coherence functions as discussed in Sect. 2.4.1.

Figure 2.20 illustrates the equivalence of regarding individual source points and regarding individual wavelengths. In both cases, the intensity distributions either of individual source points or of individual wavelengths are integrated. While the width of the spectrum  $G(\nu)$ , in this case the  $K$ -band, determines the temporal coherence, the width of the angular intensity distribution  $I(\alpha')$ , in this case the diameter of the star, determines the spatial coherence. The temporal coherence affects the visibility as a function of diffraction angle  $\alpha$  that translates into a time delay  $\tau$  (Fig. 2.20b), the spatial coherence affects the visibility as a function of baseline  $B$  (Fig. 2.20c). Thus, short baselines display a fringe pattern with higher visibility than long baselines.

In this context, a monochromatic source with infinitely narrow spectrum – or an emission line source –  $G(\nu)$  is equivalent to a point source with infinitely narrow angular intensity distribution  $I(\alpha')$ , since the fringe pattern of the monochromatic



**Fig. 2.20** Summary of the influence of source size and spectral bandwidth on the fringe pattern. The pinhole separation  $B$  is 10 cm in all figures, and we only observe at diffraction angles  $\alpha$  with  $\alpha \parallel B$ . In (a) an individual fringe pattern for an observing wavelength of  $\lambda = 2.2 \mu\text{m}$  and a point source is displayed. In (b) the  $K$ -band fringe pattern is shown when observing a point source (as in Fig. 2.3). In (c) the monochromatic illumination of a source with diameter 2 arcsec produces a fringe pattern with reduced contrast (as in Fig. 2.19). In (d) the resulting fringe pattern in  $K$ -band illumination with a 2 arcsec source is displayed. The visibility is reduced around  $\alpha = 0$  due to the source diameter and it is further reduced for increasing diffraction angles  $\alpha$  due to the finite spectral bandwidth

source has a constant visibility for any value of the diffraction angle  $\alpha$ , and the fringe pattern of a point source displays a constant visibility for any baseline  $\mathbf{B}$ .

It is important to note the different starting point of the discussion in this section compared to the preceding sections. Before, we discussed the propagation of the coherence functions from the source to the aperture plane dealing with the coherence properties along the way. Here, on the contrary, we have regarded a monochromatic point source, for which we calculated the diffraction pattern, and we have drawn our conclusions from integrating the diffraction patterns of many point sources, thus introducing heuristically the concept of coherence at the very end of the process. This is a very convenient method to explain and understand the impact of coherence properties. We will see in Chap. 3 that by summarising integrals in different ways one can focus on different aspects of the imaging process.

In the real world, when stars are neither point sources nor monochromatic one could calculate two sets of fringe patterns, for each point and for each wavelength, and then do the sums as displayed in Fig. 2.20. Or one uses the coherence functions that elegantly combine both cases.

### Young's Experiment Revisited: Summary

The spectral intensity distribution of the diffraction pattern in Young's experiment is a fringe pattern described essentially by a  $1 + \cos(\cdot)$  function with the contrast being determined by the modulus of the visibility function  $\mu_\nu(\mathbf{B})$  and the position of the white-light fringe by its phase,  $\phi_\nu(\mathbf{B})$ ,

$$I(\alpha, \nu) = 2G(\nu)I'_0 (1 + |\mu_\nu(\mathbf{B})| \cos(\phi_\nu(\mathbf{B}) - k\alpha \cdot \mathbf{B})), \quad (2.59)$$

with  $I'_0 = \frac{\lambda_0^2}{z_1^2} I_0$ ,  $G(\nu)$  the spectrum and  $\mathbf{B} = \xi_{p1} - \xi_{p2}$  the baseline vector between the two pinholes at positions  $\xi_{p1}$  and  $\xi_{p2}$ . The fringe spacing is  $\lambda/B$ .

The quasi-monochromatic approximation is valid for narrow spectral bands,  $\Delta\nu \ll \nu$  that means implicitly that the sources must be sufficiently small so that their visibility functions  $\mu_\nu(\mathbf{B})$  do not vary over the spectral band. The intensity distribution of the fringe pattern for small  $\alpha$  can be written as

$$I_{\text{qm}}(\alpha) = 2I'_0 (1 + g_B(\alpha) |\mu_{\nu_0}(\mathbf{B})| \cos(\phi_{\nu_0}(\mathbf{B}) - k_0 \alpha \cdot \mathbf{B})), \quad (2.68)$$

with  $k_0 \alpha \cdot \mathbf{B} = 2\pi \nu_0 \tau$ ,  $\nu_0$  the average frequency,  $\mu_{\nu_0}(\mathbf{B})$  the visibility function at  $\nu_0$ , (2.51), and  $g_B(\alpha) = g_B(\tau c/B) = g(\tau)$  the envelope function of the fringe pattern, which is the Fourier transform of the spectrum  $G(\nu)$ .

The visibility function  $\mu_{\nu_0}(\mathbf{B})$  is a function of baseline vector  $\mathbf{B}$  and not of the individual coordinates  $\xi_{p1}$  and  $\xi_{p2}$  (see Fig. 2.14). For very small diffraction angles around the white-light fringe, when  $g_B(\alpha)$  is approximately



constant, the visibility of the fringe pattern is given by  $|\mu(\mathbf{B})|$ . The phase  $\phi_{v_0}(\mathbf{B})$  determines the position of the fringe pattern.

