

LECTURE NOTES OF ASTRONOMICAL INTERFEROMETRY  
Prof. Mauro D'Onofrio

Patrizia Bussatori & Fabrizio Muratore

March 2022

# Premise

This book contains the notes taken from the lessons of the Astronomical Interferometry course held by prof. Mauro D'Onofrio for the master degrees in Cosmology and Astrophysics, in the academic year 2021/2022. As these are not official professor approved notes, we do not guarantee the correctness of the contents. Indeed, any error reporting is welcome! Enjoy the reading!

*Patrizia Bussatori & Fabrizio Muratore*

# Contents

<b>1 Basic principles and types of interferometry</b>	<b>2</b>
1.1 Nature or light . . . . .	2
1.2 What is interference? . . . . .	2
1.2.1 Optical Path Length (OPL) and Optical Path Difference (OPD) . . . . .	2
1.2.2 Principle of superposition . . . . .	3
1.2.3 Visibility of interference fringes . . . . .	5
1.3 What is an interferometer? . . . . .	5
1.4 Wave front division interferometers . . . . .	6
1.4.1 Young experience . . . . .	6
1.4.2 Fabry-Perot interferometer . . . . .	7
1.5 Amplitude-division interferometers . . . . .	8
1.5.1 Newton's Ring . . . . .	8
1.5.2 Michelson interferometer . . . . .	8
1.5.3 Pohl interferometer . . . . .	10
1.5.4 Twyman-Green interferometer . . . . .	11
1.5.5 Mach-Zehnder interferometer . . . . .	11
1.6 Type and positions of fringes . . . . .	12
1.6.1 Fringes of same inclination . . . . .	12
1.6.2 Fringes of same thickness . . . . .	13
1.7 Formation theory of images by Abbe (1873) . . . . .	13
1.8 Fourier transform FT . . . . .	15
1.8.1 FT . . . . .	15
1.8.2 DFT and $[u,v]$ plane . . . . .	16
1.8.3 Bidimensional transform . . . . .	17
1.9 Linear systems and convolution . . . . .	17
1.10 Point Spread Function . . . . .	18
1.11 Single dish limitations . . . . .	20
1.12 Summery . . . . .	20
<b>2 Two-element interferometers</b>	<b>22</b>
2.1 Coherence . . . . .	22
2.1.1 Temporal and spatial coherence . . . . .	22
2.1.2 Temporal and spatial coherence functions . . . . .	23
2.2 Temporal coherence function . . . . .	23
2.3 Spatial coherence function . . . . .	25
2.4 What should we draw from all this? . . . . .	25
2.5 Spatial interferometry . . . . .	25
2.5.1 Spatial coherence revisited . . . . .	26
2.6 Two-element interferometers . . . . .	26
2.6.1 Resolution of two-element interferometry . . . . .	28
2.6.2 The correlator of the two-element interferometry . . . . .	29

<b>3 Measurements of fringes</b>	<b>32</b>
3.1 Measurements of fringes . . . . .	32
3.1.1 First procedure . . . . .	32
3.1.2 Second procedure - snapshots . . . . .	33
3.1.3 Third procedure . . . . .	33
<b>4 Imaging with interferometers</b>	<b>34</b>
4.1 Simple source . . . . .	34
4.2 Image reconstruction . . . . .	36
4.3 Data reduction . . . . .	36
4.4 Aperture synthesis . . . . .	37
<b>5 Radio Astronomy</b>	<b>39</b>
5.1 Brief history . . . . .	39
5.2 Radio window . . . . .	39
5.3 How works a radio telescope . . . . .	40
5.4 Radio physics . . . . .	42
5.4.1 Intensity and flux density . . . . .	42
5.4.2 Surface brightness . . . . .	42
5.4.3 Brightness temperature . . . . .	42
5.4.4 Radiative transfer . . . . .	43
5.4.5 Black Body Radiation . . . . .	44
5.5 What does a radio telescope detect? . . . . .	46
5.5.1 Types of antennas . . . . .	47
5.6 Descriptive antenna parameters . . . . .	48
5.6.1 The power pattern $P(\theta, \phi)$ and the gain . . . . .	48
5.6.2 The main beam solid angle . . . . .	48
5.6.3 The effective aperture and the directivity . . . . .	50
5.6.4 Antenna temperature . . . . .	51
5.7 Summery . . . . .	52
5.8 Sensitivity and Noise . . . . .	53
5.8.1 Receiver calibration . . . . .	54
5.9 Noise equations . . . . .	54
<b>6 The ideal interferometer</b>	<b>56</b>
6.1 Simplification 1 . . . . .	59
6.1.1 Geometrical delay . . . . .	59
6.1.2 Multiple spectral channels . . . . .	59
6.2 Simplification 2 . . . . .	60
6.2.1 Polarization . . . . .	60
6.2.2 Describing cross-correlations . . . . .	61
6.3 Simplification 3 . . . . .	62
6.4 Simplification 4 . . . . .	62
6.5 Simplification 5 and 6 . . . . .	62
6.5.1 Primary beam . . . . .	62
6.6 Simplification 7 . . . . .	63
6.7 Simplification 8 . . . . .	63
6.8 Sampling and imaging . . . . .	63
6.8.1 Deconvolution . . . . .	64
6.8.2 Resolution, maximum scale and field size . . . . .	64
<b>7 Difference between a CCD and a interferometer</b>	<b>65</b>
7.1 How do we characterize what a source looks like? . . . . .	65
7.1.1 Transfer function of circular aperture . . . . .	66

7.2	What/how do interferometers measure? . . . . .	66
7.2.1	What happens with polychromatic light? . . . . .	69
7.2.2	Issues with optical/IR delay lines . . . . .	71
7.2.3	Timeout . . . . .	71
7.3	Planning and undertaking interferometric science . . . . .	72
7.3.1	A perfectly unresolved source, not on-axis . . . . .	73
7.3.2	An unequal binary star . . . . .	73
7.3.3	A uniform stellar disc . . . . .	74
7.3.4	How does this help planning observations? . . . . .	74
7.4	Imaging and sensitivity . . . . .	74
7.4.1	How maps are actually recovered . . . . .	75
7.4.2	What does "sensitivity" mean for interferometry? . . . . .	75
7.5	How is what we have learnt impacted by the atmosphere? . . . . .	76
7.5.1	How can we mitigate against these perturbations? . . . . .	77
7.5.2	Temporal effects . . . . .	78
<b>8</b>	<b>Sensitivity and noise</b> . . . . .	<b>80</b>
8.1	Radio sensitivity . . . . .	81
8.1.1	What affects aperture efficiency $\eta_A$ ? . . . . .	82
8.2	Radiometers . . . . .	82
8.2.1	Ideal radiometer equation . . . . .	83
8.3	System temperature . . . . .	83
8.4	Surface brightness sensitivity . . . . .	84
8.5	Radio frequency interference . . . . .	84
<b>9</b>	<b>Main scientific goals of VLTI and ALMA</b> . . . . .	<b>85</b>
9.1	Very Large Telescope Interferometer (VLTI) . . . . .	85
9.1.1	Extra-solar Planets . . . . .	85
9.1.2	Active galactic nuclei (AGN) . . . . .	86
9.1.3	AGB, Post-AGB stars and PNe . . . . .	86
9.2	Atacama Large Millimeter/submillimeter Array (ALMA) . . . . .	86
<b>10</b>	<b>Thermal and nonthermal radiation</b> . . . . .	<b>88</b>
10.1	Thermal radiation mechanisms . . . . .	89
10.1.1	Black-body radiation and its spectrum . . . . .	89
10.2	Nonthermal radiation mechanisms . . . . .	89
10.2.1	Synchrotron radiation and its spectrum . . . . .	89
10.2.2	Bremsstrahlung radiation and its spectrum . . . . .	91
10.3	Summary . . . . .	92
<b>A</b>	<b>Polarization of EM waves</b> . . . . .	<b>93</b>
A.1	Vector waves . . . . .	93
A.2	The Poincarè Sphere and the Stokes Parameters . . . . .	97
A.3	Quasi-Monochromatic Plane waves . . . . .	99
A.4	The Stokes Parameters for Quasi-Monochromatic Waves . . . . .	99

# Introduction

Interferometry is a technique which uses the interference of **superimposed waves** to extract information. Interferometry typically uses electromagnetic waves and is an important investigative technique in the fields of astronomy, fiber optics, engineering metrology, oceanography, seismology, spectroscopy (and its applications to chemistry), quantum mechanics, nuclear and particle physics.

Interferometers are devices that extract information from **interference**. They are widely used in science and industry for the measurement of microscopic displacements, refractive index changes and surface irregularities. In the case with most interferometers, light from a single source is split into two beams that travel in different optical paths, which are then combined again to produce interference; two incoherent sources can also be made to interfere under some circumstances though. The resulting interference **fringes** give information about the difference in optical path lengths. In analytical science, interferometers are used to measure lengths and the shape of optical components with **nanometer precision**; they are the highest precision length measuring instruments in existence. In Fourier transform spectroscopy they are used to analyze light containing features of absorption or emission associated with a substance or mixture. An astronomical interferometer consists of two or more separate telescopes that combine their signals, offering a resolution equivalent to that of a telescope of diameter equal to the largest separation between its individual elements. Usually those type of systems are used to observe with a huge amount of details very low temperature sources that form the *cold Universe*.

# Chapter 1

## Basic principles and types of interferometry

### 1.1 Nature of light

For many centuries, light was considered a stream of particles, however it can be seen also as waves. In particular, light wave exhibits various behaviours which can not be interpreted through the particles theory of light such as, refraction, diffraction and interference. To be more specific, light waves are transverse waves with two components; magnetic and electric field each one of them oscillating perpendicular to the other and to the propagation direction. In vacuum, those waves travel at a speed  $c$ , the speed of light but inside a medium of refractive index  $n$ , they travel at a speed of  $v = c/n$ .

According to the different wavelength we are considering, we can construct the electromagnetic spectrum. The electromagnetic spectrum covers electromagnetic waves with frequencies ranging from below one hertz to above  $10^{25}$  hertz, corresponding to wavelengths from thousands of kilometers down to a fraction of the size of an atomic nucleus. This frequency range is divided into separate bands, and the electromagnetic waves within each frequency band are called by different names; beginning at the low frequency (long wavelength) end of the spectrum these are: radio waves, microwaves, infrared, visible light, ultraviolet, X-rays, and gamma rays at the high-frequency (short wavelength) end. In particular the visible light extends from 750nm for the red color to 380nm for the violet color.

Taking into account those basic information about electromagnetic waves, let's focus on interference phenomena.

### 1.2 What is interference?

Interference is a light phenomenon. It is the net effect of the combination of two or more wave trains moving on intersecting or coincident paths.

In electromagnetic waves, interference between two or more waves is just an addition or **superposition process** and this can result in a new wave pattern: the interference pattern.

Let's now analyse interference pattern in terms of the Optical Path Difference (OPD) first and in terms of superposition of waves second, in the following next two subsections. After that we will talk about the visibility of the interference fringes.

#### 1.2.1 Optical Path Length (OPL) and Optical Path Difference (OPD)

When a light beam travels in space from one point to another, the path length is the geometric length  $d$  multiplied by  $n$ , the refractive index of the medium in which it is traveling. So  $OPL = nd$ . In particular, if the beam travels in the air, where  $n = 1$ , then  $OPL = d$ .

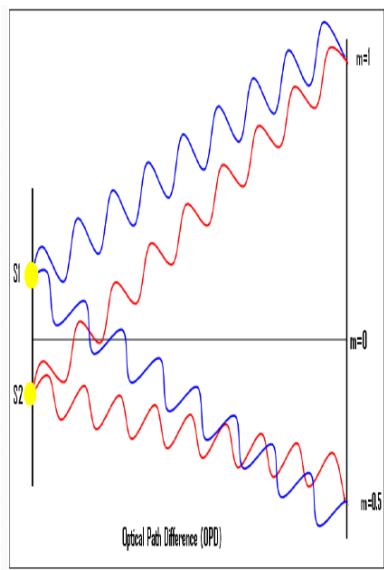


Figure 1.1: According to the value of  $m$ , there can be a different type of fringe, a bright one or a dark one.

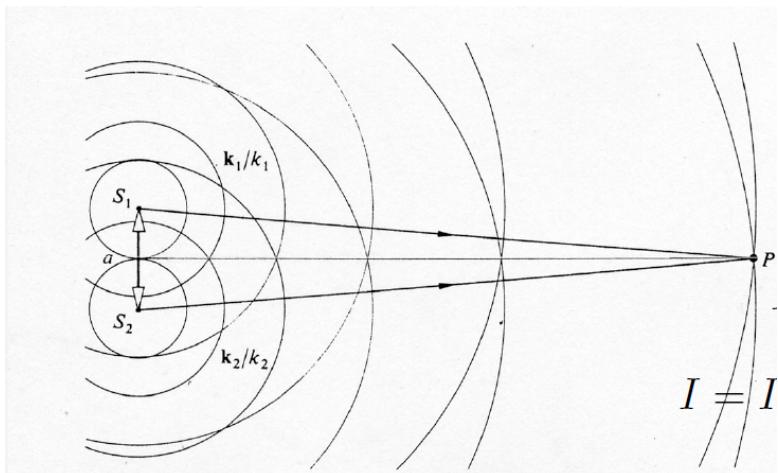


Figure 1.2: Spherical waves emitted from  $S_1$  and  $S_2$ .

Consider now two light beams with the same wavelength  $\lambda$  i.e same frequency, that travel from two different points towards the same destination. Taking different paths, there will be a difference in their optical path and this one is called Optical Path Difference (OPD). This one is a fundamental factor in determining the fringes intensity. Indeed, assuming the integer number  $m$  as the order, if  $OPD = m\lambda$  there is a bright fringe. On the contrary, if  $OPD = (m - 1/2)\lambda$  there is a dark fringe. This can be seen in figure 1.1.

### 1.2.2 Principle of superposition

We can see the previous discussion also in terms of superposition of waves. Indeed, interferometry makes use of the **principle of superposition** to combine waves in a way that will provide information about the sources their self. This works because when two waves with the *same frequency* combine, the resulting intensity pattern is determined by the phase difference between the two waves: waves that are in phase will undergo **constructive interference** while waves that are out of phase will undergo **destructive interference**. Waves which are not completely in phase nor completely out of phase will have an intermediate intensity pattern, which can be used to determine their relative phase difference.

To be more specific, let's consider image 1.2.

Here there are two sources,  $S_1$  and  $S_2$  at a distance  $a$  one from the other, that are emitting isotropic spherical waves. Assuming that the electric field of the first source is  $\vec{E}_1$  and the one of the second source is  $\vec{E}_2$ , than the total electric field is:

$$\vec{E} = \vec{E}_1 + \vec{E}_2 \quad (1.1)$$

The total electric field in vectorial sense can be used to define the intensity  $I$ :

$$I = \left\langle \vec{E} \cdot \vec{E} \right\rangle_T \quad (1.2)$$

which can be also defined in terms of individual intensities  $I_1$  and  $I_2$ :

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2} \cos \delta \quad (1.3)$$

where  $\delta$  is the phase difference between the two waves, also defined as function of observational wavelength  $\lambda$  and the angle  $\theta$  between the normal and the line joining  $S_1$  and  $P$  (and the same for  $S_2$ ):

$$\delta = \frac{2\pi}{\lambda} a \sin \theta \quad (1.4)$$

According to the values of  $\delta$ , there can be different type of interference in point P. Indeed if  $\delta = 0, \pm 2\pi, \pm 4\pi, \dots$  there is **positive interference**, while if  $\delta = 0, \pm \pi, \pm 3\pi, \dots$  there is **negative interference**.

For example, if  $\delta = 0$  then  $I_{max} = I_1 + I_2 + 2\sqrt{I_1 I_2}$ . In this case if  $I_1 = I_2 = I$  then  $I_{max} = 4I$ . Or eventually, if  $\delta = \pi$  then  $I_{min} = I_1 + I_2 - 2\sqrt{I_1 I_2}$ . In this case if  $I_1 = I_2 = I$  then  $I_{min} = 0$ .

Assuming spherical waves as stated before and wave number  $|\vec{k}| = \frac{2\pi}{\lambda}$ , the difference in terms of phase becomes:

$$\delta = k(r_1 - r_2) + (\epsilon_1 - \epsilon_2) \quad (1.5)$$

and as consequence the intensity becomes:

$$I = 4I_0 \cos^2 \frac{1}{2} [k(r_1 - r_2) + (\epsilon_1 - \epsilon_2)] \quad (1.6)$$

in which  $\epsilon_1$  and  $\epsilon_2$  are the eccentricity of hyperboloids of revolution with focuses in  $S_1$  and  $S_2$  and  $r_1$  and  $r_2$  are the radius of waves emitted by the two sources.

The superposition of waves from two sources can usually only result in an observable fixed (stationary) interference pattern if the sources are **coherent**. This means that the waves from the sources have both the same frequency and the phase difference between them is constant. In particular in P are observed maximums for  $r_1 - r_2 = [2\pi m + (\epsilon_1 - \epsilon_2)]/k$  and minimums for  $r_1 - r_2 = [(2m+1)\pi + (\epsilon_1 - \epsilon_2)]/k$ . See the interference pattern in figure 1.3.

**PAY ATTENTION!** Interference pattern require coherent waves in a very small wavelength range. The ideas of coherent light have become very prominent since the invention of lasers, which are almost ideal sources of coherent light. But the concept of coherence dates back to the work of Van Cittert in 1934 and Zernike in 1938. Two waves are mutually coherent if the phase difference between them at a particular point in space is constant during a **long period of time**. If this is so, then interference between the waves will be observed. But of course nothing in the real world can be exactly constant, so we should really ask how much this phase difference changes during the period of an observation.

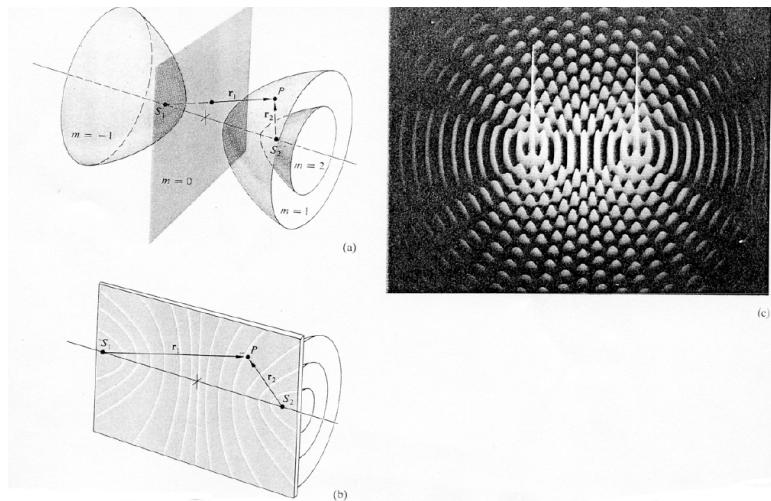


Figure 1.3: Interference phenomena in 3D and bi-dimensional representation.

If the answer is that the change is much smaller than  $\pi$ , then clear interference effects will be seen. If not, then interference won't be seen. Suppose for example that we have two independent sources, with frequencies  $\omega_1$  and  $\omega_2$ . Then the phase difference between them is changing as  $(\omega_1 - \omega_2)t$ . So if the observation time (say, the exposure time for a photograph of an interference pattern) is 0.01 second, the difference  $(\omega_1 - \omega_2)$  has to be  $\pi/0.01 = 314s^{-1}$  for fringes to be recorded. Since at optical frequencies  $\omega$  is of order  $10^{16}s^{-1}$ , it follows that no two independent sources could be stable enough (to one part in  $10^{13}$ ) for this to be possible; so in practice interference patterns can only be observed when the interfering waves derive from the same source.

### 1.2.3 Visibility of interference fringes

Suppose to observe an interference pattern. One of the main feature of it is the **visibility**. Visibility determines the ability to resolve interference fringes and it depends on the coherence degree between the recombined light waves. It is define as:

$$V = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \quad (1.7)$$

- if  $I_{min} = 0$  then  $V = 1$ ;
- if  $I_{min} = I_{max}$  then  $V = 0$ .

From this we infer that  $0 \leq V \leq 1$ .

## 1.3 What is an interferometer?

Assuming all what we did in the previous sections, we can now give a definition of an interferometer. An **interferometer** is an optical instrument that can produced two beams interference or multiple beam interference.

Those instruments follow the Fresnel-Arago laws:

- two beams of coherent light with polarizations orthogonal one to each other can NOT interfere and fringes do NOT form;
- two beams of coherent light with polarizations parallel one to each other interfere always, also in case of natural light;
- the two states of polarization perpendicular one to each other that constitutes natural light can not interfere and so can not form fringes observable, even if one of the two is rotated artificially

and aligned to the other. This because they are incoherent between them.

Taking into account those laws, interferometers can be divided into two main types:

- **wave front division interferometers** in which two light beams from the same wave front are made to interfere to produce an interference fringe pattern;
- **amplitude-division interferometers** in which a light beam from one source point is divided into two beams using a beam splitter.

## 1.4 Wave front division interferometers

### 1.4.1 Young experience

Young's interference experiment, also called Young's double-slit interferometer, was the original version of the modern double-slit experiment, performed at the beginning of the nineteenth century by Thomas Young. This experiment played a major role in the general acceptance of the wave theory of light.

**First experiment** - The figure 1.4 shows the geometry for a far-field viewing plane. It is seen that the relative paths of the light travelling from the two points sources to a given point in the viewing plane varies with the angle  $\theta$ , so that their relative phases also vary. When the path difference is equal to an integer number of wavelengths, the two waves add together to give a maximum in the brightness, whereas when the path difference is equal to half a wavelength, or one and a half etc., then the two waves cancel, and the intensity is at a minimum.

The linear separation (distance) -  $\Delta y$  between fringes (lines with maximum brightness) on the screen is given by the equation :

$$\Delta y = L\lambda/d \quad (1.8)$$

where  $L$  is the distance between the slit and screen,  $\lambda$  is the wavelength of light and  $d$  is the slit separation as shown in figure. The angular spacing of the fringes,  $\theta_f$ , is then given by:

$$\theta_f \approx \lambda/d \quad (1.9)$$

where  $\theta_f \ll 1$ . It can be seen that the spacing of the fringes depends on the wavelength, the separation of the holes, the distance between the slits and the observation plane.

This expression applies when the light source has a single wavelength, whereas Young used sunlight, and was therefore looking at white-light fringes which he describes above. A white light fringe pattern can be considered to be made up of a set of individual fringe patterns of different colours. These all have a maximum value in the centre, but their spacing varies with wavelength, and the superimposed patterns will vary in colour, as their maximums will occur in different places. Only two or three fringes can normally be observed.

**Modern double-slit interferometer** In 1801, Young presented a famous paper which described various interference phenomena and in which he also mentioned the possibility of passing light through two slits. So a modern version of Young experience is the **double-slit experiment** which is a demonstration that light and matter can display characteristics of both classically defined waves and particles; moreover, it displays the fundamentally probabilistic nature of quantum mechanical phenomena.

In the basic version of this experiment, a coherent light source, such as a laser beam, first illuminates a plate pierced by two parallel slits, and then the light passing through the slits is observed on a screen behind the plate. The wave nature of light causes the light waves passing through the two slits to

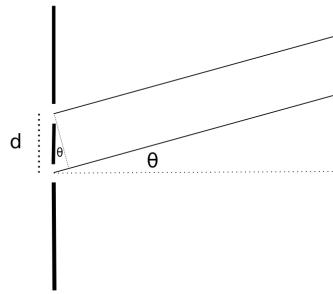


Figure 1.4: Young experiment

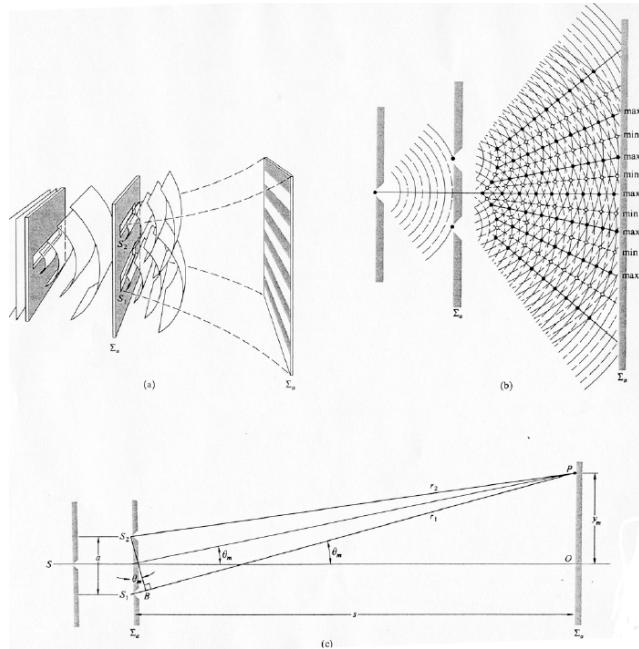


Figure 1.5: Double-slit experiment.

interfere, producing bright and dark bands on the screen – a result that would not be expected if light consisted of classical particles, as shown in figure 1.5.

### 1.4.2 Fabry-Perot interferometer

This type of interferometer consists of two parallel high reflecting glass plates separated by a distance  $p$  of several millimeters, a focusing lens and a display screen. In practice the beam falls on  $L_1$  with a specific angle  $\theta$  with respect to the surface. Part of the beam is transmitted to  $L_2$ , other part is reflected. The transmitted part is partially reflected back to  $L_1$ . Then it is again reflected to the  $L_2$  which partially reflects and transmits each incident light. The transmitted lights from  $L_2$  falls on the focusing lens and the beams are focused on the screen at point P. These beams interfere constructively or destructively according to the phase difference between them.

This system allows to have high resolution fringes (especially in optical spectroscopy) but it has also a high sensitivity to wave length changes. For this last reason it is often used in laser to select wavelength.

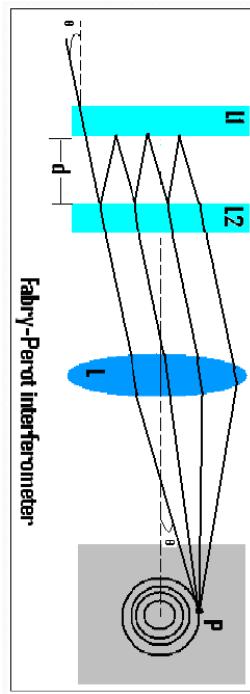


Figure 1.6: Fabry-Perot interferometer

## 1.5 Amplitude-division interferometers

### 1.5.1 Newton's Ring

Newton's rings is a phenomenon in which an interference pattern is created by the reflection of light between two surfaces; a spherical surface and an adjacent touching flat surface. It is named after Isaac Newton, who investigated the effect in 1666. When viewed with monochromatic light, Newton's rings appear as a series of concentric, alternating bright and dark rings centered at the point of contact between the two surfaces. When viewed with white light, it forms a concentric ring pattern of rainbow colors, because the different wavelengths of light interfere at different thicknesses of the air layer between the surfaces. According to image 1.7, radius of bright -ring is at:

$$x_m = \sqrt{(m + \frac{1}{2})\lambda n_f R} \quad (1.10)$$

### 1.5.2 Michelson interferometer

A Michelson interferometer consists minimally of mirrors  $M1$  &  $M2$  and a beam splitter  $M$ . In figure 1.8, a source  $S$  emits light that hits the beam splitter (in this case, a plate beamsplitter) surface  $M$  at point C.  $M$  is partially reflective, so part of the light is transmitted through to point B while some is reflected in the direction of A. Both beams recombine at point C' to produce an interference pattern incident on the detector at point E (or on the retina of a person's eye). If there is a slight angle between the two returning beams, for instance, then an imaging detector will record a sinusoidal fringe pattern as shown in 1.9 figure b. If there is perfect spatial alignment between the returning beams, then there will not be any such pattern but rather a constant intensity over the beam dependent on the differential pathlength; this is difficult, requiring very precise control of the beam paths.

Image 1.8 shows use of a coherent (laser) source. Narrowband spectral light from a discharge or even white light can also be used, however to obtain significant interference contrast it is required that the differential pathlength is reduced below the coherence length of the light source. This can be only micrometers for white light, as discussed below.

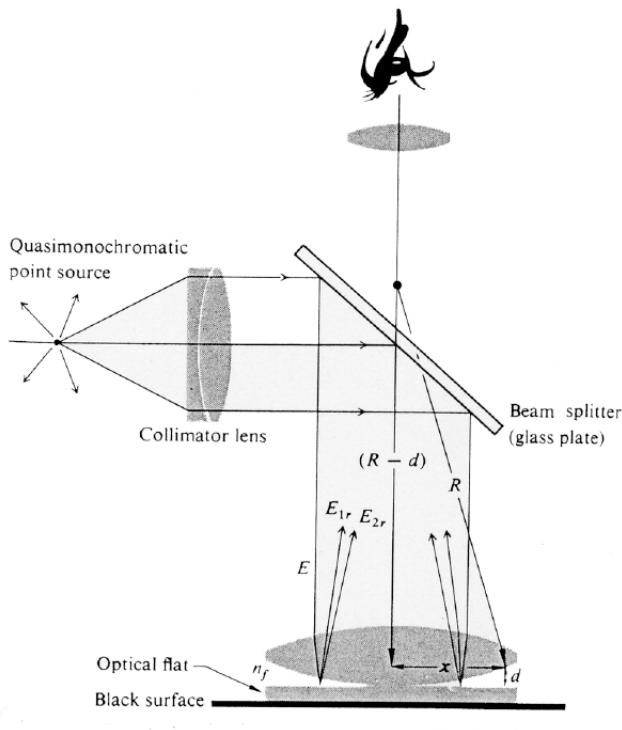


Figure 1.7: System to create Newton's ring.

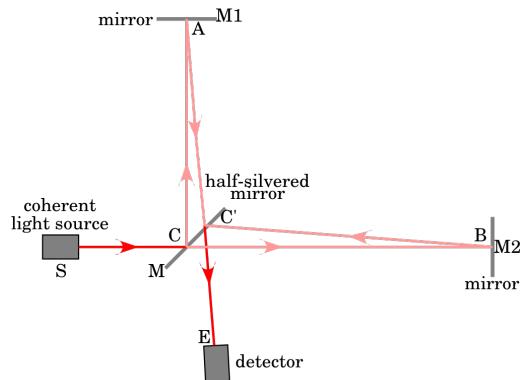


Figure 1.8: Michelson interferometer.

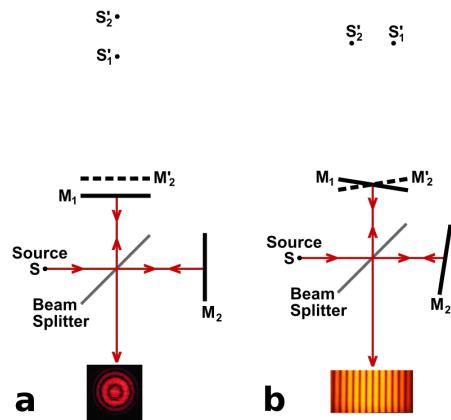


Figure 1.9: Formation of fringes in a Michelson interferometer.

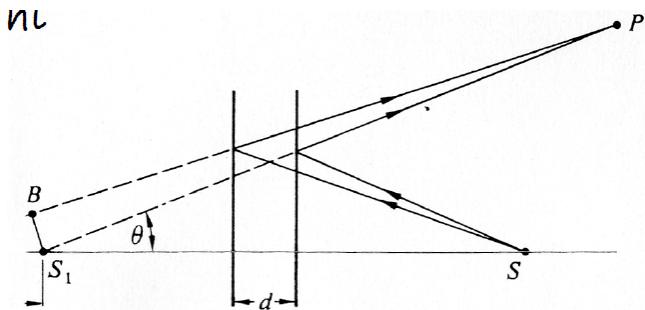


Figure 1.10: Case of parallel plates.

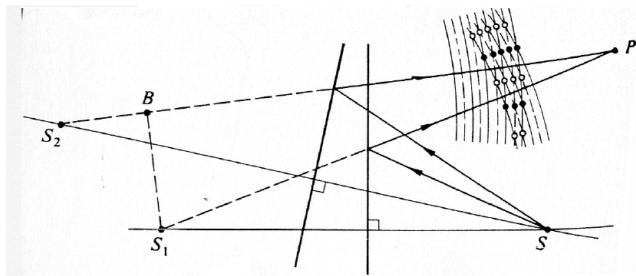


Figure 1.11: Case of NON parallel plates.

As shown in figure 1.9 a and b, the observer has a direct view of mirror  $M1$  seen through the beam splitter, and sees a reflected image  $M'2$  of mirror  $M2$ . The fringes can be interpreted as the result of interference between light coming from the two virtual images  $S'1$  and  $S'2$  of the original source  $S$ . The characteristics of the interference pattern depend on the nature of the light source and the precise orientation of the mirrors and beam splitter. In image 1.9 a, the optical elements are oriented so that  $S'1$  and  $S'2$  are in line with the observer, and the resulting interference pattern consists of circles centered on the normal to  $M1$  and  $M'2$  (fringes of equal inclination). If, as in image 1.9 b,  $M1$  and  $M'2$  are tilted with respect to each other, the interference fringes will generally take the shape of conic sections (hyperbolas), but if  $M1$  and  $M'2$  overlap, the fringes near the axis will be straight, parallel, and equally spaced (fringes of equal thickness).

From a mathematical point of view, if all the system are perfect aliened (circular fringes), for an observer located on the detector the light seems to come from two coherent sources separated by a path difference of  $2d \cos \theta$ . The total phase difference is  $k_0 2d \cos \theta + \pi$ . So destructive interference is for  $2d \cos \theta_m = m\lambda$ .

The Michelson interferometer (among other interferometer configurations) is employed in many scientific experiments and became well known for its use by Michelson and Edward Morley in the famous Michelson–Morley experiment in a configuration which would have detected the Earth's motion through the supposed ether that most physicists at the time believed was the medium in which light waves propagated. The null result of that experiment essentially disproved the existence of such an ether, leading eventually to the special theory of relativity and the revolution in physics at the beginning of the twentieth century. In 2015, another application of the Michelson interferometer, LIGO, made the first direct observation of gravitational waves. That observation confirmed an important prediction of general relativity.

### 1.5.3 Pohl interferometer

The Pohl interferometer is an amplitude splitting interferometer in which light beam from a point source is reflected from the surfaces with different optical properties (refractive index, reflectivity) forming an interference pattern. The Pohl configuration offers a simple, quick and efficient approach for measuring the wavelength. It has been applied for shop testing conditions, phase measurement and parallelism measurement of transparent surfaces.

