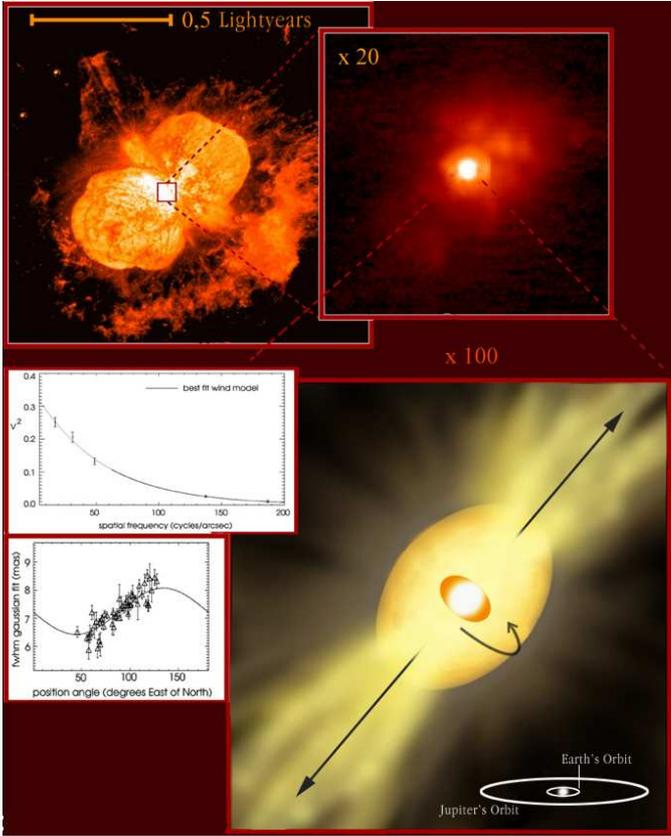
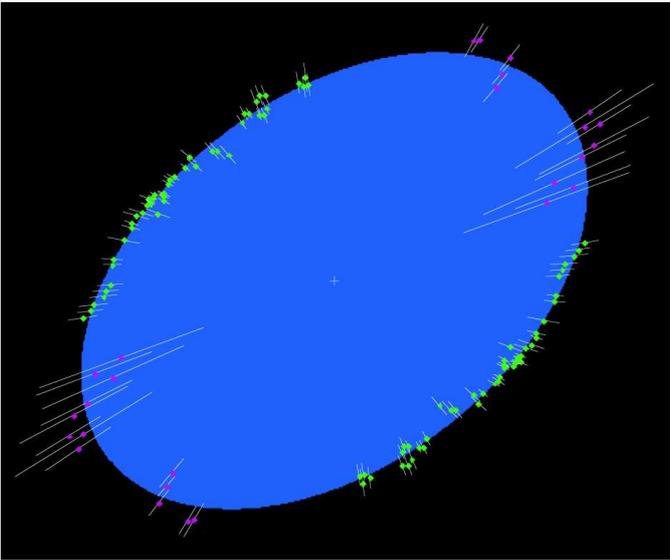


Introduction to Spatial Interferometry



Andreas Glindemann

ESO Garching

Contents

1	Why Spatial Interferometry?	3
2	Introduction	4
3	Preliminaries and Definitions	4
4	Young's Experiment	6
5	The Visibility and the Mutual Coherence Function	10
5.1	Example: Venus	13
6	Image Formation	14
6.1	Image Formation in a Single Telescope	14
6.2	Image Formation in a Spatial Interferometer	16
6.2.1	Example: A Narrow Binary Star	19
7	Practical Issues	20
8	Visibility Measurement	21
9	Some Further reading	22

1 Why Spatial Interferometry?

Over many centuries, since the times of Galileo Galilei (1564-1642), astronomers have relied on increasing the size of their telescopes to improve the quality of the observations. Telescope diameters have grown from Galileo's telescope with a few centimetres to the current 10-m class of telescopes represented by the Keck Telescopes and ESO's Unit Telescopes. In all this time, the basic idea has not changed: the telescope as an optical system produces an image in the focal plane that is taken as a representation of the object shape.

In the 19th century, Hippolyte Fizeau (1819-1896) suggested to improve the measurement of stellar diameters by masking the telescope aperture with two small pinholes. Light passing through these pinholes would then interfere in the telescope focal plane. Hence the term *spatial interferometry*. The first successful measurement using this principle was performed by Albert A. Michelson and Francis G. Pease in 1920 determining the diameter of α Orionis to 0.047 arcsec. This was at a time when the smallest diameter that could be measured with a full aperture was about 1 arcsec, equivalent to the angular resolution of the telescope.

Although the measurement of a stellar diameter is not the same as an image, the dramatic increase in angular resolution sparked enough interest in the new method so that it was soon understood how such contrast measurements with different pairs of pinholes – different in separation and orientation – can be combined to form a high resolution image not only of stars but of any type of object.

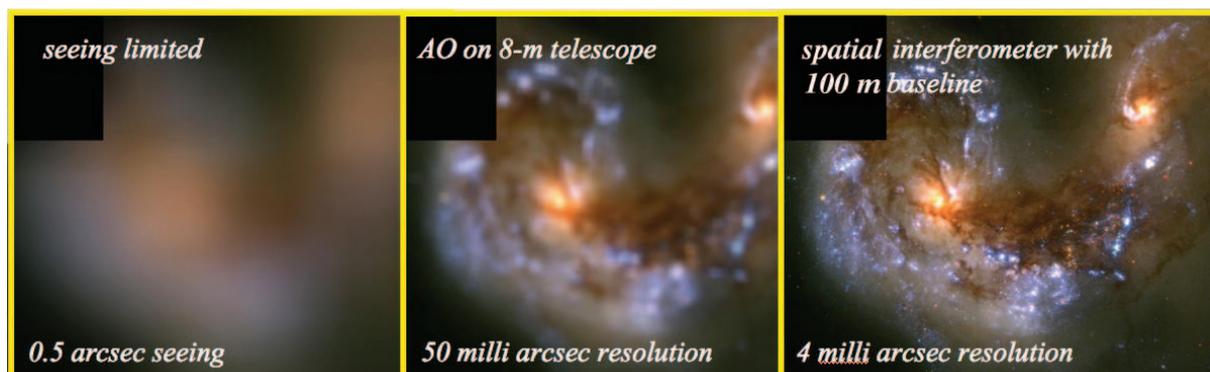


Figure 1: Comparison of the image quality on an 8-m telescope at an observing wavelength of $2 \mu\text{m}$. On the left, the seeing limited image is displayed for a seeing of 0.5 arcsec. In the middle, the diffraction limited image of an 8-m telescope with adaptive optics is shown. On the right, the reconstructed image of a spatial interferometer with a baseline of 100 m demonstrates the improvement over single telescope observations. These images are only simulations for comparison of the achievable angular resolution. Especially for spatial interferometers, an image with this amount of detail and this size has yet to be achieved.

However, due to insurmountable technical problems with the mechanical stability at larger pinhole separations, spatial interferometry was abandoned in the late 1920's. It took until 1974 when Antoine Labeyrie combined the light from two independent telescopes (instead of two pinholes in one telescope) demonstrating that spatial interferometry was feasible.

Over the last decade, high angular resolution methods have matured considerably so that spatial interferometry and adaptive optics (see experiment adaptive optics) have become standard observing techniques, and spatial interferometry has gone beyond the measurement of stellar diameters. However, it is still a long way to go before 'real' high resolution images as in Fig. 1 can be produced.

2 Introduction

There are mainly two parameters determining the performance of astronomical observations: sensitivity and angular resolution. The sensitivity increases with the telescope size, with larger telescopes simply collecting more photons. The angular resolution depends on the telescope diameter if the telescope image is diffraction limited. To achieve this, the optical aberrations of the telescope have to be smaller than the diffraction effects, and an adaptive optics system has to be available to eliminate the image blur due to atmospheric turbulence.

Without adaptive optics the angular resolution of even the largest telescopes in the 10-m class is equivalent to that with telescopes of 10-cm (in the visible) to 1-m (in the near infrared) diameter. Then the resolution is called *seeing* limited and its value is called the seeing. The advent of reliable adaptive optics systems at the end of the 1990s was a big step forward increasing the angular resolution from about 0.5 arcsec to 50 milli arcsec in the near infrared (see Fig. 1).

However, even today's largest telescopes cannot resolve features on the surface of individual stars. The diffraction limit is still so much larger than the stars that their images in the telescope focus are undistinguishable from point sources. For example, the angular resolution of 50 milli arcsec is only just about the angular size of Betelgeuse, the biggest star (in terms of angular size as seen from the Earth) in our galaxy.

Switching together individual telescopes to form a *spatial interferometer* the resolution is no longer determined by the individual telescope diameter but by the distance between the telescopes, called the *baseline*, B . Combining two 8-m telescopes that are separated by $B = 130$ m, like the Unit Telescopes of *ESO's VLTI*, improves the angular resolution by a factor of $\text{baseline/telescope diameter} = 130/8 = 16$ to about 3 milli arcsec. Then, a large number of stars can be resolved revealing their shape that is not necessarily circular (see front page with the reconstructed shape of Achernar on the left). Thus, with spatial interferometry, stars will never be points again.

This improvement, however, comes for a price. First, one has to do many observations at different baselines, different in length and orientation, to achieve an image quality as in Fig.1. Second, in spatial interferometry as in adaptive optics one needs a bright guide star to eliminate the effects of atmospheric turbulence in order to be able to observe faint objects. Since the guide star has to be within 1 arcmin of the faint object, the number of faint objects that can be observed is limited to the immediate neighbourhood of bright stars.