The *ABCD method* can be used to derive both modulus and phase of the complex visibility from the fringe pattern [214,254]. It relies on measuring the intensity around the white-light fringe at four different points that are spaced by  $1/4$  of the fringe spacing  $\lambda/B$ . If we denote the intensities by  $I_A, I_B, I_C, I_D$  and their sum by  $I_{\text{tot}}$ , we write

$$|\mu_{v_0}(\mathbf{B})| = \mathcal{V} = C \frac{\sqrt{(I_A - I_C)^2 + (I_B - I_D)^2}}{I_{\text{tot}}} \quad (2.70)$$

$$\phi_{v_0}(\mathbf{B}) = \tan^{-1} \left( \frac{I_A - I_C}{I_B - I_D} \right).$$

The value of the constant  $C$  is  $\pi/\sqrt{2}$  if the intensities are measured with pixels that are  $\lambda/(4B)$  wide. In practice, the visibility always has to be calibrated, for instance by observing a point source with a nominal visibility  $\mathcal{V} = 1$ .

Applying the ABCD method to any other fringe than the white-light fringe provides too small an estimate for  $|\mu_{v_0}|$  since the fringe visibility  $\mathcal{V}$  is reduced by the temporal coherence.

Returning to polychromatic illumination with spectrum  $G(v)$ , the Fourier transform (2.77) of the fringe pattern – expressed as a function of  $\tau$ , using  $\tau c = \alpha \mathbf{B}$  – provides information on the temporal and the spatial spectrum of the source (see Fig. 2.17) without the restriction of the quasi-monochromatic approximation to the white-light fringe:

$$\hat{I}(v) = I'_0 (2\delta(v) + G(v)\mu_v(\mathbf{B}) + G(-v)\mu_v^*(\mathbf{B})), \quad (2.77)$$

displaying a peak at  $v = 0$ , and the visibility function  $\mu_v(\mathbf{B})$  and its mirror function  $\mu_v^*(\mathbf{B})$  (see Fig. 2.16). In the absence of spectral information of the source, one can only measure the integral of the visibility function over the spectral band.

Under the assumption that the visibility function is linear over the spectral band the value of the integral is approximately equal to the value of the visibility function at the average frequency  $v_0$  as requested in quasi-monochromatic approximation.

The ratio of the integral of the visibility function over the spectral band in the Fourier transform, the interferometric peak, (2.77) and the integral of the peak at  $v = 0$ , the photometric peak, is a measure for the average visibility function  $\mu_{v_0}(\mathbf{B})$ , with

$$\frac{\int I'_0 G(v)\mu_v(\mathbf{B}, v) dv}{\int 2I'_0 \delta(v) dv} \approx \frac{\mu_{v_0}(\mathbf{B})}{2}. \quad (2.78)$$

If the signals are noisy the power spectrum of the fringe pattern provides an unbiased estimate of the visibility as

$$\frac{\int |I'_0 G(v) \mu_v(\mathbf{B}, v)|^2 dv}{\int 4 I_0'^2 \delta^2(v) dv} \approx \frac{|\mu_{v_0}(\mathbf{B})|^2}{4}. \quad (2.80)$$

Up to now, we have focused on determining the visibility function that we expressed as the Fourier transform of the source brightness distribution according to the van Cittert–Zernike theorem. This incorporates the possibility to reconstruct the source brightness distribution from the visibility function through a Fourier back transform, i.e., to form an image. For this it is required to measure the visibility function for many different vectors  $\mathbf{B}$ . However, without the phase of the visibility function the imaging capability is very limited.

We have, thus, introduced the topic of imaging while originally discussing Young's experiment to explain the basic principles of the measurement of the coherence function. We will see in Chap. 3 that the conceptual step towards an imaging system like a telescope and a stellar interferometer is relatively small.

## 2.5 Higher Order Correlation Functions: Intensity Interferometry

In 1949, R. Hanbury Brown developed the idea of correlating the intensities measured by individual telescopes rather than the amplitudes. The motivation was to determine stellar diameters using very long baselines avoiding – coming from radio interferometry – the use of local oscillators. First results at radio wavelengths were published only three years after the first idea [94]. Together with R.Q. Twiss, Hanbury Brown developed the theory, first for radio interferometers [95] and then for electromagnetic waves in general [97]. An intensity interferometer for the visible was eventually built by them near Narrabri in Australia in the early 1960s [96]. The history and theory of intensity interferometry are competently summarised in Hanbury Brown's book *The Intensity Interferometer* [93].

We will concentrate in the following on the concept of intensity interferometers in the general context of coherence functions, and we will discuss the example of a binary star to explain the concept.