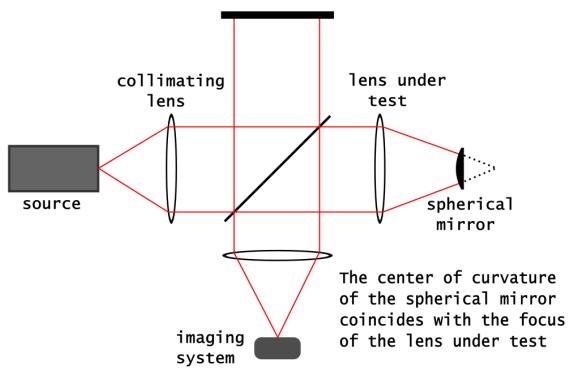


Figure 1.12: Twyman-Green interferometer.

Let's analyse image 1.10 and image 1.11: parallel or inclined surfaces represent the possible configurations for mirrors or transparent films used in Pohl interferometer. In the first parallel case (figure 1.10, the light is emitted from source  $S$ , then it is reflected by the two plates forming a virtual image  $S_1$  and then the light reflected hit point  $P$  where there are no interference patterns. On the contrary, if the two surfaces have an inclination one with respect to the other (figure 1.11), two virtual images are created ( $S_1$  and  $S_2$ ) and due to the fact that light beams have a small difference in path, then in  $P$  fringes are observed.

#### 1.5.4 Twyman-Green interferometer

A Twyman-Green interferometer is a variant of the Michelson interferometer principally used to test optical components.

Figure 1.12 illustrates a Twyman-Green interferometer set up to test a lens. Light from a laser is expanded by a diverging lens (not shown), then is collimated into a parallel beam. A convex spherical mirror is positioned so that its center of curvature coincides with the focus of the lens being tested. The emergent beam is recorded by an imaging system for analysis.

The fixed mirror in the Michelson interferometer is rotatable in the Twyman-Green interferometer, and while the light source is usually an extended source (although it can also be a laser) in a Michelson interferometer, the light source is always a point-like source in the Twyman-Green interferometer. The rotation of one mirror results in straight fringes appearing in the interference pattern, a fringing which is used to test the quality of optical components by observing changes in the fringe pattern when the component is placed in one arm of the interferometer.

#### 1.5.5 Mach-Zehnder interferometer

The Mach-Zehnder interferometer is a particularly simple device for demonstrating interference by division of amplitude. It consists of a light source, a detector, two mirrors to control the beams directions and two beam splitters to split and recombine the incident beam.

See figure 1.13. Practically, a light beam is first split into two parts by a beam splitter, reaching mirror  $M1$  and  $M2$ . Then the two beams are recombined by a second beam splitter ( $SB2$ ). Depending on the relative phase acquired by the beam along the two paths the second beam splitter will reflect the beam with efficiency between 0 and 100% and interference fringes are produced depending on the path difference.

This system can be used to measure thickness at constant refractive index or it can measure refractive index at constant thickness. Indeed, main applications of this type pf interferometer are the measure of refractive index of fluids and study the heat transfer.

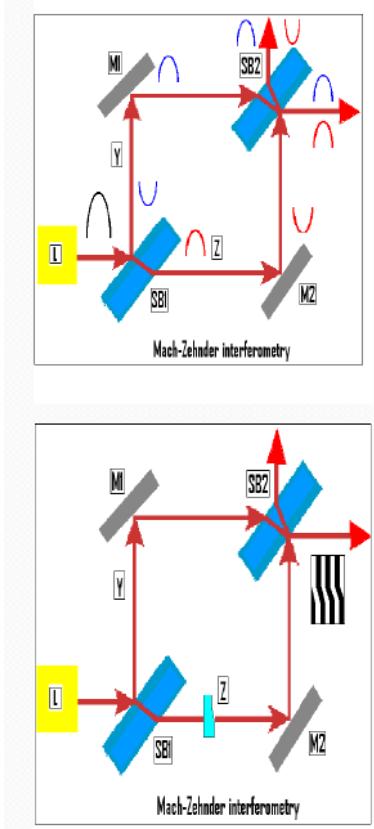


Figure 1.13: Mach-Zehnder interferometer.

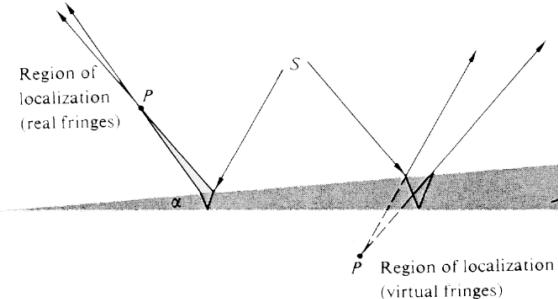


Figure 1.14: Region of localization of real and virtual fringes.

## 1.6 Type and positions of fringes

We have seen previously different type of interferometers that, according to the conditions and to the system, can produce interference pattern so fringes. Fringes can be classified first of all in **real** and **virtual** ones and, secondary, in **localized** or **non localized** ones.

Real fringes are the ones that can be seen on a screen without a lens. Virtual fringes, on the contrary, can not be projected on a screen without using a system to focus them. Fringes not localized are reals and they exist in every point of the space (as in the case of Young experience). Fringes localized can be observed only a specific region, for example on a surface. For clarification, let's see the figure 1.14.

### 1.6.1 Fringes of same inclination

Consider image 1.15. Here the source  $S$  hit the beam splitter with a specific angle and then this one separates the light in a fraction of reflected light and in a fraction of transmitted light. The light which is transmitted passes, then, thought a lens that converge the radiation in  $P$ , creating here fringes of same inclination. Here  $d$  is the width of the film,  $n_f$  is the refracting index of the film,  $\theta_t$  is the angle

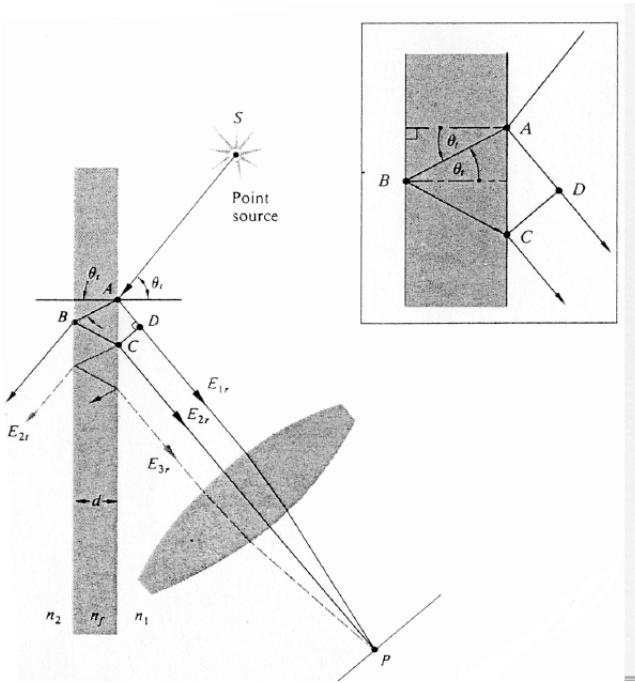


Figure 1.15: Finges of same inclination.

of incident and  $n_1$  and  $n_2$  are refractive indices before and after the beam splitter.

So the difference of optical path is:

$$\Lambda = n_f[(\bar{AB}) + (\bar{BC})] - n_1(\bar{AD}) \quad (1.11)$$

and as consequence:

$$\Lambda = \frac{2n_f d}{\cos \theta_t} (1 - \sin^2 \theta_t) = r n_f d \cos \theta_t \quad (1.12)$$

therefore the phase difference is  $\delta = k_0 \Delta \pm \pi$ . There is a maximum if  $n_f d \cos \theta_t = (2m + 1)\frac{\lambda}{4}$  or a minimum for  $n_f d \cos \theta_t = 2m\frac{\lambda}{4}$ .

### 1.6.2 Fringes of same thickness

According to figure 1.16, we can create a system with two plates of non negligible width and refractive index  $n_1$  and  $n_2$  with an inclination of  $\alpha$  one with respect to the other. In this case maximums for small angle of incident are located at:

$$(m + \frac{1}{2})\lambda = 2n_f d_m = 2\alpha x_m n_f \quad (1.13)$$

## 1.7 Formation theory of images by Abbe (1873)

An object illuminated by a plane wave front forms a figure of diffraction on the focal plane of a lens. For coherent light the formation of images is strictly correlated to the spatial frequencies in which the object can be decomposed. So long as the diffraction converts the distribution of incident luminosity intensity on an object in spacial frequencies, this phenomena can be considered mathematically like the Fourier transform, inside the space of spatial frequencies, of the intensity of the incident electric field. In other words, what you see on the focal plane, is the FT of the light made by nature so the the *focal plane = plane for FT* (image 1.17).

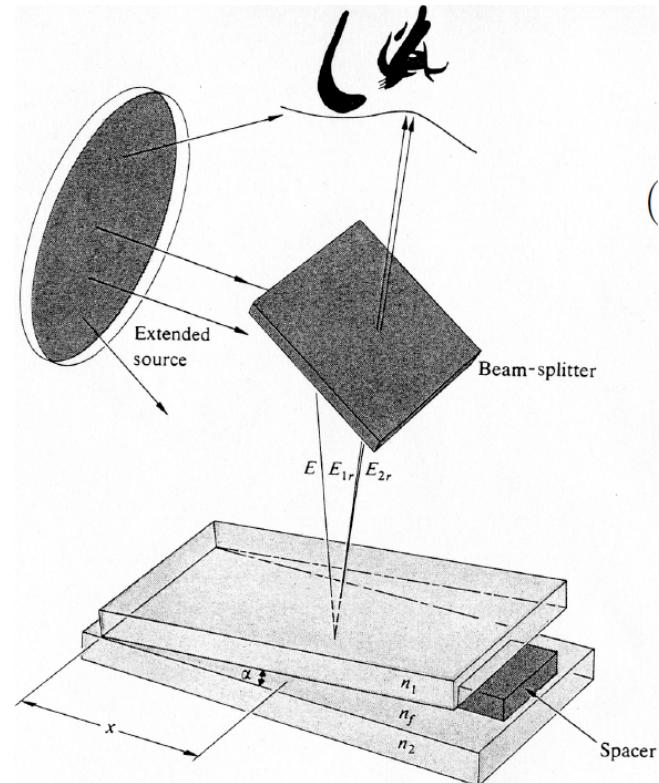


Figure 1.16: System to create fringes of same width.

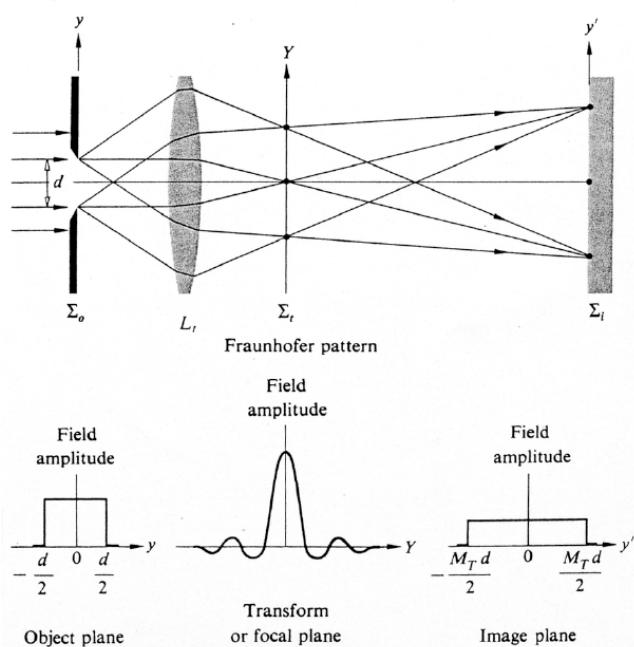


Figure 1.17: Lens realize a FT of the signal.

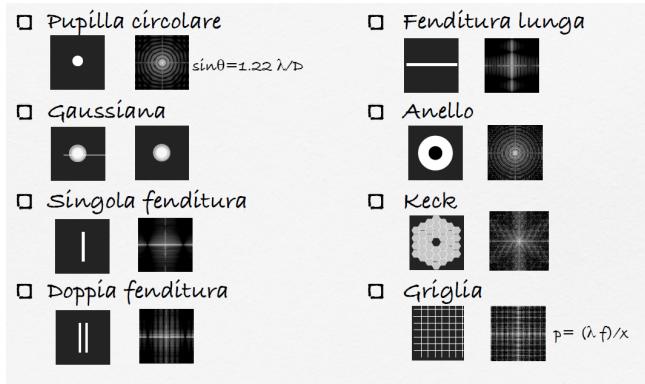


Figure 1.18: Types of apertures.

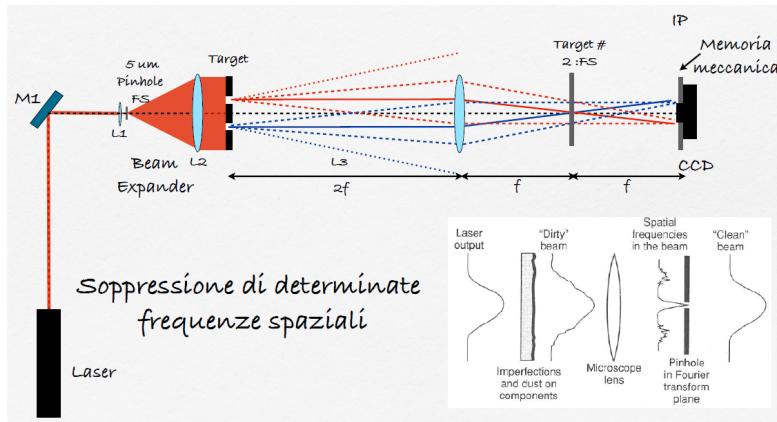


Figure 1.19: Spacial filtering.

In particular, according to the type and size of the apertures, the FT can give a different pattern of interference, as shown in figure 1.18.

For example, let's consider the laser output, a monochromatic wave, that passes thought an optical system, like in image 1.19. Even if the system is made with a great accuracy, there are of course imperfections and/or dust on the components so the result signal is instead a "dirty" beam. This one, passing thought microscope lens, selects spacial frequencies in the beam (**spatial filtering**). Then the wave passes thought a pinhole in FT plane, "cleaning" the beam. This is a type of suppression of specific spacial frequencies in order to eliminate other sources of errors. This operation is a convolution.

Now let's review some fundamental tools for this discussion, like the FT and its proprieties.

## 1.8 Fourier transform FT

### 1.8.1 FT

Consider a generic function  $f(x)$ . This can be expressed as a linear combination of an infinite number of harmonics (Fourier analysis):

$$f(x) = \frac{1}{\pi} \left[ \int_0^{+\infty} A(k) \cos kx dk + \int_0^{+\infty} B(k) \sin kx dk \right] \quad (1.14)$$

where:

$$A(k) = \int_{-\infty}^{+\infty} f(x') \cos kx' dx' \quad (1.15)$$

$$B(k) = \int_{-\infty}^{+\infty} f(x') \sin kx' dx' \quad (1.16)$$

in which  $k$  is said angular spatial frequency. In exponential notation:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(k) \exp -ikxdk \quad (1.17)$$

so we have:

$$F(k) = \int_{-\infty}^{+\infty} f(x) \exp ikxdx \quad (1.18)$$

The function  $F(k)$  inside the frequency space is called Fourier transform of the function  $f(x)$ . In other words, starting from a continuum function, we get a **discrete spectrum** in the frequency field. Due to the fact that  $F(k)$  is a complex quantity, can be written in terms of the real amplitude and of one phase:

$$F(k) = |F(k)| \exp i\phi(k) \quad (1.19)$$

It's a reversible operation, indeed the function  $f(x)$  is called **antittransform** or inverse transform.

### 1.8.2 DFT and [u,v] plane

The discrete Fourier transform (DFT) converts a finite sequence of equally-spaced samples of a function into a same-length sequence of equally-spaced samples of the discrete-time Fourier transform (DTFT), which is a complex-valued function of frequency. It is defined as:

$$\tilde{I}[u, v] = T(I)[u, v] = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} I[x, y] \exp -i2\pi(ux/M + vy/N) \quad (1.20)$$

in which  $u = 0, \dots, M - 1$  and  $v = 0, \dots, N - 1$  are the frequencies and  $[u, v]$  is the frequency plane,  $[x, y]$  are the coordinates and the two sum concern images of size  $M \times N$ . Defining  $T(I)[u, v] = |T(I)[u, v]| \exp i\phi[x, y]$ , we have:

- the **phase**:

$$\phi[x, y] = \arg(T(I)[x, y]) \quad (1.21)$$

- the **spectral energy**:

$$P[x, y] = |T(I)[x, y]|^2 \quad (1.22)$$

Moreover DFT has different properties. For example if we made a **translation** of the coordinates, then:

$$\begin{aligned} J[x, y] &= I[x, y] e^{i2\pi(u_0x/M + v_0y/N)} \\ T(J)[u, v] &= T(I)[u - u_0, v - v_0] \end{aligned} \quad (1.23)$$

$$\begin{aligned} J[x, y] &= I[x - x_0, y - y_0] \\ T(J)[u, v] &= T(I)[u, v] e^{i2\pi(x_0u/M + y_0v/N)} \end{aligned} \quad (1.24)$$

**Rotation** of angle  $\theta$  of the image is:

$$T(I_\theta)[u, v] = T(I_\theta)[u, v] \quad (1.25)$$

Finally **periodicity** is:

$$\begin{aligned} I[x, y] &= I[x + k_1 M, y + k_2 N] \\ T(I)[u, v] &= T(I)[u + k_1 M, v + k_2 N] \end{aligned} \quad (1.26)$$

### EXAMPLE - FT of a Gaussian

The Gaussian is described by  $f(x) = Ce^{-ax^2}$ . The FT is:

$$F(k) = \int_{-\infty}^{+\infty} (C \exp -ax^2) \exp ikx dx = \frac{C}{\sqrt{a}} \exp -k^2/4a \int_{-\infty}^{+\infty} \exp -\beta^2 dx \quad (1.27)$$

where  $\beta = x\sqrt{a} - ik/2\sqrt{a}$  so the final result is  $F(k) = \exp -k^2/4a$  which is again a Gaussian!

### 1.8.3 Bidimensional transform

Let's consider a bidimensional function  $f(x, y)$ , the FT of it is:

$$F(k_x, k_y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \exp i(k_x x + k_y y) dx dy \quad (1.28)$$

where  $k_x$  and  $k_y$  are the angular spatial frequencies along the two axis and they can be expressed as function of the spatial period:

$$\begin{aligned} k_x &= 2\pi/\lambda_x \\ k_y &= 2\pi/\lambda_y \end{aligned} \quad (1.29)$$

## 1.9 Linear systems and convolution

First of all let's introduce the **sifting property** that implies:

$$\int_{-\infty}^{+\infty} \delta(x) f(x) dx = f(0) \quad (1.30)$$

where  $\delta(x) = 0$  for  $x \neq 0$  and  $\delta(x) = \infty$  for  $x = 0$ . Consider now  $f(y, z)$  the entrance signal of a system and  $g(Y, Z)$  the exit signal of the same system. A system is linear if:

$$\begin{cases} af(y, z) = ag(Y, Z) \\ af_1 + bf_2 = ag_1 + bg_2 \end{cases} \quad (1.31)$$

Imposing  $g(Y, Z) = L[f(y, z)]$  and using the sifting property, then:

$$g(Y, Z) = \int \int_{-\infty}^{+\infty} f(y', z') L[\delta(y' - y)\delta(z' - z)] dy' dz' \quad (1.32)$$

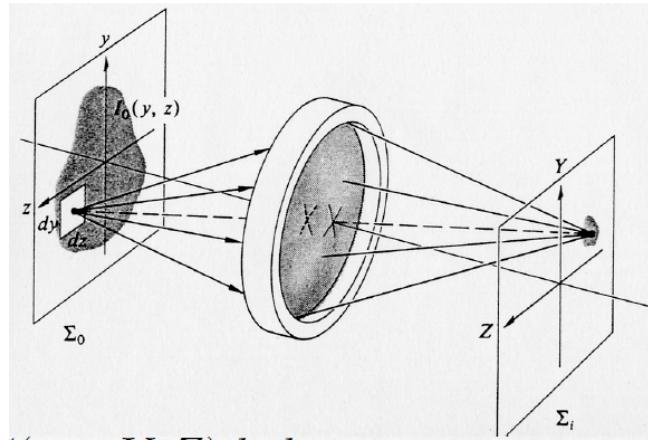


Figure 1.20: Point Spread Function.

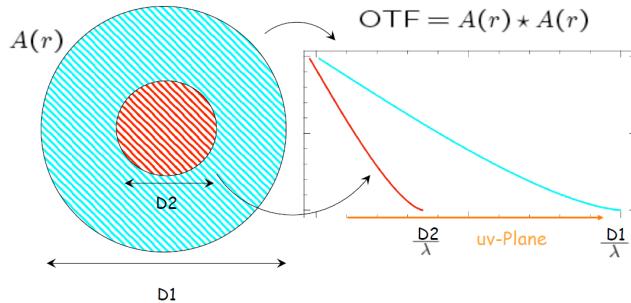


Figure 1.21: OTF as function of the diameters.

We can also define the convolution. Consider the function  $f$  and a generic impulse  $h$ , the convolution is:

$$g(X) = \int_{-\infty}^{+\infty} f(x)h(X-x)dx \quad (1.33)$$

For convolution theorem we have:

$$F\{g\} = F\{f \otimes h\} = F\{f\} \cdot F\{h\} \quad (1.34)$$

## 1.10 Point Spread Function

Analyse the following integral of convolution that refers to image 1.20 and that represents the answer of an optical system:

$$I(Y, Z) = \int \int_{-\infty}^{+\infty} I_0(y, z)S(y, z; Y, Z)dydz \quad (1.35)$$

The function  $S$  is called **point spread function (PSF)**. The PSF in many contexts can be thought of as the extended blob in an image that represents a single point object. In functional terms, it is the spatial domain version of the optical transfer function of the imaging system. In an ideal system without aberrations, the PSF coincides with the Airy Disk.

The ideal lens (or antenna) samples all spatial frequencies up to  $D/\lambda$ , so the lower frequencies are favored. Let's consider image 1.21.

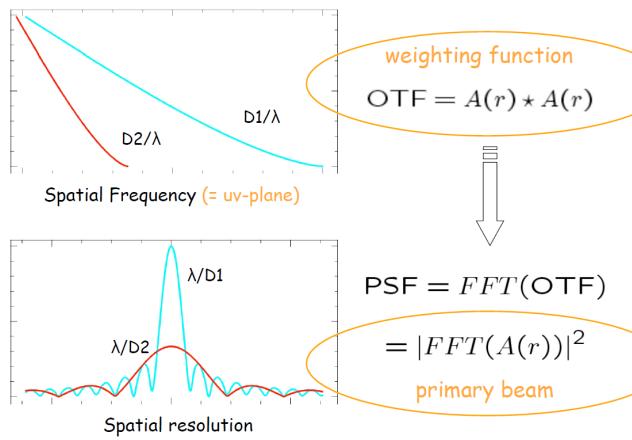


Figure 1.22: Spatial frequency and spatial resolution.

Here optical transfer function OTF is given by the convolution of  $A(r)$  with it-self and specifies how different spatial frequencies are handled by the system. In the  $[u, v]$  plane it is possible to observe the different trend according to the diameter  $D_1$  and  $D_2$ . This is a weighting function and the FFT (Fast Fourier Transform, an algorithm to implement the DFT) of the OTF is exactly the PSF. What is clear from image 1.22, is that the spatial resolution is higher with diameter  $D_1$ .

What should we draw from all this?

- An image can be decomposed into a series of spatially separated compact PSFs and can be described in terms of its Fourier components.
- Due to the Fourier decomposition, this is equivalence to a superposition of non-localized sinusoids.
- Incoherent imaging system acts as a filter for the true spatial Fourier spectrum of the source.
- Each Fourier component (or spatial frequency) is associated with a distinct physical baseline in the aperture that samples the light.
- PSF forms as arising from the relative sampling (and hence weighting given to) the different spatial frequencies (and hence baselines) measured by the pupil of the imaging system.

Moreover we have to remember that:

- The baseline is the most important thing in interferometry: depending on that, we can express all the spatial frequencies as a function of the baseline. As long as the baseline is big, we have high spatial resolution.
- Baselines as huge as the Earth diameter can be used. In this case, the technique is very different from the one in the VLT and here atomic clocks are needed. The signals are superimposed using a correlator; but this can be done only if we have a very precise information about time. In fact, a very high precision of correlation is needed.
- In short-baseline interferometers, the signals can be carried in cables and summed after being delayed of the right amount of time.
- Each image is the sum of snapshots from different antennas that maps all the  $[u, v]$  plane (that should be completely filled). If the  $[u, v]$  plane is not completely filled, the anti-transformation cannot be performed with precision. This is why many antennas are needed. An interferometer collects a huge amount of data: something like 1 TB per hour for ALMA data.

## 1.11 Single dish limitations

A **single-dish radio telescope** basically consists of a parabolic reflector which focuses incoming radio frequency energy onto a receiver/detector. These specially-designed telescopes observe the longest wavelengths of light, ranging from 1 millimeter to over 10 meters long. For comparison, visible light waves are only a few hundred nanometers long.

Radio telescopes are built in all shapes and sizes based on the kind of radio waves they pick up. However, every radio telescope has an antenna on a mount and at least one piece of receiver equipment to detect the signals.

Because radio waves are so long and cosmic radio sources are extremely weak, radio telescopes are the largest telescopes in the world, and only the most sensitive radio receivers are used inside them. Indeed the size of a dish determines the amount of incoming radiation that can be collected. The larger the collecting area, the fainter the source that can be detected. Unfortunately, these huge antennas also pick up radio interference from modern electronics, and great effort is taken to protect radio telescopes from radio frequency interference. For a single-dish radio telescope the size of the dish also determines the field-of-view of the telescope.

**Antenna** - The most versatile and powerful type of radio telescope is the parabolic dish antenna. The parabola is a useful mathematical shape that forces incoming radio waves to bounce up to a single point above it, called a focus. Dish antennae bounce many different wavelengths at once, and we need different receivers to tune to different frequency channels for the different kinds of research we do. To observe a specific wavelength range, we select a specific size funnel to grab the radio waves we want.

**Dish surface** - If the lengths of the radio waves we're studying are very small, such as the millimeter waves collected by ALMA, then the perfection of the telescope's dish surface is critical. Any warp, bump, or ding in the parabola will scatter these tiny waves away from the focus, and we'll lose information.

If the size of the radio wavelength being observed is very long, such as the centimeter waves picked up by the VLA and the VLBA, then the perfection of the dish's shape is not as critical to keep excellent observations of the radio sky. Those dishes are made rigid and tough and withstand the rigors of moving and working in various conditions.

**Angular resolution** - Angular resolution of antennas depends on wavelength of observation and on the diameter of the antenna  $D$  in the following way:

$$\sim \frac{\lambda}{D} \quad (1.36)$$

So to have a relevant angular resolution is necessary to increase the diameter of the antenna indeed to have resolution compare to optical telescopes, a radio telescope's antenna size needs to be much, much larger. Of course this can be quite challenging. To solve partially this problem we use, therefore, interferometry.

## 1.12 Summery

Summarizing, advantages of interferometers are:

- high angular resolution;
- large collecting area;
- flatter baselines;
- astrometry;
- can filter out extended emission;

- large field of view with independent pixels;
- flexible angular resolution (different configuration).

On the contrary, disadvantages of interferometers are:

- require stable atmosphere;
- high altitude and flat site (usually difficult to access);
- lots of receivers to do;
- complex correlator;
- can filter out extended emission;
- need time and different configuration to fill in the uv-plane.

# Chapter 2

## Two-element interferometers

In the previous chapter we have seen the basic concepts and ideas of interferometry. Now we can see how it works in detail the interferometry between two elements, like two antennas. To do this it is necessary first to go deeper in the concept of coherence.

### 2.1 Coherence

Coherence of light wave is defined as the correlation between the electric field values at different locations or times. The coherent light source is able to produce a coherent waves able to interfere with each other.

Ideal coherent source is a source with one wavelength (monochromatic) which does not exist in practice. Practically, there is no fully coherent light or fully incoherent light, but there are light sources with different coherence degree.

#### 2.1.1 Temporal and spatial coherence

There are two different type of coherence.

**Spatial coherence** Spatial coherence is the degree of correlation between different points on the same wave front at the same time. It is light source dependent: as the source size extends its spatial coherence degree deteriorate. The **coherent length** is:

$$\Delta S = N\lambda \quad (2.1)$$

where  $N$  is the waves numbers contained in one wave train.

**Temporal coherence** Temporal coherence is the correlation between the electric fields at the same point but at different times. It is proportionate to the wave train length. Monochromatic sources such as laser have a high degree of temporal coherence, because of the long wave trains. The **coherence time** is:

$$\Delta t = \Delta S/c \quad (2.2)$$

where  $c$  is the light speed in space.

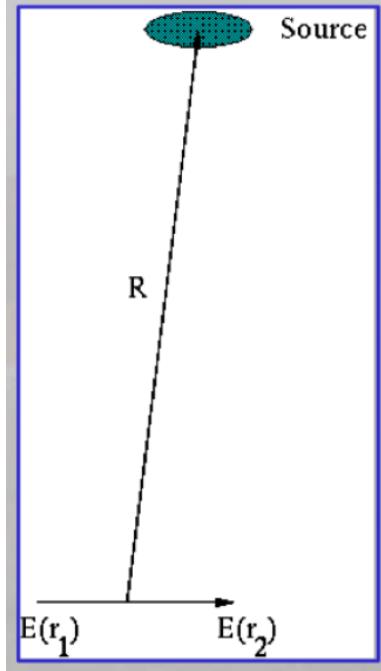


Figure 2.1: System to define temporal and spatial coherence functions.

### 2.1.2 Temporal and spatial coherence functions

Consider a source distant  $R$  from the ground where are located two devices that measures electric field at two locations  $r_1$  and  $r_2$  at times  $t_1$  and  $t_2$  (figure 2.1). Each field ( $E(r_1)$  and  $E(r_2)$ ) is composed of contributions from each element of the source. Assuming this, we define the **spatial-temporal coherence function** as:

$$V(r_1, t_1, r_2, t_2) = \langle E(r_1, t_1) \times E^*(r_2, t_2) \rangle \quad (2.3)$$

In particular we are interested in two special cases seen before:

- $t_1 = t_2$ : spatial coherent function
- $r_1 = r_2$ : temporal coherence function

## 2.2 Temporal coherence function

For astronomical sources, the temporal coherence function can be written as:

$$\langle E(r_1, t_1) \times E^*(r_2, t_2) \rangle = V(t_1 - t_2) = V(\tau) \quad (2.4)$$

In this case we should note that the coherence function does not depend on  $r_1$  but it is a function of a **time delay**  $\tau = t_1 - t_2$ . It quantifies the extent to which the fields along a given wave train are correlated. Practically, it is related to the quantity that a laboratory Michelson interferometer measures.

The temporal coherence function is associated to the **correlation coefficient**:

$$\gamma(\tau) = f(t)f^*(t + \tau)/|f(t)|^2 \quad (2.5)$$

In particular, if the function  $f(t) = Ae^{i\omega t}$ , then  $\gamma(\tau) = e^{-i\omega\tau}$  and, as consequence,  $|\gamma(\tau)| = 1$ .

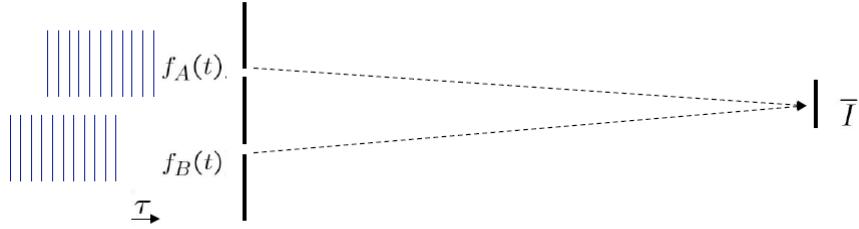


Figure 2.2: Two sub-apertures.

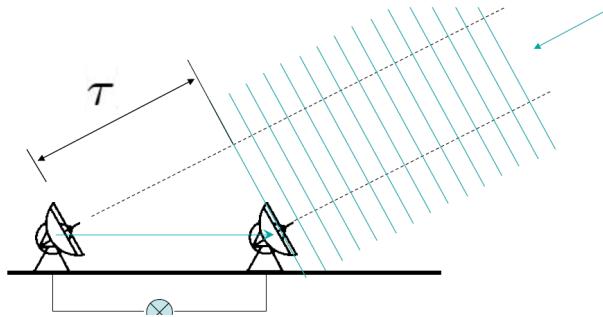


Figure 2.3: System of two antennas. Consider these two antennas separated by the baseline  $B$  and by the time delay  $\tau$ . The correlator that connects the two antennas measure the degree of correlation between the two beams received. However an additional device is necessary: a delay devise that compensate the time delay and communicate this to the computer. In the case of optical telescopes the correlator is provided by the nature and there is no time delay: in this case if the source is sufficiently coherent then fringes are visible. In both cases to have a higher resolution you have to increase the baseline so the distance between antennas but pay attention! If the baseline is to small then you don't observe small details or you can not resolve the source but if the baseline is too large you over-resolve the source! So there must be the right compromise. For example to have a resolution of milliarcseconds a baseline like the diameter of the Earth is necessary.

Let's now calculate the correlation coefficient of two apertures, like in the case of Young experiment (figure 2.2):

$$\gamma_{AB} = f_A(t)f_B^*(t + \tau)/\sqrt{|f_A(t)|^2|f_B(t)|^2} \quad (2.6)$$

Then the visibility, defined in the previous chapter, became:

$$V = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} = 2|\gamma_{AB}|(\vec{I}_A \vec{I}_B)^{1/2}/(\vec{I}_A + \vec{I}_B) \quad (2.7)$$

In particular if  $\vec{I}_A = \vec{I}_B$  then the visibility became  $V = |\gamma_{AB}|$ .

From this discussion we are now able to infer a new definition for interferometer: they are able to measure the temporal coherence of the incoming wave front. Consider, for example, two antenna with a time delay  $\tau$  and connected by the correlator (figure 2.3). This last one measures the visibility given by  $V = \gamma(\tau)$ .