Thus, while the angular resolution as one parameter of the observing performance is improved drastically with spatial interferometers, the sensitivity as the second parameter suffers, unless a bright guide star is available nearby.

In the following, we will introduce the theory of image formation in spatial interferometers and the influence of atmospheric turbulence on this process. The tasks of the experiment are (1) to measure a 'stellar' diameter and (2) the distance of a 'binary' when the stars are represented by a masked light source at a large distance. The goal is to grasp the general idea of spatial interferometry and to be able to properly interpret data taken with spatial interferometers.

3 Preliminaries and Definitions

Light as an electromagnetic wave can be represented by a dimensionless scalar $v(\vec{r}, t)$, called the *optical disturbance* that is proportional to one component, e.g. E_x , of the electric field vector \vec{E} of the electromagnetic wave. The orthogonal component, E_y , can be treated independently. It is customary to write $v(\vec{r}, t)$ as a complex quantity. However, the physically relevant part representing the electromagnetic wave is the real part.

A monochromatic, plane wave propagating along the z-axis of a cartesian coordinate system is then

described by

$$v(z, t) = v_0 e^{-i(\omega t - kz)}, \quad (1)$$

with $\omega = 2\pi\nu$ and $k = 2\pi/\lambda$, when ν is the frequency and λ the wavelength of the monochromatic wave.

To discuss the propagation of light in space it is convenient to introduce the time independent dimensionless *amplitude* $V(z)$ at frequency ν so that the monochromatic optical disturbance can be written as

$$v(z, t) = V(z) e^{-i\omega t}. \quad (2)$$

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 $\vec{S} = \vec{E} \times \vec{H}$. The Poynting vector is perpendicular both to \vec{E} and \vec{H} , and it points into the direction of propagation of the electromagnetic wave, which is the z-direction in our example of a plane wave.

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

$$\langle S_z \rangle = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T S_z dt. \quad (3)$$

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 W/m^2 . The measurable quantity with an optical detector is the integral of the flux over the detector area, i.e. the power in units of Watt.

In the notation with the dimensionless optical disturbance $v(z, t)$ we define the dimensionless *intensity* as the time average of the product vv^*

$$I(z) := \langle v(z, t)v^*(z, t) \rangle = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T v(z, t)v^*(z, t) dt = v_0^2. \quad (4)$$

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.

In terms of the amplitude $V(z)$ the intensity, using (2) and (4), is written as

$$I(z) := |V(z)|^2. \quad (5)$$

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* $V(z, \nu)$ (compare to (2)) such that it is

$$v(z, t) = \int_0^\infty V(z, \nu) e^{-i2\pi\nu t} d\nu. \quad (6)$$

$V(z, \nu)$ has the dimension of Hz^{-1} .

The *spectral intensity* – sometimes called the *power spectral density* – $I(\vec{r}, \nu)$ is defined such that it is

$$I(z) = \int_0^\infty I(z, \nu) d\nu. \quad (7)$$

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

Hint: With the polychromatic intensity $I(z)$ being the integral of the spectral intensities $I(z, \nu)$, 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. This is sometimes helpful if the number of integrals starts being overwhelming.

4 Young's Experiment

This is the classical diffraction experiment named after Thomas Young who conducted it in 1802, providing the experimental cornerstone for the demonstration of the wave nature of light. In its simplicity it is perfectly suited to explain the concept of spatial interferometry.

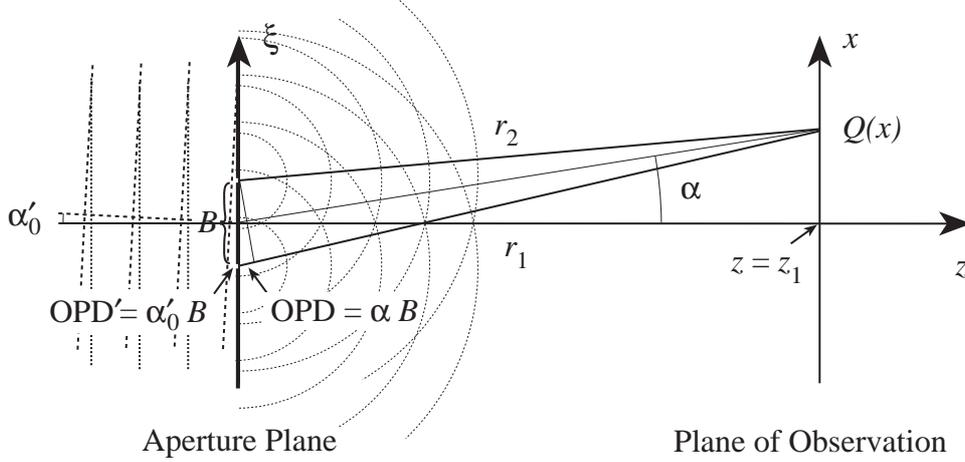


Figure 2: The geometry of Young's experiment in a plane across the pinholes. A monochromatic plane wave from an on-axis point at a large distance illuminates the two pinholes separated by a distance B , called the *baseline*, in the aperture plane. For small angles, it is $\alpha = x/z_1$ with x the coordinate of Q , the point of observation. The difference $r_1 - r_2$ is called the optical path difference (OPD), with $\text{OPD} = \alpha B$ for small α . The OPD is related to the difference in arrival time τ between the light from the two pinholes by $\tau = \text{OPD}/c$. A plane wave from a point source at α'_0 is also displayed.

Young illuminated a screen with two pinholes by light from a single point source at a large distance. By passing through the pinholes the light was diffracted and the waves from the two pinholes interfered. On a second screen, in the plane of observation, the diffraction pattern could be observed showing the characteristic fringe pattern. In Fig. 2 the experiment is depicted schematically for an illuminating point source at very large distance from the screen so that an approximately plane wave, with $V(\xi) = V_0$, illuminates the two pinholes. A plane wave from a light source at an angle α'_0 is also displayed.

The amplitude V at point Q as a function of diffraction angle α in the plane of observation at distance z_1 is then the sum of the two spherical waves originating from the pinholes,

$$\begin{aligned} V(\alpha) &= \frac{V_0}{r_1} e^{ikr_1} + \frac{V_0}{r_2} e^{ikr_2} \\ &= \frac{V_0}{z_1} e^{ik(r_1+r_2)/2} 2 \cos(k(r_1 - r_2)/2). \end{aligned} \quad (8)$$

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 of the spherical waves. The difference between the optical path lengths, $r_1 - r_2$, is called the optical path difference, OPD.

The **intensity** is the squared modulus of the amplitude,

$$I(\alpha) = |V(\alpha)|^2 = \left(\frac{V_0}{z_1}\right)^2 2 \left(1 + \cos(k(r_1 - r_2))\right) = I_0 \left(1 + \cos(k\alpha B)\right), \quad (9)$$

with $I_0 = 2\left(\frac{V_0}{z_1}\right)^2$. This intensity distribution is called the *fringe pattern*, when the fringe spacing, the distance between the maxima of the cosine function, is λ/B (see Fig. 3a). The notation for the fringe spacing in milli arcsec (mas), $\alpha/\text{mas} = 206 \frac{\lambda/\mu\text{m}}{B/\text{m}}$, with λ in μm is sometimes quite handy.

The OPD, $r_1 - r_2 = \alpha B$, is at the same time a difference in arrival time called the time delay τ between the light from the two pinholes, with $\tau = \alpha B/c$ and c the speed of light. We will see later that the diffraction pattern for increasing diffraction angles α – corresponding to increasing time delays τ – is constrained by the temporal coherence of the incoming light.

For a light source at position α'_0 we have to add $\text{OPD}' = \alpha'_0 B$ to $r_1 - r_2$ obtaining the total OPD $(\alpha + \alpha'_0)B$. This results in a fringe pattern that is shifted by $-\alpha'_0$, yielding

$$I(\alpha) = I_0 \left(1 + \cos(k(\alpha + \alpha'_0)B) \right). \quad (10)$$

The **visibility**, \mathcal{V} , is defined as the contrast of the fringe pattern. The monochromatic fringe patterns (10) have an excellent contrast since the intensity oscillates between 0 and 1. This can be expressed more formally by defining the visibility as

$$\mathcal{V} = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}. \quad (11)$$

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

Spatial 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 plane monochromatic wave illuminating the screen with the two pinholes and, thus, the light 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 but only a homogeneous illumination as a result of the diffraction of light at each individual aperture. Then, the light is called incoherent.

It should be emphasised that by measuring the contrast of the fringe pattern in the plane of observation we determine the coherence in the aperture plane with the two pinholes. Although this sounds trivial at this stage we will see later that this has important repercussions.

Temporal coherence

First, we 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 (9) at different frequencies to obtain the observed intensity distribution. This is displayed in Fig. 3b for the *K*-band,¹ with $2.2 \pm 0.2 \mu\text{m}$, $\Delta\lambda = 0.4 \mu\text{m}$ and a baseline B 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 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 (black curve in Fig. 3b) for increasing diffraction angles α . Since α is related to the difference in arrival time, the time delay τ , through $\tau = \alpha B/c$, this effect could be reformulated by stating that the contrast of the resulting fringe pattern is reduced with increasing τ . That time delay that is related to the quasi loss of fringe contrast is called the coherence time τ_c , which is proportional to the reciprocal of the spectral bandwidth $\Delta\nu$. Consequently, the coherence length is defined as $l_0 = c/\tau_0$. The exact relationship depends on the form of the spectral band.