We start by defining the instantaneous intensity

$$i(\xi, t) = v(\xi, t) v^*(\xi, t) \quad (2.82)$$

**Table 2.1** This table summarizes the propagation of light in Young's experiment. Three different quantities, the MSDF, the spectral and the white-light intensity, are regarded in three planes (see Fig. 2.14): the source plane with an incoherent source, the aperture plane with two pinholes, and the plane of observation. In the source plane, the situation is completely described by the spectral intensity distribution  $I(\alpha', \nu)$  since the MSDF  $\hat{F}(\alpha'_1, \alpha'_2, \nu)$  of an incoherent source only has non-zero values for  $\alpha'_1 = \alpha'_2$ , (2.44). The MSDF in the aperture plane is the product of the spectrum  $G(\nu)$  and the Fourier transform of the source brightness distribution  $I_b(\alpha')$ , (2.45). The latter is called the visibility function  $\mu_\nu$ , (2.52), which – taken at the average frequency  $\nu_0$  of a narrow spectral band – forms the core of the van Cittert–Zernike theorem in quasi-monochromatic approximation providing the visibility function  $\mu_{\nu_0}(\xi_1 - \xi_2)$  in the aperture plane, (2.51). The MSDF in the plane of observation  $\hat{F}(\alpha_1, \alpha_2, \nu)$  will be given in (3.48) in Sect. 3.3.1. This quantity is only of limited practical interest, and one would have to design another interferometer experiment to measure it. The white-light intensity  $I(\alpha)$  is the quantity that is measured by an optical detector. In the source plane, the white-light intensity is given by the source brightness distribution. The aperture plane is illuminated homogeneously by the source, and the intensity takes the constant value  $I_0$ . In the image plane, the quasi-monochromatic approximation has to be applied to obtain a solution of the integral over the spectrum, (2.59), yielding a fringe pattern proportional to  $1 + \cos(\cdot)$ , when the contrast of the fringes is determined by the modulus of the visibility  $|\mu_{\nu_0}(\mathbf{B})|$ , (2.68). The baseline vector is defined as  $\mathbf{B} = \xi_{p1} - \xi_{p2}$

Source plane Incoherent star	Aperture plane with two Pinholes at $\xi_{p1}$ and $\xi_{p2}$	Plane of observation with $\mathbf{B} = \xi_{p1} - \xi_{p2}$
$\hat{F}(\alpha'_1, \alpha'_2, \nu)$ $= \frac{\lambda^2}{z_0} G(\nu) I_b(\alpha'_1) \delta(\alpha'_1 - \alpha'_2)$ <p>(2.44) with <math>\alpha' = \mathbf{x}/z_0</math></p>	<p>MSDF</p> $\hat{F}(\xi_1 - \xi_2, \nu)$ $= G(\nu) \int I_b(\alpha') e^{-ik(\xi_1 - \xi_2) \cdot \alpha'} d\alpha'$ <p>(2.45)</p> $= G(\nu) I_0 \mu_\nu(\xi_1 - \xi_2) \quad (2.52)$ <p>Quasi-monochromatic approx.: Visibility function <math>\mu_{\nu_0}(\xi_1 - \xi_2)</math></p> $= \frac{\int I_b(\alpha') e^{-ik_0(\xi_1 - \xi_2) \cdot \alpha'} d\alpha'}{\int I_b(\alpha') d\alpha'} \quad (2.51)$	$\hat{F}(\alpha_1, \alpha_2, \nu)$ $= 2 \frac{\lambda^2}{z_1} \left( G(\nu) I_0 \cos(2\pi \frac{\mathbf{R}_B}{z_2} \cdot (\alpha_1 - \alpha_2)) \right.$ $\left. +  \hat{F}(\mathbf{R}_B, \nu)  \cos(\hat{\phi}(\mathbf{R}_B, \nu) - 2\pi \frac{\mathbf{R}_B}{z_2} \cdot (\alpha_1 + \alpha_2)) \right)$ <p>(3.48) in Sect. 3.3.1, with <math>\mathbf{R}_B = \mathbf{B}/\lambda</math></p>

(continued)

Table 2.1 (Continued)

Source plane Incoherent star	Aperture plane with two Pinholes at $\xi_{p1}$ and $\xi_{p2}$	Plane of observation with $\mathbf{B} = \xi_{p1} - \xi_{p2}$
Spectral intensity		
$I(\boldsymbol{\alpha}', \nu)$ $= G(\nu) I_b(\boldsymbol{\alpha}')$	$I(\xi, \nu) = \hat{I}(\xi - \xi, \nu)$ $= G(\nu) \int I_b(\boldsymbol{\alpha}') d\boldsymbol{\alpha}'$ $= G(\nu) I_0$	$I(\boldsymbol{\alpha}, \nu) = \hat{I}(\boldsymbol{\alpha}, \nu)$ $= 2G(\nu) I'_0$ $\times (1 +  \mu_\nu(\mathbf{B})  \cos(\phi_\nu(\mathbf{B}) - k\boldsymbol{\alpha} \cdot \mathbf{B}))$ (2.59)
White-light intensity		
$I(\boldsymbol{\alpha}')$ $= \int I(\boldsymbol{\alpha}', \nu) d\nu$ $= \int G(\nu) I_b(\boldsymbol{\alpha}') d\nu = I_b(\boldsymbol{\alpha}')$	$I(\xi)$ $= \int I(\xi, \nu) d\nu$ $= \int G(\nu) I_0 d\nu = I_0$	$I(\boldsymbol{\alpha}) = \int I(\boldsymbol{\alpha}, \nu) d\nu = 2I'_0$ $\times (1 + \int G(\nu)  \mu_\nu(\mathbf{B})  \cos(\phi_\nu(\mathbf{B}, \nu) - k\boldsymbol{\alpha} \cdot \mathbf{B}) d\nu)$ (2.64)
Quasi-monochromatic approximation:		
		$I_{\text{qm}}(\boldsymbol{\alpha}) = 2I'_0$ $\times (1 + g_B(\boldsymbol{\alpha})  \mu_{\nu_0}(\mathbf{B})  \cos(\phi_{\nu_0}(\mathbf{B}) - k_0\boldsymbol{\alpha} \cdot \mathbf{B}))$ (2.68)

as the product of optical disturbances. The instantaneous intensity like the optical disturbance is a fast oscillating signal that cannot be measured directly. Therefore, in Sect. 2.1, the intensity was defined as the time average (2.5) of this product, reading

$$I(\xi) = \langle i(\xi, t) \rangle. \quad (2.83)$$

In the present context however, we will discuss the instantaneous intensity before calculating the time average. This is mathematically correct since the ergodicity of the random emission process of light allows us first to calculate statistical averages like the correlation and then to apply temporal averaging [87].