**Use of the temporal coherence function** The importance of the temporal coherence function arises from an important result in physics, the **Wiener-Khintchine theorem**. This theorem says that the normalized value of the temporal coherence function  $V(\tau)$  is equal to the normalized Fourier transform of the spectral energy distribution  $B(\omega)$ , of the source:

$$V(\tau) = \frac{\int B(\omega)e^{-i\omega\tau}d\omega}{\int B(\omega)d\omega} \quad (2.8)$$

In particular, a broad spectral energy distribution leads to a coherence function that decays rapidly since  $\tau$  and  $\omega$  are reciprocal coordinates.

We can also define a **coherence time**  $\tau_{coh} \sim 1/\Delta\nu$  where  $\Delta\nu = \Delta\omega/2\pi$  is the spectral bandwidth of the radiation. Moreover, measurements of  $V(\tau)$  allow to recovery of the source spectrum.

## 2.3 Spatial coherence function

For astronomical sources, this coherence function can be written as:

$$\langle E(r_1, t_1) \times E^*(r_2, t_2) \rangle = V(r_1 - r_2) = V(\rho) \quad (2.9)$$

In this case we see that, this coherence function does not depend on  $t_1$  and it is a function of a vector separation  $\rho = r_1 - r_2$ . It quantifies the correlations between different spatial locations on a wave front. Considering the Young experiment, it corresponds to the quantity that a Young's slit experiment investigates (on axis).

**Use of the spatial coherence function** The importance of the spatial coherence function arises from another important result in physics, the **van Cittert-Zernike theorem**. This states that, for incoherent sources in the far-field, the normalized value of the spatial coherence function  $V(\rho)$  is equal to the normalized Fourier transform of the brightness distribution in the sky  $I(\alpha)$ :

$$V(\rho) = \frac{\int I(\alpha) e^{-i2\pi/\lambda(\alpha,\rho)} d\rho}{\int I(\alpha) d\rho} \quad (2.10)$$

or in slightly different notation:

$$V(u, v) = \frac{\int \int I(l, m) e^{-i2\pi(ul+vm)} dl dm}{\int \int I(l, m) dl dm} \quad (2.11)$$

where  $u$  and  $v$  are the components of the baseline  $\rho$  measured in wavelengths, and  $l$  and  $m$  are angular coordinates on the sky.

## 2.4 What should we draw from all this?

Measurements of these coherence functions allow us to interrogate a source without using a conventional imaging telescope. This brings to measurements of **time-averaged products** of field quantities like  $\langle E(r_1) \times E^*(r_2) \rangle$ . The relationships between the source parameters and the coherence functions is a Fourier transform. Hence it is *linear, invertible* and *complex*.

We note the mathematical equivalence of the spatial coherence function  $V(\tau = 0, \rho)$  and the Fourier decomposition of an image we referred to earlier.

## 2.5 Spatial interferometry

We can put this all together in the following form:

- we can describe a source in the sky as a superposition of co-sinusoids, each of which corresponds to a given spatial frequency;
- measurements of the coherence function are in fact measurements of the strength of each of those Fourier components;
- interferometers are merely devices to measure the coherence function;

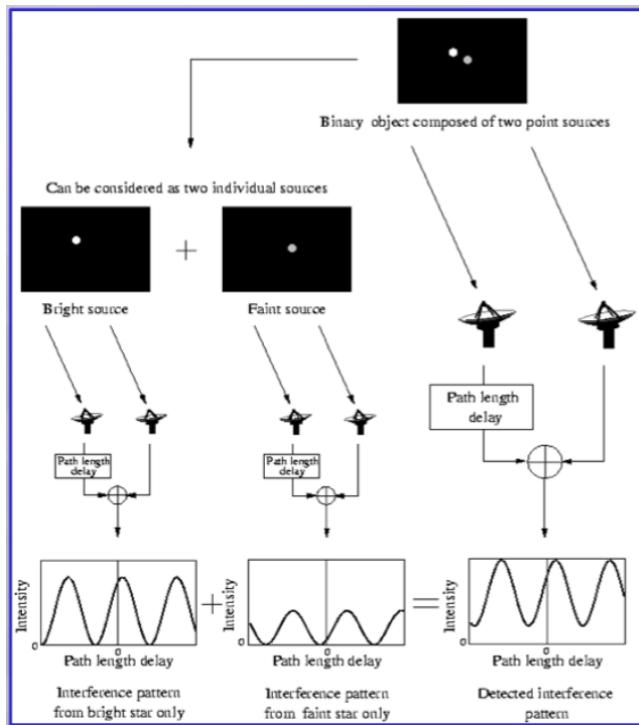


Figure 2.4: Response of two-element interferometry.

- two telescopes with a projected separation  $B$  will measure the value of the Fourier transform of the source brightness distribution at a spatial frequency  $u = B/\lambda$ ;
- telescopes do all of this for you for a range of baselines at once for free.

### 2.5.1 Spatial coherence revisited

Consider the response of an two-element interferometer observing a star comprising two separate infinitesimally small sources (figure 2.4).

As before, the resulting fringe pattern has a **modulation depth** that is reduced with respect to that from each source individually. Note how the positions of the sources are encoded in the fringe phase.

## 2.6 Two-element interferometers

The coherence function  $\Gamma(u, \tau)$  is measured by correlating the outputs of two antenna systems. The simplest example of this process is a two-element interferometer. Let us assume that the interferometer consists of two antennas  $A_1$  and  $A_2$  separated by the distance  $B$  (directed from  $A_2$  to  $A_1$ ), and that both antennas are sensitive only to radiation of the same state of polarization (see figure 2.5).

A plane electromagnetic wave (from a very distant source) of amplitude  $E$  induces the voltage  $U_1$  at the output of antenna  $A_1$ :

$$U_1 \propto E e^{i\omega t} \quad (2.12)$$

while at  $A_2$  we obtain:

$$U_2 \propto E e^{i\omega(t-\tau)} \quad (2.13)$$

where  $\tau$  is the geometric delay caused by the orientation of the interferometer baseline  $B$  relative to the direction of the wave propagation. The outputs will be correlated. In a correlation the signals are

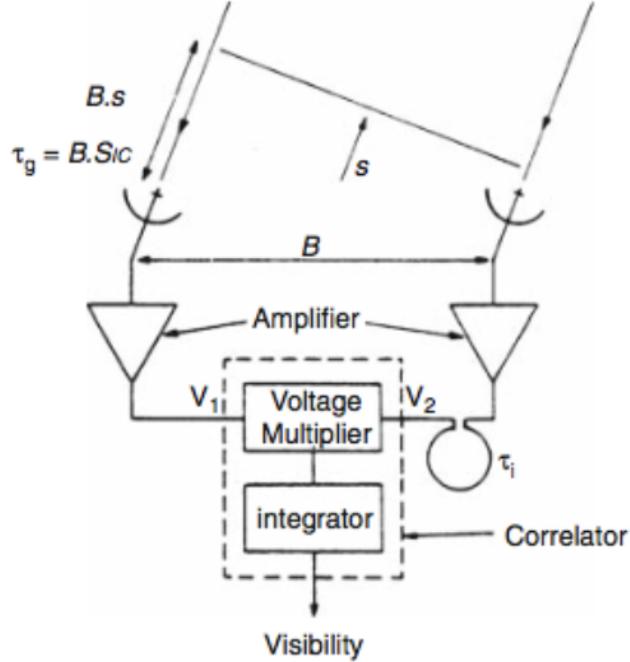


Figure 2.5: Two-element interferometry

input to a multiplying device followed by an integrator. The output is proportional to:

$$R(\tau) \propto \frac{E^2}{T} \int_0^T e^{i\omega t} e^{-i\omega(t-\tau)} dt \quad (2.14)$$

If \$T\$ is a time much longer than the time of a single full oscillation, i.e., \$T \gg 2\pi/\omega\$ then the average over time \$T\$ will not differ much from the average over a single full period, that is:

$$R(\tau) \propto \frac{\omega}{2\pi} E^2 \int_0^{2\pi/\omega} e^{i\omega\tau} dt \propto \frac{\omega}{2\pi} E^2 e^{i\omega\tau} \int_0^{2\pi/\omega} dt \propto \frac{1}{2} E^2 e^{i\omega\tau} \quad (2.15)$$

The output of the correlator+integrator thus varies periodically with \$\tau\$, the delay time; this output is the mutual coherence function of the received wave. If the relative orientation of interferometer baseline \$B\$ and wave propagation direction \$s\$ remain invariable, \$\tau\$ remains constant, so does \$R(\tau)\$. But since \$s\$ is slowly changing due to the rotation of the earth, \$\tau\$ will vary, and we will measure interference fringes as a function of time. In order to understand the response of interferometers in terms of measurable quantities, we consider a two-element system. The basic constituents are shown in figure 2.5. If the radio brightness distribution is given by \$I\_\nu(s)\$, the power received per bandwidth \$d\nu\$ from the source element \$d\Omega\$ is \$A(s)I\_\nu(s)d\Omega d\nu\$, where \$A(s)\$ is the effective collecting area in the direction \$s\$; we will assume the same \$A(s)\$ for each of the antennas. The amplifiers are assumed to have constant gain and phase factors which we neglect for simplicity. The output of the correlator for radiation from the direction \$s\$ is:

$$r_{12} = A(s)I_\nu(s)e^{i\omega\tau} d\Omega d\nu \quad (2.16)$$

where \$\tau\$ is the difference between the geometrical and instrumental delays \$\tau\_g\$ and \$\tau\_i\$. If \$B\$ is the baseline vector for the two antennas:

$$\tau = \tau_g - \tau_i = \frac{1}{c} \mathbf{B} \cdot \mathbf{s} - \tau_i \quad (2.17)$$

and the total response is obtained by integrating over the source \$S\$:

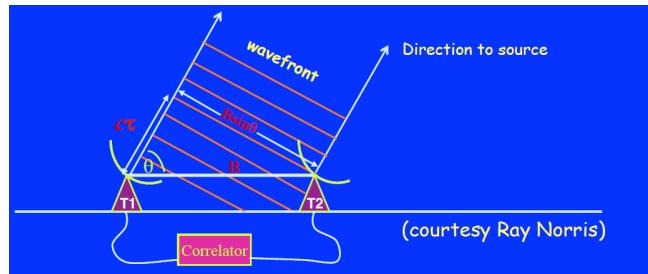


Figure 2.6: Two-element interferometry.

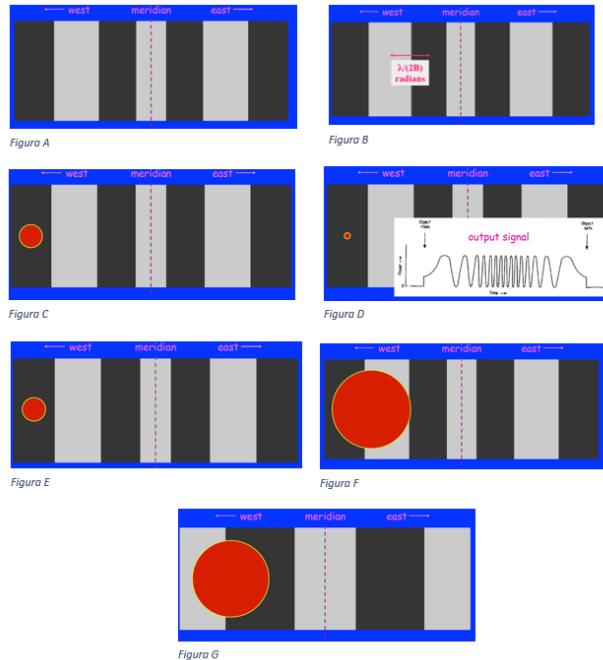


Figure 2.7: Fringe pattern "in the sky".

$$R(\mathbf{B}) = \int \int_{\Omega} A(s) I_{\nu}(s) \exp \left[ i 2\pi \nu \left( \frac{1}{c} \mathbf{B} \cdot s - \tau_i \right) \right] d\Omega d\nu \quad (2.18)$$

This function  $R(B)$ , the Visibility Function is closely related to the mutual coherence function of the source but, due to the power pattern  $A(s)$  of the individual antennas, it is not identical to  $\Gamma(B, \tau)$ . For parabolic antennas it is usually assumed that  $A(s) = 0$  outside the main beam area so that the previous relation is integrated only over this region. A one dimensional version of the last equation, with a baseline  $B$ ,  $\nu = \nu_0$  and  $\tau_i = 0$ , is:

$$R(B) = \int A(\theta) I_{\nu}(\theta) \exp \left[ i 2\pi \nu_0 \left( \frac{1}{c} B \cdot \theta \right) \right] d\theta \quad (2.19)$$

### 2.6.1 Resolution of two-element interferometry

As seen previously, an interferometer measures coherence in the electric field between two pairs of points (the baseline). Consider, for example, image 2.6. As seen many times, because of the geometric path difference  $c\tau$ , the incoming wave-front arrives at each antenna at a different phase.

In particular, if the interferometer is a two-element East-West system, by the analogy to the double slit experiment, regions which would cause constructive and destructive interference can be considered like "stripes" in the sky (see figure 2.7, panel A). The angular resolution of the interferometer is given by fringe half-spacing  $\lambda/(2B)$  (panel B). However, as the source moves though the fringe pattern, it

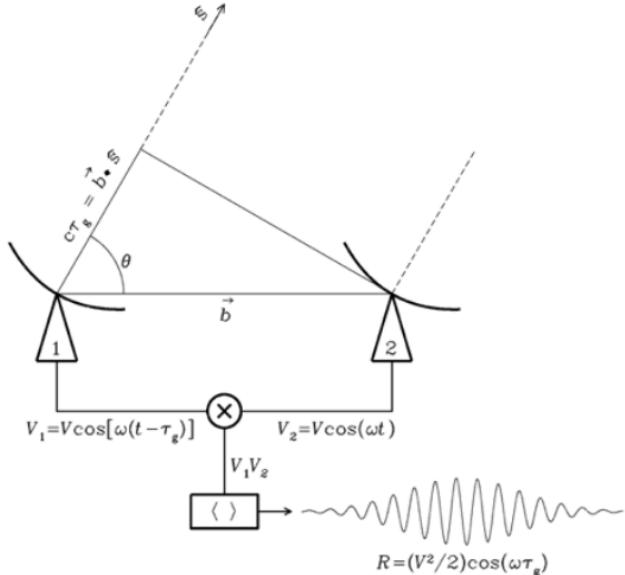


Figure 2.8: This block diagram shows the components of the simplest two-element interferometer observing in a very narrow frequency range centered on  $\nu = \omega/(2\pi)$ . The correlator multiplies and averages the voltage outputs  $V_1$  and  $V_2$  of the two dishes.  $\vec{s}$  is the unit vector in the direction of a distant point source and  $\vec{b}$  is the vector baseline from antenna 1 to antenna 2. The output voltage  $V_1$  of antenna 1 is the same as the output voltage  $V_2$  of antenna 2, but it is retarded by the geometric delay  $\tau_g = \vec{b} \cdot \vec{s}/c$ . These voltages are multiplied and time averaged by the correlator to yield an output response whose amplitude  $R$  is proportional to the point source flux density and whose phase depends on the delay and the frequency. The quasi-sinusoidal output fringe shown occurs if the source direction in the interferometer frame is changing at a constant rate  $d\theta/dt$ . The broad Gaussian envelope of the fringe is caused by primary-beam attenuation if the individual dishes do not track the source.

produces an oscillating output signal from the interferometer (panel C). If the source is very small compared to the fringe half-spacing  $\lambda/(2B)$ , we say it is **unresolved**. The output signal is just the fringe pattern, and the source structure cannot be determined (panel D). On the contrary, if the source is comparable to the fringe half-spacing  $\lambda/(2B)$ , then the output signal is the fringe pattern smoothed by the finite size of the source (pattern E). If the source is large enough to span both a peak and a trough in the fringe pattern, the output signal is nearly constant. The source is **over-resolved** or **resolved out**, and its structure poorly determined (panel F). If you are interested in source structure that is being resolved out, then observe with a shorter baseline  $B$  to make the fringe pattern  $\lambda/B$  larger (panel G).

### 2.6.2 The correlator of the two-element interferometry

As seen previously, the basic interferometer is a pair of radio telescopes whose voltage outputs are *correlated* (multiplied and averaged). Even the most elaborate interferometers with  $N \gg 2$  elements can be treated as  $N(N - 1)/2$  independent interferometer pairs, so we begin by analysing the simplest case, a two-element narrow-band interferometer. Let's consider image 2.8.

The **correlator** first multiplies these two voltages to yield the product:

$$V_1 V_2 = V^2 \cos \omega t \cos [\omega(t - \tau_g)] = (V^2/2)[\cos(2\omega t - \omega \tau_g) + \cos \omega \tau_g] \quad (2.20)$$

and then takes a time average long enough [ $\Delta t \gg (2\omega)^{-1}$ ] to remove the high-frequency term  $\cos 2\omega t - \omega \tau_g$  from the final output  $R$ :

$$R = \langle V_1 V_2 \rangle = (V^2/2) \cos \omega \tau_g \quad (2.21)$$

The amplitudes  $V_1$  and  $V_2$  are proportional to the electric field produced by the source multiplied by the voltage gains of antennas 1 and 2. Thus the output amplitude  $V^2$  is proportional to the point-source flux density  $S$  multiplied by  $(A_1 A_2)^{1/2}$ , where  $A_1$  and  $A_2$  are the effective collecting areas of the two antennas. Uncorrelated noise from the receivers and the atmosphere over the two telescopes does not appear in the correlator output, so fluctuations in receiver gain or atmospheric emission are much less significant than for a total-power observation with a single dish. Pulsed interference with duration  $t \ll |\vec{b}|/c$  is also suppressed because it usually does not reach both telescopes simultaneously.

The correlator output voltage  $R = (V^2/2) \cos \omega \tau_g$  varies sinusoidally with the change of source direction in the interferometer frame. These sinusoids are called **fringes** and the **fringes phase** is:

$$\phi = \omega \tau_g = \frac{\omega}{c} b \cos \theta \quad (2.22)$$

and it depends on  $\theta$  as follows:

$$\frac{d\phi}{d\theta} = \frac{\omega}{c} b \sin \theta = 2\pi \left( \frac{b \sin \theta}{\lambda} \right) \quad (2.23)$$

The fringe period ( $\Delta\phi = 2\pi$ ) corresponds to an angular change  $\Delta\theta = \lambda/(b \sin \theta)$ . The fringe phase is exquisitely measure of source position if the **projected baseline**  $b \sin \theta$  is many wavelengths long. Note that fringe phase and hence measured source position is not affected by small tracking errors of the individual telescopes. It depends on time, and times can be measured by clocks with much higher accuracy than angles (ratios of lengths of moving telescopes parts) can be measured by rulers. Also, an interferometer whose baseline is horizontal is not affected by the plane-parallel component of atmospheric refraction, which delays the signals reaching both telescopes equally. Consequently, interferometers can determine the positions of compact radio sources with unmatched accuracy. Absolute positions with errors as small as  $\sigma_\theta \sim 10^{-3}$  arcsec and differential positions with errors down to  $\sigma \sim 10^{-5}$  arcsec  $< 10^{-10}$  rad have frequently been measured.

A resume of math behind the signals received by the two antennas and about the rule of the correlator is provided in images 2.9 and 2.10.

The signal arrives at one of the antennas first and then, after  $\tau_g$ , gets to the other. In case it gets into  $i$  first, then in  $j$  the signal is

$$(V^i V^j) = V^2 \sin(\omega t + \phi^i) \sin[\omega(t + \tau_g) + \phi^j]$$

after some (boring) algebra and approximations we obtain:

$$V^i V^j \approx V^2 \cos(\omega \tau_g) = V^2 \cos\left(2\pi\nu \frac{D}{c} \sin\theta\right) = V^2 \cos\left(2\pi \frac{D}{\lambda} \sin\theta\right)$$

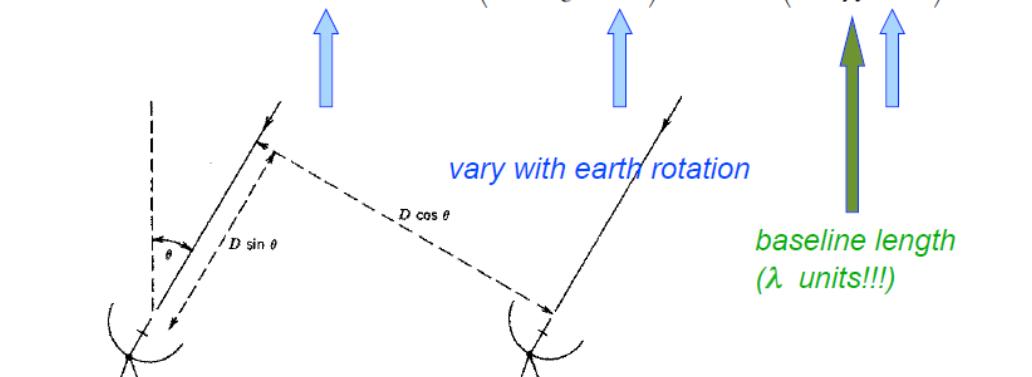


Figure 2.9: Mathematical summary of the way in which the two signals are treated by the two antennas. Pay attention on what depends on Earth Rotation!

Indeed the **CORRELATOR** performs a more complicated operation (i.e. the true cross-correlation) to deliver **VISIBILITIES**:

$$V^{ij}(\tau_g) = (V^i V^j) = \lim_{T \rightarrow \infty} \int_{-T/2}^{T/2} V^i(t) V^{j*}(t + \tau_g) dt$$

In the (2-D) uv-plane each visibility samples the FT of the (2-D)  $B(\theta, \phi)$

**Modern correlators:**

Are special computing devices  
Handle and deliver a **HUGE** amount of data

**Visibilities are:**

Complex numbers (amplitude & phase), with ancillary information  
Computed over the integration time =  $T$

Figure 2.10: Summary of how works a correlator of two-element interferometry.

# Chapter 3

## Measurements of fringes

Let's consider the different types of interferometers we discussed in chapter 1. In what sense do laboratory set-ups like a Michelson or Young's slit experiment measure coherence functions?

The detector receives contributions from each slit: it receives the electric field  $E_1$  and  $E_2$ . The measured **time average intensity** is:

$$\langle (E_1 + E_2) \times (E_1 + E_2)^* \rangle = \langle |E_1|^2 \rangle \langle |E_2|^2 \rangle \langle 2|E_1||E_2| \cos \phi \rangle \quad (3.1)$$

where  $\phi$  is the phase difference between  $E_1$  and  $E_2$ . So the properties of the fringe pattern encode the coherence function.

Consider the response, for example, the response of a Michelson interferometer to a range of wavelength. It gives a resulting fringe pattern whose modulation depth decreases as the delay between the interfering beams increase. Note that the fringe modulation disappears when the delay is  $D = \lambda^2/\Delta\lambda = l_{coh}$ .

### 3.1 Measurements of fringes

From an interferometric point of view, the key features of any interference fringe are its modulation and its location with respect to some reference point. In particular we can identify:

- the **fringe visibility** defined in the previous chapter;
- the **fringe phase**: it is the location of the white light fringe as measured from some reference (in radians).

These measure the amplitude and phase of the complex coherence function, respectively.

Usually measurements are carried out in 1 of 3 ways.

#### 3.1.1 First procedure

In the first procedure, measurements of the source of interest and a calibrator are made. This is as in the case of single telescope position switching. One significant difference with single dish measurements is that the interferometer measurement extends over a wide range of hour angles to provide a better coverage of the  $[u, v]$  plane, if the baseline is East-West. One first measures a calibration source or reference source, which has a known position and size, to remove the effect of instrumental phases in the instrument and atmosphere and to calibrate the amplitudes of the sources in question. Sources and calibrators are usually observed alternately. The time variations caused by instrumental and weather effects must be slower than the time between measurements of source and calibrator. If, as is the case for millimeter and sub-mm wavelength measurements, weather is an important influence, one must

switch frequently between on and off. In fast switching one might spend 10 s on a nearby calibrator, then a few minutes on-source (with single exposures of 10 – 20 milliseconds). This method will reduce the amount of phase fluctuations (in this small time interval the atmosphere can be considered frozen), but also the amount of time available for source measurements. For more rapid changes in the earth's atmosphere, one can correct the phase using measurements of atmospheric water vapor, or changes in the system noise temperature of the individual receivers caused by atmospheric effects. The corrections for instrumental amplitudes and phases are assumed to be constant over the times when the source is observed. The ratio of amplitudes of source and calibrator are taken to be the normalized source amplitudes. Since the calibrators have known flux densities and positions, the flux densities and positions of the sources can be determined. The reference source should be as close to the on-source as possible, but must have a large enough intensity to guarantee a good signal-to-noise ratio after a short time. Frequently nearby calibrators are time variable over months, so a more distant calibrator with a known or fixed flux density is measured at the beginning or end of the session. This source is usually rather intense, so may also serve as a bandpass calibrator for spectral line measurements. The length of time spent on the off-source measurement is usually no more than a few minutes.

### 3.1.2 Second procedure - snapshots

In the next procedure, the so-called **snapshots**, one makes a series of short observations (at different hour angles) of a large number of sources. For sensitivity reasons these are usually made in the radio continuum or intense maser lines. As in the first observing method, one intersperses measurements of a calibration source which has a known position and size to remove the effect of instrumental phases in the instrument and atmosphere and to calibrate the amplitudes of the sources in question. The images are affected by the shape of the synthesized beam of the interferometer system. If the size of the source to be imaged is comparable to the primary beam of the individual telescopes, the power pattern of the primary beams will have a large effect. This effect can be corrected easily.

### 3.1.3 Third procedure

In the third procedure, one aims to produce a high-resolution image of a source where the goal is either high dynamic range or high sensitivity. The dynamic range is the ratio of the highest to the lowest brightness level of reliable detail in the image. This may depend on the signal-to-noise ratio for the data, but for centimeter aperture synthesis observations, spurious features in the image caused by the incomplete sampling of the  $[u, v]$  plane are usually more important than the noise. Frequently one measures the source in a number of different interferometer configurations to better fill the  $[u, v]$  plane. These measurements are taken at different times and after calibration, the visibilities are entered into a common data set. In order to eliminate the loss of source flux density due to missing short spacings, one could supplement the interferometer data with single dish measurements. The diameter of the single dish telescope should be larger than the shortest spacing between interferometer dishes. This single dish image must extend to the FWHP of the smallest of the interferometer antennas. When Fourier transformed and appropriately combined with the interferometer response, this data set has no missing flux density. Such “missing spacings” are frequently a problem with interferometer images. Usually, interferometer images have shortcomings.

# Chapter 4

## Imaging with interferometers

Imaging with interferometers is based on van Cittert-Zernike theorem. Indeed, according to this last one, the Fourier transform of the brightness distribution is the coherence or the visibility function  $V(u, v) = V(B_x/\lambda, B_y/\lambda)$ . So in principle the strategy is straightforward:

- measure  $V$  for as many values of  $B$  as possible;
- perform an inverse Fourier transform so perform the image of the source.

But we need to consider the following topics:

- typical visibility functions - what do they look like?
- how complete do the measurement of  $V(u, v)$  have to be?
- what is the nature of the images that can be recovered?

Note that all of this will assume the absence of a turbulent atmosphere.

### 4.1 Simple source

To go deeper in imaging with interferometry, let's consider three different cases, taking into account that here we explore one-dimensional examples for simplicity so we assume that:

$$V(u) = \frac{\int I(l)e^{-i2\pi(ul)}dl}{\int I(l)dl} \quad (4.1)$$

**A single point source** Consider a point source of strength  $A_1$  and located at angle  $l_1$  relative to the optical axis. Then:

$$V(u) = \frac{\int A_1\delta(l - l_1)e^{-i2\pi(ul)}dl}{\int A_1\delta(l - l_1)dl} = e^{-i2\pi(ul_1)} \quad (4.2)$$

Here the visibility amplitude is unity per  $u$  and visibility phase varies linearly with  $u$  ( $= B/\lambda$ ). See figure 4.1. Sources such as this are easy to observe, but of little interest if you have built an interferometer for high-angular resolution imaging.

**Double source** Consider now a double source comprising point sources of strength  $A_1$  and  $A_2$  located at angles 0 and  $l_2$  relative to the optical axis. Then:

$$V(u) = \frac{[\int A_1\delta(l) + A_2\delta(l - l_2)]e^{-i2\pi(ul)}dl}{\int [A_1\delta(l) + A_2\delta(l - l_2)]dl} = [A_1 + A_2e^{-i2\pi(ul_2)}]/[A_1 + A_2] \quad (4.3)$$

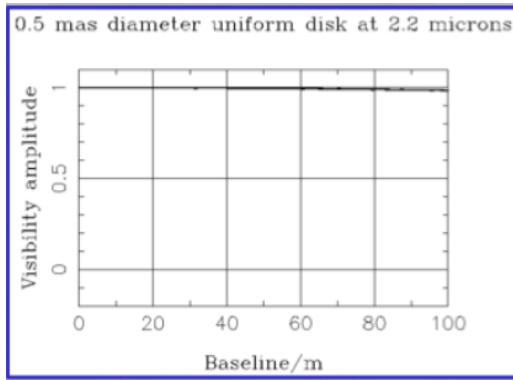


Figure 4.1: Single point source.

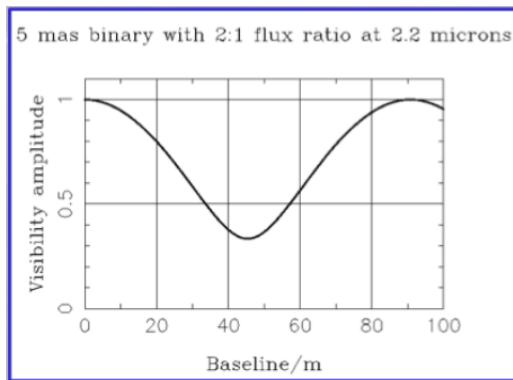


Figure 4.2: Double source.

The visibility amplitude and phase oscillate as functions of  $u$ . To identify this as binary, baselines from 0 to  $\lambda/l_2$  are required. If the ratio of component fluxes is large the modulation of the visibility becomes increasingly difficult to measure. See figure 4.2.

**Uniform source** Consider now a uniform on-axis disc source of diameter  $\theta$ . In this case we have:

$$V(u_r) \propto \int^{\theta/2} \rho J_0(2\pi\rho u_r) d\rho = 2J_1(\pi\theta u_r)/(\pi\theta u_r) \quad (4.4)$$

In this case, to identify this as a disc requires baselines from 0 to  $\lambda/\theta$  at least. The visibility amplitude falls rapidly as  $u_r$  increases. Information on scales smaller than the disc diameter correspond to values of  $u_r$  where  $V \ll 1$ , and is hence difficult to measure. See figure 4.3.

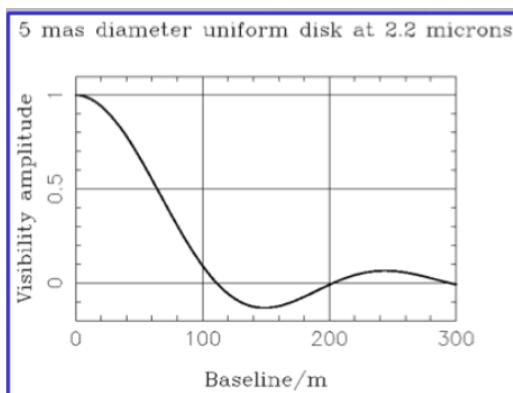


Figure 4.3: Uniform source.

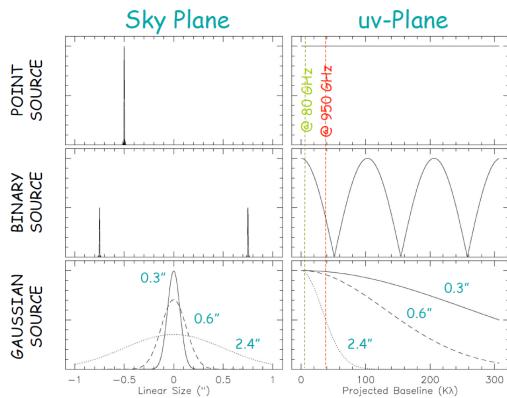


Figure 4.4: Summery of the three previous cases.

A summery of all these cases is provided in image 4.4.

## 4.2 Image reconstruction

To proceed with the image reconstruction, we start with the fundamental relationship between the visibility function and the normalized sky brightness:

$$I_{norm}(l, m) = \int \int V(u, v) e^{+i2\pi(ul+vm)} du dv \quad (4.5)$$

In practice what we measure is a sampled version of  $V(u, v)$ , so the image we have access to is the so-called *dirty map*:

$$I_{dirty}(l, m) = \int \int S(u, v) V(u, v) e^{+i2\pi(ul+vm)} du dv = B_{dirty}(l, m) * I_{norm}(l, m) \quad (4.6)$$

where  $B_{dirty}(l, m)$  is the Fourier transform of the sampling distribution or *dirty beam*. The dirty beam is the interferometer PSF, and while it in general is far less attractive than an Airy pattern, its shape is completely determined by the samples of the visibility function that are measured.

## 4.3 Data reduction

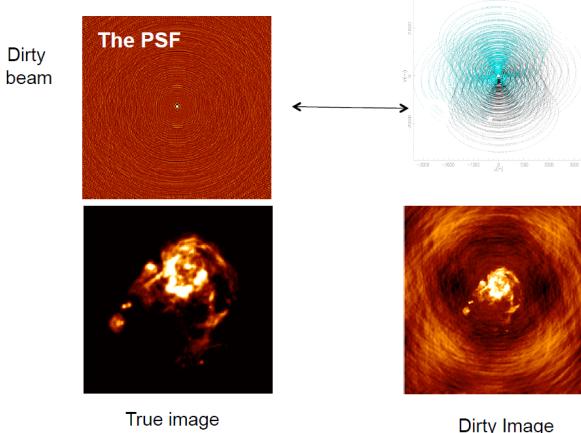
After obtaining raw visibilities, the usual procedure (described in the main points) is the following:

- **calibration** of visibilities using data from one or more bright point sources, observed at regular intervals during the observation;
- establish the **flux density scale**  $J_y$  using a standard source;
- **inverse Fourier transform** to make a dirty map;
- **deconvolution** to remove artefacts due to the PSF.

To be more specific, it is applied the so-called **Hogbom's CLEAN algorithm**. It consists in, first of all, locating the peak in the map. Then it is subtracted off a scaled version of the PSF, so the dirty beam. This operation is repeated until only noise is left in image. Then it is necessary to add back the subtracted components in the form of Gaussians (**clean beams**) with size comparable to the centre of the PSF.