The number of fringes visible in a fringe pattern is thus limited to about $2\lambda/\Delta\lambda$ since the diffraction angle when the fringes have faded is $\alpha = \pm\tau_c c/B = \pm 1/\Delta\nu \lambda\nu/B = \pm\nu/\Delta\nu \lambda/B$. With a fringe spacing of λ/B , and $\Delta\nu/\nu = \Delta\lambda/\lambda$ we arrive at a number of $2\lambda/\Delta\lambda$ fringes.

¹The atmosphere transmits only certain bands in the infrared. One frequently used band in the near infrared is the *K*-band at $2.2 \pm 0.2 \mu\text{m}$. Most of the numerical examples will be given for this band.

Spatial coherence

While the spectral bandwidth affects the temporal coherence, the size of the light source affects the fringe contrast and, thus, the spatial coherence of the light in the aperture plane. First we take an assumption on the nature of the light source. We assume that the source is spatially incoherent, meaning that each point on its surface radiates independent, i.e. uncorrelated, from its neighbour. A thermal source like a star is a typical example of an incoherent source. Then we compute the monochromatic fringe pattern for each source point using (10). Having an incoherent light source we add up the individual fringe patterns, i.e. the intensity distributions, that are slightly shifted against each other. The result is a fringe pattern (black curve) with reduced contrast as displayed in Fig. 3c for $\lambda = 2.2 \mu\text{m}$.

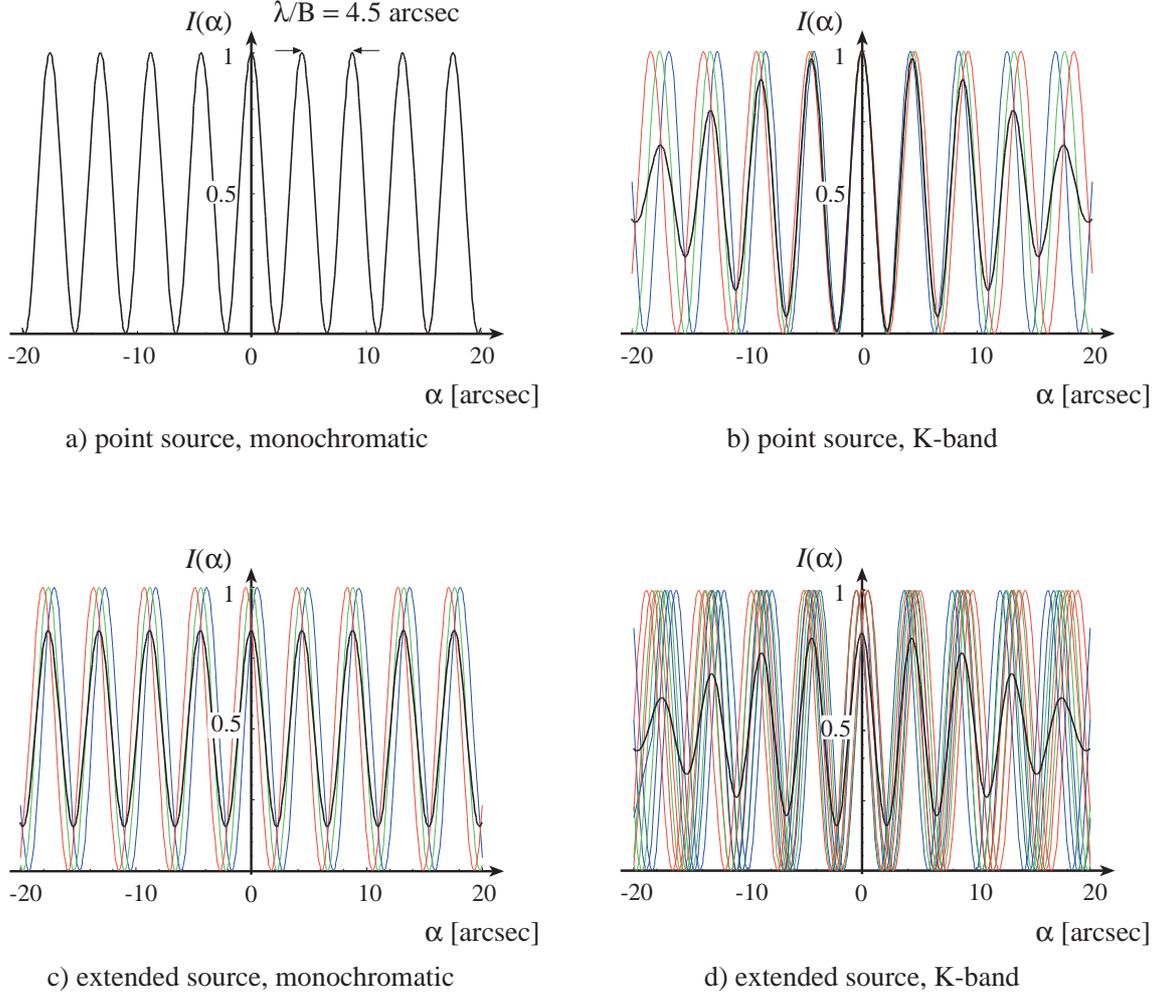


Figure 3: Summary of the influence of source size and spectral bandwidth on the fringe pattern. The pinhole separation, the baseline B , is 10 cm in all figures. In (a) an individual fringe pattern for an observing wavelength of $\lambda = 2.2 \mu\text{m}$ and a point source is displayed. In (b) the K -band fringe pattern is shown (black curve) when observing a point source. In (c) the monochromatic illumination of a source with diameter 2 arcsec produces a fringe pattern (black curve) with reduced contrast. In (d) the resulting fringe pattern (black curve) 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 α due to the finite spectral bandwidth.

To compute this result, we write the fringe pattern (10) as a function both of the diffraction angle α and of the position of the source point α' , $I(\alpha, \alpha') = I_0(1 + \cos(k(\alpha + \alpha')B))$, yielding the resulting fringe pattern for a source of diameter α'_0 as

$$\begin{aligned} I(\alpha) &= \int_{-\alpha'_0/2}^{\alpha'_0/2} I(\alpha, \alpha') d\alpha' \\ &= \int_{-\alpha'_0/2}^{\alpha'_0/2} I_0 d\alpha' + \int_{-\alpha'_0/2}^{\alpha'_0/2} I_0 \cos(k(\alpha + \alpha')B) d\alpha' \\ &= \int I(\alpha') d\alpha' + \int I(\alpha') \cos(k(\alpha + \alpha')B) d\alpha', \end{aligned} \quad (12)$$

with $I(\alpha')$ the source intensity distribution of the source with diameter α'_0 , incorporating the integration limits. The second term is the real part of the complex Fourier transform $\int I(\alpha') \exp(-ik(\alpha + \alpha')B) d\alpha'$. We extract $\exp(-ik\alpha B)$ since this term does not depend on α' . The remaining expression $\int I(\alpha') \exp(-ik\alpha' B) d\alpha'$ is the Fourier transform of the source intensity distribution $I(\alpha')$.

In Young's experiment, the baseline B is the distance between the pinholes. Discussing now a Fourier transform between the coordinate spaces α' and B we have to interpret B as a coordinate that is the distance between any two points in the aperture plane.

In two-dimensional planes, we use coordinate vectors, $\vec{\alpha}'$ in the source plane and \vec{B} in the plane of the difference coordinates.

Denoting the integral over the source intensity by $I_0 = \int I(\vec{\alpha}') d\vec{\alpha}'$, and the Fourier transform of the source intensity by $\mu(\vec{B})$, we obtain the fringe pattern of an extended source as

$$\begin{aligned} I(\vec{\alpha}) &= I_0 \left(1 + \text{Re} \left(\mu(\vec{B}) e^{-ik\vec{\alpha} \cdot \vec{B}} \right) \right) \\ &= I_0 \left(1 + |\mu(\vec{B})| \cos(\phi(\vec{B}) - k\vec{\alpha} \cdot \vec{B}) \right). \end{aligned} \quad (13)$$

The contrast, i.e. the visibility, of the fringe pattern for a pair of pinholes separated by \vec{B} is determined by $|\mu(\vec{B})|$. This is why $\mu(\vec{B})$ is called the *visibility function* – in general a complex function with phase $\phi(\vec{B})$. It is straightforward to see that for $|\mu(\vec{B})| = 1$ the fringe contrast is 1 and that for a value of zero there is no fringe pattern, i.e. the contrast is zero. The phase $\phi(\vec{B})$ defines the position of the central fringe, called the *white-light fringe*.

Observing with a spatial interferometer, it is exactly the two quantities fringe visibility $|\mu(\vec{B})|$ and its phase $\phi(\vec{B})$, that we are chasing.

The definition of the visibility function:

$$\mu(\vec{B}) = \frac{\int I(\vec{\alpha}') e^{-ik\vec{\alpha}' \cdot \vec{B}} d\vec{\alpha}'}{I_0} \quad (14)$$

is also known as the *van-Cittert-Zernike theorem*. We have deduced it here in a heuristic manner by interpreting the result of Young's experiment. One can also base the deduction on the general theory of the propagation of light and on the propagation of its statistical properties described by coherence functions. We will treat this topic in the next section.

Regarding (14), it is obvious that the visibility function $\mu(\vec{B})$ is a complex function. Due to the intensity distribution $I(\vec{\alpha}')$ being real and positive by definition, the modulus $|\mu|$ of the visibility is symmetric and its phase ϕ is anti-symmetric:

$$\begin{aligned} |\mu(\vec{B})| &= |\mu(-\vec{B})| \text{ and} \\ \phi(\vec{B}) &= -\phi(-\vec{B}). \end{aligned}$$

The two formulae (13) and (14) are the cornerstones of spatial interferometry. They link the intensity distribution in the object $I(\vec{\alpha}')$ through a Fourier transform, (14), to the visibility of the fringe pattern

(13). The inversion of this argument – measuring the visibilities for many different baseline vectors \vec{B} , and reconstructing the object intensity $I(\vec{\alpha}')$ through a Fourier back transform of $\mu(\vec{B})$ – was the important step to take spatial interferometry from the mere determination of stellar diameters to an imaging tool of extremely high angular resolution.