When introducing the coherence function as the second order moment of optical disturbances in Sect. 2.3.1 it was not necessary to make any assumptions on the form of the probability density function of the optical disturbances. Now, computing the correlation of instantaneous intensities and, thus, the *fourth order moment* of optical disturbances, we make use of the common assumption that optical disturbances  $v(\xi, t)$  follow a circular Gaussian random process [87] meaning that the real and imaginary part are independent, identically distributed zero-mean Gaussian random numbers. Then, higher order moments are reduced to second order moments [147], and the correlation of instantaneous intensities, measured at two different points  $\xi_1, \xi_2$  in the aperture plane (e.g., at two telescopes) and at different times  $t_1, t_2$ , is written as

$$\begin{aligned} \langle i(\xi_1, t_1) i(\xi_2, t_2) \rangle &= \langle v(\xi_1, t_1) v^*(\xi_1, t_1) v^*(\xi_2, t_2) v(\xi_2, t_2) \rangle \\ &= \langle v(\xi_1, t_1) v^*(\xi_1, t_1) \rangle \langle v^*(\xi_2, t_2) v(\xi_2, t_2) \rangle \\ &\quad + \langle v(\xi_1, t_1) v^*(\xi_2, t_2) \rangle \langle v^*(\xi_1, t_1) v(\xi_2, t_2) \rangle \\ &= \langle i(\xi_1, t_1) \rangle \langle i(\xi_2, t_2) \rangle + |\langle v(\xi_1, t_1) v^*(\xi_2, t_2) \rangle|^2 \\ &= I(\xi_1) I(\xi_2) + |\Gamma(\xi_1 - \xi_2, \tau)|^2, \end{aligned} \quad (2.84)$$

with  $\tau$  the time difference  $t_1 - t_2$ . Being a stationary random process the correlation only depends on the time difference  $\tau$ .

Since we are interested in the fluctuations of the intensity we write down the intensity covariance as

$$\begin{aligned} \langle i(\xi_1, t + \tau) i(\xi_2, t) \rangle_{\text{cov}} &= \langle i(\xi_1, t + \tau) i(\xi_2, t) \rangle \\ &\quad - \langle i(\xi_1, t + \tau) \rangle \langle i(\xi_2, t) \rangle \\ &= |\langle v(\xi_1, t + \tau) v^*(\xi_2, t) \rangle|^2 \\ &= |\Gamma(\xi_1 - \xi_2, \tau)|^2. \end{aligned} \quad (2.85)$$

Thus, the intensity covariance is the squared modulus of the MCF  $\Gamma(\xi_1 - \xi_2, \tau)$ , which is the correlation function of the optical disturbances (2.27).

We should keep in mind that in practice the intensities are measured in the focal plane of each telescope individually. There is no interference of the amplitudes in a common plane of observation as in Young's experiment. The coordinate  $\xi_i$  denotes

the center of each telescope aperture, and the fact that the average of the intensity over the telescope aperture is actually measured does not affect the conclusion of this discussion.

One major shortcoming of the covariance of two intensities is that the phase of the MCF is lost when the squared modulus of the MCF is computed, and real images cannot be obtained. For binary stars this means that their separation can be measured with high resolution but one cannot determine which one of the two is the “brighter” star if they are of unequal intensity. More generally speaking, the symmetric content of the image can be determined but not the asymmetric content (see Sect. A.1).

Computing the **triple correlation** of intensities, the phase of the MCF can be recovered. We measure the intensities at three points  $\xi_1, \xi_2, \xi_3$  (three telescopes) and at three moments in time  $t_1, t_2$  and  $t_3$ , writing the triple correlation as

$$\begin{aligned} \langle i_1(t_1)i_2(t_2)i_3(t_3) \rangle &= \int i_1(t + \tau_1)i_2(t + \tau_2)i_3(t)dt \\ &= \langle |v_1(t + \tau_1)|^2 |v_2(t + \tau_2)|^2 |v_3(t)|^2 \rangle, \quad (2.86) \end{aligned}$$

when the subscript  $i$  indicates the position  $\xi_i$  where the intensity is measured, and  $t_{1/2}$  are replaced by  $t + \tau_{1/2}$  and  $t_3$  by  $t$ .