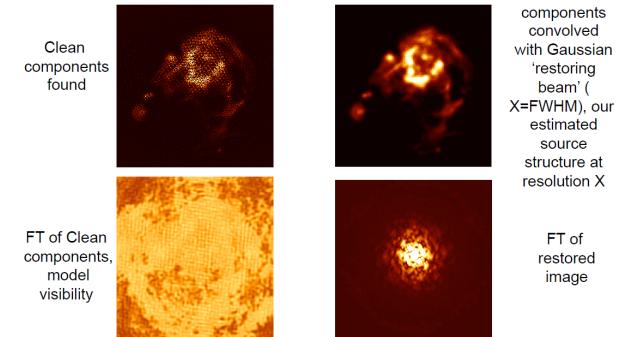
A synthetic scheme is shown in figure 8.2.

## The Dirty Image



(a) Dirty image

## Clean: Example



(b) CLEAN procedure

Figure 4.5: Data reduction

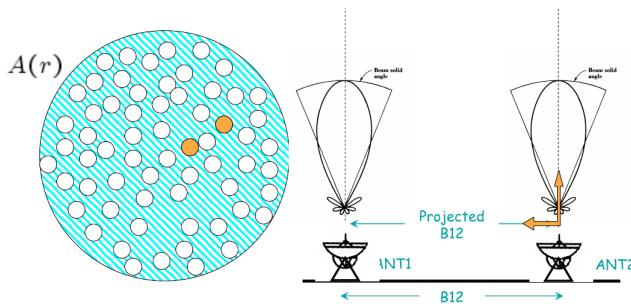


Figure 4.6: Aperture synthesis.

### Important general rules

- The number of visibility data is bigger than the number of filled pixels in the recovered image indeed:
$$N(N - 1)/2 \times (\text{number of reconfigurations}) \geq (\text{number of filled pixels}) \quad (4.7)$$
- The distribution of samples should be as *uniform* as possible to aid (aiutare) the de-convolution process.
- The range of interferometer baselines ( $B_{max}/B_{min}$ ) will govern the range of spatial scales in the map.
- There is no need to sample the visibility function too finely. For a source of maximum extent  $\theta_{max}$ , sampling very much finer than  $\Delta \sim 1/\theta_{max}$  is unnecessary.

## 4.4 Aperture synthesis

Aperture synthesis or synthesis imaging is a type of interferometry that mixes signals from a collection of telescopes to produce images having the same angular resolution as an instrument the size of the entire collection. At each separation and orientation, the lobe-pattern of the interferometer produces an output which is one component of the Fourier transform of the spatial distribution of the brightness of the observed object. The image (or "map") of the source is produced from these measurements. In other words, it is a technique by which a large telescope is replaced by a number of smaller telescopes for a total aperture  $A(r)$ . See figure 4.6.

Aperture synthesis is possible only if both the amplitude and the phase of the incoming signal are

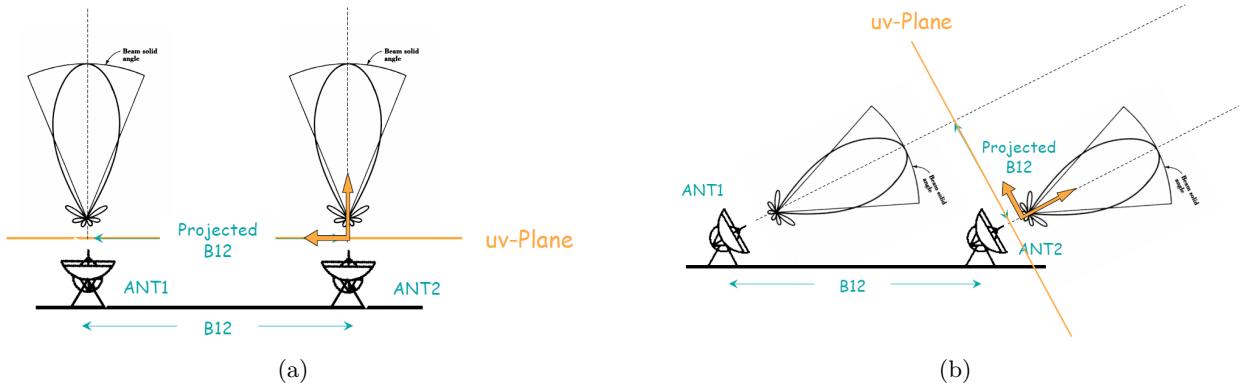
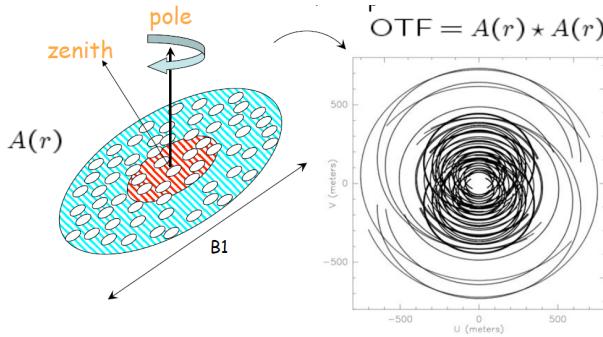
Figure 4.7:  $[u, v]$  plane according to the direction of observation

Figure 4.8: Earth Rotation Synthesis.

measured by each telescope. For radio frequencies, this is possible by electronics, while for optical frequencies, the electromagnetic field cannot be measured directly and correlated in software, but must be propagated by sensitive optics and interfered optically. Accurate optical delay and atmospheric wavefront aberration correction are required, a very demanding technology that became possible only in the 1990s. This is why imaging with aperture synthesis has been used successfully in radio astronomy since the 1950s and in optical/infrared astronomy only since the turn of the millennium.

In order to produce a high quality image, a large number of different separations between different telescopes is required so different values of baselines. The number of baselines  $n_b$  for an array of  $n$  telescopes is given by  $n_b = (n^2 - n)/2$ . For example, the Very Large Array has 27 telescopes giving 351 independent baselines at once, and can give high quality images. In contrast to radio arrays, the largest optical arrays currently have only 6 telescopes, giving poorer image quality from the 15 baselines between the telescopes.

Moreover, we have to take in account that the projected baseline changes with elevation i.e. with source declination and hour angle. For example, as seen in figure 4.7, observing with antennas positioned perpendicularly to the ground, the  $[u, v]$  plane is parallel to the ground. While observing with a specific inclination, also the  $[u, v]$  plane is inclined.

Most aperture synthesis interferometers use the rotation of the Earth to increase the number of different baselines included in an observation. Taking data at different times provides measurements with different telescope separations and angles without the need for buying additional telescopes or moving the telescopes manually, as the rotation of the Earth moves the telescopes to new baselines. This technique, called **Earth Rotation Synthesis**, is a method by which the elements of an interferometer sweep out the aperture of a large telescopes. According to this, the Optical Transfer Function OTF is given by the convolution of  $A(r)$ :  $OTF = A(r) \star A(r)$  (figure 4.8).

# Chapter 5

## Radio Astronomy

### 5.1 Brief history

Karl Jansky made the discovery of the first astronomical radio source in the early 1930s. Using a large directional antenna, Jansky noticed that his analog pen-and-paper recording system kept recording a persistent repeating signal or "hiss" of unknown origin. Since the signal peaked about every 24 hours, Jansky first suspected the source of the interference was the Sun crossing the view of his directional antenna. Continued analysis, however, showed that the source was not following the 24-hour daily cycle of the Sun exactly, but instead repeating on a cycle of 23 hours and 56 minutes. Jansky discussed the puzzling phenomena with his friend, astrophysicist Albert Melvin Skellett, who pointed out that the observed time between the signal peaks was the exact length of a sidereal day; the time it took for "fixed" astronomical objects, such as a star, to pass in front of the antenna every time the Earth rotated. By comparing his observations with optical astronomical maps, Jansky eventually concluded that the radiation source peaked when his antenna was aimed at the densest part of the Milky Way in the constellation of Sagittarius.

Jansky concluded that since the Sun (and therefore other stars) were not large emitters of radio noise, the strange radio interference may be generated by interstellar gas and dust in the galaxy. Current thinking is that these are ions in orbit around a massive Black hole at the center of the galaxy at a point now designated as Sagitarius A\*. The asterisk indicates that the particles at Sagitarius A are ionized.

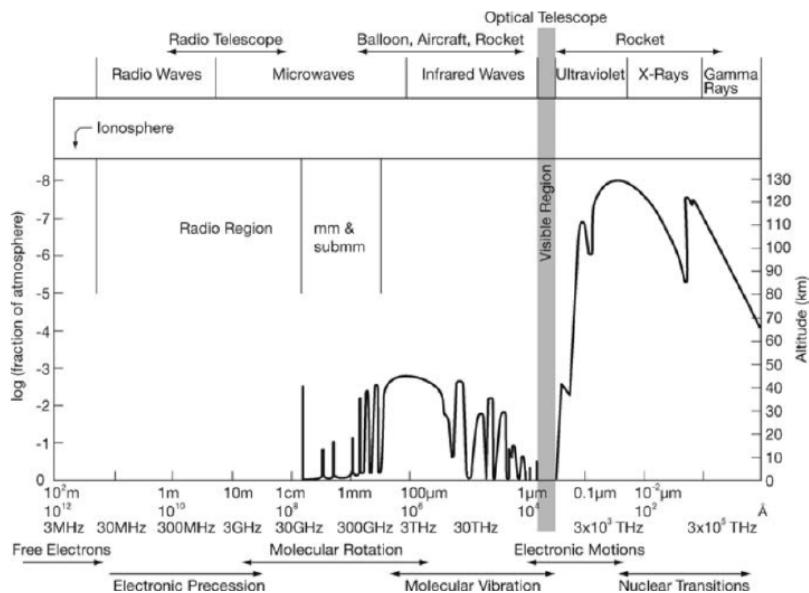
Jansky's pioneering efforts in the field of radio astronomy have been recognized by the naming of the fundamental unit of flux density, the jansky (Jy), after him.

From that moment to now radio astronomy has developed very much and very fast and nowadays there a lot of antennas all around the world analysing the sky in radio range.

### 5.2 Radio window

Earth's atmosphere presents an opaque barrier to much of the electromagnetic spectrum. The atmosphere absorbs most of the wavelengths shorter than ultraviolet, most of the wavelengths between infrared and microwaves, and most of the longest radio waves. That leaves only visible light, some ultraviolet and infrared, and short wave radio to penetrate the atmosphere and bring information about the universe to our Earth-bound eyes and instruments.

The main frequency ranges allowed to pass through the atmosphere are referred to as the radio window and the optical window. The radio window is the range of frequencies from about 1  $THz$  to over 15  $GHz$  (wavelengths of almost 20  $m$  down to about 0.03  $mm$ ). The low-frequency end of the window is limited by signal absorption in the ionosphere, while the upper limit is determined by signal attenuation caused by water vapor and carbon dioxide in the atmosphere.



**Fig. 1.1** The transmission of the earth's atmosphere for electromagnetic radiation. The diagram gives the height in the atmosphere at which the radiation is attenuated by a factor 1/2

Figure 5.1: Atmospheric windows and radiation sources.

The optical window, and thus optical astronomy, can be severely limited by atmospheric conditions such as clouds and air pollution, as well as by interference from artificial light and the literally blinding interference from the Sun's light. Radio astronomy is not hampered by most of these conditions. For one thing, it can proceed even in broad daylight. However, at the higher frequencies in the atmospheric radio window, clouds and rain can cause signal attenuation. For this reason, radio telescopes used for studying sub-millimeter wavelengths are built on the highest mountains, where the atmosphere has had the least chance for attenuation. Conversely, most radio telescopes are built in low places to alleviate problems with human-generated interference.

Atmospheric windows and radiation sources are shown in figure 5.1.

### 5.3 How works a radio telescope

Radio telescopes vary widely, but they all have two basic components: (1) a large radio antenna and (2) a sensitive radiometer, or radio receiver. The sensitivity of a radio telescope — i.e., the ability to measure weak sources of radio emission—depends both on the area and efficiency of the antenna and on the sensitivity of the radio receiver used to amplify and to detect the signals. For broadband continuum emission over a range of wavelengths, the sensitivity also depends on the bandwidth of the receiver. Because cosmic radio sources are extremely weak, radio telescopes are usually very large—up to hundreds of metres across—and use the most sensitive radio receivers available. Moreover, weak cosmic signals can be easily masked by terrestrial radio interference, and great effort is taken to protect radio telescopes from man-made emissions.

The most familiar type of radio telescope is the radio reflector consisting of a parabolic antenna, which operates in the same manner as a television satellite dish to focus the incoming radiation onto a small antenna called the feed, a term that originated with antennas used for radar transmissions. This type of telescope is also known as the dish, or filled-aperture, telescope. In a radio telescope the feed is typically a waveguide horn and transfers the incoming signal to the sensitive radio receiver. Solid-state amplifiers that are cooled to very low temperatures to reduce significantly their internal noise are used to obtain the best possible sensitivity.

In the simplest form of radio telescope, the receiver is placed directly at the focal point of the parabolic reflector, and the detected signal is carried by cable along the feed support structure to a point near the

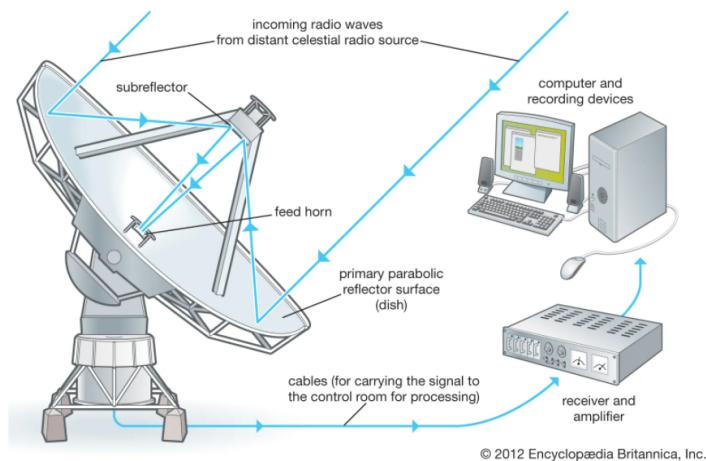


Figure 5.2: How works a radio telescope.

ground where it can be recorded and analyzed. However, it is difficult in this type of system to access the instrumentation for maintenance and repair, and weight restrictions limit the size and number of individual receivers that can be installed on the telescope. More often, a secondary reflector is placed in front of (Cassegrain focus) or behind (Gregorian focus) the focal point of the paraboloid to focus the radiation to a point near the vertex, or centre, of the main reflector. Multiple feeds and receivers may be located at the vertex where there is more room, where weight restrictions are less stringent, and where access for maintenance and repair is more straightforward. Secondary focus systems also have the advantage that both the primary and secondary reflecting surfaces may be carefully shaped so as to improve the gain over that of a simple parabolic antenna.

The performance of a radio telescope is limited by various factors. The accuracy of a reflecting surface may depart from the ideal shape because of manufacturing irregularities. Wind load can exert force on the telescope. Thermal deformations cause differential expansion and contraction. As the antenna is pointed to different parts of the sky, deflections occur due to changes in gravitational forces. Departures from a perfect parabolic surface become important when they are a few percent or more of the wavelength of operation. Since small structures can be built with greater precision than larger ones, radio telescopes designed for operation at millimetre wavelengths are typically only a few tens of metres across, whereas those designed for operation at centimetre wavelengths range up to 300 metres in diameter. For operation at relatively long metre wavelengths where the reflecting surface need not have an accuracy better than a few centimetres, it becomes practical to build very large fixed structures in which the reflecting surface can be made of simple “chicken wire” fencing or even parallel rows of wires.

Radio telescopes are used to measure broad-bandwidth continuum radiation as well as narrow-bandwidth spectroscopic features due to atomic and molecular lines found in the radio spectrum of astronomical objects. In early radio telescopes, spectroscopic observations were made by tuning a receiver across a sufficiently large frequency range to cover the various frequencies of interest. Because the spectrometer had a narrow frequency range, this procedure was extremely time-consuming, and it greatly restricted observations. Modern radio telescopes observe simultaneously at a large number of frequencies by dividing the signals up into as many as several thousand separate frequency channels that can range over a much larger total bandwidth of tens to hundreds of megahertz.

The most straightforward type of radio spectrometer employs a large number of filters, each tuned to a separate frequency and followed by a separate detector that combines the signal from the various filters to produce a multichannel, or multifrequency, receiver. Alternatively, a single broad-bandwidth signal may be converted into digital form and analyzed by the mathematical process of autocorrelation and Fourier transforms. In order to detect faint signals, the receiver output is often averaged over periods of up to several hours to reduce the effect of noise generated by thermal radiation in the receiver.

## 5.4 Radio physics

Let's now see some important definitions and concepts for radio astronomy, in part similar to the ones studied for optical astronomy.

### 5.4.1 Intensity and flux density

As seen previously, the antenna, the main part a radio telescope, receive the weak electromagnetic signal from astronomical sources. The electromagnetic power huddled up in bandwidth  $\delta\nu$  from solid angle  $\delta\Omega$  intercepted by surface  $\delta A$  is:

$$\delta W = I_\nu \delta\Omega \delta A \delta\nu \quad (5.1)$$

which leads to the definition of the **surface brightness**  $I_\nu$  measured in  $Wm^{-2}Hz^{-1}sr^{-1}$ . As consequence it is possible to define also the **flux density**  $S_\nu$  [ $Wm^{-2}Hz^{-1}$ ] as the integration of the brightness over solid angle of source:

$$S_\nu = \int_{\Omega} I_\nu d\Omega \quad (5.2)$$

However a more used and convenient unit for the flux density is the Jansky, in honor of the first man that discovered the first radio source. In particular  $1 Jy = 10^{-26} Wm^{-2} Hz^{-1} = 10^{-23} ergs^{-1} cm^{-2} Hz^{-1}$ .

Note that the flux density  $S_\nu$ , also defined as  $S_\nu = L_\nu / (4\pi d^2)$  is distance dependent while the surface brightness  $I_\nu$  is distance independent indeed  $\Omega \propto 1/d^2$  so  $I_\nu \propto S_\nu/\Omega$ .

### 5.4.2 Surface brightness

On the contrary surface brightness is position dependent indeed, given the angles  $\theta$  and  $\phi$ ,  $I_\nu$  is given by:

$$I_\nu(\theta, \phi) = \frac{2k\nu^2 T(\theta, \phi)}{c^2} \quad (5.3)$$

So, going back to the flux density,  $S_\nu$  can be described as:

$$S_\nu = \int_{\Omega} I_\nu(\theta, \phi) d\Omega = \frac{2k\nu^2}{c^2} \int T(\theta, \phi) d\Omega \quad (5.4)$$

### 5.4.3 Brightness temperature

In general, a radio telescope maps the temperature distribution of the sky so we can define the brightness temperature. The **brightness temperature**  $T_B$  of a source is defined as the temperature of a blackbody with the same surface brightness at a given frequency:

$$I_\nu = \frac{2k\nu^2 T_B}{c^2} \quad (5.5)$$

This implies that the flux density is:

$$S_\nu = \int_{\Omega} I_\nu d\Omega = \frac{2k\nu^2}{c^2} \int T_B d\Omega \quad (5.6)$$

Another useful quantity related to the brightness is the radiation energy density  $u_\nu$  in units of  $\text{ergcm}^{-3}$ . From dimensional analysis  $u_\nu$  is intensity divided by speed. Since radiation propagates with the velocity of light, we have for the **spectral energy density per solid angle**:

$$u_\nu(\Omega) = \frac{1}{c} I_\nu \quad (5.7)$$

If integrated over the whole sphere,  $4\pi$  steradian, equation 5.7 results in the **total spectrum energy density**:

$$u_\nu = \int_{4\pi} u_\nu(\Omega) d\Omega = \frac{1}{c} \int_{4\pi} I_\nu d\Omega \quad (5.8)$$

#### 5.4.4 Radiative transfer

For a change in  $I_\nu$  along the line of sight, a loss term  $dI_{\nu-}$  and a gain term  $dI_{\nu+}$  are introduced, and we adopt the form  $dI_{\nu-} = -k_\nu I_\nu ds$  and  $dI_{\nu+} = \epsilon_\nu ds$  so that the change of intensity in a slab of material of the thickness  $ds$  will be:

$$[I_\nu(s + ds) - I_\nu(s)]d\sigma d\Omega d\nu = [-k_\nu I_\nu + \epsilon_\nu]d\sigma d\Omega d\nu ds \quad (5.9)$$

resulting in the **equation of transfer**:

$$\frac{dI_\nu}{ds} = -k_\nu I_\nu + \epsilon_\nu \quad (5.10)$$

From general experience, the linear absorption coefficient  $k_\nu$  is independent of the intensity  $I_\nu$  leading to the adoption of the above form for  $dI_{\nu-}$ ; similar arguments hold for the emissivity  $\epsilon_\nu$ .

There may be situations for which  $\epsilon_\nu$  depends strongly on  $I_\nu$ , such as an environment in which radiation is strongly scattered. However, there are many other important situations where  $\epsilon_\nu$  is independent of  $I_\nu$ .

There are several limiting cases for which the solution of the differential equation 5.10 is especially simple.

##### 1. Emission only - $k_\nu = 0$

$$\begin{aligned} \frac{dI_\nu}{ds} &= \epsilon_\nu \\ I_\nu(s) &= I_\nu(s_0) + \int_{s_0}^s \epsilon_\nu(s) ds \end{aligned} \quad (5.11)$$

##### 2. Absorption only - $\epsilon_\nu = 0$

$$\begin{aligned} \frac{dI_\nu}{ds} &= -k_\nu I_\nu \\ I_\nu(s) &= I_\nu(s_0) \exp - \int_{s_0}^s k_\nu(s) ds \end{aligned} \quad (5.12)$$

##### 3. Thermodynamic equilibrium (TE) -

If there is complete equilibrium of the radiation with its surroundings, the brightness distribution is described by the Planck function, which depends only on the thermodynamic temperature  $T$ , of the surroundings:

$$\begin{aligned} \frac{dI_\nu}{ds} &= 0 \\ I_\nu(s) &= B_\nu(T) = \epsilon_\nu/k_\nu \end{aligned} \quad (5.13)$$

where  $B_\nu$  is given by the following relation (we talk about this in the next section):

$$B_\nu(T) = \frac{2h\nu^3}{c^2} \frac{1}{e^{h\nu/kT} - 1} \quad (5.14)$$

4. **Local thermodynamic equilibrium (LTE)** - Full thermodynamic equilibrium will be realized only in very special circumstances such as in a black enclosure or say, in stellar interiors. This is described by **Kirchhoff's law**:

$$\frac{\epsilon_\nu}{k_\nu} = B_\nu(T) \quad (5.15)$$

applicable independent of the material, as is the case with complete thermodynamic equilibrium. In general however,  $I_\nu$  will differ from  $B_\nu(T)$ . If we define the **optical depth**  $d\tau_\nu$  by  $d\tau_\nu = -k_\nu ds$  or:

$$\tau_\nu(s) = \int_{s_0}^s k_\nu(s) ds \quad (5.16)$$

then the equation of transfer 5.10 can be written as:

$$-\frac{1}{k_\nu} \frac{dI_\nu}{ds} = \frac{dI_\nu}{d\tau_\nu} = I_\nu - B_\nu(T) \quad (5.17)$$

The solution of 5.17 is obtained by first multiplying 5.17 by  $\exp -\tau_\nu$  and then integrating  $\tau_\nu$  by parts. The final result is:

$$I_\nu(s) = I_\nu(0)e^{-\tau_\nu(s)} + \int_0^{\tau_\nu(s)} B_\nu(T(\tau))e^{-\tau} d\tau \quad (5.18)$$

If the medium is isothermal, that is, if  $T(\tau) = T(s) = T = \text{const}$ , the integral in 5.18 can be computed explicitly resulting in:

$$I_\nu(s) = I_\nu(0)e^{-\tau_\nu(s)} + B_\nu(T)(1 - e^{-\tau_\nu(s)}) \quad (5.19)$$

For a large optical depth, that is for  $\tau_\nu(0)$  tends to  $\infty$ , equation 5.19 in LTE approaches the limit  $I_\nu = B_\nu(T)$ . The observed brightness  $I_\nu$  for the optically thick case is equal to the Planck black-body brightness distribution independent of the material. If the intensity is to be compared with the result obtained in the absence of an intervening medium,  $I_\nu(0)$ , we have:

$$\Delta I_\nu(s) = I_\nu(s) - I_\nu(0) = (B_\nu(T) - I_\nu(0))(1 - e^{-\tau}) \quad (5.20)$$

#### 5.4.5 Black Body Radiation

The spectral distribution of the radiation of a black body in thermodynamic equilibrium is given by the **Planck law**:

$$B_\nu(T) = \frac{2h\nu^3}{c^2} \frac{1}{e^{h\nu/kT} - 1} \quad (5.21)$$

It gives the power per unit frequency interval (see figure 5.3).

Converting this to the wavelength scale, we obtain  $B_\lambda(T)$ . Because  $B_\nu(T)d\nu = -B_\lambda(T)d\lambda$  and  $d\nu = (-c/\lambda^2)d\lambda$  this is:

$$B_\lambda(T) = \frac{2hc^2}{\lambda^5} \frac{1}{e^{hc/kT\lambda} - 1} \quad (5.22)$$

Integrating 5.14 over  $\nu$  or 5.22 over  $\lambda$ , the total brightness of a black body is obtained:

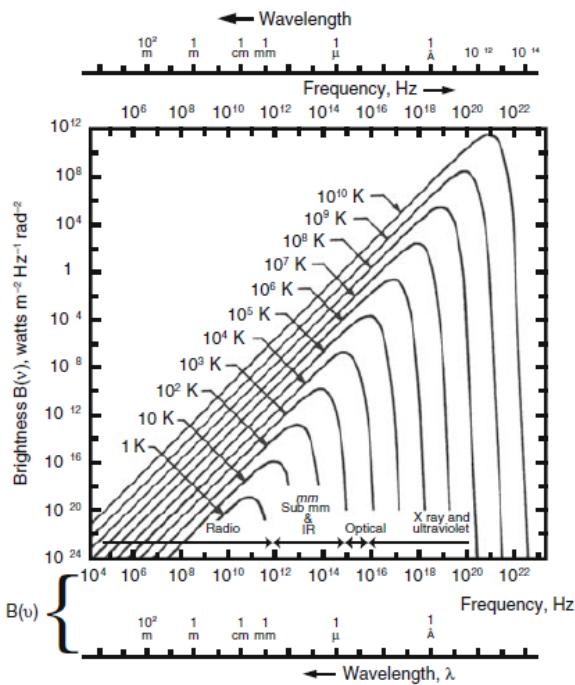


Fig. 1.6 Planck spectra for black bodies of different temperatures

Figure 5.3: Planck spectra for black bodies.

$$B(T) = \frac{2h}{c^2} \int_0^\infty \frac{\nu^3}{e^{h\nu/kT} - 1} \quad (5.23)$$

This integral has also equal to  $B(T) = \sigma T^4$  where  $\sigma = 1.8047 \times 10^{-5} \text{ erg cm}^{-2} \text{ s}^{-1} \text{ K}^{-4}$ . The maximum in terms of frequency is given by:

$$\frac{\nu_{max}}{\text{GHz}} = 58.789 \frac{T}{K} \quad (5.24)$$

while in terms of wavelength the maximum is:

$$\frac{\lambda_{max}}{\text{cm}} \frac{T}{K} = 0.28978 \quad (5.25)$$

These two last equations are both known as **Wien's displacement law**. If the ratio  $h\nu/kT$  is far from the maximum, equation 5.14 can be approximated by simpler expressions.

- $h\nu \ll kT$ : **Rayleigh-Jeans Law** - An expansion of the exponential  $e^{h\nu/kT} \sim 1 + \frac{h\nu}{kT} + \dots$  results in:

$$B_{RJ}(\nu, T) = \frac{2\nu^2}{c^2} kT \quad (5.26)$$

This is the classical limit of the Planck law since it does not contain Planck's constant. In the millimeter and submillimeter range, one frequently defines a radiation temperature,  $J(T)$  as:

$$J(T) = \frac{c^2}{2k\nu^2} I = \frac{h\nu}{k} \frac{1}{e^{h\nu/kT} - 1} \quad (5.27)$$

Inserting numerical values for  $k$  and  $h$ , we see that the Rayleigh-Jeans relation holds for frequencies:

$$\frac{\nu}{\text{GHz}} \ll 20.84 \frac{T}{K} \quad (5.28)$$

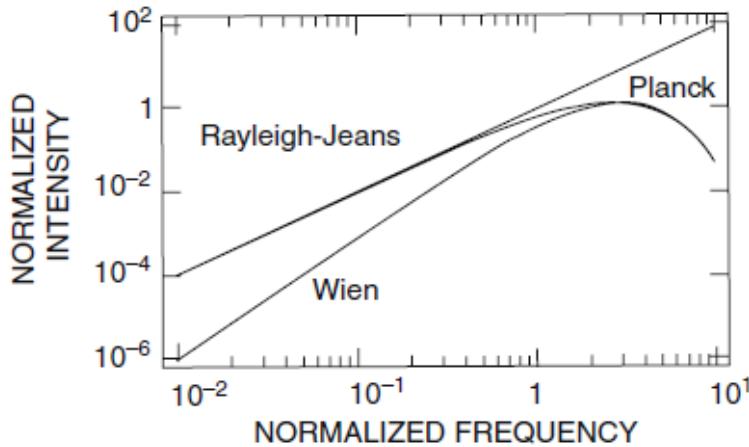


Figure 5.4: Normalized Planck curve and the Rayleigh-Jeans and Wien approximation.

It can thus be used for all thermal radio sources except perhaps for low temperatures in the millimeter or sub-mm range.

- $h\nu \gg kT$ : **Wien's Law** - In this case  $e^{h\nu/kT} \gg 1$ , so that:

$$B_W(\nu, T) = \frac{2h\nu^3}{c^2} e^{h\nu/kT} \quad (5.29)$$

While this limit is quite useful for stellar measurements in the visual and ultraviolet range, it plays no role in radio astronomy.

One of the important features of the Rayleigh-Jeans law is the implication that the brightness and the thermodynamic temperature of the black body that emits this radiation are strictly proportional (5.26). This feature is so useful that it has become the custom in radio astronomy to measure the brightness of an extended source by its brightness temperature  $T_b$ . This is the temperature which would result in the given brightness if inserted into the Rayleigh-Jeans law:

$$T_b = \frac{c^2}{2k} \frac{1}{\nu^2} I_\nu = \frac{\lambda^2}{2k} I_\nu \quad (5.30)$$

Combining the flux density  $S_\nu$  and the last relation, we have:

$$S_\nu = \frac{2k\nu^2}{c^2} T_b \Delta\Omega \quad (5.31)$$

The two previous conditions are represented in figure 5.4.

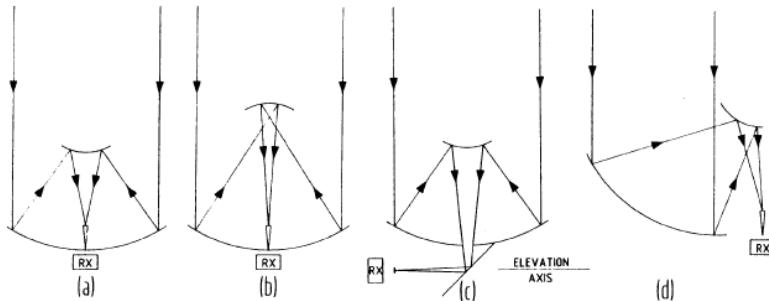
## 5.5 What does a radio telescope detect?

Consider a radio telescope of effective area  $A_e$ . This telescope receives power  $P_{rec}$  per unit frequency from an unpolarized source (it is only sensitive to one mode of polarization):

$$P_{rec} = \frac{1}{2} I_\nu A_e \delta\Omega \quad (5.32)$$

However, it is sensitive to radiation from more than one direction with relative sensitivity given by the normalized antenna pattern  $P_N(\theta, \phi)$ . Then the power received is given by:

$$P_{rec} = \frac{1}{2} A_e \int_{4\pi} I_\nu(\theta, \phi) P_N(\theta, \phi) d\Omega \quad (5.33)$$



**Fig. 7.6** The geometry of (a) Cassegrain, (b) Gregory, (c) Nasmyth and (d) offset Cassegrain systems

Figure 5.5: Different types of antenna configurations.

### 5.5.1 Types of antennas

As said before, the antenna collects the electric field over the aperture at the focus. In general the feed horn at the focus adds the fields together and guides the signal to the front end however there are different types of antenna (see figure 5.5).

A summary of the main features to take in mind is provided below.

1. Fully steerable paraboloids of revolution have become the standard antennas in the centimeter, millimeter and sub-mm wavelength regions. In the mm and submm ranges, these were housed in shelters, but recently designs have allowed high performance paraboloids to operate in the open air.
2. All symmetric paraboloids have designs that make use of the homology principle. That is, with changing elevation, the surfaces deform freely from one symmetric parabolic shape to another.
3. All modern designs use Altitude-Azimuth mounts. The control is carried out with digital computer systems.
4. The Cassegrain and Nasmyth foci are preferred for millimeter and sub-mm telescopes, since these help to minimize the reception of noise from the ground, have additional reflecting surfaces that allow optimization of telescope gain, and provide larger amounts of space for receivers.

### Components of a heterodyne system

Radio telescopes use a Heterodyne system. A heterodyne is a signal frequency that is created by combining or mixing two other frequencies using a signal processing technique called heterodyning. Heterodyning is used to shift one frequency range into another, new frequency range. The two input frequencies are combined in a nonlinear signal-processing device such as a vacuum tube, transistor, or diode, usually called a mixer.

In the most common application, two signals at frequencies  $f_1$  and  $f_2$  are mixed, creating two new signals, one at the sum of the two frequencies  $f_1 + f_2$ , and the other at the difference between the two frequencies  $f_1 - f_2$ . The new signal frequencies are called heterodynes. Typically, only one of the heterodynes is required and the other signal is filtered out of the output of the mixer.

In the case of the radio telescopes, these type of systems are composed by:

- **amplifier** - it is a device that amplifies a very weak radio frequency (RF) signal, stable and with low noise;
- **mixer** - it produces a stable lower, intermediate frequency (IF) signal by mixing the RF signal with a stable local oscillator (LO) signal which is tunable;
- **filter** - it selects a narrow signal band out of the IF;

- **backend** - it contains a total power detector or more typically today, a correlator.