Figure 3d shows the fringe pattern for a star with a diameter of 2 arcsec as in Fig. 3c and the K -band spectrum as in Fig. 3b. The visibility is reduced at $\vec{\alpha} = 0$ because of the finite size of the source and goes further down for increasing diffraction angle $\vec{\alpha}$. This mixture of spatial and temporal coherence is unavoidable and sets tight constraints to the data processing since the contrast variation due to the shape of the source has to be clearly separated from the contrast variation due to the spectral bandwidth. In the following section, the quasi-monochromatic approximation will be introduced to deal with this problem.

Hint: *If the Fourier transform of an intensity distribution $I(\vec{\alpha}')$ as a function of a coordinate difference \vec{B} is too abstract one can always look at the situation in terms of 'fringe pattern per source point and per wavelength' and then add up the intensity distributions of these fringe patterns as in (12), if the source (not the light in the aperture plane!) is spatially incoherent. The same argument was used in the Sect. 3 when monochromatic intensity distributions were regarded and then summed up to obtain the polychromatic intensity. This illustrates the equivalence of temporal and spatial coherence that we have discussed here heuristically. In the next section, the general coherence theory will be used to explain the phenomena.*

5 The Visibility and the Mutual Coherence Function

In the context of Young's experiment we introduced the coherence as a phenomenon related to the contrast of the fringe pattern. Here, we will use a more general definition based on statistical properties. Before defining the coherence function, the nature of this random process needs to be discussed. So far, the optical disturbance v has been regarded as a plane or as a spherical wave providing a deterministic signal throughout the propagation and diffraction process. 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 wavefront² 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 $v(\vec{x}, t)$ in a given plane at distance z , that is fed by light from individual, independently radiating, polychromatic 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 the same.

We will regard the individual wavefronts as possible realizations or members of the ensemble of the random process, and the optical disturbance $v(\vec{x}, t)$ as the random variable. 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 all 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 wavefront in 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 wavefront 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

²A *wavefront* is the virtual surface of the same phase of a propagating wave.

over a sufficient number of realisations is defined as $T \gg 1/\nu$, i.e. much longer than the oscillation of the electromagnetic wave at the average frequency.

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

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

Due to the ergodic statistical process, the MCF is defined as a time average:

$$\Gamma(\vec{x}_1, \vec{x}_2, \tau) := \langle v(\vec{x}_1, t + \tau)v^*(\vec{x}_2, t) \rangle \quad (15)$$

$$= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T v(\vec{x}_1, t + \tau)v^*(\vec{x}_2, t) dt, \quad (16)$$

where \vec{x}_i are the coordinate vectors and τ is the time difference.

With (4), it is straightforward to see that the intensity can be expressed by

$$I(\vec{x}) = \Gamma(\vec{x}, \vec{x}, 0). \quad (17)$$

In the case of **incoherent light** the correlation between the optical disturbances is zero, they are completely independent. The MCF on the surface of a thermal source for instance is then given by

$$\Gamma(\vec{x}_1, \vec{x}_2, \tau) = I(\vec{x}_1)\delta(\vec{x}_1 - \vec{x}_2)\delta(\tau), \quad (18)$$

when $\delta(\cdot)$ denotes the *Dirac function* having zero value unless the argument is zero. In this case, illustrating the independence of the optical disturbances, the correlation is zero unless the value at the same position ($\vec{x}_1 = \vec{x}_2$) and at the same moment in time ($\tau = 0$) is taken. Thus, an incoherent source is completely described by its intensity.

The other extreme is that of **coherent light**, when the optical disturbances are well defined – for instance a plane wave – throughout space and time, and the MCF can be replaced by the product of the optical disturbances,

$$\Gamma(\vec{x}_1, \vec{x}_2, \tau) = v(\vec{x}_1, t + \tau)v^*(\vec{x}_2, t) = V(\vec{x}_1)V^*(\vec{x}_2)e^{-2\pi\nu_0\tau}. \quad (19)$$

Using (2), the optical disturbances are replaced by the amplitudes V yielding the MCF as the product of amplitudes at positions \vec{x}_1 and \vec{x}_2 multiplied by the complex exponential as a function of the time difference τ .

The MCF in the immediate neighbourhood of the light source, e.g. close to the surface of a star, is different from the MCF at a very large distance, for instance some light-years away. This means that the statistical properties of the light, its coherence, change while the light propagates. In this sense, we speak of the propagation of the MCF keeping in mind that it is actually the electromagnetic wave that propagates.

The **propagation of the MCF** in space is described by applying a formalism that was derived from the Rayleigh-Sommerfeld diffraction formula. To simplify the propagation process, we assume that the light sources (e.g. stars) are incoherent and that all involved angles are small. The latter is called the Fresnel approximation. In addition, we apply the *quasi-monochromatic approximation* when (1) the spectral bandwidth $\Delta\nu$ is assumed to be much smaller than the average frequency ν_0 and (2) the time difference τ is much smaller than $1/\Delta\nu$.

We can now write the MCF in the aperture plane as a function of the coordinate difference, \vec{B} , of two points as

$$\Gamma_{\text{qm}}(\vec{B}, \tau) = G(\nu_0) \int I(\vec{\alpha}') e^{-ik_0\vec{B}\cdot\vec{\alpha}'} d\vec{\alpha}' e^{-i2\pi\nu_0\tau}, \quad (20)$$

with $G(\nu_0)$ the source spectrum at ν_0 .

It is very important to note that, due to the light source being incoherent, the MCF only depends on one coordinate, \vec{B} , which is the coordinate difference vector between two points in the aperture plane. The absolute positions of the two points are not relevant.

Not very surprisingly, the MCF in quasi-monochromatic approximation, (20), is very similar to the MCF of coherent light, (19). However, while the dependence on the time difference τ is exactly the same – as long as τ is much smaller than $1/\Delta\nu$ – the dependence on the difference coordinate \vec{B} is different. The latter, describing the spatial coherence, depends on the angular shape of the source $I(\vec{\alpha}')$. Therefore, the form of the MCF with respect to the spatial coordinate \vec{B} varies with the distance of the light source from the aperture plane, while it is invariant with respect to the temporal coordinate τ . This is also true if we go beyond the quasi-monochromatic approximation.

Thus, the spatial coherence of the light changes when it propagates in space while the temporal coherence remains untouched.

Regarding the MCF at $\tau = 0$,

$$\Gamma_{\text{qm}}(\vec{B}, 0) = G(\nu_0) \int I(\vec{\alpha}') e^{-ik_0 \vec{B} \cdot \vec{\alpha}'} d\vec{\alpha}', \quad (21)$$

and calibrating it by the average intensity I_0 we obtain the van Cittert-Zernike theorem (14) describing the propagation of the mutual coherence function from the source plane into the aperture plane by Fourier transforming the intensity distribution $I(\vec{\alpha}')$ to obtain the complex visibility function μ :

$$\mu(\vec{B}) = \frac{\Gamma_{\text{qm}}(\vec{B}, 0)}{I_0} = \frac{G(\nu_0) \int I(\vec{\alpha}') e^{-ik_0 \vec{B} \cdot \vec{\alpha}'} d\vec{\alpha}'}{G(\nu_0) \int I(\vec{\alpha}') d\vec{\alpha}'}, \quad (22)$$

with \vec{B} the difference coordinate vector in the aperture plane, $\vec{\alpha}'$ the angular coordinate vector in the source plane, ν_0 the average frequency, and $I(\vec{\alpha}')$ the source intensity distribution.

By rigorous application of the coherence theory we have arrived at the same formula, the van Cittert-Zernike theorem, for the propagation of the MCF, as by discussing Young's experiment in Sect. 4.

We can now proceed **from the aperture plane into the plane of observation** by computing the propagation of the MCF using again the Rayleigh-Sommerfeld diffraction formula. Since we only measure the intensity, we reduce the MCF in the plane of observation to the intensity, yielding

$$\begin{aligned} I(\vec{\alpha}) &= \left(I_0 + \text{Re}[\Gamma_{\text{qm}}(\vec{B}, \tau)] \right) \\ &= \left(I_0 + \text{Re}[\Gamma_{\text{qm}}(\vec{B}, 0) e^{-i2\pi\nu_0\tau}] \right) \\ &= I_0 \left(1 + \text{Re}[\mu(\vec{B}) e^{-i2\pi\nu_0\tau}] \right) \\ &= I_0 \left(1 + |\mu(\vec{B})| \cos(\phi(\vec{B}) - 2\pi\nu_0\tau) \right), \end{aligned} \quad (23)$$

with $\nu_0\tau = \vec{\alpha} \cdot \vec{B} / \lambda$. We found the same expression for the fringe pattern in the plane of observation when discussing Young's experiment in monochromatic illumination (13). We note again, that by measuring the intensity distribution of the fringe pattern in the plane of observation we obtain information on the MCF $\Gamma_{\text{qm}}(\vec{B}, \tau)$ in the aperture plane.