Under the same assumption as above, that the optical disturbances follow a circular Gaussian random process, we can reduce the sixth order moment of the disturbances to combinations of second order moments [147], yielding

$$\begin{aligned} \langle i_1(t + \tau_1)i_2(t + \tau_2)i_3(t) \rangle &= I_1 I_2 I_3 \\ &\quad + I_1 | \Gamma_{23}(\tau_2) |^2 + I_2 | \Gamma_{13}(\tau_1) |^2 + I_3 | \Gamma_{12}(\tau_1 - \tau_2) |^2 \\ &\quad + | \Gamma_{12}(\tau_1 - \tau_2) | | \Gamma_{23}(\tau_2) | | \Gamma_{13}(\tau_1) | \\ &\quad \times \cos(\phi_{12}(\tau_1 - \tau_2) + \phi_{23}(\tau_2) - \phi_{13}(\tau_1)). \quad (2.87) \end{aligned}$$

Similar to (2.84), we have the sum of the intensity products and of the squared MCF  $| \Gamma_{ij}(\tau_i) |^2$  plus a term depending on the sum of MCF phases  $\phi_{12}(\tau_1 - \tau_2) + \phi_{23}(\tau_2) - \phi_{13}(\tau_1)$ . The latter is called the *closure phase* because it is the sum of phases around a closed loop of three telescopes. Repeating this measurement for a large number of configurations one can recover the phases  $\phi_{ij}$  of the individual baselines from the closure phase [9, 84].

The intensity covariance in (2.85) represents the ideal result if the instantaneous intensities  $i(\xi, t)$  were measured with infinite temporal resolution. However, only the time averaged intensity is available. Applying a temporal averaging over a period  $T$  to the instantaneous intensities on the left hand side of (2.85) we have to calculate the time average of  $| \Gamma(\xi_1 - \xi_2, \tau) |^2$ , now with respect to the time difference  $\tau$ :

$$| \Gamma(\xi_1 - \xi_2, \tau) |_T^2 = \frac{1}{T} \int_{\tau-T/2}^{\tau+T/2} | \Gamma(\xi_1 - \xi_2, \tau') |^2 d\tau'. \quad (2.88)$$

If  $T$  is much smaller than the width of the MCF, the time averaged signal is very similar to the original signal. Since the width of the MCF is determined by the coherence time  $\tau_c$  the intensities would have to be measured with a temporal resolution better than the coherence time. However, even for a very narrow spectral bandwidth  $\Delta\nu$  of 1/1,000 of the frequency, e.g.,  $\nu \approx 10^{15}$  Hz in the visible, the coherence time is approximately  $\tau_c = 10^{-12}$  s, which is beyond the state of the art of today's detectors resolving signals only down to approximately  $T = 10^{-9}$  s.

If  $T$  is much larger than the width of the MCF, the time average of  $|\Gamma(\xi_1 - \xi_2, \tau)|^2$  is proportional to the coherence time  $\tau_c$  [146]. The result is that the time average of the MCF and thus the covariance of time averaged intensities (2.85) are attenuated by a factor of  $\tau_c/T = \Delta f/\Delta\nu$  compared to an ideal detector, reducing its sensitivity.  $\Delta f = 1/T$  is the detector bandwidth. Diminishing the spectral bandwidth  $\Delta\nu$  reduces the intensity so that the overall situation does not change.

Therefore, the signal-to-noise ratio (SNR) of this measurement is independent of the spectral bandwidth  $\Delta\nu$  but it is proportional (1) to the spectral intensity of the source – strictly speaking to the number of detected photo electrons per unit optical bandwidth and per unit time –, (2) to  $\sqrt{\Delta f}$ , the square root of the detector bandwidth, and (3) to the square root of the integration time, limiting the sensitivity for a  $5\sigma$  SNR to stars of approximately magnitude 5 for 10-m class telescopes and 10-min observations [97]. This does not compare very favourably to amplitude interferometry reaching stars of magnitude 10 in a few 10 msec.

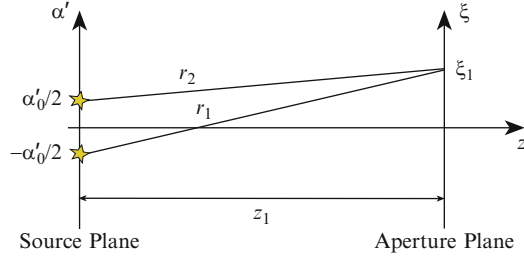
The averaging process does not affect the behaviour of the MCF with respect to the spatial coordinates  $\xi_i$  that is determined by the shape of the object. Thus, the purpose of the intensity interferometer to measure the squared MCF as a function of baseline  $B = \xi_1 - \xi_2$  can be pursued.

### Example: A Binary Star

The simple example of a binary star will give us an idea why the covariance of intensities turns out to be the square of the correlation of optical disturbances, i.e., the square of the MCF, and why only small frequency differences lead to a measurable signal. However, the reasoning in the following on the plausibility of the concept cannot replace a thorough mathematical analysis that can be found in [87, 93].

We place the individual stars at positions  $+\alpha'_0/2$  and  $-\alpha'_0/2$  and we assume that Fraunhofer conditions (see Sect. 2.2.2) apply for the setup displayed in Fig. 2.21. Then, the optical path length  $r_1$  for the light from the star at  $\alpha'_0/2$  can be approximated by  $z_1 + \xi_1\alpha'_0/2$ , and  $r_2$  by  $z_1 - \xi_1\alpha'_0/2$ . We ignore the constant distance  $z_1$ , giving rise to a constant phase term, and we write the optical disturbance with frequency  $\nu$  at position  $\xi_1$  as the sum of the individual contributions

$$v(\xi_1, t) = V_0 \cos\left(k\xi_1 \frac{\alpha'_0}{2} + 2\pi\nu t + \varphi_1\right) + V_0 \cos\left(-k\xi_1 \frac{\alpha'_0}{2} + 2\pi\nu t + \varphi_2\right). \quad (2.89)$$



**Fig. 2.21** A binary star separated by  $\alpha'_0$  at a distance  $z_1$  from the aperture plane with coordinate  $\xi$ . The optical path lengths  $r_1$  and  $r_2$  between the individual stars and the position  $\xi_1$  are displayed. The separation of the binary and the distance allows for the application of the Fraunhofer approximation with  $r_{1/2} \approx z_1 \pm \xi_1 \alpha'_0/2$

$\varphi_{1/2}$  are random phase terms considering the fact that the light from the two stars although formally of the same frequency will have a variable phase difference due to the random emission process in each star.