## 5.6 Descriptive antenna parameters

If one wants an accurate but rather simple description of antenna properties, one must use the concepts presented in the following sections, which allow one to characterize the antenna properties based on astronomical measurements.

### 5.6.1 The power pattern $P(\theta, \phi)$ and the gain

Often, the **normalized power pattern**, not the power pattern, is measured:

$$P_n(\theta, \phi) = \frac{1}{P_{max}} P(\theta, \phi) \quad (5.34)$$

Consider the power pattern of the antenna used as a transmitter. If the total spectral power,  $P_\nu$  in  $[W Hz^{-1}]$  is fed into a lossless isotropic antenna, this would transmit  $P$  power units per solid angle per Hertz. Then the total radiated power at frequency  $\nu$  is  $4\pi P_\nu$ . In a realistic, but still lossless antenna, a power  $P(\theta, \phi)$  per unit solid angle is radiated in the direction  $(\theta, \phi)$ . If we define the **directive gain**  $G(\theta, \phi)$  as the:

$$P(\theta, \phi) = G(\theta, \phi)P \quad (5.35)$$

or:

$$G(\theta, \phi) = \frac{4\pi P(\theta, \phi)}{\int \int P(\theta, \phi) d\Omega} \quad (5.36)$$

Thus the gain or directivity is also a normalized power pattern similar to equation 5.34, but with the difference that the normalizing factor is  $P(\theta, \phi)d\Omega/4\pi$ . This is the gain relative to a lossless isotropic source. Since such an isotropic source cannot be realized in practice, a measurable quantity is the gain relative to some standard antenna such as a half-wave dipole whose directivity is known from theoretical considerations.

### 5.6.2 The main beam solid angle

The **beam solid angle**  $\Omega_A$  of an antenna is given by:

$$\Omega_A = \int \int_{4\pi} P_n(\theta, \phi) d\Omega \quad (5.37)$$

this is measured in steradians (*sr*). The integration is extended over the full sphere  $4\pi$ , such that  $\Omega_A$  is the solid angle of an ideal antenna having  $P_n = 1$  for all of  $\Omega_A$  and  $P_n = 0$  everywhere else. Such an antenna does not exist; for most antennas the (normalized) power pattern has considerably larger values for a certain range of both  $\theta$  and  $\phi$  than for the remainder of the sphere. This range is called the **main beam** or **main lobe** of the antenna; the remainder are the **side lobes** or back lobes (see figure 5.6). For actual situations, the properties are well defined up to the shortest operating wavelengths. At the shortest wavelength, there is indeed a main beam, but much of the power enters through sidelobes. In addition, the main beam efficiency may vary significantly with elevation. Thus, the ability to accurately calibrate the radio telescope at the shortest wavelengths may be challenging.

In analogy to 5.37 we define the main beam solid angle  $\Omega_{MB}$  by:

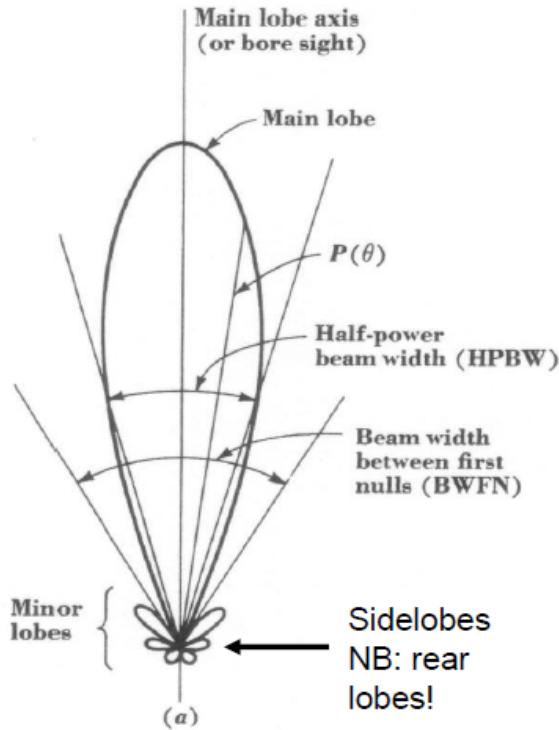


Figure 5.6: A polar power pattern showing the main beam, and near and far side lobes.

$$\Omega_{MB} = \int \int_{mainlobe} P_n(\theta, \phi) d\Omega \quad (5.38)$$

The quality of an antenna as a direction measuring device depends on how well the power pattern is concentrated in the main beam. If a large fraction of the received power comes from the side lobes it would be rather difficult to determine the location of the radiation source, the so-called *pointing*.

It is appropriate to define a **main beam efficiency** or (usually) beam efficiency,  $\eta_B$ , by:

$$\eta_B = \frac{\Omega_{MB}}{\Omega_A} \quad (5.39)$$

The main beam efficiency is not related to the angular size of the main beam. A small antenna with a wide main beam can have a high beam efficiency:  $\eta_B$  is an indication of the fraction of the power is concentrated in the main beam. Substituting equation 5.37 into 5.36 it is easy to see that the **maximum directive gain**  $G_{max}$  or **directivity**  $D$  can be expressed as:

$$D = G_{max} = \frac{4\pi}{\Omega_A} \quad (5.40)$$

The angular extent of the main beam is usually described by the *half power beam width* (HPBW), which is the angle between points of the main beam where the normalized power pattern falls to 1/2 of the maximum (see figure 5.7). This is also referred to as the *full width to half power* (FWHP). Less frequently used definitions are the *beam width between first nulls* (BWFN) or the *equivalent width of the main beam* (EWMB). The latter quantity is defined by:

$$EWMB = \sqrt{\frac{12}{\pi} \Omega_{MB}} \quad (5.41)$$

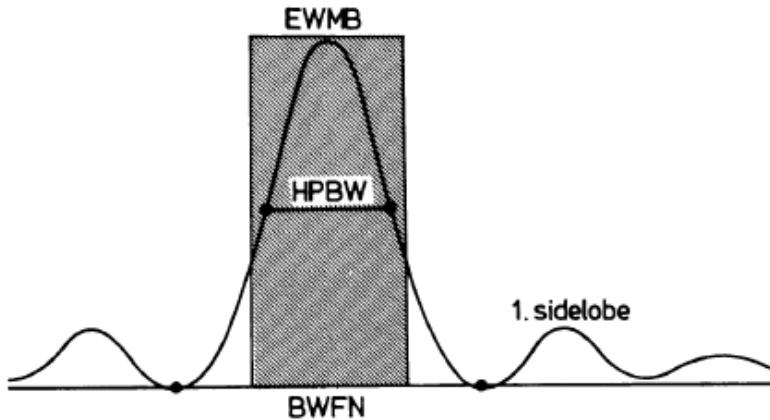


Figure 5.7: A sketch of the telescope beamwidth, together with commonly used measurements of beam size for a one-dimensional power pattern. EWMB is the equivalent width of the (full) half power beam width. The HPBW is sometimes referred to as FWHP, or full width to half power. BWFN denotes the beam width between first nulls. This is indicated by the two dots.

For elliptically shaped main beams, values for widths in orthogonal directions are needed. The beam width is related to the geometric size of the antenna and the wavelength used; the exact beam size depends on grating functions and illumination.

### 5.6.3 The effective aperture and the directivity

Let a plane wave with the power density  $|S|$  be intercepted by an antenna. A certain amount of power is then extracted by the antenna from this wave; let this amount of power be  $P_e$ . We will then call the fraction:

$$A_e = P_e / |S| \quad (5.42)$$

the **effective aperture** of the antenna.  $A_e$  is a quantity very much like a cross-section in particle physics,  $A_e$  has the dimension of  $m^2$ . Comparing this to the **geometric aperture**  $A_g$  we can define an aperture efficiency  $\eta_A$  by:

$$A_e = \eta_A A_g \quad (5.43)$$

For some antennas, such as the Hertz dipole there is no clearly defined geometric aperture; in such cases there is no simple expression for the aperture efficiency  $\eta_A$ . For a calculation of the effective aperture, the peak value of  $A_e$  is used; this is the direction of the telescope axis. **Directivity** is related to  $A_e$  by:

$$D = G_{max} = \frac{4\pi A_e}{\lambda^2} \quad (5.44)$$

which is equivalent to  $A_e \Omega_A = \lambda^2$ .

Let antenna, receiver and a radiating surface  $C$  all be enclosed by a black body at the temperature  $T$ . Let us assume thermodynamic equilibrium for the whole system. Then the antenna will radiate power into the black enclosure, and this power will be absorbed there. The black body will also radiate, and part of this radiation will be received by the antenna. Let the radiation surface  $C$  subtend the solid angle  $\Omega_A$  as seen from the antenna (figure 5.8), whose directivity is  $D$ , effective aperture  $A_e$  and receiver bandwidth  $\Delta\nu$ . According to the Rayleigh-Jeans relation, the surface  $C$  radiates with the intensity:

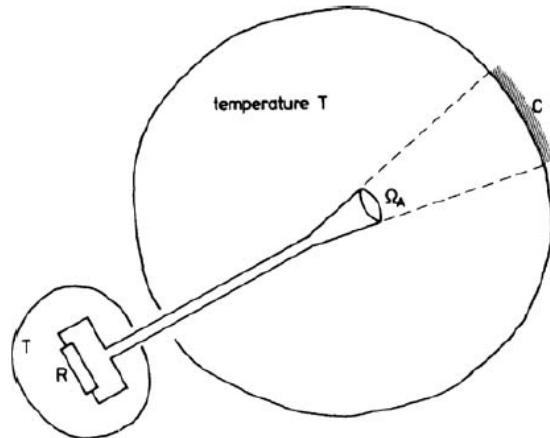


Figure 5.8: A sketch to illustrate the relation between effective aperture and directivity.

$$I_\nu = \frac{2kT}{\lambda^2} \Delta\nu \quad (5.45)$$

in units of  $Wm^{-2}Hz^{-1}$  per unit solid angle. Then the antenna collects a total power of:

$$W = A_e \frac{kT}{\lambda^2} \Delta\nu \Omega_A \quad (5.46)$$

since only one polarization component is recorded.

If the whole system is in thermal equilibrium, the principle of detailed balance holds. Then the antenna must reradiate the same amount of power that it receives. If the antenna terminals are connected by a matched resistor  $R$ , then the transmitted power according to the Nyquist theorem ( $P_\nu = kT$ ) is:

$$L\Delta\nu = kT\Delta\nu \quad (5.47)$$

The surface  $C$  intercepts the fraction  $L'\Delta\nu = kT\Delta\nu D \frac{\Omega_A}{4\pi}$ . As stated, this last relation and relation 5.46 are equal if thermodynamic equilibrium prevails; thus:

$$A_e \frac{kT}{\lambda^2} \Delta\nu \Omega_A = kT\Delta\nu D \frac{\Omega_A}{4\pi} \quad (5.48)$$

so that:

$$D = \frac{4\pi A_e}{\lambda^2} \quad (5.49)$$

Although this relation has been derived under the assumption of thermodynamic equilibrium, this relates quantities which do not involve thermodynamics, so will always be valid.

#### 5.6.4 Antenna temperature

Consider a receiving antenna with a normalized power pattern  $P_n(\theta, \phi)$  that is pointed at a brightness distribution  $B_\nu(\theta, \phi)$  in the sky. Then at the output terminals of the antenna, the total power per unit bandwidth,  $P_\nu$  is:

$$P_\nu = \frac{1}{2} A_e \int \int B_\nu(\theta, \phi) P_n(\theta, \phi) d\Omega \quad (5.50)$$

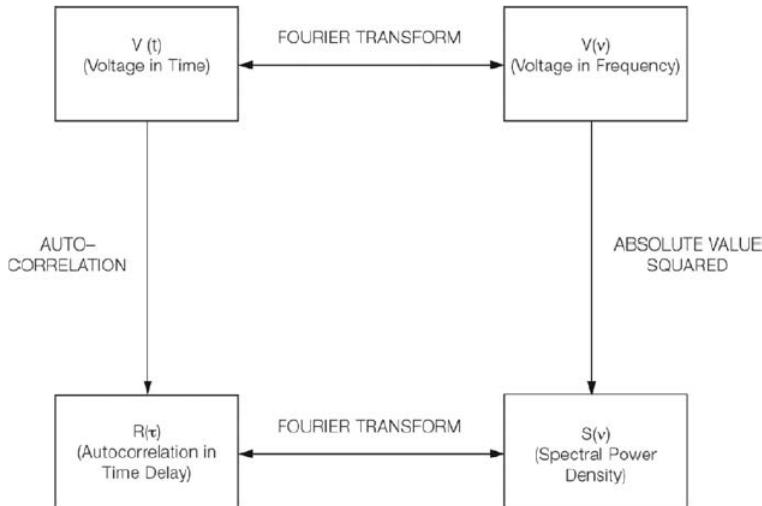


Figure 5.9: A sketch of the relation between the voltage input as a function of time,  $V(t)$ , and frequency,  $V(\nu)$ , with the autocorrelation function, ACF,  $R(\tau)$ , and corresponding power spectral density, PSD,  $S(\nu)$ . The two-headed arrows represent reversible processes

By definition, we are in the Rayleigh-Jeans limit, and can therefore exchange the brightness distribution by an equivalent distribution of brightness temperature. Using the Nyquist theorem we can introduce an equivalent antenna temperature  $T_A$  by:

$$P_\nu = kT_A \quad (5.51)$$

This definition of **antenna temperature** relates the output of the antenna to the power from a matched resistor. When these two power levels are equal, then the antenna temperature is given by the temperature of the resistor. Instead of the effective aperture  $A_e$  we can introduce the beam solid angle  $\Omega_A$ . Then equation 5.50 becomes:

$$T_A(\theta_0, \phi_0) = \frac{\int T_b(\theta, \phi) P_n(\theta - \theta_0, \phi - \phi_0) \sin \theta d\theta d\phi}{\int P_n(\theta, \phi) d\Omega} \quad (5.52)$$

which is the *convolution* of the brightness temperature with the beam pattern of the telescope. The brightness temperature  $T_b(\theta, \phi)$  corresponds to the thermodynamic temperature of the radiating material only for thermal radiation in the Rayleigh-Jeans limit from an optically thick source; in all other cases  $T_b$  is only a convenient quantity that in general depends on the frequency.

## 5.7 Summary

- The FWHP (Full Width to Half Power) defines the telescope resolution.
- The voltage response pattern is the FT of the aperture distribution.
- The power response pattern,  $P(\theta) \propto V(\theta)^2$ , is the FT of the auto-correlation function of the aperture.
- For a uniform circle,  $P(\theta)$  is the Airy pattern.

A schematic summary is shown in figure 5.9.

## 5.8 Sensitivity and Noise

The output power detected is the sum of the power of antenna and the power introduced by the system itself:

$$P_{out} = P_A + P_{sys} \quad (5.53)$$

Since the power is proportional to the temperature, the previous relation can be expressed also as:

$$T_{out} = T_A + T_{sys} \quad (5.54)$$

The system temperature  $T_{sys}$  represents the noise added by the system and it is given by many different components:

- $T_{bg}$  which corresponds to the microwave and galactic background (with maximum temperature at 3K, except below 1 GHz);
- $T_{sky}$  that is associated to the atmospheric emission, which increases with frequency and it is dominant in mm (mainly molecules of  $H_2O$  and  $O_2$ );
- $T_{spill}$  that comes from thermal ground radiation. Coming from the ground, this contribution affects mainly the sidelobes of the beam;
- $T_{loss}$  which is the losses in the feed and signal transmission system (it depends on the design of the telescope);
- $T_{cal}$  is a noise introduced by the injected calibrator signal (usually small);
- $T_{rx}$  is given by receiver system (often dominates at cm wavelength).

The overall system temperature is, then, given:

$$T_{sys} = T_{bg} + T_{sky} + T_{spill} + T_{loss} + T_{cal} + T_{rx} \quad (5.55)$$

Note that  $T_{bg}$ ,  $T_{sky}$  and  $T_{spill}$  vary with sky position and  $T_{sky}$  is also time variable. Moreover, in the millimeter and sub-millimeter regime, especially at low elevation,  $T_{sky}$  is a relevant contribution. In general  $T_{rx} < T_{sky}$ . For example, to minimise  $T_{rx}$ , ALMA observes in many different bands.

So a good question can be: how can you detect the antenna temperature  $T_A$  which means the signal if there is the system temperature  $T_{sys}$  so the noise? The answer is quite simple: the signal is indeed correlated from one sample to the next but the noise is not. In particular, for bandwidth  $\Delta\nu$ , samples taken less than  $\Delta\tau = \frac{1}{\Delta\nu}$  are not independent while time  $\tau$  contains  $N = \tau/\Delta t = \tau\Delta\nu$  independent samples. In general for Gaussian noise, total error for  $N$  samples  $1/\sqrt{N}$  that of single sample is therefore:

$$\frac{\Delta T_A}{T_{sys}} = \frac{1}{\sqrt{\tau\Delta\nu}} \quad (5.56)$$

As consequence, it is possible to define the Signal to Noise Ratio as following:

$$SNR = \frac{T_A}{\Delta T_A} = \frac{T_A}{T_{sys}} \sqrt{\tau\Delta\nu} \quad (5.57)$$

This last relation is also called **radiometer equation**. This is an important relation that allows us to determine the signal-to-noise ratio from a source generating an antenna temperature  $T_A$  using a telescope with a system temperature  $T_{sys}$ , a bandwidth  $\Delta\nu$  and an integration time  $\tau$ . PAY ATTENTION: in radio range exposures are usually of few milliseconds, as consequence the amount of data obtained is very huge, as the number of samples  $N$ , that is very big.

### 5.8.1 Receiver calibration

In the calibration process, a noise power scale must be established at the receiver input. While the detailed procedures depend on the actual instruments in use, the basic principles are following. In radio astronomy the noise power of coherent receivers (those which preserve the phase of the input) is usually measured in terms of the noise temperature. To calibrate a receiver, one relates the noise temperature increment  $\Delta T$  at the receiver input to a given measured receiver output increment  $\Delta z$  (this applies to heterodyne receivers). In principle, the receiver noise temperature,  $T_R$ , could be computed from the output signal  $z$  provided the detector characteristics are known. In practice the receiver is calibrated by connecting two or more known power sources to the input. Usually matched resistive loads at the known (thermodynamic) temperatures  $T_L$  and  $T_H$  are used. The receiver outputs are then:

$$\begin{aligned} z_L &= (T_L + T_R)G \\ z_H &= (T_H + T_R)G \end{aligned} \tag{5.58}$$

from which:

$$T_{rx} = \frac{T_H - T_L y}{y - 1} \tag{5.59}$$

where:

$$y = z_H/z_L \tag{5.60}$$

## 5.9 Noise equations

In previous section we have seen all contributions to noise but this can be expressed also in terms of specific equation according to the type of system we are using. For example, in the case of a *single dish antenna* the noise equation:

$$\sigma_S = \frac{2k}{\eta_A A} \frac{T_{sys}}{\sqrt{\Delta\nu\Delta t}} \tag{5.61}$$

in which  $A$  is the total collecting area of a single antenna,  $\eta_A$  is the aperture efficiency,  $T_{sys}$  is the system temperature (as seen before),  $\Delta\nu$  is the bandwidth and  $\Delta t$  is the integration time on the source. In the case of a single antenna there are many limitations. The most important is that the sensitivity is proportional to  $\sim \frac{1}{D^2}$  where  $D$  is the diameter of the telescope so to have higher sensitivity is necessary to increase the collating area keeping a high surface quality.

Suppose now to consider a *single baseline* so a two-element interferometer. In this case the noise equation is:

$$\sigma_S = \frac{2k}{\eta_A A} \frac{\langle T_{sys} \rangle}{\sqrt{\sqrt{2}\Delta\nu\Delta t}} \tag{5.62}$$

in which the term  $\sqrt{2}$  improves the total power received from the two-element system than a single antenna with the same area.

If we instead consider an *interferometric system* with  $N$  samples observing a **point source**, the noise equation is:

$$\sigma_S = \frac{2k}{\eta_{AA}} \frac{\langle T_{sys} \rangle}{\sqrt{\eta_C \eta_J \eta_P \sqrt{N(N-1)\Delta\nu\Delta t}}} \frac{1}{\sqrt{N_P}} \quad (5.63)$$

in which  $N(N - 1)$  is the number of baselines,  $\eta_C$  is the correlator efficiency,  $\eta_J$  is the instrumental jitter so it associated to local oscillations,  $\eta_P$  is the atmospheric de-correlation (so the seeing) and  $N_P$  indicated the possible linear polarizations (it can be 1 or 2). In particular, the first term,  $\frac{2k}{\eta_{AA}}$ , indicates the single dish efficiency measured in  $Jy/K$  so it is due to the antenna.

If we instead consider an interferometric system with  $N$  samples but observing an **extended source**, then the noise equation is:

$$\sigma_{T_b} = \left( \frac{\theta_p}{\theta_s} \right)^2 \frac{T_{sys}}{\eta \sqrt{N(N-1)\Delta\nu\Delta t}} \quad (5.64)$$

where  $\eta = 0, 5$ ,  $\theta_s$  is the source angular extension and  $\theta_p$  is the angular extension of power beam. Of course, for extended sources, it is important to have main beam lobe bigger than the extension of the source it-self in order to have the highest possible sensitivity. This is due to the fact that the estimation of sensitivity is based on an assumption that source power is diluted over the beam. On the contrary there would be problems of missing flux.

All noise equations have dimensions of  $Jy$  (Jansky) and usually have order of few  $mJy$ .

# Chapter 6

## The ideal interferometer

Figure 6.1 represents a schematic image of how works a two-elements interferometer. Radiation coming from the source is received by the two telescopes. As said many times, to form the fringe pattern is necessary that the two beams have the same path (a certain degree of coherence) but practically there is a time delay due to the difference in path. This can be compensated by the delay line. Once the fringes are formed, from them we can derive the amplitude and the phase of the visibility function. In particular this is done by the correlator which *natural* in case of optical telescopes and *artificial* for radio astronomy.

This is in synthesis what happen for simple and ideal interferometry. In the following we will study assumptions, simplifications and mathematical approach for this type of observations.

Suppose now to observe a radio source. An astrophysical source at location  $R$  causes a time-variable electric field  $E(R, t)$ . Then the electromagnetic wave propagates to us at point  $r$ . It is possible to express the field as a Fourier series in which the only time-varying functions are complex exponentials, however we are interested only in the (complex) coefficients of this series,  $E_\nu(R)$ :

$$E(R, t) = \int E_\nu(R) e^{2\pi i \nu t} d\nu \quad (6.1)$$

Assuming this, many simplifications can be done.

**Simplification 1** It is possible to assume monochromatic radiation for which:

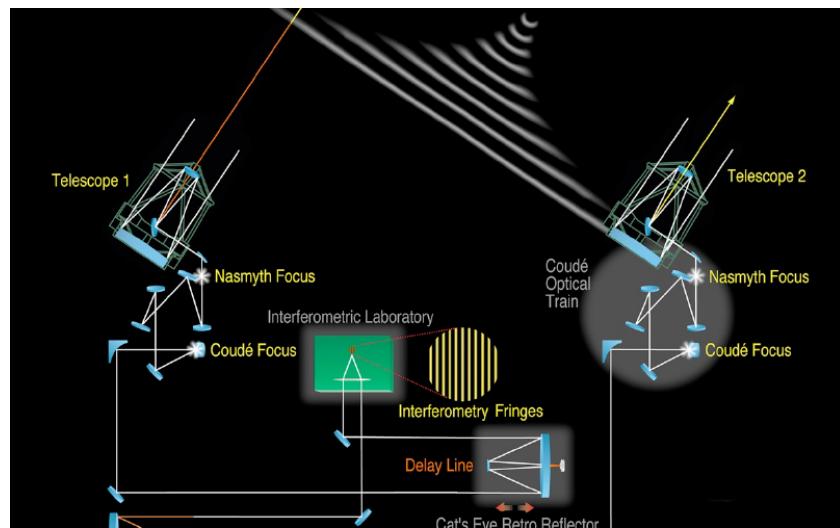


Figure 6.1: An ideal interferometer.

$$E_\nu(R) = \int \int \int P_\nu(R, r) E_\nu(r) dx dy dz \quad (6.2)$$

where  $P_\nu(r, R)$  is the propagator.

**Simplification 2** It is known that the electric field is a vector field which is always polarized but we can ignore this, assuming a scalar field and ignore the polarization

**Simplification 3** All sources are very far away therefore, it is possible to assume they have all the same fixed distance  $|R|$  (with no depth information as consequence).

**Simplification 4** The space between us and the source is empty. In this case, the propagator is quite simple and follows the Huygens' Principle so:

$$E_\nu(R) = \int E_\nu(r) \frac{\exp 2\pi i |R - r|/c}{|R - r|} dS \quad (6.3)$$

in which  $dS$  is the element of area at distance  $|R|$ . Practically, what we can measure is the correlation of the field at two different observing locations. This is:

$$C_\nu(r_1, r_2) = \langle E_\nu(r_1) E_\nu^*(r_2) \rangle \quad (6.4)$$

where  $\langle \rangle$  denotes an expectation value and  $*$  means complex conjugation.

**Simplification 5** Assume that the radiation from astronomical objects is not spatially coherent which means characterized by random noise. In this case  $\langle E_\nu(r_1) E_\nu^*(r_2) \rangle = 0$  unless the source has positions  $R_1 = R_2$  from the two antennas (if the source is resolved there is no spatial coherence). Now write  $\mathbf{s} = R/|R|$  and the observed intensity as  $I_\nu(\mathbf{s}) = |R|^2 \langle |E_\nu(\mathbf{s})|^2 \rangle$ . Using the approximation of large distance to the source again:

$$C_\nu(r_1, r_2) = \int I_\nu(\mathbf{s}) e^{\frac{-2\pi i \nu \mathbf{s} \cdot (r_1 - r_2)}{c}} d\Omega \quad (6.5)$$

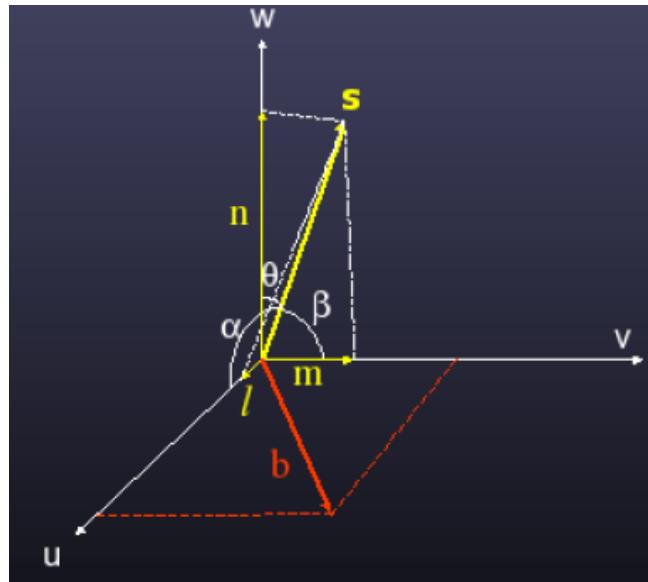
This relation defines the spatial coherence function  $C_\nu(r_1, r_2)$  that depends only on separation  $r_1 - r_2$ , so we can keep one point fixed and move the other around. In particular it is a complex function, with real and imaginary parts, or an amplitude and phase. According to this we can give again the definition of an interferometer: a device that measures the spatial coherence function.

To be more specific, the unit direction vector  $\mathbf{s}$  is defined by its projections on the  $(u, v, w)$  axes. These components are also called **direction cosines** and, according to figure 6.2, are defined as:

$$\begin{aligned} l &= \cos \alpha \\ m &= \cos \beta \\ n &= \cos \theta = \sqrt{1 - l^2 - m^2} \end{aligned} \quad (6.6)$$

The baseline vector  $\mathbf{b}$  is specified by its coordinates  $(u, v, w)$  (measured in wavelengths). In this special case:

$$\mathbf{b} = (\lambda u, \lambda v, \lambda w) = (\lambda u, \lambda v, 0) \quad (6.7)$$

Figure 6.2: Projections of vector  $\mathbf{s}$  on  $(u, v, w)$  axes.

**Simplification 6** Receiving elements have no direction dependence.

**Simplification 7A** All measurements are made in the same plane so  $w = 0$  and, as consequence, we treat only the  $(u, v)$  plane. In this case, the coherence function becomes:

$$C(r_1, r_2) = V_\nu(u, v, 0) = \int \int I_\nu(l, m) \exp \frac{-2\pi i(ul + vm)}{(1 - l^2 - m^2)^{1/2}} dldm \quad (6.8)$$

This is a Fourier transform relation between the complex visibility  $V_\nu$  (the spatial coherence function with separations expressed in wavelengths) and a modified intensity  $I_\nu(l, m)/(1 - l^2 - m^2)^{1/2}$ .

**Simplification 7B** All source are located in a small region of the sky. It is useful to fix a special coordinate system such that the phase tracking centre has  $s_0 = (0, 0, 1)$ . In this way:

$$C(r_1, r_2) = e^{-2\pi iw} V'_\nu(u, v) \quad (6.9)$$

where:

$$V'_\nu(u, v) = \int \int I_\nu(l, m) e^{-2\pi i(ul + vm)} dldm \quad (6.10)$$

In either simplified case, we can invert the Fourier transform to derive the intensity, e.g.:

$$I_\nu(l, m) = \int \int V'_\nu(u, v) e^{2\pi i(ul + vm)} dudv \quad (6.11)$$

This is the fundamental equation of synthesis imaging. In general, as said many times, interferometrists like to refer to the  $(u, v)$  plane. Remember that  $u, v$  (and  $w$ ) are measured in wavelengths. To fill this plane as much as possible is one of the main goal to have good data.

**Simplification 8** We have so far implicitly assumed that we can measure the visibility everywhere.

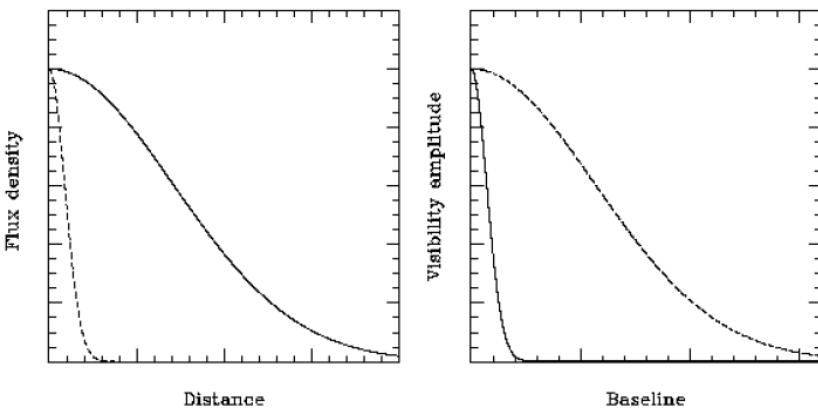


Figure 6.3: FT of Gaussians.

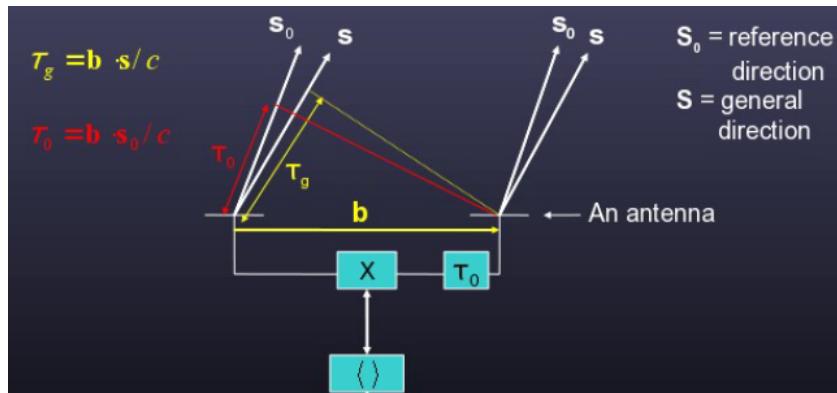


Figure 6.4: Geometrical and time delay.

**Example** Let's have a look of image 6.3. The Fourier transform of a Gaussian function is another Gaussian. In particular the FWHM on sky is inversely proportional to FWHM in spatial frequency: fat objects have thin Fourier transforms and vice versa.

## 6.1 Simplification 1

The first assumption is that radiation is monochromatic. However we are interested in observing wide bands, both for spectroscopy (e.g. *HI*, molecular lines) and for extra sensitivity for continuum imaging, so we have to get round this restriction. In fact, we can easily divide the band into **multiple spectral channels**. There are imaging restrictions only if the individual channels are too wide for the field size (often the case for older VLA continuum data). This effect, **bandwidth smearing**, restricts the usable field of view. The angular extent is roughly  $(\Delta\nu/\nu_0)(l^2 + m^2)^{1/2}$ . Modern correlators (e.g. ALMA) have many more frequency channels per unit frequency.

### 6.1.1 Geometrical delay

See figure 6.4. Assume a two-element interferometry. The two antennas observe the same source but the beams have different path and, as consequence, a geometrical delay which is translated in a time delay. Indeed a delay just corresponds to a change in arrival time of the wavefront. It is equivalent to a frequency-dependent phase change  $2\pi\tau\nu$ . On the contrary, geometrical delay is known for a given source and antenna position and can be removed by the correlator.

### 6.1.2 Multiple spectral channels

We make multiple channels by correlating with different values of lag,  $\tau$ . This is a delay introduced into the signal from one antenna with respect to another as seen before. For each quasimonochromatic

frequency channel, a lag is equivalent to a phase shift  $2\pi i\tau\nu$ , i.e.:

$$V(u, v, \tau) = \int V(u, v, \nu) e^{2\pi i\tau\nu} d\nu \quad (6.12)$$

This is another Fourier transform relation with complementary variables  $\nu$  and  $\tau$ , and can be inverted to extract the desired visibility as a function of frequency. In practice, we do this digitally, in finite frequency channels:

$$V(u, v, j\Delta\nu) = \sum_k V(u, v, k\Delta\tau) e^{-2\pi ijk\Delta\nu\Delta\tau} \quad (6.13)$$

Each spectral channel is then imaged (and deconvolved) individually. The final product is a **data cube**, regularly gridded in two spatial and one spectral coordinate.

## 6.2 Simplification 2

The second simplification is to treat the radiation field as a scalar quantity. The field is a vector, and we are interested in both components (i.e. its polarization). However this assumption makes no difference to the analysis as long as we measure two states of polarization (e.g. right and left circular or crossed linear) and account for coupling between the states. In this case there are some complications: the sky rotates and the two states are coupled (in our instrument, but also by propagation effects like Faraday rotation).