The fringe pattern in monochromatic illumination is displayed in Fig. 3c where the fringe pattern has a constant visibility for all diffraction angles. Formally, (23) also describes a fringe pattern that does not lose contrast with increasing diffraction angle α resp. τ . However, we have to keep in mind that the quasi-monochromatic approximation is restricted to very small τ only. Regarding the K -band fringe pattern in Fig. 3d this can be explained. The K -band, $2.2 \pm 0.2 \mu\text{m}$, with $\Delta\lambda = 0.4 \mu\text{m}$ fulfils the condition of a narrow bandwidth. As long as we stay on the central fringe, i.e. very small diffraction angle and very small τ , we can attribute its contrast loss in very good approximation to the spatial coherence only (see

Fig. 3d). For large α , the fringe pattern can no longer be described by (23) since the contrast loss is not accounted for in the quasi-monochromatic approximation.

Thus, the contrast and the position of the central fringe are very good approximations for the visibility $|\mu(\vec{B})|$ and its phase $\phi(\vec{B})$ given by the Fourier transform of the source intensity $I(\vec{\alpha}')$ as stated by the van Cittert-Zernike theorem.

The two pinholes can be regarded as an instrument to measure the coherence properties by probing the wavefront 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 out how to relate a measurable quantity to the complex visibility function. Second, by doing so, we derive the visibility function in the aperture plane from characteristics of the intensity distribution (fringe visibility and centre fringe position) in the plane of observation.

Hint: *Don't be overwhelmed by all the theory. What remains from the full framework of propagating coherence functions is (1) an incoherent source that is fully described by its intensity, (2) the Fourier transform of this intensity distribution that provides the coherence function in the aperture plane, and (3) again the intensity in the plane of observation since we cannot measure anything else.*

The important step is to conclude from the intensity distribution of the diffraction pattern, namely its visibility, on the coherence function in the aperture plane. Thus, we only deal with the coherence function in the aperture and regard intensities elsewhere.

The important point is the connection between the shape of the object intensity distribution and the visibility for varying baselines.

5.1 Example: Venus

An example illuminates the situation: we model Venus as a uniform disk with angular diameter α'_0 varying between about 15 and 45 arcsec depending on the mutual positions of Venus and Earth. The circular intensity distribution is described by the *circ-function* that is defined as $\text{circ}(\frac{|\vec{x}|}{R}) = 1$ if $|\vec{x}| \leq R$ and 0 elsewhere. A circular intensity distribution is then represented by $I(\vec{\alpha}') = (\pi(\alpha'_0/2)^2)^{-1} \text{circ}(\frac{\alpha'}{\alpha'_0/2})$ with $\alpha' = |\vec{\alpha}'|$.

We compute the visibility function as the Fourier transform of $I(\vec{\alpha}')$ yielding

$$\begin{aligned} \mu(\vec{B}) &= \frac{1}{\pi(\alpha'_0/2)^2} \int \text{circ}\left(\frac{\alpha'}{\alpha'_0/2}\right) e^{-ik_0\vec{B}\cdot\vec{\alpha}'} d\vec{\alpha}' \\ &= \text{Besinc}(k_0B\alpha'_0/2), \end{aligned} \quad (24)$$

with the *Besinc-function* defined as $\text{Besinc}(x) = 2 J_1(x)/x$, and $J_1(x)$ the first order *Bessel function*.

Figure 4 displays the visibility function for Venus' smallest angular diameter of $\alpha'_0 = 15$ arcsec as a function of the baseline in mm. The first zero of the visibility function is reached at $B = 1.22\lambda/\alpha'_0 = 37$ mm, or $B/\text{mm} = 252 \frac{\lambda/\mu\text{m}}{\alpha'_0/\text{arcsec}}$. For longer baselines the visibility function slowly oscillates between negative and positive values with decreasing amplitude.

One can now understand the effect of a visibility function that is shaped like a Besinc-function (24). The visibility function shows up in the varying visibility of the fringe patterns in Fig. 4. Each horizontal line shows a fringe pattern for an individual pinhole separation B whose visibility is determined by the modulus $|\mu(\vec{B})|$ 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 considered by a phase of $\phi = \pi$, since $|\mu(\vec{B})|e^{i\pi} = -|\mu(\vec{B})|$. The fringe pattern is shifted by π , 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. 4).

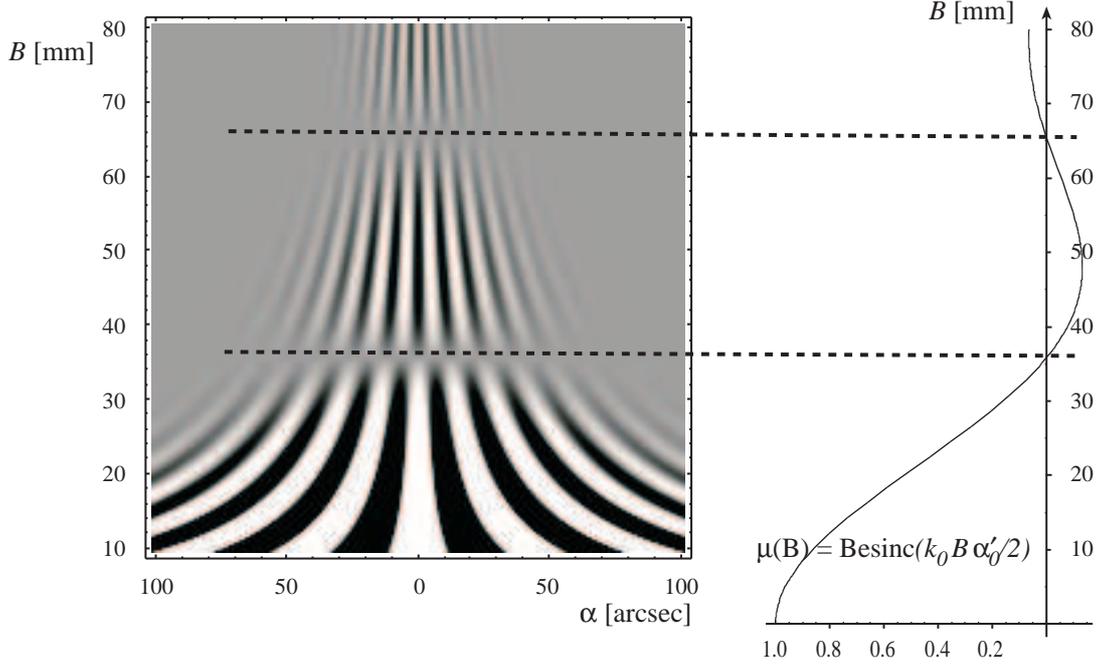


Figure 4: Fringe patterns of Venus in Young's experiment. On the left, the intensity distributions along horizontal lines display the fringe pattern for pinhole separations $B = |\vec{B}|$ between 10 and 80 mm. The fringe spacing is λ/B . The light source illuminating the pinholes is Venus with a uniform disk diameter of 15 arcsec. The spectral band is the K -band ($2.2 \pm 0.2 \mu\text{m}$), i.e. $\lambda/\Delta\lambda = 5.5$. The finite spectral bandwidth makes the fringe visibility disappear for diffraction angles $\alpha = |\vec{\alpha}|$ larger than about $5.5\lambda/B$ i.e. there are about 11 fringes in the fringe pattern. The fringe spacing decreases with increasing pinhole separation, and the fringe visibility is reduced to zero at $B = 1.22\lambda/\alpha'_0 = 37$ mm. For $37 \text{ mm} \leq B \leq 67.5$ mm, the fringe pattern inverts its sign displaying a black fringe at $\alpha = 0$. The visibility function $\mu(\vec{B})$ as a function of pinhole separation B follows a Besinc-function that is displayed on the right.

6 Image Formation

While Young's experiment is a useful tool to demonstrate coherence effects, in practise the apertures of a spatial interferometer are considerably larger than pinholes, keeping in mind that sensitivity is one of the performance parameters of an astronomical observation.

Before discussing the case of combining the light from two individual telescopes we will discuss the case of image formation in a single telescope.

6.1 Image Formation in a Single Telescope

We start again by regarding the light from a single point source at a large distance on axis. A plane wave with amplitude $V(\vec{\xi})$ arrives at the aperture $A(\vec{\xi})$, with $\vec{\xi}$ the coordinate vector in the aperture plane. We denote the wave leaving the aperture by $V_{\text{ap}}(\vec{\xi}) = V(\vec{\xi})A(\vec{\xi})$. The light is diffracted at the aperture producing the *Fraunhofer diffraction* pattern in the plane of observation that we can compute by

$$V(\vec{\alpha}) = \int V_{\text{ap}}(\vec{\xi}) e^{-ik\vec{\alpha}\cdot\vec{\xi}} d\vec{\xi}. \quad (25)$$

Fraunhofer diffraction can be observed at a very large distance behind the aperture. If we place a lens in the aperture we find the Fraunhofer diffraction pattern in the focal plane of the lens at a distance f

from the aperture. This is the situation that we have in optical systems. The intensity distribution of the diffraction pattern is given by $I(\vec{\alpha}) = |V(\vec{\alpha})|^2$.

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. Since the connection between the aperture of the optical system and its PSF is given by a Fourier transform, the formalism described in this section is also called *Fourier optics*.

For a circular aperture as the most common case of a telescope aperture with diameter D we write $A(\vec{\xi}) = \text{circ}\left(\frac{\xi}{D/2}\right)$, $\xi = |\vec{\xi}|$, with the area of the circular aperture given by $A_0 = \pi(D/2)^2$. $A(\vec{\xi})$ is illuminated by a point source at infinity with $V(\vec{\xi}) = V_0$ in the aperture plane, setting $V_0 = 1$. Using (25), the diffraction limited amplitude in the focal plane can be written as

$$\begin{aligned} V(\vec{\alpha}) &= \int \text{circ}\left(\frac{\xi}{D/2}\right) e^{-ik\vec{\alpha}\cdot\vec{\xi}} d\vec{\xi} \\ &= A_0 \text{Besinc}(k\alpha D/2), \end{aligned} \quad (26)$$

with $\alpha = |\vec{\alpha}|$. The result of the Fourier transform of the circ-function is the Besinc-function.