We can now write the instantaneous intensity as

$$i(\xi_1, t) = v^2(\xi_1, t) \\ = 4V_0^2 \cos^2 \left( k\xi_1 \frac{\alpha'_0}{2} + \frac{\varphi_1 - \varphi_2}{2} \right) \cos^2 \left( 2\pi\nu t + \frac{\varphi_1 + \varphi_2}{2} \right). \quad (2.90)$$

Since we cannot measure signals with a time resolution of the optical frequency we apply a moving time average over a period  $T$  longer than  $1/\nu$  but shorter than the typical fluctuation of the random phases  $\varphi_1$  and  $\varphi_2$ .

Then the second  $\cos^2$  term that is a function of  $\nu$  averages to  $\frac{1}{2}$ . We call the averaged signal  $i_T(\xi_1, t)$ , yielding

$$i_T(\xi_1, t) = \frac{1}{T} \int_{t-T}^t i(\xi_1, t') dt' = 2V_0^2 \cos^2 \left( k\xi_1 \frac{\alpha'_0}{2} + \frac{\Delta\varphi}{2} \right), \quad (2.91)$$

with  $\Delta\varphi = \varphi_1 - \varphi_2$ . This is a periodic signal, like a fringe pattern, as a function of the coordinate  $\xi_1$  in the aperture plane. The period length is determined by the wavelength and by the separation of the binary. The time dependence of  $i_T(\xi_1, t)$  is determined by the randomly varying values of  $\Delta\varphi$ . If the random phases were zero we formally would have a coherent binary in (2.89) – which is the same as a pair of pinholes illuminated by a plane wave – displaying a fringe pattern as in (2.59).

The random phases  $\varphi_1$  and  $\varphi_2$  were introduced because the phase of the wave with frequency  $\nu$  changes randomly each time wave trains are emitted by individual atoms. If the integration time  $T$  in (2.91) is extended beyond the typical time constant of the fluctuations of the random phases, i.e., of their difference  $\Delta\varphi$ , the  $\cos^2$  function is averaged to  $\frac{1}{2}$ , and one obtains the familiar expression of a constant intensity  $i_T(\xi_1, t) = V_0^2$  in the aperture plane that is illuminated by two stars at a large distance.



What happens if light of two different frequencies interferes? We first have to add optical disturbances of two different frequencies  $\nu_1$  and  $\nu_2$  in the first and second term of (2.89), eventually obtaining the intensity time average over  $T$  as

$$i_T(\xi_1, t) = 2V_0^2 \cos^2 \left( k\xi_1 \frac{\alpha'_0}{2} + \frac{\Delta\varphi}{2} + \pi(\nu_1 - \nu_2)t \right). \quad (2.92)$$

This means that as long as the frequency difference, the beat frequency,  $\delta\nu = \nu_1 - \nu_2$  is smaller than  $1/T$ , the contribution of these two frequencies to the intensity is temporally resolved. However, we still cannot determine the separation  $\alpha'_0$  of the binary because of the random phase  $\Delta\varphi$  taking different values for each measurement of  $i_T(\xi_1, t)$ .

If the integration time  $T$  is both longer than  $1/\delta\nu$  and longer than the time constant of the phase difference  $\Delta\varphi$  then the fluctuations average out and the intensity is constant with  $i_T(\xi_1, t) = 2V_0^2$ .

Regarding the interference of optical disturbances with different frequencies in (2.92) seems to contradict our former statements that interference processes can be computed by first determining the spectral intensity, and then by integrating these intensities over the spectrum finding the polychromatic intensity (see e.g. Sect. 2.1). However, this approximation is limited to the cases when the integration time  $T$  is longer than the time scales of the fluctuations involved. Here in (2.92), we assume explicitly that the  $T$  is shorter than the fluctuations so that extra terms have to be considered. In the preceding sections, we were not interested in high frequency fluctuations of the intensity but we discussed the (long) time average intensity of the fringe pattern in order to determine the coherence function in the aperture plane.