### 6.2.1 Polarization

Polarization is a property applying to transverse waves that specifies the geometrical orientation of the oscillations. In general classical EM waves are elliptically polarized.

Radio telescopes detect radiation from incoherent astronomical sources which is a superposition of waves and which is only partially polarized.

There are different possible descriptions. The simplest is in terms of two orthogonal basis states (e.g. crossed linear or RH and LH circular). A more relevant physics description takes into account the Stokes parameters like the total intensity  $I$ , polarization  $Q$  and  $U$  for the linear case and  $V$  for the circular.

The fractional linear polarization is given by  $(Q^2 + U^2)^{1/2}/I$ , the position angle is  $(1/2) \arctan U/Q$  and the fractional circular polarization is  $|V|/I$ .

### Measured polarization

The receiver usually measures two (nominally) orthogonal polarization states, e.g. right and left circular or crossed linear – but not exactly (hence a whole load of problems). Then the polarization is described by a  $2 \times 2$  matrix of correlations between components, which we can correlate and image separately. For example, for right and left circular polarizations (e.g. VLA, eMERLIN), if we fix polarization position angle in the antenna, the matrix is:

$$\begin{bmatrix} RR^* & RL^* \\ LR^* & LL^* \end{bmatrix} = \begin{bmatrix} I + V & Q + iU \\ Q - iU & I - V \end{bmatrix}$$

Instead for vertical and horizontal linear polarizations (e.g. ALMA, ATNF), the matrix is:

$$\begin{bmatrix} XX^* & XY^* \\ YX^* & YY^* \end{bmatrix} = \begin{bmatrix} I + Q & U + iV \\ U - iV & I - Q \end{bmatrix}$$

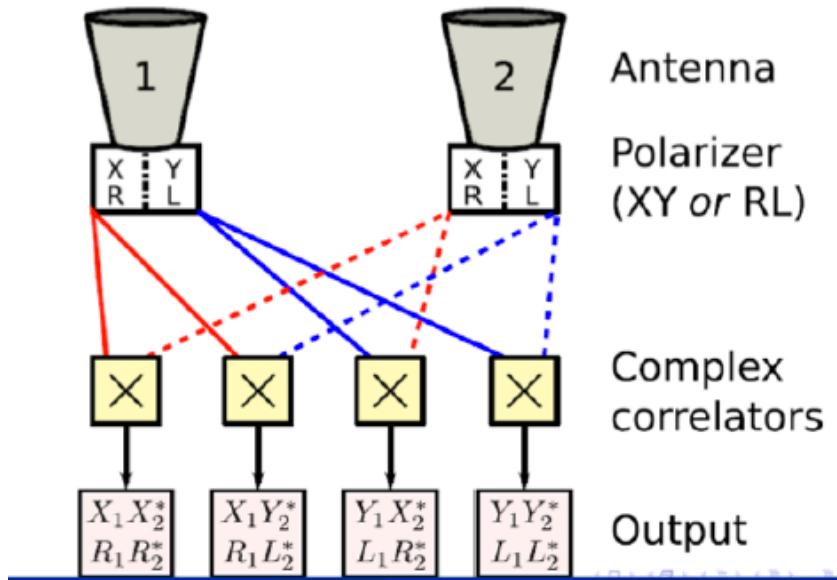


Figure 6.5: The polarization is described by a  $2 \times 2$  matrix of correlations between components, which we can correlate and image separately.

To understand better the concept, see figure 6.5.

### 6.2.2 Describing cross-correlations

Correlator output for antennas  $i$  and  $j$  is the coherence matrix  $V_{ij}$  of four elements (basis states  $p$  and  $q$  could be either linear or circular). This matrix is:

$$V_{ij} = \begin{bmatrix} XX^* & XY^* \\ YX^* & YY^* \end{bmatrix}$$

The formalism for describing linear corrupting effects which mix polarization states is:

$$V'_{ij} = J_i V_{ij} J_j^\dagger \quad (6.14)$$

where  $\dagger$  means transpose and complex conjugate.  $J'$ s ( $2 \times 2$ ) are Jones matrices for antennas  $i$  and  $j$ . This is the **measurement equation**. This is important for two reasons:

- it is invertible to get the true visibilities;
- Jones matrices can be expressed as products of individual Jones matrices for different corrupting effects.

Assumes that all corrupting effects are antenna-based and assume the following dictionary for Jones matrices.

- $S$  = sampling function
- $F$  = ionospheric effects
- $T$  = tropospheric effects
- $P$  = parallactic angle
- $X$  = linear polarization position angle
- $E$  = antenna voltage pattern
- $D$  = polarization leakage

- $G$  = time-variable gain
- $B$  = bandpass response
- $K$  = geometric compensation

Assuming this, then:

$$V = MKBG \int DEXPTFSI(l, m) e^{-2\pi i(ul+vm)} dl dm \quad (6.15)$$

### 6.3 Simplification 3

The third assumption is that sources are all far away. Strictly speaking, this means in the far field of the interferometer, so that the distance is  $> D^2/\lambda$ , where  $D$  is the interferometer baseline. This is true except in the extreme case of very long baseline observations of solar-system objects. For example, global VLBI observations of the Moon at 1 mm wavelength have distance  $\sim 4 \cdot 10^8$  m and  $D^2/\lambda \sim 10^{17}$  m!

### 6.4 Simplification 4

Another simplification is that radiation is not spatially coherent. This is generally true, even if the radiation mechanism is itself coherent (masers, pulsars). Indeed coherence can be produced by scattering (since signals from the same location in a source are spatially coherent), but travelling by different paths through the interstellar or interplanetary medium produce incoherence radiation.

This spatially incoherent radiation may become detectable in observations with extremely high spatial and spectral resolution.

### 6.5 Simplification 5 and 6

The others two simplifications are: space between us and the source is empty and the receiving elements have no direction dependence. This is not true in general. For example:

- antennas are usually designed to be highly directional;
- there are ionospheric and tropospheric fluctuations (which lead to path/phase and amplitude errors, sometimes seriously direction-dependent);
- there is ionospheric Faraday rotation, which changes the plane of polarization;
- there are interstellar or interplanetary scattering.

However, according to the assumptions made before, standard calibration deals with the case that there is no direction dependence (i.e. each antenna has an associated amplitude and phase which may be time-variable). Direction dependence is harder to deal with, but is becoming more important as field sizes increase.

#### 6.5.1 Primary beam

If the response of the antenna + atmosphere is direction-dependent, then we are measuring  $I_\nu(l, m)$   $D_{1\nu}(l, m)$   $D_{2\nu}^*(l, m)$  instead of  $I_\nu(l, m)$  (ignore polarization for now). An easier case is when the direction dependence is just due to the antennas, and they all have the same response:  $A_\nu(l, m) = |D_\nu(l, m)|^2$ . In this case:

$$V'_\nu(u, v) = \int \int A_\nu(l, m) I_\nu(l, m) e^{-2\pi i(ul+vm)} dl dm \quad (6.16)$$

We just make the standard Fourier inversion and then divide by the **primary beam**  $A_\nu(l, m)$ . This doesn't work for the atmosphere, or if antennas are different. The primary beam is shorthand for the directions over which  $|D_\nu(l, m)|$  is  $> 10\%$  of peak, or so. For a dish antenna, this is  $\sim \lambda/\text{diameter}$  (rad). For low-frequency array elements (e.g. dipoles) the primary beam covers a large fraction of the sky.

## 6.6 Simplification 7

The simplification number 7 is the following: (A) antennas are in a single plane or (B) the field is small. This is not true for wide-field imaging (except for snapshots). This is particularly relevant at low frequencies. According to these assumption, basic imaging equation becomes:

$$V_\nu(u, v, w) = \int \int I_\nu(l, m) \exp \frac{-2\pi i(ul + vm + (1 - l^2 - m^2)^{1/2}w)}{(1 - l^2 - m^2)^{1/2}} dl dm \quad (6.17)$$

This is no longer a 2D Fourier transform, so analysis becomes much more complicated. However this is not so much an issue for ALMA.

## 6.7 Simplification 8

We have implicitly assumed that we can measure the visibility function everywhere. In fact:

- we have a number of antennas at fixed locations on the Earth;
- the Earth rotates;
- we make measurements over finite (usually short) time intervals;

This means that we actually measure only at discrete  $u, v$  (and  $w$ ) positions. Of course our main goal is to fill as much as possible this plane.

## 6.8 Sampling and imaging

In 2D, this process can be described by a sampling function  $S(u, v)$  which is a delta function where we have taken data and zero elsewhere. The following relation is the **dirty image**, which is the Fourier transform of the sampled visibility data:

$$I_\nu^D(l, m) = \int \int V_\nu(u, v) S(u, v) e^{2\pi i(ul + vm)} du dv \quad (6.18)$$

Using the convolution theorem:

$$I_\nu^D(l, m) = I_\nu(l, m) \otimes B(l, m) \quad (6.19)$$

in which the **synthesised beam** is:

$$B(l, m) = \int \int S(u, v) e^{2\pi i(ul + vm)} du dv \quad (6.20)$$

The dirty image is the convolution of the true image of the sky with the dirty beam. Working out the true image of the sky from this is **deconvolution**.

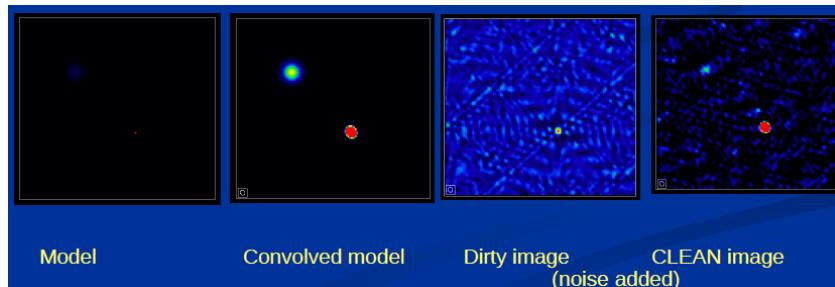


Figure 6.6: Application of the CLEAN algorithm.

### 6.8.1 Deconvolution

The next stage in the imaging process is to estimate the convolution of the sky with a well-behaved restoring beam (usually a Gaussian function) rather than the dirty beam. This is deconvolution. Methods for this include CLEAN and maximum entropy algorithms (see previous chapters). An example is shown in figure 6.6.

### 6.8.2 Resolution, maximum scale and field size

Some useful parameters:

- the *resolution/rad* is  $\sim \lambda/d_{max}$ ;
- the *maximum observable scale/rad* is  $\sim \lambda/d_{min}$ ;
- the *primary beam/rad* is  $\sim \lambda/D$ .

But some brightness distributions are in principle undetectable. Indeed uniform sources and sinusoid with Fourier transform in an unsampled part of the  $u - v$  plane are not detectable. Moreover:

- sources with all brightness on scales  $> \lambda/d_{min}$  are **resolved out**;
- sources with all brightness on scales  $< \lambda/d_{max}$  look like **points**.

As said many times, good coverage of the  $u - v$  plane (using many antennas and the Earth rotation) allows high quality imaging.

# Chapter 7

## Difference between a CCD and a interferometer

### 7.1 How do we characterize what a source looks like?

Let's think about a usual images taken with CCD and consider a perfect telescope in space observing an unresolved point source. This produced an Airy pattern with a characteristic width  $\theta = 1.22\lambda/D$ . If we observe a more complicated object made up by different points (like a cluster with many stars, for example), each point in the source produces a displaced Airy pattern and the superposition of these limits the detail visible in the final image. So even in perfect conditions, a telescope image is NOT a perfect representation of what's in the sky.

How do we describe this process mathematically? We can consider the fundamental relationship for isoplanatic imaging that is:

$$I(l, m) = \int \int P(l - l', m - m') O(l', m') dl' dm' \quad (7.1)$$

i.e. the observed brightness distribution is the true source brightness distribution convolved with a point-spread function,  $P(l, m)$ . Note that here  $l$  and  $m$  are angular coordinates on the sky, measured in radians. But now let's think to forget this and think differently. In particular, take the Fourier transform of the convolution on the previous slide to get:

$$I(u, v) = T(u, v) \times O(u, v) \quad (7.2)$$

where  $u$  and  $v$  are now spatial frequencies measured in  $\text{radians}^{-1}$  and the functions are:

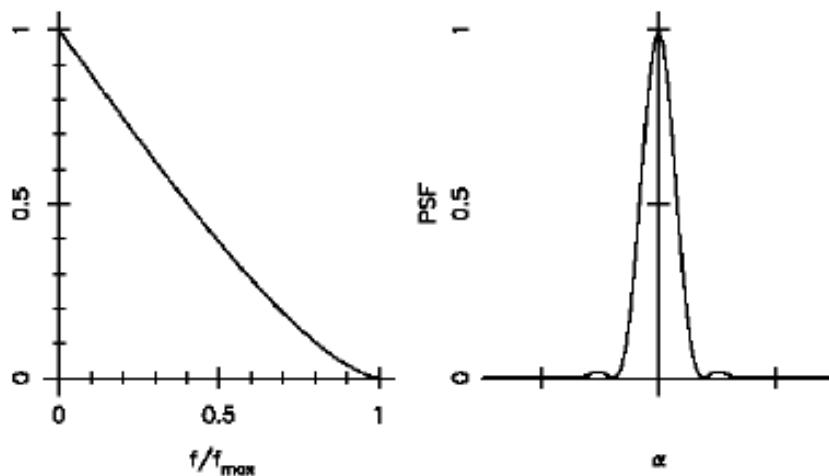
- $O(u, v)$ , the **Fourier spectrum** that encapsulates the essential properties of the target;
- $T(u, v)$ , a complex multiplicative **transfer function** that contains the essential properties of the imaging system. It is just the Fourier transform of the PSF.

So we find that measuring  $I(l, m)$  is equal to measure  $I(u, v)$ .

In general the transfer function is obtained from the auto-correlation of the complex aperture function:

$$T(u, v) = \int \int A * (x, y) A(x + u, y + v) dx dy \quad (7.3)$$

Here  $x$  and  $y$  denote co-ordinates in the aperture. In the absence of aberrations  $A(x, y)$  is equal to 1 where the aperture is transmitting and 0 otherwise. Some key features of this formalism worth noting are:

Figure 7.1: Behavior of  $T(f)$  and of the PSF.

- for each spatial frequency,  $u$ , there is a physical baseline,  $B$ , in the aperture, of length  $\lambda u$ ;
- different shaped apertures measure different Fourier components of the source;
- different shaped apertures give different PSFs. This is important, for example, for planet detection.

### 7.1.1 Transfer function of circular aperture

For a circularly symmetric aperture, the transfer function can be written as a function of a single co-ordinate:  $T(f)$ , with  $f^2 = u^2 + v^2$ . In this case, the PSF is the familiar Airy pattern and the full-width at half-maximum of this is at approximately  $\lambda/D$ . In particular,  $T(f)$  falls smoothly to zero at  $f_{\max} = D/\lambda$ , as seen in image 7.1. The behavior of  $T(f)$  is why even perfect telescopes don't image sources perfectly.

So, what should you really learn from all this? The decomposition of an image consists into a series of spatially separated compact PSFs and the equivalence of this is a superposition of non-localized sinusoids, i.e. Fourier components.

In particular, the action of ANY incoherent imaging system acts as a filter for the Fourier spectrum of the source. As consequence there is an association of each Fourier component (or spatial frequency) with a distinct physical baseline in the aperture that samples the light. So, the formation of the PSF arises from the mix of different spatial frequencies measured by the imaging system. Of course, if you can measure the Fourier components of the source, you should be able to do your science.

#### Brief summary

- Forget thinking about what a source looks like – start thinking about what its Fourier transform looks like.
- If you can measure all the Fourier transform, that's the same as making an image.
- Interferometers are devices to measure the Fourier content of your target.

## 7.2 What/how do interferometers measure?

Consider a two-element interferometer, like in image 7.2.

Consider now distant point source and the radiation coming from it. It is very distant so we can assume flat incident phasefronts that enter in the two elements. For this reason there is a geometric

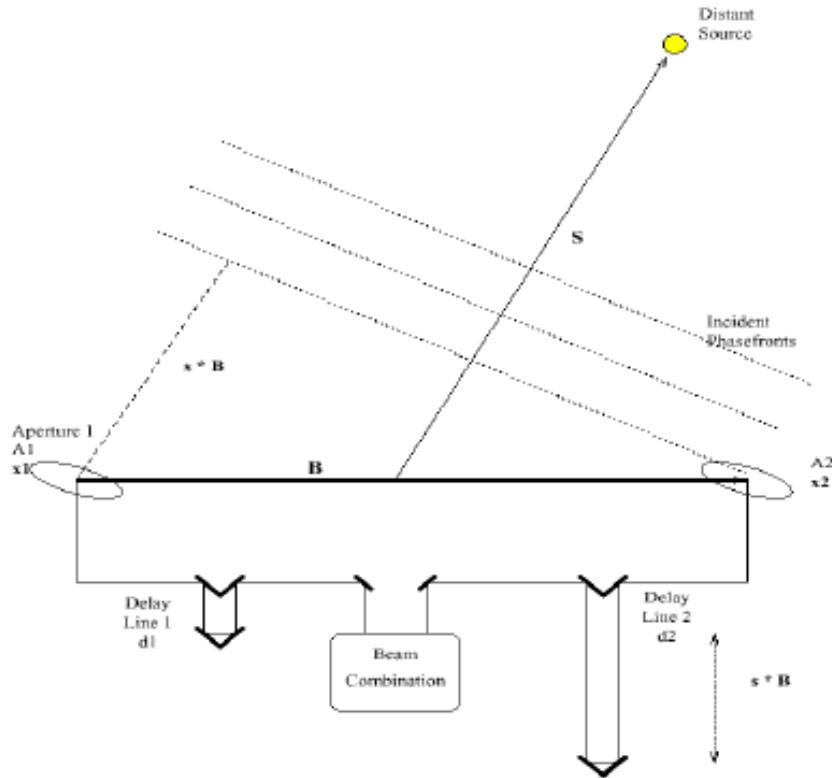


Figure 7.2: Two-element interferometer.

delay that changes according to the movement of the source and that must be compensated. After this, there is a combination of the beams and then the detection of the resulting output.

In figure 7.2 telescopes are located at  $x_1$  and  $x_2$ , the baseline  $B$  is equal to  $x_1 - x_2$ . The baseline is a fundamental parameter, it governs sensitivity to different to different angular scales. Pointing direction towards source is  $S$ , the geometric delay is  $\bar{s}B$  where  $\bar{s} = S/|S|$  and optical paths along two arms are  $d_1$  and  $d_2$ .

In particular combination of the beams produces fringes with modulation, like in figure 7.3.

Mathematically, the output of a two-element interferometer is a combination of electric fields from the two collectors that can be described as:

$$\Psi_1 = A \exp ik[\bar{s}B + d_1] \exp -i\omega t \quad (7.4)$$

and:

$$\Psi_2 = A \exp ikd_2 \exp -i\omega t \quad (7.5)$$

So, summing these at the detector we get as result:

$$\Psi = \Psi_1 + \Psi_2 = A[\exp ik[\bar{s}B + d_1] + \exp ikd_2] \exp -i\omega t \quad (7.6)$$

Here the **time averaged intensity**  $\langle \Psi \Psi^* \rangle$  will be given by:

$$\langle \Psi \Psi^* \rangle \propto \langle [\exp ik[\bar{s}B + d_1] + \exp ikd_2] \times [\exp -ik[\bar{s}B + d_1] + \exp -ikd_2] \rangle \quad (7.7)$$

which is equivalent to:

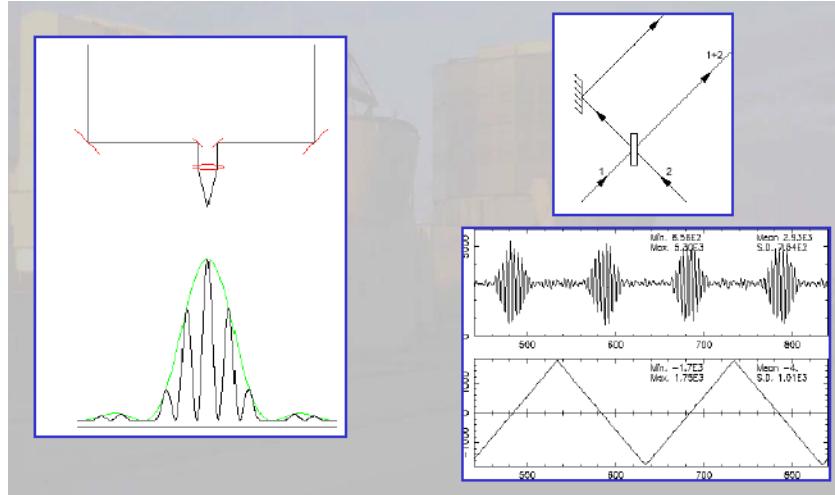


Figure 7.3: Modulated fringe pattern produced by the combination of the beams. The left image is a scheme of the resulting output. The right image is the modulated pattern of many images.

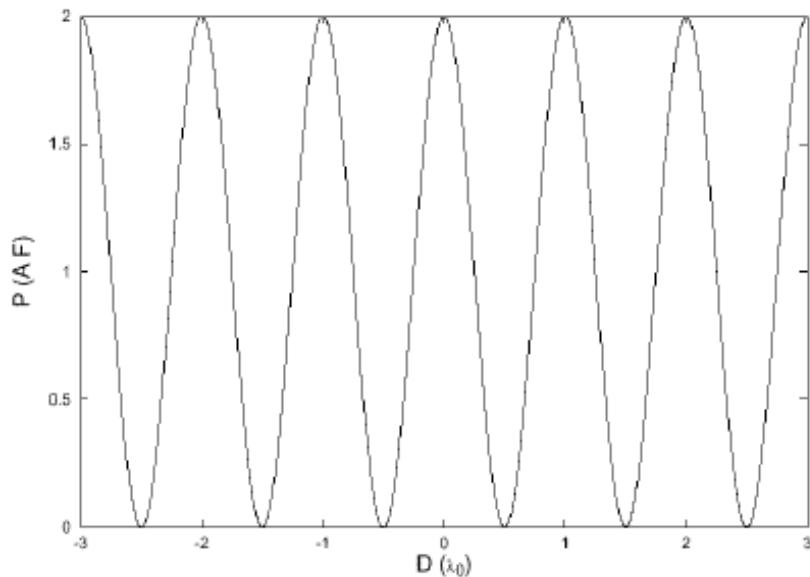


Figure 7.4: Intensity variation as function of  $D$ .

$$\langle \Psi \Psi^* \rangle \propto 2 + 2 \cos k[\bar{s}B + d_1 - d_2] \propto 2 + 2 \cos kD \quad (7.8)$$

where we define  $D = \bar{s}B + d_1 - d_2$  so  $D$  is a function of the path lengths,  $d_1$  and  $d_2$ , the pointing direction (i.e. where the target is) and the baseline. So the **detected intensity** is given by:

$$I = \langle \Psi \Psi^* \rangle \quad (7.9)$$

As shown in figure 7.4, the intensity varies sinusoidally with  $kD$  and  $k = 2\pi/\lambda$ . Here adjacent peaks are separated by:

$$\Delta d_{1,2} = \lambda = \Delta(\bar{s}B) \quad (7.10)$$

and:

$$\Delta(1/\lambda) = 1/D \quad (7.11)$$

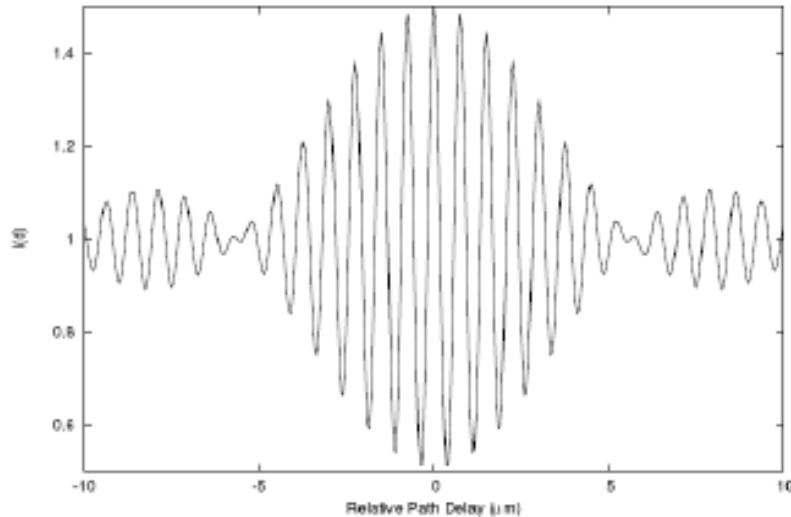


Figure 7.5: Intensity as function of relative path delay.

So, as said before, we get fringes but what should we measure? From an interferometric point of view the key observables are the contrast and location of these modulations in intensity (image 7.5). In particular we can identify:

- the **fringe visibility** at  $D = 0$ :

$$V = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \quad (7.12)$$

- the **fringe phase**. The location of the white-light fringe are measured from some reference in radians.

So, summarizing, the fringe amplitude and phase measure the amplitude and phase of the Fourier transform of the source at one spatial frequency.

### Brief summary

- All (2-element summing) interferometers produce a power output that shows a cosinusoidal variation – these are its fringes.
- Properties of these fringes encode the amplitude and phase of the Fourier transform of the target.
- For every pair of telescopes (i.e. 2-element interferometer) you get one measurement.
- Interferometers are just machines to make these single measurements.

#### 7.2.1 What happens with polychromatic light?

In case of polychromatic light, we can integrate the previous result over a range of wavelengths. For example, for a uniform bandpass of  $\lambda_0 \pm \Delta\lambda/2$  (i.e.  $\nu_0 \pm \Delta\nu/2$ ) we obtain:

$$I \propto \int_{\lambda_0 - \Delta\lambda/2}^{\lambda_0 + \Delta\lambda/2} 2[1 + \cos 2\pi D/\lambda] d\lambda \quad (7.13)$$

Experimentally, we obtain a graphic like image 7.6.

The fringes are modulated with an envelope with a characteristic width equal to the coherence length,  $\Lambda_{coh} = \lambda_0^2 / \Delta\lambda$  (see figure 7.7).

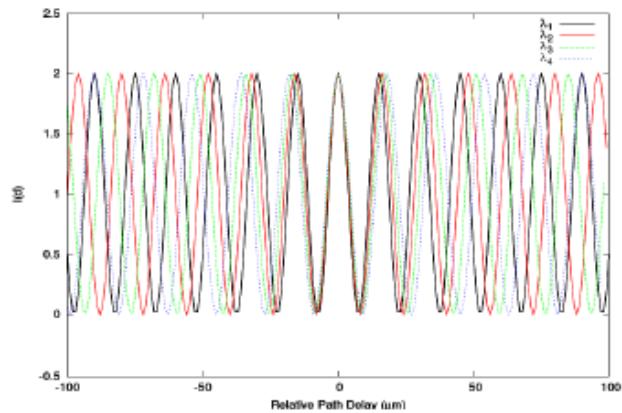
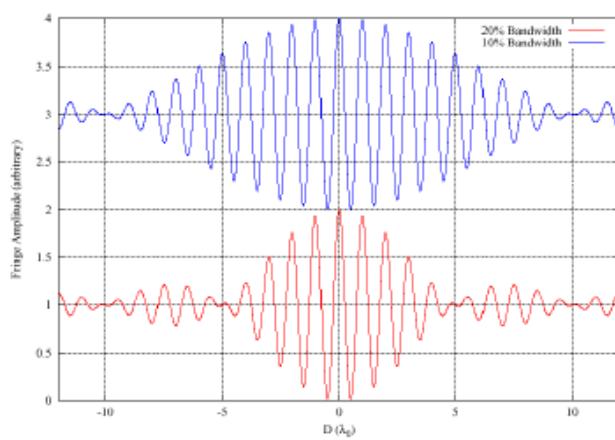


Figure 7.6: Intensity as function of relative path delay.

Figure 7.7: Fringe amplitude as function of  $D$ .

## Key ideas regarding the interferometric output

- The output of the interferometer is a time averaged intensity.
- The intensity has a co-sinusoidal variation – these are the “fringes”.
- The intensity varies a function of  $kD$ , which itself can depend on:
  1. the wavevector  $k = 2\pi/\lambda$ ;
  2. the baseline  $B$ ;
  3. the pointing direction  $s$ ;
  4. the optical path difference between the two interferometer arms.
- If things are adjusted correctly, then the interferometer output can remain fixed: in that case there will be no fringes.
- The response to a polychromatic source is given by integrating the intensity response for each color.
- This alters the interferometric response and leads to modulation of the fringe contrast:
  1. the desired response is only achieved when  $k[\bar{s}B + d_1 - d_2] = 0$ ;
  2. this is the so called white-light condition.
- This is the primary motivation for matching the optical paths in an interferometer and correcting for the geometric delay.
- The narrower the range of wavelengths detected, the smaller is the effect of this “coherence envelope”. This is usually quantified via the coherence length,  $\Lambda_{coh} = \lambda_0^2/\Delta\lambda$ .

How well do delay lines have to perform? The Optical Path Delay (OPD) added can be as large as the maximum baseline. The OPD correction varies roughly as  $B\cos(\theta)d\theta/dt$ , with  $\theta$  the zenith angle while the correction has to be better than  $l_{coh} \sim \lambda^2/\Delta\lambda$ .

### 7.2.2 Issues with optical/IR delay lines

Unless very specialized beam-combining optics are used it is only possible to correct the OPD for a single direction in the sky. This gives rise to a FOV limitation:  $\theta_{max} \leq [\lambda/B][\lambda/\Delta\lambda]$ .

For an optical train in air, the OPD is actually different for different wavelengths since the refractive index  $n = n(\lambda)$ . This longitudinal dispersion implies that different locations of the delay line carts will be required to equalize the OPD at different wavelengths!

**Example** For a 100 m baseline and a source 50 degrees from the zenith this  $\delta OPD$  corresponds to  $\sim 10 \mu m$  between  $2.0 - 2.5 \mu m$ . More precisely, this implies the use of a spectral resolution,  $R > 5$  to ensure good fringe contrast ( $> 90\%$ ) in the  $K(J)$  band.

### 7.2.3 Timeout

The fringe modulation encodes information about the Fourier spectrum of the target. Exactly how does this work? Let's think about the heuristic operation of an interferometer (figure 7.8).

Each unresolved element of the source produces its own fringe pattern. These have unit visibilities and phases that are associated with the location of that source element in the sky: this is the basis for astrometric measurements with interferometers. The observed fringe pattern from a distributed source is just the intensity superposition of these individual fringe pattern. This relies upon the individual elements of the source being “spatially incoherent”. The resulting fringe pattern has a contrast that is reduced with respect to that from each source individually. This means we detect less correlated

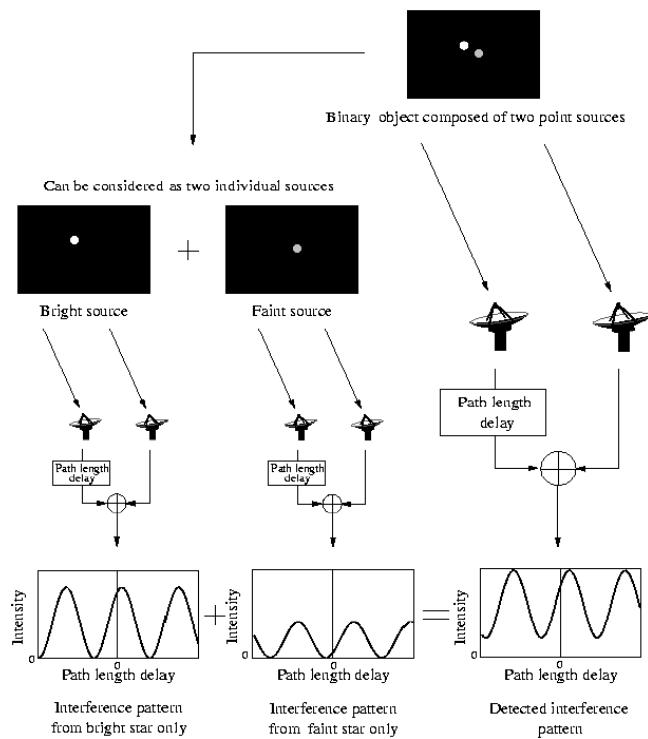


Figure 7.8: Operations of an interferometer.

flux. Moreover the positions of the sources are encoded (in a scrambled manner) in the resulting fringe phase.

### Brief summary

- For a given baseline, measurements of the output of an interferometer are made for different values of  $kD$ .
- These measurements allow you to recover the Fourier transform  $V(u)$ , of the target at a single spatial frequency,  $u$ , determined by the projected baseline:  $u = B_{proj}/\lambda$ .
- If the projected baseline is large the interferometer probes small scale structures (spatial frequencies), if short, it probes larger structures.
- Even though the Fourier transform is complex and you measure a real signal you recover it fully.

## 7.3 Planning and undertaking interferometric science

Planning interferometric observations, astronomers must answer to some question.

- Which bits of the Fourier spectrum of the target,  $V(u, v)$ , are the ones you wish to measure?
- How easy will it be to measure these parts?
- How easy will it be for you to interpret these measurements?
- What do we do with the measurements of  $V(u, v)$  if we wish to make a map of the sky?
- How faint can you go?

You can not answer to them unless you can do FTs in your head. For example, think about simple 1d FT and take into account that  $V$  is the Fourier transform of source brightness while  $\nu$  is the spatial frequency derived from projected baseline divided by  $\lambda$ .

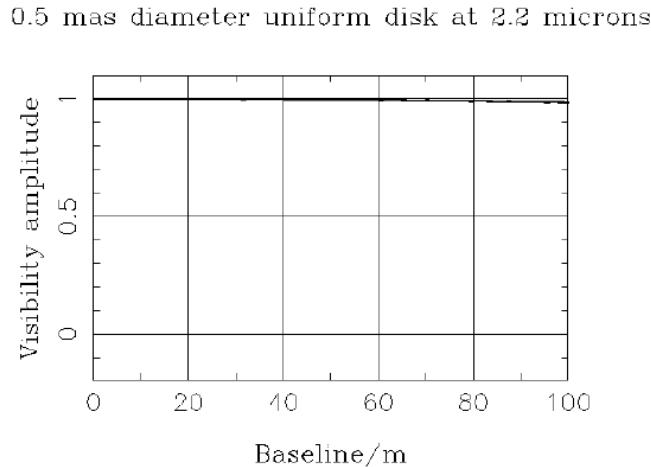


Figure 7.9: Visibility function as function of the baseline for 0.5 mas diameter uniform disk at 2.2 microns.