In telescopes, the intensity distribution of the diffraction limited PSF is called the *Airy disk*. The squared modulus of the amplitude $V(\vec{\alpha})$ yields the Airy disk as

$$\text{PSF}(\vec{\alpha}) = V(\vec{\alpha})V^*(\vec{\alpha}) = A_0^2 \text{Besinc}^2(k\alpha D/2). \quad (27)$$

The first minimum of the PSF is at $\alpha_{\min} = 1.22 \lambda/D$, or $\alpha_{\min}/\text{mas} = 252 \frac{\lambda/\mu\text{m}}{D/\text{m}}$, with α in milli arcsec (mas), the observing wavelength λ in μm and the telescope diameter D in m. For a binary star with a separation α_{\min} , the resulting image, which is the sum of two individual Airy disks, shows a local minimum between the peaks of the Airy disk. Therefore the two stars of the binary can be identified as individual objects in the image. This criterion of angular resolution, when the smallest resolvable angle is α_{\min} , is called the *Rayleigh criterion* of resolution of a telescope.

Discussing the visibility function in the last section, we found a Besinc function as visibility function in the aperture plane assuming that the light source at a large distance was circular. Here, the Besinc function describes the intensity distribution in the focal plane of a telescope, assuming that a point source illuminates a circular aperture, an altogether different physical quantity. One has to carefully consider which quantity is discussed in order to avoid confusion.

The **imaging process** can now be described by summing up the PSF of each individual object point – all slightly shifted according to the position of the object point – assuming that the object is incoherent and, thus, the intensities add up. In Sect. 4 discussing the loss in fringe contrast in Young’s experiment when illuminating the pinholes with an extended object we used the same principle.

Denoting the object intensity distribution by $O(\vec{\alpha}')$ we obtain the image intensity distribution as the integral of individual PSF weighted by the object intensity at each object point:

$$\begin{aligned} I(\vec{\alpha}) &= \int O(\vec{\alpha}') \text{PSF}(\vec{\alpha} - \vec{\alpha}') d\vec{\alpha}' \\ &= O(\vec{\alpha}) * \text{PSF}(\vec{\alpha}). \end{aligned} \quad (28)$$

This operation is a convolution of the object intensity $O(\vec{\alpha})$ with the PSF, denoted by $*$.

If the object has some very fine structure a rather broad PSF might wash out the detail. Or, if there is a binary star with a separation smaller than the Rayleigh limit of $1.22\lambda/D$, the image intensity $I(\vec{\alpha})$ is undistinguishable from a single star. By the same token, one cannot determine the diameter of a star if its disk is much smaller than the PSF. Thus the width of the PSF, or its reciprocal the diameter D of the aperture, determine the angular resolution. We will see in the following that combining the light from two apertures increases the limit of resolution proportional to the baseline that is much larger than the individual aperture.

6.2 Image Formation in a Spatial Interferometer

We begin by regarding an aperture that consists of **two pinholes** like in Young's experiment. This can be described by two δ -functions at ξ_p and at $-\xi_p$:

$$A(\vec{\xi}) = \delta(\vec{\xi} - \vec{\xi}_p) + \delta(\vec{\xi} + \vec{\xi}_p).$$

In Young's experiment we let a plane wave pass through the pinholes and we computed the diffraction pattern by having two spherical waves propagate from each pinhole. In the formalism of the imaging process we use (25) to calculate the aperture in the focal plane as

$$\begin{aligned} V(\vec{\alpha}) &= \int (\delta(\vec{\xi} - \vec{\xi}_p) + \delta(\vec{\xi} + \vec{\xi}_p)) e^{-ik\vec{\alpha}\cdot\vec{\xi}} d\vec{\xi} \\ &= e^{-ik\vec{\alpha}\cdot\vec{\xi}_p} + e^{+ik\vec{\alpha}\cdot\vec{\xi}_p} = 2 \cos(k\vec{\alpha} \cdot \vec{\xi}_p). \end{aligned}$$

The intensity distribution of this diffraction pattern shows the familiar fringe pattern $I(\vec{\alpha}) = 4 \cos^2(k\vec{\alpha} \cdot \vec{\xi}_p) = 2(1 + \cos(k\vec{\alpha} \cdot \vec{B}))$ with $\vec{B} = 2\vec{\xi}_p$ the baseline vector, like in (9).

Enlarging the pinholes to small **sub-apertures with diameter** D , we write the aperture as a convolution of a sub-aperture with two δ -functions,

$$A(\vec{\xi}) = a(\vec{\xi}) * (\delta(\vec{\xi} - \vec{\xi}_p) + \delta(\vec{\xi} + \vec{\xi}_p)),$$

yielding the amplitude in the focal plane as

$$V(\vec{\alpha}) = \int a(\vec{\xi}) * (\delta(\vec{\xi} - \vec{\xi}_{p1}) + \delta(\vec{\xi} - \vec{\xi}_{p2})) e^{-ik\vec{\alpha}\cdot\vec{\xi}} d\vec{\xi}.$$

This formula can be simplified since the Fourier transform of a convolution is the product of the Fourier transform of the individual functions. If we assume circular sub-apertures of diameter D we obtain again a Besinc function for the Fourier transform of $a(\vec{\xi})$, and a cosine function for the Fourier transform of $\delta(\vec{\xi} - \vec{\xi}_p) + \delta(\vec{\xi} + \vec{\xi}_p)$,

$$\begin{aligned} V(\vec{\alpha}) &= \int a(\vec{\xi}) e^{-ik\vec{\alpha}\cdot\vec{\xi}} d\vec{\xi} \times \int (\delta(\vec{\xi} - \vec{\xi}_p) + \delta(\vec{\xi} + \vec{\xi}_p)) e^{-ik\vec{\alpha}\cdot\vec{\xi}} d\vec{\xi} \\ &= A_0 \text{Besinc}(k\alpha D/2) \times 2 \cos(k\vec{\alpha} \cdot \vec{B}/2). \end{aligned} \quad (29)$$

The intensity distribution in the focal plane is given by $|V(\vec{\alpha})|^2$. Since this is the intensity distribution for a point source at infinity, we call it the **PSF of the interferometer**:

$$\text{PSF}(\vec{\alpha}) = |V(\vec{\alpha})|^2 = 2A_0^2 \text{Besinc}^2(k\alpha D/2) (1 + \cos(k\vec{\alpha} \cdot \vec{B})). \quad (30)$$

The PSF consists of the Besinc-function of an individual sub-aperture that is multiplied by the fringe pattern of two pinholes. Since the Besinc-function is wider than the fringe pattern, the PSF looks like a fringe pattern with the Besinc-function as an envelope, see Fig. 5. If the sub-apertures were infinitely small, the Besinc-function would be infinitely wide and we would see a fringe pattern like in Fig. 3.

Before writing down the final image intensity distribution we have to distinguish two cases, that of a large object – larger than the Besinc-function of the PSF – and that of a small object that looks like a point-source when observing it with an individual telescope.

Assuming we observe **a binary star that is clearly resolved by an individual telescope**, we would see two distinct PSF, the image of this binary, in the image plane. If we observe this binary with an interferometer combining two of these telescopes, we would again see two PSF but now each PSF has fringes as described by (30). Figure 6 shows the image intensity distribution of a stellar cluster, with each PSF displaying a fringe pattern.

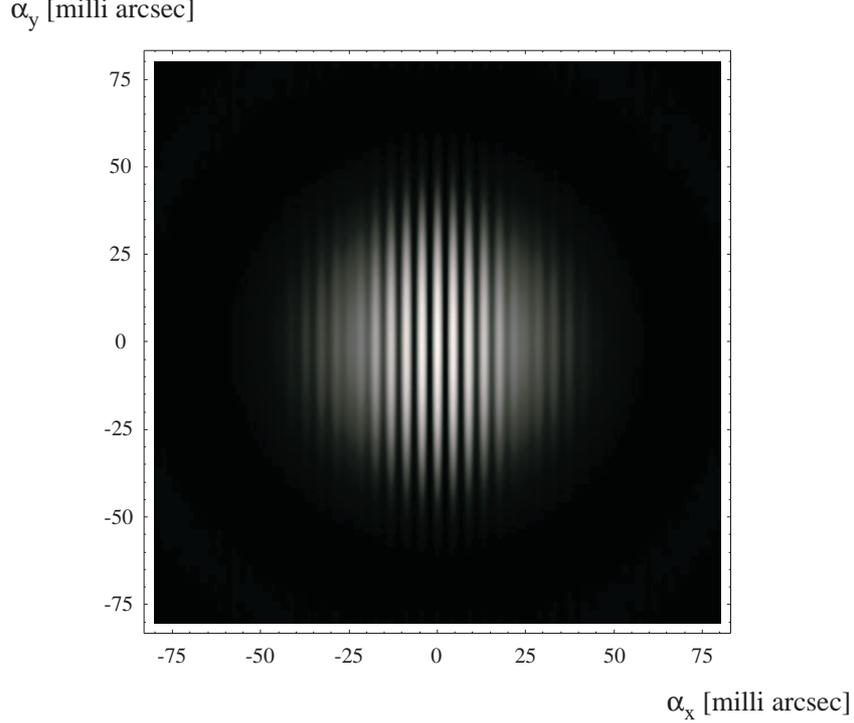


Figure 5: The PSF of a spatial interferometer (30) with two sub-apertures of $D = 8$ m and a baseline B of 100 m in the K-band, with $\lambda = 2.2 \pm 0.2 \mu\text{m}$. The radius of the Airy disk – only the central core is displayed – is $1.22\lambda/D = 69$ milli arcsec and the fringe spacing $\lambda/B = 4.5$ milli arcsec. The consequence of observing a spectral band instead of a single spectral line is a loss of contrast after $\lambda/\Delta\lambda = 2.2/0.4 = 5.5$ periods, resp. after about ± 25 milli arcsec from the white-light fringe at the centre.