The computation of the correlation between the two measured signals  $i_T(\xi_1, t + \tau)$  and  $i_T(\xi_2, t)$  with frequency  $\nu$  – at positions  $\xi_1$  and  $\xi_2$  and at times  $t + \tau$  and  $t$  – is now straightforward. With (2.91) and  $V_0^2 = 1$  we obtain the product of the intensities as

$$\begin{aligned} i_T(\xi_1, t + \tau) i_T(\xi_2, t) &= 4 \cos^2 \left( k\xi_1 \frac{\alpha'_0}{2} + \frac{\Delta\varphi}{2} \right) \cos^2 \left( k\xi_2 \frac{\alpha'_0}{2} + \frac{\Delta\varphi}{2} \right) \\ &= (1 + \cos(k\xi_1 \alpha'_0 + \Delta\varphi)) (1 + \cos(k\xi_2 \alpha'_0 + \Delta\varphi)) \\ &= 1 + \cos(k\xi_1 \alpha'_0 + \Delta\varphi) \\ &\quad + \frac{1}{2} \cos(k(\xi_1 - \xi_2) \alpha'_0) + \frac{1}{2} \cos(k(\xi_1 + \xi_2) \alpha'_0 + 2\Delta\varphi) \\ &\quad + \cos(k\xi_2 \alpha'_0 + \Delta\varphi). \end{aligned} \quad (2.93)$$

The correlation is the time average over a period that is long enough to average over all fluctuations of the random process, i.e., much longer than  $T$  (see Sect. 2.3.1). Then, all terms containing  $\Delta\varphi$  in (2.93) disappear since their average is zero, yielding

$$\begin{aligned}
\langle i_T(\xi_1, t + \tau) i_T(\xi_2, t) \rangle &= 1 + \frac{1}{2} \cos(k(\xi_1 - \xi_2)\alpha'_0) \\
&= \frac{1}{2} + \cos^2\left(k(\xi_1 - \xi_2)\frac{\alpha'_0}{2}\right). \quad (2.94)
\end{aligned}$$

This is the correlation of the intensities at  $\xi_1$  and  $\xi_2$  when observing a binary star separated by  $\alpha'_0$  in monochromatic illumination at frequency  $\nu$ . As in the general result in (2.84) we have a constant term and we found the squared modulus of the MCF of a binary (see (3.75)).

If we now look again at the signal with two different frequencies  $\nu_1$  and  $\nu_2$  with a frequency difference  $\delta\nu$  we obtain the intensity correlation as

$$\langle i_T(\xi_1, t + \tau) i_T(\xi_2, t) \rangle = \frac{1}{2} + \cos^2\left(k(\xi_1 - \xi_2)\frac{\alpha'_0}{2} + \pi\delta\nu\tau\right). \quad (2.95)$$

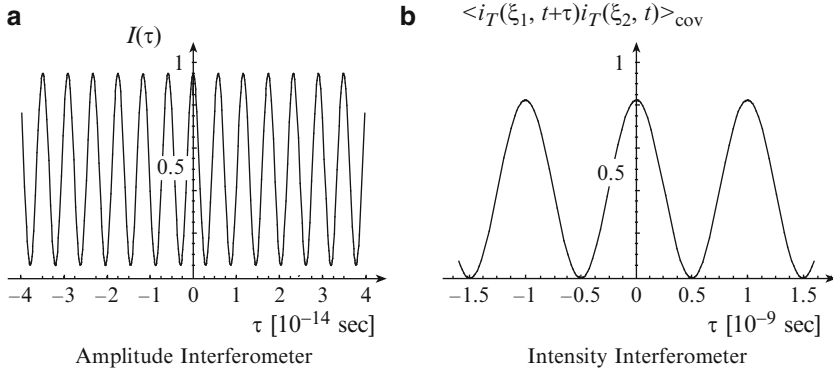
This is the result if the integration time  $T$  is shorter than  $1/\delta\nu$ . If the beat frequency  $\delta\nu$  is larger than  $1/T$  then the individual intensities  $i_T(\xi_i, t)$  are constant and the intensity correlation is itself constant (see discussion after (2.92)) providing a constant background signal.

In our deduction, we assumed that one star radiates with frequency  $\nu_1$  and its neighbour with  $\nu_2$ . If the spectrum were properly considered by having both stars emit light of two frequencies,  $\nu_1$  and  $\nu_2$ , the  $\cos^2$  function in (2.95) would be replaced by a product of two  $\cos^2$  functions yielding

$$\begin{aligned}
\langle i_T(\xi_1, t + \tau) i_T(\xi_2, t) \rangle &= 1 + \cos^2\left(k(\xi_1 - \xi_2)\frac{\alpha'_0}{2}\right) \cos^2(\pi\delta\nu\tau) \text{ and} \\
\langle i_T(\xi_1, t + \tau) i_T(\xi_2, t) \rangle_{\text{cov}} &= \cos^2\left(k(\xi_1 - \xi_2)\frac{\alpha'_0}{2}\right) \cos^2(\pi\delta\nu\tau). \quad (2.96)
\end{aligned}$$

Thus, the intensity covariance as a measure for the intensity fluctuations is exactly the squared modulus of the MCF of a binary observed at two frequencies. The MCF of a binary as the Fourier transform of its intensity distribution is  $\cos(kB\alpha'_0/2)$ , with  $B = \xi_1 - \xi_2$ , and  $\cos(\pi\delta\nu\tau)$  is the Fourier transform of the spectrum consisting of two spectral lines at  $\nu_1$  and  $\nu_2$ . For this result, we assumed that the integrals over  $\nu$  and  $\alpha'$  in the van Cittert–Zernike theorem can be treated individually (see discussion following (2.47)).

As discussed above, only those frequencies contribute to the intensity fluctuations that are separated by less than  $\delta\nu = 1/T$  and all those that are further apart contribute to the background signal. The temporal resolution of the available detectors limits  $T$  to about  $10^{-9}$  s. Observing a finite spectrum that is usually much wider than  $10^9$  Hz, the measured signal of the intensity covariance is the squared MCF attenuated by  $\delta\nu/\Delta\nu$ , with  $\Delta\nu$  the width of the observed spectrum [87]. Discussing the time average of the squared MCF (2.88) we came to the same conclusion stating that the fluctuations are reduced by a factor of  $\tau_c/T$ , with  $\tau_c = 1/\Delta\nu$  the coherence time.