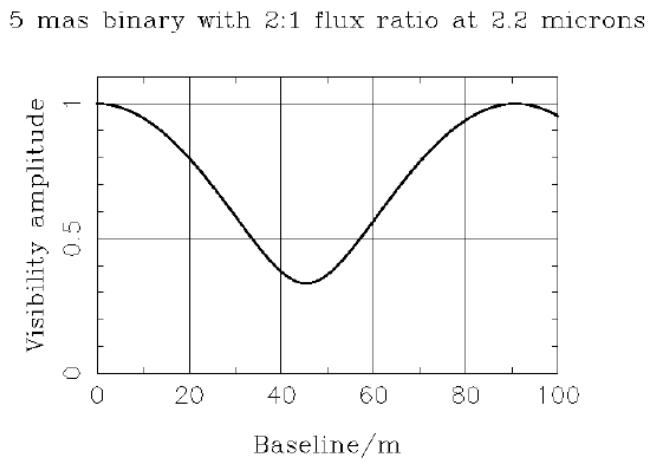


Figure 7.10: Visibility function as function of the baseline for 5 mas binary with 2.1 flux ratio at 2.2 microns.

### 7.3.1 A perfectly unresolved source, not on-axis

Consider a perfectly unresolved source, not on-axis of strength  $A_1$  and located at angle  $l_1$  relative to the optical axis, then:

$$V(u) = e^{-i2\pi(ul_1)} \quad (7.14)$$

which is the *visibility amplitude* measured in unity of  $u$ . The *visibility phase* varies linearly with  $u = B/\lambda$ . Sources such as this are easy to observe since the interferometer output gives fringes with high contrast, as visible in figure 7.9.

### 7.3.2 An unequal binary star

Consider now a binary system, a double source comprising point sources of strength  $A_1$  and  $A_2$  located at angles 0 and  $l_2$  relative to the optical axis, then:

$$V(u) \propto A_1 + A_2 e^{-i2\pi(ul_2)} \quad (7.15)$$

The visibility amplitude and phase oscillate as functions of  $u$ , as seen in figure 7.10. To identify this as a binary, baselines from  $0 \rightarrow \lambda/l_2$  are required. If the ratio of fluxes is large the modulation of the visibility becomes difficult to measure, i.e. the contrast of the interferometric fringes is similar for all baselines.

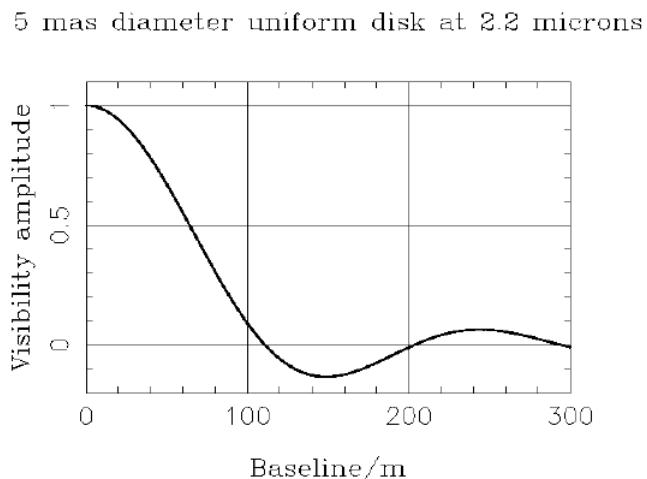


Figure 7.11: Visibility function as function of the baseline for 5 mas diameter uniform disk at 2.2 microns.

### 7.3.3 A uniform stellar disc

If instead we consider a uniform stellar disc source of diameter  $\theta$ , then the visibility amplitude falls rapidly as  $u_r$  increases, as visible in figure 7.11. To identify this as a disc requires baselines from  $0 \rightarrow \lambda/\theta$  at least. Information on scales smaller than the disc correspond to values of  $u_r$  where  $V \ll 1$ , and is difficult to measure. This is because the interferometer output gives fringes with very low contrast.

### 7.3.4 How does this help planning observations?

Compact sources have visibility functions that remain high whatever the baseline, and produce high contrast fringes all the time. On the contrary, resolved sources have visibility functions that fall to low values at long baselines, giving fringes with very low contrast. So *fringe parameters for fully resolved sources will be difficult to measure*.

In general, to usefully constrain a source, the visibility function must be measured adequately. Measurements on a single, or small number of, baselines may not be enough for unambiguous interpretation. Indeed, imaging – which necessarily requires information on both small and large scale features in a target – will generally need measurements where the fringe contrast is both high and low.

#### Brief summary

- When planning an interferometric measurement you must have some idea what the target looks like.
- You need to have thought which bits of the Fourier transform of the source are most valuable to measure.
- Successful interferometry demands a lot more of the user than conventional imaging.

## 7.4 Imaging and sensitivity

In order to make maps with interferometers there are some "rules". First of all the number of visibility data must be bigger (or eventually equal) to the number of filled pixels in the recovered image so  $N_{tel}(N_{tel} - 1)/2 \times \text{number of re-configurations}$  must be bigger than the number of pixels. Moreover, the distribution of samples taken of the Fourier plane should be as uniform as possible in order to aid deconvolution of the interferometric PSF.

Remember! The range of interferometer baselines, i.e.  $B_{max}/B_{min}$ , will govern the range of spatial scales in the map. However there is no need to sample too finely in the Fourier plane: for a source of

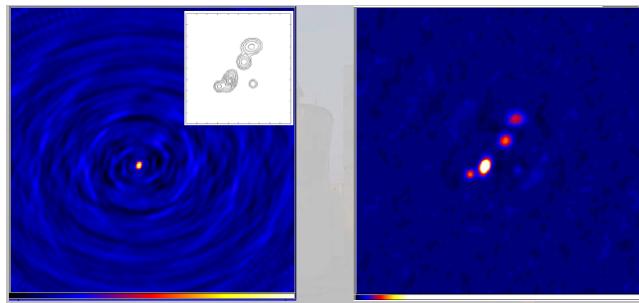


Figure 7.12: Interferometric maps.

maximum extent  $\theta_{max}$ , sampling very much finer than  $\Delta u \sim 1/\theta_{max}$  is unnecessary.

#### 7.4.1 How maps are actually recovered

The measurements of the visibility function are secured and calibrated. These can be represented as a sampled version of the Fourier transform of the sky brightness:

$$V_{meas}(u, v) = V_{true}(u, v) \times S(u, v) \quad (7.16)$$

These data are inverse Fourier transformed to give a representation of the sky, similar to that of a normal telescope, albeit with a strange PSF:

$$\int \int S(u, v) V_{true}(u, v) e^{+i2\pi(ul+vm)} du dv = I_{norm}(l, m) * B_{dirty}(l, m) \quad (7.17)$$

where  $B_{dirty}(l, m)$  is the Fourier transform of the sampling distribution, or the so-called dirty-beam, and the image is usually referred to as the “dirty” map. So you get a map that has a “crazy” point spread function.

For example, look at figure 7.12. The looks quite strange indeed interferometric PSF’s can often be horrible. Correcting an interferometric map for the Fourier plane sampling function is known as **deconvolution** and is broadly speaking straightforward. Of course more filled is the  $(U, v)$  plane, better is for the imaging.

#### 7.4.2 What does ”sensitivity” mean for interferometry?

In case of interferometry, sensitivity depends on the presence of a guide star, bright enough to be used as reference and possibly close to the target, and depends on AO system used for the telescope in case of optical interferometry. Of course depends on the sensitivity of the detector and on the exposure time.

So, how do we assess interferometric sensitivity? The “source” has to be bright enough to allow stabilisation of the interferometer against any atmospheric fluctuations. Moreover, it has to be bright enough to allow a reasonable signal-to-noise for the fringe parameters to be build up over some total convenient integration time. This will be measured in minutes.

Once this achieved, the faintest features one will be able to interpret reliably will be governed by  $S/N$  ratio and number of visibility data measured.

In most cases the sensitivity of an interferometer will scale like some power of the measured fringe contrast  $\times$  another power of the number of photons detected while the fringe is being measured.

This highlights fringe contrast and throughput as both being critical. This also highlights the difficulty of measuring resolved targets where  $V$  is low.

So, how do we assess interferometric sensitivity? The “source” has to be bright enough to:

- allow stabilisation of the interferometer against any atmospheric fluctuations;
- allow a reasonable signal-to-noise for the fringe parameters to be build up over some total convenient integration time. This will be measured in minutes.

Once this achieved, the faintest features one will be able to interpret reliably will be governed by  $S/N$  ratio and number of visibility data measured. In most cases the sensitivity of an interferometer will scale like some power of the measured fringe contrast  $\times$  another power of the number of photons detected while the fringe is being measured. This highlights fringe contrast and throughput as both being critical. This also highlights the difficulty of measuring resolved targets where  $V$  is low.

The  $S/N$  ratio is then given by:

$$S/N \propto (VN)^2 / [(N + N_{dark})^2 + 2(N + N_{dark})N^2V^2 + 2(N_{pix})^2(\sigma_{read})^4]^{1/2} \quad (7.18)$$

in which  $V$  is apparent visibility,  $N$  is the number of detected photons,  $N_{dark}$  is the dark current,  $N_{pix}$  is the number of pixels and  $\sigma_{read}$  is the readout noise/pixel. So  $S/N$  depends on the signal coming from the target, on the amount of read noise on the detector, on the amount of dark/thermal background and on the source visibility, i.e. its structure. This last one is not too unusual really. It just means that what matters is not the integrated brightness of the target but the surface brightness, i.e. the brightness per unit solid angle on the sky.

Once we have measured the visibility function of the source, we have to interpret these data. This can take many forms:

- small amount of Fourier data  $\rightarrow$  model-fitting;
- moderate amount of Fourier data  $\rightarrow$  model-fitting and rudimentary imaging;
- large amount of Fourier data  $\rightarrow$  model-fitting and model-independent imaging.

Of course increasing the number of telescopes, also resolution and quality of the images increase.

### Brief summary

- There are alternative methods to describe images: Fourier decomposition, spatial frequencies, physical baselines.
- Interferometer make fringes and the fringe amplitude and phase are what is important. More precisely, these measure the FT of the sky brightness distribution. A measurement with a given interferometer baseline measures a single Fourier component (usually the square of  $V$  and its phase are measured).
- Multiple baselines are obligatory for studying a source reliably. Resolved targets produce fringes with low contrast – these are difficult to measure well. Once a number of visibilities have been measured, reliable interpretation can take multiple forms – making an image is only required if the source is complex.

## 7.5 How is what we have learnt impacted by the atmosphere?

We visualise the atmosphere altering the phase (but not amplitude) of the incoming wavefronts (see figure 7.13). We know that this impacts the instantaneous image but we need to understand how this impacts the fringe contrast and phase, which are what we are actually interested in measuring.

In particular, we characterize spatial fluctuations in the wavefront using Fried's parameter  $r_0$ . This is the circular aperture size over which the mean square wavefront error is approximately 1 radian $^2$  and this scales as  $\lambda^{6/5}$ . The dimension of the telescope is a fundamental parameter. If the diameter is bigger or smaller than  $r_0$ , this delimit different regimes of instantaneous image structure:

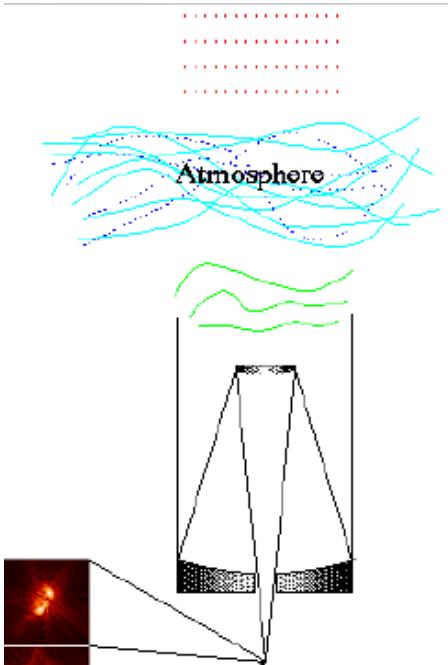
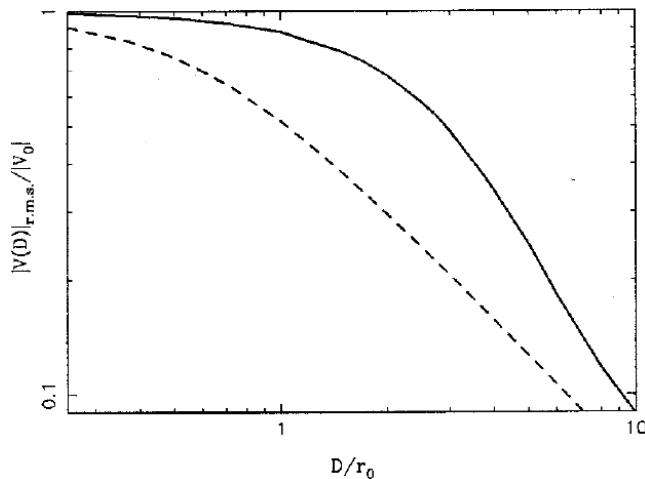


Figure 7.13: Atmosphere alters the phase of the incoming wavefront.

Figure 7.14: Visibility as function of  $D/r_0$ .

- if  $D < r_0 \rightarrow$  quasi-diffraction limited images with image motion;
- if  $D > r_0 \rightarrow$  high contrast speckled (distorted) images.

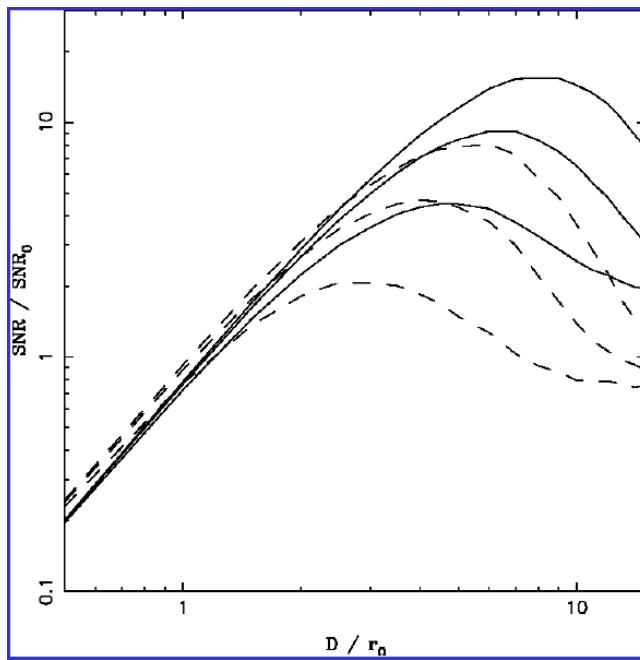
while if  $D \sim r_0$ , spatial fluctuations are not very important.

So, spatial corrugations affect the instantaneous fringe contrast. In particular, it reduces the rms visibility amplitude (dashed line in figure 7.14) as  $D/r_0$  increases. It leads to increases fluctuations in  $V$  but both imply a loss in sensitivity and calibration becomes less reliable. Moderate improvement is possible with tip-tilt correction (black line in figure 7.14). Higher order corrections improve things but more slowly.

So in general, the mean fringe contrast now is a function of both the source structure and the atmospheric conditions.

### 7.5.1 How can we mitigate against these perturbations?

To mitigate these perturbations, in principle there are two approaches to deal with spatial fluctuations for telescopes of finite size.

Figure 7.15:  $S/N$  ratio as function of  $D/r_0$ .

First of all, we can use an **adaptive optics system** correcting higher order Zernike modes. It can use either the source or an off-axis reference star to sense atmosphere but need to worry about how bright and how far off axis is sensible.

On the other hand, we can use **spatial filters** to spatially filter the light coming from the collectors. The light passes through either a monomode optical fibre or a pinhole and this trades off a fluctuating visibility for a variable throughput.

Many interferometers use both strategies. Image 7.15 shows how the  $S/N$  for fringe contrast measurement scales with telescope size. In this graph solid line represents results with spatial filters while dashed line represents measurements without spatial filters. Different curves are for 2, 5 and 9 Zernike mode correction.

Implications of this are perfect wavefronts for  $S/N \propto D$ . Of course spatial filtering always helps. These methods can work with large  $D/r_0$  but if this ratio is too large for the AO system, it is necessary to make  $D$  smaller. In general Natural Guide Stars (NGS) adaptive optics systems basically offer modest improvements in sky coverage, and allow photons to be collected faster for bright sources.

### Brief summary

- Spatial fluctuations in the atmosphere lead to a reduction in the mean fringe contrast from the intrinsic source-dependent value.
- We have to calibrate this effect by looking at source whose visibility we know a priori – usually an unresolved target.
- We have to rely upon the instrumental and atmospheric characteristics being identical (in a statistical sense) for the two observations.
- It is possible to moderate the effects of the atmosphere using AO and spatial filtering.

### 7.5.2 Temporal effects

We again visualise the atmosphere altering the phase (but not amplitude) of the incoming wavefronts. We know that these perturbations change with time. Again, we need to understand how this impacts

the fringe contrast and phase. In particular, temporal fluctuations in the atmosphere lead to a reduction in the mean fringe contrast from the intrinsic source-dependent value if the exposure time is too long. This cannot easily be calibrated, so we try to avoid it. More importantly, temporal fluctuations change the measured fringes phases in a random manner. One solution is to monitor (and possibly correct) the atmospheric perturbations in real time. A second strategy is to try to measure combinations of phases that are immune to these perturbations.

# Chapter 8

## Sensitivity and noise

As seen in previous chapters, radio telescopes convert EM waves into output power as a function of radio freq  $\nu$  and time  $t$ . The sources they observe are characterized by astrophysical signals incredibly weak and measured in Janskys ( $10^{-26} \text{ W/m}^2/\text{Hz}$ ). However almost of all power we measure is **noise**.

In particular, we usually talk about power in terms of *temperature* as the units are better, as converted using Boltzmann's constant ( $k = 1.38 \times 10^{-16} \text{ erg/K}$ ). At most radio frequencies and for most thermal emission sources, we do not need full Plank blackbody law:

$$B_\nu(\nu, T) = \frac{2h\nu^3}{c^2} \frac{1}{\exp \frac{h\nu}{kT} - 1} \quad (8.1)$$

because in case of radio sources  $h\nu/(kT) \ll 1$  so we can use an approximation but be careful in sub-mm and for cold sources. Indeed the denominator can be expanded and we can get a simpler relation:

$$B_\nu(\nu, T) = \frac{2h\nu^3}{c^2} \frac{kT}{h\nu} = \frac{2kT\nu^2}{c^2} \quad (8.2)$$

that in terms of wavelength, becomes:

$$B_\nu(\lambda, T) = \frac{2kT}{\lambda^2} \quad (8.3)$$

This approximation is called **Rayleigh-Jeans approximation** that is brightness (or intensity) of a thermal source. But we usually want **flux density**, and for that, for small sources of angular size  $\Omega$ :

$$S_\nu = B_\nu \Omega \quad (8.4)$$

Most radio sources do not emit only blackbody radiation, yet we can still use Rayleigh-Jeans equation to describe a property of them. In particular, brightness temperature  $T_b$  is the BB temperature that a source would have to have such that it's brightness at frequency  $\nu$  matches that of a blackbody:

$$T_b(\nu) = \frac{I_\nu c^2}{2k\nu^2} \quad (8.5)$$

in which  $I_\nu$  is the spectral brightness or intensity. ATTENTION: brightness temperature is *not* usually a physical temperature!

Similar thing happens with power per unit bandwidth emitted by a warm resistor:

$$P_\nu = \frac{h\nu}{\exp \frac{h\nu}{kT} - 1} \quad (8.6)$$

that can be approximated in  $P_\nu = kT$ , also called **Nyquist approximation**. This lets us measure power at the back of a radio receiver in terms of temperature. Associated to these concepts, there is also the concept of *antennateperature*. It is defined as the temperature of an ideal resistor that would produce the same Rayleigh-Jeans power per unit frequency as the antenna output:

$$T_A = \frac{P_\nu}{k} \quad (8.7)$$

In general, the instantaneous signal-to-noise in the radio might be:

$$S/N \sim \frac{T_A}{T_{sys}} \quad (8.8)$$

where  $T_{sys}$  is the system temperature. If a source has temperature  $T_b$  and is much larger than the beam of the telescope, then  $T_A = T_b$ . However, if the source angular size  $\Omega_s$  is smaller than the beam size  $\Omega_b$  then:

$$\frac{T_A}{T_b} = \frac{\Omega_s}{\Omega_A} \quad (8.9)$$

The ratio  $\Omega_s/\Omega_b$  is called **beam filling factor**.

## 8.1 Radio sensitivity

We can guess some things about how the sensitivity of radio telescopes must scale. First of all,  $S/N$  is what we want to determine:

- for **signal**: higher is better because we are collecting radio waves and/or photons. For telescopes this is the effective area  $A_{eff}$ ;
- for **noise**: we must want as little as possible. We describe noise in radio astronomy in terms of temperature, and in this case, the system temperature,  $T_{sys}$ .

In particular, we define the **radio sensitivity**  $\propto A_{eff}/T_{sys}$ .

Flux density  $S_\nu$  is in units of  $W/m^2/Hz$  and antennas collect power  $P_\nu$  in  $Hz$ . From an un-polarized source,  $P_\nu = AS_\nu/2$  and re-arranging, we can define the effective area of an antenna, where  $\eta_A$  is the aperture efficiency, and  $A_{geom}$  is the geometric area:

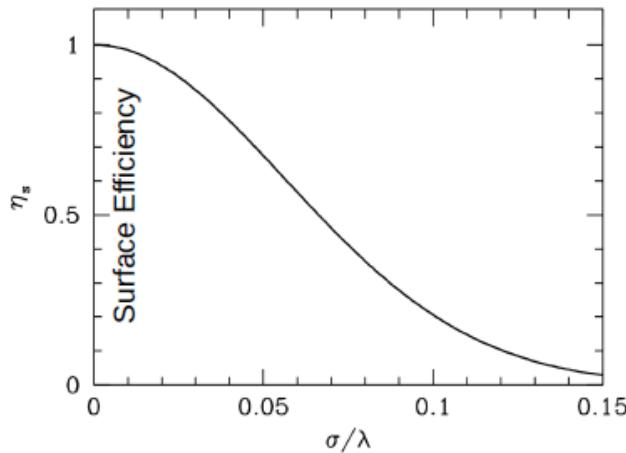
$$A_e = \frac{2P_\nu}{S_\nu} \sim \eta_A A_{geom} \quad (8.10)$$

Astonishingly, for all antennas, the average effective area is:

$$\bar{A}_e = \frac{\lambda^2}{4\pi} \quad (8.11)$$

that really only usefully applies to near omni-directional antennas like dipoles. For telescopes with large apertures  $A_0$ , and therefore strong directionality,  $\Omega_A \propto (\lambda/D)^2$  and then:

$$A_0 \Omega_A = \lambda^2 \quad (8.12)$$

Figure 8.1:  $\eta_s$  as function of  $\sigma/\lambda$ .

in which  $\Omega_A$  is called the *beam solid angle* of the telescope. ATTENTION: the dishes are NOT the antennas, they are only collectors. The antenna is the dipole system present at the focus to receive the signal.  $A_{effe}$  is often written ad the gain of the telescope, measured in  $K/Jy$ . For example, a gain of 1  $K/Jy$  telescope therefore implies an effective area of:

$$A_e = \frac{2kT_A}{S} = 2761m^2 \quad (8.13)$$

### 8.1.1 What affects aperture efficiency $\eta_A$ ?

There are different aspects that can affects aperture efficiency  $\eta_A$ .

First of all, how well the feed “illuminates” collecting area. Feed horns are typically designed to have a pseudo-gaussian illumination pattern of the dish or other reflector. For a good design,  $\eta_A$  can be 70%. For phased array systems,  $\eta_A$  can be 80 + %

Secondary, how good the surface is at wavelength  $\lambda$ . Surface efficiency is defined by “Ruze” Equation in which  $\sigma$  is the size of the RMS surface error:

$$\eta_s = \exp -((4\pi\sigma)/(\lambda))^2 \quad (8.14)$$

As seen in figure 8.1, surface efficiency  $\eta_s$  decreases as  $\sigma/\lambda$  increases.

It is also possible to define a **system equivalent Flux Density**. It is the flux density that would generate the amount of power per unit bandwidth that we see at the output of the system:

$$SEFD = \frac{T_{sys}}{G} \quad (8.15)$$

In general, smaller it is, better it is and note that it is  $\nu$  dependent.

## 8.2 Radiometers

Antennas produce noise in the form of voltages. Those voltages have mean of 0 and are hard to average to a measurable value. Radiometers “detect” a signal, typically by squaring it (a so called “square-law” detector), so that it can be measured and/or integrated.

The sequence is the following.

1. Band-limited signal voltages enter.

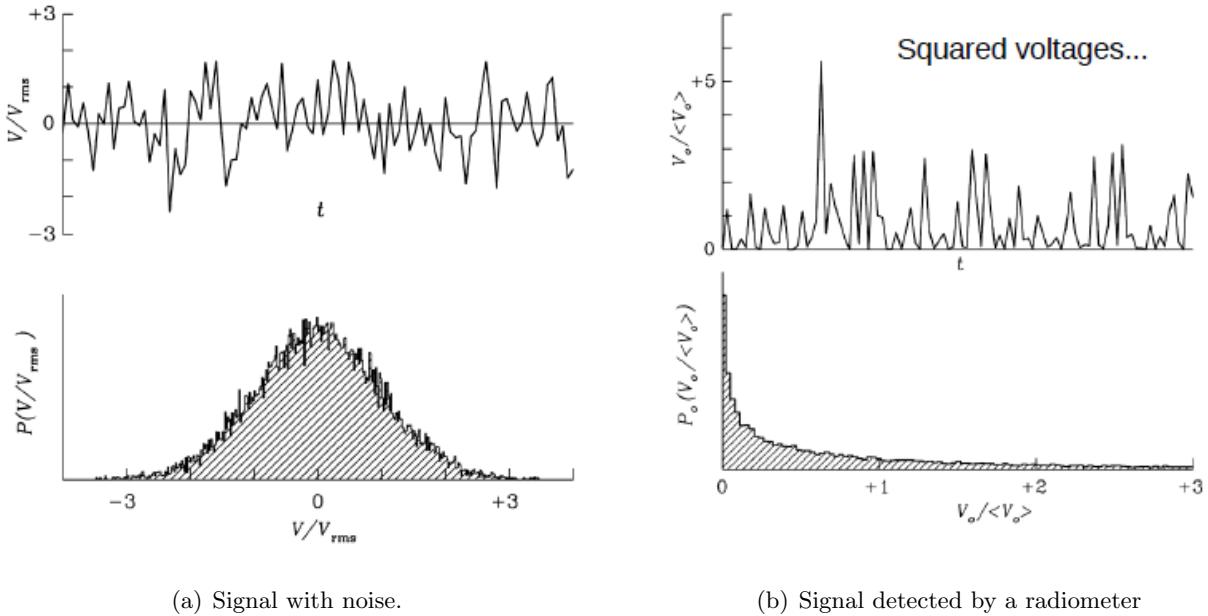


Figure 8.2: Signal detected.

2. Nyquist-sampled, it has  $N = 2\Delta\nu$  samples per second (i.e. sample rate is twice the bandwidth).
3. Square-law detector squares voltages.
4. Integrator averages them. Becomes more gaussian with time via central limit theorem.
5. Standard deviation goes down as  $N^{1/2}$ .

### 8.2.1 Ideal radiometer equation

For band-limited signal in bandwidth  $\Delta\nu$  and integrated for time  $\tau$ , the RMS error on the measured noise temperature of a signal (i.e.  $T_{sys}$ ) is:

$$\sigma_T \sim \frac{T_s}{\sqrt{\Delta\nu\tau}} \quad (8.16)$$

Note that  $\Delta\nu\tau$  is often very large,  $10^8$  or more.

## 8.3 System temperature

Noise powers are represented by temperatures ( $P = kT$ ), which add linearly:

$$T_s = T_{cmb} + T_{rsb} + \Delta T_{source} + [1 - \exp -\tau_A] T_{atm} + T_{spill} + T_r + \dots \quad (8.17)$$

in which:

- $\Delta T_{source}$  is from the astrophysical source, and is usually tiny ( $\ll T_{sys}$ );
- $T_{cmb}$  is the  $2.7\text{ K}$  cosmic microwave background;
- $T_r$  is the receiver temperature, and is typically  $10 - 20\text{ K}$ ;
- $T_{rsb}$  is the radio source background, including Galactic synchrotron and the integrated extra-galactic radio background. For frequencies  $< 1\text{ GHz}$ , our Galaxy can dominate  $T_{sys}$ . At  $\sim 1\text{ GHz}$ , typical  $T_{gal}$  is a couple  $K$ . For example, at  $400\text{ MHz}$ , it can be  $50 - 300\text{ K}$ . It scales with  $\nu^{-2.7}$ ;

- $T_{spill}$  refers to imperfect illumination of dish or subreflector picks up background power. Heavily dependent on optics design and shape of illumination from feed (i.e. “taper”). Typical values are 10 K;
- for  $T_{atm}$  see previous chapter.

## 8.4 Surface brightness sensitivity

Too much spatial resolution from interferometers means that extended sources bigger than the synthesized beam ( $\sim \lambda/b$ , where  $b$  is max baseline) lose flux. This is in addition to the high-pass spatial filtering of interferometry. Loss in sensitivity is proportional to area filling factor, which is  $\sim (D/b)^2$ , where  $D$  is dish diam.

ATTENTION: Single dish telescopes are much better at measuring extended sources than interferometers (100s – 1000s of times better) so don’t use more resolution than you need!

## 8.5 Radio frequency interference

At  $cm$  and  $m$  wavelengths, fractional observing bandwidths are increasing and use of the spectrum is increasing as well interference (meaning legal use of the spectrum!) can be extremely strong, highly variable in time, and narrow or broadband. It can cause system to behave non-linearly and can dominate over rest of  $T_{sys}$ .

# Chapter 9

## Main scientific goals of VLTI and ALMA

### 9.1 Very Large Telescope Interferometer (VLTI)

The Very Large Telescope Interferometer (VLTI) consists in the coherent combination of the four VLT Unit Telescopes or the four moveable 1.8 *m* Auxiliary Telescopes. The VLTI provides milli-arcsec angular resolution at low and intermediate ( $R = 5000$ ) spectral resolution at near and mid-infrared wavelengths.

With the instruments PIONIER and GRAVITY in the near-infrared and MATISSE in the mid-infrared, valuable data can be produced for a broad range of research areas which are the following.

- The structure and composition of the outer solar system.
- The mass function of low mass stars, brown dwarfs (BD) and planets.
- The detection of extra-solar planets.
- The formation mechanism of stars and planetary systems.
- The surface structure of stars.
- The accurate distance to galactic Cepheids, the Large Magellanic Cloud and globular clusters.
- The physical mechanisms responsible for stellar pulsation, mass loss and dust formation in stellar envelopes and evolution to the Planetary Nebula and White Dwarf stages.
- The structure and evolution of stellar and galactic nuclear accretion disks and associated features (jets, dust torii, NarrowLine Regions, Broad Line Regions etc).
- The nature of the Milky Way nucleus surrounding the central black hole (BH).
- Interacting binary evolution and mass transfer mechanisms.
- The structure of the circum-stellar environment of stellar BH and neutron stars.
- The evolution of the expanding shells of novae and supernovae and their interaction with the interstellar medium and its chemical enrichment.

#### 9.1.1 Extra-solar Planets

The search for extra-solar planets and their more precise description and characterisation constitute an area of major interest not only for scientists but also for a broader public. In the border zone between real stars, brown dwarfs and planets the VLTI has the ability to substantially enlarge the sample of stars searched for planets. The VLTI PRIMA facility will be able to perform high precision

astrometry and direct imaging with high spectral resolution of the planets and the gaseous and dust disks out of which they are most likely to form. These techniques combined to the radial velocity method will considerably improve the precision of the planets mass determination.

### 9.1.2 Active galactic nuclei (AGN)

Active galactic nuclei (AGN) are one of the most energetic and mysterious phenomena in the universe. In some galaxies indeed, the core generates huge amounts of energy which are orders of magnitudes higher than for normal galaxies, such as the Milky Way. AGNs are thought to be powered by accretion onto a massive black hole.

The VLTI is very well suited to study objects far away from our galaxy. With its high sensitivity to thermal radiation, MIDI is ideally adapted to study heated cosmic material around a radiating nucleus. The ultraviolet and optical radiation from the hot material surrounding the black hole indeed heats the dust torus to several hundred degrees. The absorbed energy is then re-radiated in the thermal infrared between 5 and 100 microns.

The MIDI instrument on the VLTI is thus a quite appropriate instrument to peer at the enigmatic dust and gas tori believed to be located around giant black holes at the centres of quasars and Active Galactic Nuclei.

### 9.1.3 AGB, Post-AGB stars and PNe

At the late stage of stellar evolution, stars return most of their mass to the interstellar space before they end their life in supernova explosions for massive stars ( $M > 8 M_{\text{solar}}$ ) or as planetary nebulae (PNe) in the case of low- and intermediate-mass stars. While mass loss plays an important role not only in stellar evolution but also in the chemical evolution of the Galaxy, the mass loss mechanism of evolved stars is still not completely understood. In the circumstellar environment of cool evolved stars such as asymptotic giant branch (AGB) stars or red supergiants (RSGs), complicated physical and chemical processes take place, being mutually coupled: stellar pulsation, molecule and dust formation, chromospheric heating, and acceleration of mass outflows. For a better understanding of the mass loss mechanism in evolved stars, it is essential to obtain information on the physical properties of the region between the photosphere and the expanding dust shell – exactly the region where mass outflows are expected to be initiated.

Interferometric observations with high-spatial resolution combined with spectroscopy (using the VLTI MID-infrared Interferometric instrument MIDI ), turn out to be particularly suitable for investigations in these domains.

## 9.2 Atacama Large Millimeter/submillimeter Array (ALMA)

The Atacama Large Millimeter/submillimeter Array (ALMA) is an astronomical interferometer of 66 radio telescopes in the Atacama Desert of northern Chile, which observe electromagnetic radiation at millimeter and submillimeter wavelengths. The array has been constructed on the 5000 m elevation Chajnantor plateau - near the Llano de Chajnantor Observatory and the Atacama Pathfinder Experiment.

These capabilities will enable ALMA to achieve the following level one science goals.