Since the fringe patterns in the two PSF do not overlap we cannot determine the influence of the binary on *the* visibility of the fringe pattern. One could also say that the information that we are seeking – the separation of the binary – is already available by the distance between the two separated PSF.

What we do find in the visibility of each individual fringe pattern is information on the size or shape of each individual star of the binary. We could observe the fringe pattern in each PSF individually ignoring the other PSF. This is in fact what is done very often in spatial interferometry, when the PSF is fed into an optical fibre. This technique has the advantage that the influence of atmospheric turbulence is greatly reduced.

Thus, the coarse detail of the object, in this case the separation of the binary, is determined by an individual telescope since the PSF is smaller than the separation. The fine detail, the shape of the individual star, which is much smaller than the PSF, cannot be determined by an individual aperture, but it is measurable through the visibility of the fringe pattern.

Using the assumption that our **object intensity distribution** $O(\vec{\alpha})$ **is much narrower than the PSF of an individual sub-aperture** we write the convolution of object intensity and PSF as

$$\begin{aligned} I(\vec{\alpha}) &= O(\vec{\alpha}) * \text{PSF}(\vec{\alpha}) \\ &\approx 2A_0^2 \text{Besinc}^2(k\alpha D/2) \left(O(\vec{\alpha}) * (1 + \cos(k\vec{\alpha} \cdot \vec{B})) \right). \end{aligned}$$

Thus, we have a convolution of the object intensity $O(\vec{\alpha})$ with a fringe pattern proportional to $1 + \cos()$ like in Young's experiment. The resulting fringe pattern is enveloped by the Besinc-function.

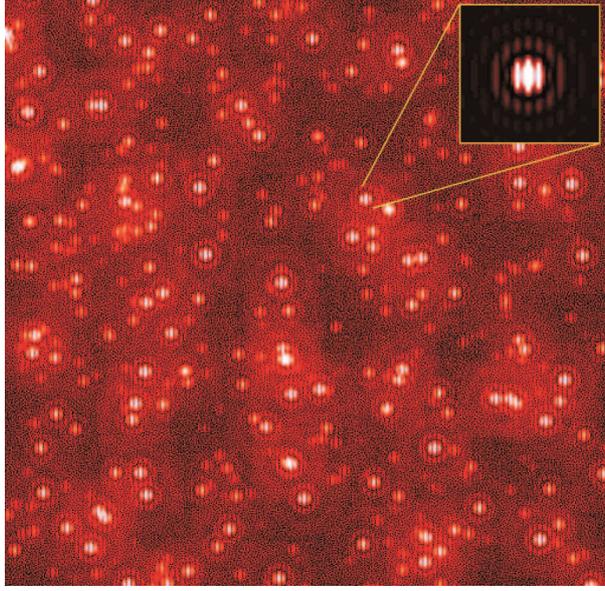


Figure 6: Simulated intensity distribution in the image of the Large Binocular Telescope (LBT) on Mt. Graham in Arizona. The parameters are: telescope aperture $D = 8.4$ m, baseline $B = 14.4$ m, $\lambda = 2.2 \pm 0.2 \mu\text{m}$ (K-band). The first minimum of the PSF is at $1.22\lambda/D = 66$ mas and the fringe spacing is 32 mas. Each star in this crowded field shows an Airy disk with fringes. The inset displays an individual PSF. Due to the combination of values for D and B , there are only about three visible fringes across the Airy disk, and the loss of contrast due to the width of the spectral band is barely visible. Courtesy T. Herbst, MPIA, Heidelberg

For Young's experiment in Sect. 4 we computed in (12) the resulting fringe pattern as an integral of individual fringe patterns that were shifted according to the respective source point. Formally, this is identical to the convolution between the object intensity and the fringe pattern. The result was a fringe pattern with a visibility $\mu(\vec{B})$ determined by the Fourier transform of the object intensity. Inserting this result here we obtain the image intensity distribution in a spatial interferometer as

$$I(\vec{\alpha}) = 2O_0A_0^2 \text{Besinc}^2(k\alpha D/2) \left(1 + |\mu(\vec{B})| \cos(\phi(\vec{B}) - k\vec{\alpha} \cdot \vec{B})\right), \quad (31)$$

with $O_0 = \int O(\vec{\alpha}') d\vec{\alpha}'$ and

$$\mu(B) = \frac{\int O(\vec{\alpha}') e^{-ik\vec{\alpha}' \cdot \vec{B}} d\vec{\alpha}'}{O_0}, \quad (32)$$

the van-Cittert-Zernike theorem.

These two formulae form the basis of our experiment. As long as the object is smaller than the PSF of an individual telescope, given by the square of the Besinc-function, we observe a fringe pattern with a visibility according to the van-Cittert-Zernike theorem. Combining visibility measurements for many baselines \vec{B} allows to reconstruct the object intensity $O(\vec{\alpha}')$ as the Fourier back-transform of $\mu(\vec{B})$. The smallest detail that can be resolved in the object is then determined by the longest baseline of all measurements, which in turn is equivalent to the finest fringe pattern with fringe spacing λ/B . One should note that the baselines need not only vary in length but also in orientation since we have two-dimensional images.

The coordinate space that is opened by different baselines \vec{B} – remembering that \vec{B} is the difference coordinate between the two sub-apertures – is not the aperture plane but a virtual plane called the uv -

plane, when u and v are the two coordinates of the plane. A vector in the uv -plane is given by the baseline calibrated by the wavelength, $(u, v) = \vec{B}/\lambda$. One should emphasise again that only the distance and the orientation of two telescopes determine the coordinate in the uv -plane, not their absolute position.

Hint: *The step from Young's experiment with an unlimited fringe pattern to an interferometer with a fringe pattern within an Airy disk in the image plane is often a source of confusion. One should keep the following points in mind. We are dealing with diffraction patterns. These can be caused by pinholes, like in Young's experiment, or by real sized apertures, like in a telescope or spatial interferometer. Increasing the size of a pinhole from 'zero' to a real aperture decreases the size of the diffraction pattern from 'infinite' to the limited size of e.g. an Airy disk in the case of a circular aperture. Lenses allow observing the diffraction pattern in their focal plane instead of at a very large distance.*

Combining two pinholes or two telescopes adds fringes to the single aperture diffraction pattern. Thus, the unlimited fringe pattern is confined to the Airy disk.

The imaging aspect comes in when looking at Airy disks created by light sources at different (angular) positions. The resulting distribution of Airy disks in the focal plane resembles – in terms of position and relative brightness – the distribution of the light sources, therefore it is called the image.

If the light sources are so close that their Airy disks are undistinguishable from a single Airy disk, they cannot be resolved as individual sources. However, in a spatial interferometer the contrast of the fringe pattern in the resulting Airy disk might be affected by the distribution of the sources and we can resolve the object by processing the fringe contrast resp. the fringe visibility.

6.2.1 Example: A Narrow Binary Star

In Sect. 5.1, we computed the visibility function of Venus as an example for an object shaped like uniform disk. Here, we will discuss the visibility function of a narrow binary star with an intensity distribution $I(\vec{\alpha}') = \frac{1}{2}(\delta(|\vec{\alpha}' - \vec{\alpha}'_s/2|) + \delta(|\vec{\alpha}' + \vec{\alpha}'_s/2|))$. The separation vector of the binary is $\vec{\alpha}'_s$ that is supposed to be much smaller than the PSF of an individual sub-aperture.

The visibility function is computed by Fourier transforming the intensity distribution obtaining

$$\begin{aligned}\mu(\vec{B}) &= \frac{1}{2} \int \left(\delta \left(\left| \vec{\alpha}' - \frac{\vec{\alpha}'_s}{2} \right| \right) + \delta \left(\left| \vec{\alpha}' + \frac{\vec{\alpha}'_s}{2} \right| \right) \right) e^{-ik_0 \vec{B} \cdot \vec{\alpha}'} d\vec{\alpha}' \\ &= \cos(k_0 \vec{B} \cdot \vec{\alpha}'_s/2).\end{aligned}\quad (33)$$

Writing the visibility in complex notation, the negative values of the cosine are accounted for by a phase of $\phi(\vec{B}) = \pi$ like in Sect. 5.1.

We have now the slightly confusing situation that the visibility function is a cosine function that, when inserted into (31), is multiplied by the cosine describing the fringe pattern. What does this mean? The visibility function determines the contrast of the fringe pattern. Thus, for very small values of the baseline the cosine has values close to 1 and the fringe pattern has almost maximum contrast as in Fig. 5.

Assuming that the baseline vector is parallel to the separation vector, the visibility function will decrease with increasing baseline, the fringes will disappear and come back again for another maximum for a baseline of $B = \lambda/\alpha'_s$. For this baseline the cosine has the value -1 , which means that there is a minimum in the fringe pattern where there was a maximum before and vice versa.