**Fig. 2.22** The measurement of the MCF (a) with an amplitude interferometer, e.g., Young's experiment, and (b) with an intensity interferometer. The observed binary star is separated by  $\alpha'_0 = 6$  mas and two spectral lines are accounted for, centred around  $\lambda = 2.2 \mu\text{m}$ , i.e.,  $\nu = 1.37 \times 10^{14}$  Hz, and separated by  $\delta\nu = 10^9$  Hz. The separation of the two pinholes in the aperture plane is the baseline  $B = \xi_1 - \xi_2 = 10$  m. The fringe pattern in Young's experiment, (a), which is the intensity distribution in the plane of observation due to interference of the amplitudes, here as a function of time delay  $\tau$ , displays a visibility of about 0.9 according to the binary's MCF  $\cos(kB\alpha'_0/2)$  for the given parameters. With an intensity interferometer, (b), the measured signal is the covariance between the intensities when the two intensity detectors replace the pinholes in the aperture plane. Displayed is the covariance signal (2.96) given by  $\cos^2(\pi\delta\nu\tau)$  attenuated by  $\cos^2(kB\alpha'_0/2) = 0.9^2$ , the latter being the squared modulus of the MCF

Figure 2.22 shows the comparison of interferometric measurements of a binary star, separated by  $\alpha'_0 = 6$  mas, with an amplitude interferometer (Fig. 2.22a) and with an intensity interferometer (Fig. 2.22b). Out of the K-band, two spectral lines, separated by  $\delta\nu = 10^9$  Hz, are used. The separation of the pinholes is  $B = 10$  m. The amplitude interferometer as in Young's experiment shows a fringe pattern as a function of time delay  $\tau$  with a fringe period of  $1/\nu = 0.73 \times 10^{-14}$  s (see Fig. 2.22a) equivalent to  $\lambda/B = 45$  mas fringe spacing. For these parameters the MCF of the binary,  $\cos(kB\alpha'_0/2)$ , has a value of 0.9 displayed by the reduced fringe contrast. The beat frequency of  $\delta\nu = 10^9$  Hz is about  $10^5$  times smaller than the average frequency of  $\nu = 1.37 \times 10^{14}$  Hz at  $2.2 \mu\text{m}$ , and gives rise to a periodic envelope,  $\cos(\pi\delta\nu\tau)$ , of the fringe pattern with a period length of about  $10^5$  fringes, i.e.,  $2 \times 10^{-8}$  s that cannot be displayed in Fig. 2.22a.

The square  $\cos^2(\pi\delta\nu\tau)$  of this envelope with a period length of the reciprocal of the beat frequency of  $1/\delta\nu = 10^{-9}$  s describes the intensity covariance (2.96) in an intensity interferometer as displayed in Fig. 2.22b. Here, the detectors are placed at the positions of the pinholes. The 6 mas separation of the binary reduces the maximum of the covariance to  $\cos^2(kB\alpha'_0/2) = 0.9^2$ . The latter, the spatial part of the MCF, is the quantity that we want to determine.

Writing these results as a function of OPD instead of time delay  $\tau$ , using  $\text{OPD} = \tau c$ , we see that while the fringe spacing in the amplitude interferometer corresponds to an OPD of  $\lambda$ , i.e.,  $2.2 \mu\text{m}$  in the example above, the spacing of two maxima in

Fig. 2.22b corresponds to an OPD of  $c/\delta\nu = 30$  cm. This comparison illustrates the advantage of the intensity interferometer over the amplitude interferometer when it comes to real interferometers. If two telescopes are used and the detectors are placed at the end of an optical system, the optical path lengths can be several hundred metres long including a large number of optical surfaces. In both the intensity and the amplitude interferometer one has to make sure that the optical path difference, OPD, is about zero so that we know how to define  $\tau = 0$  in the measurement setup. However, the requirements for an intensity interferometer are much more relaxed compared to an amplitude interferometer.

For precise measurements with an amplitude interferometer, the OPD has to be stable to a fraction of the wavelength, since the fringe pattern moves by one fringe if the OPD varies by  $\lambda$ . With an intensity interferometer, the relevant fringe spacing is about 30 cm (see Fig. 2.22) so that the required OPD accuracy is in the centimetre range.

By the same token, the random variations of the optical path lengths of several micrometers caused by atmospheric turbulence (see Chap. 4) do not disturb the measurement with an intensity interferometer since they are far below the relevant stability requirement of some centimetres.

At the time when Hanbury Brown came up with his idea this was an enormous advantage since the technology was not available to control the OPD with sub- $\mu\text{m}$  accuracy as required for an amplitude interferometer.

In the meantime, the problems of OPD control have been resolved, and amplitude interferometers are in operation. Considering the severely limited sensitivity of intensity interferometry (see discussion following (2.88)) there have been no new attempts of building intensity interferometers. However, some new ideas based on using a large number of telescopes, that are relatively cheap due to the low requirement on optical quality, might lift the sensitivity to useful levels [173, 174].



<http://www.springer.com/978-3-642-15027-2>

Principles of Stellar Interferometry

Glindemann, A.

2011, XVIII, 346 p., Hardcover

ISBN: 978-3-642-15027-2