- The ability to detect spectral line emission from  $CO$  or  $C+$  in a normal galaxy like the Milky Way at a redshift of  $z = 3$ .
- The ability to image the gas kinematics in a solar-mass protostellar/ protoplanetary disk at a distance of 150 pc, enabling one to study the physical, chemical, and magnetic field structure of the disk and to detect the tidal gaps created by planets undergoing formation.

- The ability to provide precise images at an angular resolution of  $0.1''$ . Here the term "precise image" means an accurate representation of the sky brightness at all points where the brightness is greater than 0.1% of the peak image brightness. This requirement applies to all sources visible to ALMA that transit at an elevation greater than 20 degrees.

ALMA will be a complete astronomical imaging and spectroscopic instrument for the millimeter-/submillimeter regime, providing scientists with capabilities and wavelength coverage that complement those of other research facilities of its era, such as the Expanded Very Large Array (EVLA), James Webb Space Telescope (JWST), Thirty Metre Telescope (TMT), European Extremely Large Telescope (E-ELT), and Square Kilometer Array (SKA). Specifically, ALMA will fill in a crucial scientific gap by providing a sensitive, high-resolution probe of the cold gas and dust properties of star-forming region in our Galaxy and other galaxies out of high-redshift, and in protoplanetary disks. These region are obscured at shorter wavelengths, thus ALMA complements shorter wavelengths observations by providing a complete picture of these cold regions in which stars and planets are formed.

# Chapter 10

## Thermal and nonthermal radiation

**First classification** In radio astronomy it is possible to disentangle between two different families of sources: **galactic sources**, concentrated towards the galactic plane and **extragalactic sources** distributed more or less uniformly in space. The unresolved, spatially continuous radiation belongs to the galactic component. In addition, there is the  $2.7\text{ K}$  thermal background radiation which is cosmological in origin.

While the flux density of one type of source is roughly constant with increasing frequency, the other type is more intense at lower frequencies. Those sources which show an increasing flux density with increasing frequency could be identified with objects well known from the optical range of the spectrum. Both the moon and the sun are radio sources of this kind. The moon is an example of a black body and its spectrum is an almost exact representation of the Rayleigh-Jeans law for a temperature of  $T \sim 225\text{ K}$ .

**Second classification** There is another possible classification of radio sources. Indeed, radio source can be classified into two categories: those which radiate by **thermal mechanisms** and the others, which radiate by **nonthermal processes**. In principle many different radiation mechanisms could be responsible for nonthermal emission, but in practice one single mechanism seems to dominate: *synchrotron emission* or *magnetic bremsstrahlung*.

**Are the two classifications independent?** The other division of the discrete radio sources into galactic and extragalactic ones, is in principle completely independent of this classification. However, we find predominantly nonthermal sources among the extragalactic sources. This is simply a result of the fact that the most intense emitters are nonthermal in origin. Even if thermal sources are abundant in extragalactic objects, these will not easily be detected.

Thermal means that the radiation is dependent solely on the temperature of the emitter. Non-thermal radiation involves other processes. In this chapter we are going to look at two specific examples of continuum radiation: thermal blackbody radiation and non-thermal synchrotron radiation.

At short radio wavelengths, thermal emission sources dominate the sky, while non-thermal process dominate at long radio wavelengths. As we shall see, the shape of the spectrum of thermal and non-thermal radiation differs, making it easy to determine the emission mechanism of a source. Other types of radio thermal radiation include free-free radiation, and other types of non-thermal radio radiation include maser emission. Note that there are many of types of thermal and non-thermal radiation mechanisms that do not emit radio photons, such as thermal Compton scattering and non-thermal Bremsstrahlung, both of which emit X-ray photons.

## 10.1 Thermal radiation mechanisms

### 10.1.1 Black-body radiation and its spectrum

Black-body radiation is the thermal electromagnetic radiation within, or surrounding, a body in thermodynamic equilibrium with its environment, emitted by a black body (an idealized opaque, non-reflective body). It has a specific, continuous, spectrum of wavelengths, inversely related to intensity, that depend only on the body's temperature, which is assumed, for the sake of calculations and theory, to be uniform and constant.

Black-body radiation has a characteristic, continuous frequency spectrum that depends only on the body's temperature, called the Planck spectrum or Planck's law. The spectrum is peaked at a characteristic frequency that shifts to higher frequencies with increasing the temperature. We talked about Planck spectrum in chapter 5, see this one.

All normal (baryonic) matter emits electromagnetic radiation when it has a temperature above absolute zero. The radiation represents a conversion of a body's internal energy into electromagnetic energy, and is therefore called thermal radiation. It is a spontaneous process of radiative distribution of entropy.

Conversely, all normal matter absorbs electromagnetic radiation to some degree. An object that absorbs all radiation falling on it, at all wavelengths, is called a black body. When a black body is at a uniform temperature, its emission has a characteristic frequency distribution that depends on the temperature. Its emission is called black-body radiation.

The concept of the black body is an idealization, as perfect black bodies do not exist in nature.

**Sources of BB radiation** Of particular importance, although planets and stars (including our own earth and sun) are neither in thermal equilibrium with their surroundings nor perfect black bodies, black-body radiation is a good first approximation for the energy they emit. The sun's radiation, after being filtered by the earth's atmosphere, thus characterises "daylight", which humans (also most other animals) have evolved to use for vision.

## 10.2 Nonthermal radiation mechanisms

While thermal emission depends on the temperature of the emitting source, non-thermal emission depends on other things, such as the relative proportions of excited states of atoms and magnetic field strength.

### 10.2.1 Synchrotron radiation and its spectrum

Synchrotron radiation is the electromagnetic radiation emitted when relativistic charged particles are subject to an acceleration perpendicular to their velocity. It is produced artificially in some types of particle accelerators, or naturally by fast electrons moving through magnetic fields. The radiation produced in this way has a characteristic polarization and the frequencies generated can range over a large portion of the electromagnetic spectrum.

Synchrotron radiation is similar to bremsstrahlung radiation, which is emitted by a charged particle when the acceleration is parallel to the direction of motion. The general term for radiation emitted by particles in a magnetic field is gyromagnetic radiation, for which synchrotron radiation is the ultra-relativistic special case. Radiation emitted by charged particles moving non-relativistically in a magnetic field is called cyclotron emission. For particles in the mildly relativistic range ( $\sim 85\%$  of the speed of light), the emission is termed gyro-synchrotron radiation.

**Description** A direct consequence of Maxwell's equations is that accelerated charged particles always emits electromagnetic radiation. Synchrotron radiation is the special case of charged particles

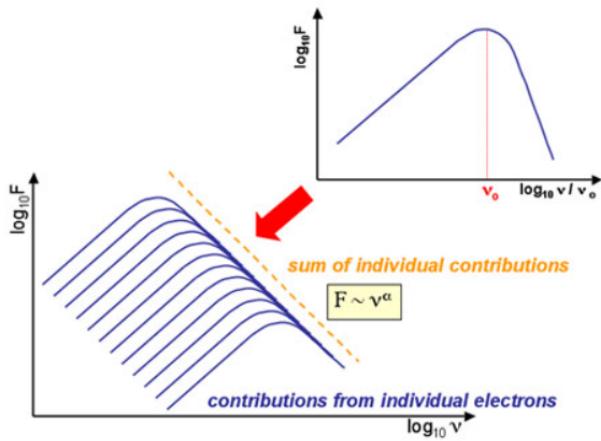


Figure 10.1: The synchrotron emission spectrum follows a power-law decay, and is constructed by adding the contributions from individual electrons.

moving at relativistic speed undergoing acceleration perpendicular to their direction of motion, typically in a magnetic field. In such a field, the force due to the field is always perpendicular to both the direction of motion and to the direction of field, as shown by the Lorentz force law.

The power carried by the radiation is found by the relativistic Larmor formula:

$$P_\gamma = \frac{1}{6\pi\epsilon_0} \frac{q^2 a^2}{c^3} \gamma^4 \quad (10.1)$$

where  $\epsilon_0$  is the vacuum permittivity,  $q$  is the particle charge,  $a$  is the magnitude of the acceleration,  $c$  the speed of light and  $\gamma$  is the Lorentz factor. When the radiation is emitted by a particle moving in a plane, the radiation is linearly polarized when observed in that plane, and circularly polarized when observed at a small angle.

The spectrum of synchrotron emission results from summing the emission spectra of individual electrons. As the electron spirals around the magnetic field, it emits radiation over a range of frequencies peaking at  $\nu_0$ , the critical frequency. The longer the electron travels around the magnetic field, the more energy it loses, the narrower the spiral it makes, and the longer the wavelength of the critical frequency. See figure 10.1.

### Source of synchrotron radiation

- In astrophysics, synchrotron emission occurs, for instance, due to ultra-relativistic motion of a charged particle around a black hole. When the source follows a circular geodesic around the black hole, the synchrotron radiation occurs for orbits close to the photosphere where the motion is in the ultra-relativistic regime.
- A class of astronomical sources where synchrotron emission is important is pulsar wind nebulae, also known as plerions, of which the Crab nebula and its associated pulsar are archetypal. Pulsed emission gamma-ray radiation from the Crab has recently been observed up to  $> 25$  GeV, probably due to synchrotron emission by electrons trapped in the strong magnetic field around the pulsar. Polarization in the Crab nebula at energies from 0.1 to 1.0 MeV, illustrates this typical property of synchrotron radiation.
- Much of what is known about the magnetic environment of the interstellar medium and intergalactic medium is derived from observations of synchrotron radiation. Cosmic ray electrons moving through the medium interact with relativistic plasma and emit synchrotron radiation which is detected on Earth.

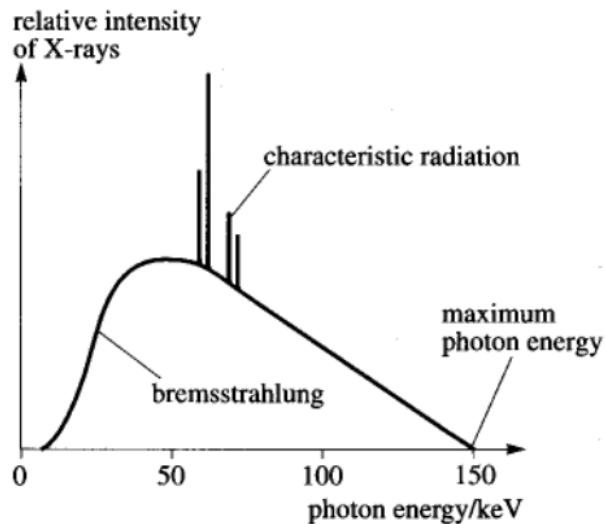


Figure 10.2: Bremsstrahlung spectrum.

### 10.2.2 Bremsstrahlung radiation and its spectrum

Bremsstrahlung is electromagnetic radiation produced by the deceleration of a charged particle when deflected by another charged particle, typically an electron by an atomic nucleus. The moving particle loses kinetic energy, which is converted into radiation (i.e., photons), thus satisfying the law of conservation of energy. The term is also used to refer to the process of producing the radiation. Bremsstrahlung has a continuous spectrum, which becomes more intense and whose peak intensity shifts toward higher frequencies as the change of the energy of the decelerated particles increases. Bremsstrahlung emitted from plasma is sometimes referred to as free-free radiation.

**Description** According to Maxwell's equations, accelerated charges emit electromagnetic radiation. In particular, when an electron hits a material, it is subjected to a scattering by the coulomb field of an atomic nucleus, so it can be thought that it is "braked". If the energy of the bombarding electrons is high enough, the emitted radiation lies in the X-ray region of the electromagnetic spectrum.

The energy loss for bremsstrahlung is significant – that is, over the ionization and nucleus excitation processes – for highly energy electrons (in the order of hundreds of *MeV* in air and water, and tens of *MeV* in heavy materials such as lead or iron). The average energy loss per length unit can be roughly calculated with the following:

$$-\frac{dE}{dx} \sim \frac{4N_a Z^2 \alpha^3 h^2 c^2}{m_e^2 c^4} \ln(183/Z^{1/3}) E \quad (10.2)$$

Where  $N_a$  is the number of atoms per volume unit,  $Z$  is the atomic number of the target material,  $\alpha$  is the fine structure constant and  $m_e$  is the electron mass. It is therefore clear that the loss of energy is proportional to  $Z^2$ , to the energy of particle  $E$  and inversely proportional to the mass of the particle. For particles heavier than the electron the bremsstrahlung radiation is negligible. The logarithmic term is due to the partial shielding of nuclear charge by atomic electrons.

This continuous spectrum overlaps even single strong lines, as bombarding electrons can expel electrons from the most internal atomic shells of the target, and the rapid filling of these gaps by other electrons of the upper layers produces characteristic X-ray for each atom (called "X ray fluorescence"). Alternatively, the energy difference between the two orbits will result in the further expulsion of electrons. This phenomenon is the Auger effect. See figure 10.2.

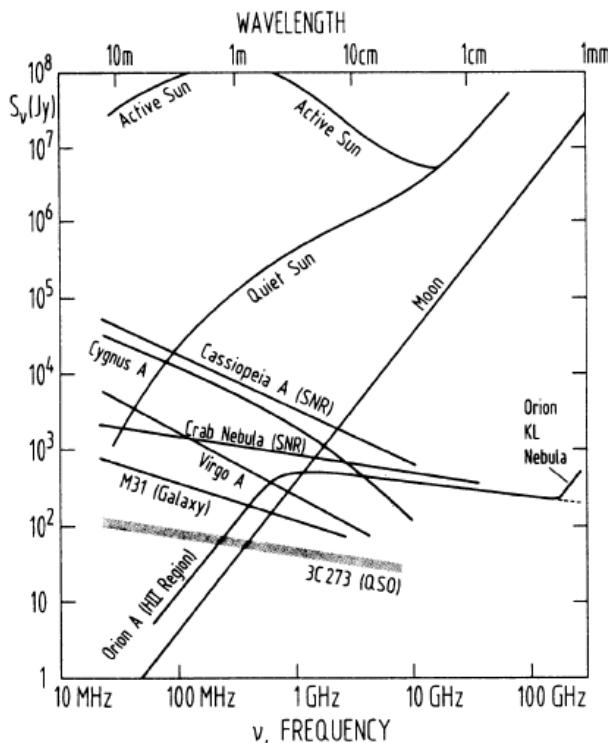


Figure 10.3: Different types of spectrum.

**Sources of bremsstrahlung radiation** All rich clusters of galaxies produce extended X-ray emission due to thermal bremsstrahlung radiation from a hot intracluster gas.

Moreover, HII regions are ionized phases of the Interstellar Medium that emits bremsstrahlung. These types of regions are localized around O and B stars, able to emit photo-ionization radiation, given their high temperature.

### 10.3 Summery

See figure 10.3. The spectral distributions of various radio sources. The Moon, the quiet Sun and (at lower frequencies) the H II region Orion A are examples of Black Bodies. Close to 300 GHz there is additional emission from dust in the molecular cloud Orion KL. The active Sun, supernova remnants such as Cassiopeia A, the radio galaxies Cygnus A, Virgo A (Messier 87, 3C274) and the Quasi Stellar radio source (QSO) 3C273 are nonthermal emitters. The hatching around the spectrum of 3C273 is meant to indicate rapid time variability.

# Appendix A

## Polarization of EM waves

### A.1 Vector waves

In general both the  $x$  and the  $y$  component have to be specified but, in a strictly monochromatic wave, they are not independent, since both share the same harmonic dependence, although with a different phase:

$$\begin{aligned} E_x &= E_1 \cos kz - \omega t + \delta_1 \\ E_y &= E_2 \cos kz - \omega t + \delta_2 \\ E_z &= 0 \end{aligned} \quad (\text{A.1})$$

Here  $k = 2\pi/\lambda$ , where  $\lambda$  is the wavelength in  $cm$ , and  $\omega = 2\pi\nu$ , where  $\nu$  is frequency in  $Hz$ . Regarding  $(E_x, E_y, z)$  as the coordinates of a point in a rectangular coordinate system we find that A.1 describes a helical path on the surface of a **cylinder**. The cross section of this cylinder can be determined by eliminating the phase of this wave, abbreviated by  $\tau = kz - \omega t$ . Rewriting the first two equations of A.1 as:

$$\begin{aligned} E_x/E_1 &= \cos \tau \cos \delta_1 - \sin \tau \sin \delta_1 \\ E_y/E_2 &= \cos \tau \cos \delta_2 - \sin \tau \sin \delta_2 \end{aligned} \quad (\text{A.2})$$

gives:

$$\begin{aligned} \frac{E_x}{E_1} \sin \delta_2 - \frac{E_y}{E_2} \sin \delta_1 &= \cos \tau \sin \delta_2 - \delta_1 \\ \frac{E_x}{E_1} \cos \delta_2 - \frac{E_y}{E_2} \cos \delta_1 &= \sin \tau \sin \delta_2 - \delta_1 \end{aligned} \quad (\text{A.3})$$

Squaring and adding we obtain:

$$\left(\frac{E_x}{E_1}\right)^2 + \left(\frac{E_y}{E_2}\right)^2 - 2\frac{E_x}{E_1}\frac{E_y}{E_2} \cos \delta = \sin \delta_1 - \delta_2^2 = \sin \delta^2 \quad (\text{A.4})$$

This is the equation of an ellipse, since the discriminant is not negative. The wave is said to be **elliptically polarized**, and this applies to both the electric and the magnetic field of the wave; *sin δ* determines the sense in which the electric vector rotates. The ellipse A.4 usually is arbitrarily oriented

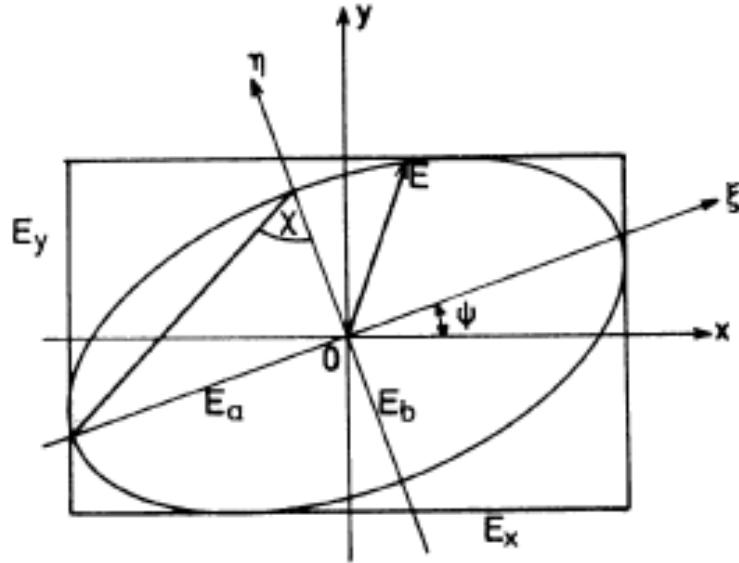


Figure A.1: The polarization ellipse for the electric vector,  $E$ , of an elliptically polarized wave.

with respect to the coordinate system. Its geometric properties are seen best by selecting a coordinate system oriented along the major and minor axes (figure A.1). In this system the ellipse equation is:

$$\begin{aligned} E_\xi &= E_a \cos \tau + \delta \\ E_\eta &= E_b \sin \tau + \delta \end{aligned} \quad (\text{A.5})$$

and the relation between the coordinate systems  $(x, y)$  and  $(\xi, \eta)$  is given by the linear transformation:

$$\begin{aligned} E_\xi &= E_x \cos \Psi + E_y \sin \Psi \\ E_\eta &= -E_x \sin \Psi + E_y \cos \Psi \end{aligned} \quad (\text{A.6})$$

The intrinsic parameters of the polarization ellipse  $E_a$  and  $E_b$ , as well as the angle  $\Psi$  by which the major axis is tilted with respect to the  $x$  axis, can then be determined by requiring that A.4 transformed by A.6 should lead to equation A.5. Substituting A.2 and A.5 into A.6 while simultaneously expanding the  $\cos(\tau + \delta)$  term leads to:

$$E_a(\cos \tau \cos \delta - \sin \tau \sin \delta) = E_1(\cos \tau \cos \delta_1 - \sin \tau \sin \delta_1) \cos \Psi + E_2(\cos \tau \cos \delta_2 - \sin \tau \sin \delta_2) \sin \Psi \quad (\text{A.7})$$

and:

$$E_b(\sin \tau \cos \delta + \cos \tau \sin \delta) = -E_1(\cos \tau \cos \delta_1 - \sin \tau \sin \delta_1) \sin \Psi + E_2(\cos \tau \cos \delta_2 - \sin \tau \sin \delta_2) \cos \Psi \quad (\text{A.8})$$

These equations are valid for all  $\tau$ , i.e. also for  $\tau = 0$  and  $\tau = \pi/2$ , resulting in:

$$E_a \cos \delta = E_1 \cos \delta_1 \cos \Psi + E_2 \cos \delta_2 \sin \Psi \quad (\text{A.9})$$

$$-E_a \sin \delta = -E_1 \sin \delta_1 \cos \Psi - E_2 \sin \delta_2 \sin \Psi \quad (\text{A.10})$$

$$E_b \cos \delta = E_1 \sin \delta_1 \sin \Psi - E_2 \sin \delta_2 \cos \Psi \quad (\text{A.11})$$

$$E_b \sin \delta = -E_1 \cos \delta_1 \sin \Psi + E_2 \cos \delta_2 \cos \Psi \quad (\text{A.12})$$

Squaring these equations and adding we obtain:

$$S_0 = E_a^2 + E_b^2 = E_1^2 + E_2^2 \quad (\text{A.13})$$

we find that this can be interpreted that the **total Poynting flux** of the polarized wave is equal to the sum of the fluxes of two orthogonal, but otherwise arbitrary directions.

Multiplying A.9 by A.11 and A.10 by A.12 and subtracting the results, we obtain:

$$E_a E_b = E_1 E_2 \sin \delta \quad (\text{A.14})$$

while division and addition of the same pairs of equations result in:

$$\begin{aligned} -(E_1^2 - E_2^2) \sin \Psi \cos \Psi &= E_1 E_2 \cos \delta (\sin^2 \Psi - \cos^2 \Psi) \\ (E_1^2 - E_2^2) \sin 2\Psi &= 2E_1 E_2 \cos \delta \cos 2\Psi \end{aligned} \quad (\text{A.15})$$

If we now define  $\alpha$  by:

$$\frac{E_1}{E_2} = \tan \alpha \quad (\text{A.16})$$

A.15 can be rewritten as:

$$\tan 2\Psi = -\tan 2\alpha \cos \delta \quad (\text{A.17})$$

Defining:

$$\frac{E_a}{E_b} = \tan \chi \quad (\text{A.18})$$

then A.17 is equivalent to:

$$\sin 2\chi = \sin 2\alpha \sin \delta \quad (\text{A.19})$$

Equations (A.13, A.16, A.17, A.18 and A.19) now permit the computation of all intrinsic polarization properties of the elliptically polarized wave from the intensities specified in an arbitrary coordinate system. Values for  $E_1$ ,  $E_2$  and  $\delta$  give  $S_0$  the total intensity, while A.17 combined with A.16 allows the determination of the angle  $\Psi$ , while the angle  $\chi$  is determined from A.19.  $E_a$  and  $E_b$  can be computed from A.18 and A.13.

The phase difference  $\delta$  is important in several respects. Its sign determines the sense in which the wave vector is rotating. If  $\sin \delta > 0$  or equivalently  $\tan \chi > 0$ , the polarization is called **right-handed**; conversely  $\sin \delta < 0$  or  $\tan \chi < 0$  describes **left-handed** elliptical polarization. For right-handed polarization, the rotation of the  $E$  vector and the direction of propagation form a right-handed screw. According to this definition, right-handed helical beam antennas radiate or receive right-circular polarization, a result which is easy to remember.

If the phase difference is:

$$\delta = \delta_1 - \delta_2 = m\pi, m = 0, \pm 1, \pm 2, \dots \quad (\text{A.20})$$

the polarization ellipse degenerates into a straight line and  $E$  is *linearly polarized*. As we have seen, an elliptically polarized wave can be regarded as the superposition of two orthogonal linearly polarized waves.

Another important special case is that of a *circularly polarized* wave. For this:

$$E_1 = E_2 = E \quad (\text{A.21})$$

and:

$$\delta = \frac{\pi}{2}(1 + m), m = 0, \pm 1, \pm 2, \pm 3, \dots \quad (\text{A.22})$$

so that:

$$E_x^2 + E_y^2 = E \quad (\text{A.23})$$

with the orthogonal linear components:

$$\begin{aligned} E_x &= E \cos \tau \\ E_y &= \pm E \cos \tau - \frac{\pi}{2} \end{aligned} \quad (\text{A.24})$$

From this we see that an arbitrary elliptically polarized wave can be decomposed into the sum of two circularly polarized waves:

$$\begin{aligned} E_r &= \frac{1}{2}(E_a + E_b) \\ E_l &= \frac{1}{2}(E_a - E_b) \end{aligned} \quad (\text{A.25})$$

and, for the total Poynting flux of the wave, we obtain:

$$S_0 = E_a^2 + E_b^2 = E_r^2 + E_l^2 \quad (\text{A.26})$$

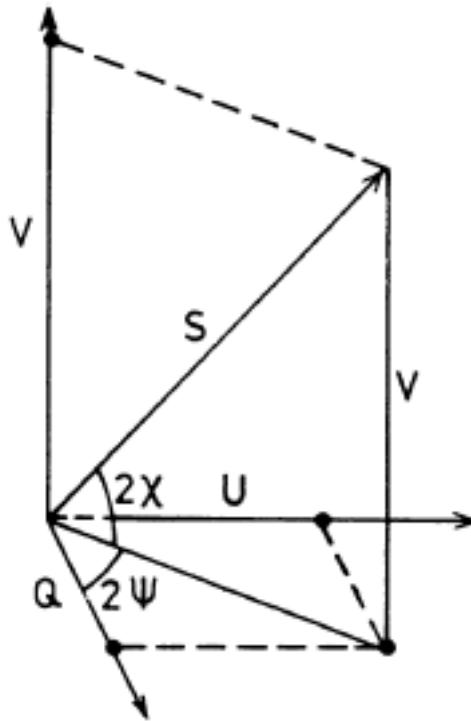


Figure A.2: A sketch which illustrates the definition of the Stokes parameters.

## A.2 The Poincarè Sphere and the Stokes Parameters

The results of the preceding section show that three independent parameters are needed to describe the state of the polarization of a monochromatic vector wave. For this we have introduced several sets of parameters:

- the amplitudes  $E_1$ ,  $E_2$  and the relative phase  $\delta$  of two orthogonal, linearly polarized waves;
- the amplitudes  $E_r$  and  $E_l$ , and the relative phase  $\delta$  of a right- and a left-hand circularly polarized wave;
- the major and minor axis  $E_a$ ,  $E_b$  and the position angle  $\Psi$  of the polarization ellipse.

Poincarè introduced another representation that permits an easy visualization of all the different states of polarization of a vector wave. If we interpret the angles  $2\Psi$  of A.17 and  $2\chi$  of A.19 as longitude and latitude on a sphere with the radius  $S_0$  of A.13 there is a one-to-one relation between polarization states and points on the sphere (figure A.2). The equator represents linear polarization; the north pole corresponds to right-circular and the south pole to left-circular polarization (figure A.3).

There is a natural relation between the Poincarè sphere and the Stokes parameters. These are the Cartesian coordinates of the points on the sphere with the definitions:

$$\begin{aligned} S_0 &= I = E_a^2 + E_b^2 \\ S_1 &= Q = S_0 \cos 2\chi \cos 2\Psi \\ S_2 &= U = S_0 \cos 2\chi \sin 2\Psi \\ S_3 &= V = S_0 \sin 2\chi \end{aligned} \tag{A.27}$$

Only three of these parameters are independent, since according to the construction of the Poincarè sphere:

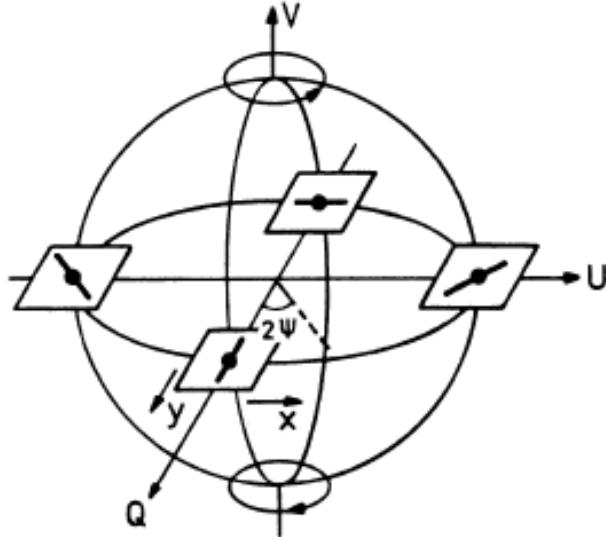


Figure A.3: Polarization and the Poincaré sphere. Considering the angles  $2\Psi$  and  $2\chi$  as angles in a polar coordinate system, each point on the surface of the resulting sphere corresponds to a unique state of polarization. The positions on the equator ( $2\chi = 0$ ) correspond to linear polarization, those at the northern latitudes ( $2\chi > 0$ ) contain right-handed circular polarization, while those on the southern hemisphere contain left-handed. If we orient the  $(x, y)$  coordinate system parallel to  $Q$  and  $U$ , the linear polarization of the waves are oriented as indicated.

$$\begin{aligned} S_0^2 &= S_1^2 + S_2^2 + S_3^2 \\ I^2 &= Q^2 + U^2 + V^2 \end{aligned} \quad (\text{A.28})$$

The Stokes parameters can also be directly expressed by the parameters of the polarization ellipse:

$$\begin{aligned} S_0 &= I = E_1^2 + E_2^2 \\ S_1 &= Q = E_1^2 - E_2^2 \\ S_2 &= U = 2E_1E_2 \cos \delta \\ S_3 &= V = 2E_1E_2 \sin \delta \end{aligned} \quad (\text{A.29})$$

These equations permit us to express the Stokes parameters directly in terms of observable quantities. A few special cases will illustrate the principle.

- For a *right-handed circularly* polarized wave we have  $E_1 = E_2$  and  $\delta = \pi/2$ , so that:

$$\begin{aligned} S_0 &= I = S \\ S_1 &= Q = 0 \\ S_2 &= U = 0 \\ S_3 &= V = S \end{aligned} \quad (\text{A.30})$$

- For a *left-handed circularly* polarized wave we have:

$$\begin{aligned} S_0 &= I = S \\ S_1 &= Q = 0 \\ S_2 &= U = 0 \\ S_3 &= V = -S \end{aligned} \quad (\text{A.31})$$

- For a *linearly* polarized wave we have  $E_b = E$  and  $E_a = 0$ , so that  $\chi = 0$  and:

$$\begin{aligned} S_0 &= I = E_2 = S \\ S_1 &= Q = I \cos 2\Psi \\ S_2 &= U = I \sin 2\Psi \\ S_3 &= V = 0 \end{aligned} \tag{A.32}$$

Finally, one should note that so far we have implied that a strictly monochromatic wave is always polarized; there is no such thing as an unpolarized monochromatic wave. This becomes evident if we remember that for a monochromatic plane harmonic wave,  $E_1, E_2, \delta_1$  and  $\delta_2$  are always constants. This situation will be different when we consider quasi-monochromatic radiation, in which  $\omega$  is restricted to some small but finite bandwidth. Radiation of this kind can be un-polarized or partially polarized. To analyze this, one must have a convenient way to describe such radiation. This will be done in the next section.

### A.3 Quasi-Monochromatic Plane waves

The previous description of the polarization properties applies only to monochromatic waves. Now, let us modify the description for a quasi-monochromatic waves. Both the electric and the magnetic field intensity of the wave at a given fixed position can then be expressed by an integral of the form:

$$V(t) = \int_0^{\infty} a(\nu) e^{i[\phi(\nu) - 2\pi\nu t]} d\nu \tag{A.33}$$

This can be decomposed in the real and imaginary part:

$$\begin{aligned} V^{(r)}(t) &= \int_0^{\infty} a(\nu) \cos(\phi(\nu) - 2\pi\nu t) d\nu \\ V^{(i)}(t) &= \int_0^{\infty} a(\nu) \sin(\phi(\nu) - 2\pi\nu t) d\nu \end{aligned} \tag{A.34}$$

$V^{(i)}$  does not contain information not already contained in  $V(r)$ .  $V$  is referred to as the analytic signal associated with  $V(r)$ . This allows phase to be determined. Frequently  $a(\nu)$  has a form that effectively limits this range to an interval  $\Delta\nu$  which is small compared with the mean frequency  $\bar{\nu}$ ;

$$\Delta\nu/\bar{\nu} \ll 1 \tag{A.35}$$

If this condition is fulfilled, the signal is said to be quasi-monochromatic.

### A.4 The Stokes Parameters for Quasi-Monochromatic Waves

The Stokes parameters of a quasi-monochromatic wave are:

$$\begin{aligned} S_0 &= I = \langle a_1^2 \rangle + \langle a_2^2 \rangle \\ S_1 &= Q = \langle a_1^2 \rangle - \langle a_2^2 \rangle \\ S_2 &= U = 2\langle a_1 a_2 \cos \delta \rangle \\ S_3 &= V = 2\langle a_1 a_2 \sin \delta \rangle \end{aligned} \tag{A.36}$$

and these can be calculated from 6 intensity measurements. On the contrary, in terms of angles, these parameters can be re-written as:

$$\begin{aligned} S_0 &= I = I(0^\circ, 0) + I(90^\circ, 0) \\ S_1 &= Q = I(0^\circ, 0) - I(90^\circ, 0) \\ S_2 &= U = I(45^\circ, 0) - I(135^\circ, 0) \\ S_3 &= V = I(45^\circ, \frac{\pi}{2}) - I(135^\circ, \frac{\pi}{2}) \end{aligned} \quad (\text{A.37})$$

These are the relationships used to analyze the outputs of radio polarimeters. We will return to this later. For partially polarized light we find from A.36:

$$\begin{aligned} S_0^2 &\geq S_1^2 + S_2^2 + S_3^2 \\ I^2 &\geq Q^2 + U^2 + V^2 \end{aligned} \quad (\text{A.38})$$

which is valid for strictly monochromatic waves. It is then easy to express the degree of polarization:

$$p = \frac{\sqrt{S_1^2 + S_2^2 + S_3^2}}{S_0} \quad (\text{A.39})$$

The Stokes parameters of the superposition of several independent vector waves will be the sum of the Stokes parameters of the individual waves.