Further increasing the baseline brings back another swap of black and white fringes and another maximum of the visibility for $B = 2\lambda/\alpha'_s$. If the stars were truly point-like this periodicity would be unlimited. In the real world (and in our experiment) when stars have a finite size we have to compose our object of two disk-like structures both with diameter α'_0 that result in a visibility function of

$$\mu(\vec{B}) = \text{Besinc}(k_0 \vec{B} \cdot \vec{\alpha}'_0/2) \cos(k_0 \vec{B} \cdot \vec{\alpha}'_s/2).\quad (34)$$

This means that the visibility that was varying periodically with the baseline, is now damped by the Besinc function.

The last point to be discussed is a baseline that is perpendicular to the binary separation vector. In this case, the scalar product $\vec{B} \cdot \vec{\alpha}'_s$ is always zero and the cosine equal to 1, and the variation of fringe contrast with the baseline is solely determined by the diameter α'_0 of the individual stars. Thus, observing a binary with unsuitable baseline orientation would not reveal its binary nature but it would appear like a single star.

7 Practical Issues

In real spatial interferometers, the main problem is to control the optical path difference (OPD). Given that OPD variations of λ move the fringe pattern by one period of the fringe spacing, it is easy to see that the OPD has to be stable to better than $\lambda/10$ during the exposure time if we want to have a reliable measurement of the visibility. Otherwise the fringe pattern is simply smeared with randomly reduced visibility.

The OPD is affected by the optical path before the interferometer – mostly passing through vacuum but unfortunately also through turbulent air over the last few dozen kilometres – and inside the interferometer when the light is reflected by many mirrors before it interferes in the focus of the science camera.

In addition, we must not forget that we observe celestial objects that are in constant (apparent) motion due to the rotation of the Earth. Thus, the angle between the time dependent position vector $\vec{\alpha}'$ and the baseline vector \vec{B} , and therefore the $OPD = \vec{\alpha}' \cdot \vec{B}$ are in permanent motion.

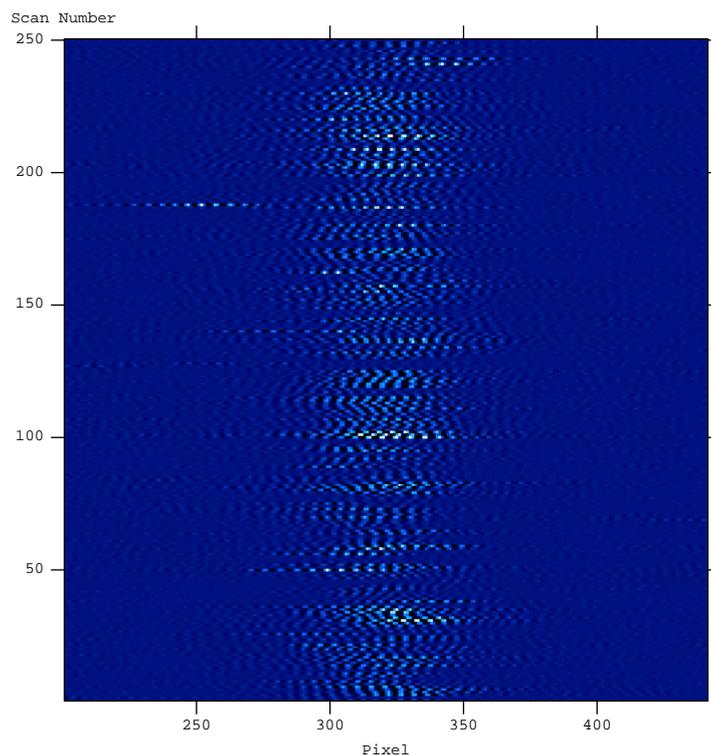


Figure 7: First fringes of the VLTI with two 8-m telescopes on October 30, 2001. Each horizontal line represents the interferometric fringes in the K-band registered during a single scan (resp. exposure). Due to atmospheric turbulence, the fringes are slightly shifted sideways

This implies that two independent telescopes form the spatial interferometer so that the baseline vector is immobile, horizontal to the ground. For shorter baselines one can use a common telescope

mount for the telescopes of the interferometer so that the baseline vector rotates in the same way as the telescopes, tracking the position vector $\vec{\alpha}'$. In this case the angle between the baseline and the position vector is constant and the OPD is invariant.

In long baseline interferometers with independent telescopes, the OPD in motion is compensated for by adding a *delay line* in the optical path that adds an OPD of exactly $-\vec{\alpha}' \cdot \vec{B}$. A delay line is an optical cat's eye system on rails that is moving with an accuracy of several 10nm over a distance up to 100m depending on the length of the baseline and on the maximum zenith angle that is used for observations. Since this kind of high-tech equipment was unavailable in the 1920s, spatial interferometry had been abandoned until the 1970s when modern interferometry started.

With delay lines, the fringes are kept in a stable position on average but the observations still suffer from random OPD variations due to atmospheric turbulence and vibrations. These random OPD variations have an amplitude of several microns, resulting in a fringe wobble of several fringes. However, the power spectrum of the OPD variations shows that the maximum amplitude is at very low frequencies (≤ 1 Hz). The tolerable OPD variation of about $\lambda/10$ has a typical timescale of about 1/10 seconds so that exposure times in this range produce a fringe pattern that is only moderately smeared.

The fringe patterns in Fig. 7 give an idea about the short time variations of the fringe positions – look at the fringe motion from scan to scan – and the long time variations indicated by the total variations over all scans.

Unfortunately, there is nothing else one can do but reducing the exposure time. This reduces the sensitivity to rather bright sources. However, there is a technique that allows observing faint sources at least in the vicinity of bright stars. This technique, called *fringe tracking* uses the measurements of the bright star fringe position to compensate for the OPD variations with a fast moving mirror so that the fringe motion is stabilised to an acceptable level³. Since the OPD variations are correlated over several 10 arcsec in the sky, a faint star within this distance can be observed taking advantage of the stabilised fringes using a long exposure time without smearing the fringe pattern. Instead of a few 100 milli seconds the exposure time can be a few 10 seconds, increasing the sensitivity by a factor of 100. Since two stars – the bright guide star and the faint science object – are observed this is also called a *dual-feed system*.

Adding some extra equipment, for instance a laser metrology system to measure the OPD inside the interferometer, a dual-feed system allows to determine the phase $\phi(\vec{B})$ of the visibility function $\mu(\vec{B})$ so that we can attempt to reconstruct a real image from the visibility measurements.

8 Visibility Measurement

The number of independent parameters in the reconstructed object intensity is determined by the number of visibilities (modulus and phase) that are measured. Thus, even a reconstruction of a 10x10 pixel image of e.g. a galaxy requires about 100 visibility measurements. Given that each baseline requires either to move the telescopes or to have a sufficient number of telescopes that are conveniently distributed, and that the baselines have to be distributed equally between the shortest and the longest lengths, one understands that it is not easy to collect a suitable set of visibilities. In the jargon of interferometrists this process of collecting visibility measurements is called *filling the uv-plane*.

Therefore, almost all interferometric observations attempt to fit model parameters rather than reconstructing an image. The measurement of stellar diameters is a good example for this technique. By assuming that the shape of the star can be modelled as a uniform disk, the visibility function is a Bessel-function as displayed in Fig. 4. Then one measurement would be sufficient to determine the one parameter, the diameter α'_0 that we are looking for (unless the visibility is very low and could be either in

³Correcting a signal with an actuator, in this case the fast moving mirror, before the sensor so that the sensor 'sees' the corrected signal, is called a *closed-loop system*. There are many applications for closed loop systems, for instance telescope guiding or adaptive optics.

the first or the second side lobe of the function). More measurements allow to further refine the model by assuming (the physically justified) limb-darkening effect, i.e. a slight reduction of the intensity towards larger radii of the disk. On the cover page, the reconstructed shape of Achernar is displayed. Different orientations of baselines revealed different diameters of the star, and it turned out that the star is elliptical and not circular.

Narrow binaries form another class of objects that can be successfully observed with a few visibility measurements. A binary has a visibility function that is proportional to a cosine function (see Sect. 6.2.1). Then we have to make sure that the measurements are unambiguous.

9 Some Further reading

- Florentin Millour, "All you ever wanted to know about optical long baseline stellar interferometry, but were too shy to ask your adviser", http://arxiv.org/PS_cache/arxiv/pdf/0804/0804.2368v1.pdf.
- Andreas Glindemann, "Das Sterninterferometer auf dem Paranal", *Physik in unserer Zeit* 2003, **34**, 2, 64–71
- Course Notes from the 2002 Michelson Interferometry Summer School, <http://olbin.jpl.nasa.gov/iss2002/index.html>
- Peter R. Lawson (Ed.), "Principles of Long Baseline Stellar Interferometry", Course Notes from the 1999 Michelson Interferometry Summer School, <http://olbin.jpl.nasa.gov/iss1999/coursenotes.html>

Index

Airy disk, 15
amplitude, 5
angular resolution, 15
baseline \vec{B} , 4
Besinc-function, 13
circ-function, 13
coherence, 7
delay line, 21
Dirac function $\delta(\cdot)$, 11
dual-feed system, 21
energy flow density, 5
ESO – European Southern Observatory, 3, 4
flux, 5
flux density, 5
Fourier optics, 15
Fraunhofer diffraction, 14
fringe pattern, 6
fringe tracking, 21
intensity, 5
irradiance, 5
optical disturbance, 4
point-spread function (PSF), 15
power spectral density, 5
quasi-monochromatic approximation, 11
Rayleigh criterion, 15
seeing, 4
spectral amplitude, 5
spectral bandwidth, 7
spectral intensity, 5
uv-plane, 19
van-Cittert-Zernike theorem, 9
visibility \mathcal{V} , 7
visibility function, 9
VLTI – Very Large Telescope Interferometer, 4, 20
wavefront, 10
white-light fringe, 7, 9